

Università degli studi di Padova

Dipartimento di Scienze Statistiche
Corso di Laurea Magistrale in
Scienze Statistiche



**MODELLI QUANTILICI DINAMICI PER DATI
SPAZIO-TEMPORALI**

Relatore: Prof. Mauro Bernardi
Dipartimento di Scienze Statistiche

Laureando: Cristian Castiglione
Matricola: 1149656

Anno Accademico 2017/2018

Indice

Introduzione	9
1 Dati e processi spazio-temporali	13
1.1 Problemi ed esempi	13
1.2 Processi spazio-temporali	17
2 Modello spazio-temporale in media	25
2.1 Il modello teorico	26
2.2 Rappresentazione <i>State Space</i>	29
2.3 Estrazione del segnale	30
2.4 Inferenza a posteriori	33
2.5 Previsione nel tempo	36
2.6 Possibili estensioni	37
3 Modelli quantilici spazio-temporali	39
3.1 <i>Quantile regression</i>	39
3.2 <i>Quantile regression</i> Bayesiana	41
3.2.1 Rappresentazione stocastica	42
3.2.2 Inferenza a posteriori	43
3.3 <i>Quantile regression</i> dinamica a disturbi indipendenti	45
3.3.1 Rappresentazione stocastica e modello <i>State Space</i>	45
3.3.2 Inferenza ed estrazione del segnale	46
3.4 <i>Quantile regression</i> dinamica a disturbi correlati	48
3.4.1 Distribuzione Laplace asimmetrica multivariata	48
3.4.2 Rappresentazione stocastica e modello <i>State Space</i>	49
3.4.3 Inferenza ed estrazione del segnale	49
3.5 Previsione dei quantili nel tempo	52
4 Applicazione su Dati Reali	55
4.1 I dati	55

4.2	Analisi esplorative	55
4.3	Modello in media	59
4.4	Modello quantilico a disturbi indipendenti	68
4.5	Modello quantilico a disturbi correlati	73
4.6	Previsione spazio-temporale	78
5	Estensioni e quantili multipli	83
5.1	<i>Multiple quantile regression</i>	84
5.2	Modello quantilico congiunto	85
5.3	<i>Constrained Kalman Filter</i>	88
5.4	Inferenza stocastica MCMC	91
5.5	Inferenza Bayesiana approssimata	92
5.5.1	<i>Variational approximations</i>	92
5.5.2	Inferenza approssimata per modelli <i>State Space</i>	96
5.6	Inferenza Bayesiana o frequentista?	99
A	Appendice matematica	105
A.1	Distribuzioni <i>full conditional</i> per il modello 2.2.1	106
A.2	Distribuzioni <i>full conditional</i> per il modello 3.2.4	107
A.3	Distribuzioni <i>full conditional</i> per il modello 3.3.3	110
A.4	Distribuzioni <i>full conditional</i> per il modello 3.4.4	111
A.5	Distribuzioni <i>full conditional</i> per il modello 5.2.3	112
A.6	Distribuzioni variazionali per il modello 5.2.3	114
B	Codice Python	119
B.1	Algoritmi di filtraggio, lisciamiento e simulazione	119
B.2	Algoritmi per il modello 4.3.4	127
B.3	Algoritmi per il modello 4.4.1	134
B.4	Algoritmi per il modello 4.5.1	141
	Bibliografia	151

Elenco dei codici

B.1	<i>Kalman filter</i>	119
B.2	<i>Kalman disturbance smoother</i>	122
B.3	<i>Disturbance simulation smoother</i>	124
B.4	Matrici della rappresentazione <i>State Space</i> per il modello 4.3.4	127
B.5	Calcolo della matrice <i>kernel</i>	128
B.6	Calcolo della matrice di correlazione spaziale	129
B.7	<i>Full conditional</i> collassata per σ_ε^2 nel modello 4.3.4	130
B.8	Riparametrizzazione <i>Inverse Gamma</i>	131
B.9	<i>Gibbs sampling</i> per il modello 4.3.4	131
B.10	Esempio di simulazione per il modello 4.3.4	133
B.11	Matrici della rappresentazione <i>State Space</i> per il modello 4.4.1	134
B.12	Generatore di valori da <i>Inverse Gaussian</i>	135
B.13	Funzione di perdita quantilica	136
B.14	<i>Full conditional</i> collassata per σ_ε^2 nel modello 4.4.1	136
B.15	<i>Full conditional</i> per il fattore latente ω nel modello 4.4.1	137
B.16	<i>Partially collapsed Gibbs sampling with data augmentation</i> per il modello 4.4.1	138
B.17	Esempio di simulazione per il modello 4.4.1	140
B.18	Matrici della rappresentazione <i>State Space</i> per il modello 4.5.1	141
B.19	Generatore di valori dalla <i>Generalized Inverse Gaussian</i>	142
B.20	<i>Full conditional</i> per σ_ε^2 nel modello 4.5.1	145
B.21	<i>Full conditional</i> per il fattore latente ω nel modello 4.5.1	146
B.22	<i>Gibbs sampling with data augmentation</i> per il modello 4.5.1	147
B.23	Esempio di simulazione per il modello 4.5.1	149

Elenco delle tabelle

1.1	Funzioni di covarianza isotropiche	19
2.1	Stati e varianze di filtraggio e lisciamiento	32
4.1	Stazioni di rilevazione del vento in Irlanda.	56
4.2	Funzioni di covarianza stimate	58
4.3	Stime a posteriori per il modello 4.4.1	68
4.4	Diagnostiche sui quantili stimati per il modello 4.4.1	72
4.5	Stime a posteriori per il modello 4.5.1	75
4.6	Diagnostiche sui quantili stimati per il modello 4.5.1	77

Introduzione

La modellazione di fenomeni spazio-temporali sta assumendo un'importanza sempre più rilevante nel corso degli ultimi decenni, soprattutto in contesti applicativi meteorologici e ambientali: Wikle e Cressie 1999, Wikle, Berliner e Cressie 2004, Sahu e Mardia 2005, Lowe et al. 2011. In questi ambiti tipicamente si è interessati a costruire modelli di previsione dinamica nello spazio e nel tempo; dove con l'aggettivo "dinamico" facciamo riferimento a strumenti statistici in grado di adattarsi e modificarsi nel tempo in base all'evoluzione naturale del fenomeno di interesse osservabile attraverso i dati.

Per affrontare problemi di questa natura un approccio adottabile è quello di combinare modelli per serie storiche multivariate con tecniche di previsione spaziale provenienti dall'ambito geostatistico. Questo tipo di modellistica è nota in letteratura come *Kriged Kalman Filtering* (Mardia et al. 1998), ovvero la combinazione di *Kriging* (Cressie 1993), nota tecnica di previsione e interpolazione spaziale, e di *Kalman Filter* (Durbin e Koopman 2012), celebre algoritmo di estrazione dinamica del segnale per modelli *State Space*.

La complessità dei fenomeni dipendenti da una doppia natura, spaziale e temporale, spesso limita lo spettro di applicabilità dei classici metodi di regressione dinamica basati su assunzioni del primo e secondo ordine, e talvolta su ipotesi distributive parametriche (come il succitato *Kriged Kalman Filter*). Dati geolocalizzati e variabili nel tempo non di rado presentano distribuzioni devianti dalla normalità, caratterizzate, ad esempio, da forme di asimmetria, curtosi ed eteroschedasticità.

Per superare le restrizioni imposte dalle assunzioni usuali, si può andare a caratterizzare la distribuzione di un processo mediante tecniche di *quantile regression* (Koenker e Bassett 1978), le quali hanno il vantaggio di non richiedere assunzioni distributive sull'ignota legge di probabilità sottostante ai dati.

In letteratura l'unica proposta modellistica per i quantili di una distribuzione spazio-temporale è stata fatta da Reich, Fuentes e Dunson 2011. Nel loro lavoro gli autori adottano metodi di regressione non parametrici e semiparametrici per ottenere una stima liscia dei quantili del processo. Questo approccio, seppur efficace

dal punto di vista della previsione spaziale, non permette di costruire agevolmente previsioni nel tempo in modo dinamico, tenendo conto della naturale unidirezionalità del tempo.

Come alternativa, in questo lavoro viene proposto un approccio generale alla *quantile regression* dinamica in ambito *spatio-temporal*, basato su opportune estensioni dei classici modelli *State Space*. In ottica modellistica, questa tesi si colloca nel punto di congiunzione di tre grandi temi di ricerca statistica: modellistica di dati e processi spazio-temporali, *quantile regression* e *dynamic state space modelling*.

Il contributo innovativo e originale che si è sviluppato in questo lavoro ruota attorno a quattro argomenti centrali. Il primo elemento di novità presentato è lo sviluppo di un modello quantilico multivariato per dati longitudinali geolocalizzati, come generalizzazione degli usuali metodi legati al *Kriged Kalman Filter*. Quest'estensione è stata costruita combinando l'approccio geostatistico di Wikle e Cressie 1999 con la proposta di Bernardi, Gayraud e Petrella 2015 per la specificazione e la stima di modelli lineari dinamici di *quantile regression* a coefficienti tempo dipendenti. Il modello così sviluppato permette di calcolare contemporaneamente i quantili di livello fissato τ per n serie storiche geolocalizzate, assumendo incorrelazione spaziale e temporale tra i disturbi di misura della formulazione *State Space*.

Con l'obiettivo di proporre una specificazione più generale, è stato poi sviluppato e discusso un modello quantilico a disturbi di misura spazialmente correlati. Questo secondo approccio permette di rispettare maggiormente la struttura di dipendenza spaziale presente nei dati e può portare a diversi vantaggi in termini di precisione di stima dei quantili, come mostrato nell'applicazione empirica su dati reali discussa nel capitolo 4.

Al di là del contesto spaziale, la previsione di quantili fra loro correlati costituisce di per sé un elemento di novità rispetto alla letteratura esistente (De Rossi e Harvey 2009, Kurose e Omori 2012, Bernardi, Gayraud e Petrella 2015) e va a definire il secondo elemento originale di questa tesi.

Come terzo argomento di ricerca si è sviluppato il tema della *multiple quantile regression* dinamica per dati spazio-temporali, andando a costruire un modello congiunto per la gestione di k quantili in corrispondenza di n serie storiche. La stima contemporanea di più quantili può infatti prevenire il verificarsi di problemi di incrocio tra le curve stimate, mediante l'imposizione di vincoli espliciti (Koenker 1984, Zou e Yuan 2008). A tal proposito è stata adottata una modifica ai classici algoritmi di estrazione del segnale per gli stati latenti del sistema, basata sul *Constrained Kalman Filter* proposto da Simon 2006.

Per la stima di tutti i modelli quantilici presentati sono stati proposti dei metodi

Markov Chain Monte Carlo (MCMC), in un'ottica di inferenza Bayesiana, basata su *Gibbs Sampling* con *Data Augmentation* (Yu e Moyeed 2001, Bernardi, Gayraud e Petrella 2015, Sriram e Ramamoorthi 2017).

In aggiunta alla proposta MCMC, nel contesto di *multiple quantile* è stato sviluppato anche un secondo algoritmo di stima a posteriori, le cui radici poggiano sulla teoria del calcolo variazionale e delle *mean-field approximations* (Bishop 2006, Ormerod e Wand 2010).

L'applicazione di un approccio *variational Bayes* in ambito *quantile*, quarto e ultimo contributo originale di questa tesi, si può rivelare particolarmente conveniente in presenza di grandi masse di dati e per modelli di dimensioni considerevoli, e costituisce senz'altro una buona alternativa a metodi di stima Bayesiani basati su algoritmi MCMC.

La tesi è strutturata in cinque capitoli, articolati nel modo seguente: il primo capitolo introduce alcuni strumenti probabilistici elementari per lo studio di processi stocastici definiti su spazi euclidei di dimensione generica, con particolare enfasi sulla costruzione di processi Gaussiani.

Nel secondo capitolo viene presentata la specificazione di un modello spazio-temporale Gaussiano, rappresentabile in forma *State Space*.

Nel terzo capitolo, dopo una breve introduzione alla *quantile regression*, si mostra come generalizzare la classica formulazione *State Space* per stimare i quantili di un processo dinamico, prima assumendo disturbi di misura indipendenti e poi andando ad introdurre una struttura di correlazione spaziale tra questi. Il quarto capitolo mostra un'applicazione su dati reali, che mette in luce alcune delle potenzialità offerte dalla *quantile regression* in ambito spazio-temporale. Infine, nel quinto capitolo vengono ulteriormente generalizzati i modelli precedentemente presentati, in modo da stimare congiuntamente più quantili (*multiple quantile regression*).

In conclusione di ogni capitolo vengono presentati i metodi inferenziali proposti per la stima dei parametri e per l'estrazione del segnale latente.

Capitolo 1

Dati e processi spazio-temporali

1.1 Problemi ed esempi

I dati rilevati nello spazio e nel tempo presentano caratteristiche complesse, che dipendono da molteplici fattori, come lo schema di campionamento, la morfologia del territorio, la frequenza di rilevazione, e chiaramente la natura del fenomeno di interesse.

In geostatistica tipicamente si classificano i dati in base alla topologia dello spazio sul quale sono collocati. Le osservazioni possono essere rilevate su spazi discreti, come griglie, reticoli e grafi, o su spazi continui, come sottoinsiemi dello spazio Euclideo bidimensionale, porzioni di superfici curve o regioni dello spazio tridimensionale.

Generalmente la natura dello spazio di campionamento dipende dalle caratteristiche intrinseche del fenomeno di interesse o da considerazioni derivanti dagli obiettivi di ricerca.

Esempio: immagini digitali

Ad esempio, nell'analisi delle immagini digitali si ipotizza che lo spazio di campionamento sia discreto e giaccia su una griglia regolare. A tale reticolo può essere associato un grafo, dove ogni nodo corrisponde ad un *pixel* e ogni arco definisce una connessione tra *pixel* adiacenti.

Esempio: dati regionalizzati

In econometria, demografia, ed epidemiologia è spesso di interesse indagare il modo in cui diverse regioni, come, ad esempio, gli stati, interagiscono tra loro sulla base di specifici indicatori aggregati. La locazione nello spazio degli stati può essere

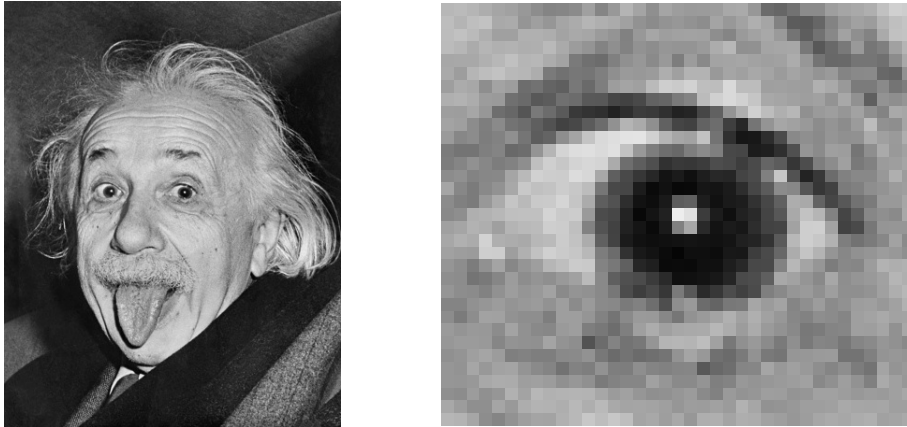


Figura 1.1: Esempio di immagine digitale: Einstein (a sinistra) e zoom sull'occhio destro (a destra).

descritta mediante un grafo, dove i nodi corrispondono ai paesi e gli archi descrivono le connessioni tra questi. Le nazioni tuttavia non sono collocabili all'interno di una griglia regolare, dunque la definizione di spazio discreto, in questo caso, non è equivalente a quella che si assume quando si lavora con delle immagini digitali.

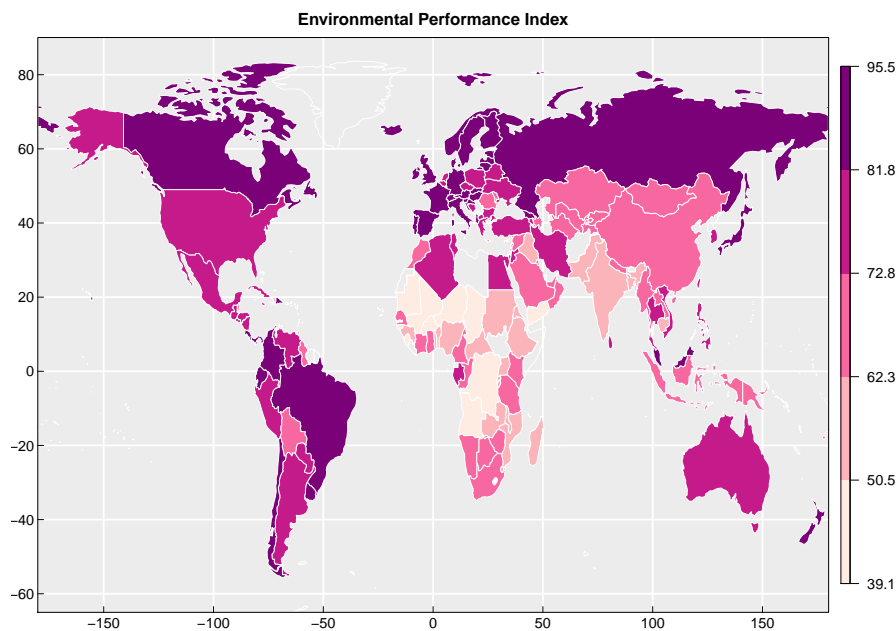


Figura 1.2: Esempio di dati spaziali su spazi discreti: EPI.

Esempio: dati geolocalizzati

Si parla di dati geolocalizzati quando lo spazio di campionamento si ipotizza essere continuo e le osservazioni sono disposte su n siti, fissati o stocastici, localizzabili

attraverso un sistema di coordinate. I siti di campionamento possono essere disposti secondo schemi regolari, o in modo apparentemente disordinato sulla mappa, come si può osservare in figura 1.3.

Le misurazioni di fenomeni ambientali, il monitoraggio degli inquinanti e la rilevazione delle condizioni meteo costituiscono alcuni esempi di dati geolocalizzati.

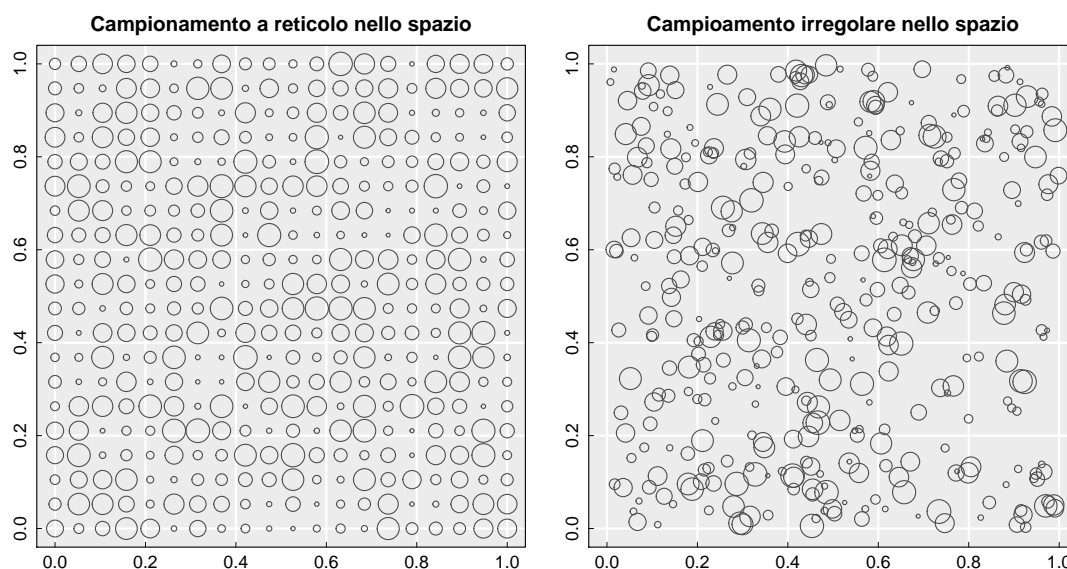


Figura 1.3: Esempio simulato di dati spaziali su una griglia regolare (a sinistra) ed esempio simulato di dati spaziali campionati casualmente sulla mappa (a destra).

Dati spaziali

Qualsiasi sia lo spazio su cui giacciono i dati, per caratterizzarne adeguatamente la struttura di dipendenza è necessario definire una metrica di distanza adeguata. Questa fornisce uno strumento per indurre una topologia sullo spazio di rilevazione e per modellare le principali fonti di variabilità che si presentano nello studio di fenomeni geostatistici. In molte applicazioni infatti è sensato ipotizzare che rilevazioni vicine nello spazio siano fra loro fortemente correlate, e che tale connessione vada scemando all'aumentare della distanza.

Le forme di dipendenza che possono manifestarsi per fenomeni spaziali sono molteplici. Un approccio modellistico largamente diffuso in letteratura prevede di distinguere la variabilità di un processo spaziale in almeno due componenti, dette di larga e piccola scala. La prima è quella che determina il comportamento macroscopico del processo ed è tipicamente caratterizzata da forme di correlazione a forte persistenza nello spazio. La seconda è quella che determina le fluttuazioni locali del

processo ed è spesso associata a funzioni di correlazione stazionarie e con una rapida decrescita all'aumentare della distanza.

Lo scopo di un'analisi di statistica spaziale è quello di identificare in modo quanto più preciso possibile le diverse strutture che compongono il processo sulla base dei dati, della conoscenza pregressa e dello schema di campionamento adottato. Una buona ricostruzione del processo e della sua distribuzione può, in secondo luogo, essere sfruttata per produrre previsioni o fare interpolazione.

Per dati spaziali, e così anche per dati temporali, la classica assunzione di indipendenza non è realistica ed è quindi necessario modellare adeguatamente le dinamiche con cui i dati interagiscono tra loro.

Dati temporali

A differenza dello spazio, il tempo è una dimensione unidirezionale, questo implica che nell'analisi di dati temporali non è importante solo la distanza che separa due eventi, ma anche il loro ordinamento. Questa è la principale caratteristica che differenzia un processo spaziale da uno temporale.

Anche nell'ambito delle serie storiche è possibile distinguere varie dinamiche che vanno a comporre la variabilità totale del processo e, anche in questo caso, si può distinguere tra larga e piccola scala, secondo la stessa logica vista in precedenza.

La componente di larga scala in ambito temporale è dominata da fenomeni tipicamente non stazionari, quali *trend*, periodicità e *shock* strutturali. La componente di piccola scala invece raccoglie fenomeni stazionari e di breve memoria. Nella piccola scala ciò che domina la variabilità del processo sono le funzioni di autocovarianza, che quantificano la persistenza nel tempo di un processo, ovvero quanto il passato influenzi il presente ed il futuro.

In ottica predittiva, al netto della componente di grande scala, le funzioni di covarianza spaziale e autocovarianza temporale rivestono un ruolo di fondamentale importanza. La funzione di covarianza spaziale permette infatti di mettere in relazione rilevazioni del processo in diversi punti dello spazio, mentre la funzione di autocovarianza permette di stabilire un legame tra passato, presente e futuro di un processo.

Le caratteristiche che sono state intuitivamente introdotte per dati spaziali e dati temporali separatamente si possono combinare per dare origine a fenomeni più complessi con dinamiche sia spaziali che temporali. Dati uno spazio di campionamento spaziale ed uno temporale, avremo dei dati che ad ogni tempo fissato descrivono un processo spaziale e ad ogni sito fissato descrivono un serie storica. Nel suo complesso i dati possono essere visti come una serie storica multivariata a componenti

spazialmente correlate. Avremo dunque una fattorizzazione del processo il larga e piccola scala spazio-temporale. La larga scala raccoglierà comportamenti medi e forme di non stazionarietà nel tempo e nello spazio, mentre la piccola scala ingloberà variazioni stazionarie, locali e con memoria a breve termine.

I modelli e gli algoritmi che verranno proposti in questa tesi sono pensati per dati raccolti su un numero di siti fissato con geolocalizzazione statica, ma sparsa nello spazio e con misurazioni temporali effettuate a cadenza regolare.

Nel seguito di questo capitolo verranno introdotti alcuni strumenti probabilistici elementari relativi alla definizione di processi stocastici su spazi euclidei multidimensionali, che costituiscono la base modellistica su cui poggiano tutti gli sviluppi che verranno presentati nei capitoli successivi.

1.2 Processi spazio-temporali

Processi stocastici

Siano $(\Omega, \mathcal{F}, \mathbb{P})$ uno spazio di probabilità, $(\mathbf{E}, \mathcal{E})$ uno spazio misurabile e $\mathbf{X} \subseteq \mathbb{R}^d$ uno spazio parametrico.

Definizione 1 (Processo Stocastico) *Un processo stocastico $(Y(\mathbf{x}) \in \mathbf{E}, \mathbf{x} \in \mathbf{X})$ definito su $(\Omega, \mathcal{F}, \mathbb{P})$ a valori in $(\mathbf{E}, \mathcal{E})$ è un'applicazione misurabile*

$$\begin{aligned} Y &: \Omega \times \mathbf{X} \longrightarrow \mathbf{E} \\ (\omega, \mathbf{x}) &\longrightarrow Y_\omega(\mathbf{x}) \end{aligned}$$

dove $(\mathbf{E}, \mathcal{E})$ è detto spazio degli stati, mentre \mathbf{X} è l'insieme dei siti su cui è definito il processo.

Più informalmente possiamo dire che un processo stocastico è una famiglia di variabili aleatorie $Y(\cdot)$ a valori in $(\mathbf{E}, \mathcal{E})$ indicizzate da un parametro $\mathbf{x} \in \mathbf{X}$. Per ogni evento fissato $\omega \in \Omega$ diremo che $(Y_\omega(\mathbf{x}), \mathbf{x} \in \mathbf{X})$ è una traiettoria del processo, ovvero una sua realizzazione $Y : \mathbf{X} \longrightarrow \mathbf{E}$.

Definizione 2 (Distribuzione) *Per ogni valore $n \in \mathbb{N}$, per ogni n -upla $(\mathbf{x}_1, \dots, \mathbf{x}_n) \in \mathbf{X}^n$, definiamo distribuzione di $\mathbf{Y} = (Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))$ l'immagine di \mathbb{P} sotto la mappa $\omega \longrightarrow \mathbf{Y}(\omega)$. Per ogni evento $\mathbf{A} = (A_1 \cap \dots \cap A_n) \in \mathcal{E}^n$ si ha che:*

$$\mathbb{P}(\mathbf{Y} \in \mathbf{A}) = \mathbb{P}(\{\omega \in \Omega : \mathbf{Y}(\omega) \in \mathbf{A}\})$$

tale si dice distribuzione finito-dimensionale del processo $Y(\cdot)$.

La famiglia di tutte le distribuzioni finito-dimensionali del processo ne caratterizza completamente la distribuzione.

Se $\mathbf{X} \subseteq \mathbb{R}^+$ si parla di distribuzione temporale, mentre se $\mathbf{X} \subseteq \mathbb{R}^d$, con $d > 1$ si parla di distribuzione spaziale. Infine se $\mathbf{X} \subseteq \mathbb{R}^d \times \mathbb{R}^+$ si dice che $Y(\cdot)$ è un processo spazio-temporale.

D'ora in avanti considereremo solo processi a valori reali, $\mathbf{E} \subseteq \mathbb{R}$, dotati di σ -algebra di Borel $\mathcal{E} = \mathcal{B}(\mathbf{E})$ e con distribuzione di probabilità assolutamente continua.

Processi del secondo ordine

Definizione 3 (Processo del secondo ordine) Diremo che $Y(\cdot)$ è un processo del secondo ordine se per ogni $\mathbf{x} \in \mathbf{X}$, $\mathbb{E}[|Y(\mathbf{x})|^2] < \infty$. Se $Y(\cdot)$ è del secondo ordine indicheremo $Y(\cdot) \in L^2$, dove $L^2 = L^2(\Omega, \mathcal{F}, \mathbb{P})$ è lo spazio di tutte le variabili aleatorie a valori reali, a varianza finita su $(\mathbf{E}, \mathcal{E})$.

Ogni processo in L^2 ammette media $m : \mathbf{X} \rightarrow \mathbb{R}$ e covarianza $C : \mathbf{X} \times \mathbf{X} \rightarrow \mathbb{R}$ finite, definibili come $m(\mathbf{x}) = \mathbb{E}[Y(\mathbf{x})]$ e $C(\mathbf{x}', \mathbf{x}'') = \mathbb{E}(Y(\mathbf{x}'), Y(\mathbf{x}''))$.

Diremo che un processo è centrato se ha funzione media nulla $m(\mathbf{x}) = 0$ per ogni $\mathbf{x} \in \mathbf{X}$ e che ha funzione di covarianza (o di correlazione) omogenea rispetto a traslazioni se:

$$\forall \mathbf{x}', \mathbf{x}'', \mathbf{h} \in \mathbf{X} : C(\mathbf{x}', \mathbf{x}' + \mathbf{h}) = C(\mathbf{x}'', \mathbf{x}'' + \mathbf{h}) = C(\mathbf{h})$$

Una funzione $C : \mathbf{X} \times \mathbf{X} \rightarrow \mathbb{R}$ è una covarianza se rispetta la proprietà

$$\forall n \in \mathbb{N}, \forall \mathbf{a} \in \mathbb{R}^n, \forall (\mathbf{x}_1, \dots, \mathbf{x}_n) \in \mathbf{X}^n : \sum_{i=1}^n \sum_{j=1}^n a_i a_j C(\mathbf{x}_i, \mathbf{x}_j) \geq 0.$$

e in tal caso si dice che $C(\cdot, \cdot)$ è semi-definita positiva.

Per ogni processo del secondo ordine è definibile una funzione di correlazione $R(\cdot, \cdot)$ semi-definita positiva, tale che $C(\mathbf{x}', \mathbf{x}'') = C(\mathbf{x}', \mathbf{x}')C(\mathbf{x}'', \mathbf{x}'')R(\mathbf{x}', \mathbf{x}'')$. Come l'usuale indice di correlazione, la funzione di correlazione definisce una misura normalizzata di covarianza e fornisce uno strumento utile per scomporre la variabilità del processo (la varianza), dalla sua struttura di dipendenza del secondo ordine nello spazio (la correlazione).

Le funzioni di correlazione e covarianza rivestono un ruolo di fondamentale importanza nella modellistica di fenomeni aleatori reali. Una prima approssimazione della distribuzione di un processo può essere ottenuta caratterizzandone adeguatamente le funzioni di media e covarianza, che in molti casi definiscono le maggiori fonti di variabilità strutturata di un processo.

Se una funzione di covarianza (o di correlazione) è omogenea rispetto a traslazioni e inoltre dipende solo dalla distanza $\|\mathbf{h}\|$ e non dalla direzione $\mathbf{h}/\|\mathbf{h}\|$, ovvero per ogni $\mathbf{h} \in \mathbf{X}$, $C(\mathbf{h}) = C(\|\mathbf{h}\|)$, allora $C(\cdot)$ si dice isotropica.

La semplicità di costruzione e di interpretazione rende le funzioni di covarianza isotropiche le più utilizzate e le più maneggevoli, soprattutto in contesti applicativi e modellistici.

Tabella 1.1: Funzioni di covarianza stazionarie ed isotropiche.

Modello	Funzione di Covarianza
Esponenziale	$C(\ \mathbf{h}\ ; \sigma^2, \lambda) = \sigma^2 \exp(-\lambda\ \mathbf{h}\)$
Gaussiano	$C(\ \mathbf{h}\ ; \sigma^2, \lambda) = \sigma^2 \exp(-\lambda\ \mathbf{h}\ ^2)$
Sferico	$C(\ \mathbf{h}\ ; \sigma^2, \lambda) = \sigma^2 \left\{ 1 - \frac{3}{2} \left(\frac{\ \mathbf{h}\ }{\lambda} \right) + \frac{1}{2} \left(\frac{\ \mathbf{h}\ }{\lambda} \right)^3 \right\} \mathbb{I}(\ \mathbf{h}\ < \lambda)$
<i>Powered Exponential</i>	$C(\ \mathbf{h}\ ; \sigma^2, \lambda, \alpha) = \sigma^2 \exp(-\lambda\ \mathbf{h}\ ^\alpha)$
Matern	$C(\ \mathbf{h}\ ; \sigma^2, \lambda, \nu) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{\ \mathbf{h}\ }{\lambda} \right)^\nu K_\nu \left(\sqrt{2\nu} \frac{\ \mathbf{h}\ }{\lambda} \right)$

In tabella 1.1 sono riportate alcune delle funzioni di covarianza più largamente utilizzate in letteratura, mentre in figura 1.4 sono riportati i grafici delle medesime funzioni al variare dei parametri.

Come si può notare dai grafici in figura 1.4 e dalla tabella 1.1, ogni funzione di covarianza ha almeno due parametri: $\sigma^2 > 0$ che definisce il valore della varianza del processo e $\lambda > 0$, che controlla la dispersione della funzione di covarianza. I parametri aggiuntivi $0 < \alpha < 2$ e $\nu > 0$ per le funzioni di covarianza *powered exponential* e di Matern costituiscono dei parametri di forma.

Le funzioni di covarianza isotropiche sono uno strumento utile e semplice da usare; tale semplicità è però controbilanciata in termini di flessibilità. Le ipotesi di omogeneità e isotropia pongono costrizioni molto stringenti sulle funzioni di covarianza, e spesso non trovano riscontro empirico.

Un modo per generalizzare il concetto di covarianza isotropica, mantenendone la semplicità nella costruzione e nell'interpretazione, è quello di introdurre anisotropie indotte da trasformazioni lineari sullo spazio dei siti. Diremo che la funzione di covarianza mostra anisotropie geometriche se $C(\mathbf{h}) = C(\|\mathbf{A}\mathbf{h}\|)$, dove \mathbf{A} è una generica matrice di deformazione lineare.

Un'importante classe di processi del secondo ordine sono i processi gaussiani, i quali hanno distribuzione completamente caratterizzabile attraverso le funzioni di media e covarianza.

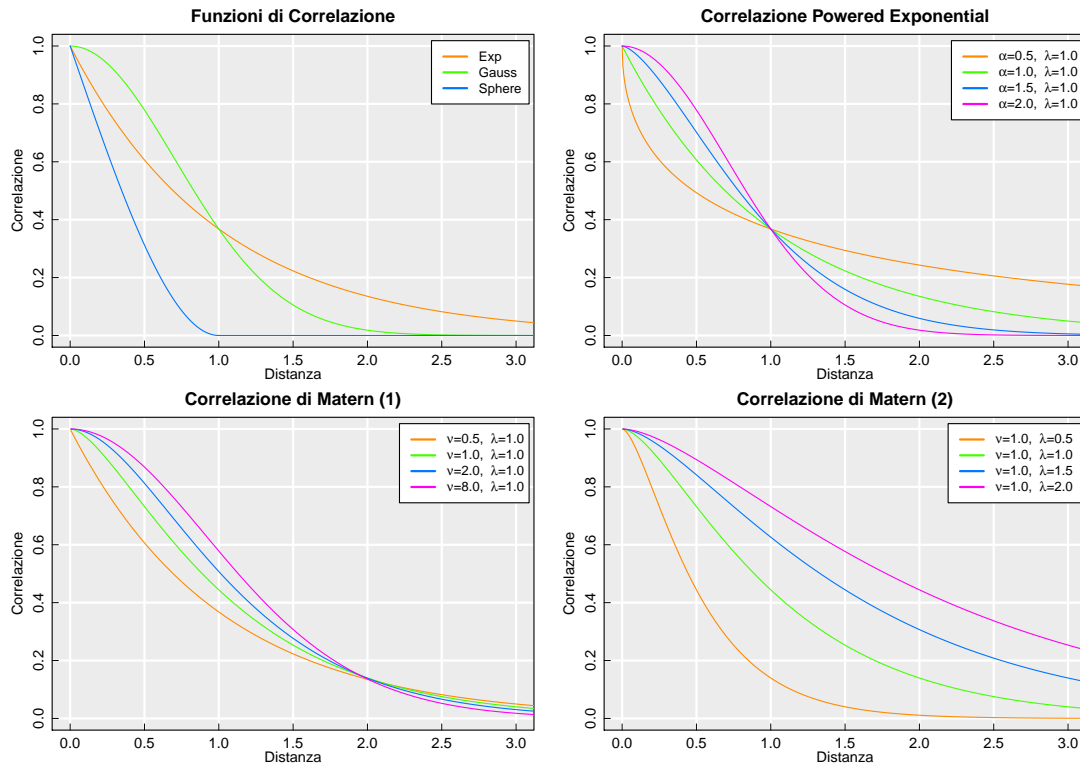


Figura 1.4: Grafico delle funzioni di correlazione spaziale. Ove non indicato diversamente il parametri di varianza σ^2 e di dispersione sono fissati pari a 1.

Definizione 4 (Processi Gaussiani) Diremo che $Y(\cdot)$ è un processo Gaussiano su \mathbf{X} se per ogni n -upla $(\mathbf{x}_1, \dots, \mathbf{x}_n) \in \mathbf{X}^n$ e per ogni $\mathbf{a} \in \mathbb{R}^n$, il vettore aleatorio $\mathbf{Y} = (Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))$ è tale che $Z = \langle \mathbf{a}, \mathbf{Y} \rangle \sim \mathcal{N}(\cdot, \cdot)$, dove $\langle \cdot, \cdot \rangle$ indica il prodotto scalare standard tra vettori di dimensione comparabile e \mathcal{N} denota la distribuzione normale.

Definiti \mathbf{m} e \mathbf{C} come il vettore medio e la matrice di varianza e covarianza di \mathbf{Y} , diremo che $\mathcal{N}(\cdot | \mathbf{m}, \mathbf{C})$ è la funzione di densità n -dimensionale del processo $Y(\cdot)$:

$$\mathcal{N}(\mathbf{y} | \mathbf{m}, \mathbf{C}) = (2\pi)^{-\frac{n}{2}} \cdot |\mathbf{C}|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2}(\mathbf{y} - \mathbf{m})^\top \mathbf{C}^{-1}(\mathbf{y} - \mathbf{m})\right\} \quad (1.2.1)$$

e indicheremo con la notazione $\mathbf{Y} \sim \mathcal{N}(\mathbf{m}, \mathbf{C})$ che \mathbf{Y} è un vettore aleatorio Gaussiano con media \mathbf{m} e matrice di covarianza \mathbf{C} .

Si vedano Cressie 1993 e Rue e Held 2005 per una più ampia trattazione di processi stocastici Gaussiani.

Processi stazionari

Definizione 5 (Stazionarietà forte) $Y(\cdot)$ è un processo stazionario in senso forte se tutte le sue distribuzioni finito-dimensionali sono invarianti rispetto a

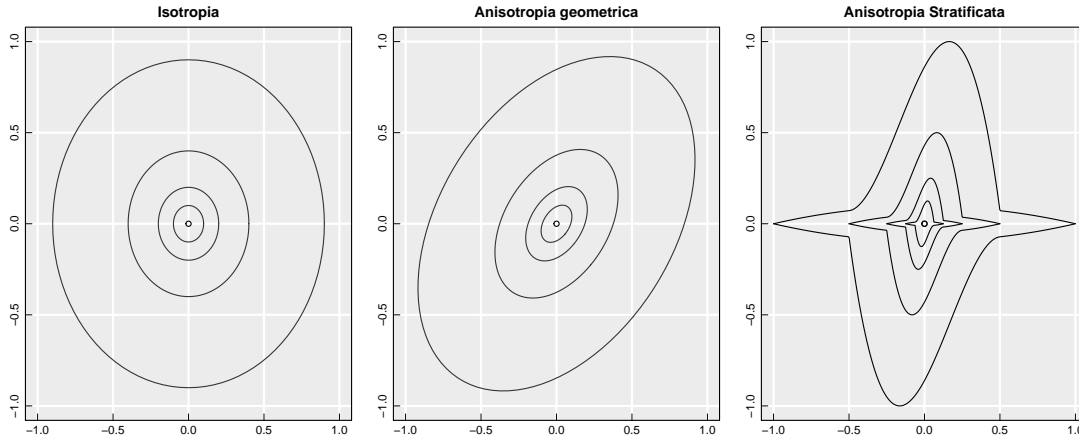


Figura 1.5: Orbite di correlazione spaziale per un processo isotropico (a sinistra), geometricamente anisotropico (al centro) e con anisotropia stratificata, ovvero con geometrica (a destra).

traslazioni:

$$\forall n \in \mathbb{N}, \forall \mathbf{A} \in \mathcal{E}^n, \forall \mathbf{h} \in \mathbf{X}, \forall (\mathbf{x}_1, \dots, \mathbf{x}_n) \in \mathbf{X}^n : \\ \mathbb{P}((Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)) \in \mathbf{A}) = \mathbb{P}((Y(\mathbf{x}_1 + \mathbf{h}), \dots, Y(\mathbf{x}_n + \mathbf{h})) \in \mathbf{A})$$

La stazionarietà forte è una proprietà estremamente stringente e difficilmente verificabile nella pratica, il che la rende poco maneggiabile a fini applicativi. Questo ci spinge ad introdurre una seconda definizione di stazionarietà, più debole, ma più semplice ed applicabile.

Definizione 6 (Stazionarietà debole) $Y(\cdot)$ è un processo stazionario del secondo ordine, o debolmente stazionario, se la sua media è costante e se la sua funzione di covarianza è omogenea rispetto a traslazioni:

$$\forall \mathbf{x}, \mathbf{h} \in \mathbf{X} : m(\mathbf{x}) = m \wedge C(\mathbf{x}, \mathbf{x} + \mathbf{h}) = C(\mathbf{h}).$$

Come conseguenza dell'omogeneità nello spazio, ogni processo debolmente stazionario è dotato di funzione di covarianza fattorizzabile nel prodotto $C(\mathbf{h}) = C(\mathbf{0}) \cdot R(\mathbf{h})$, dove $\sigma^2 = C(\mathbf{0})$ è il parametro che determina la varianza del processo per ogni valore di \mathbf{x} .

Se un processo $Y(\cdot)$ è stazionario in senso forte allora è stazionario anche in senso debole. Il contrario invece non è vero, se non per processi Gaussiani, per i quali stazionarietà forte e debole coincidono.

Se $Y(\cdot)$ è un processo stazionario del secondo ordine si dimostra (Gaetan e Guyon 2010) che valgono le seguenti proprietà:

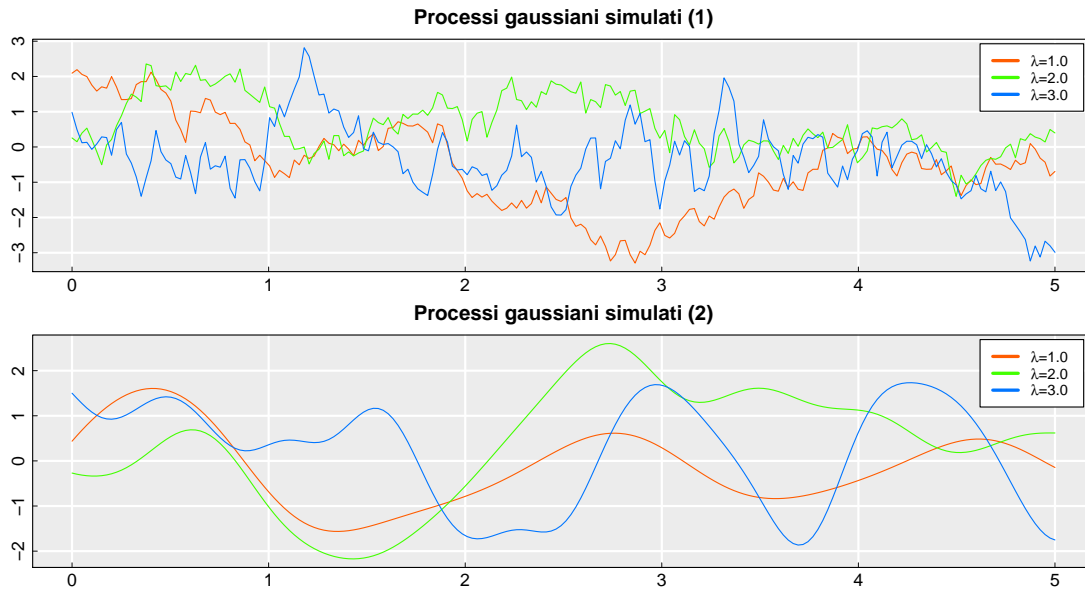


Figura 1.6: Processi Gaussiani simulati con funzione di covarianza esponenziale (in alto) e con funzione di covarianza Gaussiana (in basso) sulla retta reale. I processi sono stati simulati a media nulla e con funzione di covarianza di parametri: $\sigma^2 = 1$, $\lambda = 1$ (arancione), $\lambda = 2$ (verde) e $\lambda = 4$ (blu).

- $\forall \mathbf{x}, \mathbf{h} \in \mathbf{X}, |C(\mathbf{h})| \leq C(\mathbf{0})$;
- $\forall \mathbf{A} \in \mathbb{R}^d \times \mathbb{R}^d, Y^*(\mathbf{x}) = Y(\mathbf{Ax}) : C^*(\mathbf{x}) = C(\mathbf{Ax})$;
- Se $C(\cdot)$ è continua nell'origine, allora è uniformemente continua ovunque.
- Se $(C_n(\cdot), n \in \mathbb{N})$ sono funzioni di covarianza stazionarie, allora lo sono anche anche:
 - $C(\mathbf{h}) = \sum_{n \in \mathbb{N}} a_n C_n(\mathbf{h})$ per ogni successione $a_n \geq 0$;
 - $C(\mathbf{h}) = \prod_{n \in \mathbb{N}} C_n(\mathbf{h})$
 - $C(\mathbf{h}) = \lim_{n \rightarrow \infty} C_n(\mathbf{h})$

Esistono molti ambiti nei quali i processi con cui è necessario operare non sono stazionari, neanche debolmente, è quindi necessario allargare la famiglia dei processi stazionari per modellare fenomeni più complessi.

Wahba 1990 e Cressie 1993 mostrano come la classe dei processi stazionari può essere generalizzata attraverso la definizione di processi stazionari intrinseci e di funzioni casuali intrinseche di ordine k , che però in questa tesi non verranno approfondite.

Convoluzione di processi Gaussiani

Un modo naturale per costruire processi Gaussiani su sottoinsiemi di \mathbb{R}^d è quello di operare una convoluzione:

$$Y(\mathbf{x}) = \int_{\mathbf{x}} K(\mathbf{x}, \mathbf{u}) W(\mathbf{u}) d\mathbf{u}, \quad (1.2.2)$$

dove $K(\cdot, \cdot)$ è una funzione *kernel* non stocastica su $\mathbb{R}^d \times \mathbb{R}^d$, ovvero una funzione a valori reali, non negativa, continua, integrabile e simmetrica rispetto ai suoi argomenti. $W(\cdot)$ è invece un processo stocastico Gaussiano, centrato e con funzione di covarianza $C_w(\cdot, \cdot)$.

Il processo così costruito è ben definito in L^2 ed ha funzione di covarianza

$$C_y(\mathbf{x}', \mathbf{x}'') = \iint_{\mathbf{x} \times \mathbf{x}} K(\mathbf{x}', \mathbf{u}) K(\mathbf{x}'', \mathbf{v}) C_w(\mathbf{u}, \mathbf{v}) d\mathbf{u} d\mathbf{v}. \quad (1.2.3)$$

Se $W(\cdot)$ è scelto a componenti ortogonali su \mathbb{R}^d , quindi tale che:

$$\mathbb{E}[W(\mathbf{u}) \cdot W(\mathbf{v})] = \delta(\mathbf{u}, \mathbf{v}),$$

allora la convoluzione 1.2.2 ha funzione di covarianza $C(\mathbf{x}', \mathbf{x}'') = K(\mathbf{x}', \mathbf{x}'')$, grazie alle proprietà delle funzioni *kernel*:

$$\begin{aligned} C_y(\mathbf{x}', \mathbf{x}'') &= \iint_{\mathbf{x} \times \mathbf{x}} K(\mathbf{x}', \mathbf{u}) K(\mathbf{x}'', \mathbf{v}) \delta(\mathbf{u}, \mathbf{v}) d\mathbf{u} d\mathbf{v} \\ &= \int_{\mathbf{x}} K(\mathbf{x}', \mathbf{u}) \cdot K(\mathbf{x}'', \mathbf{u}) d\mathbf{u} = K(\mathbf{x}', \mathbf{x}''). \end{aligned}$$

Scelte comuni per il processo latente $W(\cdot)$ sono gli incrementi di un moto Browniano standard o un *white noise* Gaussiano su \mathbb{R}^d . Se la funzione *kernel* $K(\cdot, \cdot)$ e la funzione di covarianza $C_w(\cdot, \cdot)$ sono stazionarie e isotropiche, tale sarà anche la covarianza del processo $C_y(\cdot, \cdot)$.

La convoluzione è un potente strumento per specificare modelli Gaussiani secondo un approccio costruttivo, che si rivela particolarmente utile nella definizione di processi spazio-temporali. Adottando un *kernel* spaziale in $\mathbb{R}^d \times \mathbb{R}^d$ ed un fattore latente in $\mathbb{R}^d \times \mathbb{R}^+$, il processo risultante dall'operazione di convoluzione è un processo definito nello spazio e nel tempo, con funzione di covarianza spaziale determinata da $K(\cdot, \cdot)$ e con dinamica temporale governata da $W(\cdot, t)$:

$$Y(\mathbf{s}, t) = \int_{\mathbf{x}} K(\mathbf{x}, \mathbf{s}) W(\mathbf{x}, t) d\mathbf{x}$$

I metodi di generazione e analisi di processi Gaussiani basati su sviluppi e trasformate sono generalmente trattati in tutti i testi di statistica spaziale, di cui nominiamo, a titolo di esempio, Ripley 1981, Cressie 1993 e Gaetan e Guyon 2010.

A meno di casi particolari, il calcolo dell'integrale di convoluzione 1.2.2 non ammette soluzione in forma chiusa e pertanto è necessario adottare approssimazioni tramite convoluzione discreta:

$$Y(\mathbf{x}) = \sum_{i=1}^n \alpha_i \cdot K(\mathbf{x}, \mathbf{u}_i) = \sum_{i=1}^n W(\mathbf{u}_i) \cdot K(\mathbf{x}, \mathbf{u}_i) \quad (1.2.4)$$

dove $(\mathbf{u}_1, \dots, \mathbf{u}_n)$ sono un insieme di nodi di controllo, $\mathbf{u}_i \in \mathbf{X}$ e $\alpha_i = W(\mathbf{u}_i)$ sono variabili aleatorie Gaussiane.

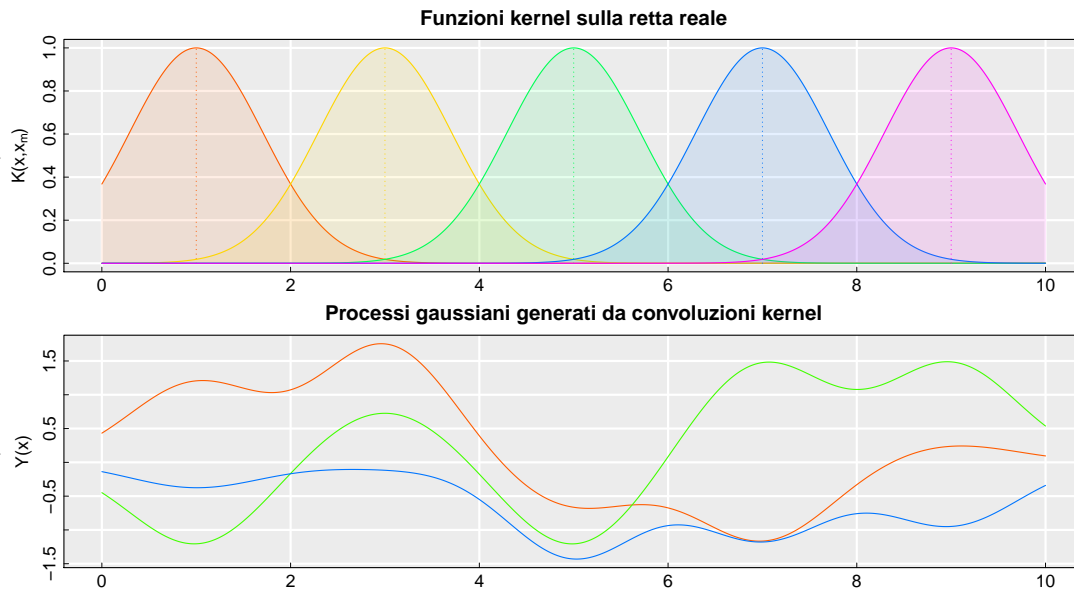


Figura 1.7: Funzioni *kernel* Gaussiani di base (in alto) e processi Gaussiani simulati via convoluzione discreta (in basso).

Questa rappresentazione discreta può essere interpretata come uno sviluppo in basi di funzioni *kernel* del processo $Y(\cdot)$ ed assume un'importanza rilevante non solo nella teoria dei processi stocastici, ma anche nell'ambito del *Machine Learning* e del *Data Mining*, come mostrato, ad esempio, nel lavoro di Wilk, Rasmussen e Hensman 2017.

Alcune delle più moderne tecniche di regressione e classificazione, quali, ad esempio, le *Support Vector Machine*, poggiano su sviluppi funzionali come 1.2.4. Nei lavori di Wahba 1990, Ruppert, Wand e Carroll 2003 e Wang 2011 si trova una trattazione sistematica e approfondita di approssimazioni funzionali mediante rappresentazioni *kernel* e metodi ad esse collegati; mentre in Hastie, Tibshirani e Friedman 2009 e Azzalini e Scarpa 2012 sono descritti ad analizzati alcuni tra i più importanti modelli statistici che fanno uso tali sviluppi per l'analisi e la previsione di dati complessi.

Capitolo 2

Modello spazio-temporale in media

Per analizzare processi con dinamiche sia spaziali, che temporali sono state proposte diverse metodologie statistiche negli anni. Un primo approccio percorribile è quello di trattare la variabilità spazio-temporale considerando i dati come una collezione di serie storiche fra loro correlate, per poi andare ad applicare su queste classici modelli per serie storiche multivariate, come VAR e VARMA. Questa procedura fornisce un utile mezzo per produrre previsioni nel tempo, ma di contro non permette di estendere i risultati inferenziali anche a regioni dello spazio non ancora esplorate e dunque non permette di fare previsione o interpolazione nello spazio.

Un secondo approccio prevede di adottare i mezzi inferenziali propri del paradigma geostatistico per prevedere il processo congiuntamente nello spazio e nel tempo. I metodi di previsione ed interpolazione tradizionali, come il *kriging*, possono essere facilmente estesi al dominio spazio-temporale aggiungendo una dimensione al sistema di coordinate spaziali, passando da \mathbb{R}^d a $\mathbb{R}^d \times \mathbb{R}^+$. Questo secondo approccio tuttavia non permette di considerare le sostanziali differenze che sussistono fra processi spaziali e processi temporali, prevalentemente dovuti all'unidirezionalità della dimensione temporale. Inoltre, l'esplicita specificazione di un modello congiunto per la covarianza, necessario per adottare metodologie predittive geostatistiche, pone un problema non banale, difficilmente affrontabile attraverso i modelli parametrici standard.

Una possibile soluzione è quella di combinare i due approcci, producendo un modello dinamico a componenti spazialmente correlate, dove la struttura di dipendenza temporale è gestita mediante modelli per serie storiche multivariate, mentre la componente spaziale è parametrizzata da opportune funzioni di covarianza. Questi modelli si dicono dinamici, nel senso che tengono conto dell'ordinamento unidirezionale del tempo, e descrittivi, poiché permettono di modellare la struttura di

correlazione spaziale, ma senza un'interpretazione causale della stessa.

In quest'ottica, la classe dei modelli *State Space* e l'algoritmo del filtro di Kalman forniscono dei flessibili strumenti per combinare le informazioni provenienti dalle dimensioni temporale e spaziale. L'utilizzo di questi modelli nel dominio della geostatistica è diventato ormai pratica comune, come dimostrano i lavori di Wikle e Cressie 1999, Wikle, Berliner e Cressie 2004, Sahu e Mardia 2005 e Cressie e Wikle 2011. A giorno d'oggi l'algoritmo *Kriged Kalman filter* trova applicazione in un gran numero di problemi reali, in particolare negli ambiti meteorologici (Wikle e Cressie 1999, Stroud, Müller e Sansó 2001, Wikle, Berliner e Cressie 2004) e di monitoraggio di agenti inquinanti (Scarpa 2001, Sahu e Mardia 2005).

In questo capitolo verrà presentato un modello dinamico per la previsione di processi spazio-temporali, inquadrabile all'interno della letteratura classica del *Kriged Kalman filter*. Si discuteranno poi alcuni metodi e algoritmi utili a fare inferenza su parametri del modello in un ottica Bayesiana.

2.1 Il modello teorico

Consideriamo un processo spazio-temporale $(Y(\mathbf{s}, t) \in L^2, \mathbf{s} \in \mathbf{S}, t \in \mathbf{T})$, dove $\mathbf{S} \subseteq \mathbb{R}^2$ e $\mathbf{T} \subseteq \mathbb{N}$. Lo scopo dell'analisi è quello di scomporre in forma strutturale la variabilità di $Y(\cdot, \cdot)$ e di prevederne l'evoluzione in funzione di spazio e tempo.

Assumiamo che il processo osservabile $Y(\mathbf{s}, t)$ sia la composizione di un processo latente $Z(\mathbf{s}, t)$ e di un errore $\varepsilon(\mathbf{s}, t)$, fra loro indipendenti, collegati dall'equazione di misura

$$Y(\mathbf{s}, t) = Z(\mathbf{s}, t) + \varepsilon(\mathbf{s}, t). \quad (2.1.1)$$

Assumiamo che $\varepsilon(\mathbf{s}, t)$ sia un processo di piccola scala stazionario e Gaussiano, con dipendenza spaziale ridotta e senza dinamiche temporali coerenti. Il processo $Z(\mathbf{s}, t)$ rappresenta invece la componente di variabilità di grande scala; seguendo l'approccio di Wikle e Cressie 1999 e Wikle, Berliner e Cressie 2004, si assume che $Z(\mathbf{s}, t)$ evolva nel tempo in accordo con l'equazione di stato

$$Z(\mathbf{s}, t + 1) = \int_{\mathbf{S}} K(\mathbf{u}, \mathbf{s}) Z(\mathbf{u}, t) d\mathbf{u} + \eta(\mathbf{s}, t), \quad (2.1.2)$$

dove $\eta(\cdot, \cdot)$ è un processo Gaussiano stazionario a componenti indipendenti nel tempo e correlate nello spazio, mentre $K(\cdot, \cdot)$ è una funzione *kernel* nota e non stocastica che descrive l'evoluzione autoregressiva spazio-temporale della componente di larga scala.

Assumiamo inoltre

$$\begin{aligned}
\mathbb{E}[\varepsilon(\mathbf{s}', t')\varepsilon(\mathbf{s}'', t'')] &= 0, \quad \forall \mathbf{s}', \mathbf{s}'' \in \mathbf{S}, \quad \forall t' \neq t'' \in \mathbf{T}, \\
\mathbb{E}[\eta(\mathbf{s}', t')\eta(\mathbf{s}'', t'')] &= 0, \quad \forall \mathbf{s}', \mathbf{s}'' \in \mathbf{S}, \quad \forall t' \neq t'' \in \mathbf{T}, \\
\mathbb{E}[\varepsilon(\mathbf{s}', t')\eta(\mathbf{s}'', t'')] &= 0, \quad \forall \mathbf{s}', \mathbf{s}'' \in \mathbf{S}, \quad \forall t', t'' \in \mathbf{T}, \\
\mathbb{E}[Z(\mathbf{s}', t')\varepsilon(\mathbf{s}'', t'')] &= 0, \quad \forall \mathbf{s}', \mathbf{s}'' \in \mathbf{S}, \quad \forall t', t'' \in \mathbf{T}, \\
\mathbb{E}[Z(\mathbf{s}', t')\eta(\mathbf{s}'', t'')] &= 0, \quad \forall \mathbf{s}', \mathbf{s}'' \in \mathbf{S}, \quad \forall t' \neq t'' \in \mathbf{T}.
\end{aligned} \tag{2.1.3}$$

Definiamo ora il processo $Z(\mathbf{s}, t)$ come una convoluzione *kernel*, ovvero come uno sviluppo in basi di funzioni, in accordo con l'equazione

$$Z(\mathbf{s}, t) = \sum_{j=1}^m K(\mathbf{u}_j, \mathbf{s}) \alpha(\mathbf{u}_j, t), \tag{2.1.4}$$

dove $\alpha(\cdot, \cdot)$ è un processo latente a media nulla e varianza finita e $\{\mathbf{u}_j\}_{j=1}^m$ sono i nodi di posizionamento della funzione *kernel* $K(\cdot, \cdot)$.

Dal teorema di Mercer (Wahba 1990), sappiamo che se $K(\cdot, \cdot)$ è una funzione *kernel* continua e simmetrica rispetto ai suoi argomenti è possibile espanderla in serie di funzioni:

$$K(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^{\infty} \phi_i(\mathbf{u})\phi_i(\mathbf{v}). \tag{2.1.5}$$

dove $\{\phi_i(\mathbf{u}), i \in \mathbb{N}\}$ è una base di autofunzioni complete e ortonormali:

$$\int_{\mathbf{S}} \phi_i(\mathbf{u}) \phi_j(\mathbf{u}) d\mathbf{u} = \delta_{ij}, \tag{2.1.6}$$

e δ_{ij} è la funzione delta di Dirac, pari a 1 se $i = j$, 0 altrimenti. Sostituendo 2.1.5 in 2.1.4 otteniamo

$$Z(\mathbf{s}, t) = \sum_{i=1}^{\infty} \phi_i(\mathbf{s}) \gamma_i(t), \tag{2.1.7}$$

dove $\gamma_i(t)$ è un processo a media nulla e varianza limitata, definito come

$$\gamma_i(t) = \sum_{j=1}^m \phi_i(\mathbf{u}_j) \alpha(\mathbf{u}_j, t). \tag{2.1.8}$$

Sostituendo 2.1.5 e 2.1.7 in 2.1.2 e sfruttando la proprietà di ortonormalità 2.1.6 otteniamo

$$\begin{aligned}
Z(\mathbf{s}, t+1) &= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \phi_j(\mathbf{s}) \gamma_i(t) \int_{\mathbf{S}} \phi_i(\mathbf{u}) \phi_j(\mathbf{u}) d\mathbf{u} + \eta(\mathbf{s}, t) \\
&= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \phi_j(\mathbf{s}) \gamma_i(t) \delta_{ij} + \eta(\mathbf{s}, t) \\
&= \sum_{i=1}^{\infty} \phi_i(\mathbf{s}) \gamma_i(t) + \eta(\mathbf{s}, t).
\end{aligned}$$

Infine sfruttando l'equivalenza tra 2.1.4 e 2.1.7 otteniamo

$$Z(\mathbf{s}, t+1) = \sum_{j=1}^m K(\mathbf{u}_j, \mathbf{s}) \alpha(\mathbf{u}_j, t) + \eta(\mathbf{s}, t), \quad (2.1.9)$$

che può essere riscritta in forma vettoriale come

$$\mathbf{k}(\mathbf{s})^\top \boldsymbol{\alpha}(t+1) = \mathbf{k}(\mathbf{s})^\top \boldsymbol{\alpha}(t) + \eta(\mathbf{s}, t), \quad (2.1.10)$$

dove

$$\begin{aligned} \mathbf{k}(\mathbf{s}) &= (K(\mathbf{u}_1, \mathbf{s}), \dots, K(\mathbf{u}_m, \mathbf{s}))^\top, \\ \boldsymbol{\alpha}(t) &= (\alpha(\mathbf{u}_1, t), \dots, \alpha(\mathbf{u}_m, t))^\top. \end{aligned}$$

Passando ad una forma matriciale dell'equazione 2.1.10 ed esplicitando $\boldsymbol{\alpha}(t+1)$ si ottiene

$$\begin{aligned} \mathbf{K}\boldsymbol{\alpha}(t+1) &= \mathbf{K}\boldsymbol{\alpha}(t) + \boldsymbol{\eta}(t), \\ \boldsymbol{\alpha}(t+1) &= \boldsymbol{\alpha}(t) + (\mathbf{K}^\top \mathbf{K})^{-1} \mathbf{K}^\top \boldsymbol{\eta}(t), \end{aligned} \quad (2.1.11)$$

dove

$$\begin{aligned} \mathbf{K} &= (\mathbf{k}(\mathbf{s}_1)^\top, \dots, \mathbf{k}(\mathbf{s}_n)^\top), \\ \boldsymbol{\eta}(t) &= (\eta(\mathbf{s}_1, t), \dots, \eta(\mathbf{s}_n, t))^\top. \end{aligned}$$

Sostituendo i risultati ottenuti nell'equazione 2.1.1 e definendo $\mathbf{R} = (\mathbf{K}^\top \mathbf{K})^{-1} \mathbf{K}^\top$ otteniamo il modello dinamico spazio-temporale

$$\begin{aligned} Y(\mathbf{s}, t) &= \mathbf{k}(\mathbf{s})^\top \boldsymbol{\alpha}(t) + \varepsilon(\mathbf{s}, t), \\ \boldsymbol{\alpha}(t+1) &= \boldsymbol{\alpha}(t) + \mathbf{R} \boldsymbol{\eta}(t), \end{aligned} \quad (2.1.12)$$

e passando poi alla forma matriciale:

$$\begin{aligned} \mathbf{Y}(t) &= \mathbf{K} \boldsymbol{\alpha}(t) + \boldsymbol{\varepsilon}(t), \\ \boldsymbol{\alpha}(t+1) &= \boldsymbol{\alpha}(t) + \mathbf{R} \boldsymbol{\eta}(t), \end{aligned} \quad (2.1.13)$$

dove

$$\begin{aligned} \mathbf{Y}(t) &= (Y(\mathbf{s}_1, t), \dots, Y(\mathbf{s}_n, t))^\top, \\ \boldsymbol{\varepsilon}(t) &= (\varepsilon(\mathbf{s}_1, t), \dots, \varepsilon(\mathbf{s}_n, t))^\top. \end{aligned}$$

Con le equazioni 2.1.13 è stata caratterizzata completamente la dinamica temporale del processo. Per completare la specificazione assumiamo distribuzioni gaussiane per i termini di errore

$$\begin{aligned} \boldsymbol{\varepsilon}(t) &\sim \mathcal{N}(\mathbf{0}, \mathbf{H}), \quad \mathbf{H} = \mathbb{E}[\boldsymbol{\varepsilon}(t) \boldsymbol{\varepsilon}(t)^\top], \\ \boldsymbol{\eta}(t) &\sim \mathcal{N}(\mathbf{0}, \mathbf{Q}), \quad \mathbf{Q} = \mathbb{E}[\boldsymbol{\eta}(t) \boldsymbol{\eta}(t)^\top], \end{aligned} \quad (2.1.14)$$

e definiamo una legge iniziale per il processo latente

$$\boldsymbol{\alpha}(0) \sim \mathcal{N}(\mathbf{0}, \mathbf{A}_0), \quad \mathbf{A}_0 = \mathbb{E}[\boldsymbol{\alpha}(0) \boldsymbol{\alpha}(0)^\top]. \quad (2.1.15)$$

Le matrici *time-invariant* \mathbf{H} e \mathbf{Q} determinano la struttura di dipendenza spaziale dei disturbi e vanno dunque parametrizzate in accordo con un'opportuna funzione di covarianza.

Nelle applicazioni pratiche di rado si hanno informazioni puntuali sulla distribuzione iniziale del segnale latente. Per riflettere questa scarsa informazione a priori sul modello usualmente si pone $\mathbf{A}_0 = k \mathbf{I}_m$ con $k \rightarrow \infty$, o, più operativamente, fissando per k un valore finito, ma molto grande e in questo caso si parla di inizializzazione diffusa.

Il modello così formulato rientra nell'ampia classe dei modelli gerarchici per l'analisi di dati spazio-temporali, la cui formulazione generale è stata esaustivamente trattata nella monografia di Cressie e Wikle 2011.

2.2 Rappresentazione *State Space*

La formulazione appena descritta è riconducibile all'ampia classe dei modelli *State Space* ed in particolare dei modelli lineari e Gaussiani, la cui specificazione generale è riassumibile attraverso le seguenti equazioni strutturali

$$\mathbf{y}_t = \mathbf{c}_t + \mathbf{Z}_t \cdot \boldsymbol{\alpha}_t + \mathbf{G}_t \cdot \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{H}_t) \quad (2.2.1)$$

$$\boldsymbol{\alpha}_{t+1} = \mathbf{d}_t + \mathbf{T}_t \cdot \boldsymbol{\alpha}_t + \mathbf{R}_t \cdot \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_t) \quad (2.2.2)$$

$$\boldsymbol{\alpha}_0 \sim \mathcal{N}(\mathbf{a}_0, \mathbf{P}_0) \quad (2.2.3)$$

dove 2.2.1 si dice equazione di misura e collega le osservazioni \mathbf{y}_t con il vettore di stati latenti $\boldsymbol{\alpha}_t$, mediante la matrice di misura \mathbf{Z}_t e il vettore di posizione \mathbf{c}_t .

L'equazione 2.2.2 è detta di transizione, o di stato, e descrive la dinamica auto-regressiva del processo latente, dove \mathbf{d}_t è un vettore di traslazione e \mathbf{T}_t una matrice di transizione.

Le matrici \mathbf{H}_t e \mathbf{Q}_t sono rispettivamente le varianze dei disturbi di misura e di stato, mentre \mathbf{G}_t ed \mathbf{R}_t costituiscono matrici di scala o di selezione sui medesimi disturbi.

Completano la specificazione del modello le condizioni iniziali sul fattore latente descritte in 2.2.3.

Assumendo che $\{\mathbf{y}_t\}_{t=1}^T$ siano un insieme di realizzazioni del processo ($\mathbf{Y}(t)$, $t \in \mathbf{T}$), campionato su n siti distinti ($\mathbf{s}_1, \dots, \mathbf{s}_n$), e ridefinendo le matrici del modello

State Space come in 2.2.4 otteniamo esattamente la specificazione del modello spazio-temporale descritta nella sezione precedente:

$$\begin{aligned} \mathbf{c}_t &= \mathbf{0}, & \mathbf{Z}_t &= \mathbf{K}, & \mathbf{G}_t &= \mathbf{I}_n, & \mathbf{H}_t &= \mathbf{H}, \\ \mathbf{d}_t &= \mathbf{0}, & \mathbf{T}_t &= \mathbf{I}_m, & \mathbf{R}_t &= \mathbf{R}, & \mathbf{Q}_t &= \mathbf{Q}, \\ \mathbf{a}_0 &= \mathbf{0}, & \mathbf{P}_0 &= k \mathbf{I}_m. \end{aligned} \quad (2.2.4)$$

La rappresentazione *State Space* è un flessibile strumento che permette di generalizzare facilmente il modello inserendo dinamiche temporali aggiuntive, come *trend*, periodicità, e componenti autoregressive. Per farlo è sufficiente ampliare a blocchi il vettore di stati e le quantità ad esso collegate. Questo aspetto rende particolarmente interessante questa classe di modelli, che negli ultimi anni ha riscosso sempre più successo e ha trovato applicazione in un numero di problemi sempre maggiore.

Fattorizzando il vettore di stati in k componenti $\boldsymbol{\alpha}_t = (\boldsymbol{\alpha}_{1,t}, \dots, \boldsymbol{\alpha}_{k,t})$ e ridefinendo le matrici del modello come in 2.2.5

$$\begin{aligned} \mathbf{d}_t &= (\mathbf{d}_{1,t}^\top, \dots, \mathbf{d}_{k,t}^\top)^\top, & \mathbf{Z}_t &= (\mathbf{Z}_{1,t}, \dots, \mathbf{Z}_{k,t}), \\ \mathbf{T}_t &= \text{diag}(\mathbf{T}_{1,t}, \dots, \mathbf{T}_{k,t}), & \mathbf{R}_t &= \text{diag}(\mathbf{R}_{1,t}, \dots, \mathbf{R}_{k,t}), \\ \mathbf{Q}_t &= \text{diag}(\mathbf{Q}_{1,t}, \dots, \mathbf{Q}_{k,t}), & \mathbf{P}_0 &= \text{diag}(\mathbf{P}_{1,0}, \dots, \mathbf{P}_{k,0}), \end{aligned} \quad (2.2.5)$$

dove $\mathbf{d}_{1,t} = \mathbf{0}$, $\mathbf{Z}_{1,t} = \mathbf{K}$, $\mathbf{T}_{1,t} = \mathbf{I}_m$, $\mathbf{R}_{1,t} = \mathbf{R}$, $\mathbf{Q}_{1,t} = \mathbf{Q}$ e $\mathbf{P}_{1,0} = k \mathbf{I}_m$, otteniamo un modello esteso, nel quale viene preservata e generalizzata la dinamica spazio-temporale descritta in 2.1.13. Il primo blocco del vettore di stati $\boldsymbol{\alpha}_{1,t}$ permette di specificare il processo di larga scala coerentemente con 2.1.13, mentre le restanti $k - 1$ componenti del vettore possono essere gestite in maniera tale da modellare altre caratteristiche dei dati (come *trend* e periodicità) attraverso una costruzione modulare del modello.

2.3 Estrazione del segnale

Supponendo note e fissate le quantità \mathbf{c}_t , \mathbf{Z}_t , \mathbf{G}_t , \mathbf{H}_t , \mathbf{d}_t , \mathbf{T}_t , \mathbf{R}_t , \mathbf{Q}_t , \mathbf{a}_0 e \mathbf{P}_0 per ogni t , il vettore degli stati $\{\boldsymbol{\alpha}_t\}_{t=0}^T$ ha distribuzione a posteriori normale multivariata.

Di particolare interesse a fini pratici sono le distribuzioni filtrate, lisciate e di previsione del processo, rispettivamente, $p(\boldsymbol{\alpha}_t | \mathbf{y}_1, \dots, \mathbf{y}_t)$, $p(\boldsymbol{\alpha}_t | \mathbf{y}_1, \dots, \mathbf{y}_T)$ e $p(\boldsymbol{\alpha}_{T+k} | \mathbf{y}_1, \dots, \mathbf{y}_T)$. Sotto le usuali assunzioni distributive Gaussianhe per gli stati e le osservazioni, queste leggi a posteriori risultano normali multivariate e sono derivabili in forma chiusa.

Kalman filter e Kalman smoother

Facendo uso del celebre algoritmo del filtro di Kalman, introdotto per la prima volta da Kalman e Bucy 1961, è possibile estrarre i primi due momenti della distribuzione filtrata $p(\boldsymbol{\alpha}_t | \mathbf{y}_1, \dots, \mathbf{y}_t)$ attraverso l'esecuzione sequenziale dei passi descritti nelle equazioni 2.3.1 per $t = 1, \dots, T$:

$$\begin{aligned} \mathbf{v}_t &= \mathbf{y}_t - \mathbf{c}_t - \mathbf{Z}_t \mathbf{a}_t, & \mathbf{F}_t &= \mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t^\top + \mathbf{G}_t \mathbf{H}_t \mathbf{G}_t^\top, \\ \mathbf{a}_{t|t} &= \mathbf{a}_t + \mathbf{P}_t \mathbf{Z}_t^\top \mathbf{F}_t^{-1} \mathbf{v}_t, & \mathbf{P}_{t|t} &= \mathbf{P}_t - \mathbf{P}_t \mathbf{Z}_t^\top \mathbf{F}_t^{-1} \mathbf{Z}_t \mathbf{P}_t, \\ \mathbf{a}_{t+1} &= \mathbf{d}_t + \mathbf{T}_t \mathbf{a}_{t|t}, & \mathbf{P}_{t+1} &= \mathbf{T}_t \mathbf{P}_{t|t} \mathbf{T}_t^\top + \mathbf{R}_t \mathbf{H}_t \mathbf{R}_t^\top, \end{aligned} \quad (2.3.1)$$

dove le quantità \mathbf{v}_t , \mathbf{a}_t , $\mathbf{a}_{t|t}$, \mathbf{F}_t , \mathbf{P}_t e $\mathbf{P}_{t|t}$ sono definite nella tabella 2.1.

Sotto l'ipotesi di distribuzione normale, i valori filtrati restituiti delle equazioni 2.3.1 coincidono con le stime di massima verosimiglianza e con il valore di massimo della legge a posteriori del processo latente. Inoltre, da un punto di vista frequentista, queste godono di proprietà ottimali tra gli stimatori lineari e non distorti. Per maggiori dettagli si vedano Durbin e Koopman 2012, Simon 2006 e Anderson e Moore 1979.

L'esecuzione del filtro di Kalman permette inoltre di marginalizzare analiticamente gli stati latenti dalla distribuzione congiunta $p(\{\mathbf{y}_t\}_{t=1}^T, \{\boldsymbol{\alpha}_t\}_{t=1}^T)$ e dunque di ottenere in forma chiusa la verosimiglianza marginale del modello, pari a:

$$\begin{aligned} \mathcal{L}(\{\mathbf{y}_t\}_{t=1}^T) &= \int_{\mathbb{R}^{m+T}} \left(p(\boldsymbol{\alpha}_0) \prod_{t=0}^{T-1} p(\boldsymbol{\alpha}_{t+1} | \boldsymbol{\alpha}_t) \right) \left(\prod_{t=1}^T p(\mathbf{y}_t | \boldsymbol{\alpha}_t) \right) d\boldsymbol{\alpha}_0 \dots d\boldsymbol{\alpha}_T \\ &= (2\pi)^{-\frac{1}{2}nT} \exp \left\{ -\frac{1}{2} \sum_{t=1}^T \log |\mathbf{F}_t| - \frac{1}{2} \sum_{t=1}^T \mathbf{v}_t^\top \mathbf{F}_t^{-1} \mathbf{v}_t \right\}. \end{aligned} \quad (2.3.2)$$

La verosimiglianza 2.3.2 riveste un ruolo di primaria importanza per la stima dei parametri di un modello *State Space*, e costituisce la base dei moderni metodi di inferenza, sia in contesto di inferenza Bayesiano che frequentista.

Eseguita una prima fase di filtraggio della serie, i momenti della distribuzione predittiva possono essere ottenuti mediante l'aggiornamento k passi in avanti del filtro di Kalman, risolvendo sequenzialmente le equazioni in 2.3.3:

$$\begin{aligned} \tilde{\mathbf{v}}_\ell &= \mathbf{0}, & \tilde{\mathbf{F}}_\ell &= \mathbf{Z}_\ell \tilde{\mathbf{P}}_\ell \mathbf{Z}_\ell^\top + \mathbf{G}_\ell \mathbf{H}_\ell \mathbf{G}_\ell^\top, \\ \tilde{\mathbf{K}}_\ell &= \mathbf{T}_\ell \tilde{\mathbf{P}}_\ell \mathbf{Z}_\ell^\top \tilde{\mathbf{F}}_\ell^{-1}, & \tilde{\mathbf{L}}_\ell &= \mathbf{T}_\ell - \tilde{\mathbf{K}}_\ell \mathbf{Z}_\ell, \\ \tilde{\mathbf{a}}_{\ell+1} &= \mathbf{d}_\ell + \mathbf{T}_\ell \tilde{\mathbf{a}}_\ell, & \tilde{\mathbf{P}}_{\ell+1} &= \mathbf{T}_\ell \tilde{\mathbf{P}}_\ell \mathbf{L}_\ell^\top + \mathbf{R}_\ell \mathbf{H}_\ell \mathbf{R}_\ell^\top, \end{aligned} \quad (2.3.3)$$

dove $\ell = T + k$, con $k > 0$ e dove il primo passo di previsione viene inizializzato ponendo $\tilde{\mathbf{a}}_{T+1} = \mathbf{a}_{T+1}$ e $\tilde{\mathbf{P}}_{T+1} = \mathbf{P}_{T+1}$, che sono i momenti della distribuzione

predittiva un tempo in avanti forniti da una preventiva fase di filtraggio dei dati. Per la notazione si faccia riferimento alla tabella 2.1.

A partire da media e varianza della distribuzione filtrata del processo, è possibile ottenere media e varianza della distribuzione lisciata di $\boldsymbol{\alpha}_t$ applicando a ritroso le regole di aggiornamento del *Kalman smoother* descritte in 2.3.4 per $t = T, \dots, 1$:

$$\begin{aligned} \mathbf{K}_t &= \mathbf{T}_t \mathbf{P}_t \mathbf{Z}_t^\top \mathbf{F}_t^{-1}, & \mathbf{L}_t &= \mathbf{T}_t - \mathbf{K}_t \mathbf{Z}_t^\top, \\ \mathbf{r}_{t-1} &= \mathbf{Z}_t \mathbf{F}_t^{-1} \mathbf{v}_t + \mathbf{L}_t^\top \mathbf{r}_t, & \mathbf{N}_{t-1} &= \mathbf{Z}_t^\top \mathbf{F}_t^{-1} \mathbf{Z}_t + \mathbf{L}_t^\top \mathbf{N}_t \mathbf{L}_t, \\ \hat{\mathbf{a}}_t &= \mathbf{a}_t + \mathbf{P}_t \mathbf{r}_{t-1}, & \mathbf{V}_t &= \mathbf{P}_t - \mathbf{P}_t \mathbf{N}_{t-1} \mathbf{P}_t, \end{aligned} \quad (2.3.4)$$

dove $\mathbf{r}_T = \mathbf{0}$ e $\mathbf{N}_T = \mathbf{O}$.

Tabella 2.1: Vettori e matrici di aggiornamento per *Kalman filter* e *Kalman smoother*

Vettori		Matrici	
\mathbf{v}_t	$\mathbf{y}_t - \mathbb{E} \left[\mathbf{y}_t \mid \{\mathbf{y}_j\}_{j=1}^{t-1} \right]$	\mathbf{F}_t	$\text{Var} \left(\mathbf{v}_t \mid \{\mathbf{y}_j\}_{j=1}^{t-1} \right)$
\mathbf{a}_t	$\mathbb{E} \left[\boldsymbol{\alpha}_t \mid \{\mathbf{y}_j\}_{j=1}^{t-1} \right]$	\mathbf{P}_t	$\text{Var} \left(\boldsymbol{\alpha}_t \mid \{\mathbf{y}_j\}_{j=1}^{t-1} \right)$
$\mathbf{a}_{t t}$	$\mathbb{E} \left[\boldsymbol{\alpha}_t \mid \{\mathbf{y}_j\}_{j=1}^t \right]$	$\mathbf{P}_{t t}$	$\text{Var} \left(\boldsymbol{\alpha}_t \mid \{\mathbf{y}_j\}_{j=1}^t \right)$
$\tilde{\mathbf{v}}_\ell$	$\mathbb{E} \left[\mathbf{v}_\ell \mid \{\mathbf{y}_t\}_{t=1}^T \right]$	$\tilde{\mathbf{F}}_\ell$	$\text{Var} \left(\mathbf{v}_\ell \mid \{\mathbf{y}_t\}_{t=1}^T \right)$
$\tilde{\mathbf{a}}_\ell$	$\mathbb{E} \left[\boldsymbol{\alpha}_\ell \mid \{\mathbf{y}_t\}_{t=1}^T \right]$	$\tilde{\mathbf{P}}_\ell$	$\text{Var} \left(\boldsymbol{\alpha}_\ell \mid \{\mathbf{y}_t\}_{t=1}^T \right)$
\mathbf{r}_t		\mathbf{N}_t	$\text{Var} \left(\mathbf{r}_t \mid \{\mathbf{y}_t\}_{t=1}^T \right)$
$\hat{\mathbf{a}}_t$	$\mathbb{E} \left[\boldsymbol{\alpha}_t \mid \{\mathbf{y}_t\}_{t=1}^T \right]$	\mathbf{V}_t	$\text{Var} \left(\boldsymbol{\alpha}_t \mid \{\mathbf{y}_t\}_{t=1}^T \right)$

Le equazioni ricorsive 2.3.1, 2.3.3 e 2.3.4 per derivare le distribuzioni a posteriori del segnale latente sono state esaustivamente discusse nei lavori di Durbin e Koopman 2012, Brockwell e Davis 2016 e Simon 2006. La teoria collegata al filtro di Kalman è stata poi ampiamente studiata ed estesa negli ultimi decenni sia in ambito statistico, che ingegneristico, ad esempio nei lavori di Koopman 1993, Koopman 1997, Koopman e Durbin 2000, Durbin e Koopman 2002, Koopman e Durbin 2003.

Simulation smoother

L'esecuzione sequenziale di *Kalman filter* (*forward step*) e *Kalman smoother* (*backward step*) fornisce la distribuzione a posteriori degli stati, condizionatamente alle matrici del modello, supposte note. Questa, in ottica Bayesiana, corrisponde alla distribuzione *full conditional* degli stati latenti e può essere efficientemente utilizzata

per fare inferenza a posteriori sul modello, o mediante algoritmi MCMC, o mediante metodi variazionali Bayesiani.

Alla base dei metodi MCMC per modelli *State Space* esistono vari algoritmi di simulazione dalla distribuzione *full conditional* degli stati latenti. In questa tesi considereremo l'efficiente metodo di *Simulation Smoothing* proposto da Durbin e Koopman 2002, ma in letteratura esistono anche altre possibilità, come l'algoritmo di *forward filtering backward sampling* proposto da Carter e Kohn 1994, e ampiamente discusso in Carter 1994.

A partire dalla distribuzione lisciata del processo, ottenibile dopo l'esecuzione di *Kalman filter* e *Kalman smoother*, l'algoritmo di *Simulation Smoothing* di Durbin e Koopman procede secondo i seguenti passi:

- Si generi $\boldsymbol{\omega}^+$ dalla distribuzione congiunta $p(\boldsymbol{\omega}) \propto \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$, dove $\boldsymbol{\omega} = (\boldsymbol{\varepsilon}_1^\top, \boldsymbol{\eta}_1^\top, \dots, \boldsymbol{\varepsilon}_T^\top, \boldsymbol{\eta}_T^\top)^\top$ e $\boldsymbol{\Omega} = \text{diag}(\mathbf{H}_1, \mathbf{Q}_1, \dots, \mathbf{H}_T, \mathbf{Q}_T)$.
- Si utilizzi $\boldsymbol{\omega}^+$ per generare $\boldsymbol{\alpha}^+$ e \mathbf{y}^+ , applicando iterativamente le equazioni 2.2.1 e 2.2.2, dopo aver inizializzato $\boldsymbol{\alpha}_0^+$ dalla distribuzione 2.2.3.
- Si applichino *Kalman filter* e *Kalman smoother* sulle serie \mathbf{y} (originale) e \mathbf{y}^+ (simulata) per estrarre le medie di lisciamento $\hat{\mathbf{a}} = \mathbb{E}[\boldsymbol{\alpha} | \mathbf{y}]$ e $\hat{\mathbf{a}}^+ = \mathbb{E}[\boldsymbol{\alpha}^+ | \mathbf{y}^+]$
- Si applichi la correzione in media $\tilde{\boldsymbol{\alpha}} = \hat{\mathbf{a}} + (\boldsymbol{\alpha}^+ - \hat{\mathbf{a}}^+)$, dove i valori in $\tilde{\boldsymbol{\alpha}}$ costituiscono una traiettoria del processo generata dalla distribuzione *full conditional* degli stati latenti.

Per maggiori dettagli su questo algoritmo di *Simulation Smoothing* e una presentazione di altri metodi MCMC per modelli *State Space* presenti in letteratura rimandiamo a Durbin e Koopman 2012.

2.4 Inferenza a posteriori

Nelle applicazioni reali tipicamente le matrici del modello non sono completamente note e dipendono da un vettore di parametri incogniti.

Adottando la specificazione spazio-temporale 2.1.13 e definendo le matrici del modello in accordo con 2.2.4, le quantità \mathbf{c}_t , \mathbf{Z}_t , \mathbf{G}_t , \mathbf{d}_t , \mathbf{T}_t e \mathbf{R}_t risultano completamente note, mentre \mathbf{H}_t e \mathbf{Q}_t vanno adeguatamente parametrizzate.

Nella letteratura dei modelli *State Space* applicati a dati spazio-temporali è pratica comune (si vedano, ad esempio, Mardia et al. 1998, Wikle e Cressie 1999, Stroud, Müller e Sansó 2001, Wikle, Berliner e Cressie 2004 e Sahu e Mardia 2005) assumere

per le varianze dei disturbi la seguente struttura:

$$\mathbf{H} = \sigma_\varepsilon^2 \mathbf{V}, \quad \mathbf{Q} = \sigma_\eta^2 \mathbf{W}^*,$$

dove $\sigma_\varepsilon^2 > 0$ e $\sigma_\eta^2 > 0$ sono parametri da determinare, mentre \mathbf{V} e \mathbf{W}^* sono matrici di correlazione spaziale stazionarie, omogenee nel tempo, note e fissate.

Completiamo la specificazione Bayesiana del modello assumendo distribuzioni a priori coniugate indipendenti per i parametri σ_ε^2 e σ_η^2 :

$$\pi(\sigma_\varepsilon^2, \sigma_\eta^2) = \pi(\sigma_\varepsilon^2) \pi(\sigma_\eta^2), \quad (2.4.1)$$

con $\pi(\sigma_\varepsilon^2) = \mathcal{IG}(\sigma_\varepsilon^2 | a_\varepsilon^0, b_\varepsilon^0)$ e $\pi(\sigma_\eta^2) = \mathcal{IG}(\sigma_\eta^2 | a_\eta^0, b_\eta^0)$, e dove $\mathcal{IG}(\cdot | a, b)$ indica la distribuzione Gamma Inversa di parametri a e b . La scelta di distribuzioni coniugate è alquanto comune, e permette di ottenere delle espressioni in forma chiusa per le distribuzioni *full conditional* dei parametri a posteriori, si vedano, ad esempio, i lavori di Wikle e Cressie 1999 e Stroud, Müller e Sansó 2001.

Dalle assunzioni sul modello, possiamo costruire la funzione di densità congiunta per i parametri e per il fattore latente componendo le distribuzioni a priori con la verosimiglianza come segue

$$\begin{aligned} \pi(\sigma_\varepsilon^2, \sigma_\eta^2, \{\boldsymbol{\alpha}_t\}_{t=0}^T | \{\mathbf{y}_t\}_{t=1}^T) &\propto \\ &\propto \pi(\sigma_\varepsilon^2) \pi(\sigma_\eta^2) p(\{\boldsymbol{\alpha}_t\}_{t=0}^T | \sigma_\eta^2) \mathcal{L}(\sigma_\varepsilon^2, \sigma_\eta^2, \{\boldsymbol{\alpha}_t\}_{t=0}^T | \{\mathbf{y}_t\}_{t=1}^T) \\ &\propto \pi(\sigma_\varepsilon^2) \pi(\sigma_\eta^2) \left(p(\boldsymbol{\alpha}_0) \prod_{t=1}^{T-1} p(\boldsymbol{\alpha}_{t+1} | \boldsymbol{\alpha}_t, \sigma_\eta^2) \right) \left(\prod_{t=1}^T p(\mathbf{y}_t | \boldsymbol{\alpha}_t, \sigma_\varepsilon^2) \right), \end{aligned}$$

dove

$$\begin{aligned} p(\mathbf{y}_t | \boldsymbol{\alpha}_t, \sigma_\varepsilon^2) &= \mathcal{N}(\mathbf{y}_t | \mathbf{K}\boldsymbol{\alpha}_t, \sigma_\varepsilon^2 \mathbf{V}), \\ p(\boldsymbol{\alpha}_{t+1} | \boldsymbol{\alpha}_t, \sigma_\eta^2) &= \mathcal{N}(\boldsymbol{\alpha}_{t+1} | \boldsymbol{\alpha}_t, \sigma_\eta^2 \mathbf{R} \mathbf{W}^* \mathbf{R}^\top). \end{aligned}$$

Esplicitando le densità dei parametri, del fattore latente e dei dati otteniamo

$$\begin{aligned} \pi(\sigma_\varepsilon^2, \sigma_\eta^2, \{\boldsymbol{\alpha}_t\}_{t=0}^T | \{\mathbf{y}_t\}_{t=1}^T) &\propto \\ &\propto (\sigma_\varepsilon^2)^{-(a_\varepsilon^0+1)} (\sigma_\eta^2)^{-(a_\eta^0+1)} \exp \left\{ -\frac{b_\eta^0}{\sigma_\eta^2} - \frac{b_\varepsilon^0}{\sigma_\varepsilon^2} - \frac{1}{2k} \|\boldsymbol{\alpha}_0\|_2^2 \right\} \\ &\times (\sigma_\eta^2)^{-\frac{m}{2}(T-1)} \exp \left\{ -\frac{1}{2\sigma_\eta^2} \sum_{t=1}^{T-1} \left\| \mathbf{W}^{-\frac{1}{2}} (\boldsymbol{\alpha}_{t+1} - \boldsymbol{\alpha}_t) \right\|_2^2 \right\} \\ &\times (\sigma_\varepsilon^2)^{-\frac{n}{2}T} \exp \left\{ -\frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^T \left\| \mathbf{V}^{-\frac{1}{2}} (\mathbf{y}_t - \mathbf{K}\boldsymbol{\alpha}_t) \right\|_2^2 \right\}, \end{aligned} \quad (2.4.2)$$

dove $\mathbf{W} = \mathbf{R}\mathbf{W}^*\mathbf{R}^\top$ e $\|\mathbf{A}^{-\frac{1}{2}}(\mathbf{u} - \mathbf{v})\|_2^2 = (\mathbf{u} - \mathbf{v})^\top \mathbf{A}^{-1}(\mathbf{u} - \mathbf{v})$ denota la norma indotta dalla distanza di Mahalanobis.

Dalla densità a posteriori non è possibile ottenere direttamente una stima in forma chiusa per i parametri ed il segnale latente, dunque è necessario ricorrere ad algoritmi MCMC per fare inferenza. In questo caso si può adottare un *Gibbs Sampling* come efficiente metodo di simulazione, sfruttando così l'esistenza di leggi a priori coniugate per i parametri.

L'algoritmo di *Gibbs Sampling* tradizionale prevede di estrarre iterativamente valori (pseudo-)casuali dalle distribuzioni *full conditional* dei parametri del modello, in modo tale che la successione di vettori così simulati generi un catena di Markov ergodica, la cui distribuzione stazionaria converga alla legge a posteriori obiettivo (Geman e Geman 1984).

Per costruire un algoritmo di *Gibbs Sampling* è quindi necessario identificare le leggi *full conditional* per i parametri del modello, che nel nostro caso sono σ_ε^2 , σ_η^2 e $\{\boldsymbol{\alpha}_t\}_{t=0}^T$.

Come abbiamo visto, la distribuzione a posteriori condizionata di $\{\boldsymbol{\alpha}_t\}_{t=0}^T$ è ottenibile analiticamente attraverso l'esecuzione degli algoritmi di filtraggio e lisciamiento discussi nella sezione 2.3. Per la derivazione delle altre distribuzioni *full conditional* si rimanda all'appendice A.1.

Trovate le distribuzioni *full conditional* per i parametri, l'algoritmo di simulazione procede secondo lo schema iterativo nel modo seguente.

Fissati i valori iniziali $(\sigma_\varepsilon^2)^{(0)}$ e $(\sigma_\eta^2)^{(0)}$, si generano, condizionatamente a questi, le traiettorie iniziali degli stati latenti $\{\boldsymbol{\alpha}_t^{(0)}\}_{t=0}^T$ adottando l'algoritmo di *Simulation Smoothing* descritto nella sottosezione 2.3. Poi le simulazioni dalla distribuzione a posteriori alla $(k+1)$ -esima iterazione sono ottenute eseguendo sequenzialmente i seguenti passi di aggiornamento e simulazione:

1. Si generi $(\sigma_\eta^2)^{(k+1)}$ da $\pi(\sigma_\eta^2 | \{\boldsymbol{\alpha}_t\}_{t=0}^T) \propto \mathcal{IG}(a_\eta^*, b_\eta^*)$, dove:

$$a_\eta^* = a_\eta^0 + \frac{m}{2}(T-1), \quad b_\eta^* = b_\eta^0 + \frac{1}{2} \sum_{t=1}^{T-1} \left\| \mathbf{W}^{-\frac{1}{2}} (\boldsymbol{\alpha}_{t+1}^{(k)} - \boldsymbol{\alpha}_t^{(k)}) \right\|_2^2.$$

2. Si generi $(\sigma_\varepsilon^2)^{(k+1)}$ da $\pi(\sigma_\varepsilon^2 | \{\boldsymbol{\alpha}_t\}_{t=0}^T, \{\mathbf{y}_t\}_{t=1}^T) \propto \mathcal{IG}(a_\varepsilon^*, b_\varepsilon^*)$, dove:

$$a_\varepsilon^* = a_\varepsilon^0 + \frac{n}{2}T, \quad b_\varepsilon^* = b_\varepsilon^0 + \frac{1}{2} \sum_{t=1}^T \left\| \mathbf{V}^{-\frac{1}{2}} (\mathbf{y}_t - \mathbf{K}\boldsymbol{\alpha}_t^{(k)}) \right\|_2^2.$$

3. Per ogni $t = 0, \dots, T$, si generi $\boldsymbol{\alpha}_t^{(k+1)}$ da $\pi(\boldsymbol{\alpha}_t | \sigma_\eta^2, \{\mathbf{y}\}_{t=1}^T)$. Applicando in sequenza *Kalman filter* e *Kalman smoother* si derivano i momenti della distri-

buzione *full conditional* degli stati latenti. Successivamente è possibile simulare congiuntamente le traiettorie di tutti gli stati latenti usando l'algoritmo di *Simulation Smoothing* proposto da Durbin e Koopman 2002.

L'algoritmo di *Gibbs Sampling* appena descritto permette di generare valori casuali da una catena di Markov la cui distribuzione stazionaria converge alla distribuzione a posteriori congiunta $\pi\left(\sigma_\eta^2, \sigma_\varepsilon^2, \{\boldsymbol{\alpha}_t\}_{t=0}^T \mid \{\mathbf{y}_t\}_{t=1}^T\right)$.

2.5 Previsione nel tempo

Dopo la stima dei parametri, il secondo grande problema inferenziale a cui tipicamente si vuole rispondere quanto si specifica un modello *State Space* è quello predittivo.

Da un punto di vista Bayesiano fare previsioni avanti nel tempo sul processo \mathbf{y}_t equivale a calcolarne la distribuzione predittiva a posteriori, marginalizzando gli effetti dei parametri e degli stati latenti.

Ipotizziamo di voler prevedere il comportamento di \mathbf{y}_t per k unità di tempo successive a T e definiamo il processo osservato e il processo predetto come, rispettivamente, $\mathbf{y}^{oss} = \{\mathbf{y}_t\}_{t=1}^T$ e $\mathbf{y} = \{\mathbf{y}_t\}_{t=T+1}^{T+k}$. Indichiamo poi con $\boldsymbol{\theta} = (\sigma_\varepsilon^2, \sigma_\eta^2)$ il vettore di parametri del modello e con $\boldsymbol{\alpha} = \{\boldsymbol{\alpha}_t\}_{t=T+1}^{T+k}$ l'insieme degli stati latenti sull'insieme di tempi *out of sample*, definiti rispettivamente sopra lo spazio parametrico Θ e sopra lo spazio degli stati \mathbf{A}^k .

Per mezzo del principio di condizionamento rispetto ai parametri e agli stati latenti, possiamo esprimere la distribuzione predittiva del processo come

$$p(\mathbf{y} \mid \mathbf{y}^{oss}) = \iint_{\Theta \times \mathbf{A}^k} p(\mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\alpha} \mid \mathbf{y}^{oss}) d\boldsymbol{\theta} d\boldsymbol{\alpha} \quad (2.5.1)$$

$$= \iint_{\Theta \times \mathbf{A}^k} p(\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\alpha}) \pi(\boldsymbol{\theta}, \boldsymbol{\alpha} \mid \mathbf{y}^{oss}) d\boldsymbol{\theta} d\boldsymbol{\alpha}, \quad (2.5.2)$$

dove, passando da 2.5.1 a 2.5.2, si è sfruttata la proprietà di indipendenza condizionata tra il processo predetto e quello osservato, ovvero $p(\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\alpha}) = p(\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{y}^{oss})$.

L'integrale 2.5.2 il più delle volte non è risolvibile analiticamente, dunque è necessario trovarne una stima in forma approssimata.

Ipotizzando di aver già eseguito una prima volta R iterazioni dell'algoritmo di *Gibbs Sampling*, descritto nella sezione precedente 2.4, e di aver salvato in memoria i risultati della simulazione, possiamo sfruttare tali valori per produrre una stima di 2.5.2.

Indichiamo con $\{(\boldsymbol{\theta}^{(r)}, \boldsymbol{\alpha}_0^{(r)}, \dots, \boldsymbol{\alpha}_T^{(r)})\}_{r=1}^R$ l'insieme di valori simulati, ottenuti dopo una prima fase di *burnin* della catena. Per ognuna di queste R tuple possiamo utilizzare le equazioni del modello, e quindi il filtro di Kalman, per effettuare k passi di previsione avanti nel tempo, ottenendo così le distribuzioni predittive per ogni singola traiettoria simulata. Disponendo dei risultati ottenuti dal filtro, si può effettuare una fase di lisciamento e successivamente una di simulazione dei processi latenti per $t = T + 1, \dots, T + k$, ottenendo così un campione simulato $\{(\boldsymbol{\alpha}_{T+1}^{(r)}, \dots, \boldsymbol{\alpha}_{T+k}^{(r)})\}_{r=1}^R$ degli stati per ogni replicazione del *Gibbs Sampling*.

Per R sufficientemente grande possiamo approssimare l'integrale 2.5.2 vai Monte Carlo, attraverso una semplice media aritmetica:

$$p(\mathbf{y} | \mathbf{y}^{oss}) \approx \frac{1}{R} \sum_{r=1}^R p(\mathbf{y} | \boldsymbol{\theta}^{(r)}, \boldsymbol{\alpha}^{(r)}).$$

Oltre alla distribuzione predittiva a posteriori possiamo produrre una stima puntuale del processo k passi in avanti calcolando il valore atteso a posteriori

$$\begin{aligned} \mathbb{E}[\mathbf{y} | \mathbf{y}^{oss}] &= \iint_{\Theta \times \mathbf{A}} \mathbb{E}[\mathbf{y} | \boldsymbol{\theta}, \boldsymbol{\alpha}] \pi(\boldsymbol{\theta}, \boldsymbol{\alpha} | \mathbf{y}^{oss}) d\boldsymbol{\theta} d\boldsymbol{\alpha} \\ &= \iint_{\Theta \times \mathbf{A}} (\mathbf{c} + \mathbf{Z} \cdot \boldsymbol{\alpha}) \pi(\boldsymbol{\theta}, \boldsymbol{\alpha} | \mathbf{y}^{oss}) d\boldsymbol{\theta} d\boldsymbol{\alpha} \end{aligned}$$

che può essere convenientemente stimato come

$$\mathbb{E}[\mathbf{y} | \mathbf{y}^{oss}] = \mathbf{c} + \mathbf{Z} \cdot \mathbb{E}[\boldsymbol{\alpha} | \mathbf{y}^{oss}] \approx \mathbf{c} + \mathbf{Z} \cdot \frac{1}{R} \sum_{r=1}^R \boldsymbol{\alpha}^{(r)}.$$

dove $\mathbf{c} = (\mathbf{c}_{T+1}^\top, \dots, \mathbf{c}_{T+k}^\top)^\top$ e $\mathbf{Z} = \text{diag}(\mathbf{Z}_{T+1}, \dots, \mathbf{Z}_{T+k})$.

Questo approccio predittivo risulta spesso molto oneroso, specialmente se R è grande, quindi tipicamente, se non è di interesse caratterizzare completamente la distribuzione predittiva, si possono produrre delle stime puntuali utilizzando solo un piccolo sottoinsieme di repliche del *Gibbs Sampler* su cui fare poi estrapolazione, simulazione e stima Monte Carlo.

2.6 Possibili estensioni

Il metodo di simulazione proposto in 2.4 può essere facilmente generalizzato anche a modelli più complessi. Consideriamo, ad esempio, il caso in cui le matrici $\mathbf{V} = \mathbf{V}(\boldsymbol{\gamma}_\varepsilon)$ e $\mathbf{W} = \mathbf{W}(\boldsymbol{\gamma}_\eta)$ siano parametrizzate, quindi parzialmente note, a meno di un vettore di parametri $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_\eta, \boldsymbol{\gamma}_\varepsilon)$.

Generalmente non è possibile determinare delle distribuzioni a priori coniugate per $(\gamma_\eta, \gamma_\varepsilon)$, e dunque un algoritmo di *Gibbs Sampling* tradizionale, come quello proposto, non è direttamente applicabile. Tuttavia è sempre possibile specificare delle a priori per i nuovi parametri e adottare un algoritmo di simulazione ibrido per generare valori casuali dalla distribuzione congiunta $\pi\left(\sigma_\eta^2, \gamma_\eta, \sigma_\varepsilon^2, \gamma_\varepsilon, \{\alpha_t\}_{t=0}^T \mid \{\mathbf{y}_t\}_{t=1}^T\right)$. Ad esempio, si potrebbe implementare un *Metropolis Hastings within Gibbs*, dove si mantiene lo schema generale del *Gibbs Sampler* per i parametri di scala e per il fattore latente, ma si introduce un passo addizionale di campionamento *Metropolis* per i parametri di correlazione.

Quest'approccio risulta talvolta molto oneroso dal punto di vista computazionale, soprattutto quando la dimensione campionaria è molto elevata. La calibrazione di un *Metropolis Hastings* spesso non è un'operazione banale e a volte anche gli algoritmi automatici più sofisticati possono trovare difficoltà a raggiungere la convergenza e ad avere un buon *mixing*.

Per evitare di dover costruire e calibrare un *Metropolis Hastings ad hoc*, alcuni autori consigliano come criterio operativo di stimare preventivamente i parametri di disturbo $(\gamma_\varepsilon, \gamma_\eta)$ sui dati, per poi fissarli e considerarli noti all'interno di un algoritmo di stima semplificato. Si vedano, ad esempio, Stroud, Müller e Sansó 2001 e Sahu e Mardia 2005.

Questo secondo metodo risulta semplice, intuitivo e poco oneroso. Di contro non fornisce garanzie matematiche sulle proprietà delle stime e non rispetta, né a livello teorico, né interpretativo, la formalizzazione tipica del paradigma inferenziale Bayesiano.

Capitolo 3

Modelli quantilici spazio-temporali

Nel precedente capitolo è stata presentata una flessibile classe di modelli *State Space* attraverso la quale è possibile fare inferenza e produrre previsioni per processi spazio-temporali Gaussiani.

In questo terzo capitolo andiamo ad estendere e generalizzare questi modelli per poter fare inferenza non sulla media del processo ma sui suoi quantili.

3.1 *Quantile regression*

Nel *framework* della *quantile regression* il quantile di ordine τ di una variabile dipendente Y è esprimibile come funzione di un insieme di covariate \mathbf{X} e viene indicato con $Q^\tau(\mathbf{x})$, dove:

$$Q^\tau(\mathbf{x}) = \left\{ q \in \mathbb{R} : \mathbb{P}(Y \leq q \mid \mathbf{X} = \mathbf{x}) = \tau \right\}.$$

In letteratura sono state proposte diverse specificazioni per la funzione quantile $Q^\tau(\mathbf{x})$, tra le quali la più comune è la forma lineare: $Q^\tau(\mathbf{x}) = \mathbf{x}^\top \boldsymbol{\beta}^\tau$.

Il problema di stima dei parametri $\boldsymbol{\beta}^\tau$ e più in generale della funzione $Q^\tau(\mathbf{x})$ è affrontabile sia da un punto di vista frequentista che Bayesiano. Nel primo caso Koenker e Bassett 1978 hanno dimostrato che questo può essere risolto attraverso la seguente minimizzazione:

$$Q^\tau(\cdot) = \arg \min_{Q^\tau \in \mathcal{Q}} \sum_{t=1}^T \rho_\tau(y_t - Q^\tau(\mathbf{x}_t)), \quad (3.1.1)$$

dove (y_t, \mathbf{x}_t) per $t = 1, \dots, T$ sono realizzazioni indipendenti ed identicamente distribuite e $\rho_\tau(y) = y \cdot (\tau - \mathbb{1}(y < 0))$ denota la funzione di perdita quantilica.

Il problema di stima della funzione quantile $Q^\tau(\mathbf{x})$ può essere affrontato anche mediante massima verosimiglianza, modellando la legge condizionata $Y|\mathbf{x}$ secondo una distribuzione non correttamente specificata Laplace Asimmetrica, che indicheremo $\mathcal{AL}(\tau, \mu, \sigma)$, la cui funzione di densità è data da:

$$\mathcal{AL}(y|\tau, \mu, \sigma) = \frac{\tau(1-\tau)}{\sigma} \exp\left\{-\frac{\rho_\tau(y-\mu)}{\sigma}\right\}, \quad (3.1.2)$$

dove $\tau \in (0, 1)$, $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}^+$. Una caratteristica interessante del modello distributivo $\mathcal{AL}(\tau, \mu, \sigma)$ è che impone l'esatta corrispondenza tra la funzione di regressione $\mu = Q^\tau(\mathbf{x})$ e il quantile di ordine τ di $Y|\mathbf{x}$.

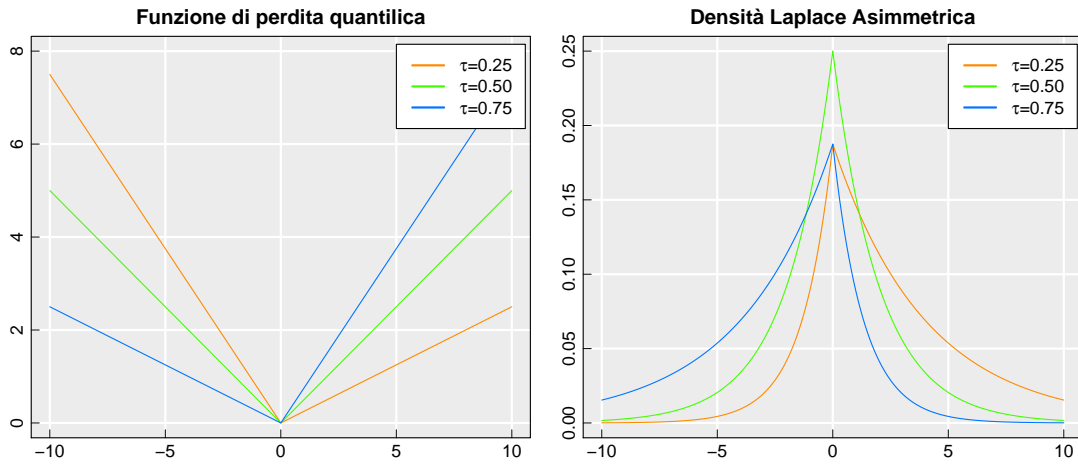


Figura 3.1: Funzione di perdita quantilica (a sinistra) e funzione di densità Laplace Asimmetrica (a destra).

Assumendo di disporre di un campione casuale semplice $(\mathbf{y}_1, \dots, \mathbf{y}_T)$, e di fissare un valore di τ , la funzione di verosimiglianza associata alla distribuzione Laplace asimmetrica per i dati sarà pari a

$$\mathcal{L}(\sigma, \mu | \mathbf{y}) \propto \sigma^{-T} \exp\left\{-\frac{1}{\sigma} \sum_{t=1}^T \rho_\tau(y_t - \mu)\right\}. \quad (3.1.3)$$

La massimizzazione della funzione di verosimiglianza 3.1.3, o equivalentemente la minimizzazione della funzione di perdita 3.1.1, definisce un problema non regolare di stima, derivante dalla non differenziabilità della funzione di perdita quantilica, nonché dalla non corretta specificazione del modello distributivo per i dati. Questo porta a complicazioni in fase stima e, più in generale, in fase di inferenza sui parametri del modello.

In letteratura sono stati proposti diversi approcci per risolvere il problema 3.1.1, ad esempio, Tian, Tian e Zhu 2014 hanno proposto un metodo di stima basato

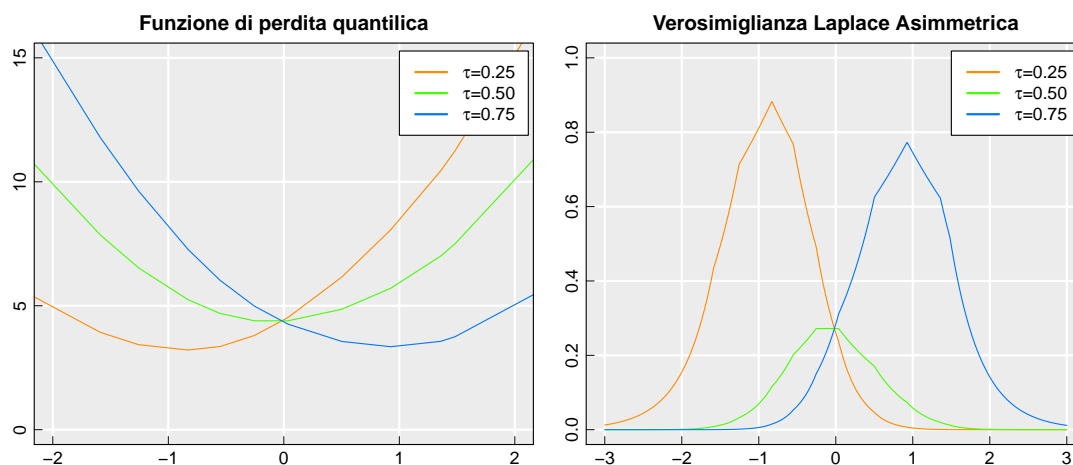


Figura 3.2: Funzione di perdita quantilica (a sinistra) e funzione di verosimiglianza (a destra) calcolate su un campione di venti valori simulati da una variabile aleatoria normale standard.

sull’algoritmo EM, mentre Hunter e Lange 2000 hanno risolto il medesimo problema di ottimizzazione facendo uso dell’algoritmo MM.

D’altra parte l’esistenza di una verosimiglianza propria associata al problema quantilico apre le porte anche ad un approccio modellistico Bayesiano, che permette di adottare metodi inferenziali e algoritmi di stima alternativi, che non risentono della non regolarità della funzione 3.1.1, quali MCMC e *Variational Bayes Approximation*.

Per motivi tecnici, che sono stati meglio discussi nella sezione 5.6, in questa tesi si è preferito adottare un approccio di inferenza Bayesiano, in modo da poter far leva sui mezzi computazionali che questo paradigma mette a disposizione.

3.2 *Quantile regression* Bayesiana

L’adozione di un approccio Bayesiano alla *quantile regression* è ormai pratica comune e porta diversi vantaggi, sia dal punto di vista inferenziale che algoritmico. In particolare permette di sfruttare metodi di simulazione MCMC per caratterizzare la distribuzione a posteriori.

Seguendo l’approccio di Yu e Moyeed 2001, si ipotizza che i dati y_t , condizionatamente a \mathbf{x}_t , provengano da un distribuzione \mathcal{AL} e che le osservazioni $(y_t, \mathbf{x}_t)_{t=1}^T$ siano realizzazioni indipendenti da una qualche legge congiunta non specificata. Per stimare il quantile τ -esimo di $Y|\mathbf{x}$ specifichiamo il modello:

$$y_t = \mathbf{x}_t^\top \boldsymbol{\beta} + \varepsilon_t, \quad (3.2.1)$$

per $t = 1, \dots, T$, dove $\boldsymbol{\beta}$ è un vettore di parametri ignoti di dimensione p e \mathbf{x}_t è un vettore di covariate il cui primo elemento è pari ad uno per includere una costante nella funzione di regressione.

Per ogni t assumiamo che ε_t siano variabili indipendenti e identicamente distribuite in accordo con la legge $\mathcal{AL}(\tau, 0, \sigma)$, con parametro $\sigma > 0$. Per le proprietà della distribuzione \mathcal{AL} , il predittore lineare $\mathbf{x}_t^\top \boldsymbol{\beta}$ corrisponde al τ -esimo quantile di $Y|\mathbf{x}$.

Per completare la specificazione Bayesiana del modello è necessario definire delle leggi a priori sul vettore dei parametri $(\boldsymbol{\beta}, \sigma)$. Assumendo distribuzioni indipendenti per un valore di τ fissato, la densità a priori congiunta sarà data da: $\pi(\boldsymbol{\beta}, \sigma) = \pi(\boldsymbol{\beta}) \pi(\sigma)$, dove $\boldsymbol{\beta} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{S})$, $\sigma \sim \mathcal{IG}(a^0, b^0)$, con \mathbf{m} , \mathbf{S} , $a^0 > 0$, $b^0 > 0$ iperparametri fissati, con \mathbf{S} matrice quadrata definita positiva di dimensioni $p \times p$. La scelta di distribuzioni a priori coniugate con il modello Gaussiano è una pratica comune nell'ambito della *quantile regression* Bayesiana, in quanto semplifica notevolmente l'inferenza operata via MCMC e permette di adottare metodi di simulazione computazionalmente efficienti, come mostrato da Yu e Moyeed 2001.

3.2.1 Rappresentazione stocastica

Seguendo l'approccio di Kozumi e Kobayashi 2011, si può adottare la ben nota rappresentazione stocastica $\varepsilon \sim \mathcal{AL}(\tau, 0, \sigma)$ come mistura di posizione e scala di distribuzioni Gaussiane (Kotz, Kozubowski e Podgórski 2001):

$$\varepsilon = \lambda \omega + \delta \sqrt{\sigma \omega} \nu, \quad (3.2.2)$$

dove $\omega \sim \mathcal{Exp}(\sigma^{-1})$ e $\nu \sim \mathcal{N}(0, 1)$ sono variabili aleatorie indipendenti e \mathcal{Exp} denota la distribuzione esponenziale. I parametri λ e δ sono fissati e pari a

$$\lambda = \frac{1 - 2\tau}{\tau(1 - \tau)}, \quad \delta^2 = \frac{2}{\tau(1 - \tau)}, \quad (3.2.3)$$

in modo da garantire che il quantile di ordine τ di ε sia nullo. Il modello definito nell'equazione 3.2.1 ammette, condizionatamente a ω , la seguente rappresentazione Gaussiana:

$$y_t = \mathbf{x}_t^\top \boldsymbol{\beta} + \lambda \omega_t + \delta \sqrt{\sigma \omega_t} \nu_t, \quad (3.2.4)$$

per $t = 1, \dots, T$ dove ν_t sono variabili indipendenti e identicamente distribuite, e dove ω_t sono i fattori latenti aumentati, anch'essi con distribuzione indipendente ed identicamente distribuita.

La rappresentazione aumentata in equazione 3.2.4 induce, condizionatamente a $\boldsymbol{\omega}$, la seguente funzione di verosimiglianza per i parametri $(\boldsymbol{\beta}, \sigma)$:

$$\mathcal{L}(\boldsymbol{\beta}, \sigma, \boldsymbol{\omega} | \mathbf{x}, \mathbf{y}) = \prod_{t=1}^T \mathcal{N}(y_t | \mathbf{x}_t^\top \boldsymbol{\beta} + \lambda \omega_t, \delta^2 \sigma \omega_t), \quad (3.2.5)$$

dove ricordiamo che $\mathcal{N}(\cdot | m, s^2)$ denota la funzione di densità di un variabile aleatoria Gaussiana di media m e varianza s^2 .

Componendo verosimiglianza e distribuzioni a priori dei parametri e del fattore latente si ottiene la distribuzione a posteriori:

$$\pi(\boldsymbol{\beta}, \sigma, \boldsymbol{\omega} | \mathbf{x}, \mathbf{y}) \propto \pi(\boldsymbol{\beta}) \pi(\sigma) \pi(\boldsymbol{\omega} | \sigma) \mathcal{L}(\boldsymbol{\beta}, \sigma, \boldsymbol{\omega} | \mathbf{x}, \mathbf{y}).$$

Esplicitando le densità a priori e la verosimiglianza si ottiene la forma analitica:

$$\begin{aligned} \pi(\boldsymbol{\beta}, \sigma, \boldsymbol{\omega} | \mathbf{x}, \mathbf{y}) \propto & \sigma^{-(a^0 + T + 1)} \exp \left\{ -\frac{b^0}{\sigma} - \sum_{t=1}^T \frac{\omega_t}{\sigma} \right\} \exp \left\{ -\frac{1}{2} \boldsymbol{\beta}^\top \mathbf{S}^{-1} \boldsymbol{\beta} \right\} \\ & \times (\delta^2 \sigma)^{-\frac{T}{2}} \left(\prod_{t=1}^T \omega_t^{-\frac{1}{2}} \right) \exp \left\{ -\frac{1}{2} \sum_{t=1}^T \frac{(y_t - \mathbf{x}_t^\top \boldsymbol{\beta} - \lambda \omega_t)^2}{\delta^2 \sigma \omega_t} \right\}. \end{aligned} \quad (3.2.6)$$

Le caratteristiche della distribuzione a posteriori non consentono di ottenere stime in forma chiusa per i parametri, ma permettono invece la costruzione di un algoritmo di simulazione efficiente per fare inferenza sul modello.

3.2.2 Inferenza a posteriori

Grazie alla rappresentazione condizionatamente Gaussiana 3.2.4, possiamo implementare un *Partially Collapsed Gibbs Sampler* con *Data Augmentation*, per generare valori casuali dalla distribuzione a posteriori $\pi(\boldsymbol{\beta}, \sigma, \boldsymbol{\omega} | \mathbf{y}, \mathbf{x})$, come proposto da Bernardi, Gayraud e Petrella 2015.

L'idea chiave dietro ad un campionamento totalmente collassato, proposto per la prima volta da Liu 1994, è di proporre nuovi valori della catena dalle distribuzioni *full conditional* dei parametri, opportunamente marginalizzate, in modo che la composizione di queste risulti essere la distribuzione congiunta a posteriori.

In questo caso non è possibile costruire un algoritmo completamente collassato, poiché non tutte le *full conditional* sono marginalizzabili analiticamente. Tuttavia è possibile integrare il fattore aumentato $\boldsymbol{\omega}$ dalla distribuzione *full conditional* congiunta $(\sigma, \boldsymbol{\omega})$ e sfruttare questa nuova legge a posteriori parzialmente collassata per simulare da $\pi(\sigma | \boldsymbol{\beta}, \mathbf{y}, \mathbf{x})$ (si veda la dimostrazione in A.2).

Un algoritmo collassato tipicamente presenta proprietà preferibili rispetto ad un *Gibbs Sampling* tradizionale. L'operazione di marginalizzazione infatti porta ad

aumentare la velocità di convergenza della catena, riduce l'autocorrelazione della stessa e ne migliora il *mixing*, come mostrato in Dyk e Park 2008.

A differenza di un *Gibbs Sampling* tradizionale i passi di simulazione dalle distribuzioni *full conditional* non sono permutabili e devono seguire un ordinamento appropriato, affinché la distribuzione ergodica della catena converga alla legge a posteriori desiderata.

Le distribuzioni *full conditional* per i parametri e il fattore aumentato della rappresentazione stocastica del modello di regressione quantilica Bayesiana sono discusse in appendice, alla sezione A.2.

Fissati i valori iniziali dei parametri $(\boldsymbol{\beta}^{(0)}, \sigma^{(0)})$, possiamo simulare il primo fattore aumentato $\omega^{(0)}$ sfruttando la sua legge a priori. Le simulazioni dalla distribuzione a posteriori alla $(k+1)$ -esima iterazione sono ottenute secondo il seguente schema di *Gibbs Sampling*:

1. Si generi $\sigma^{(k+1)}$ da $\pi(\sigma | \boldsymbol{\beta}, \mathbf{x}, \mathbf{y}) \propto \mathcal{IG}(a^*, b^*)$, dove:

$$a^* = a^0 + \frac{3}{2}T + \frac{1}{2}, \quad b^* = b^0 + \sum_{t=1}^T \rho_\tau(y_t - \mathbf{x}_t^\top \boldsymbol{\beta}^{(k)}).$$

2. Per ogni $t = 1, \dots, T$, si generi $\omega_t^{-1(k+1)}$ da $\pi(\omega_t^{-1} | \sigma, \boldsymbol{\beta}, y_t, \mathbf{x}_t) \propto \mathcal{IN}(\psi_t^*, \varphi_t^*)$, dove:

$$\psi_t^* = \sqrt{\frac{\lambda^2 + 2\delta^2}{(y_t - \mathbf{x}_t^\top \boldsymbol{\beta}^{(k)})^2}}, \quad \varphi_t^* = \frac{\lambda^2 + 2\delta^2}{\delta^2 \sigma^{(k+1)}}.$$

La scrittura $\mathcal{IN}(\psi, \varphi)$ denota la distribuzione Gaussiana Inversa di parametri ψ e φ . Questa può efficientemente essere generata adottando il metodo proposto da Michael, Schucany e Haas 1976.

3. Si generi $\boldsymbol{\beta}^{(k+1)}$ da $\pi(\boldsymbol{\beta} | \sigma, \boldsymbol{\omega}, \mathbf{x}, \mathbf{y}) \propto \mathcal{N}(\mathbf{m}^*, \mathbf{S}^*)$, dove:

$$\mathbf{m}^* = \mathbf{S}^* \left(\sum_{t=1}^T \frac{\mathbf{x}_t (y_t - \lambda \omega_t^{(k+1)})}{\delta^2 \sigma \omega_t^{(k+1)}} \right), \quad \mathbf{S}^* = \left(\mathbf{S}^{-1} + \sum_{t=1}^T \frac{\mathbf{x}_t \mathbf{x}_t^\top}{\delta^2 \sigma \omega_t^{(k+1)}} \right)^{-1}.$$

L'aggiornamento dei parametri effettuato in quest'ordine genera una catena di Markov la cui distribuzione stazionaria converge alla distribuzione a posteriori dei parametri. Combinando gli step 1. e 2. si genera un'estrazione dalla distribuzione a posteriori condizionata

$$\pi(\sigma, \boldsymbol{\omega} | \boldsymbol{\beta}, \mathbf{x}, \mathbf{y}) = \pi(\sigma | \boldsymbol{\beta}, \mathbf{x}, \mathbf{y}) \pi(\boldsymbol{\omega} | \sigma, \boldsymbol{\beta}, \mathbf{x}, \mathbf{y}),$$

ed in questo caso il *Partially Collapsed Gibbs Sampler* risulta essere un *Gibbs Sampler* ordinario con campionamento a blocchi.

3.3 *Quantile regression* dinamica a disturbi indipendenti

La *quantile regression* Bayesiana può essere facilmente estesa a modelli dinamici rappresentabili in forma *State Space*, come dimostrato nei lavori di De Rossi e Harvey 2009, Kurose e Omori 2012 e Bernardi, Gayraud e Petrella 2015.

3.3.1 Rappresentazione stocastica e modello *State Space*

Consideriamo il modello dinamico

$$\begin{aligned} \mathbf{y}_t &= \mathbf{c}_t + \mathbf{Z}_t \cdot \boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t, & \boldsymbol{\varepsilon}_{i,t} &\sim \mathcal{AL}(\tau, 0, \sigma_\varepsilon^2) \\ \boldsymbol{\alpha}_{t+1} &= \mathbf{d}_t + \mathbf{T}_t \cdot \boldsymbol{\alpha}_t + \mathbf{R}_t \cdot \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_t) \\ & & \boldsymbol{\alpha}_0 &\sim \mathcal{N}(\mathbf{a}_0, \mathbf{P}_0) \end{aligned} \quad (3.3.1)$$

dove assumiamo indipendenza tra le componenti di $\boldsymbol{\varepsilon}_t$.

Per le proprietà della distribuzione \mathcal{AL} il modello 3.3.1 impone l'uguaglianza tra $Q^\tau(y_{i,t}) = \mathbf{c}_{i,t} + \mathbf{z}_{i,t}^\top \cdot \boldsymbol{\alpha}_t$, dove $\mathbf{z}_{i,t}^\top$ è la riga i -esima della matrice di misura \mathbf{Z}_t . Stimando i parametri del modello ed estraendo il segnale latente è dunque possibile costruire una funzione quantile dinamica per ogni componente $y_{i,t}$ del vettore \mathbf{y}_t .

Il modello *State Space* così specificato risulta essere lineare ma non Gaussiano, a causa delle assunzioni fatte sul termine di errore. In queste circostanze le tecniche di filtraggio ottimale per marginalizzare analiticamente il processo latente basate sul filtro di Kalman non sono direttamente utilizzabili.

Tuttavia, anche in questo caso, è possibile ricorrere ad una rappresentazione stocastica che renda agevole una stima Bayesiana basata su algoritmi MCMC. Considerando la seguente rappresentazione aumentata per il termine di innovazione

$$\begin{aligned} \boldsymbol{\varepsilon}_{i,t} &= \lambda \omega_{i,t} + \delta \sqrt{\sigma_\varepsilon^2} \omega_{i,t} \nu_{i,t}, \\ \omega_{i,t} &\stackrel{i.i.d.}{\sim} \text{Exp}(\sigma_\varepsilon^{-2}), \\ \nu_{i,t} &\stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1), \end{aligned} \quad (3.3.2)$$

dove $\omega_{i,t}$ e $\nu_{i,t}$ sono variabili aleatorie indipendenti e λ e δ sono definiti come in 3.2.3, è possibile dimostrare che il modello non Gaussiano 3.3.1 ammette una rappresentazione condizionatamente Gaussiana:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{c}_t + \mathbf{Z}_t \cdot \boldsymbol{\alpha}_t + \lambda \cdot \boldsymbol{\omega}_t + \mathbf{G}_t \cdot \boldsymbol{\nu}_t, & \boldsymbol{\nu}_t &\sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n) \\ \boldsymbol{\alpha}_{t+1} &= \mathbf{d}_t + \mathbf{T}_t \cdot \boldsymbol{\alpha}_t + \mathbf{R}_t \cdot \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_t) \\ & & \boldsymbol{\alpha}_0 &\sim \mathcal{N}(\mathbf{a}_0, \mathbf{P}_0) \end{aligned} \quad (3.3.3)$$

dove $\boldsymbol{\omega}_t = (\omega_{1,t}, \dots, \omega_{n,t})$ e $\mathbf{G}_t = \text{diag}(\delta\sqrt{\sigma_\varepsilon^2\omega_{1,t}}, \dots, \delta\sqrt{\sigma_\varepsilon^2\omega_{n,t}})$.

Assumendo ora che $(\mathbf{y}_1, \dots, \mathbf{y}_T)$ siano realizzazioni del processo $(\mathbf{Y}(t), t \in \mathbf{T})$ campionato su n siti fissati $(\mathbf{s}_1, \dots, \mathbf{s}_n)$, possiamo adottare il modello spazio-temporale descritto in 2.1.13 per fare inferenza sui quantili di $Y(\mathbf{s}, t)$.

Specifichiamo le quantità \mathbf{c}_t , \mathbf{Z}_t , \mathbf{d}_t , \mathbf{T}_t , \mathbf{R}_t , \mathbf{Q}_t , \mathbf{a}_0 e \mathbf{P}_0 della rappresentazione *State Space* in accordo con le equazioni 2.2.4 e assumiamo che le distribuzioni a priori per i parametri siano quelle descritte in 2.4.1.

3.3.2 Inferenza ed estrazione del segnale

La densità a posteriori si ottiene dalla composizione delle distribuzioni a priori dei parametri, del fattore aumentato e del processo latente con la verosimiglianza condizionata:

$$\begin{aligned} & \pi\left(\sigma_\eta^2, \sigma_\varepsilon^2, \{\boldsymbol{\omega}_t\}_{t=0}^T, \{\boldsymbol{\alpha}_t\}_{t=0}^T \mid \{\mathbf{y}_t\}_{t=0}^T\right) \propto \\ & \propto \pi\left(\sigma_\eta^2\right) \pi\left(\sigma_\varepsilon^2\right) \pi\left(\{\boldsymbol{\omega}_t\}_{t=1}^T \mid \sigma_\varepsilon^2\right) p\left(\{\boldsymbol{\alpha}_t\}_{t=0}^T \mid \sigma_\eta^2\right) \mathcal{L}\left(\sigma_\varepsilon^2, \{\boldsymbol{\omega}_t\}_{t=1}^T, \{\boldsymbol{\alpha}_t\}_{t=0}^T \mid \{\mathbf{y}_t\}_{t=0}^T\right) \\ & \propto \pi\left(\sigma_\eta^2\right) \pi\left(\sigma_\varepsilon^2\right) \left(\prod_{t=1}^T \prod_{i=1}^n p\left(\omega_{i,t} \mid \sigma_\varepsilon^2\right)\right) \left(p\left(\boldsymbol{\alpha}_0\right) \prod_{t=1}^{T-1} p\left(\boldsymbol{\alpha}_{t+1} \mid \boldsymbol{\alpha}_t, \sigma_\eta^2\right)\right) \left(\prod_{t=1}^T p\left(\mathbf{y}_t \mid \sigma_\varepsilon^2, \boldsymbol{\omega}_t, \boldsymbol{\alpha}_t\right)\right). \end{aligned}$$

Come mostrato da Bernardi, Gayraud e Petrella 2015, esplicitando tutte le densità otteniamo la forma analitica

$$\begin{aligned} & \pi\left(\sigma_\eta^2, \sigma_\varepsilon^2, \{\boldsymbol{\omega}_t\}_{t=1}^T, \{\boldsymbol{\alpha}_t\}_{t=0}^T \mid \{\mathbf{y}_t\}_{t=1}^T\right) \propto \\ & \propto \left(\sigma_\eta^2\right)^{-(a_\eta^0+1)} \exp\left\{-\frac{b_\eta^0}{\sigma_\eta^2}\right\} \left(\sigma_\varepsilon^2\right)^{-(a_\varepsilon^0+1)} \exp\left\{-\frac{b_\varepsilon^0}{\sigma_\varepsilon^2}\right\} \\ & \times \left(\sigma_\varepsilon^2\right)^{-nT} \exp\left\{-\sum_{t=1}^T \sum_{i=1}^n \frac{\omega_{i,t}}{\sigma_\varepsilon^2}\right\} \exp\left\{-\frac{1}{2k} \|\boldsymbol{\alpha}_0\|_2^2\right\} \\ & \times \left(\sigma_\eta^2\right)^{-\frac{m}{2}(T-1)} \exp\left\{-\frac{1}{2\sigma_\eta^2} \sum_{t=1}^{T-1} \left\|\mathbf{W}^{-\frac{1}{2}}(\boldsymbol{\alpha}_{t+1} - \boldsymbol{\alpha}_t)\right\|_2^2\right\} \\ & \times \left(\delta^2\sigma_\varepsilon^2\right)^{-\frac{nT}{2}} \left(\prod_{t=1}^T \prod_{i=1}^n (\omega_{i,t})^{-\frac{1}{2}}\right) \exp\left\{-\frac{1}{2} \sum_{t=1}^T \sum_{i=1}^n \frac{(y_{i,t} - \lambda\omega_{i,t} - \mathbf{k}_i^\top \boldsymbol{\alpha}_t)^2}{\delta^2\sigma_\varepsilon^2\omega_{i,t}}\right\}. \end{aligned} \quad (3.3.4)$$

La complessità della densità a posteriori non rende possibile ottenere una stima congiunta per i parametri $(\sigma_\eta^2, \sigma_\varepsilon^2)$ ed il processo latente $\boldsymbol{\alpha}_t$ in forma chiusa, marginalizzando l'effetto del fattore aumentato $\boldsymbol{\omega}_t$. Dunque per ottenere la distribuzione a posteriori per i parametri e gli stati è necessario ricorrere ad un modello di simulazione MCMC.

Dalla densità a posteriori congiunta si possono calcolare facilmente le distribuzioni *full conditional* per i parametri σ_ε^2 e σ_η^2 e per i fattori latenti $\{\boldsymbol{\omega}_t\}_{t=1}^T$ isolando

i *kernel* delle distribuzioni condizionate, come mostrato in A.3 e come discusso da Bernardi, Gayraud e Petrella 2015.

Derivate tutte le distribuzioni *full conditional* (collassate e non), si può costruire un'efficiente algoritmo di campionamento dalla distribuzione a posteriori. Seguendo il medesimo approccio di calcolo utilizzato nella sezione A.2, è possibile dimostrare che la distribuzione *full conditional* collassata $\pi\left(\sigma_\varepsilon^2 \mid \{\boldsymbol{\alpha}_t\}_{t=1}^T, \{\mathbf{y}_t\}_{t=1}^T\right)$ per σ_ε^2 risulta proporzionale ad una Gamma Inversa di parametri

$$a_\varepsilon^* = a_\varepsilon^0 + nT, \quad b_\varepsilon^* = b_\varepsilon^0 + \sum_{t=1}^T \sum_{i=1}^n \rho_\tau \left(y_{i,t} - \mathbf{k}_i^\top \boldsymbol{\alpha}_t^{(k)} \right),$$

come mostrato in A.3

Sfruttando questo risultato possiamo simulare prima dalla legge collassata di σ_ε^2 e poi dalla distribuzione *full conditional* di $\{\boldsymbol{\omega}_t\}_{t=1}^T$, ottenendo così un campionamento a blocchi dalla distribuzione congiunta $\pi\left(\sigma_\varepsilon^2, \{\boldsymbol{\omega}_t\}_{t=1}^T \mid \{\boldsymbol{\alpha}_t\}_{t=1}^T, \{\mathbf{y}_t\}_{t=1}^T\right)$, rendendo più efficiente ed efficace l'algoritmo.

Per fare inferenza a posteriori sul modello adottiamo quindi un algoritmo di simulazione MCMC ed in particolare un *Partially Collapsed Gibbs Sampling with Data Augmentation* strutturato come segue.

Fissati i valori iniziali dei parametri $(\sigma_\eta^2)^{(0)}$ e $(\sigma_\varepsilon^2)^{(0)}$, simuliamo prima il fattore aumentato $\{\boldsymbol{\omega}_t^{(0)}\}_{t=1}^T$ sfruttando la sua distribuzione a priori, e poi gli stati latenti $\{\boldsymbol{\alpha}_t^{(0)}\}_{t=0}^T$ dalla distribuzione *full conditional*, sfruttando l'algoritmo di *Simulation Smoothing*. Dati i valori iniziali, le simulazioni dalla distribuzione a posteriori alla $(k+1)$ -esima iterazione sono ottenute secondo il seguente schema del *Gibbs Sampling* con *Data Augmentation*:

1. Si generi $(\sigma_\eta^2)^{(k+1)}$ da $\pi\left(\sigma_\eta^2 \mid \{\boldsymbol{\alpha}_t\}_{t=1}^T\right) \propto \mathcal{IG}(a_\eta^*, b_\eta^*)$, dove:

$$a_\eta^* = a_\eta^0 + \frac{m}{2}(T-1), \quad b_\eta^* = b_\eta^0 + \frac{1}{2} \sum_{t=1}^{T-1} \left\| \mathbf{W}^{-\frac{1}{2}} \left(\boldsymbol{\alpha}_{t+1}^{(k)} - \boldsymbol{\alpha}_t^{(k)} \right) \right\|_2^2.$$

2. Si generi $(\sigma_\varepsilon^2)^{(k+1)}$ da $\pi\left(\sigma_\varepsilon^2 \mid \{\boldsymbol{\alpha}_t\}_{t=1}^T, \{\mathbf{y}_t\}_{t=1}^T\right) \propto \mathcal{IG}(a_\varepsilon^*, b_\varepsilon^*)$, dove:

$$a_\varepsilon^* = a_\varepsilon^0 + nT, \quad b_\varepsilon^* = b_\varepsilon^0 + \sum_{t=1}^T \sum_{i=1}^n \rho_\tau \left(y_{i,t} - \mathbf{k}_i^\top \boldsymbol{\alpha}_t^{(k)} \right).$$

3. Per ogni $t = 1, \dots, T$ e per ogni $i = 1, \dots, n$, si generi $(\omega_{i,t}^{-1})^{(k+1)}$ da $\pi\left(\omega_{i,t}^{-1} \mid \sigma_\varepsilon^2, \boldsymbol{\alpha}_t, y_{i,t}\right) \propto \mathcal{IN}(\psi_{i,t}^*, \varphi_{i,t}^*)$, dove:

$$\psi_{i,t}^* = \sqrt{\frac{\lambda^2 + 2\delta^2}{\left(y_{i,t} - \mathbf{k}_i^\top \boldsymbol{\alpha}_t^{(k)}\right)^2}}, \quad \varphi_{i,t}^* = \frac{\lambda^2 + 2\delta^2}{\delta^2 \sigma_\varepsilon^{2(k+1)}}.$$

4. Per ogni $t = 0, \dots, T$, si generi $\boldsymbol{\alpha}_t^{(k+1)}$ da $\pi\left(\boldsymbol{\alpha}_t \mid \sigma_\eta^2, \sigma_\varepsilon^2, \{\omega_t\}_{t=1}^T, \{\mathbf{y}_t\}_{t=1}^T\right)$. Condizionatamente al fattore aumentato $\{\omega_t\}_{t=1}^T$, il modello è lineare e Gaussiano, dunque le dinamiche latenti $\{\boldsymbol{\alpha}_t\}_{t=0}^T$ possono essere efficientemente simulate sfruttando l'algoritmo di *Simulation Smoothing* proposto da Durbin e Koopman 2002.

3.4 Quantile regression dinamica a disturbi correlati

L'assunzione di mutua indipendenza tra i disturbi di misura adottata nelle precedenti sezioni spesso non è realistica nello studio di fenomeni spazio-temporali. In questa sezione generalizziamo il modello presentato nella sezione 3.3 per introdurre una struttura di correlazione spaziale flessibile tra i disturbi di misura, analoga a quella presentata nel capitolo 2, ma collocata nel contesto della *quantile regression*. Per farlo dobbiamo introdurre una generalizzazione della distribuzione \mathcal{AL} e della sua rappresentazione stocastica.

3.4.1 Distribuzione Laplace asimmetrica multivariata

Diremo che il vettore aleatorio $\mathbf{Y} \in \mathbb{R}^n$ ha distribuzione Laplace Asimmetrica Multivariata, e indicheremo $\mathbf{Y} \sim \mathcal{AL}_n(\boldsymbol{\mu}, \boldsymbol{\xi}, \boldsymbol{\Sigma})$, se la sua funzione di densità è data da

$$\begin{aligned} \mathcal{AL}_n(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\xi}, \boldsymbol{\Sigma}) &= \frac{2 \exp\{\boldsymbol{\xi}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\Delta}^{-1} (\mathbf{y} - \boldsymbol{\mu})\}}{(2\pi)^{n/2} |\boldsymbol{\Sigma}|^{1/2}} \left[\frac{(\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})}{2 + \boldsymbol{\xi}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\xi}} \right]^{\nu/2} \\ &\times K_\nu \left(\sqrt{(2 + \boldsymbol{\xi}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\xi}) (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})} \right), \end{aligned} \quad (3.4.1)$$

dove $\boldsymbol{\mu} \in \mathbb{R}^n$ e $\boldsymbol{\xi} \in \mathbb{R}^n$ sono, rispettivamente, parametri di posizione e forma, $\boldsymbol{\Sigma} = \boldsymbol{\Delta} \boldsymbol{\Omega} \boldsymbol{\Delta}$ è una matrice di scala simmetrica e definita positiva, $\boldsymbol{\Delta} = \text{diag}(\sigma_1, \dots, \sigma_n)$ dove $\sigma_j > 0$ per ogni $j = 1, \dots, n$ e $\boldsymbol{\Omega}$ è una matrice di correlazione. In ultimo $K_\nu(\cdot)$ denota la funzione di Bessel del terzo tipo di indice $\nu = 1 - n/2$.

Come dimostrato in Kotz, Kozubowski e Podgórski 2001, tutte le distribuzioni marginali del vettore aleatorio \mathbf{Y} sono Laplace Asimmetriche. In particolare le marginali univariate sono tali che $Y_j \sim \mathcal{AL}(\mu_j, \xi_j, \sigma_j)$ per ogni $j = 1, \dots, n$.

Kotz e Nadarajah 2004 hanno inoltre dimostrato che anche la distribuzione \mathcal{AL}_n può essere espressa come una mistura di posizione e scala di una variabile aleatoria normale e definendo $\mathbf{Y} \sim \mathcal{AL}_n(\boldsymbol{\mu}, \boldsymbol{\xi}, \boldsymbol{\Sigma})$, si ha che

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\Delta} \boldsymbol{\xi} \omega + \boldsymbol{\Sigma}^{\frac{1}{2}} \sqrt{\omega} \boldsymbol{\nu}, \quad (3.4.2)$$

dove $\omega \sim \mathcal{E}xp(1)$ e $\boldsymbol{\nu} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I}_n)$, con $\omega \perp \boldsymbol{\nu}$. Dall'equazione 3.4.2 si deriva che la distribuzione condizionata \mathbf{Y} dato ω è tale che $\mathbf{Y}|\omega \sim \mathcal{N}_n(\boldsymbol{\mu} + \boldsymbol{\Delta} \boldsymbol{\xi} \omega, \boldsymbol{\Sigma} \omega)$ e che la sua distribuzione incondizionata sia quella descritta in 3.4.1.

3.4.2 Rappresentazione stocastica e modello *State Space*

Definiamo ora la variabile $\boldsymbol{\varepsilon}$ in accordo con l'equazione

$$\boldsymbol{\varepsilon} = \boldsymbol{\lambda} \omega + \delta \sqrt{\sigma \omega} \boldsymbol{\Omega}^{\frac{1}{2}} \boldsymbol{\nu}, \quad (3.4.3)$$

dove $\boldsymbol{\lambda} = \lambda \mathbf{1}_n$, $\omega \sim \mathcal{E}xp(\sigma^{-1})$ e $\boldsymbol{\nu} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I}_n)$, con λ e δ scelti come in 3.2.3, per un valore di τ fissato.

Osserviamo che la rappresentazione 3.4.3 è un caso particolare della 3.4.2, dove $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\xi} = \delta^{-1} \lambda \mathbf{1}_n$ e $\boldsymbol{\Delta} = \delta \sigma \mathbf{I}_n$, dunque $\boldsymbol{\varepsilon} \sim \mathcal{AL}_n(\boldsymbol{\lambda}, \mathbf{0}, \delta^2 \sigma^2 \boldsymbol{\Omega})$, da cui segue che $\varepsilon_j \sim \mathcal{AL}(\lambda, 0, \delta^2 \sigma^2)$ per ogni $j = 1, \dots, n$. Per le proprietà della distribuzione \mathcal{AL} , il quantile di ordine τ di ε_j è centrato in 0 e la rappresentazione 3.4.3 può essere adottata per specificare la forma dei disturbi in un modello di *quantile regression*.

Riconsideriamo ora il modello dinamico 3.3.1, nel quale specifichiamo gli errori di misura come in equazione 3.4.3, ottenendo in questo modo la seguente forma *State Space*

$$\begin{aligned} \mathbf{y}_t &= \mathbf{c}_t + \mathbf{Z}_t \cdot \boldsymbol{\alpha}_t + \boldsymbol{\lambda} \cdot \omega_t + \mathbf{G}_t \cdot \boldsymbol{\nu}_t, & \boldsymbol{\nu}_t &\sim \mathcal{N}(\mathbf{0}, \mathbf{H}_t) \\ \boldsymbol{\alpha}_{t+1} &= \mathbf{d}_t + \mathbf{T}_t \cdot \boldsymbol{\alpha}_t + \mathbf{R}_t \cdot \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_t) \\ & & \boldsymbol{\alpha}_0 &\sim \mathcal{N}(\mathbf{a}_0, \mathbf{P}_0) \end{aligned} \quad (3.4.4)$$

dove $\boldsymbol{\lambda} \omega_t$ è un fattore di traslazione e $\mathbf{G}_t = \delta \sqrt{\omega_t} \mathbf{I}_n$ è la matrice di scala per i disturbi nella rappresentazione aumentata.

Il modello così specificato permette di gestire due distinte forme di correlazione spaziale tra i quantili del processo osservabile. La prima è la correlazione di grande scala, modellata attraverso la convoluzione discreta descritta dall'interazione tra la matrice di misura e il processo latente. La seconda, di piccola scala, entra nella specificazione del termine d'errore di misura e della sua matrice covarianza.

3.4.3 Inferenza ed estrazione del segnale

Assumendo ancora una volta che le matrici strutturali della forma *State Space* siano specificate come nelle equazioni 2.2.4, e che le distribuzioni a priori dei parametri siano indipendenti e coniugate con la verosimiglianza aumentata, possiamo

scrivere la funzione di densità a posteriori congiunta del modello secondo la seguente forma analitica:

$$\begin{aligned}
& \pi \left(\sigma_\eta^2, \sigma_\varepsilon^2, \{\omega_t\}_{t=1}^T, \{\boldsymbol{\alpha}_t\}_{t=0}^T \mid \{\mathbf{y}_t\}_{t=1}^T \right) \propto \\
& \propto (\sigma_\eta^2)^{-(a_\eta^0+1)} \exp \left\{ -\frac{b_\eta^0}{\sigma_\eta^2} \right\} (\sigma_\varepsilon^2)^{-(a_\varepsilon^0+1)} \exp \left\{ -\frac{b_\varepsilon^0}{\sigma_\varepsilon^2} \right\} \\
& \times (\sigma_\varepsilon^2)^{-T} \exp \left\{ -\sum_{t=1}^T \frac{\omega_t}{\sigma_\varepsilon^2} \right\} \exp \left\{ -\frac{1}{2k} \|\boldsymbol{\alpha}_0\|_2^2 \right\} \\
& \times (\sigma_\eta^2)^{-\frac{m}{2}(T-1)} \exp \left\{ -\frac{1}{2\sigma_\eta^2} \sum_{t=1}^{T-1} \left\| \mathbf{W}^{-\frac{1}{2}} (\boldsymbol{\alpha}_{t+1} - \boldsymbol{\alpha}_t) \right\|_2^2 \right\} \\
& \times (\delta^2 \sigma_\varepsilon^2)^{-\frac{n}{2}T} \left(\prod_{t=1}^T \omega_t^{-\frac{n}{2}} \right) \exp \left\{ -\frac{1}{2} \sum_{t=1}^T \frac{1}{\delta^2 \sigma_\varepsilon^2 \omega_t} \left\| \mathbf{V}^{-\frac{1}{2}} (\mathbf{y}_t - \boldsymbol{\lambda} \omega_t - \mathbf{K} \boldsymbol{\alpha}_t) \right\|_2^2 \right\}.
\end{aligned} \tag{3.4.5}$$

Per fare inferenza sui parametri anche in questo caso è necessario ricorrere ad algoritmi di simulazione MCMC, e grazie all'esistenza di a priori coniugate per i parametri possiamo costruire un efficiente algoritmo di *Gibbs Sampling* per simulare congiuntamente parametri, fattore aumentato e processo latente.

Definite tutte le distribuzioni *full conditional* (si veda A.4), l'algoritmo di *Gibbs Sampling* con *Data Augmentation* per la simulazione dalla distribuzione a posteriori procede secondo lo schema seguente. Fissati i valori iniziali dei parametri $(\sigma_\eta^2)^{(0)}$ e $(\sigma_\varepsilon^2)^{(0)}$, per ogni $t = 1, \dots, T$ si simulano i fattori aumentati $\omega_t^{(0)}$ dalla distribuzione a priori. Dati i parametri iniziali e i fattori aumentati, si esegue una prima simulazione degli stati latenti $\boldsymbol{\alpha}_t^{(0)}$ per ogni $t = 0, \dots, T$ attraverso l'algoritmo di *Simulation Smoothing* di Durbin e Koopman 2012. Poi le successive simulazioni dalla distribuzione a posteriori alla $(k+1)$ -esima iterazione sono ottenute secondo il seguente schema del *Gibbs Sampling*:

1. Si generi $(\sigma_\eta^2)^{(k+1)}$ da $\pi \left(\sigma_\eta^2 \mid \{\boldsymbol{\alpha}_t\}_{t=1}^T \right) \propto \mathcal{IG} (a_\eta^*, b_\eta^*)$, dove:

$$a_\eta^* = a_\eta^0 + \frac{m}{2}(T-1), \quad b_\eta^* = b_\eta^0 + \frac{1}{2} \sum_{t=1}^{T-1} \left\| \mathbf{W}^{-\frac{1}{2}} \left(\boldsymbol{\alpha}_{t+1}^{(k)} - \boldsymbol{\alpha}_t^{(k)} \right) \right\|_2^2.$$

2. Si generi $(\sigma_\varepsilon^2)^{(k+1)}$ da $\pi \left(\sigma_\varepsilon^2 \mid \{\boldsymbol{\alpha}_t\}_{t=1}^T, \{\omega_t\}_{t=1}^T, \{\mathbf{y}_t\}_{t=1}^T \right) \propto \mathcal{IG} (a_\varepsilon^*, b_\varepsilon^*)$, dove:

$$\begin{aligned}
a_\varepsilon^* &= a_\varepsilon^0 + T + \frac{n}{2}T, \\
b_\varepsilon^* &= b_\varepsilon^0 + \sum_{t=1}^T \omega_t^{(k)} + \sum_{t=1}^T \frac{1}{2\delta^2 \omega_t^{(k)}} \left\| \mathbf{V}^{-\frac{1}{2}} \left(\mathbf{y}_t - \boldsymbol{\lambda} \omega_t^{(k)} - \mathbf{K} \boldsymbol{\alpha}_t^{(k)} \right) \right\|_2^2.
\end{aligned}$$

3. Per ogni $t = 1, \dots, T$, si generi $\omega_t^{(k+1)}$ da $\pi(\omega_t | \sigma_\varepsilon^2, \boldsymbol{\alpha}_t, \mathbf{y}_t) \propto \mathcal{GIG}(\chi_t^*, \psi_t^*, \varphi_t^*)$, dove:

$$\begin{aligned}\chi_t^* &= \frac{2-n}{2}, \\ \psi_t^* &= \frac{1}{\delta^2 \sigma_\varepsilon^{2(k+1)}} \left(\boldsymbol{\lambda}^\top \mathbf{V}^{-1} \boldsymbol{\lambda} + 2\delta^2 \right), \\ \varphi_t^* &= \frac{1}{\delta^2 \sigma_\varepsilon^{2(k+1)}} \left\| \mathbf{V}^{-\frac{1}{2}} \left(\mathbf{y}_t - \mathbf{K} \boldsymbol{\alpha}_t^{(k)} \right) \right\|_2^2.\end{aligned}$$

La notazione \mathcal{GIG} indica la distribuzione *Generalized Inverse Gaussian* a tre parametri, che può essere efficientemente simulata attraverso l'algoritmo proposto da Devroye 2014.

4. Per ogni $t = 0, \dots, T$, si generi $\boldsymbol{\alpha}_t^{(k+1)}$ da $\pi\left(\boldsymbol{\alpha}_t | \sigma_\eta^2, \sigma_\varepsilon^2, \{\omega_t\}_{t=1}^T, \{\mathbf{y}_t\}_{t=1}^T\right)$. Condizionatamente al fattore aumentato $\{\omega_t\}_{t=1}^T$, il modello è lineare e Gaussiano, dunque le dinamiche latenti $\{\boldsymbol{\alpha}_t\}_{t=1}^T$ possono essere efficientemente simulate usando l'algoritmo di *Simulation Smoothing* proposto da Durbin e Koopman 2002.

L'esecuzione dei precedenti passi permette di generare una catena di Markov la cui distribuzione ergodica converge alla distribuzione a posteriori dei parametri e dei processi latenti.

L'introduzione di una matrice di correlazione tra i disturbi di misura porta a due differenze sostanziali rispetto all'algoritmo presentato in sezione 3.3.2. Anzitutto non è più possibile marginalizzare analiticamente il fattore latente dalla distribuzione *full conditional* congiunta di σ_ε^2 e ω_t e dunque non è nemmeno possibile adottare un passo di campionamento collassato. In secondo luogo, non è più possibile ricondurre la distribuzione condizionata di ω_t ad una \mathcal{IN} mediante trasformazioni biettive. Questo rende necessario simulare valori casuali direttamente dalla distribuzione \mathcal{GIG} mediante metodi di campionamento accetto-rifiuto adattivi.

Il guadagno in flessibilità del modello è quindi controbilanciato da una perdita in efficienza dal punto di vista computazionale, che porta inevitabilmente ad un aumento dei tempi di calcolo.

Si noti che in questo capitolo si è sempre fatta l'assunzione che la matrici di correlazione \mathbf{V} e \mathbf{W} fossero note e fissate. Nei casi in cui non lo fossero è possibile estendere le metodologie inferenziali proposte secondo gli stessi criteri descritti nella sezione 2.6.

3.5 Previsione dei quantili nel tempo

Anche per modelli di *quantile regression* dinamica è di grande importanza poter produrre delle previsioni *out of sample* e dunque fare estrapolazione nel tempo.

L'obiettivo e il metodo da adottare sono piuttosto simili a quanto visto nella sezione 2.5 del capitolo 2. Chiaramente in questo caso non è di interesse produrre delle previsioni per il processo \mathbf{y}_t , ma per i suoi quantili \mathbf{q}_t^τ .

In questa sezione non faremo distinzione tra modelli a disturbi di misura indipendenti e correlati, poiché gli strumenti di previsione sono comuni all'intera classe di modelli quantilici dinamici specificati attraverso una forma *State Space*.

Ipotizzando di aver specificato un modello di *quantile regression* dinamica in accordo con 3.3.3 o 3.4.4 per un valore di τ fissato e definite le quantità $\mathbf{y}^{oss} = \{\mathbf{y}_t\}_{t=1}^T$, $\mathbf{q} = \{\mathbf{q}_t^\tau\}_{t=T+1}^{T+k}$, $\boldsymbol{\theta} = (\sigma_\varepsilon^2, \sigma_\eta^2)$, $\boldsymbol{\omega} = \{\boldsymbol{\omega}_t\}_{t=T+1}^{T+k}$ e $\boldsymbol{\alpha} = \{\boldsymbol{\alpha}_t\}_{t=T+1}^{T+k}$, la distribuzione predittiva per i quantili del processo \mathbf{y}_t è esprimibile secondo l'equazione

$$p(\mathbf{q} | \mathbf{y}^{oss}) = \iint_{\Theta \times \mathbf{A}^k} p(\mathbf{q} | \boldsymbol{\theta}, \boldsymbol{\alpha}) \pi(\boldsymbol{\theta}, \boldsymbol{\alpha} | \mathbf{y}^{oss}) d\boldsymbol{\theta} d\boldsymbol{\alpha}, \quad (3.5.1)$$

dove $\boldsymbol{\theta} \in \Theta$, $\boldsymbol{\alpha} \in \mathbf{A}^k$ e $\mathbf{q}_t^\tau = \mathbf{c}_t + \mathbf{Z}_t \cdot \boldsymbol{\alpha}_t$. Con

$$\pi(\boldsymbol{\theta}, \boldsymbol{\alpha} | \mathbf{y}^{oss}) = \int_{\Omega^k} \pi(\boldsymbol{\theta}, \boldsymbol{\omega}, \boldsymbol{\alpha} | \mathbf{y}^{oss}) d\boldsymbol{\omega}$$

indichiamo la distribuzione a posteriori marginale dei parametri e degli stati latenti predetti, nella quale è stato integrato l'effetto del fattore aumentato $\boldsymbol{\omega} \in \Omega^k$.

La complessità della funzione integranda e la dimensione dello spazio di integrazione non permettono di trovare una soluzione in forma chiusa per la distribuzione 3.5.1. Per approssimare l'integrale 3.5.1 è quindi necessario ricorrere a metodi Monte Carlo, basati sulla simulazione di valori casuali dalla distribuzione a posteriori $\pi(\boldsymbol{\theta}, \boldsymbol{\omega}, \boldsymbol{\alpha} | \mathbf{y}^{oss})$. Analogamente a quanto visto nella sezione 2.5 possiamo utilizzare i valori simulati via *Gibbs Sampling* come base da cui costruire le previsioni.

Definendo con $\{\boldsymbol{\theta}^{(r)}\}_{r=1}^R$ i parametri a posteriori simulati, si possono costruire delle traiettorie casuali dalla legge predittiva degli stati e dei fattori latenti seguendo il seguente schema di campionamento.

Per ogni $r = 1, \dots, R$ si generi $\boldsymbol{\omega}^{(r)}$ dalla distribuzione a priori $\pi(\boldsymbol{\omega} | \boldsymbol{\theta}^{(r)})$; successivamente si generi $\boldsymbol{\alpha}^{(r)}$ dalla distribuzione *full conditional* predittiva $\pi(\boldsymbol{\alpha} | \boldsymbol{\theta}^{(r)}, \boldsymbol{\omega}^{(r)})$. Questa è ottenibile eseguendo *Kalman filter* e *Kalman smoother*, secondo le equazioni 2.3.3 e 2.3.4, e poi simulando il processo attraverso il *Simulation Smoother* di Durbin e Koopman 2002.

Ottenute le traiettorie simulate per gli stati si può approssimare l'integrale 3.5.1 come

$$p(\mathbf{q} | \mathbf{y}^{oss}) \approx \frac{1}{R} \sum_{r=1}^R p(\mathbf{q} | \boldsymbol{\theta}^{(r)}, \boldsymbol{\alpha}^{(r)}).$$

Per ottenere delle stime puntuali per i quantili si possono poi calcolare i valori attesi a posteriori

$$\begin{aligned} \mathbb{E}[\mathbf{q} | \mathbf{y}^{oss}] &= \iint_{\Theta \times \mathbf{A}} \mathbb{E}[\mathbf{q} | \boldsymbol{\theta}, \boldsymbol{\alpha}] \pi(\boldsymbol{\theta}, \boldsymbol{\alpha} | \mathbf{y}^{oss}) d\boldsymbol{\theta} d\boldsymbol{\alpha} \\ &= \iint_{\Theta \times \mathbf{A}} (\mathbf{c} + \mathbf{Z} \cdot \boldsymbol{\alpha}) \pi(\boldsymbol{\theta}, \boldsymbol{\alpha} | \mathbf{y}^{oss}) d\boldsymbol{\theta} d\boldsymbol{\alpha} \end{aligned}$$

dove $\mathbf{c} = (\mathbf{c}_{T+1}^\top, \dots, \mathbf{c}_{T+k}^\top)^\top$ e $\mathbf{Z} = \text{diag}(\mathbf{Z}_{T+1}, \dots, \mathbf{Z}_{T+k})$. La media a posteriori può essere convenientemente stimata come

$$\mathbb{E}[\mathbf{q} | \mathbf{y}^{oss}] = \mathbf{c} + \mathbf{Z} \cdot \mathbb{E}[\boldsymbol{\alpha} | \mathbf{y}^{oss}] \approx \mathbf{c} + \mathbf{Z} \frac{1}{R} \sum_{r=1}^R \boldsymbol{\alpha}^{(r)}.$$

Capitolo 4

Applicazione su Dati Reali

4.1 I dati

I modelli e gli algoritmi presentati nei precedenti capitoli sono stati testati su un insieme di dati relativi alla velocità dei venti in Irlanda, originariamente raccolti ed analizzati da Haslett e Raftery 1989. I dati utilizzati sono pubblici e gratuitamente scaricabili dal *software opensource* R all'interno delle librerie `gstat` e `statlib`.

I dati sono costituiti da una raccolta di misurazioni medie giornaliere della velocità del vento (in radice quadrata) effettuate da dodici stazioni meteorologiche nella Repubblica d'Irlanda, come mostrato in tabella 4.1.

Il periodo di rilevazione va dal 1 gennaio 1961 al 31 dicembre 1978. La velocità del vento è misurata in nodi, dove 1 nodo equivale a 0.5418 metri al secondo.

I dati erano stati raccolti originariamente dal governo Irlandese col fine di identificare le regioni del paese più esposte a venti di forte intensità, in modo tale da sfruttare quest'informazione per posizionare strategicamente delle centrali eoliche di conversione energetica.

Per ulteriori informazioni si vedano i lavori di Gneiting 2002, Stein 2005a, Stein 2005b, Luna e Genton 2005, nei quali sono descritti in modo più approfondito i dati e in cui sono stati anche trattati metodi di analisi alternativa per processi con dipendenza spazio-temporale.

4.2 Analisi esplorative

I dati constano di 12 serie storiche geolocalizzate di lunghezza 6574, nessuna delle quali presenta valori mancanti. In figura 4.1 si mostrano le locazioni delle diverse stazioni di rilevazione entro i confini Irlandesi e alcune delle serie storiche come

Tabella 4.1: Stazioni di rilevazione del vento in Irlanda.

Stazioni	Latitudine	Longitudine	Velocità Media
Valentia (Val)	51 56' N	10 15' W	3.15
Belmullet (Bel)	54 14' N	10 00' W	3.53
Claremorris (Cla)	53 43' N	8 59' W	2.80
Shannon(Sha)	52 42' N	8 55' W	3.13
Roche's Point (Rpt)	51 48' N	8 15' W	3.42
Birr (Bir)	53 05' N	7 53' W	2.53
Mullingar (Mul)	53 32' N	7 22' W	2.81
Malin Head (Mal)	55 22' N	7 20' W	3.85
Kilkenny (Kil)	52 40' N	7 16' W	2.40
Clones (Clo)	54 11' N	7 14' W	2.84
Dublin (Dub)	53 26' N	6 15' W	3.02
Roslare (Ros)	52 16' N	6 21' W	3.33

esempio della struttura del *dataset*. Per le quattro serie mostrate sono anche state sovrainposte ai dati delle curve di liscio ottenute stimando delle *smoothing spline* per rendere più evidenti gli andamenti medi delle diverse serie, altrimenti non individuabili visivamente a causa del gran numero di tempi rilevati. La dimensione dei punti che identificano ogni sito è proporzionale alla velocità del vento media in quel luogo.

Come si può osservare dalla mappa in figura 4.1 e dalla quarta colonna della tabella 4.1, le stazioni posizionate nelle zone costiere sono soggette a venti con velocità mediamente più intensa, mentre nell'entroterra i venti assumono generalmente velocità più contenute.

Fenomeni di natura meteorologica come questo sono spesso caratterizzati da una marcata periodicità annuale, accompagnata da una forte persistenza a breve termine, che ne determina il comportamento caotico e spesso imprevedibile nel medio periodo. Entrambi questi aspetti sono facilmente interpretabili da un punto di vista climatico, ma a volte difficilmente trattabili da un punto di vista statistico.

Le dinamiche meteorologiche sono influenzate da un gran numero di fattori, che introducono una forte componente stocastica nel fenomeno. Quindi anche forme di regolarità e periodicità non sono spesso riconducibili a una qualche forma funzionale parametrica e deterministica, ma vanno trattate come processi stocastici dinamici. Osserviamo infatti che ampiezza e forma delle oscillazioni periodiche non risulta omogenea nel tempo, ma perturbata da comportamenti stocastici locali.

Dallo studio delle funzioni di autocorrelazione totale delle serie in figura 4.2 emer-

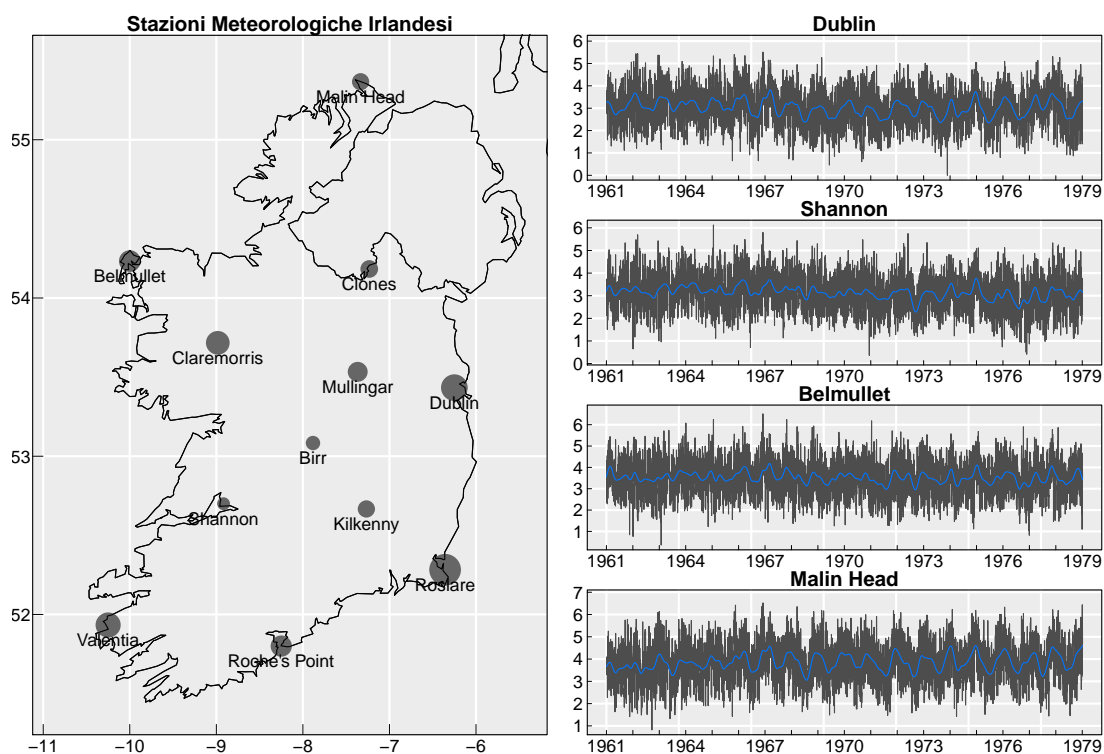


Figura 4.1: Mappa delle stazioni di rilevazione in Irlanda (a sinistra) e serie storiche della velocità del vento (a destra) collocate a Dublino (prima riga), Shannon (seconda riga), Belmullet (terza riga) e Mallin Head (quarta riga).

ge un andamento pseudo-sinusoidale, tipico di fenomeni periodici. Osserviamo però che l'intensità della correlazione periodica non è comune per tutti i siti. Ad esempio, a Dublino sembra esserci una persistenza annuale più forte che non a Shannon. Questo suggerisce che esista un forma di interazione tra l'evoluzione temporale e la correlazione spaziale che lega le diverse stazioni di rilevazione.

A riconferma di ciò, possiamo notare che le zone soggette a venti più intensi, come Dublino e Malin Head, presentano anche un'autocorrelazione stagionale più persistente, mentre nelle zone di vento più calmo, come Shannon e Belmullet, questa risulta assumere valori più contenuti.

Al netto di andamenti periodici, le serie appaiono stazionarie in media e con distribuzione approssimativamente centrata intorno a quest'ultima.

Da un punto di vista spaziale osserviamo una forte correlazione positiva tra i venti misurati in tutti i siti. Questa correlazione tende a diminuire con l'aumentare della distanza secondo un andamento non lineare, come mostrato in figura 4.3.

Per scegliere un'opportuna funzione di correlazione stazionaria e isotropica con cui modellare la dipendenza nello spazio sono stati adattati i modelli descritti in tabella 1.1 all'insieme di correlazioni empiriche misurate sui dati.

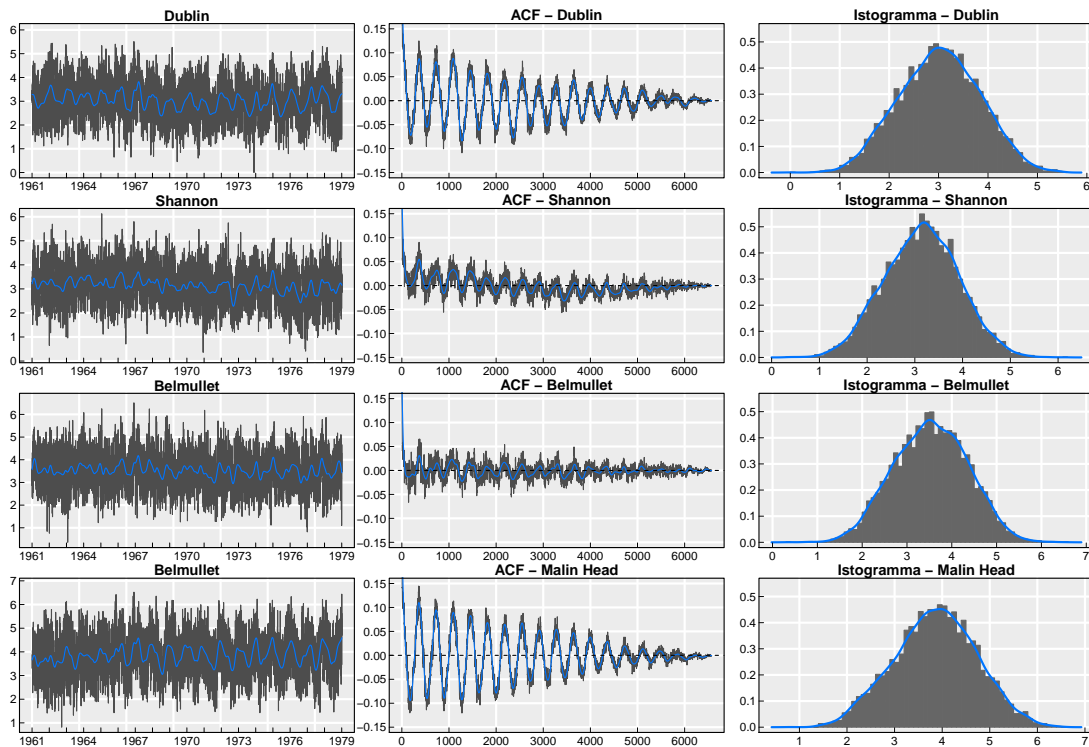


Figura 4.2: *Trace plot* (a sinistra), autocorrelazioni totali (al centro) e istogrammi (a destra) dei processi a Dublino (prima riga), Shannon (seconda riga), Belmullet (terza riga) e Mallin Head (quarta riga).

Per i modelli considerati sono stati stimati i parametri, minimizzando la distanza euclidea tra la curva predetta e le correlazioni empiriche. I valori ottenuti sono riportati in tabella 4.2.

Tabella 4.2: Parametri stimati per le funzioni di correlazione spaziale.

	Exp.	Gauss.	Sphere.	Power.	Exp.	Matern
λ	0.2021	0.0888	9.1275	0.2535	7.3891	
α	-	-	-	0.7879	-	
ν	-	-	-	-	0.3679	

Anche da una semplice analisi grafica risulta evidente che il modello Gaussiano non sia appropriato per trattare questi dati, mentre gli altri sembrano avere un adattamento buono e quasi equivalente fra loro, come si può vedere in figura 4.3. Per le analisi successive verrà adottato un modello esponenziale per la funzione di correlazione, come suggerito da Haslett e Raftery 1989, poiché di facile interpretazione e largamente utilizzato nell'ambito meteorologico.

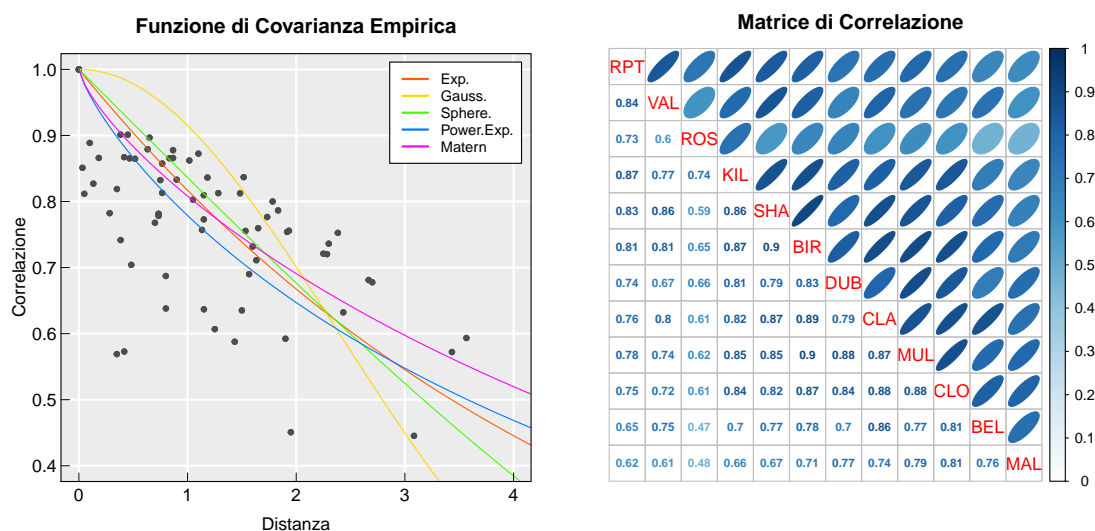


Figura 4.3: Grafico delle funzioni di correlazione spaziale (a sinistra) e matrice di correlazioni empiriche tra le serie geolocalizzate (a destra).

4.3 Modello in media

Specificazione del modello

Per modellare l'andamento medio dei venti in Irlanda si è specificato un modello di convoluzione dinamica in accordo con quanto descritto nel capitolo 2.

Questo tipo di formulazione risulta particolarmente adatta in tutte quelle situazioni in cui la struttura di correlazione spazio-temporale del fenomeno non risulta separabile e vi è la necessità di tenere conto dell'interazione tra le dimensioni spaziale e temporale.

Ipotizzando che il processo di larga scala sia regolare ed abbia delle traiettorie lisce nello spazio, si è scelto di adottare una funzione *kernel* Gaussiana bivariata

$$K(\mathbf{u}, \mathbf{v}) = (2\pi)^{-1} |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{u} - \mathbf{v})^T \Sigma^{-1} (\mathbf{u} - \mathbf{v}) \right\}. \quad (4.3.1)$$

Come è noto in letteratura geostatistica (si vedano ai lavori di Cressie 1993 e Gaetan e Guyon 2010), un processo stocastico generato da una convoluzione di *kernel* Gaussiani è uniformemente differenziabile nel suo dominio e rispetta dunque le ipotesi di regolarità richieste per un processo di larga scala spaziale.

La forma di un *kernel* Gaussiano bivariato è determinata dalla sua matrice di scala Σ . Una scelta comune è quella di assumere che la funzione *kernel* sia stazionaria e isotropica nello spazio. In accordo con l'ipotesi di stazionarietà si è scelto di considerare Σ come una matrice diagonale costante nello spazio, le cui componenti

sono determinate in maniera tale da coincidere con le varianze delle coordinate di latitudine e longitudine delle stazioni di rilevazione: $\Sigma = \text{diag}(s_{Lat}^2, s_{Long}^2)$. Questa scelta è stata fatta per rendere confrontabili i due insiemi di coordinate nel calcolo di $K(\cdot, \cdot)$, evitando quindi di assegnare pesi diversi ad una o l'altra componente. Il *kernel* così generato non è isotropico nello spazio originale, ma lo è in un spazio trasformato, genera dunque un'anisotropia geometrica indotta dalla matrice Σ .

Dato il numero esiguo di siti campionati si è ritenuto opportuno porre per ognuno di essi un nodo dell'espansione *kernel*:

$$Z(\mathbf{s}, t) = \sum_{j=1}^m K(\mathbf{u}_j, \mathbf{s}) \alpha(\mathbf{u}_j, t) = \sum_{i=1}^n K(\mathbf{s}_i, \mathbf{s}) \alpha(\mathbf{s}_i, t), \quad (4.3.2)$$

dove $m = n$ e $\{\mathbf{u}_j\}_{j=1}^m = \{\mathbf{s}_i\}_{i=1}^n$. In questo modo i diversi processi latenti diventano più interpretabili, potendo essere associati ognuno ad una specifica stazione di rilevazione. Inoltre i processi latenti localizzati nelle regioni più informative verranno automaticamente pesati maggiormente rispetto ai processi localizzati nelle regioni più periferiche e per i quali è disponibile meno informazione, come possiamo osservare in figura 4.4.

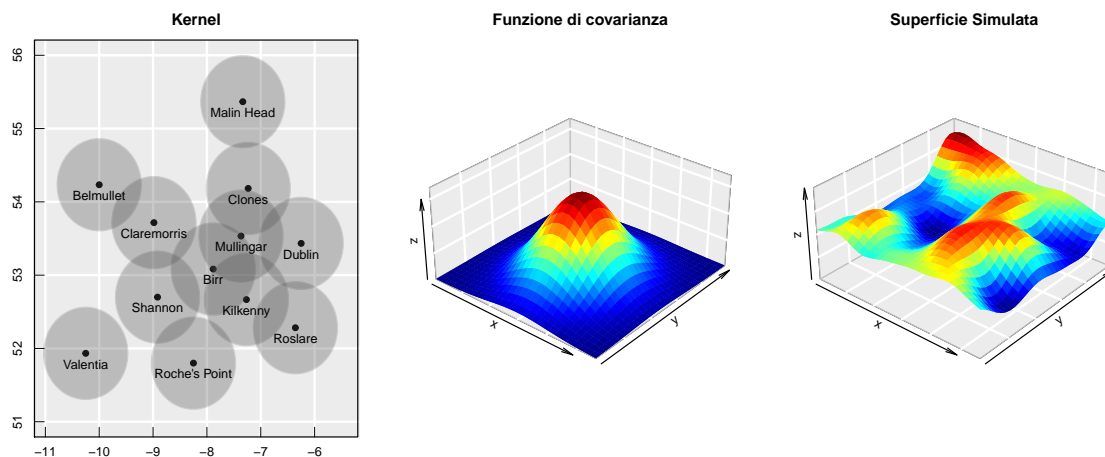


Figura 4.4: Posizione e ampiezza dei *kernel* (a sinistra), funzione di correlazione Gaussiana (al centro) e superficie spaziale simulata corrispondente (a destra).

Nel caso in cui il numero di siti fosse molto superiore sarebbe opportuno selezionare un numero inferiore di nodi su cui espandere lo sviluppo *kernel*, in modo tale da controllare la complessità del modello.

Per completare la specificazione del processo latente assumiamo indipendenza per i disturbi di stato, ottenendo quindi una matrice di correlazione $\mathbf{W}^* = \mathbf{I}_n$, da cui segue che $\mathbf{W} = (\mathbf{K}\mathbf{K})^{-1}$.

Infine per i disturbi di misura è stata adottata una funzione di correlazione esponenziale, dando luogo ad una matrice di covarianza $\mathbf{V} = (v_{ij})_{i,j=1}^n$, dove $v_{ij} = \exp(-\lambda d_{ij})$ e $d_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\|_2$ corrisponde alla distanza euclidea tra i diversi siti.

A titolo di esempio e per alleggerire il costo computazionale si è preferito fissare il parametro λ ad un valore pari a 0.17, che è stato stimato sui dati originali adottando un approccio Bayesiano empirico, anziché stimarne la densità a posteriori via *Metropolis Hastings*.

Come emerso nelle analisi esplorative i dati presentano una forma di periodicità annuale comune e non deterministica. Per modellare adeguatamente questa caratteristica possiamo formulare un modello *State Space* analogo a quello discusso nel secondo capitolo, con vettore degli stati ampliato opportunamente in modo da inserire un termine periodico.

Possiamo scrivere l'equazione di misura come segue

$$\mathbf{y}_t = \mathbf{X} \boldsymbol{\psi}_t + \mathbf{K} \boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t,$$

dove $\mathbf{X} = (\mathbf{1}_n, \mathbf{0}_n)$ e $\boldsymbol{\psi}_t = (\psi_t, \psi_t^*)$. Il vettore latente $\boldsymbol{\psi}_t$ definisce la prima armonica della rappresentazione di Fourier per processi periodici discussa in Koopman e Lee 2009 ed evolve secondo la dinamica descritta nelle equazioni 4.3.3:

$$\begin{aligned} \psi_{t+1} &= \cos(\vartheta) \cdot \psi_t + \sin(\vartheta) \cdot \psi_t^* + \kappa_t, & \kappa_t &\sim \mathcal{N}(0, \sigma_\psi^2), \\ \psi_{t+1}^* &= -\sin(\vartheta) \cdot \psi_t + \cos(\vartheta) \cdot \psi_t^* + \kappa_t^*, & \kappa_t^* &\sim \mathcal{N}(0, \sigma_\psi^2), \end{aligned} \quad (4.3.3)$$

dove $\vartheta = \pi/2n$ è la frequenza fondamentale di campionamento, e $n = 365$ è il numero di osservazioni campionate in un anno. Si assume che i disturbi (κ_t, κ_t^*) siano indipendenti tra loro e nel tempo.

Secondo le ipotesi formulate sul processo possiamo scrivere la rappresentazione *State Space* del modello completo nel modo seguente:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{Z} \cdot \boldsymbol{\xi}_t + \boldsymbol{\varepsilon}_t, & \boldsymbol{\varepsilon}_t &\sim \mathcal{N}(\mathbf{0}, \mathbf{H}), \\ \boldsymbol{\xi}_{t+1} &= \mathbf{T} \cdot \boldsymbol{\xi}_t + \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim \mathcal{N}(\mathbf{0}, \mathbf{Q}), \\ & & \boldsymbol{\xi}_0 &\sim \mathcal{N}(\mathbf{0}, k \mathbf{I}_{n+2}), \end{aligned} \quad (4.3.4)$$

dove

- $\mathbf{y}_t = (y_{1,t}, \dots, y_{n,t})$ è il vettore di osservazioni spaziali misurate al tempo t ;
- $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \dots, \varepsilon_{n,t})$ è il vettore di disturbi di misura;
- $\boldsymbol{\xi}_t = (\alpha_{1,t}, \dots, \alpha_{n,t}, \psi_t, \psi_t^*)$ è il vettore di stati latenti;

- $\boldsymbol{\eta}_t = (\eta_{1,t}, \dots, \eta_{n,t}, \kappa_t, \kappa_t^*)$ è il vettore di disturbi di stato;
- $\mathbf{Z} = (\mathbf{K}, \mathbf{X})$ è la matrice di misura;
- $\mathbf{H} = \sigma_\varepsilon^2 \mathbf{V}$ è la matrice di covarianza dei disturbi di misura;
- $\mathbf{T} = \begin{pmatrix} \mathbf{I}_n & \mathbf{O} \\ \mathbf{O} & \mathbf{C} \end{pmatrix}$ è la matrice di transizione, con $\mathbf{C} = \begin{pmatrix} \cos(\vartheta) & \sin(\vartheta) \\ -\sin(\vartheta) & \cos(\vartheta) \end{pmatrix}$ matrice della rappresentazione di Fourier per il processo periodico latente;
- $\mathbf{Q} = \begin{pmatrix} \sigma_\alpha^2 \mathbf{W} & \mathbf{O} \\ \mathbf{O} & \sigma_\psi^2 \mathbf{I}_2 \end{pmatrix}$ è la matrice di covarianza dei disturbi di stato.

Elicitazione delle a priori

I parametri incogniti del modello sono $(\sigma_\varepsilon^2, \sigma_\alpha^2, \sigma_\psi^2, \{\boldsymbol{\xi}_t\}_{t=0}^T)$, di questi $\{\boldsymbol{\xi}_t\}_{t=0}^T$ costituiscono parametri di interesse, mentre $(\sigma_\varepsilon^2, \sigma_\alpha^2, \sigma_\psi^2)$ sono parametri di disturbo strumentali, utili a stimare adeguatamente i processi latenti e conseguentemente il processo di grande scala spazio-temporale.

Tra i parametri di disturbo $(\sigma_\alpha^2, \sigma_\psi^2)$ rivestono un ruolo molto delicato, poiché governano il grado di adattamento del modello ai dati e sono interpretabili come parametri di regolazione nella stima dei processi latenti non stazionari $\boldsymbol{\alpha}_t$ e $\boldsymbol{\psi}_t$. Più le varianze dei disturbi di stato aumentano, più aumenta il grado di flessibilità dei processi latenti, specialmente se questi seguono una dinamica non stazionaria. In molti casi il massimo della verosimiglianza, e conseguentemente della distribuzione a posteriori congiunta, porta ad un sovradattamento, specialmente se il numero di osservazioni a disposizione è molto elevato.

Per evitare che questo avvenga esistono alcune soluzioni: la prima è di calibrare i parametri delle distribuzioni a priori in modo tale da comprimerne la variabilità. In questo modo la legge a posteriori verrà aggiornata dalla verosimiglianza, ma sotto il vincolo imposto da una a priori stringente. Questo approccio si dice di *tuning* Bayesiano, poiché prevede di adottare tecniche di inferenza Bayesiana per selezionare un valore opportuno per dei parametri di regolazione.

Una seconda possibilità è quella di trattare questi parametri effettivamente come parametri di calibrazione e selezionarli attraverso criteri di scelta automatica, come, ad esempio, attraverso convalida incrociata.

Una terza possibilità è quella di trattare $(\sigma_\alpha^2, \sigma_\psi^2)$ come iperparametri della distribuzione di probabilità a priori che governa la dinamica dei processi latenti, come

consigliato da Stroud, Müller e Sansó 2001. Questo approccio fornisce un metodo immediato per porre vincoli sull'evoluzione dei processi latenti a fronte di una conoscenza a priori o di specifiche assunzioni fatte sul modello.

Nel seguito di questa tesi verrà adottato questo terzo metodo, poiché si è voluto introdurre un vincolo di regolarità sulle traiettorie dei processi latenti non stazionari, in modo tale che queste risultino quanto più lisce e poco erratiche.

Per il parametro σ_ε^2 è stata selezionata una legge a priori Gamma Inversa di parametri $a_\varepsilon = 2.01$ e $b_\varepsilon = 1.01$, che corrisponde ad una distribuzione di media 1 e varianza 100. Per i disturbi di stato sono stati fissati $\sigma_\alpha^2 = 0.01$ e $\sigma_\psi^2 = 0.001$, in modo tale che le fluttuazioni stocastiche attorno al processo periodico siano contenute e quest'ultimo risulti quanto più regolare possibile. Anche per i processi di convoluzione dinamica si assume varianza ridotta, ma con maggiore libertà in modo che possano cogliere meglio dinamiche locali e cambiamenti climatici improvvisi.

Simulazione e risultati

La distribuzione congiunta a posteriori per $(\sigma_\varepsilon^2, \{\xi_t\}_{t=0}^T)$ è stata stimata via *Gibbs Sampling* secondo l'algoritmo descritto nel capitolo 2. Sono state effettuate 1200 replicazioni dell'algoritmo di simulazione e di queste sono state scartate le prime 200, come fase di *burnin* della catena.

In figura 4.5 osserviamo alcune diagnostiche di convergenza della catena per il parametro di scala σ_ε^2 . Dai grafici osserviamo che la catena raggiunge la convergenza molto velocemente e presenta una ridotta autocorrelazione totale. Inoltre la distribuzione a posteriori marginale a cui converge si presenta approssimativamente simmetrica e molto concentrata intorno al valore modale, pari a circa 0.4315. La deviazione standard a posteriori stimata è pari a 0.00237.

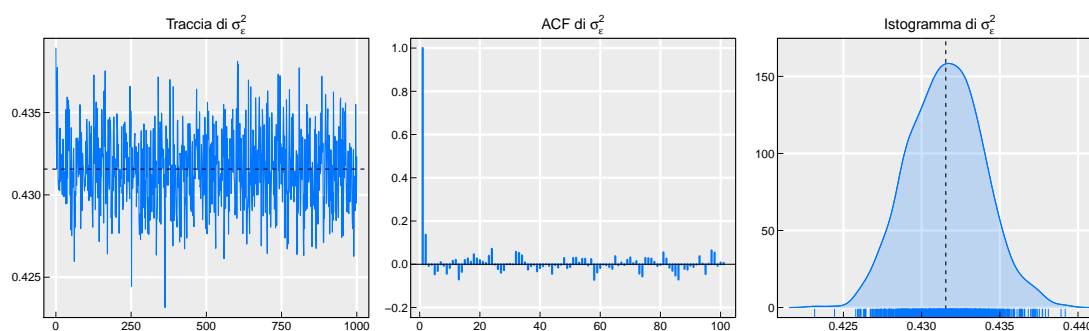


Figura 4.5: *Trace plot* della catena simulata dopo il periodo di *burnin* (a sinistra), grafico di autocorrelazione totale (al centro) e densità a posteriori marginale stimata (a destra).

Il secondo set di parametri simulati dal *Gibbs Sampler* è costituito dagli stati dinamici, le cui traiettorie stimate sono state riportate in figura 4.6. Ognuna di queste è stata ottenuta mediando i processi simulati ad ogni tempo.

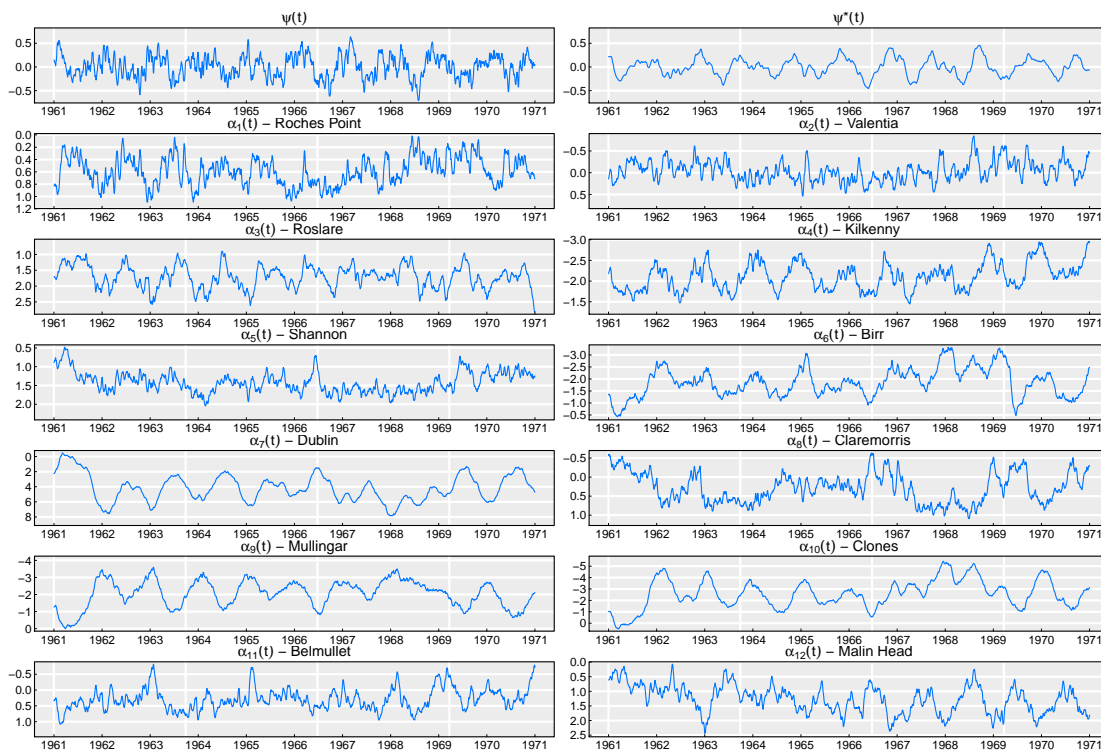


Figura 4.6: Stati dinamici latenti medi ottenuti via simulazione.

Dai grafici 4.6 osserviamo un fenomeno interessante. Le dinamiche dominanti nei processi latenti sono tutte periodiche o pseudo-periodiche. I processi di Fourier ψ_t sembrano catturare una parte della periodicità comune, ma di ampiezza contenuta, mentre i processi dinamici α_t hanno il ruolo di adattarsi alle fluttuazioni locali, ampliando o riducendo l'effetto della periodicità di base. Per esempio, il processo localizzato a Dublino ha una magnitudine decisamente superiore in modulo rispetto a quello situato a Shannon e questo perché a Dublino la persistenza stagionale è più forte, così come anche l'intensità media del vento.

Considerazioni simili possono valere anche per altri siti, è però importante ricordare che il processo spaziale complessivo è dato dalla composizione e dall'interazione *kernel* tra tutti gli stati latenti e non da singoli processi e questo chiaramente cambia l'interpretazione globale del modello.

In figura 4.7 possiamo osservare l'adattamento generale del modello, che complessivamente sembra cogliere bene l'andamento generale dei dati, i quali hanno distribuzione approssimativamente simmetrica intorno ad un valore centrale, come possiamo vedere in figura 4.9, dall'analisi dei residui.

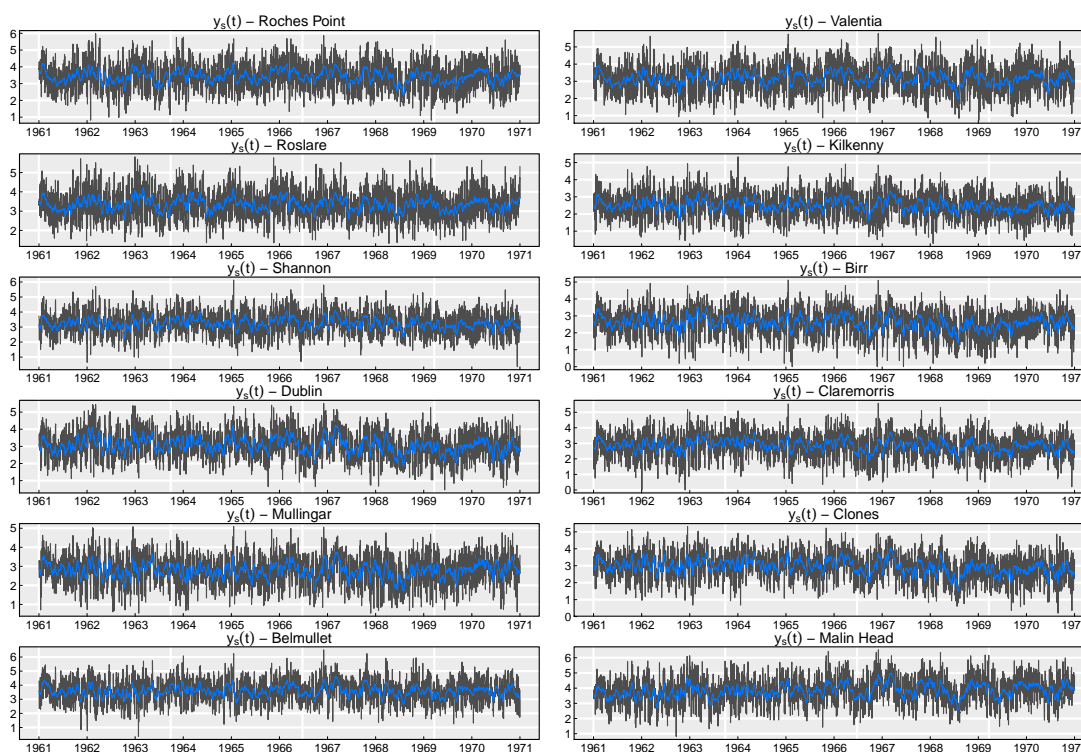


Figura 4.7: Velocità dei venti originale contro il processo stimato come media delle previsioni ottenute via MCMC.

Andando a studiare il comportamento dei residui in 4.9, ci si accorge che per alcune serie si manifesta un fenomeno di eteroschedasticità periodica. Per meglio indagare questo aspetto è utile applicare una trasformazione quadratica ai residui, ed utilizzare questi ultimi come surrogato delle varianze dinamiche nel tempo. Applicando questa trasformazione la volatilità periodica diventa immediatamente più evidente, sia dai grafici dei residui trasformati contro il tempo, sia dalle funzioni di autocorrelazione di questi ultimi, come possiamo notare dai grafici in figura 4.10.

Questo comportamento eteroschedastico, maggiormente evidente in alcuni siti, piuttosto che in altri, fa decadere una delle ipotesi di base del modello *State Space* adottato, ovvero di stazionarietà in varianza delle innovazioni.

L'aspetto in questione potrebbe essere dovuto ad un adattamento non ottimale del modello ai dati oppure ad una caratteristica intrinsecamente legata alle dinamiche del fenomeno.

Per modellare forme di *volatility clustering* e di eteroschedasticità condizionale esistono molti metodi statistici largamente studiati in letteratura, in particolare provenienti da ambiti di ricerca economico-finanziari. La classe di modelli più popolare per gestire questo tipo di fenomeni è costituita dalla famiglia degli ARCH (*AutoRegressive Conditional Heteroskedasticity*) e dei GARCH (*Generalized ARCH*). Come

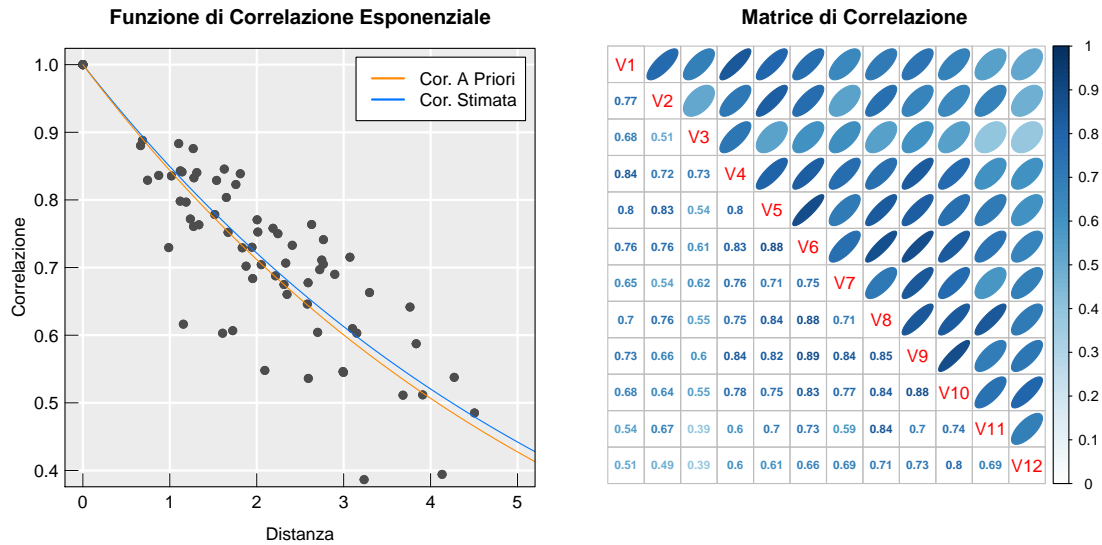


Figura 4.8: Funzione di correlazione spaziale tra i residui (a sinistra) e matrice di correlazione corrispondente (a destra).

per i modelli ARMA, anche i modelli ARCH e GARCH ammettono una rappresentazione *State Space*, ma secondo una formulazione non lineare e non Gaussiana (si veda Durbin e Koopman 2012).

La perdita delle ipotesi di linearità e Gaussianità in un modello *State Space* comporta una serie di complicazioni; gli algoritmi di filtraggio e liscio perdono le proprietà di ottimalità e portano ad avere soluzioni non più in forma esatta, ma soltanto approssimata (si veda, ad esempio, Durbin e Koopman 2000). Inoltre il peso computazionale richiesto per ottenere una stima può aumentare considerevolmente.

Un approccio alternativo alla specificazione di un modello appositamente studiato per gestire fenomeni di instabilità in varianza è quello di modellare direttamente i quantili della distribuzione spazio-temporale del processo, attraverso un modello di *quantile regression* dinamica.

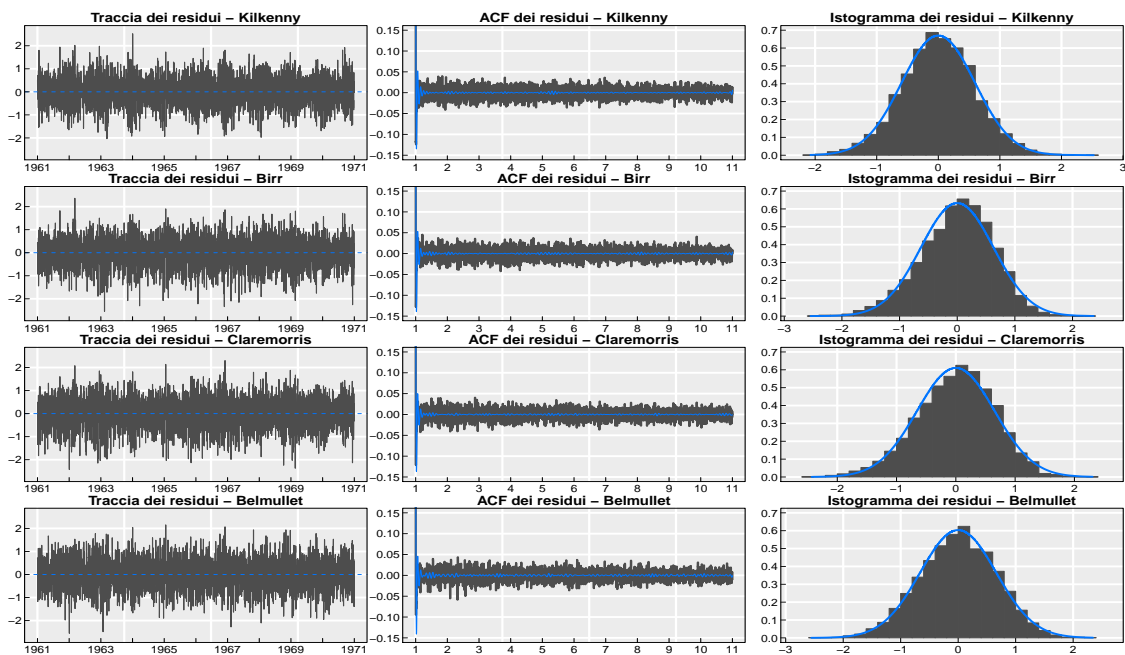


Figura 4.9: Residui del modello *State Space* in media (a sinistra), autocorrelazioni totali (al centro) e istogrammi (a sinistra) per le seguenti località: Kilkenny (prima riga), Birr (seconda riga), Claremorris (terza riga) e Belmullet (quarta riga).

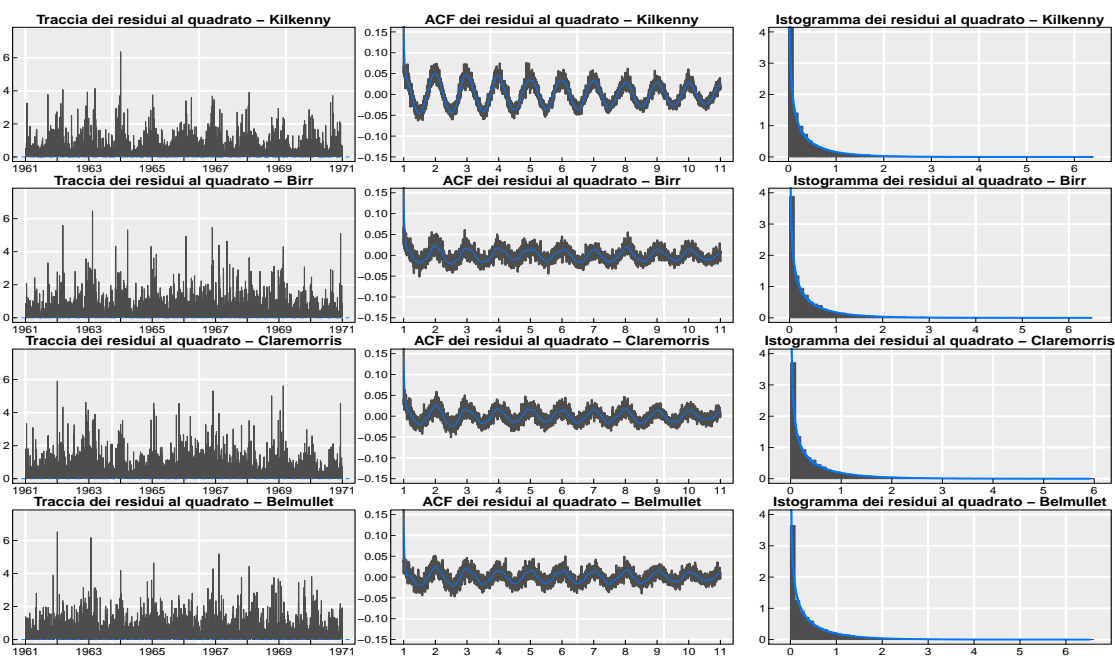


Figura 4.10: Residui al quadrato del modello *State Space* in media (a sinistra), autocorrelazioni totali (al centro) e istogrammi (a sinistra) per le seguenti località: Kilkenny (prima riga), Birr (seconda riga), Claremorris (terza riga) e Belmullet (quarta riga).

4.4 Modello quantilico a disturbi indipendenti

Specificazione del modello

Il modello di *quantile regression* spazio-temporale che si è deciso di utilizzare assume la forma *State Space*:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{Z} \cdot \boldsymbol{\xi}_t + \lambda \cdot \boldsymbol{\omega}_t + \mathbf{G}_t \cdot \boldsymbol{\nu}_t, & \boldsymbol{\nu}_t &\sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n) \\ \boldsymbol{\xi}_{t+1} &= \mathbf{T} \cdot \boldsymbol{\xi}_t + \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim \mathcal{N}(\mathbf{0}, \mathbf{Q}) \\ \boldsymbol{\xi}_0 &\sim \mathcal{N}(\mathbf{0}, k \mathbf{I}_{n+2}) \end{aligned} \quad (4.4.1)$$

dove $\boldsymbol{\omega}_t = (\omega_{1,t}, \dots, \omega_{n,t})$ e $\mathbf{G}_t = \text{diag}(\delta\sqrt{\sigma_\varepsilon^2\omega_{1,t}}, \dots, \delta\sqrt{\sigma_\varepsilon^2\omega_{n,t}})$ con $\omega_{i,t} \sim \mathcal{Exp}(\sigma_\varepsilon^{-2})$ e $\omega_{i,t'} \perp \omega_{j,t''}$ per ogni $i \neq j$ e $t' \neq t''$.

Il processo di larga scala e la rappresentazione *State Space* sono specificate con gli stessi criteri utilizzati per il modello in media, così come le leggi a priori e l'elicitazione degli iperparametri.

Simulazione e risultati

Per caratterizzare adeguatamente la distribuzione del processo si è deciso di stimare cinque quantili, corrispondenti ai valori di τ pari a 0.1, 0.25, 0.5, 0.75 e 0.9.

Per ogni quantile stimato si è effettuata una simulazione adottando l'algoritmo di *Gibbs Sampling* descritto in sezione 3.4.3, e replicando la procedura per 1200 iterazioni, di cui le prime 200 sono state utilizzate come periodo di *burnin* della catena.

In figura 4.11 troviamo i grafici diagnostici relativi alla simulazione del parametro di scala σ_ε^2 per i cinque quantili stimati, dopo aver scartato le prime 200 iterazioni di *burnin*. Osserviamo che la catena presenta un buon *mixing* e una ridotta autocorrelazione. La distribuzione a posteriori marginale è approssimativamente simmetrica e concentrata attorno ad un valore modale. Valor medio e deviazione standard stimati per le cinque catene risulta sono riportati in tabella 4.3.

Tabella 4.3: Medie e Deviazioni Standard a posteriori per i parametri di scala σ_ε^2 del modello 4.4.1.

	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.9$
Valore Atteso	0.0562	0.0988	0.1222	0.0960	0.0538
Deviazione Standard	0.000246	0.000398	0.000475	0.000379	0.000228

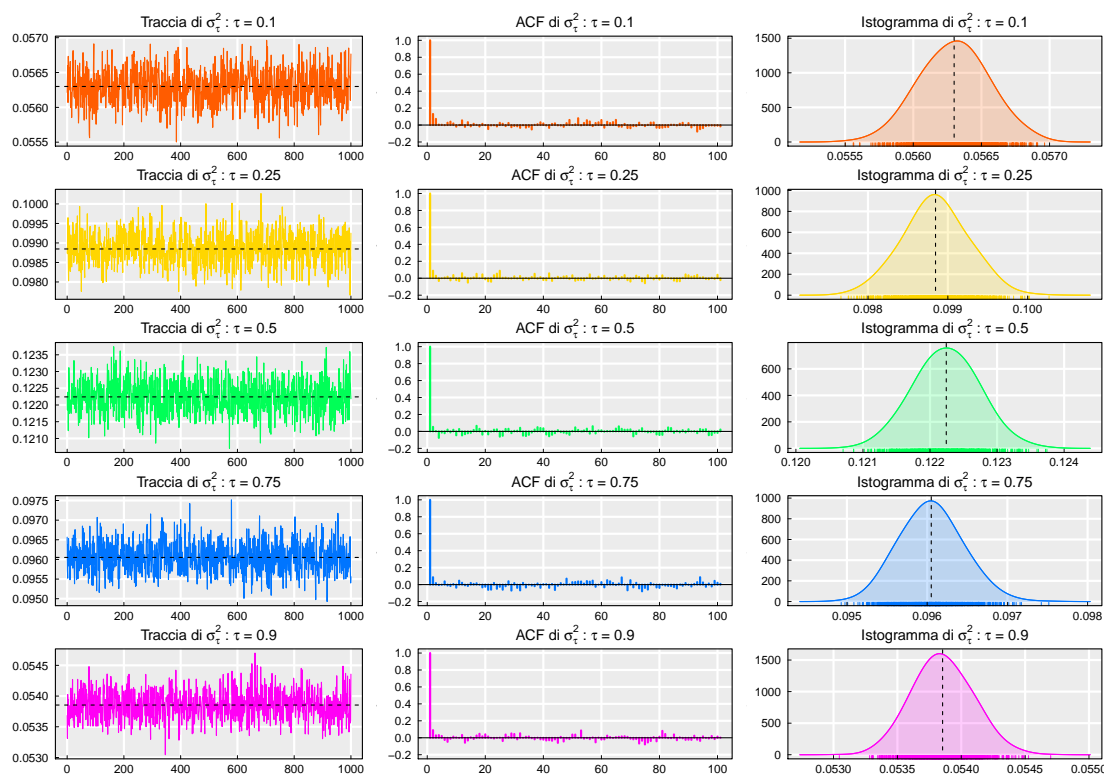


Figura 4.11: *Trace plot* delle catene simulate (a sinistra), autocorrelazione totale (al centro) e densità a posteriori marginale stimata (a destra) per i quantili di ordine: $\tau = 0.1$ (prima riga, arancione), $\tau = 0.25$ (seconda riga, giallo), $\tau = 0.5$ (terza riga, verde), $\tau = 0.75$ (quarta riga, azzurro), $\tau = 0.9$ (quinta riga, viola).

Si noti che in un contesto di *quantile regression* il parametro σ_ε^2 non è più interpretabile come varianza dei disturbi di misura, che complessivamente è data dal prodotto $\delta^2 \sigma_\varepsilon^2 \omega_{i,t}$. Non sorprende quindi che il valore centrale della distribuzione a posteriori per quest'ultimo non sia direttamente confrontabile con il valore ottenuto dalla stima del modello in media.

Andando poi a studiare gli stati latenti in figura 4.12 osserviamo un comportamento simile a quanto visto per il modello in media. Alcuni stati tendono a rafforzare la periodicità di base modellata dal processo di Fourier, mentre altri permettono di catturare fluttuazioni di altra natura. L'andamento generale dei processi latenti è simile per tutti i cinque quantili stimati e talvolta la differenza tra questi è data da una semplice traslazione, che permette di definire un corretto ordinamento tra i quantili.

In figura 4.13 si può osservare l'adattamento del modello ai dati per il primo anno di rilevazione. I quantili stimati sembrano caratterizzare adeguatamente la distribuzione dinamica della velocità del vento nei diversi siti. In particolare si nota come

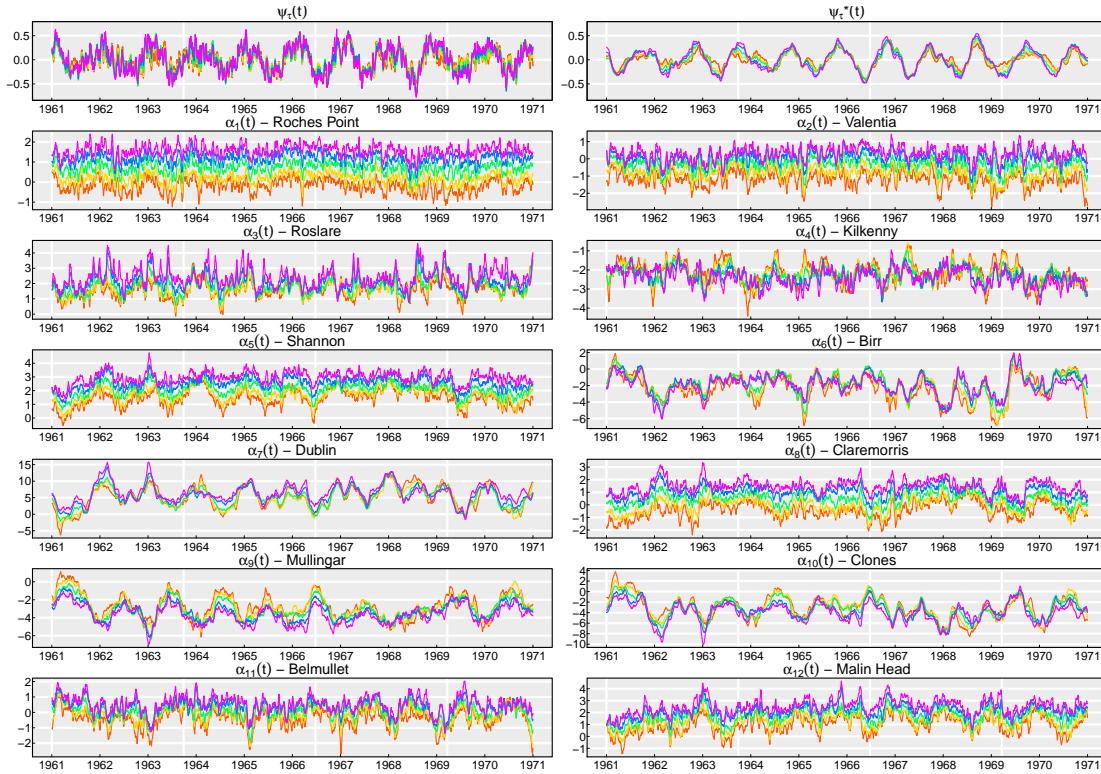


Figura 4.12: Stati latenti medi ottenuti via simulazione per i primi 10 anni di osservazione e per i quantili di ordine: $\tau = 0.1$ (arancione), $\tau = 0.25$ (giallo), $\tau = 0.5$ (verde), $\tau = 0.75$ (azzurro), $\tau = 0.9$ (viola).

questi riescano a catturare non solo dinamiche collegate alla media del processo, ma anche alla sua variabilità.

I quantili rispettano i vincoli di non incrocio, pur non essendo stati imposti esplicitamente in fase di stima. Questo fatto permette di interpretare effettivamente le curve stimate come dei quantili. In generale nulla garantisce che quantili stimati separatamente non si incrocino, dunque bisogna prestare molta attenzione nel momento in cui si vanno a trarre conclusioni sul modello ed è sempre opportuno fare dei controlli per individuare eventuali problemi e risolverli.

Per valutare la bontà di adattamento del modello ai dati è stato calcolato il numero medio di volte in cui un'osservazione di trova al di sotto del livello teorico τ fissato per il quantile, ovvero

$$\hat{p}^\tau(\mathbf{s}) = \frac{1}{T} \sum_{t=1}^T \mathbb{1}[y_t(\mathbf{s}) < \hat{Q}_t^\tau(\mathbf{s})] \quad (4.4.2)$$

In tabella 4.4 e in figura 4.18 sono stati riportati i valori ottenuti per la probabilità stimata 4.4.2. Dai risultati ottenuti si può notare che, in generale, i quantili stimati sono leggermente compressi verso la mediana e questo è tanto più evidente, quanto

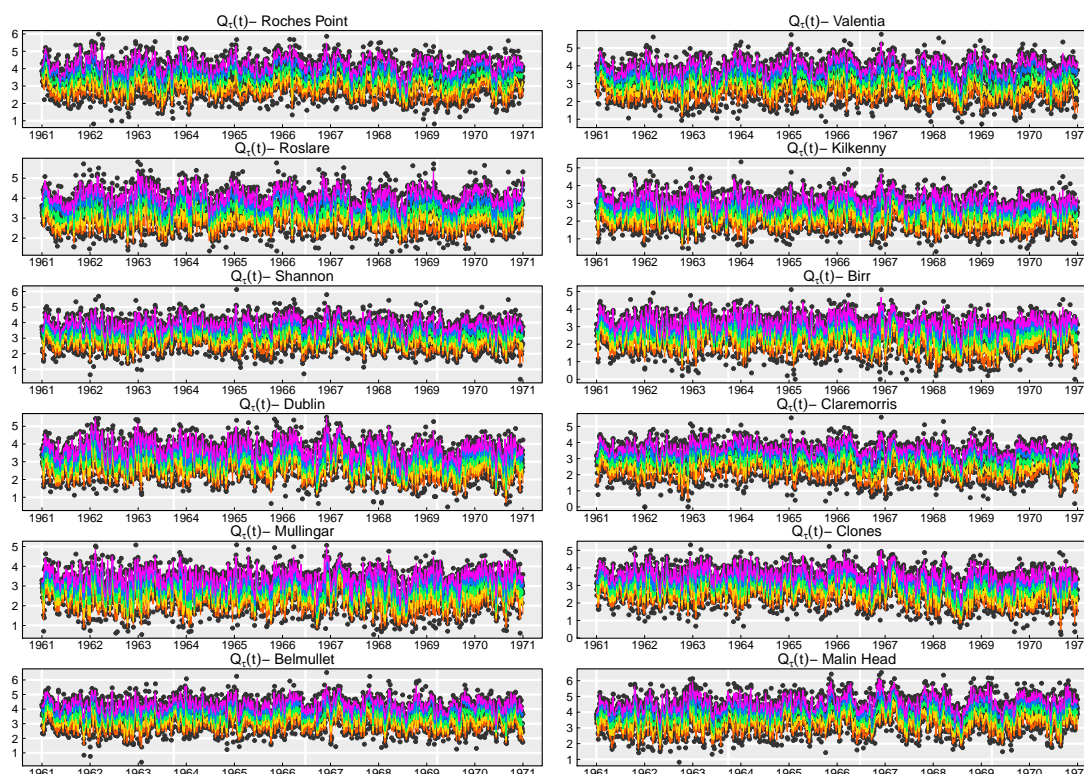


Figura 4.13: Velocità dei venti originale per i primi dieci anni di osservazione contro i quantili stimati di ordine: $\tau = 0.1$ (arancione), $\tau = 0.25$ (giallo), $\tau = 0.5$ (verde), $\tau = 0.75$ (azzurro), $\tau = 0.9$ (viola).

più il valore di τ è estremo, ovvero vicino a 0 o 1. Dei cinque quantili considerati, la mediana è chiaramente la funzione che viene stimata in maniera più precisa. Questo fenomeno è tipico nell'ambito della *quantile regression* e può essere dovuto ad una specificazione non ottimale del modello, ma anche da un'oggettiva difficoltà che si incontra nel momento in cui si cerca di stimare quantità legate alle code di una distribuzione ignota.

Disponendo di una stima per i quantili dinamici si possono costruire degli indicatori per studiare aspetti importanti della distribuzione in funzione del tempo. In genere si è interessati a studiare caratteristiche della distribuzione quali posizione, variabilità, asimmetria e curtosi.

In questo caso è di particolare interesse studiare la variabilità del processo per verificare, ed eventualmente quantificare, se effettivamente sussista una qualche forma di non stazionarietà periodica in varianza, come emerso dalle analisi dei residui del modello in media. Seguendo l'approccio proposto da Koenker e Bassett 1982, si possono studiare le serie costruite come differenza tra quantili simmetrici attraverso

Tabella 4.4: Numero medio di volte che un'osservazione si trova al di sotto del quantile stimato di ordine τ , per i quantili stimati usando il modello 4.4.1.

Stazioni	Probabilità stimate				
	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.9$
Valentia (Val)	0.1326	0.2650	0.5006	0.7274	0.8666
Belmullet (Bel)	0.1411	0.2756	0.5003	0.7271	0.8552
Claremorris (Cla)	0.1336	0.2694	0.4994	0.7270	0.8593
Shannon(Sha)	0.1463	0.2796	0.4998	0.7241	0.8515
Roche's Point (Rpt)	0.1357	0.2686	0.4983	0.7311	0.8661
Birr (Birr)	0.1617	0.2939	0.4970	0.7069	0.8324
Mullingar (Mul)	0.1585	0.2922	0.4962	0.7038	0.8333
Malin Head (Mal)	0.1427	0.2782	0.5078	0.7263	0.8567
Kilkenny (Kil)	0.1562	0.2851	0.5003	0.7169	0.8445
Clones (Clo)	0.1580	0.2857	0.5018	0.7098	0.8394
Dublin (Dub)	0.1701	0.2984	0.5008	0.7092	0.8302
Roslare (Ros)	0.1441	0.2770	0.5024	0.7250	0.8485

la funzione:

$$\Delta_{\tau}Q(t) = |Q_{\tau}(t) - Q_{1-\tau}(t)|$$

dove $Q_{\tau}(t)$ è il quantile di ordine τ al tempo t .

Dai quantili stimati si sono quindi calcolate le serie $\Delta^{(1)}Q(t) = \Delta_{0.9}Q(t)$ e $\Delta^{(2)}Q(t) = \Delta_{0.75}Q(t)$ per ogni sito e sono state riportate in figura 4.21.

Si può notare che effettivamente in *range* interquantilico tende ad avere un comportamento oscillatorio, con un picco ogni fine anno e un minimo approssimativamente a fine estate. Questo fenomeno non si ripete sempre con la stessa intensità, ma presenta una forma di periodicità stocastica correlata al livello medio della velocità del vento, che anch'essa tende ad aumentare in inverno, per decrescere in estate.

Il modello appena proposto permette di estrarre molta più informazione dai dati rispetto al più elementare modello *State Space* in media. La stima di un buon numero di quantili può infatti permettere di caratterizzare non solo media e varianza del processo, ma tutta la sua distribuzione di probabilità.

Con il modello 4.4.1 stiamo ipotizzando che, al netto del processo di larga scala, il vento delle diverse regioni sia indipendente e che quindi non esistano forme di correlazione di piccola scala. Quest'assunzione in ambito geostatistico spesso non è credibile e va dunque generalizzata. La non corretta specificazione della matrice di

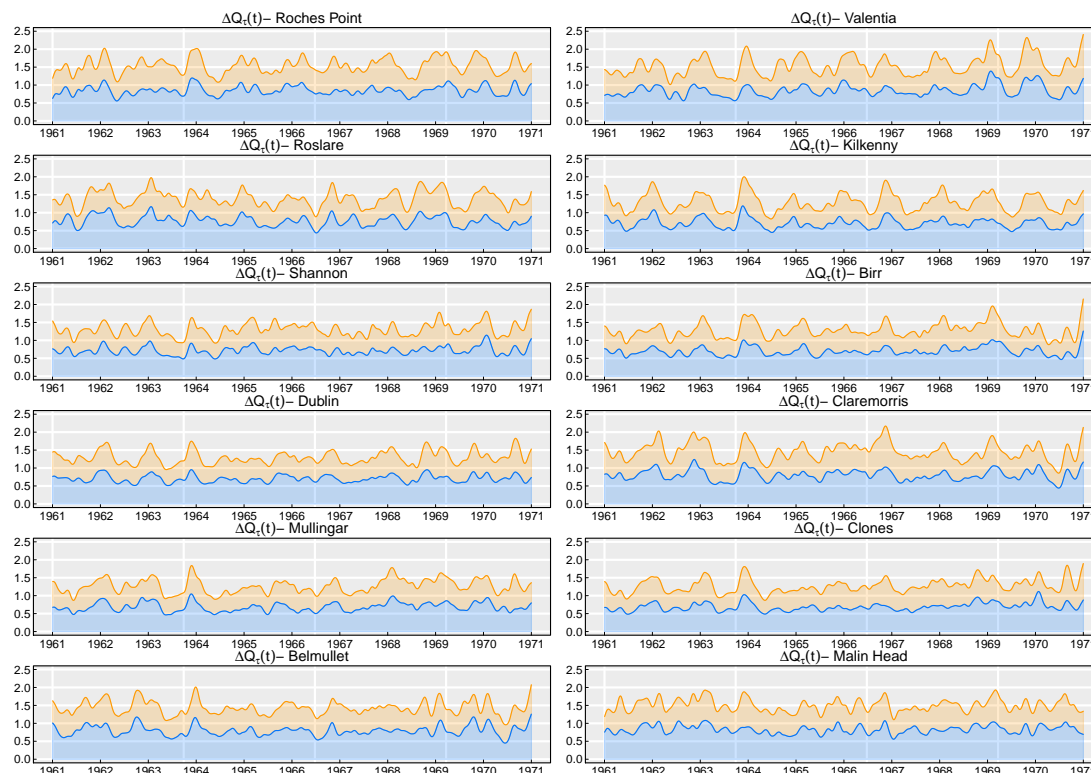


Figura 4.14: *Interquantile range* stimato per i primi dieci anni di osservazione. La linea arancione raffigura $\Delta^{(1)}Q(t)$ e la linea azzurra tratteggiata $\Delta^{(2)}Q(t)$.

correlazione residua può inficiare non solo la stima dei parametri e degli stati latenti, ma anche l'interpretazione stessa del modello e dei risultati.

4.5 Modello quantilico a disturbi correlati

Specificazione del modello

Per tenere in considerazione la correlazione tra i disturbi di misura possiamo ricorrere alla rappresentazione stocastica per la distribuzione Laplace Asimmetrica Multivariata introdotta nella sottosezione 3.4.2 e specificare un modello *State Space* più generale, descritto dalle equazioni strutturali 4.5.1:

$$\begin{aligned}
 \mathbf{y}_t &= \mathbf{Z} \cdot \boldsymbol{\xi}_t + \boldsymbol{\lambda} \cdot \omega_t + \mathbf{G}_t \cdot \boldsymbol{\nu}_t, & \boldsymbol{\nu}_t &\sim \mathcal{N}(\mathbf{0}, \mathbf{H}), \\
 \boldsymbol{\xi}_{t+1} &= \mathbf{T} \cdot \boldsymbol{\xi}_t + \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim \mathcal{N}(\mathbf{0}, \mathbf{Q}), \\
 & & \boldsymbol{\xi}_0 &\sim \mathcal{N}(\mathbf{0}, k \mathbf{I}_{n+2}),
 \end{aligned} \tag{4.5.1}$$

dove $\mathbf{G}_t = \delta \sqrt{\omega_t} \mathbf{I}_n$, con $\omega_t \sim \text{Exp}(\sigma_\varepsilon^{-2})$ e $\omega_{t'} \perp \omega_{t''}$ per ogni $t' \neq t''$.

Simulazione e risultati

I quantili che si è deciso di stimare sono gli stessi adottati nella sezione precedente, così come le matrici del modello sono rimaste inalterate, eccezion fatta per la matrice di covarianza dei disturbi di misura, ora non più diagonale, ma pari a $\mathbf{H} = \sigma_\varepsilon^2 \mathbf{V}$, dove \mathbf{V} è parametrizzata da una funzione di covarianza esponenziale come descritto nella sottosezione 4.3.

Anche in questo caso sono state effettuate 1200 replicazioni del *Gibbs Sampling* descritto in sezione 3.4.3, di cui le prime 200 sono state utilizzate come periodo di *burnin* della catena e quindi scartate.

L'introduzione di una struttura di dipendenza spaziale tra i disturbi porta ad avere un'autocorrelazione più marcata nelle catene simulate per i parametri σ_ε^2 sui diversi quantili, e di conseguenza un *mixing* meno erratico, come si può notare in figura 4.15. I valori a cui convergono le medie e le deviazioni standard a posteriori di σ_ε^2 sono riportati in tabella 4.5.

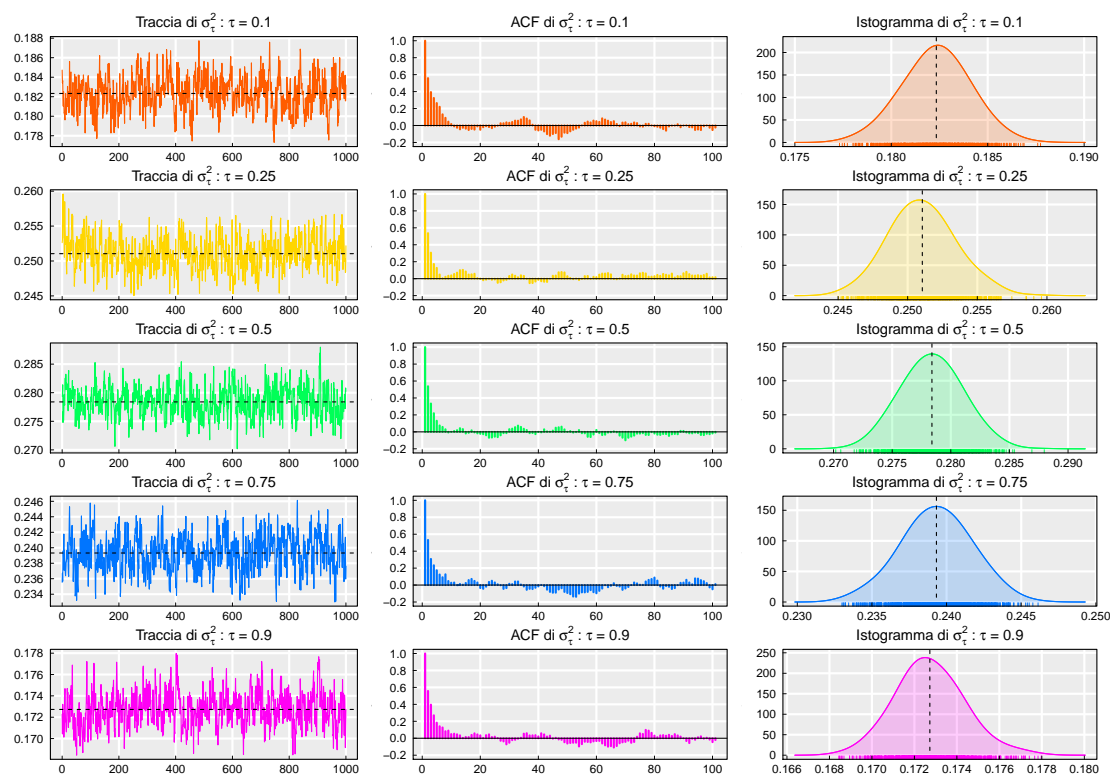


Figura 4.15: *Trace plot* delle catene simulate (a sinistra), autocorrelazione totale (al centro) e densità a posteriori marginale stimata (a destra) per i quantili di ordine: $\tau = 0.1$ (prima riga, arancione), $\tau = 0.25$ (seconda riga, giallo), $\tau = 0.5$ (terza riga, verde), $\tau = 0.75$ (quarta riga, azzurro), $\tau = 0.9$ (quinta riga, viola).

Tabella 4.5: Medie e Deviazioni Standard a posteriori per i parametri di scala σ_ε^2 del modello 4.5.1.

	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.9$
Valore Atteso	0.1823	0.2510	0.2783	0.2393	0.1727
Deviazione Standard	0.001722	0.002343	0.002586	0.002328	0.001594

Confrontando le tabelle 4.3 e 4.5 osserviamo che i valori ottenuti per le medie a posteriori nei due modelli quantilici sono sostanzialmente diverse, soprattutto in relazione alla variabilità delle due distribuzioni a posteriori.

In questo modello più generale osserviamo anche una sostanziale differenza nei processi latenti simulati (si veda la figura 4.16). I processi della rappresentazione di Fourier risultano avere delle traiettorie più lisce e regolari, mentre i processi di convoluzione dinamica presentano andamenti più irregolari e non necessariamente legati ad una qualche forma di periodicità locale. Inoltre quasi tutti gli stati latenti sono ordinati secondo il livello del quantile a loro associato e di rado si incrociano, sebbene alcuni siano molto vicini tra loro.

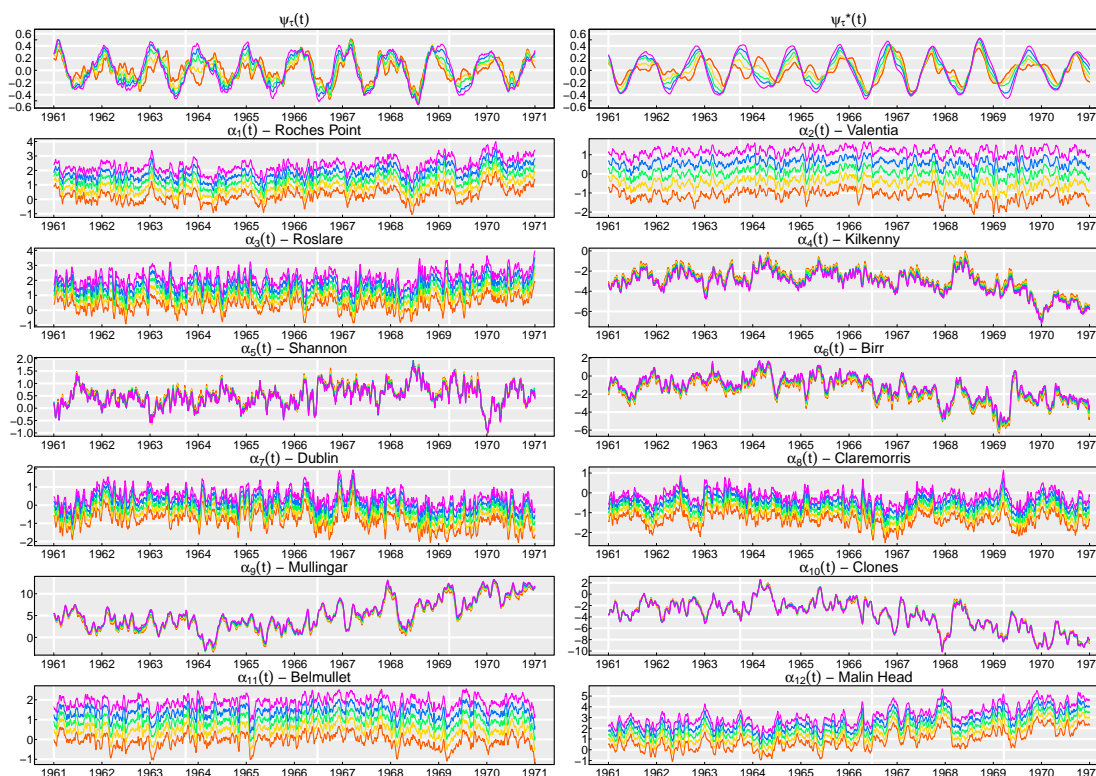


Figura 4.16: Stati latenti medi ottenuti via simulazione per i primi 10 anni di osservazione e per i quantili di ordine: $\tau = 0.1$ (arancione), $\tau = 0.25$ (giallo), $\tau = 0.5$ (verde), $\tau = 0.75$ (azzurro), $\tau = 0.9$ (viola).

Le caratteristiche osservabili sugli stati latenti si riflettono anche sui quantili, che, come possiamo osservare in figura 4.17, risultano avere traiettorie più regolari rispetto a quelle stimate su un modello a disturbi indipendenti.

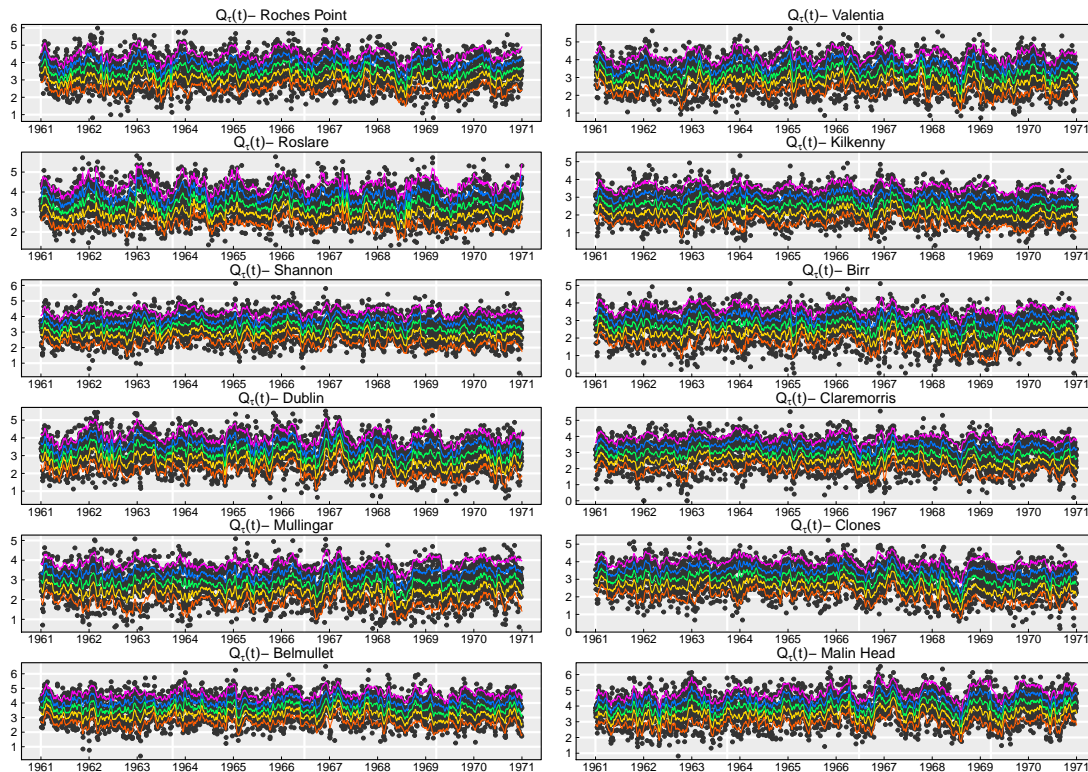


Figura 4.17: Velocità dei venti originale per i primi dieci anni di osservazione contro i quantili stimati di ordine: $\tau = 0.1$ (arancione), $\tau = 0.25$ (giallo), $\tau = 0.5$ (verde), $\tau = 0.75$ (azzurro), $\tau = 0.9$ (viola).

L'introduzione di una funzione di correlazione spaziale per i disturbi fa in modo che i quantili stimati siano meno legati all'andamento specifico della singola serie storica e siano invece guidati dalle variazioni comuni del processo in diverse regioni dello spazio. Questo comportamento è chiaramente influenzato alla funzione di correlazione scelta e dai parametri da cui questa dipende.

Come possiamo osservare dalla tabella 4.6 e dalla figura 4.18, il livelli nominale ed effettivo dei quantili stimati per il modello 4.5.1 risultano mediamente più vicini ed inoltre risulta scomparso l'effetto di *shrinkage* dei quantili estremi verso la mediana, che si osservava per il modello 4.4.1

Anche con questo secondo modello non vengono violati i vincoli di non incrocio tra i quantili, che, in virtù della maggiore regolarità, appaiono ben separati.

La possibilità di specificare un modello con maggiore struttura permette (in questo caso) di regolarizzare le curve stimate, e dà la possibilità di migliorare l'a-

Tabella 4.6: Numero medio di volte che un'osservazione si trova al di sopra del quantile stimato di ordine τ , per i quantili stimati usando il modello 4.5.1.

Stazioni	Probabilità stimate				
	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.9$
Valentia (Val)	0.0981	0.2861	0.5205	0.7601	0.9369
Belmullet (Bel)	0.0916	0.2870	0.5284	0.7632	0.9331
Claremorris (Cla)	0.0945	0.2857	0.5187	0.7659	0.9384
Shannon(Sha)	0.0823	0.2845	0.5389	0.7779	0.9382
Roche's Point (Rpt)	0.0896	0.2939	0.5321	0.7633	0.9309
Birr (Birr)	0.0923	0.2784	0.5141	0.7738	0.9480
Mullingar (Mul)	0.0774	0.2749	0.5216	0.7858	0.9525
Malin Head (Mal)	0.0957	0.2948	0.5265	0.7586	0.9303
Kilkenny (Kil)	0.0685	0.2639	0.5326	0.7869	0.9419
Clones (Clo)	0.0835	0.2775	0.5287	0.7784	0.9425
Dublin (Dub)	0.0770	0.2797	0.5368	0.7733	0.9408
Roslare (Ros)	0.0640	0.2581	0.5187	0.7686	0.9355

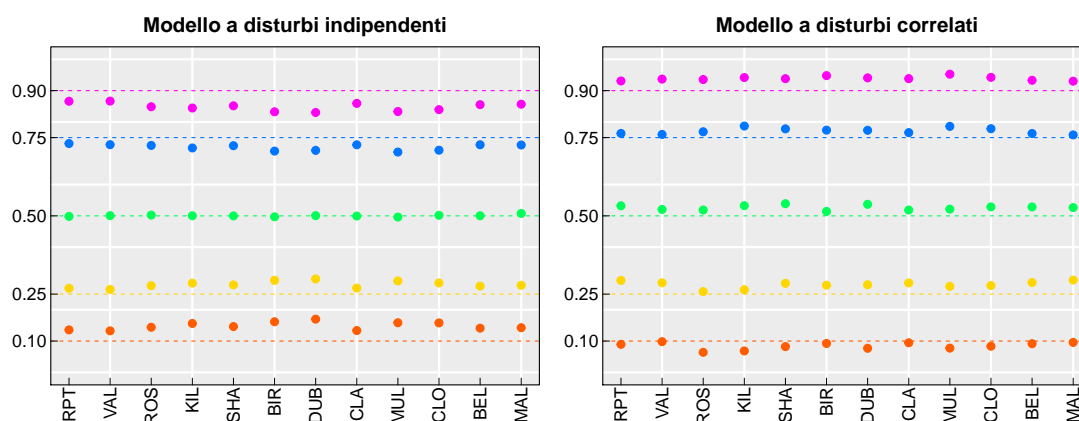


Figura 4.18: Numero medio di volte che un'osservazione si trova al di sopra del quantile stimato di ordine τ , confrontata col livello di probabilità nominale assunto nel modello. Le probabilità stimate riportate fanno riferimento al modello 4.4.1 (a sinistra) e al modello 4.5.1 (a destra).

dattamento del modello introducendo conoscenza a priori o derivante da evidenze empiriche in modo più flessibile.

Andando poi a studiare gli scarti interquantilici in figura 4.19, osserviamo che, anche con questo secondo modello, vengono sostanzialmente confermate le considerazioni relative all'eteroschedasticità periodica del fenomeno, che anzi appaiono anche più evidenti.

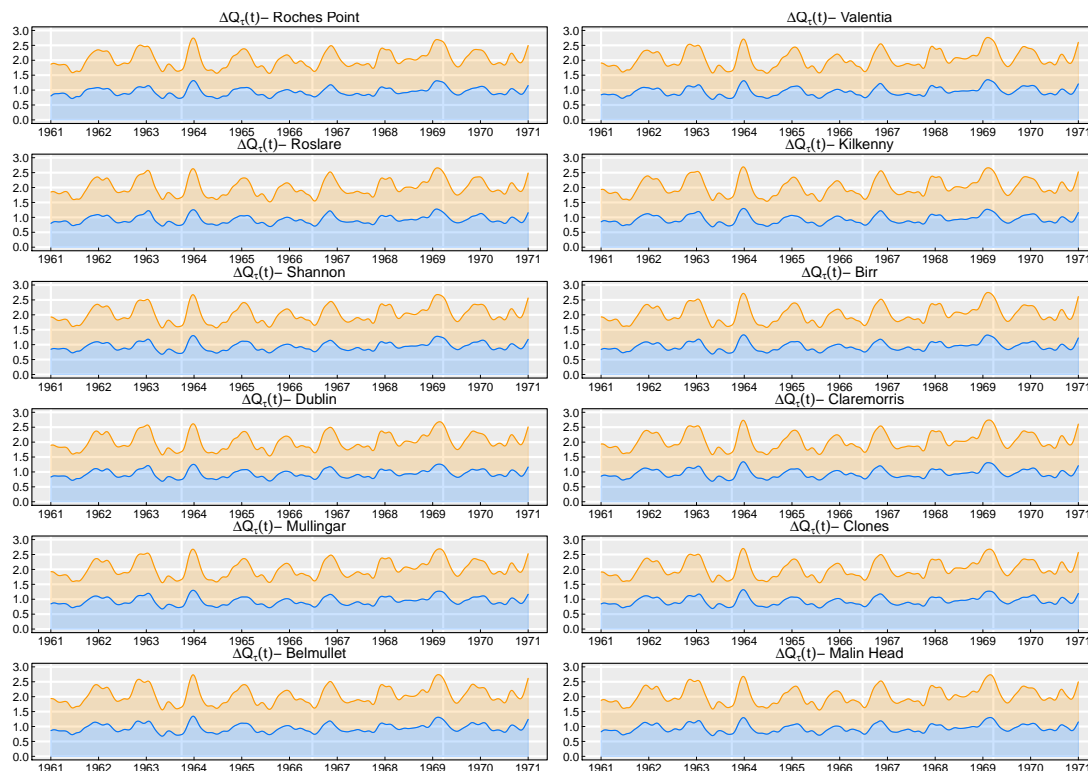


Figura 4.19: *Interquantile range* stimato per i primi 10 anni di osservazione. La linea arancione raffigura $\Delta^{(1)}Q(t)$, mentre la linea azzurra $\Delta^{(2)}Q(t)$.

4.6 Previsione spazio-temporale

Ipotizziamo ora nota la distribuzione a posteriori dei processi $\boldsymbol{\psi}_t$ e $\boldsymbol{\alpha}_t$, o perché stimata precedentemente, o perché derivante da estrapolazione temporale, e definiamo un metodo per produrre previsioni in un nuovo sito \mathbf{s} per un tempo fissato t . Per ottenere una stima del processo spaziale a posteriori in un luogo non ancora osservato è sufficiente calcolarne il valore atteso a posteriori, pari a

$$\begin{aligned} \mathbb{E}\left[y(\mathbf{s}, t) \mid \mathbf{y}^{oss}\right] &= \iint_{\boldsymbol{\Theta} \times \boldsymbol{\Psi} \times \mathbf{A}} \mathbb{E}\left[y(\mathbf{s}, t) \mid \mathbf{y}^{oss}\right] \pi(\boldsymbol{\theta}, \boldsymbol{\psi}_t, \boldsymbol{\alpha}_t \mid \mathbf{y}^{oss}) d\boldsymbol{\theta} d\boldsymbol{\psi}_t d\boldsymbol{\alpha}_t \\ &= \iint_{\boldsymbol{\Theta} \times \boldsymbol{\Psi} \times \mathbf{A}} \left(\mathbf{x}^\top \boldsymbol{\psi}_t + \mathbf{k}(\mathbf{s})^\top \boldsymbol{\alpha}_t\right) \pi(\boldsymbol{\theta}, \boldsymbol{\psi}_t, \boldsymbol{\alpha}_t \mid \mathbf{y}^{oss}) d\boldsymbol{\theta} d\boldsymbol{\psi}_t d\boldsymbol{\alpha}_t \\ &= \mathbf{x}^\top \mathbb{E}\left[\boldsymbol{\psi}_t \mid \mathbf{y}^{oss}\right] + \mathbf{k}(\mathbf{s})^\top \mathbb{E}\left[\boldsymbol{\alpha}_t \mid \mathbf{y}^{oss}\right]. \end{aligned}$$

Disponendo dei valori ottenuti tramite simulazione possiamo approssimare tale valore atteso via Monte Carlo come

$$\mathbb{E}\left[y(\mathbf{s}, t) \mid \mathbf{y}^{oss}\right] \approx \mathbf{x}^\top \left[\frac{1}{R} \sum_{r=1}^R \boldsymbol{\psi}_t^{(r)}\right] + \mathbf{k}(\mathbf{s})^\top \left[\frac{1}{R} \sum_{r=1}^R \boldsymbol{\alpha}_t^{(r)}\right]. \quad (4.6.1)$$

Analogamente, fissato un ordine τ , si possono calcolare i quantili puntuali del processo a posteriori nello spazio secondo la seguente relazione integrale

$$\begin{aligned}\mathbb{E}\left[Q^\tau(\mathbf{s}, t) | \mathbf{y}^{oss}\right] &= \iint_{\Theta \times \Psi \times \mathbf{A}} \mathbb{E}\left[Q^\tau(\mathbf{s}, t) | \mathbf{y}^{oss}\right] \pi(\boldsymbol{\theta}, \boldsymbol{\psi}_t, \boldsymbol{\alpha}_t | \mathbf{y}^{oss}) d\boldsymbol{\theta} d\boldsymbol{\psi}_t d\boldsymbol{\alpha}_t \\ &= \iint_{\Theta \times \Psi \times \mathbf{A}} \left(\mathbf{x}^\top \boldsymbol{\psi}_t^\tau + \mathbf{k}(\mathbf{s})^\top \boldsymbol{\alpha}_t^\tau\right) \pi(\boldsymbol{\theta}, \boldsymbol{\psi}_t, \boldsymbol{\alpha}_t | \mathbf{y}^{oss}) d\boldsymbol{\theta} d\boldsymbol{\psi}_t d\boldsymbol{\alpha}_t \\ &= \mathbf{x}^\top \mathbb{E}\left[\boldsymbol{\psi}_t^\tau | \mathbf{y}^{oss}\right] + \mathbf{k}(\mathbf{s})^\top \mathbb{E}\left[\boldsymbol{\alpha}_t^\tau | \mathbf{y}^{oss}\right],\end{aligned}$$

la cui approssimazione Monte Carlo è data da

$$\mathbb{E}\left[Q^\tau(\mathbf{s}, t) | \mathbf{y}^{oss}\right] \approx \mathbf{x}^\top \left[\frac{1}{R} \sum_{r=1}^R \boldsymbol{\psi}_t^{\tau(r)}\right] + \mathbf{k}(\mathbf{s})^\top \left[\frac{1}{R} \sum_{r=1}^R \boldsymbol{\alpha}_t^{\tau(r)}\right], \quad (4.6.2)$$

dove $Q^\tau(\mathbf{s}, t)$ rappresenta il quantile a posteriori di ordine τ per $y(\mathbf{s}, t)$ e dove è stata esplicitata la dipendenza da τ ad apice degli stati latenti $\boldsymbol{\psi}_t^\tau$ e $\boldsymbol{\alpha}_t^\tau$ per non creare ambiguità tra la previsione in media e quella quantilica.

Definendo una griglia sufficientemente fitta di valori nello spazio e sfruttando le equazioni 4.6.1 e 4.6.2 si possono costruire mappe e superfici di previsione dinamiche per fare inferenza su regioni dello spazio non ancora esplorate.

A titolo di esempio in figura 4.20 sono state rappresentate le mappe di calore che descrivono i quantili di ordine $\tau = 0.1, 0.5, 0.9$ per la velocità del vento in Irlanda in data 1 gennaio 1961. Ognuna di queste è stata ottenuta considerando una griglia 50×50 di coordinate geografiche equispaziate, su cui è stata calcolata la superficie $Q^\tau(\mathbf{s}, t)$ come descritto nell'equazione 4.6.2.

Come possiamo notare la mappa per il quantile di ordine 0.1 è dominata da colori freddi, poiché è associata alla coda sinistra della distribuzione di probabilità e dunque a valori più bassi. Viceversa il quantile di ordine 0.9 è dominato da colori caldi, perché descrive la coda destra della legge di probabilità. La superficie mediana, che si colloca circa a metà via tra gli altri due quantili presenta invece colori più tenui.

Osserviamo che i diversi quantili non rappresentano una mera traslazione l'uno dell'altro, ma che anche la forma della superficie risulta adattarsi in modo specifico ai diversi quantili per descrivere in modo flessibile la distribuzione di probabilità in funzione dello spazio.

In generale notiamo che nelle regioni di costa meridionali, occidentali e settentrionali tutti i tre quantili considerati tendono a prevedere valori di vento più alti, mentre nell'entroterra e nelle regioni costiere orientali questi tendono ad attenuarsi. Considerando l'esposizione dell'isola al mare aperto e la copertura fornita dal Regno Unito a oriente, questi risultati appaiono sensati.

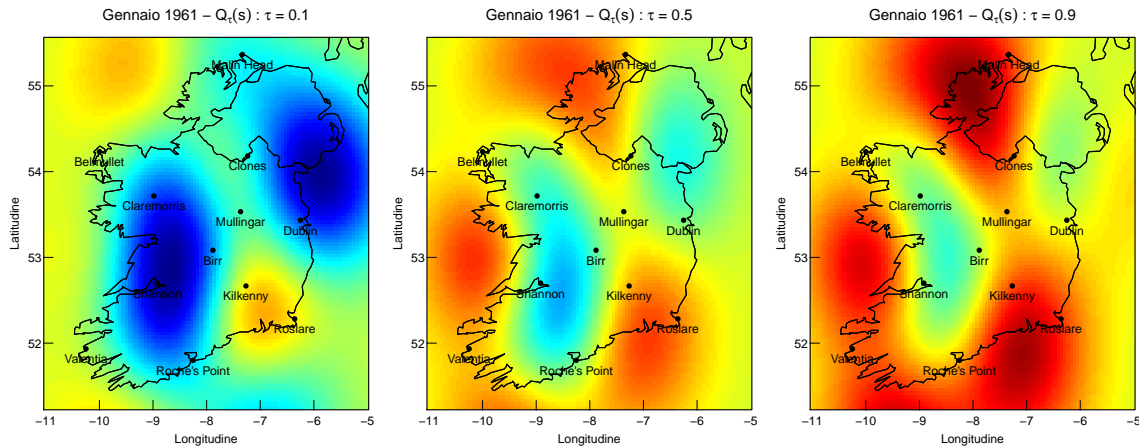


Figura 4.20: Superficie quantilica stimata in data 1 Gennaio 1961 per $\tau = 0.1$ (a sinistra), $\tau = 0.5$ (al centro) e $\tau = 0.9$ (a destra). Il gradiente di colore descrive l'intensità del vento: colori più freddi indicano regioni a bassa intensità di vento, mentre colori più caldi denotano vento più intenso.

Un secondo aspetto interessante è l'evoluzione del processo spaziale nel tempo. Per mostrare come la distribuzione del processo si modifichi dinamicamente sono state costruite le superfici di previsione spaziale per la mediana per ogni primo del mese dell'anno 1961. Le previsioni sono state ottenute utilizzando gli stati simulati del modello quantilico a disturbi correlati descritto dalla equazioni 4.5.1. Nelle figure 4.21 e 5.1 possiamo osservare il risultato.

Dai grafici 4.21 e 5.1 possiamo notare l'effetto dell'espansione *kernel* a coefficienti dinamici sulle superfici predette. Questa permette di ottenere una modellazione flessibile di strutture di dipendenza spazio-temporali non separabili. Osserviamo infatti come la forma del processo spaziale predetto cambi dinamicamente nel tempo adeguandosi ai campi di stagione.

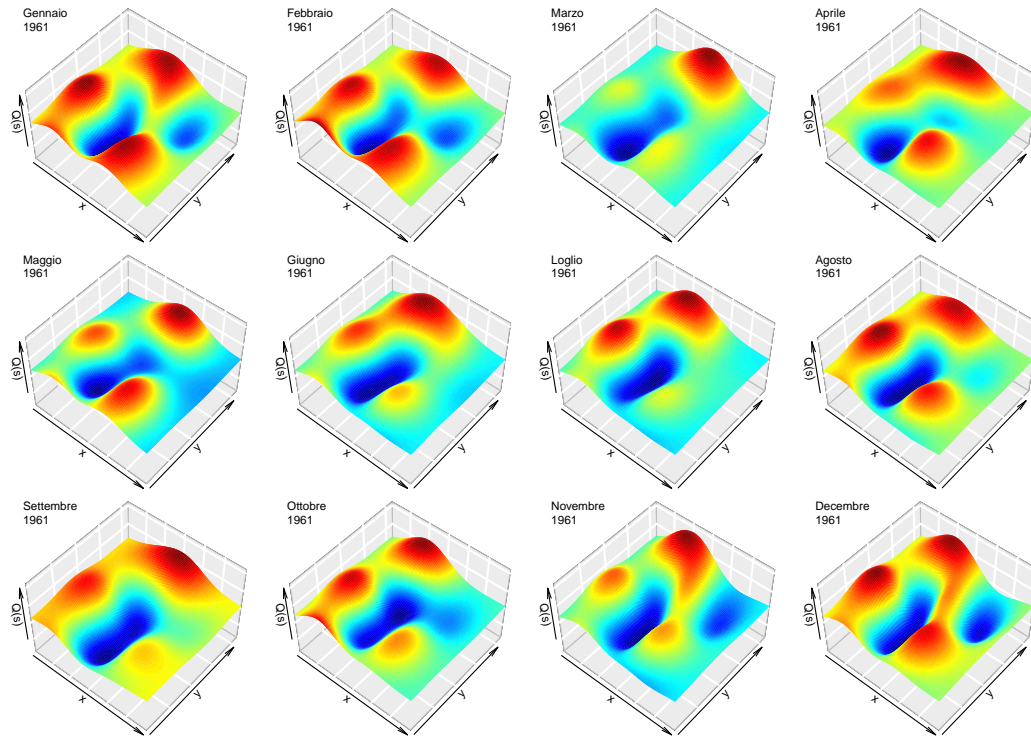


Figura 4.21: *Perspective plot* della superficie mediana per il processo spaziale nel tempo.

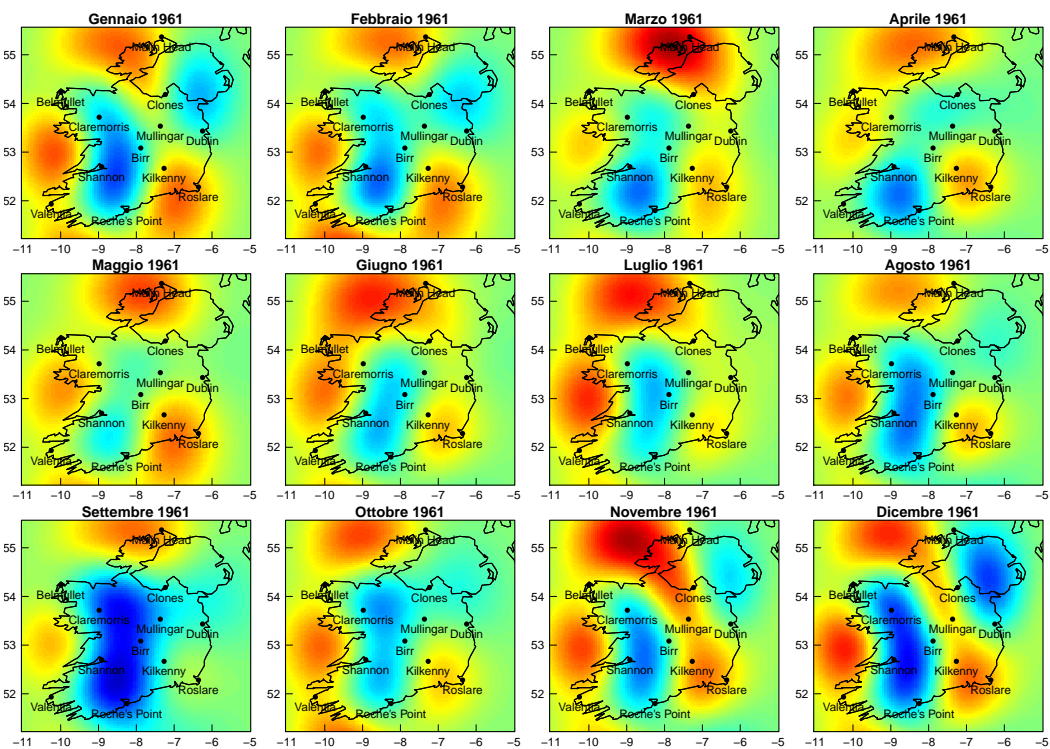


Figura 4.22: *Heatmap* della superficie mediana per il processo spaziale nel tempo.

Capitolo 5

Estensioni e quantili multipli

Nel capitolo 2 è stato proposto un modello *State Space* lineare per analizzare dati collezionati nello spazio e nel tempo, sotto ipotesi di Gaussianità dei disturbi. Come abbiamo visto, questa classe di modelli presenta dei limiti e non risulta spesso adeguata per trattare aspetti distributivi come, ad esempio, asimmetria, code pesanti e eteroschedasticità.

Per superare questi problemi, si possono andare a studiare direttamente i quantili della distribuzione spazio-temporale del processo, attraverso modelli di *quantile regression* dinamica.

Stimando contemporaneamente un gran numero di quantili è infatti possibile caratterizzare la distribuzione del processo punto per punto, secondo un approccio non parametrico. Per fare ciò, in prima approssimazione, si possono adottare direttamente i metodi e gli algoritmi presentati nelle sezioni 3.3 e 3.4 del capitolo 3, stimando in parallelo k modelli, in corrispondenza di k diversi valori di τ . Asintoticamente i quantili così stimati convergono ai veri quantili della distribuzione dei dati (si vedano Koenker e Bassett 1978, Koenker 2005 e Davino, Furno e Vistocco 2014).

Sfortunatamente la stima di k modelli separati non garantisce che vengano rispettati i vincoli di non incrocio tra i quantili. Nel caso in cui alcuni quantili si incrociassero, le stime non sarebbero utilizzabili a fini inferenziali e di conseguenza risulterebbero inutili.

Un modo per ottenere quantili stimati che non si incrocino per costruzione è quello di specificare un modello congiunto, che tenga conto esplicitamente dei vincoli necessari.

Nell'ambito dei modelli spazio-temporali, l'unica proposta esistente di *quantile regression* Bayesiana per la stima di quantili multipli senza incroci è stata fatta

da Reich, Fuentes e Dunson 2011. L'approccio adottato dagli autori prevede di trattare i quantili della distribuzione spaziale come una funzione non decrescente del parametro τ , andando poi a produrre delle previsioni dei quantili nello spazio e nel tempo utilizzando metodi non parametrici e semiparametrici di liscio.

In questa tesi si vuole proporre un approccio differente alla *quantile regression* spazio-temporale, attraverso cui tenere conto, non solo dei vincoli di *non-crossing*, ma anche dell'ordinamento unidirezionale del tempo, attraverso la formulazione di modelli *State Space* e di algoritmi di filtraggio, appositamente studiati per produrre stime dinamiche nel tempo.

In questo capitolo verrà quindi presentata una formulazione aumentata dei modelli precedentemente discussi e di seguito verranno proposti due algoritmi di stima vincolata.

5.1 Multiple quantile regression

Generalizzando l'approccio di stima frequentista per un singolo quantile, definito in 3.1.1, Zou e Yuan 2008 hanno proposto di stimare congiuntamente k quantili di ordine $0 < \tau_1 < \dots < \tau_k < 1$, con $k \geq 2$, attraverso la soluzione del seguente problema di ottimizzazione vincolata

$$\begin{aligned} \boldsymbol{\mu} &= \arg \min_{\boldsymbol{\mu}} \sum_{j=1}^k \sum_{t=1}^T \rho_{\tau_j} \left(y_t - \mu_{\tau_j} \right), \\ \text{s.t. } &\left\{ \mu_{\tau_j} < \mu_{\tau_{j+1}}, j = 1, \dots, k-1 \right\}. \end{aligned} \quad (5.1.1)$$

dove $\boldsymbol{\mu} = (\mu_{\tau_1}, \dots, \mu_{\tau_k})$ e $\mu_{\tau_j} = Q^{\tau_j}(\mathbf{x})$ è il quantile di ordine τ_j della distribuzione condizionata $Y|\mathbf{x}$.

Volendo calare il problema di stima di quantili multipli in un ottica inferenziale Bayesiana, è necessario specificare una verosimiglianza associata al modello di ottimizzazione vincolata 5.1.1. Osservando che

$$\arg \min_{\boldsymbol{\mu}} \sum_{j=1}^k \sum_{t=1}^T \rho_{\tau_j} \left(y_t - \mu_{\tau_j} \right) = \arg \max_{\boldsymbol{\mu}} \exp \left\{ - \sum_{j=1}^k \sum_{t=1}^T \rho_{\tau_j} \left(y_t - \mu_{\tau_j} \right) \right\},$$

il problema di ottimizzazione 5.1.1 può essere riformulato in termini di massima verosimiglianza, assumendo per i dati un modello distributivo con funzione di densità proporzionale a

$$p(y_t | \boldsymbol{\mu}) \propto \exp \left\{ - \sum_{j=1}^k \rho_{\tau_j} \left(y_t - \mu_{\tau_j} \right) \right\}, \quad (5.1.2)$$

a cui corrisponde una funzione di pseudo-verosimiglianza composta

$$\mathcal{L}_p(y | \boldsymbol{\mu}) \propto \exp \left\{ - \sum_{t=1}^T \sum_{j=1}^k \rho_{\tau_j} \left(y_t - \mu_{\tau_j} \right) \right\}. \quad (5.1.3)$$

Sriram, Ramamoorthi e Ghosh 2016 hanno mostrato che la funzione 5.1.3 non definisce una verosimiglianza propria, in quanto non tiene conto della costante di normalizzazione della densità 5.1.2, pari a

$$C(\boldsymbol{\mu}) = \left(\int_{-\infty}^{\infty} \exp \left\{ - \sum_{j=1}^k \rho_{\tau_j} \left(y - \mu_{\tau_j} \right) \right\} dy \right)^{-1}, \quad (5.1.4)$$

che dipende dai quantili $\boldsymbol{\mu}$.

Tuttavia, sempre Sriram, Ramamoorthi e Ghosh 2016 hanno dimostrato che, in un'ottica di inferenza Bayesiana, le stime a posteriori ottenute adottando per i dati la funzione di pseudo-verosimiglianza 5.1.3 sono asintoticamente consistenti. Di contro, l'utilizzo di una funzione di verosimiglianza propria, $\mathcal{L}(y | \boldsymbol{\mu}) = C(\boldsymbol{\mu}) \cdot \mathcal{L}_p(y | \boldsymbol{\mu})$ in luogo di 5.1.3 porta a conclusioni inferenziali inconsistenti e asintoticamente compresse verso la mediana, qualsiasi sia il valore di τ fissato inizialmente.

Quindi d'ora in avanti si considererà la stima di quantili congiunti adottando 5.1.3 come funzione di pseudo-verosimiglianza, alla quale corrisponde un modello distributivo \mathcal{AL} per i dati: $Y_t - \mu_{\tau_j} \stackrel{i.i.d.}{\sim} \mathcal{AL}(\tau_j, 0, \sigma_{\tau_j})$, $j = 1, \dots, k$.

5.2 Modello quantilico congiunto

Consideriamo ora la specificazione di k modelli spazio-temporali per la stima dei quantili di ordine $0 < \tau_1 < \dots < \tau_k < 1$, con $k \geq 2$ ed esplicitiamo i vincoli lineari di *non-crossing* sugli stati

$$\begin{aligned} \mathbf{y}_t^{(\tau)} &= \mathbf{c}_t + \mathbf{Z}_t \cdot \boldsymbol{\alpha}_t^{(\tau)} + \boldsymbol{\varepsilon}_t^{(\tau)}, & \boldsymbol{\varepsilon}_t^{(\tau)} &\sim \mathcal{AL}_n \left(\boldsymbol{\xi}^{(\tau)}, \mathbf{0}, \mathbf{H}_t^{(\tau)} \right) \\ \boldsymbol{\alpha}_{t+1}^{(\tau)} &= \mathbf{d}_t + \mathbf{T}_t \cdot \boldsymbol{\alpha}_t^{(\tau)} + \mathbf{R}_t \cdot \boldsymbol{\eta}_t^{(\tau)}, & \boldsymbol{\eta}_t^{(\tau)} &\sim \mathcal{N} \left(\mathbf{0}, \mathbf{Q}_t^{(\tau)} \right) \\ & & \boldsymbol{\alpha}_0^{(\tau)} &\sim \mathcal{N} \left(\mathbf{a}_0, \mathbf{P}_0 \right) \end{aligned} \quad (5.2.1)$$

s.t. $\left\{ \mathbf{Z}_t \cdot \boldsymbol{\alpha}_t^{(\tau_\ell)} < \mathbf{Z}_t \cdot \boldsymbol{\alpha}_t^{(\tau_{\ell+1})} : \ell = 1, \dots, k-1 \right\}, \quad t = 1, \dots, T,$

dove $\boldsymbol{\xi}^{(\tau)} = (\delta^{(\tau)})^{-1} \lambda^{(\tau)} \mathbf{1}_n$, $\mathbf{H}_t^{(\tau)} = \left(\sigma_{\varepsilon,(\tau)}^2 \right) \mathbf{V}$ e $\mathbf{Q}_t^{(\tau)} = \left(\sigma_{\eta,(\tau)}^2 \right) \mathbf{W}^*$, per $\tau = \tau_1, \dots, \tau_k$ e $k \geq 2$.

Riscrivendo il modello secondo la rappresentazione stocastica 3.4.4 e ipotizzando per le matrici strutturali una forma coerente con le equazioni 2.2.4, otteniamo

$$\begin{aligned}
\mathbf{y}_t^{(\tau)} &= \mathbf{c}_t + \mathbf{Z}_t \cdot \boldsymbol{\alpha}_t^{(\tau)} + \boldsymbol{\lambda}^{(\tau)} \cdot \boldsymbol{\omega}_t^{(\tau)} + \mathbf{G}_t^{(\tau)} \cdot \boldsymbol{\nu}_t^{(\tau)}, & \boldsymbol{\nu}_t^{(\tau)} &\sim \mathcal{N}(\mathbf{0}, \mathbf{H}_t^{(\tau)}) \\
\boldsymbol{\alpha}_{t+1}^{(\tau)} &= \mathbf{d}_t + \mathbf{T}_t \cdot \boldsymbol{\alpha}_t^{(\tau)} + \mathbf{R}_t \cdot \boldsymbol{\eta}_t^{(\tau)}, & \boldsymbol{\eta}_t^{(\tau)} &\sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_t^{(\tau)}) \\
& & \boldsymbol{\alpha}_0^{(\tau)} &\sim \mathcal{N}(\mathbf{a}_0, \mathbf{P}_0)
\end{aligned} \tag{5.2.2}$$

$$s.t. \quad \left\{ \mathbf{Z}_t \cdot \boldsymbol{\alpha}_t^{(\tau_\ell)} < \mathbf{Z}_t \cdot \boldsymbol{\alpha}_t^{(\tau_{\ell+1})} : \ell = 1, \dots, k-1 \right\}, \quad t = 1, \dots, T,$$

dove assumiamo le seguenti distribuzioni a priori per i parametri e i fattori latenti della rappresentazione aumentata:

$$\begin{aligned}
\sigma_{\eta,(\tau)}^2 &\sim \mathcal{IG}(a_{\eta,(\tau)}^0, b_{\eta,(\tau)}^0), & \tau &= \tau_1, \dots, \tau_k, \\
\sigma_{\varepsilon,(\tau)}^2 &\sim \mathcal{IG}(a_{\varepsilon,(\tau)}^0, b_{\varepsilon,(\tau)}^0), & \tau &= \tau_1, \dots, \tau_k, \\
\omega_t^{(\tau)} &\sim \mathcal{EXP}(\sigma_{\varepsilon,(\tau)}^{-2}), & \tau &= \tau_1, \dots, \tau_k.
\end{aligned}$$

Accorpendo tutti i modelli quantilici e riscrivendo i vincoli in forma matriciale, possiamo riesprimere le relazioni descritte in 5.2.2 secondo una formulazione piú compatta, come mostrato nelle equazioni 5.2.3.

$$\begin{aligned}
\tilde{\mathbf{y}}_t &= \tilde{\mathbf{c}}_t + \tilde{\mathbf{Z}}_t \cdot \tilde{\boldsymbol{\alpha}}_t + \tilde{\boldsymbol{\lambda}} \cdot \tilde{\boldsymbol{\omega}}_t + \tilde{\mathbf{G}}_t \cdot \tilde{\boldsymbol{\nu}}_t, & \tilde{\boldsymbol{\nu}}_t &\sim \mathcal{N}(\mathbf{0}, \tilde{\mathbf{H}}_t) \\
\tilde{\boldsymbol{\alpha}}_{t+1} &= \tilde{\mathbf{d}}_t + \tilde{\mathbf{T}}_t \cdot \tilde{\boldsymbol{\alpha}}_t + \tilde{\mathbf{R}}_t \cdot \tilde{\boldsymbol{\eta}}_t, & \tilde{\boldsymbol{\eta}}_t &\sim \mathcal{N}(\mathbf{0}, \tilde{\mathbf{Q}}_t) \\
& & \tilde{\boldsymbol{\alpha}}_0 &\sim \mathcal{N}(\tilde{\mathbf{a}}_0, \tilde{\mathbf{P}}_0)
\end{aligned} \tag{5.2.3}$$

$$s.t. \quad \left\{ \mathbf{D}_t \cdot \tilde{\boldsymbol{\alpha}}_t < \mathbf{0} : t = 0, \dots, T \right\}$$

dove

- $\tilde{\mathbf{y}}_t = (\mathbf{y}_t^{\tau_1}, \dots, \mathbf{y}_t^{\tau_k})^\top$, $\tilde{\boldsymbol{\alpha}}_t = \left((\boldsymbol{\alpha}_t^{(\tau_1)})^\top, \dots, (\boldsymbol{\alpha}_t^{(\tau_k)})^\top \right)^\top$
- $\tilde{\boldsymbol{\nu}}_t = \left((\boldsymbol{\nu}_t^{(\tau_1)})^\top, \dots, (\boldsymbol{\nu}_t^{(\tau_k)})^\top \right)^\top$, $\tilde{\boldsymbol{\eta}}_t = \left((\boldsymbol{\eta}_t^{(\tau_1)})^\top, \dots, (\boldsymbol{\eta}_t^{(\tau_k)})^\top \right)^\top$
- $\tilde{\boldsymbol{\omega}}_t = \left((\omega_t^{(\tau_1)} \mathbf{1}_n)^\top, \dots, (\omega_t^{(\tau_k)} \mathbf{1}_n)^\top \right)^\top$, $\tilde{\boldsymbol{\lambda}} = \left((\boldsymbol{\lambda}^{(\tau_1)})^\top, \dots, (\boldsymbol{\lambda}^{(\tau_k)})^\top \right)^\top$
- $\tilde{\mathbf{c}}_t = (\mathbf{c}_t^{\tau_1}, \dots, \mathbf{c}_t^{\tau_k})^\top$, $\tilde{\mathbf{d}}_t = (\mathbf{d}_t^{\tau_1}, \dots, \mathbf{d}_t^{\tau_k})^\top$
- $\tilde{\mathbf{Z}}_t = \text{diag}(\mathbf{Z}_t, \dots, \mathbf{Z}_t)$, $\tilde{\mathbf{T}}_t = \text{diag}(\mathbf{T}_t, \dots, \mathbf{T}_t)$

- $\tilde{\mathbf{H}}_t = \text{diag}(\mathbf{H}_t^{(\tau_1)}, \dots, \mathbf{H}_t^{(\tau_k)})$, $\tilde{\mathbf{Q}}_t = \text{diag}(\mathbf{Q}_t^{(\tau_1)}, \dots, \mathbf{Q}_t^{(\tau_k)})$
- $\tilde{\mathbf{G}}_t = \text{diag}(\mathbf{G}_t^{(\tau_1)}, \dots, \mathbf{G}_t^{(\tau_k)})$, $\tilde{\mathbf{R}}_t = \text{diag}(\mathbf{R}_t, \dots, \mathbf{R}_t)$
- $\tilde{\mathbf{a}}_0 = (\mathbf{a}_0^\top, \dots, \mathbf{a}_0^\top)^\top$, $\tilde{\mathbf{P}}_0 = \text{diag}(\mathbf{P}_0, \dots, \mathbf{P}_0)$
- $\mathbf{D}_t = \begin{pmatrix} \mathbf{Z}_t & -\mathbf{Z}_t & & & \\ & \mathbf{Z}_t & -\mathbf{Z}_t & & \\ & & \ddots & \ddots & \\ & & & \mathbf{Z}_t & -\mathbf{Z}_t \end{pmatrix}$ dove \mathbf{D}_t è una matrice $(nk - n) \times (mk)$.

Secondo le ipotesi fatte sul modello e sui parametri la distribuzione a posteriori aumentata risulta proporzionale a

$$\begin{aligned}
& \pi \left(\tilde{\boldsymbol{\sigma}}_\eta, \tilde{\boldsymbol{\sigma}}_\varepsilon, \{\tilde{\boldsymbol{\omega}}_t\}_{t=1}^T, \{\tilde{\boldsymbol{\alpha}}_t\}_{t=0}^T \mid \{\tilde{\mathbf{y}}_t\}_{t=1}^T \right) \propto \\
& \propto \left(\prod_{j=1}^k \left(\sigma_{\eta,(\tau_j)}^2 \right)^{-\left(a_{\eta,(\tau_j)}^0 + \frac{m}{2}(T-1)+1 \right)} \left(\sigma_{\varepsilon,(\tau_j)}^2 \right)^{-\left(a_{\varepsilon,(\tau_j)}^0 + T + \frac{n}{2}T+1 \right)} \right) \\
& \times \exp \left\{ - \sum_{j=1}^k \left(\frac{b_{\eta,(\tau_j)}^0}{\sigma_{\eta,(\tau_j)}^2} + \frac{b_{\varepsilon,(\tau_j)}^0}{\sigma_{\varepsilon,(\tau_j)}^2} \right) - \sum_{t=1}^T \sum_{j=1}^k \left(\frac{\omega_t^{(\tau_j)}}{\sigma_{\varepsilon,(\tau_j)}^2} \right) \right\} \\
& \times \exp \left\{ - \frac{1}{2s} \sum_{j=1}^k \left\| \boldsymbol{\alpha}_0^{(\tau_j)} \right\|_2^2 \right\} \exp \left\{ - \sum_{t=1}^{T-1} \sum_{j=1}^k \left\| \frac{\mathbf{W}^{-\frac{1}{2}} \left(\boldsymbol{\alpha}_{t+1}^{(\tau_j)} - \boldsymbol{\alpha}_t^{(\tau_j)} \right)}{2 \sigma_{\eta,(\tau_j)}^2} \right\|_2^2 \right\} \\
& \times \left(\prod_{t=1}^T \prod_{j=1}^k \left(\omega_t^{(\tau_j)} \right)^{-\frac{n}{2}} \right) \exp \left\{ - \frac{1}{2} \sum_{t=1}^T \sum_{j=1}^k \left\| \frac{\mathbf{V}^{-\frac{1}{2}} \left(\mathbf{y}_t - \boldsymbol{\lambda}^{(\tau_j)} \omega_t^{(\tau_j)} - \mathbf{K} \boldsymbol{\alpha}_t^{(\tau_j)} \right)}{\delta_{(\tau_j)}^2 \sigma_{\varepsilon,(\tau_j)}^2 \omega_t^{(\tau_j)}} \right\|_2^2 \right\}.
\end{aligned} \tag{5.2.4}$$

Osserviamo che la distribuzione a posteriori indotta dal *kernel* 5.2.4 è data dalla composizione di k densità 3.4.5, ognuna riferita ad un diverso quantile di ordine τ_j . Tale distribuzione è definita per tutti gli elementi dello spazio degli stati tali per cui sono rispettati i vincoli di *non-crossing* tra quantili. Ciò significa che all'insieme di tutti gli eventi nello spazio degli stati tali per cui le traiettorie dei quantili si incrociano verrà assegnata probabilità nulla.

Più formalmente avremo che

$$\int_{\mathcal{A}} \pi \left(\tilde{\boldsymbol{\sigma}}_\eta, \tilde{\boldsymbol{\sigma}}_\varepsilon, \{\tilde{\boldsymbol{\omega}}_t\}_{t=1}^T, \{\tilde{\boldsymbol{\alpha}}_t\}_{t=0}^T \mid \{\tilde{\mathbf{y}}_t\}_{t=1}^T \right) d\tilde{\boldsymbol{\alpha}}_0 \dots d\tilde{\boldsymbol{\alpha}}_T = 0,$$

dove $\mathcal{A} = \bigcap_{t=0}^T \left\{ \tilde{\boldsymbol{\alpha}}_t \in \mathbb{R}^{m+k} : \mathbf{D}_t \cdot \tilde{\boldsymbol{\alpha}}_t \geq \mathbf{0} \right\}$.

Per trovare delle stime Bayesiane per i parametri e gli stati latenti basate sulla distribuzione 5.2.4 vorremmo poter adottare una semplice estensione dell'algoritmo di *Gibbs Sampling* descritto in 3.4.3. Per farlo è necessario apportare delle modifiche agli algoritmi di lisciamiento e di *Simulation Smoothing* per gli stati latenti, in modo tale da includere i vincoli di non incrocio nelle procedure legate al filtro di Kalman.

5.3 Constrained Kalman Filter

Condizionatamente alle matrici strutturali del modello, gli stati hanno distribuzione Gaussiana soggetta a vincoli lineari di disuguaglianza. Sotto queste assunzioni i normali algoritmi di filtraggio e lisciamiento non sono applicabili e vanno pertanto modificati in modo tale che il segnale latente venga estratto rispettando l'ordinamento imposto sui quantili.

In letteratura sono stati proposte diverse estensioni del filtro di Kalman per problemi di filtraggio vincolato, come il metodo delle proiezioni (Simon e Simon 2006), il metodo di regolarizzazione (Simon e Simon 2006) e il metodo di troncamento in distribuzione (Simon e Simon 2010).

Seguendo l'approccio proposto da Simon e Simon 2010 possiamo incorporare i vincoli di disuguaglianza all'interno delle equazioni del filtro di Kalman (ed equivalentemente nello *Smoother*) andando a modificare il valore atteso a posteriori degli stati coerentemente con la distribuzione normale troncata indotta dai vincoli lineari sugli stati.

Supponiamo che \mathbf{x} sia un vettore aleatorio Gaussiano di lunghezza n , con media \mathbf{m} e matrice di covarianza \mathbf{C} , la cui distribuzione è soggetta ai seguenti vincoli di disuguaglianza:

$$\left\{ \mathbf{a} < \mathbf{Z} \mathbf{x} < \mathbf{b} : \mathbf{a} < \mathbf{b} \right\} = \left\{ a_j < \tilde{\mathbf{z}}_j^\top \mathbf{x} < b_j : a_j < b_j, j = 1, \dots, m \right\}, \quad (5.3.1)$$

dove $\tilde{\mathbf{z}}_j^\top$ è la riga j -esima della matrice non stocastica \mathbf{Z} , di dimensioni $m \times n$.

Il problema è come troncatura opportunamente la legge $\mathcal{N}(\mathbf{m}, \mathbf{C})$ secondo i vincoli 5.3.1, per ottenere media $\tilde{\mathbf{m}}$ e varianza $\tilde{\mathbf{C}}$ della nuova distribuzione vincolata.

Per rendere il problema trattabile definiamo $\tilde{\mathbf{m}}_{(j)}$ come il vettore medio degli stati filtrati ottenuto dopo l'imposizione di j vincoli, e $\tilde{\mathbf{C}}_{(j)}$ come la matrice di covarianze corrispondente.

Per inizializzare l'algoritmo di troncamento poniamo

$$j \leftarrow 0, \quad \tilde{\mathbf{m}}_{(j)} \leftarrow \mathbf{m}, \quad \tilde{\mathbf{C}}_{(j)} \leftarrow \mathbf{C},$$

e definiamo la variabile trasformata

$$\mathbf{w}_{(j)} \leftarrow \mathbf{R} \mathbf{\Lambda}^{-1/2} \mathbf{\Gamma}^\top (\mathbf{x} - \tilde{\mathbf{m}}_{(j)}), \quad (5.3.2)$$

dove $\mathbf{\Lambda}$ e $\mathbf{\Gamma}$ sono matrici ottenibili tramite scomposizione canonica di Jordan della matrice $\tilde{\mathbf{C}}_{(j)}$, e dunque tali che $\mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^\top = \tilde{\mathbf{C}}_{(j)}$. $\mathbf{\Lambda}$ è una matrice diagonale, mentre $\mathbf{\Gamma}$ è ortogonale. \mathbf{R} è anch'essa una matrice ortogonale di dimensioni $n \times n$, ottenibile tramite il metodo di ortogonalizzazione di Gram-Schmidt e tale che

$$\mathbf{R}\mathbf{\Lambda}^{1/2}\mathbf{\Gamma}^\top\tilde{\mathbf{z}}_j = \left[\left(\tilde{\mathbf{z}}_j^\top \tilde{\mathbf{C}}_{(j)} \tilde{\mathbf{z}}_j \right)^{1/2} \ 0 \ \dots \ 0 \right]^\top. \quad (5.3.3)$$

Come mostrato da Simon 2006, combinando le equazioni 5.3.2 e 5.3.3, si dimostra che i vincoli 5.3.1 nello spazio trasformato sono equivalenti a

$$c_j < \mathbf{J}^\top \mathbf{w}_{(j)} < d_j, \quad (5.3.4)$$

dove

$$\mathbf{J} = (1 \ 0 \ \dots \ 0)^\top, \quad c_j = \frac{(a_j - \tilde{\mathbf{z}}_j^\top \tilde{\mathbf{m}}_{(j)})}{\sqrt{\tilde{\mathbf{z}}_j^\top \tilde{\mathbf{C}}_{(j)} \tilde{\mathbf{z}}_j}}, \quad d_j = \frac{(b_j - \tilde{\mathbf{z}}_j^\top \tilde{\mathbf{m}}_{(j)})}{\sqrt{\tilde{\mathbf{z}}_j^\top \tilde{\mathbf{C}}_{(j)} \tilde{\mathbf{z}}_j}}.$$

Si osservino due aspetti importanti: per costruzione la distribuzione libera di $\mathbf{w}_{(j)}$ è una normale multivariata a media nulla e matrice di covarianza identitaria, dunque il vettore aleatorio $\mathbf{w}_{(j)}$ è a componenti indipendenti. Inoltre il vincoli di disuguaglianza 5.3.4 impongono una restrizione solo sul primo elemento del vettore trasformato $\mathbf{w}_{(j)}$, che quindi avrà distribuzione vincolata normale standard univariata con massa di probabilità non nulla solo nell'intervallo aperto (c_j, d_j) .

Per le proprietà della distribuzione normale troncata, il primo elemento del vettore $\mathbf{w}_{(j)}$ ha media e varianza pari a

$$\begin{aligned} \mu &= \kappa \left(e^{-c_j^2/2} - e^{-d_j^2/2} \right), \\ \sigma^2 &= \kappa \left((c_j - 2\mu) e^{-c_j^2/2} - (d_j - 2\mu) e^{-d_j^2/2} \right) + \mu^2 + 1, \\ \kappa &= \left(\int_{-\infty}^{d_j} e^{-w^2/2} dw - \int_{-\infty}^{c_j} e^{-w^2/2} dw \right)^{-1}, \end{aligned}$$

dunque media e varianza del vettore trasformato vincolato $\mathbf{w}_{(j)}$ sono rispettivamente pari a $\boldsymbol{\mu} = (\mu \ 0 \ \dots \ 0)^\top$ e $\boldsymbol{\Sigma} = \text{diag}(\sigma^2 \ 1 \ \dots \ 1)$.

Operando la trasformazione inversa di 5.3.2 si ottengono i momenti filtrati degli stati soggetti ai vincoli 5.3.1:

$$\tilde{\mathbf{m}}^{(j)} \leftarrow \mathbf{\Gamma}\mathbf{\Lambda}^{1/2}\mathbf{R}^\top \mathbf{w}_{(j)} + \tilde{\mathbf{m}}_{(j)}, \quad \tilde{\mathbf{C}}_{(j)} \leftarrow \mathbf{\Gamma}\mathbf{\Lambda}^{1/2}\mathbf{R}^\top \boldsymbol{\Sigma} \mathbf{R}\mathbf{\Lambda}^{1/2}\mathbf{\Gamma}^\top.$$

Iterando questa procedura per ogni $j = 1, \dots, m$, all'ultima iterazione otteniamo i momenti della distribuzione congiunta del vettore \mathbf{x} sotto tutti i vincoli di disuguaglianza $\mathbf{a} < \mathbf{Z}\mathbf{x} < \mathbf{b}$:

$$\tilde{\mathbf{m}} = \tilde{\mathbf{m}}_{(m)}, \quad \tilde{\mathbf{C}} = \tilde{\mathbf{C}}_{(m)}, \quad (5.3.5)$$

che sono media e varianza della distribuzione $\mathcal{MTN}_n(\mathbf{m}, \mathbf{C}, \mathbf{a}, \mathbf{b}, \mathbf{Z})$, ovvero la distribuzione normale multivariata troncata per tutti i valori che non rispettano il vincolo indotto dai vettori \mathbf{a} e \mathbf{b} e dalla matrice di trasformazione \mathbf{Z} .

Imponendo i vincoli $\mathbf{D}_t \cdot \boldsymbol{\alpha}_t < \mathbf{0}$ e applicando quest'algoritmo sugli stati del processo (filtrato, o liscio) ad ogni tempo, si ottiene una stima dei primi due momenti della distribuzione vincolata.

Se si desidera ottenere una stima liscia per media e varianza del processo vincolato e sufficiente eseguire prima il filtro di Kalman, poi il *Kalman Smoother* ed infine applicare la correzione per troncamento alle stime ottenute dal *Kalman Smoother*.

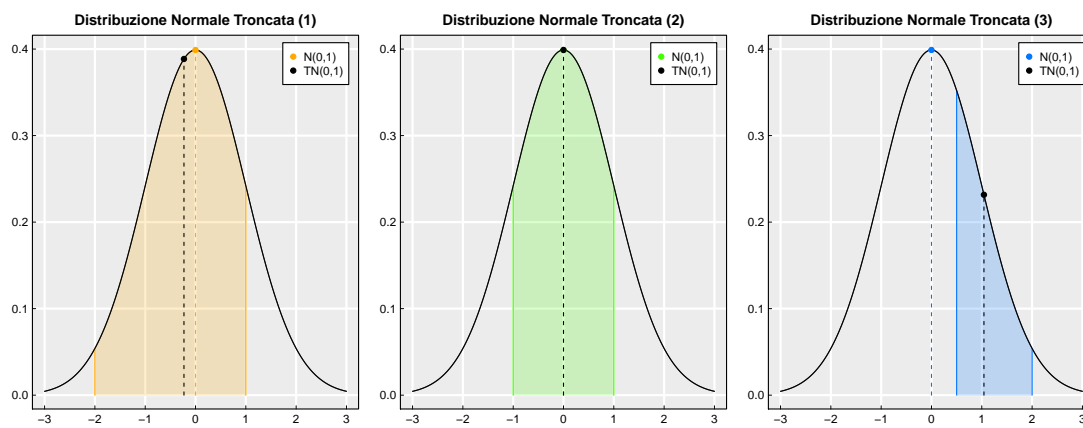


Figura 5.1: Confronto tra valori attesi di una distribuzione Gaussiana e di una Gaussiana troncata, sugli intervalli $(-2, 1)$ (a sinistra), $(-1, 1)$ (al centro), $(0.5, 2)$ (a destra).

Andando a vincolare gli stati e modificando i risultati del *Kalman Smoother*, non risulta più possibile simulare le traiettorie degli stati adottando l'algoritmo proposto da Durbin e Koopman 2002.

In letteratura sono stati proposti anche altri *Simulation Smoother*, che permettono di simulare in modo efficiente gli stati a partire dalla distribuzione congiunta liscia. I più diffusi sono quelli di Frühwirth-Schnatter 1994, Carter e Kohn 1994, Jong e Shephard 1995, Carter e Kohn 1996, Gerlach, Carter e Kohn 2000, Koopman e Durbin 2000.

Tuttavia questi metodi risultano poco maneggevoli nel momento in cui la distribuzione a posteriori degli stati risulta soggetta a restrizioni.

Una strategia alternativa che si può adottare in questi casi è di simulare gli stati del processo direttamente dalle distribuzioni fornite dal *Simulation Smoother* tempo per tempo.

Chiaramente se si vuole simulare una traiettoria del processo vincolato è necessario adottare come distribuzione *proposal* una Gaussiana opportunamente troncata. Questa distribuzione ci viene automaticamente fornita dal processo di correzione per troncamento.

Arrivati all'ultima iterazione del processo di imposizione dei vincoli si ottiene un vettore aleatorio trasformato $\mathbf{w}_{(m)} = (w_1^{(m)}, \dots, w_n^{(m)})$ tale che $w_1^{(m)} \sim \mathcal{TN}_{(c_m, d_m)}(0, 1)$ e $(w_2^{(m)}, \dots, w_n^{(m)}) \sim \mathcal{N}_{n-1}(\mathbf{0}, \mathbf{I}_{n-1})$, dove $\mathcal{TN}_{(c,d)}$ denota la distribuzione normale univariata troncata sull'intervallo (c, d) . Per simulare dei valori dalla distribuzione vincolata originaria è sufficiente generare da $\mathbf{w}_{(m)}$, per poi applicare una trasformazione lineare inversa e tornare sullo spazio originale.

5.4 Inferenza stocastica MCMC

Anche per il modello congiunto 5.2.3 è possibile costruire un algoritmo di *Gibbs Sampling* con *Data Augmentation* per simulare i valori di parametri, fattori aumentati e stati latenti dalla distribuzione a posteriori. La costruzione delle distribuzioni *full conditional* segue la stessa logica proposta nelle sezioni 3.3.2 e 3.4.3, dal momento che i singoli blocchi del modello sono indipendenti, a meno dei vincoli lineari sugli stati. Per maggiori dettagli si veda A.5.

L'algoritmo può essere strutturato come segue. Fissati i valori iniziali dei parametri $(\sigma_{\eta,(\tau)}^2)^{(0)}$ e $(\sigma_{\varepsilon,(\tau)}^2)^{(0)}$ per ogni $\tau = \tau_1, \dots, \tau_k$, si simulano i fattori aumentati $(\omega_t^{(\tau)})^{(0)}$ dalla distribuzione a priori per ogni $t = 1, \dots, T$ e per ogni $\tau = \tau_1, \dots, \tau_k$. Dati i parametri iniziali e i fattori aumentati, si esegue una prima simulazione degli stati latenti $(\boldsymbol{\alpha}_t^{(\tau)})^{(0)}$ per ogni $t = 0, \dots, T$ e per ogni $\tau = \tau_1, \dots, \tau_k$ attraverso l'algoritmo di *Simulation Smoothing* modificato proposto nella sezione 5.3. Poi le successive simulazioni dalla distribuzione a posteriori alla $(k+1)$ -esima iterazione sono ottenute secondo il seguente schema del *Gibbs Sampling*:

1. Per ogni $\tau = \tau_1, \dots, \tau_k$ si generi $(\sigma_{\eta,(\tau)}^2)^{(k+1)}$ da $\pi(\sigma_{\eta,(\tau)}^2 | \{\boldsymbol{\alpha}_t^{(\tau)}\}_{t=1}^T) \propto \mathcal{IG}(a_{\eta,(\tau)}^*, b_{\eta,(\tau)}^*)$, dove:

$$\begin{aligned} a_{\eta,(\tau)}^* &= a_{\eta,(\tau)}^0 + \frac{m}{2}(T-1), \\ b_{\eta,(\tau)}^* &= b_{\eta,(\tau)}^0 + \frac{1}{2} \sum_{t=1}^{T-1} \left\| \mathbf{W}^{-\frac{1}{2}} \left((\boldsymbol{\alpha}_{t+1}^{(\tau)})^{(k)} - (\boldsymbol{\alpha}_t^{(\tau)})^{(k)} \right) \right\|_2^2. \end{aligned}$$

2. Per ogni $\tau = \tau_1, \dots, \tau_k$ si generi $(\sigma_{\varepsilon,(\tau)}^2)^{(k+1)}$ dalla distribuzione $\pi \left(\sigma_{\varepsilon,(\tau)}^2 \mid \{\boldsymbol{\alpha}_t^{(\tau)}\}_{t=1}^T, \{\omega_t^{(\tau)}\}_{t=1}^T, \{\mathbf{y}_t^{(\tau)}\}_{t=1}^T \right) \propto \mathcal{IG} \left(a_{\varepsilon,(\tau)}^*, b_{\varepsilon,(\tau)}^* \right)$, dove:

$$\begin{aligned} a_{\varepsilon,(\tau)}^* &= a_{\varepsilon,(\tau)}^0 + T + \frac{n}{2}T, \\ b_{\varepsilon,(\tau)}^* &= b_{\varepsilon,(\tau)}^0 + \sum_{t=1}^T \left[\left(\omega_t^{(\tau)} \right)^{(k)} + \frac{1}{2} \left(\delta_{(\tau)}^2 \left(\omega_t^{(\tau)} \right)^{(k)} \right)^{-1} \left(\kappa_t^{(\tau)} \right)^{(k)} \right], \\ \left(\kappa_t^{(\tau)} \right)^{(k)} &= \left\| \mathbf{V}^{-\frac{1}{2}} \left(\mathbf{y}_t^{(\tau)} - \boldsymbol{\lambda}^{(\tau)} \left(\omega_t^{(\tau)} \right)^{(k)} - \mathbf{K} \left(\boldsymbol{\alpha}_t^{(\tau)} \right)^{(k)} \right) \right\|_2^2. \end{aligned}$$

3. Per ogni $\tau = \tau_1, \dots, \tau_k$ e per ogni $t = 1, \dots, T$, si generi $(\omega_t^{(\tau)})^{(k+1)}$ da $\pi \left(\omega_t^{(\tau)} \mid \sigma_{\varepsilon,(\tau)}^2, \boldsymbol{\alpha}_t^{(\tau)}, \mathbf{y}_t^{(\tau)} \right) \propto \mathcal{GIG} \left(\chi_{t,(\tau)}^*, \psi_{t,(\tau)}^*, \varphi_{t,(\tau)}^* \right)$, dove:

$$\begin{aligned} \chi_{t,(\tau)}^* &= \frac{2-n}{2}, \\ \psi_{t,(\tau)}^* &= \left(\delta_{(\tau)}^2 \left(\sigma_{\varepsilon,(\tau)}^2 \right)^{(k+1)} \right)^{-1} \left(\left(\boldsymbol{\lambda}^{(\tau)} \right)^\top \mathbf{V}^{-1} \left(\boldsymbol{\lambda}^{(\tau)} \right) + 2 \delta_{(\tau)}^2 \right), \\ \varphi_{t,(\tau)}^* &= \left(\delta_{(\tau)}^2 \left(\sigma_{\varepsilon,(\tau)}^2 \right)^{(k+1)} \right)^{-1} \left\| \mathbf{V}^{-\frac{1}{2}} \left(\mathbf{y}_t^{(\tau)} - \mathbf{K} \left(\boldsymbol{\alpha}_t^{(\tau)} \right)^{(k)} \right) \right\|_2^2. \end{aligned}$$

4. Per ogni $t = 0, \dots, T$, si generi $\tilde{\boldsymbol{\alpha}}_t^{(k+1)}$ da $\pi \left(\tilde{\boldsymbol{\alpha}}_t \mid \tilde{\boldsymbol{\sigma}}_\eta, \tilde{\boldsymbol{\sigma}}_\varepsilon, \{\tilde{\boldsymbol{\omega}}_t\}_{t=1}^T, \{\tilde{\mathbf{y}}_t\}_{t=1}^T \right)$. Condizionatamente al fattore aumentato $\{\tilde{\boldsymbol{\omega}}_t\}_{t=1}^T$, il modello è lineare e Gaussiano, dunque le dinamiche latenti $\{\tilde{\boldsymbol{\alpha}}_t\}_{t=1}^T$ possono essere simulate direttamente dalle distribuzioni lisce vincolate, come descritto in 5.3.

5.5 Inferenza Bayesiana approssimata

Per un modello *State Space* di grandi dimensioni, come 5.2.3, può non essere sempre possibile fare inferenza tramite algoritmi MCMC, che tipicamente sono molto onerosi ad punto di vista computazionale. Una valida alternativa è quella di adottare metodi di approssimazione variazionale della legge a posteriori.

5.5.1 Variational approximations

Consideriamo un generico modello Bayesiano $p(\mathbf{y}|\boldsymbol{\vartheta})$ per la variabile \mathbf{y} e indichiamo con $\boldsymbol{\vartheta} \in \Theta$ l'insieme di tutti i parametri e di tutti i fattori latenti presenti nel modello. Denotando con $q(\boldsymbol{\vartheta})$ una generica legge di probabilità su Θ e, prendendo il

logaritmo naturale della legge marginale $p(\mathbf{y})$, è facile mostrare che l'identità 5.5.1 è sempre verificata (Ormerod e Wand 2010):

$$\log p(\mathbf{y}) = \int_{\Theta} q(\boldsymbol{\theta}) \log \frac{p(\mathbf{y}, \boldsymbol{\theta})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta} - \int_{\Theta} q(\boldsymbol{\theta}) \log \frac{p(\boldsymbol{\theta}|\mathbf{y})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta}. \quad (5.5.1)$$

Il logaritmo della distribuzione marginale dei dati può quindi essere espresso come la somma di due quantità:

$$\log p(\mathbf{y}) = \mathcal{L}(q) + KL(q \| p), \quad (5.5.2)$$

$$\mathcal{L}(q) = \int_{\Theta} q(\boldsymbol{\theta}) \log \frac{p(\mathbf{y}, \boldsymbol{\theta})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta}, \quad (5.5.3)$$

$$KL(q \| p) = - \int_{\Theta} q(\boldsymbol{\theta}) \log \frac{p(\boldsymbol{\theta}|\mathbf{y})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta}, \quad (5.5.4)$$

dove $KL(q \| p)$ è la divergenza di Kullback-Leibler, la quale identifica una misura di distanza asimmetrica tra le distribuzioni q e p . Mentre $\mathcal{L}(q)$ rappresenta un funzionale di q tale che $\mathcal{L}(q) \leq \log p(\mathbf{y})$.

Ora l'obiettivo è di trovare una funzione $q(\cdot)$ trattabile che approssimi bene $p(\cdot|\mathbf{y})$. Per trovare q possiamo restringere la ricerca ad un conveniente spazio di funzioni \mathcal{Q} e determinare q come soluzione del seguente problema variazionale

$$q(\cdot) = \arg \min_{q \in \mathcal{Q}} \left\{ \mathcal{L}(q) + KL(q \| p) \right\} \quad (5.5.5)$$

Si può dimostrare (Ormerod e Wand 2010, Bishop 2006) che il problema di ottimizzazione funzionale 5.5.5 è anche esprimibile in forma semplificata come

$$q(\cdot) = \arg \max_{q \in \mathcal{Q}} \mathcal{L}(q) = \arg \min_{q \in \mathcal{Q}} KL(q \| p) \quad (5.5.6)$$

Chiaramente se $p(\cdot|\mathbf{y}) \in \mathcal{Q}$, allora il problema è banalmente risolto per $q(\cdot) = p(\cdot|\mathbf{y})$, ovvero la funzione tale da annullare la divergenza di Kullback-Leibler (che, per costruzione, è un funzionale sempre maggiore o uguale a zero).

Una delle ipotesi più adottate in letteratura è che q sia una funzione fattorizzabile nel prodotto

$$q(\boldsymbol{\vartheta}) = \prod_{i=1}^m q_i(\boldsymbol{\vartheta}_i) \quad (5.5.7)$$

dove $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_1, \dots, \boldsymbol{\vartheta}_m)$ è un'opportuna partizione del vettore dei parametri, a cui corrisponde una partizione dello spazio parametrico $\Theta = (\Theta_1, \dots, \Theta_m)$, tale che $\boldsymbol{\vartheta}_i \in \Theta_i$ per ogni $i = 1, \dots, m$.

Sotto l'ipotesi di fattorizzazione degli elementi di q , si ottiene

$$\begin{aligned}
\mathcal{L}(q) &= \int_{\Theta} \prod_{i=1}^m q_i(\boldsymbol{\vartheta}_i) \left(\log p(\mathbf{y}, \boldsymbol{\vartheta}) - \sum_{i=1}^m \log q_i(\boldsymbol{\vartheta}_i) \right) d\boldsymbol{\vartheta} \\
&= \int_{\Theta_i} q_i(\boldsymbol{\vartheta}_i) \left\{ \int_{\Theta_{-i}} \log p(\mathbf{y}, \boldsymbol{\vartheta}) \prod_{j \neq i} q_j(\boldsymbol{\vartheta}_j) d\boldsymbol{\vartheta}_j \right\} d\boldsymbol{\vartheta}_i \\
&\quad - \int_{\Theta_i} q_i(\boldsymbol{\vartheta}_i) \log q_i(\boldsymbol{\vartheta}_i) d\boldsymbol{\vartheta}_i \\
&\quad - \int_{\Theta_i} q_i(\boldsymbol{\vartheta}_i) \left\{ \int_{\Theta_{-i}} \left(\sum_{i=1}^m \log q_i(\boldsymbol{\vartheta}_i) \right) \prod_{j \neq i} q_j(\boldsymbol{\vartheta}_j) d\boldsymbol{\vartheta}_j \right\} d\boldsymbol{\vartheta}_i \\
&= \int_{\Theta_i} q_i(\boldsymbol{\vartheta}_i) \log \tilde{p}(\mathbf{y}, \boldsymbol{\vartheta}_i) d\boldsymbol{\vartheta}_i + \int_{\Theta_i} q_i(\boldsymbol{\vartheta}_i) \log q_i(\boldsymbol{\vartheta}_i) d\boldsymbol{\vartheta}_i \\
&\quad - \int_{\Theta_i} q_i(\boldsymbol{\vartheta}_i) \left\{ \int_{\Theta_{-i}} \left(\sum_{j \neq i} \log q_j(\boldsymbol{\vartheta}_j) \right) \prod_{j \neq i} q_j(\boldsymbol{\vartheta}_j) d\boldsymbol{\vartheta}_j \right\} d\boldsymbol{\vartheta}_i \\
&= \int_{\Theta_i} q_i(\boldsymbol{\vartheta}_i) \log \frac{\tilde{p}(\mathbf{y}, \boldsymbol{\vartheta}_i)}{q_i(\boldsymbol{\vartheta}_i)} d\boldsymbol{\vartheta}_i + \mathcal{C}(\boldsymbol{\vartheta}_{-i})
\end{aligned}$$

dove $\boldsymbol{\vartheta}_{-i} = \boldsymbol{\vartheta} \setminus \boldsymbol{\vartheta}_i$, $\Theta_{-i} = \Theta \setminus \Theta_i$ e

$$\log \tilde{p}(\mathbf{y}, \boldsymbol{\vartheta}_i) = \mathbb{E}_{-i} \left[\log p(\mathbf{y}, \boldsymbol{\vartheta}) \right] = \int_{\Theta_{-i}} \log p(\mathbf{y}, \boldsymbol{\vartheta}) \prod_{j \neq i} q_j(\boldsymbol{\vartheta}_j) d\boldsymbol{\vartheta}_j,$$

che è il valore atteso di $\log p(\mathbf{y}, \boldsymbol{\vartheta})$ calcolato rispetto a tutte le distribuzioni $q_j(\boldsymbol{\vartheta}_j)$ con $j \neq i$.

Come si può notare $\mathcal{C}(\boldsymbol{\vartheta}_{-i})$ è un addendo che non dipende da q_i , quindi, volendo massimizzare $\mathcal{L}(q)$ rispetto a q_i , $\mathcal{C}(\boldsymbol{\vartheta}_{-i})$ costituisce un elemento trascurabile. Si osservi inoltre che, a meno di $\mathcal{C}(\boldsymbol{\vartheta}_{-i})$, $\mathcal{L}(q)$ definisce la divergenza di Kullback-Leibler negativa tra q_i e $\tilde{p}(\mathbf{y}, \boldsymbol{\vartheta}_i)$, la quale è massimizzata quando $q_i^*(\boldsymbol{\vartheta}_i) = \tilde{p}(\mathbf{y}, \boldsymbol{\vartheta}_i)$.

Dunque, sotto l'ipotesi di fattorizzazione 5.5.7, le soluzioni del problema variazionale si ottengono in corrispondenza delle distribuzioni

$$q_i^*(\boldsymbol{\vartheta}_i) = \frac{\exp \left\{ \mathbb{E}_{-i} \left[\log p(\mathbf{y}, \boldsymbol{\vartheta}) \right] \right\}}{\int_{\Theta_i} \exp \left\{ \mathbb{E}_{-i} \left[\log p(\mathbf{y}, \boldsymbol{\vartheta}) \right] \right\} d\boldsymbol{\vartheta}_i}, \quad i = 1, \dots, m. \quad (5.5.8)$$

Con la stessa logica adottata per trovare $q_i^*(\boldsymbol{\vartheta}_i)$, si può scomporre $\log \tilde{p}(\mathbf{y}, \boldsymbol{\vartheta}_i)$

nella somma di due termini, uno che dipende da $\boldsymbol{\vartheta}_i$ e una costante:

$$\begin{aligned}
\mathbb{E}_{-i} \left[\log p(\mathbf{y}, \boldsymbol{\vartheta}) \right] &= \int_{\boldsymbol{\Theta}_{-i}} \log p(\mathbf{y}, \boldsymbol{\vartheta}) \prod_{j \neq i} q_j(\boldsymbol{\vartheta}_j) d\boldsymbol{\vartheta}_j \\
&= \int_{\boldsymbol{\Theta}_{-i}} \left\{ \log p(\boldsymbol{\vartheta}_i | \mathbf{y}, \boldsymbol{\vartheta}_{-i}) + \log p(\mathbf{y}, \boldsymbol{\vartheta}_{-i}) \right\} \prod_{j \neq i} q_j(\boldsymbol{\vartheta}_j) d\boldsymbol{\vartheta}_j \\
&= \int_{\boldsymbol{\Theta}_{-i}} \log p(\boldsymbol{\vartheta}_i | \mathbf{y}, \boldsymbol{\vartheta}_{-i}) \prod_{j \neq i} q_j(\boldsymbol{\vartheta}_j) d\boldsymbol{\vartheta}_j \\
&\quad + \int_{\boldsymbol{\Theta}_{-i}} \log p(\mathbf{y}, \boldsymbol{\vartheta}_{-i}) \prod_{j \neq i} q_j(\boldsymbol{\vartheta}_j) d\boldsymbol{\vartheta}_j \\
&= \int_{\boldsymbol{\Theta}_{-i}} \log p(\boldsymbol{\vartheta}_i | \mathbf{y}, \boldsymbol{\vartheta}_{-i}) \prod_{j \neq i}^m q_j(\boldsymbol{\vartheta}_j) d\boldsymbol{\vartheta}_j + \tilde{\mathcal{C}}(\boldsymbol{\vartheta}_{-i}).
\end{aligned}$$

Sostituendo questo risultato in $\mathcal{L}(q)$ otteniamo

$$\mathcal{L}(q) = \int_{\boldsymbol{\Theta}_i} q_i(\boldsymbol{\vartheta}_i) \log \frac{\tilde{p}(\boldsymbol{\vartheta}_i | \mathbf{y})}{q_i(\boldsymbol{\vartheta}_i)} d\boldsymbol{\vartheta}_i + \mathcal{C}(\boldsymbol{\vartheta}_{-i}) + \tilde{\mathcal{C}}(\boldsymbol{\vartheta}_{-i}), \quad (5.5.9)$$

dove $\log \tilde{p}(\boldsymbol{\vartheta}_i | \mathbf{y}) = \mathbb{E}_{-i} \left[\log p(\boldsymbol{\vartheta}_i | \mathbf{y}, \boldsymbol{\vartheta}_{-i}) \right]$.

Con la formulazione 5.5.9 il funzionale $\mathcal{L}(q)$ è massimizzato in corrispondenza della funzione $\bar{q}_i^*(\boldsymbol{\vartheta}_i) = \tilde{p}(\boldsymbol{\vartheta}_i | \mathbf{y})$.

Adottando $\{\bar{q}_i^*(\boldsymbol{\vartheta}_i)\}_{i=1}^m$ come distribuzioni variazionali per i parametri $\boldsymbol{\vartheta}$ si può costruire un algoritmo di stima iterando i passi di aggiornamento delle distribuzioni variazionali 5.5.10 fino a convergenza:

$$\bar{q}_i^*(\boldsymbol{\vartheta}_i) \leftarrow \frac{\exp \left\{ \mathbb{E}_{-i} \left[\log p(\boldsymbol{\vartheta}_i | \mathbf{y}, \boldsymbol{\vartheta}_{-i}) \right] \right\}}{\int_{\boldsymbol{\Theta}_i} \exp \left\{ \mathbb{E}_{-i} \left[\log p(\boldsymbol{\vartheta}_i | \mathbf{y}, \boldsymbol{\vartheta}_{-i}) \right] \right\} d\boldsymbol{\vartheta}_i}, \quad i = 1, \dots, m. \quad (5.5.10)$$

Se la funzione di densità a posteriori è log-concava, allora l'iterazione dei passi 5.5.10 garantisce il raggiungimento di un massimo locale della funzione obiettivo (Ormerod e Wand 2010).

Nella letteratura di inferenza stocastica, basata su algoritmi MCMC, $p(\boldsymbol{\vartheta}_i | \mathbf{y}, \boldsymbol{\vartheta}_{-i})$ rappresenta la distribuzione *full conditional* per $\boldsymbol{\vartheta}_i$. Questo risultato permette di recuperare tutte le proprietà fin'ora analizzate per i modelli di *quantile regression* Bayesiana, per andare a sfruttarle nella costruzione di un algoritmo di inferenza approssimata a posteriori.

5.5.2 Inferenza approssimata per modelli *State Space*

Riconsiderando il modello congiunto 5.2.3, ipotizziamo che le distribuzioni approssimanti a posteriori fattorizzino secondo lo schema 5.5.11:

$$\begin{aligned}
q\left(\tilde{\boldsymbol{\sigma}}_\eta, \tilde{\boldsymbol{\sigma}}_\varepsilon, \{\tilde{\boldsymbol{\omega}}_t\}_{t=1}^T, \{\tilde{\boldsymbol{\alpha}}_t\}_{t=0}^T\right) &= q_\eta\left(\tilde{\boldsymbol{\sigma}}_\eta\right) q_\varepsilon\left(\tilde{\boldsymbol{\sigma}}_\varepsilon\right) q_\omega\left(\{\tilde{\boldsymbol{\omega}}_t\}_{t=1}^T\right) q_\alpha\left(\{\tilde{\boldsymbol{\alpha}}_t\}_{t=0}^T\right) \\
q_\eta\left(\tilde{\boldsymbol{\sigma}}_\eta\right) &= \prod_{j=1}^k q_\eta^{(\tau_j)}\left(\sigma_{\eta,(\tau_j)}^2\right), & q_\varepsilon\left(\tilde{\boldsymbol{\sigma}}_\varepsilon\right) &= \prod_{j=1}^k q_\varepsilon^{(\tau_j)}\left(\sigma_{\varepsilon,(\tau_j)}^2\right) \\
q_\omega\left(\tilde{\boldsymbol{\omega}}_t\right) &= \prod_{t=1}^T \prod_{j=1}^k q_{\omega_t}^{(\tau_j)}\left(\omega_t^{(\tau_j)}\right), & q_\alpha\left(\tilde{\boldsymbol{\alpha}}\right) &= \prod_{t=1}^T q_{\alpha_t}\left(\tilde{\boldsymbol{\alpha}}_t\right)
\end{aligned} \tag{5.5.11}$$

Sotto queste assunzioni le distribuzioni variazionali che minimizzano la divergenza di Kullback-Leibler tra q e p sono quelle descritte in 5.5.10, ovvero le *full conditional* marginali:

$$\begin{aligned}
\log q_\eta^{(\tau)}\left(\sigma_{\eta,(\tau)}^2\right) &= \mathbb{E}_{\eta,\varepsilon,\omega,\alpha} \left[\log \pi \left(\sigma_{\eta,(\tau)}^2 \mid \tilde{\boldsymbol{\sigma}}_\eta^{(-\tau)}, \tilde{\boldsymbol{\sigma}}_\varepsilon, \{\tilde{\boldsymbol{\omega}}_t\}_{t=1}^T, \{\tilde{\boldsymbol{\alpha}}_t\}_{t=1}^T, \{\tilde{\boldsymbol{y}}_t\}_{t=1}^T \right) \right] + \mathcal{C} \\
\log q_\varepsilon^{(\tau)}\left(\sigma_{\varepsilon,(\tau)}^2\right) &= \mathbb{E}_{\eta,\varepsilon,\omega,\alpha} \left[\log \pi \left(\sigma_{\varepsilon,(\tau)}^2 \mid \tilde{\boldsymbol{\sigma}}_\eta, \tilde{\boldsymbol{\sigma}}_\varepsilon^{(-\tau)}, \{\tilde{\boldsymbol{\omega}}_t\}_{t=1}^T, \{\tilde{\boldsymbol{\alpha}}_t\}_{t=1}^T, \{\tilde{\boldsymbol{y}}_t\}_{t=1}^T \right) \right] + \mathcal{C} \\
\log q_{\omega_t}^{(\tau)}\left(\omega_t^{(\tau)}\right) &= \mathbb{E}_{\eta,\varepsilon,\omega,\alpha} \left[\log \pi \left(\omega_t^{(\tau)} \mid \tilde{\boldsymbol{\sigma}}_\eta, \tilde{\boldsymbol{\sigma}}_\varepsilon, \{\tilde{\boldsymbol{\omega}}_\ell^{(-\tau)}\}_{\ell \neq t}, \{\tilde{\boldsymbol{\alpha}}_t\}_{t=1}^T, \{\tilde{\boldsymbol{y}}_t\}_{t=1}^T \right) \right] + \mathcal{C} \\
\log q_{\alpha_t}\left(\tilde{\boldsymbol{\alpha}}_t\right) &= \mathbb{E}_{\eta,\varepsilon,\omega,\alpha} \left[\log \pi \left(\tilde{\boldsymbol{\alpha}}_t \mid \tilde{\boldsymbol{\sigma}}_\eta, \tilde{\boldsymbol{\sigma}}_\varepsilon, \{\tilde{\boldsymbol{\omega}}_t\}_{t=1}^T, \{\tilde{\boldsymbol{\alpha}}_\ell\}_{\ell \neq t}, \{\tilde{\boldsymbol{y}}_t\}_{t=1}^T \right) \right] + \mathcal{C}
\end{aligned}$$

per ogni $t = 1, \dots, T$ e per ogni $\tau = \tau_1, \dots, \tau_k$.

Recuperando le distribuzioni *full conditional* adottate per costruire l'algoritmo di *Gibbs Sampling* in 5.4 e marginalizzando analiticamente gli effetti che non dipendono dal parametro di interesse (si veda A.6), otteniamo le seguenti *Variational Bayesian Approximations*:

$$\begin{aligned}
\log q_\eta^{(\tau)}\left(\sigma_{\eta,(\tau)}^2\right) &= -\left(a_{\eta,(\tau)}^0 + \frac{m}{2}T - \frac{m}{2} + 1\right) \log\left(\sigma_{\eta,(\tau)}^2\right) - \frac{b_{\eta,(\tau)}^0}{\sigma_{\eta,(\tau)}^2} - \sum_{t=0}^{T-1} \frac{\Lambda_{\alpha,t}^{(\tau)}}{2\sigma_{\eta,(\tau)}^2} + \mathcal{C}, \\
\log q_\varepsilon^{(\tau)}\left(\sigma_{\varepsilon,(\tau)}^2\right) &= -\left(a_{\varepsilon,(\tau)}^0 + T + \frac{n}{2}T + 1\right) \log\left(\sigma_{\varepsilon,(\tau)}^2\right) - \frac{b_{\varepsilon,(\tau)}^0}{\sigma_{\varepsilon,(\tau)}^2} - \frac{1}{2\delta_{(\tau)}^2 \sigma_{\varepsilon,(\tau)}^2} \times \\
&\quad \times \sum_{t=1}^T \left\{ \Lambda_{\omega,1,t}^{(\tau)} \Lambda_{y,t}^{(\tau)} - 2(\boldsymbol{\lambda}^{(\tau)})^\top \mathbf{V}^{-1} \left(\mathbf{y}_t - \mathbf{K} \hat{\mathbf{a}}_t^{(\tau)} \right) + \Gamma^{(\tau)} \Lambda_{\omega,2,t}^{(\tau)} \right\} + \mathcal{C}, \\
\log q_{\omega_t}^{(\tau)}\left(\omega_t^{(\tau)}\right) &= -\frac{1}{2} \log\left(\omega_t^{(\tau)}\right) - \frac{1}{2\delta_{(\tau)}^2} \Lambda_\varepsilon^{(\tau)} \Lambda_{y,t}^{(\tau)} \frac{1}{\omega_t^{(\tau)}} - \frac{1}{2\delta_{(\tau)}^2} \Gamma^{(\tau)} \Lambda_\varepsilon^{(\tau)} \omega_t^{(\tau)} + \mathcal{C},
\end{aligned}$$

dove

$$\begin{aligned}
\Lambda_{\alpha,t}^{(\tau)} &= \text{Tr} \left\{ \mathbf{W}^{-1} \left(\hat{\mathbf{P}}_{t+1}^{(\tau)} + \hat{\mathbf{P}}_t^{(\tau)} - 2\hat{\mathbf{P}}_{t+1,t}^{(\tau)} + \hat{\mathbf{a}}_{t+1}^{(\tau)} \hat{\mathbf{a}}_{t+1}^{(\tau)\top} + \hat{\mathbf{a}}_t^{(\tau)} \hat{\mathbf{a}}_t^{(\tau)\top} - 2\hat{\mathbf{a}}_{t+1}^{(\tau)} \hat{\mathbf{a}}_t^{(\tau)\top} \right) \right\} \\
\Lambda_{y,t}^{(\tau)} &= \mathbf{y}_t^\top \mathbf{V}^{-1} \mathbf{y}_t - 2\mathbf{y}_t^\top \mathbf{V}^{-1} \hat{\mathbf{a}}_t^{(\tau)} + \text{Tr} \left\{ \mathbf{V}^{-1} \left(\hat{\mathbf{P}}_t^{(\tau)} + \hat{\mathbf{a}}_t^{(\tau)} \hat{\mathbf{a}}_t^{(\tau)\top} \right) \right\} \\
\Lambda_{\omega,1,t}^{(\tau)} &= \frac{\sqrt{\varphi_{t,(\tau)}^*} K(\chi_{t,(\tau)}^*+1) \left(\sqrt{\psi_{t,(\tau)}^* \varphi_{t,(\tau)}^*} \right)}{\sqrt{\varphi_{t,(\tau)}^*} K(\chi_{t,(\tau)}^*) \left(\sqrt{\psi_{t,(\tau)}^* \varphi_{t,(\tau)}^*} \right)} - 2 \frac{\chi_{t,(\tau)}^*}{\varphi_{t,(\tau)}^*} \\
\Lambda_{\omega,2,t}^{(\tau)} &= \frac{\sqrt{\varphi_{t,(\tau)}^*} K(\chi_{t,(\tau)}^*+1) \left(\sqrt{\psi_{t,(\tau)}^* \varphi_{t,(\tau)}^*} \right)}{\sqrt{\varphi_{t,(\tau)}^*} K(\chi_{t,(\tau)}^*) \left(\sqrt{\psi_{t,(\tau)}^* \varphi_{t,(\tau)}^*} \right)} \\
\Lambda_\varepsilon^{(\tau)} &= \frac{a_{\varepsilon,(\tau)}^*}{b_{\varepsilon,(\tau)}^*} \\
\Gamma^{(\tau)} &= (\boldsymbol{\lambda}^{(\tau)})^\top \mathbf{V}^{-1} (\boldsymbol{\lambda}^{(\tau)}) + 2\delta_{(\tau)}^2
\end{aligned}$$

Dal *log-kernel* delle distribuzioni variazionali si riconosce che:

$$\begin{aligned}
q_\eta^{(\tau)}(\sigma_{\eta,(\tau)}^2) &\propto \mathcal{IG}(\sigma_{\eta,(\tau)}^2 | a_{\eta,(\tau)}^*, b_{\eta,(\tau)}^*) \\
q_\varepsilon^{(\tau)}(\sigma_{\varepsilon,(\tau)}^2) &\propto \mathcal{IG}(\sigma_{\varepsilon,(\tau)}^2 | a_{\varepsilon,(\tau)}^*, b_{\varepsilon,(\tau)}^*) \\
q_{\omega_t}^{(\tau)}(\omega_t^{(\tau)}) &\propto \mathcal{GIG}(\omega_t^{(\tau)} | \chi_{t,(\tau)}^*, \psi_{t,(\tau)}^*, \varphi_{t,(\tau)}^*)
\end{aligned}$$

dove

$$\begin{aligned}
a_{\eta,(\tau)}^* &= a_{\eta,(\tau)}^0 + \frac{m}{2}T - \frac{m}{2}, \\
b_{\eta,(\tau)}^* &= b_{\eta,(\tau)}^0 + \frac{1}{2} \sum_{t=0}^{T-1} \Lambda_{\alpha,t}^{(\tau)}, \\
a_{\varepsilon,(\tau)}^* &= a_{\varepsilon,(\tau)}^0 + T + \frac{n}{2}T, \\
b_{\varepsilon,(\tau)}^* &= b_{\varepsilon,(\tau)}^0 + \frac{1}{2\delta_{(\tau)}^2} \sum_{t=1}^T c_{t,(\tau)}^*, \\
c_{t,(\tau)}^* &= \Lambda_{\omega,1,t}^{(\tau)} \Lambda_{y,t}^{(\tau)} - 2(\boldsymbol{\lambda}^{(\tau)})^\top \mathbf{V}^{-1} (\mathbf{y}_t - \mathbf{K} \hat{\mathbf{a}}_t^{(\tau)}) + \Gamma^{(\tau)} \Lambda_{\omega,2,t}^{(\tau)}, \\
\chi_{t,(\tau)}^* &= 1 - \frac{n}{2}, \\
\psi_{t,(\tau)}^* &= \frac{1}{\delta_{(\tau)}^2} \Lambda_\varepsilon^{(\tau)} \Lambda_{y,t}^{(\tau)}, \\
\varphi_{t,(\tau)}^* &= \frac{1}{\delta_{(\tau)}^2} \Gamma^{(\tau)} \Lambda_\varepsilon^{(\tau)}.
\end{aligned}$$

La distribuzione variazionale per gli stati latenti si ottiene come prodotto del *Kalman Smoother* modificato, a cui in *input* vanno passati i dati, le matrici del

modello e i momenti a posteriori approssimati delle variabili $(\sigma_{\eta,(\tau)}^2)^{-1}$, $(\sigma_{\varepsilon,(\tau)}^2)^{-1}$, $(\omega_t^{(\tau)})$ e $(\omega_t^{(\tau)})^{-1}$ per ogni τ e per ogni t .

Definita la forma delle distribuzioni variazionali, e fissati dei valori iniziali per gli iperparametri, l'algoritmo di stima approssimata prevede di iterare fino a convergenza i seguenti passi di aggiornamento:

1. Per ogni $\tau = \tau_1, \dots, \tau_k$ si aggiorni $q_{\eta}^{(\tau)}(\sigma_{\eta,(\tau)}^2) \propto \mathcal{IG}(\sigma_{\eta,(\tau)}^2 | a_{\eta,(\tau)}^*, b_{\eta,(\tau)}^*)$ ponendo

$$\begin{aligned} a_{\eta,(\tau)}^* &\leftarrow a_{\eta,(\tau)}^0 + \frac{m}{2}T - \frac{m}{2}, \\ b_{\eta,(\tau)}^* &\leftarrow b_{\eta,(\tau)}^0 + \frac{1}{2} \sum_{t=0}^{T-1} (\Lambda_{\alpha,t}^{(\tau)})^{(i)}. \end{aligned}$$

2. Per ogni $\tau = \tau_1, \dots, \tau_k$ si aggiorni $q_{\varepsilon}^{(\tau)}(\sigma_{\varepsilon,(\tau)}^2) \propto \mathcal{IG}(\sigma_{\varepsilon,(\tau)}^2 | a_{\varepsilon,(\tau)}^*, b_{\varepsilon,(\tau)}^*)$ ponendo

$$\begin{aligned} a_{\varepsilon,(\tau)}^* &\leftarrow a_{\varepsilon,(\tau)}^0 + T + \frac{n}{2}T, \\ b_{\varepsilon,(\tau)}^* &\leftarrow b_{\varepsilon,(\tau)}^0 + \frac{1}{2\delta_{(\tau)}^2} \sum_{t=1}^T \left\{ (c_{\varepsilon,t}^{(\tau)})^{(i)} - 2(d_{\varepsilon,t}^{(\tau)})^{(i)} + (e_{\varepsilon,t}^{(\tau)})^{(i)} \right\}, \\ (c_{\varepsilon,t}^{(\tau)})^{(i)} &\leftarrow (\Lambda_{\omega,1,t}^{(\tau)})^{(i)} (\Lambda_{y,t}^{(\tau)})^{(i)}, \\ (d_{\varepsilon,t}^{(\tau)})^{(i)} &\leftarrow (\boldsymbol{\lambda}^{(\tau)})^T \mathbf{V}^{-1} \left(\mathbf{y}_t - \mathbf{K} (\hat{\mathbf{a}}_t^{(\tau)})^{(i)} \right), \\ (e_{\varepsilon,t}^{(\tau)})^{(i)} &\leftarrow \boldsymbol{\Gamma}^{(\tau)} (\Lambda_{\omega,2,t}^{(\tau)})^{(i)}. \end{aligned}$$

3. Per ogni $\tau = \tau_1, \dots, \tau_k$ e per ogni $t = 1, \dots, T$ si aggiorni $q_{\omega_t}^{(\tau)}(\omega_t^{(\tau)}) \propto \mathcal{GIG}(\omega_t^{(\tau)} | \chi_{t,(\tau)}^*, \psi_{t,(\tau)}^*, \varphi_{t,(\tau)}^*)$ ponendo

$$\begin{aligned} \chi_{t,(\tau)}^* &\leftarrow 1 - \frac{n}{2}, \\ \psi_{t,(\tau)}^* &\leftarrow \frac{1}{\delta_{(\tau)}^2} \Lambda_{\varepsilon}^{(\tau)} \Lambda_{y,t}^{(\tau)}, \\ \varphi_{t,(\tau)}^* &\leftarrow \frac{1}{\delta_{(\tau)}^2} \boldsymbol{\Gamma}^{(\tau)} \Lambda_{\varepsilon}^{(\tau)}. \end{aligned}$$

4. Si aggiorni $q_{\alpha}(\tilde{\boldsymbol{\alpha}})$. Condizionatamente ai parametri e ai fattori aumentati il modello è lineare e Gaussiano, quindi si possono trovare le distribuzioni variazionali per gli stati eseguendo in sequenza *Kalman Filter*, *Kalman Smoother* e correzione per troncamento dei momenti lisciati.

Risultati teorici garantiscono che, eseguendo iterativamente i passi di aggiornamento appena definiti, l'algoritmo converge alla distribuzione variazionale che minimizza la divergenza di Kullback-Leibler, dal momento che la distribuzione a posteriori originale è una funzione log-concava, data dal prodotto di funzioni log-concave. Per maggiori dettagli si veda Ormerod e Wand 2010.

5.6 Inferenza Bayesiana o frequentista?

Come si è accennato nel capitolo 3, per modelli quantilici dinamici un approccio inferenziale di tipo Bayesiano risulta particolarmente conveniente. Algoritmi di simulazione, come il *Gibbs Sampling* (Geman e Geman 1984), o di ottimizzazione, come il *variational Bayes* (Ormerod e Wand 2010), ammettono soluzioni in forma analitica per ogni passo di aggiornamento. La semplicità di calcolo rende queste alternative particolarmente attraenti dal punto di vista matematico ed estremamente efficienti dal punto di vista computazionale.

L'importanza di poter adottare soluzioni in forma chiusa assume particolare rilevanza nel momento in cui si lavora con processi latenti, la cui distribuzione a posteriori deve essere calcolata per mezzo di algoritmi di filtraggio e lisciamiento. La rappresentazione stocastica per la distribuzione Laplace Asimmetrica permette di ricondurre modelli quantilici non Gaussiani ad una formulazione lineare e Gaussiana più maneggevole. Grazie a questa rappresentazione è possibile sfruttare mezzi quali il *Kalman filter* e il *Kalman smoother* per fare inferenza a posteriori sugli stati latenti del sistema.

Nel caso in cui fosse necessario estrarre i momenti della distribuzione filtrata o lisciata per gli stati latenti di un modello *State Space* non lineare e/o non Gaussiano, si perderebbero tutte le proprietà di ottimalità del filtro, ed inoltre i risultati ottenuti varrebbero solo in forma approssimata e non esatta. Talvolta, per problemi complessi, come nel caso della *quantile regression* dinamica con disturbi Laplace Asimmetrici, gli algoritmi di filtraggio non sarebbero proprio applicabili, per la natura stessa della distribuzione dei disturbi, asimmetrica e non differenziabile.

In conseguenza di ciò, modelli di regressione dinamici portano inevitabilmente a complicazioni in fase di inferenza, rispetto a modelli di regressione statici.

Passiamo ora a considerare alcune alternative di stima in un contesto inferenziale frequentista.

In prima analisi si potrebbe pensare di generalizzare gli algoritmi di stima solitamente utilizzati per fare inferenza su modelli quantilici statici nel caso dinamico. Vediamo alcuni esempi.

Due alternative percorribili nel contesto della *quantile regression* statica fanno riferimento agli algoritmi MM (Hunter e Lange 2000) e del semplice (Chen 2004). Entrambi questi metodi di ottimizzazione permettono di ricercare efficientemente il punto di minimo della funzione di perdita quantilica, senza fare uso di rappresentazioni stocastiche e *data augmentation*. Adottando gli stessi metodi nel caso dinamico sarebbe necessario stimare come parametri incogniti un gran numero di quantità: almeno due parametri di scala, in aggiunta ad una matrice di stati latenti di dimensione $n \times T$. Alternativamente sarebbe necessario marginalizzare l'effetto dei processi latenti dalla distribuzione dei dati, facendo poi inferenza su quest'ultima.

Sfortunatamente non è possibile adottare metodi di filtraggio esatto, in virtù della non gaussianità della funzione di perdita, dunque non si può stimare in forma chiusa la distribuzione degli stati, né integrarne analiticamente l'effetto dalla verosimiglianza.

A causa di questi problemi tutti gli approcci di calcolo basati sull'ottimizzazione diretta della funzione quantilica risultano poco adeguati per stimare i parametri di un processo dinamico quantilico in forma *State Space*.

Una seconda alternativa molto popolare sia nell'ambito *quantile*, che nella letteratura dei modelli dinamici lineari, è fornita dall'algoritmo EM (Tian, Tian e Zhu 2014) e dai metodi di ottimizzazione variazionali applicati a problemi frequentisti (Tan e Nott 2018). A differenza dei metodi MM e del semplice, questi fanno leva su rappresentazioni aumentate per modelli dipendenti da fattori latenti e variabili nascoste.

Per quanto riguarda l'algoritmo EM, questo non è direttamente applicabile in un contesto di *quantile regression* dinamica e non ammette passi di aggiornamento in forma chiusa. Questo fatto è dovuto alla presenza contemporanea di due strutture latenti all'interno del modello, ovvero il fattore aumentato $\{\omega_t\}_{t=1}^T$ e gli stati $\{\alpha_t\}_{t=0}^T$.

L'algoritmo EM prevede di ricercare il massimo della funzione di log-verosimiglianza iterando un passo di *Expectation* ed uno di *Maximization* fino al raggiungimento di un qualche criterio di convergenza. Il passo di *Expectation* prevede di calcolare il valore atteso della log-verosimiglianza aumentata, marginalizzando gli effetti di tutte le strutture latenti e stocastiche del modello rispetto alla legge di queste, condizionate ai dati e ai parametri. Nel caso della *quantile regression* dinamica questo implicherebbe la necessità di calcolare il valore atteso

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}') = \mathbb{E}_{\omega, \alpha} \left[\log p \left(\{\mathbf{y}_t\}_{t=1}^T \mid \boldsymbol{\theta}, \{\omega_t\}_{t=1}^T, \{\alpha_t\}_{t=1}^T \right) \mid \boldsymbol{\theta}', \{\mathbf{y}_t\}_{t=1}^T \right] \quad (5.6.1)$$

rispetto alla distribuzione $p \left(\{\omega_t\}_{t=1}^T, \{\alpha_t\}_{t=1}^T \mid \boldsymbol{\theta}', \{\mathbf{y}_t\}_{t=1}^T \right)$. Tale valore atteso tuttavia non è derivabile analiticamente e andrebbe dunque approssimato, richiedendo

tipicamente molte risorse di calcolo. Si fa notare inoltre che il valore atteso 5.6.1 è definito come un'integrale in dimensione $n \times T \times T$, il che lo rende irrisolvibile mediante metodi di cubatura deterministica.

Non disponendo di una scrittura analitica per la funzione $Q(\boldsymbol{\theta}, \boldsymbol{\theta}')$, anche il passo di *Maximization* non può essere risolto facilmente, se non ricorrendo a *routine* iterative di calcolo numerico, che chiaramente portano a spendere un gran numero di risorse computazionali.

Per tutti questi motivi l'algoritmo EM non fornisce una strategia applicabile per risolvere il problema di stima quantilico in modelli *State Space*.

L'ultima alternativa che si è presa in considerazione per stimare i parametri del modello in ambito frequentista è costituita dalla famiglia dei metodi di ottimizzazione variazionali, di cui l'algoritmo EM costituisce un caso particolare. Senza addentrarci nei dettagli tecnici, possiamo dire che, similmente all'EM, questi algoritmi di ottimizzazione possono essere separati in tre passi distinti per ogni replicazione del processo iterativo di ricerca del massimo (o del minimo). Il primo prevede di aggiornare le distribuzioni variazionali per i processi e i fattori latenti; il secondo calcola il valore atteso della log-verosimiglianza rispetto alle distribuzioni variazionali e il terzo ottimizza la funzione obiettivo calcolata al punto precedente.

Di tutti gli algoritmi frequentisti menzionati questo è l'unico che permette di fare esplicitamente ricorso a metodi di filtraggio e lisciamiento ottimi e che permette di calcolare il valore atteso della log-verosimiglianza completa in forma analitica. Tuttavia il passo di ottimizzazione richiesto all'ultimo step di aggiornamento in molti casi non ammette forma chiusa e dunque anche in questo caso è necessario ricorrere ad approssimazioni numeriche.

Da un'accurata esplorazione della letteratura esistente, non sono stati trovati altri mezzi applicabili per operare la stima dei parametri in modelli quantilici dinamici secondo l'approccio inferenziale classico. Di conseguenza, le uniche alternative percorribili che permettono di fare inferenza in modo efficiente e con risultati teorici esatti su tutte le quantità coinvolte nei modelli sono da ricercarsi in ambito Bayesiano.

Conclusioni

In questo lavoro di tesi sono stati proposti diversi metodi e algoritmi per affrontare il problema di stima dei quantili di un processo spazio-temporale.

I modelli presentati permettono di estrarre un gran numero di informazioni dai dati, caratterizzandone dettagliatamente la distribuzione dinamica e geolocalizzata.

Dal punto di vista predittivo la *quantile regression* dinamica fornisce un'utile strumento per ottenere indicazioni più ricche rispetto ai classici metodi di estrapolazione in media sotto ipotesi di normalità dei disturbi.

Questa flessibilità è però controbilanciata da un onere computazionale non indifferente richiesto in fase di stima. Questa, per modelli particolarmente complessi, o per insiemi di dati di grandi dimensioni, può diventare insostenibile in termini di tempo e di risorse di calcolo richieste.

Qualora si volesse stimare un solo quantile per volta, i problemi di natura computazionale, possono essere parzialmente aggirati sfruttando le potenzialità del calcolo parallelo, andando quindi ad elaborare in maniera indipendente diversi quantili contemporaneamente.

Tuttavia, come si è visto, la stima separata dei quantili non garantisce a priori che i risultati di stime indipendenti non si incrocino. Andando dunque a gestire anche il problema di quantili multipli si previene la possibilità di *crossing*, ma si aggrava notevolmente la complessità computazionale degli algoritmi, che non possono più essere eseguiti in parallelo.

Sarebbe perciò interessante come sviluppo futuro studiare qualche metodo per rendere più efficienti gli algoritmi presentati. Una possibilità studiata nel quinto capitolo, ma non ancora testata empiricamente, è quella di sfruttare algoritmi *variational Bayes* di inferenza approssimata.

Una seconda possibilità per lavori futuri è quella di sviluppare metodi di aggiornamento *online* per le stime. Questi, oltre ad ampliare i campi di applicabilità dei modelli, potrebbero anche risolvere in parte le problematiche relative alla fase di stima per campioni rilevati su un gran numero di tempi.

Un aspetto che non è stato sviluppato in questa tesi riguarda l'ottimizzazione del filtro di Kalman per risolvere alcuni problemi che potenzialmente possono essere causati da un'inizializzazione diffusa. Un raffinamento di tale algoritmo può quindi essere operata andando a correggere la diffusione mediante un'inizializzazione in forma esatta.

Come ultima osservazione si vuole porre l'accento sul fatto che i modelli proposti poggiano su ipotesi distributive strumentali e su assunzioni distributive non correttamente specificate. Questo fatto, da un punto di vista Bayesiano, costituisce una forzatura in termini interpretativi e porta a delle conclusioni inferenziali che vanno trattate con la dovuta cautela.

Appendice A

Appendice matematica

In questa sezione si discute brevemente la derivazione di tutte le distribuzioni *full conditional* di cui si è fatto uso nei capitoli precedenti.

In generale, data una funzione di densità a posteriori $\pi(\boldsymbol{\vartheta}_{(1)}, \boldsymbol{\vartheta}_{(2)} | \mathbf{y})$ per il vettore di parametri $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_{(1)}, \boldsymbol{\vartheta}_{(2)}) \in \Theta_{(1)} \times \Theta_{(2)}$, supponendo fissati i dati \mathbf{y} , è sempre possibile trovare una fattorizzazione tale che

$$\pi(\boldsymbol{\vartheta}_{(1)}, \boldsymbol{\vartheta}_{(2)} | \mathbf{y}) \propto f(\boldsymbol{\vartheta}_{(1)} | \mathbf{y}) \cdot g(\boldsymbol{\vartheta}_{(2)} | \mathbf{y}) \cdot h(\boldsymbol{\vartheta}_{(1)}, \boldsymbol{\vartheta}_{(2)} | \mathbf{y}), \quad (\text{A.0.1})$$

dove f è una funzione che dipende solo dal parametro $\boldsymbol{\vartheta}_{(1)}$, g dipende solo da $\boldsymbol{\vartheta}_{(2)}$ e h dipende sia da $\boldsymbol{\vartheta}_{(1)}$, che da $\boldsymbol{\vartheta}_{(2)}$.

Sfruttando la fattorizzazione [A.0.1](#), è immediato dimostrare che le due distribuzioni *full conditional* derivanti dalla legge a posteriori congiunta saranno

$$\begin{aligned} \pi(\boldsymbol{\vartheta}_{(1)} | \boldsymbol{\vartheta}_{(2)}, \mathbf{y}) &= \frac{f(\boldsymbol{\vartheta}_{(1)} | \mathbf{y}) \cdot h(\boldsymbol{\vartheta}_{(1)}, \boldsymbol{\vartheta}_{(2)} | \mathbf{y})}{\int_{\Theta_{(1)}} f(\boldsymbol{\vartheta}_{(1)} | \mathbf{y}) \cdot h(\boldsymbol{\vartheta}_{(1)}, \boldsymbol{\vartheta}_{(2)} | \mathbf{y}) d\boldsymbol{\vartheta}_{(1)}}, \\ \pi(\boldsymbol{\vartheta}_{(2)} | \boldsymbol{\vartheta}_{(1)}, \mathbf{y}) &= \frac{g(\boldsymbol{\vartheta}_{(2)} | \mathbf{y}) \cdot h(\boldsymbol{\vartheta}_{(1)}, \boldsymbol{\vartheta}_{(2)} | \mathbf{y})}{\int_{\Theta_{(2)}} g(\boldsymbol{\vartheta}_{(2)} | \mathbf{y}) \cdot h(\boldsymbol{\vartheta}_{(1)}, \boldsymbol{\vartheta}_{(2)} | \mathbf{y}) d\boldsymbol{\vartheta}_{(2)}}, \end{aligned}$$

e dunque

$$\begin{aligned} \pi(\boldsymbol{\vartheta}_{(1)} | \boldsymbol{\vartheta}_{(2)}, \mathbf{y}) &\propto f(\boldsymbol{\vartheta}_{(1)} | \mathbf{y}) \cdot h(\boldsymbol{\vartheta}_{(1)}, \boldsymbol{\vartheta}_{(2)} | \mathbf{y}), \\ \pi(\boldsymbol{\vartheta}_{(2)} | \boldsymbol{\vartheta}_{(1)}, \mathbf{y}) &\propto g(\boldsymbol{\vartheta}_{(2)} | \mathbf{y}) \cdot h(\boldsymbol{\vartheta}_{(1)}, \boldsymbol{\vartheta}_{(2)} | \mathbf{y}). \end{aligned}$$

Da questo semplice risultato si può dedurre che, identificata un'opportuna scomposizione moltiplicativa [A.0.1](#), per trovare il *kernel* delle densità *full conditional* di

interesse è sufficiente scartare dalla densità a posteriori congiunta tutti i termini che non dipendono dal parametro desiderato.

A.1 Distribuzioni *full conditional* per il modello 2.2.1

Si consideri il modello specificato in 2.2.1, con funzione di densità a posteriori per i parametri 2.4.2.

Full conditional a posteriori per $(\sigma_\varepsilon^2 | \sigma_\eta^2, \{\boldsymbol{\alpha}_t\}_{t=0}^T, \{\mathbf{y}_t\}_{t=1}^T)$

Il nucleo non normalizzato della distribuzione *full conditional* di σ_ε^2 è pari a

$$\begin{aligned} \pi(\sigma_\varepsilon^2 | \sigma_\eta^2, \{\boldsymbol{\alpha}_t\}_{t=0}^T, \{\mathbf{y}_t\}_{t=1}^T) &\propto \\ &\propto (\sigma_\varepsilon^2)^{-(a_\varepsilon^0 + \frac{n}{2}T+1)} \exp \left\{ -\frac{1}{\sigma_\varepsilon^2} \left(b_\varepsilon^0 + \frac{1}{2} \sum_{t=1}^T \left\| \mathbf{V}^{-\frac{1}{2}} (\mathbf{y}_t - \mathbf{K}\boldsymbol{\alpha}_t) \right\|_2^2 \right) \right\}. \end{aligned} \quad (\text{A.1.1})$$

Come funzione di σ_ε^2 , il nucleo A.1.1 è proporzionale ad una densità Gamma Inversa (\mathcal{IG}) di parametri

$$a_\varepsilon^* = a_\varepsilon^0 + \frac{n}{2}T, \quad b_\varepsilon^* = b_\varepsilon^0 + \frac{1}{2} \sum_{t=1}^T \left\| \mathbf{V}^{-\frac{1}{2}} (\mathbf{y}_t - \mathbf{K}\boldsymbol{\alpha}_t) \right\|_2^2.$$

Full conditional a posteriori per $(\sigma_\eta^2 | \sigma_\varepsilon^2, \{\boldsymbol{\alpha}_t\}_{t=0}^T, \{\mathbf{y}_t\}_{t=1}^T)$

Similmente il nucleo non normalizzato della distribuzione *full conditional* di σ_η^2 è pari a

$$\begin{aligned} \pi(\sigma_\eta^2 | \sigma_\varepsilon^2, \{\boldsymbol{\alpha}_t\}_{t=0}^T, \{\mathbf{y}_t\}_{t=1}^T) &\propto \\ &\propto (\sigma_\eta^2)^{-(a_\eta^0 + \frac{m}{2}(T-1)+1)} \exp \left\{ -\frac{1}{\sigma_\eta^2} \left(b_\eta^0 + \frac{1}{2} \sum_{t=1}^{T-1} \left\| \mathbf{W}^{-\frac{1}{2}} (\boldsymbol{\alpha}_{t+1} - \boldsymbol{\alpha}_t) \right\|_2^2 \right) \right\}. \end{aligned} \quad (\text{A.1.2})$$

da cui si riconosce il nucleo di una densità \mathcal{IG} di parametri

$$a_\eta^* = a_\eta^0 + \frac{m}{2}(T-1), \quad b_\eta^* = b_\eta^0 + \frac{1}{2} \sum_{t=1}^{T-1} \left\| \mathbf{W}^{-\frac{1}{2}} (\boldsymbol{\alpha}_{t+1} - \boldsymbol{\alpha}_t) \right\|_2^2.$$

Si noti che, condizionatamente a $\{\boldsymbol{\alpha}_t\}_{t=0}^T$ e $\{\mathbf{y}_t\}_{t=1}^T$, la funzione A.1.1 non dipende da σ_η^2 , dunque

$$\pi(\sigma_\varepsilon^2 | \sigma_\eta^2, \{\boldsymbol{\alpha}_t\}_{t=0}^T, \{\mathbf{y}_t\}_{t=1}^T) \propto \pi(\sigma_\varepsilon^2 | \{\boldsymbol{\alpha}_t\}_{t=0}^T, \{\mathbf{y}_t\}_{t=1}^T).$$

Allo stesso modo, condizionatamente agli stati latenti $\{\boldsymbol{\alpha}_t\}_{t=0}^T$, A.1.2 è indipendente da σ_ε^2 e $\{\mathbf{y}_t\}_{t=1}^T$, dunque

$$\pi\left(\sigma_\eta^2 | \sigma_\varepsilon^2, \{\boldsymbol{\alpha}_t\}_{t=0}^T, \{\mathbf{y}_t\}_{t=1}^T\right) \propto \pi\left(\sigma_\eta^2 | \{\boldsymbol{\alpha}_t\}_{t=0}^T\right).$$

In generale, assumendo a priori coniugate naturali con la verosimiglianza Gaussiana, i parametri di scala, come σ_ε^2 e σ_η^2 , avranno distribuzione a posteriori *full conditional* \mathcal{IG} .

A.2 Distribuzioni *full conditional* per il modello 3.2.4

Si consideri ora il modello specificato in 3.2.4, con funzione di densità a posteriori per i parametri 3.2.6.

Full conditional a posteriori per $(\boldsymbol{\beta} | \sigma, \boldsymbol{\omega}, \mathbf{x}, \mathbf{y})$

Consideriamo anzitutto la legge condizionata a posteriori per i parametri di regressione

$$\begin{aligned} \pi(\boldsymbol{\beta} | \sigma, \boldsymbol{\omega}, \mathbf{x}, \mathbf{y}) &\propto \exp\left\{-\frac{1}{2}\boldsymbol{\beta}^\top \mathbf{S}^{-1}\boldsymbol{\beta} - \frac{1}{2}\sum_{t=1}^T \frac{(y_t - \mathbf{x}_t^\top \boldsymbol{\beta} - \lambda\omega_t)^2}{\delta^2\sigma\omega_t}\right\} \\ &\propto \exp\left\{-\frac{1}{2}\boldsymbol{\beta}^\top \left(\mathbf{S}^{-1} + \sum_{t=1}^T \frac{\mathbf{x}_t\mathbf{x}_t^\top}{2\delta^2\sigma\omega_t}\right)\boldsymbol{\beta} + \boldsymbol{\beta}^\top \left(\sum_{t=1}^T \frac{\mathbf{x}_t(y_t - \lambda\omega_t)}{\delta^2\sigma\omega_t}\right)\right\}. \end{aligned}$$

dalla quale si riconosce chiaramente il nucleo di una distribuzione Gaussiana di parametri a posteriori

$$\mathbf{m}^* = \mathbf{S}^* \left(\sum_{t=1}^T \frac{\mathbf{x}_t (y_t - \lambda\omega_t)}{\delta^2\sigma\omega_t} \right), \quad \mathbf{S}^* = \left(\mathbf{S}^{-1} + \sum_{t=1}^T \frac{\mathbf{x}_t\mathbf{x}_t^\top}{\delta^2\sigma\omega_t} \right)^{-1}.$$

Full conditional a posteriori per $(\omega_t | \sigma, \boldsymbol{\beta}, \mathbf{x}, \mathbf{y})$

Passiamo ora a considerare la legge di probabilità *full conditional* per il fattore aumentato $\boldsymbol{\omega}$, la cui densità è proporzionale a

$$\begin{aligned} \pi(\omega_t | \sigma, \boldsymbol{\beta}, \mathbf{x}, \mathbf{y}) &\propto \omega_t^{-\frac{1}{2}} \exp\left\{-\frac{\omega_t}{\sigma} - \frac{(y_t - \mathbf{x}_t^\top \boldsymbol{\beta} - \lambda\omega_t)^2}{2\delta^2\sigma\omega_t}\right\} \\ &\propto \omega_t^{-\frac{1}{2}} \exp\left\{-\frac{(y_t - \mathbf{x}_t^\top \boldsymbol{\beta})^2}{2\delta^2\sigma} \frac{1}{\omega_t} - \frac{\lambda^2 + 2\delta^2}{2\delta^2\sigma} \omega_t\right\}, \end{aligned}$$

che corrisponde al nucleo della densità di una variabile aleatoria $\mathcal{GIG}(\chi, \psi, \varphi)$ (Devroye 2014), di parametri

$$\chi = \frac{1}{2}, \quad \psi = \frac{\lambda^2 + 2\delta^2}{\delta^2\sigma}, \quad \varphi = \frac{(y_t - \mathbf{x}_t^\top \boldsymbol{\beta})^2}{\delta^2\sigma}.$$

Adottiamo ora la trasformazione $\kappa_t = \omega_t^{-1}$. Questa nuova variabile aleatoria ha densità condizionata

$$\pi(\kappa_t | \sigma, \mathbf{x}, \mathbf{y}) \propto \kappa_t^{-\frac{3}{2}} \exp \left\{ -\frac{(y_t - \mathbf{x}_t^\top \boldsymbol{\beta})^2}{2\delta^2\sigma} \kappa_t - \frac{\lambda^2 + 2\delta^2}{2\delta^2\sigma} \frac{1}{\kappa_t} \right\}$$

che da luogo, oltre che ad una distribuzione $\mathcal{GIG}(\tilde{\chi}, \psi, \varphi)$, dove $\tilde{\chi} = -1/2$, anche ad un legge di probabilità Gaussiana Inversa $\mathcal{IN}(\tilde{\gamma}, \tilde{\mu})$, di parametri $\tilde{\gamma} = \psi$ e $\tilde{\mu} = \sqrt{\psi/\varphi}$ (si veda Michael, Schucany e Haas 1976).

Full conditional a posteriori per $(\sigma | \boldsymbol{\beta}, \mathbf{x}, \mathbf{y})$

Andiamo infine a studiare la *full conditional* collassata per il parametri di scala σ . Dalla distribuzione a posteriori 3.2.6 si può ricavare la legge di probabilità *full conditional* congiunta per il parametro σ e per il fattore aumentato $\boldsymbol{\omega}$, la cui densità è proporzionale a

$$\pi(\sigma, \boldsymbol{\omega} | \boldsymbol{\beta}, \mathbf{x}, \mathbf{y}) \propto \sigma^{-(a+\frac{3}{2}T+1)} \exp \left\{ -\frac{b}{\sigma} \right\} \prod_{t=1}^T \omega_t^{-\frac{1}{2}} \exp \left\{ -\frac{\omega_t}{\sigma} - \frac{1}{2} \frac{(y_t - \mathbf{x}_t^\top \boldsymbol{\beta} - \lambda\omega_t)^2}{\delta^2\sigma\omega_t} \right\}.$$

A questo punto l'obiettivo è quello di ricavare la distribuzione *full conditional* marginale di σ , integrando analiticamente il fattore aumentato $\boldsymbol{\omega}$. Per farlo è necessario risolvere il seguente integrale:

$$\begin{aligned} \pi(\sigma | \boldsymbol{\beta}, \mathbf{y}, \mathbf{x}) &= \int_{\Omega^T} \sigma^{-(a+\frac{3}{2}T+1)} \exp \left\{ -\frac{b}{\sigma} \right\} \prod_{t=1}^T \omega_t^{-\frac{1}{2}} \exp \left\{ -\frac{\omega_t}{\sigma} - \frac{(y_t - \mathbf{x}_t^\top \boldsymbol{\beta} - \lambda\omega_t)^2}{2\delta^2\sigma\omega_t} \right\} d\boldsymbol{\omega}_t \\ &= \sigma^{-(a+\frac{3}{2}T+1)} \exp \left\{ -\frac{b}{\sigma} \right\} \prod_{t=1}^T \int_{\Omega} \omega_t^{-\frac{1}{2}} \exp \left\{ -\frac{\omega_t}{\sigma} - \frac{(y_t - \mathbf{x}_t^\top \boldsymbol{\beta} - \lambda\omega_t)^2}{2\delta^2\sigma\omega_t} \right\} d\omega_t \\ &= \sigma^{-(a+\frac{3}{2}T+1)} \exp \left\{ -\frac{b}{\sigma} \right\} \exp \left\{ \sum_{t=1}^T \frac{\lambda(y_t - \mathbf{x}_t^\top \boldsymbol{\beta})}{\delta^2\sigma} \right\} \\ &\quad \times \prod_{t=1}^T \int_{\Omega} \omega_t^{-\frac{1}{2}} \exp \left\{ -\frac{(y_t - \mathbf{x}_t^\top \boldsymbol{\beta})^2}{2\delta^2\sigma} \frac{1}{\omega_t} - \frac{\lambda^2 + 2\delta^2}{2\delta^2\sigma} \omega_t \right\} d\omega_t. \end{aligned}$$

Definiamo ora

$$\Lambda(\sigma | \boldsymbol{\beta}, \mathbf{x}_t, \mathbf{y}_t) = \int_{\Omega} \omega_t^{-\frac{1}{2}} \exp \left\{ -\frac{(y_t - \mathbf{x}_t^\top \boldsymbol{\beta})^2}{2\delta^2\sigma} \frac{1}{\omega_t} - \frac{\lambda^2 + 2\delta^2}{2\delta^2\sigma} \omega_t \right\} d\omega_t. \quad (\text{A.2.1})$$

All'interno di quest'ultimo integrale riconosciamo il nucleo di una distribuzione $\mathcal{GIG}(\chi, \psi, \varphi)$, dove

$$\chi = \frac{1}{2}, \quad \psi = \frac{\lambda^2 + 2\delta^2}{\delta^2\sigma}, \quad \varphi = \frac{(y_t - \mathbf{x}_t^\top \boldsymbol{\beta})^2}{\delta^2\sigma}.$$

Una variabile aleatoria con distribuzione $\mathcal{GIG}(\chi, \psi, \varphi)$ ha funzione di densità

$$\mathcal{GIG}(w|\chi, \psi, \varphi) = \frac{\varphi^{-\frac{\chi}{2}} \psi^{\frac{\chi}{2}}}{2 K_\chi(\sqrt{\psi\varphi})} w^{(\chi-1)} \exp\left\{-\frac{1}{2}\left(\psi w - \varphi \frac{1}{w}\right)\right\},$$

dove $K_\chi(\cdot)$ è una funzione di Bessel del terzo tipo. Ricordiamo inoltre che la funzione di Bessel modificata è tale che

$$K_{\frac{1}{2}}(u) = \sqrt{\frac{\pi}{2u}} e^{-u}.$$

Sfruttando congiuntamente le proprietà della funzione di Bessel e il fatto che A.2.1 è l'integrale di una distribuzione \mathcal{GIG} non normalizzata, operando le opportune semplificazioni, ricaviamo che

$$\begin{aligned} \Lambda(\sigma|\boldsymbol{\beta}, \mathbf{x}_t, \mathbf{y}_t) &\propto 2 \varphi^{\frac{\chi}{2}} \psi^{-\frac{\chi}{2}} K_\chi(\sqrt{\psi\varphi}) \\ &\propto \sqrt{|y_t - \mathbf{x}_t^\top \boldsymbol{\beta}|} K_{\frac{1}{2}}\left(\frac{\sqrt{\lambda^2 + 2\delta^2}}{\delta^2\sigma} |y_t - \mathbf{x}_t^\top \boldsymbol{\beta}|\right) \\ &\propto \sigma^{\frac{1}{2}} \exp\left\{-\frac{\sqrt{\lambda^2 + 2\delta^2}}{\delta^2\sigma} |y_t - \mathbf{x}_t^\top \boldsymbol{\beta}|\right\}. \end{aligned}$$

Sostituendo $\Lambda(\sigma|\boldsymbol{\beta}, \mathbf{x}_t, \mathbf{y}_t)$ in $\pi(\sigma|\boldsymbol{\beta}, \mathbf{y}, \mathbf{x})$, si ottiene

$$\begin{aligned} \pi(\sigma|\boldsymbol{\beta}, \mathbf{y}, \mathbf{x}) &= \sigma^{-(a+\frac{3}{2}T+\frac{3}{2})} \exp\left\{-\frac{b}{\sigma}\right\} \\ &\times \exp\left\{\frac{1}{\delta^2\sigma} \sum_{t=1}^T \left(\lambda (y_t - \mathbf{x}_t^\top \boldsymbol{\beta}) - \sqrt{\lambda^2 + 2\delta^2} |y_t - \mathbf{x}_t^\top \boldsymbol{\beta}|\right)\right\}. \end{aligned}$$

Per come sono definite le quantità λ e δ^2 valgono le seguenti uguaglianze:

$$\begin{aligned} \Psi(\boldsymbol{\beta}, \mathbf{x}_t, \mathbf{y}_t) &= \lambda (y_t - \mathbf{x}_t^\top \boldsymbol{\beta}) - \sqrt{\lambda^2 + 2\delta^2} |y_t - \mathbf{x}_t^\top \boldsymbol{\beta}| \\ &= (y_t - \mathbf{x}_t^\top \boldsymbol{\beta}) \left[\lambda - \sqrt{\lambda^2 + 2\delta^2} \operatorname{sgn}(y_t - \mathbf{x}_t^\top \boldsymbol{\beta})\right] \\ &= \frac{2}{\tau(1-\tau)} (y_t - \mathbf{x}_t^\top \boldsymbol{\beta}) \left[\tau - \mathbb{1}(y_t - \mathbf{x}_t^\top \boldsymbol{\beta} < 0)\right] \\ &= -\frac{2}{\tau(1-\tau)} \rho_\tau(y_t - \mathbf{x}_t^\top \boldsymbol{\beta}), \end{aligned}$$

dove $\rho_\tau(\cdot)$ è la funzione di perdita quantilica, definita in 3.1.1.

Andando a sostituire il risultato ottenuto un'ultima volta si ottiene

$$\pi(\sigma | \boldsymbol{\beta}, \mathbf{y}, \mathbf{x}) \propto \sigma^{-(a + \frac{3}{2}T + \frac{3}{2})} \exp \left\{ -\frac{b}{\sigma} - \frac{1}{\sigma} \sum_{t=1}^T \rho_{\tau}(y_t - \mathbf{x}_t^{\top} \boldsymbol{\beta}) \right\},$$

da cui riconosciamo il nucleo di una distribuzione \mathcal{IG} di parametri

$$a^* = a^0 + \frac{3}{2}T + \frac{1}{2}, \quad b^* = b^0 + \sum_{t=1}^T \rho_{\tau}(y_t - \mathbf{x}_t^{\top} \boldsymbol{\beta}^{(k)}).$$

A.3 Distribuzioni *full conditional* per il modello 3.3.3

Si consideri ora il modello specificato in 3.3.3, con funzione di densità a posteriori per i parametri 3.3.4.

Full conditional a posteriori per $(\sigma_{\eta}^2 | \sigma_{\varepsilon}^2, \{\boldsymbol{\omega}_t\}_{t=1}^T, \{\boldsymbol{\alpha}_t\}_{t=0}^T, \{\mathbf{y}_t\}_{t=1}^T)$

La distribuzione *full conditional* per σ_{η}^2 ha densità proporzionale alla funzione

$$\begin{aligned} \pi(\sigma_{\eta}^2 | \sigma_{\varepsilon}^2, \{\boldsymbol{\omega}_t\}_{t=1}^T, \{\boldsymbol{\alpha}_t\}_{t=0}^T, \{\mathbf{y}_t\}_{t=1}^T) &\propto \\ &\propto (\sigma_{\eta}^2)^{-(a_{\eta}^0 + \frac{m}{2}(T-1) + 1)} \exp \left\{ -\frac{1}{\sigma_{\eta}^2} \left(b_{\eta}^0 + \frac{1}{2} \sum_{t=1}^{T-1} \left\| \mathbf{W}^{-\frac{1}{2}} (\boldsymbol{\alpha}_{t+1} - \boldsymbol{\alpha}_t) \right\|_2^2 \right) \right\}, \end{aligned}$$

che non dipende da σ_{ε}^2 , $\{\boldsymbol{\omega}_t\}_{t=1}^T$ e $\{\mathbf{y}_t\}_{t=1}^T$, quindi si può operare la semplificazione:

$$\pi(\sigma_{\eta}^2 | \{\boldsymbol{\alpha}_t\}_{t=0}^T) \propto \pi(\sigma_{\eta}^2 | \sigma_{\varepsilon}^2, \{\boldsymbol{\omega}_t\}_{t=1}^T, \{\boldsymbol{\alpha}_t\}_{t=0}^T, \{\mathbf{y}_t\}_{t=1}^T).$$

Il *kernel* così ottenuto ha densità riconducibile ad una distribuzione \mathcal{IG} di parametri

$$a_{\eta}^* = a_{\eta}^0 + \frac{m}{2}(T-1), \quad b_{\eta}^* = b_{\eta}^0 + \frac{1}{2} \sum_{t=1}^{T-1} \left\| \mathbf{W}^{-\frac{1}{2}} (\boldsymbol{\alpha}_{t+1} - \boldsymbol{\alpha}_t) \right\|_2^2.$$

Full conditional a posteriori per $(\sigma_{\varepsilon}^2 | \sigma_{\eta}^2, \{\boldsymbol{\omega}_t\}_{t=1}^T, \{\boldsymbol{\alpha}_t\}_{t=0}^T, \{\mathbf{y}_t\}_{t=1}^T)$

Adottando la stessa strategia di calcolo per σ_{ε}^2 si ottiene il nucleo

$$\begin{aligned} \pi(\sigma_{\varepsilon}^2 | \sigma_{\eta}^2, \{\boldsymbol{\omega}_t\}_{t=1}^T, \{\boldsymbol{\alpha}_t\}_{t=0}^T, \{\mathbf{y}_t\}_{t=1}^T) &\propto \\ &\propto (\sigma_{\varepsilon}^2)^{-(a_{\varepsilon}^0 + \frac{3}{2}nT + 1)} \exp \left\{ -\frac{b_{\varepsilon}^0}{\sigma_{\varepsilon}^2} - \sum_{t=1}^T \sum_{i=1}^n \left(\frac{\omega_{i,t}}{\sigma_{\varepsilon}^2} - \frac{(y_{i,t} - \lambda \omega_{i,t} - \mathbf{k}_i^{\top} \boldsymbol{\alpha}_t)^2}{2 \delta^2 \sigma_{\varepsilon}^2 \omega_{i,t}} \right) \right\}, \end{aligned}$$

e anche in questo caso quindi la distribuzione *full conditional* è una \mathcal{IG} di parametri

$$a_\varepsilon^* = a_\varepsilon^0 + \frac{3}{2}nT, \quad b_\varepsilon^* = b_\varepsilon^0 - \sum_{t=1}^T \sum_{i=1}^n \left(\omega_{i,t} - \frac{(y_{i,t} - \lambda\omega_{i,t} - \mathbf{k}_i^\top \boldsymbol{\alpha}_t)^2}{2\delta^2 \omega_{i,t}} \right),$$

da cui possiamo marginalizzare σ_η^2 , ottenendo

$$\pi \left(\sigma_\varepsilon^2 \mid \{\boldsymbol{\omega}_t\}_{t=1}^T, \{\boldsymbol{\alpha}_t\}_{t=0}^T, \{\mathbf{y}_t\}_{t=1}^T \right) \propto \pi \left(\sigma_\varepsilon^2 \mid \sigma_\eta^2, \{\boldsymbol{\omega}_t\}_{t=1}^T, \{\boldsymbol{\alpha}_t\}_{t=0}^T, \{\mathbf{y}_t\}_{t=1}^T \right).$$

Full conditional a posteriori per $\left(\omega_{i,t} \mid \sigma_\eta^2, \sigma_\varepsilon^2, \{\boldsymbol{\alpha}_t\}_{t=0}^T, \{\mathbf{y}_t\}_{t=1}^T \right)$

Infine, la distribuzione *full conditional* non normalizzata per i fattori latenti è data da

$$\pi \left(\omega_{i,t} \mid \sigma_\eta^2, \sigma_\varepsilon^2, \{\boldsymbol{\alpha}_t\}_{t=0}^T, \{\mathbf{y}_t\}_{t=1}^T \right) \propto \omega_{i,t}^{-\frac{1}{2}} \exp \left\{ -\frac{\omega_{i,t}}{\sigma_\varepsilon^2} - \frac{(y_{i,t} - \lambda\omega_{i,t} - \mathbf{k}_i^\top \boldsymbol{\alpha}_t)^2}{2\delta^2 \sigma_\varepsilon^2 \omega_{i,t}} \right\},$$

dalla quale si riconosce il nucleo di una distribuzione \mathcal{GIG} con parametro di forma pari a $1/2$ e, come mostrato in A.2, trasformando $\omega_{i,t}$ in $\omega_{i,t}^{-1}$ otteniamo una variabile aleatoria con distribuzione condizionata \mathcal{IN} di parametri

$$\psi_{i,t}^* = \sqrt{\frac{\lambda^2 + 2\delta^2}{(y_{i,t} - \mathbf{k}_i^\top \boldsymbol{\alpha}_t)^2}}, \quad \varphi_{i,t}^* = \frac{\lambda^2 + 2\delta^2}{\delta^2 \sigma_\varepsilon^2}.$$

A.4 Distribuzioni *full conditional* per il modello 3.4.4

Si consideri ora il modello specificato in 3.4.4, con funzione di densità a posteriori per i parametri 3.4.5.

Full conditional a posteriori per $\left(\sigma_\varepsilon^2 \mid \sigma_\eta^2, \{\boldsymbol{\omega}_t\}_{t=1}^T, \{\boldsymbol{\alpha}_t\}_{t=0}^T, \{\mathbf{y}_t\}_{t=1}^T \right)$

Isolando i termini della distribuzione a posteriori che dipendono solo da σ_ε^2 otteniamo il *kernel*

$$\begin{aligned} \pi \left(\sigma_\varepsilon^2 \mid \sigma_\eta^2, \{\boldsymbol{\omega}_t\}_{t=1}^T, \{\boldsymbol{\alpha}_t\}_{t=0}^T, \{\mathbf{y}_t\}_{t=1}^T \right) &\propto \\ &\propto (\sigma_\varepsilon^2)^{-(a_\varepsilon^0 + T + \frac{n}{2}T + 1)} \exp \left\{ -\frac{1}{\sigma_\varepsilon^2} \left(b_\varepsilon^0 + \sum_{t=1}^T \left(\omega_t + \left\| \frac{\mathbf{V}^{-\frac{1}{2}}(\mathbf{y}_t - \lambda\boldsymbol{\omega}_t - \mathbf{K}\boldsymbol{\alpha}_t)}{2\delta^2 \omega_t} \right\|_2^2 \right) \right) \right\}. \end{aligned}$$

La funzione così trovata identifica il nucleo di una densità \mathcal{IG} di parametri

$$\begin{aligned} a_\varepsilon^* &= a_\varepsilon^0 + T + \frac{n}{2}T, \\ b_\varepsilon^* &= b_\varepsilon^0 + \sum_{t=1}^T \left(\omega_t + \frac{1}{2\delta^2 \omega_t} \left\| \mathbf{V}^{-\frac{1}{2}}(\mathbf{y}_t - \lambda\boldsymbol{\omega}_t - \mathbf{K}\boldsymbol{\alpha}_t) \right\|_2^2 \right). \end{aligned}$$

Full conditional a posteriori per $\left(\sigma_\eta^2 \mid \sigma_\varepsilon^2, \{\omega_t\}_{t=1}^T, \{\boldsymbol{\alpha}_t\}_{t=0}^T, \{\mathbf{y}_t\}_{t=1}^T\right)$

Adottando lo stesso procedimento otteniamo la distribuzione *full conditional* per σ_η^2 , che risulta proporzionale a

$$\begin{aligned} \pi\left(\sigma_\eta^2 \mid \sigma_\varepsilon^2, \{\omega_t\}_{t=1}^T, \{\boldsymbol{\alpha}_t\}_{t=0}^T, \{\mathbf{y}_t\}_{t=1}^T\right) &\propto \\ &\propto \left(\sigma_\eta^2\right)^{-\left(a_\eta^0 + \frac{m}{2}(T-1) + 1\right)} \exp\left\{-\frac{1}{\sigma_\eta^2}\left(b_\eta^0 + \frac{1}{2}\sum_{t=1}^{T-1}\left\|\mathbf{W}^{-\frac{1}{2}}(\boldsymbol{\alpha}_{t+1} - \boldsymbol{\alpha}_t)\right\|_2^2\right)\right\}, \end{aligned}$$

a cui corrisponde la legge di probabilità \mathcal{IG} di parametri

$$a_\eta^* = a_\eta^0 + \frac{m}{2}(T-1), \quad b_\eta^* = b_\eta^0 + \frac{1}{2}\sum_{t=1}^{T-1}\left\|\mathbf{W}^{-\frac{1}{2}}(\boldsymbol{\alpha}_{t+1} - \boldsymbol{\alpha}_t)\right\|_2^2.$$

Full conditional a posteriori per $\left(\omega_t \mid \sigma_\eta^2, \sigma_\varepsilon^2, \{\boldsymbol{\alpha}_t\}_{t=0}^T, \{\mathbf{y}_t\}_{t=1}^T\right)$

Per completare l'identificazione delle distribuzioni *full conditional* osserviamo che ω_t ha legge a posteriori condizionata pari a

$$\begin{aligned} \pi\left(\omega_t \mid \sigma_\eta^2, \sigma_\varepsilon^2, \{\boldsymbol{\alpha}_t\}_{t=0}^T, \{\mathbf{y}_t\}_{t=1}^T\right) &\propto \\ &\propto \omega_t^{-\frac{n}{2}} \exp\left\{-\frac{\omega_t}{\sigma_\varepsilon^2} - \frac{1}{2\delta^2\sigma_\varepsilon^2\omega_t}\left\|\mathbf{V}^{-\frac{1}{2}}(\mathbf{y}_t - \boldsymbol{\lambda}\omega_t - \mathbf{K}\boldsymbol{\alpha}_t)\right\|_2^2\right\} \\ &\propto \omega_t^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\delta^2\sigma_\varepsilon^2}\left(\left(\boldsymbol{\lambda}^\top\mathbf{V}^{-1}\boldsymbol{\lambda} + 2\delta^2\right)\omega_t + \left\|\mathbf{V}^{-\frac{1}{2}}(\mathbf{y}_t - \mathbf{K}\boldsymbol{\alpha}_t)\right\|_2^2\frac{1}{\omega_t}\right)\right\}, \end{aligned}$$

dalla quale si riconosce la forma funzionale associata alla densità di una variabile aleatoria $\mathcal{GIG}(\chi_t^*, \psi_t^*, \varphi_t^*)$ di parametri

$$\begin{aligned} \chi_t^* &= \frac{2-n}{2}, \\ \psi_t^* &= \frac{1}{\delta^2\sigma_\varepsilon^2}\left(\boldsymbol{\lambda}^\top\mathbf{V}^{-1}\boldsymbol{\lambda} + 2\delta^2\right), \\ \varphi_t^* &= \frac{1}{\delta^2\sigma_\varepsilon^2}\left\|\mathbf{V}^{-\frac{1}{2}}(\mathbf{y}_t - \mathbf{K}\boldsymbol{\alpha}_t)\right\|_2^2. \end{aligned}$$

A.5 Distribuzioni *full conditional* per il modello 5.2.3

Il modello 5.2.3 è costruito dalla composizione modulare di k modelli 3.4.4 indipendenti, soggetti a vincoli lineari sugli stati. A partire da questa considerazione e dall'analisi della distribuzione a posteriori 5.2.4, si derivano facilmente le distribuzioni *full conditional* per i parametri e i fattori latenti del modello congiunto, le quali avranno la stessa forma di quelle discusse nella sezione A.4, eccezion fatta per la distribuzione a posteriori vincolata degli stati, che è stata discussa in sezione 5.3.

Full conditional a posteriori per $\left(\sigma_{\eta,(\tau)}^2 \mid \tilde{\sigma}_\varepsilon, \{\tilde{\omega}_t\}_{t=1}^T, \{\tilde{\alpha}_t\}_{t=0}^T, \{\tilde{\mathbf{y}}_t\}_{t=1}^T\right)$

La densità *full conditional* per $\sigma_{\eta,(\tau)}^2$ risulta proporzionale a

$$\begin{aligned} \pi \left(\sigma_{\eta,(\tau)}^2 \mid \tilde{\sigma}_\varepsilon, \{\tilde{\omega}_t\}_{t=1}^T, \{\tilde{\alpha}_t\}_{t=0}^T, \{\tilde{\mathbf{y}}_t\}_{t=1}^T \right) &\propto \\ &\propto \left(\sigma_{\eta,(\tau)}^2 \right)^{-\left(a_{\eta,(\tau)}^0 + \frac{m}{2}(T-1)+1\right)} \exp \left\{ -\frac{1}{\sigma_{\eta,(\tau)}^2} \left(b_{\eta,(\tau)}^0 + \frac{1}{2} \sum_{t=1}^{T-1} \left\| \mathbf{W}^{-\frac{1}{2}} \left(\boldsymbol{\alpha}_{t+1}^{(\tau)} - \boldsymbol{\alpha}_t^{(\tau)} \right) \right\|_2^2 \right) \right\}. \end{aligned}$$

che è il nucleo di una legge \mathcal{IG} di parametri

$$a_{\eta,(\tau)}^* = a_{\eta,(\tau)}^0 + \frac{m}{2}(T-1), \quad b_{\eta,(\tau)}^* = b_{\eta,(\tau)}^0 + \frac{1}{2} \sum_{t=1}^{T-1} \left\| \mathbf{W}^{-\frac{1}{2}} \left(\boldsymbol{\alpha}_{t+1}^{(\tau)} - \boldsymbol{\alpha}_t^{(\tau)} \right) \right\|_2^2.$$

Full conditional a posteriori per $\left(\sigma_{\varepsilon,(\tau)}^2 \mid \tilde{\sigma}_\eta, \{\tilde{\omega}_t\}_{t=1}^T, \{\tilde{\alpha}_t\}_{t=0}^T, \{\tilde{\mathbf{y}}_t\}_{t=1}^T\right)$

Il *kernel* della distribuzione *full conditional* per $\sigma_{\varepsilon,(\tau)}^2$ risulta pari a

$$\begin{aligned} \pi \left(\sigma_{\varepsilon,(\tau)}^2 \mid \tilde{\sigma}_\eta, \{\tilde{\omega}_t\}_{t=1}^T, \{\tilde{\alpha}_t\}_{t=0}^T, \{\tilde{\mathbf{y}}_t\}_{t=1}^T \right) &\propto \\ &\propto \left(\sigma_{\varepsilon,(\tau)}^2 \right)^{-\left(a_{\varepsilon,(\tau)}^0 + T + \frac{n}{2}T + 1\right)} \exp \left\{ -\frac{1}{\sigma_{\varepsilon,(\tau)}^2} \left(b_{\varepsilon,(\tau)}^0 - \sum_{t=1}^T \omega_t^{(\tau)} \right) \right\} \\ &\times \exp \left\{ -\frac{1}{2\delta_{(\tau)}^2 \sigma_{\varepsilon,(\tau)}^2} \sum_{t=1}^T \frac{1}{\omega_t^{(\tau)}} \left\| \mathbf{V}^{-\frac{1}{2}} \left(\mathbf{y}_t - \boldsymbol{\lambda}^{(\tau)} \omega_t^{(\tau)} - \mathbf{K} \boldsymbol{\alpha}_t^{(\tau)} \right) \right\|_2^2 \right\}, \end{aligned}$$

da cui si riconosce la forma funzionale di una densità \mathcal{IG} di parametri

$$\begin{aligned} a_{\varepsilon,(\tau)}^* &= a_{\varepsilon,(\tau)}^0 + T + \frac{n}{2}T, \\ b_{\varepsilon,(\tau)}^* &= b_{\varepsilon,(\tau)}^0 + \sum_{t=1}^T \left(\omega_t^{(\tau)} + \frac{1}{2\delta_{(\tau)}^2 \omega_t^{(\tau)}} \left\| \mathbf{V}^{-\frac{1}{2}} \left(\mathbf{y}_t - \boldsymbol{\lambda} \omega_t^{(\tau)} - \mathbf{K} \boldsymbol{\alpha}_t^{(\tau)} \right) \right\|_2^2 \right). \end{aligned}$$

Full conditional a posteriori per $\left(\omega_t^{(\tau)} \mid \tilde{\sigma}_\eta, \tilde{\sigma}_\varepsilon, \{\tilde{\alpha}_t\}_{t=0}^T \mid \{\tilde{\mathbf{y}}_t\}_{t=1}^T\right)$

Completiamo la specificazione delle *full conditional* per il modello A.5 andando a studiare la distribuzione condizionata di $\omega_t^{(\tau)}$:

$$\begin{aligned} \pi \left(\omega_t^{(\tau)} \mid \tilde{\sigma}_\eta, \tilde{\sigma}_\varepsilon, \{\tilde{\omega}_t^{(-\tau)}\}_{t=1}^T, \{\tilde{\alpha}_t\}_{t=0}^T \mid \{\tilde{\mathbf{y}}_t\}_{t=1}^T \right) &\propto \\ &\propto \left(\omega_t^{(\tau)} \right)^{-\frac{n}{2}} \exp \left\{ -\frac{\omega_t^{(\tau)}}{\sigma_{\varepsilon,(\tau)}^2} - \frac{1}{2\delta_{(\tau)}^2 \sigma_{\varepsilon,(\tau)}^2 \omega_t^{(\tau)}} \left\| \mathbf{V}^{-\frac{1}{2}} \left(\mathbf{y}_t - \boldsymbol{\lambda}^{(\tau)} \omega_t^{(\tau)} - \mathbf{K} \boldsymbol{\alpha}_t^{(\tau)} \right) \right\|_2^2 \right\} \\ &\propto \left(\omega_t^{(\tau)} \right)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \left(\frac{\boldsymbol{\lambda}_{(\tau)}^\top \mathbf{V}^{-1} \boldsymbol{\lambda}_{(\tau)} + 2\delta_{(\tau)}^2}{\delta_{(\tau)}^2 \sigma_{\varepsilon,(\tau)}^2} \omega_t^{(\tau)} + \left\| \frac{\mathbf{V}^{-\frac{1}{2}} \left(\mathbf{y}_t - \mathbf{K} \boldsymbol{\alpha}_t^{(\tau)} \right)}{\delta_{(\tau)}^2 \sigma_{\varepsilon,(\tau)}^2} \right\|_2^2 \frac{1}{\omega_t^{(\tau)}} \right) \right\}. \end{aligned}$$

Dalla distribuzione a posteriori condizionata si riconosce il nucleo di una legge \mathcal{GIG} di parametri

$$\begin{aligned}\chi_{t,(\tau)}^* &= \frac{2-n}{2}, \\ \psi_{t,(\tau)}^* &= \frac{1}{\delta_{(\tau)}^2 \sigma_{\varepsilon,(\tau)}^2} \left(\boldsymbol{\lambda}_{(\tau)}^\top \mathbf{V}^{-1} \boldsymbol{\lambda}_{(\tau)} + 2\delta_{(\tau)}^2 \right), \\ \varphi_{t,(\tau)}^* &= \frac{1}{\delta_{(\tau)}^2 \sigma_{\varepsilon,(\tau)}^2} \left\| \mathbf{V}^{-\frac{1}{2}} \left(\mathbf{y}_t - \mathbf{K} \boldsymbol{\alpha}_t^{(\tau)} \right) \right\|_2^2.\end{aligned}$$

A.6 Distribuzioni variazionali per il modello 5.2.3

Sotto le ipotesi di fattorizzazione 5.5.11, le distribuzioni variazionali per il modello 5.2.3 possono essere derivate marginalizzando analiticamente i parametri e i fattori latenti dalle distribuzioni *full conditional* presentate in A.5, come mostrato in 5.5.10.

Variational approximation a posteriori per $q_\eta^{(\tau)}(\sigma_{\eta,(\tau)}^2)$

Per le proprietà del valore atteso otteniamo che la log-densità variazionale ottima per il parametro $\sigma_{\eta,(\tau)}^2$ è

$$\begin{aligned}\log q_\eta^{(\tau)}(\sigma_{\eta,(\tau)}^2) &= - \left(a_{\eta,(\tau)}^0 + \frac{m}{2}(T-1) + 1 \right) \log(\sigma_{\eta,(\tau)}^2) - \frac{b_{\eta,(\tau)}^0}{\sigma_{\eta,(\tau)}^2} \\ &\quad - \frac{1}{2\sigma_{\eta,(\tau)}^2} \sum_{t=0}^{T-1} \mathbb{E}_\alpha \left[\left\| \mathbf{W}^{-\frac{1}{2}} \left(\boldsymbol{\alpha}_{t+1}^{(\tau)} - \boldsymbol{\alpha}_t^{(\tau)} \right) \right\|_2^2 \right] + \mathcal{C},\end{aligned}$$

dove $\mathbb{E}_\alpha[\cdot]$ denota il valore atteso calcolato rispetto a $q_\alpha(\tilde{\boldsymbol{\alpha}})$.

Dalla forma del log-*kernel* variazionale per $\sigma_{\eta,(\tau)}^2$ si riconosce la densità di una distribuzione \mathcal{IG} di parametri

$$\begin{aligned}a_{\eta,(\tau)}^* &= a_{\eta,(\tau)}^0 + \frac{m}{2}(T-1), \\ b_{\eta,(\tau)}^* &= b_{\eta,(\tau)}^0 + \frac{1}{2} \sum_{t=0}^{T-1} \mathbb{E}_\alpha \left[\left\| \mathbf{W}^{-\frac{1}{2}} \left(\boldsymbol{\alpha}_{t+1}^{(\tau)} - \boldsymbol{\alpha}_t^{(\tau)} \right) \right\|_2^2 \right].\end{aligned}$$

Variational approximation a posteriori per $q_\varepsilon^{(\tau)}(\sigma_{\varepsilon,(\tau)}^2)$

Andando ad applicare lo stesso procedimento di marginalizzazione condizionata per la distribuzione *full conditional* di $\sigma_{\varepsilon,(\tau)}^2$, si ottiene

$$\begin{aligned} \log q_\varepsilon^{(\tau)}(\sigma_{\varepsilon,(\tau)}^2) &= - \left(a_{\varepsilon,(\tau)}^0 + T + \frac{n}{2}T + 1 \right) \log \sigma_{\varepsilon,(\tau)}^2 - \frac{1}{\sigma_{\varepsilon,(\tau)}^2} \left(b_{\varepsilon,(\tau)}^0 + \sum_{t=1}^T \mathbb{E}_\omega \left[\omega_t^{(\tau)} \right] \right) \\ &\quad - \frac{1}{2 \delta_{(\tau)}^2 \sigma_{\varepsilon,(\tau)}^2} \sum_{t=1}^T \mathbb{E}_{\omega, \alpha} \left[\frac{1}{\omega_t^{(\tau)}} \left\| \mathbf{V}^{-\frac{1}{2}} \left(\mathbf{y}_t - \boldsymbol{\lambda}^{(\tau)} \omega_t^{(\tau)} - \mathbf{K} \boldsymbol{\alpha}_t^{(\tau)} \right) \right\|_2^2 \right] + \mathcal{C}. \end{aligned}$$

Sviluppando la forma quadratica interna al secondo valore atteso si trova

$$\begin{aligned} \left\| \mathbf{V}^{-\frac{1}{2}} \left(\mathbf{y}_t - \boldsymbol{\lambda}^{(\tau)} \omega_t^{(\tau)} - \mathbf{K} \boldsymbol{\alpha}_t^{(\tau)} \right) \right\|_2^2 &= \left(\mathbf{y}_t - \mathbf{K} \boldsymbol{\alpha}_t^{(\tau)} \right)^\top \mathbf{V}^{-1} \left(\mathbf{y}_t - \mathbf{K} \boldsymbol{\alpha}_t^{(\tau)} \right) \\ &\quad - 2 \omega_t^{(\tau)} \boldsymbol{\lambda}_{(\tau)}^\top \mathbf{V}^{-1} \left(\mathbf{y}_t - \mathbf{K} \boldsymbol{\alpha}_t^{(\tau)} \right) \\ &\quad + \boldsymbol{\lambda}_{(\tau)}^\top \mathbf{V}^{-1} \boldsymbol{\lambda}_{(\tau)} \left(\omega_t^{(\tau)} \right)^2. \end{aligned}$$

Andando a sostituire tale sviluppo all'interno della *variational approximation*, distribuendo i valori attesi ed eseguendo le opportune semplificazioni si ottiene

$$\begin{aligned} \log q_\varepsilon^{(\tau)}(\sigma_{\varepsilon,(\tau)}^2) &= - \left(a_{\varepsilon,(\tau)}^0 + T + \frac{n}{2}T + 1 \right) \log \left(\sigma_{\varepsilon,(\tau)}^2 \right) - \frac{b_{\varepsilon,(\tau)}^0}{\sigma_{\varepsilon,(\tau)}^2} - \frac{1}{2 \delta_{(\tau)}^2 \sigma_{\varepsilon,(\tau)}^2} \times \\ &\quad \times \sum_{t=1}^T \left\{ \mathbb{E}_\omega \left[\left(\omega_t^{(\tau)} \right)^{-1} \right] \mathbb{E}_\alpha \left[\left\| \mathbf{V}^{-\frac{1}{2}} \left(\mathbf{y}_t - \mathbf{K} \boldsymbol{\alpha}_t^{(\tau)} \right) \right\|_2^2 \right] \right. \\ &\quad \left. - 2 \boldsymbol{\lambda}_{(\tau)}^\top \mathbf{V}^{-1} \left(\mathbf{y}_t - \mathbf{K} \mathbb{E}_\alpha \left[\boldsymbol{\alpha}_t^{(\tau)} \right] \right) \right. \\ &\quad \left. + \left(\boldsymbol{\lambda}_{(\tau)}^\top \mathbf{V}^{-1} \boldsymbol{\lambda}_{(\tau)} + 2 \delta_{(\tau)}^2 \right) \mathbb{E}_\omega \left[\omega_t^{(\tau)} \right] \right\} + \mathcal{C}. \end{aligned}$$

Da questa forma estesa riconosciamo il nucleo di una distribuzione \mathcal{IG} di parametri

$$\begin{aligned} a_{\varepsilon,(\tau)}^* &= a_{\varepsilon,(\tau)}^0 + \frac{2+n}{2}T, \\ b_{\varepsilon,(\tau)}^* &= b_{\varepsilon,(\tau)}^0 + \frac{1}{2 \delta_{(\tau)}^2} \sum_{t=0}^{T-1} \left\{ \mathbb{E}_\omega \left[\left(\omega_t^{(\tau)} \right)^{-1} \right] \mathbb{E}_\alpha \left[\left\| \mathbf{V}^{-\frac{1}{2}} \left(\mathbf{y}_t - \mathbf{K} \boldsymbol{\alpha}_t^{(\tau)} \right) \right\|_2^2 \right] \right. \\ &\quad \left. - 2 \boldsymbol{\lambda}_{(\tau)}^\top \mathbf{V}^{-1} \left(\mathbf{y}_t - \mathbf{K} \mathbb{E}_\alpha \left[\boldsymbol{\alpha}_t^{(\tau)} \right] \right) + \left(\boldsymbol{\lambda}_{(\tau)}^\top \mathbf{V}^{-1} \boldsymbol{\lambda}_{(\tau)} + 2 \delta_{(\tau)}^2 \right) \mathbb{E}_\omega \left[\omega_t^{(\tau)} \right] \right\}. \end{aligned}$$

Variational approximation a posteriori per $q_{\omega_t}^{(\tau)}\left(\omega_t^{(\tau)}\right)$

$$\begin{aligned} \log q_{\omega_t}^{(\tau)}\left(\omega_t^{(\tau)}\right) &= -\frac{1}{2} \log\left(\omega_t^{(\tau)}\right) - \mathbb{E}_{\varepsilon} \left[\frac{1}{\sigma_{\varepsilon,(\tau)}^2} \right] \omega_t^{(\tau)} - \frac{1}{2 \delta_{(\tau)}^2 \omega_t^{(\tau)}} \times \\ &\quad \times \mathbb{E}_{\varepsilon, \alpha} \left[\frac{1}{\sigma_{\varepsilon,(\tau)}^2} \left\| \mathbf{V}^{-\frac{1}{2}} \left(\mathbf{y}_t - \boldsymbol{\lambda}^{(\tau)} \omega_t^{(\tau)} - \mathbf{K} \boldsymbol{\alpha}_t^{(\tau)} \right) \right\|_2^2 \right] + \mathcal{C}. \end{aligned}$$

Anche in questo caso si può sviluppare la forma quadratica e apportare le dovute semplificazioni. Andando poi a distribuire il valore atteso e accorpando tutte le quantità che non dipendono da $\omega_t^{(\tau)}$ all'interno della costante \mathcal{C} , si ottiene

$$\begin{aligned} \log q_{\omega_t}^{(\tau)}\left(\omega_t^{(\tau)}\right) &= -\frac{1}{2} \log\left(\omega_t^{(\tau)}\right) - \mathbb{E}_{\varepsilon} \left[\sigma_{\varepsilon,(\tau)}^{-2} \right] \mathbb{E}_{\alpha} \left[\left\| \mathbf{V}^{-\frac{1}{2}} \left(\mathbf{y}_t - \mathbf{K} \boldsymbol{\alpha}_t^{(\tau)} \right) \right\|_2^2 \right] \times \\ &\quad \times \frac{1}{2 \delta_{(\tau)}^2 \omega_t^{(\tau)}} - \left(\frac{\boldsymbol{\lambda}_{(\tau)}^{\top} \mathbf{V}^{-1} \boldsymbol{\lambda}_{(\tau)} + 2 \delta_{(\tau)}^2}{2 \delta_{(\tau)}^2} \right) \mathbb{E}_{\varepsilon} \left[\sigma_{\varepsilon,(\tau)}^{-2} \right] \omega_t^{(\tau)} + \mathcal{C}. \end{aligned}$$

Dalla distribuzione a posteriori condizionata si riconosce il nucleo di una legge *GIG* di parametri

$$\begin{aligned} \chi_{t,(\tau)}^* &= \frac{2-n}{2}, \\ \psi_{t,(\tau)}^* &= \delta_{(\tau)}^{-2} \left(\boldsymbol{\lambda}_{(\tau)}^{\top} \mathbf{V}^{-1} \boldsymbol{\lambda}_{(\tau)} + 2 \delta_{(\tau)}^2 \right) \mathbb{E}_{\varepsilon} \left[\sigma_{\varepsilon,(\tau)}^{-2} \right], \\ \varphi_{t,(\tau)}^* &= \delta_{(\tau)}^{-2} \mathbb{E}_{\varepsilon} \left[\sigma_{\varepsilon,(\tau)}^{-2} \right] \mathbb{E}_{\alpha} \left[\left\| \mathbf{V}^{-\frac{1}{2}} \left(\mathbf{y}_t - \mathbf{K} \boldsymbol{\alpha}_t^{(\tau)} \right) \right\|_2^2 \right]. \end{aligned}$$

Variational expectations

I parametri delle distribuzioni approssimanti dipendono tipicamente dal valore atteso variazionale di diverse quantità stocastiche, andiamo ora a sviluppare tali valori attesi e a mostrarne le proprietà.

Anzitutto andiamo a sviluppare il valore atteso della forma quadratica associata

all'equazione di transizione per gli stati:

$$\begin{aligned}
& \mathbb{E}_\alpha \left[\left\| \mathbf{W}^{-\frac{1}{2}} \left(\boldsymbol{\alpha}_{t+1}^{(\tau)} - \boldsymbol{\alpha}_t^{(\tau)} \right) \right\|_2^2 \right] \\
&= \mathbb{E}_\alpha \left[\text{Tr} \left\{ \left(\boldsymbol{\alpha}_{t+1}^{(\tau)} - \boldsymbol{\alpha}_t^{(\tau)} \right)^\top \mathbf{W}^{-1} \left(\boldsymbol{\alpha}_{t+1}^{(\tau)} - \boldsymbol{\alpha}_t^{(\tau)} \right) \right\} \right] \\
&= \text{Tr} \left\{ \mathbf{W}^{-1} \mathbb{E}_\alpha \left[\left(\boldsymbol{\alpha}_{t+1}^{(\tau)} - \boldsymbol{\alpha}_t^{(\tau)} \right) \left(\boldsymbol{\alpha}_{t+1}^{(\tau)} - \boldsymbol{\alpha}_t^{(\tau)} \right)^\top \right] \right\} \\
&= \text{Tr} \left\{ \mathbf{W}^{-1} \left(\mathbb{V}_\alpha \left[\boldsymbol{\alpha}_{t+1}^{(\tau)} \right] + \mathbb{V}_\alpha \left[\boldsymbol{\alpha}_t^{(\tau)} \right] - 2 \mathbb{C}_\alpha \left[\boldsymbol{\alpha}_{t+1}^{(\tau)}, \boldsymbol{\alpha}_t^{(\tau)} \right] \right. \right. \\
&\quad \left. \left. + \mathbb{E}_\alpha \left[\boldsymbol{\alpha}_{t+1}^{(\tau)} \right] \mathbb{E}_\alpha \left[\boldsymbol{\alpha}_{t+1}^{(\tau)} \right]^\top + \mathbb{E}_\alpha \left[\boldsymbol{\alpha}_t^{(\tau)} \right] \mathbb{E}_\alpha \left[\boldsymbol{\alpha}_t^{(\tau)} \right]^\top - 2 \mathbb{E}_\alpha \left[\boldsymbol{\alpha}_{t+1}^{(\tau)} \right] \mathbb{E}_\alpha \left[\boldsymbol{\alpha}_t^{(\tau)} \right]^\top \right) \right\} \\
&= \text{Tr} \left\{ \mathbf{W}^{-1} \left(\hat{\mathbf{P}}_{t+1}^{(\tau)} + \hat{\mathbf{P}}_t^{(\tau)} - 2 \hat{\mathbf{P}}_{t+1,t}^{(\tau)} + \hat{\mathbf{a}}_{t+1}^{(\tau)} \hat{\mathbf{a}}_{t+1}^{(\tau)\top} + \hat{\mathbf{a}}_t^{(\tau)} \hat{\mathbf{a}}_t^{(\tau)\top} - 2 \hat{\mathbf{a}}_{t+1}^{(\tau)} \hat{\mathbf{a}}_t^{(\tau)\top} \right) \right\}.
\end{aligned}$$

Quest'ultima identità permette di calcolare il valore atteso utilizzando solo le quantità fornite in *output* dal *Kalman Smoother* modificato, dove $\hat{\mathbf{a}}_t^{(\tau)}$, $\hat{\mathbf{P}}_t^{(\tau)}$ e $\hat{\mathbf{P}}_{t+1,t}^{(\tau)}$ sono rispettivamente media, varianza e covarianza del blocco di stati riferiti a τ . Il vettore $\hat{\mathbf{a}}_t^{(\tau)}$, e la matrice $\hat{\mathbf{P}}_t^{(\tau)}$ vengono restituiti direttamente dallo *smoother*, mentre la matrice di covarianza $\hat{\mathbf{P}}_{t+1,t}^{(\tau)}$ può essere facilmente derivata, come mostrato da Shumway e Stoffer 1982:

$$\begin{aligned}
\hat{\mathbf{P}}_{t+1,t} &= \mathbf{P}_{t+1|t+1} \mathbf{J}_t^\top + \mathbf{J}_{t+1} \left(\mathbf{P}_{t+2,t+1} - \mathbf{T}_t \mathbf{P}_{t+1|t+1} \right) \mathbf{J}_t^\top, \\
\mathbf{J}_{t+1} &= \mathbf{P}_{t+1|t+1} \mathbf{T}_t^\top \mathbf{P}_{t+1}^{-1}.
\end{aligned}$$

Ricordiamo che $\mathbf{P}_{t|t}$ e \mathbf{P}_{t+1} sono le matrici di varianze e covarianze per gli stati filtrati e predetti e sono definite in tabella 2.1.

Andiamo ora a studiare il valore atteso associato alla forma quadratica per l'equazione di transizione:

$$\begin{aligned}
& \mathbb{E}_\alpha \left[\left\| \mathbf{V}^{-\frac{1}{2}} \left(\mathbf{y}_t - \mathbf{K} \boldsymbol{\alpha}_t^{(\tau)} \right) \right\|_2^2 \right] \\
&= \mathbf{y}_t^\top \mathbf{V}^{-1} \mathbf{y}_t - 2 \mathbf{y}_t^\top \mathbf{V}^{-1} \mathbb{E}_\alpha \left[\boldsymbol{\alpha}_t^{(\tau)} \right] + \mathbb{E}_\alpha \left[\boldsymbol{\alpha}_t^{(\tau)\top} \mathbf{V}^{-1} \boldsymbol{\alpha}_t^{(\tau)} \right] \\
&= \mathbf{y}_t^\top \mathbf{V}^{-1} \mathbf{y}_t - 2 \mathbf{y}_t^\top \mathbf{V}^{-1} \mathbb{E}_\alpha \left[\boldsymbol{\alpha}_t^{(\tau)} \right] + \text{Tr} \left\{ \mathbf{V}^{-1} \mathbb{E}_\alpha \left[\boldsymbol{\alpha}_t^{(\tau)} \boldsymbol{\alpha}_t^{(\tau)\top} \right] \right\} \\
&= \mathbf{y}_t^\top \mathbf{V}^{-1} \mathbf{y}_t - 2 \mathbf{y}_t^\top \mathbf{V}^{-1} \mathbb{E}_\alpha \left[\boldsymbol{\alpha}_t^{(\tau)} \right] \\
&\quad + \text{Tr} \left\{ \mathbf{V}^{-1} \left(\mathbb{V}_\alpha \left[\boldsymbol{\alpha}_t^{(\tau)} \right] + \mathbb{E}_\alpha \left[\boldsymbol{\alpha}_t^{(\tau)} \right] \mathbb{E}_\alpha \left[\boldsymbol{\alpha}_t^{(\tau)} \right]^\top \right) \right\} \\
&= \mathbf{y}_t^\top \mathbf{V}^{-1} \mathbf{y}_t - 2 \mathbf{y}_t^\top \mathbf{V}^{-1} \hat{\mathbf{a}}_t^{(\tau)} + \text{Tr} \left\{ \mathbf{V}^{-1} \left(\hat{\mathbf{P}}_t^{(\tau)} + \hat{\mathbf{a}}_t^{(\tau)} \hat{\mathbf{a}}_t^{(\tau)\top} \right) \right\}
\end{aligned}$$

Anche in questo caso tutte le quantità di interesse necessarie al calcolo vengono fornite dai dati, dalle matrici della forma *State Space* e dal *Kalman Smoother* modificato.

Gli ultimi valori attesi necessari sono direttamente calcolabili attraverso l'applicazione delle proprietà delle distribuzioni \mathcal{GIG} e \mathcal{IG} e risultano pari a

$$\begin{aligned}\mathbb{E}_\omega \left[(\omega_t^{(\tau)})^{-1} \right] &= \frac{\sqrt{\varphi_{t,(\tau)}^*} K(\chi_{t,(\tau)}^* + 1) \left(\sqrt{\psi_{t,(\tau)}^* \varphi_{t,(\tau)}^*} \right)}{\sqrt{\varphi_{t,(\tau)}^*} K(\chi_{t,(\tau)}^*) \left(\sqrt{\psi_{t,(\tau)}^* \varphi_{t,(\tau)}^*} \right)} - 2 \frac{\chi_{t,(\tau)}^*}{\varphi_{t,(\tau)}^*}, \\ \mathbb{E}_\omega \left[\omega_t^{(\tau)} \right] &= \frac{\sqrt{\varphi_{t,(\tau)}^*} K(\chi_{t,(\tau)}^* + 1) \left(\sqrt{\psi_{t,(\tau)}^* \varphi_{t,(\tau)}^*} \right)}{\sqrt{\varphi_{t,(\tau)}^*} K(\chi_{t,(\tau)}^*) \left(\sqrt{\psi_{t,(\tau)}^* \varphi_{t,(\tau)}^*} \right)}, \\ \mathbb{E}_\varepsilon \left[\sigma_{\varepsilon,(\tau)}^{-2} \right] &= \frac{a_{\varepsilon,(\tau)}^*}{b_{\varepsilon,(\tau)}^*}.\end{aligned}$$

Appendice B

Codice Python

In questa appendice vengono riportati alcuni dei principali codici sviluppati in Python 3.6.5 per implementare i modelli e gli algoritmi descritti nel corso della tesi.

B.1 Algoritmi di filtraggio, lisciamiento e simulazione

Codice B.1: *Kalman filter*

```
# YY : matrice delle osservazioni
# c : intercetta per le osservazioni
# ZZ : matrice di misura
# HH : matrice di varianza dei disturbi di misura
# d : intercetta per gli stati
# TT : matrice di transizione
# RR : matrice di selezione
# QQ : matrice di varianza dei disturbi di stato
# a : media iniziale degli stati
# PP : matrice di varianze iniziale degli stati
# h : scalare/vettore di correzione in media quantilica
# WW : matrice di fattori latenti esponenziali

def SSM_DQ_KalmanFilter(YY, d, ZZ, HH, c, TT, RR, QQ, a, PP, h, WW):

    # Librerie
    import numpy as np
    from vec import vech
```

```

# Dimensioni
dimT = YY.shape[0]
dimN = ZZ.shape[0]
dimS = TT.shape[1]

# Verosimiglianza
LogFF = 0
SumSQ = 0
dimNP = 0

# Allocazione della memoria
yt = np.zeros(( dimT, dimN ))
vt = np.zeros(( dimT, dimN ))
FFt = np.zeros(( dimT, dimN * ( dimN+1 ) // 2 ))
IFFt = np.zeros(( dimT, dimN * ( dimN+1 ) // 2 ))
KKt = np.zeros(( dimT, dimS * dimN ))
at = np.zeros(( dimT, dimS ))
PPt = np.zeros(( dimT, dimS * ( dimS+1 ) // 2 ))
att = np.zeros(( dimT, dimS ))
PPtt = np.zeros(( dimT, dimS * ( dimS+1 ) // 2 ))

# Ciclo di filtraggio
for t in np.arange(0, dimT):

    # Archiviazione degli stati predetti in memoria
    at[t,:] = a.flatten()
    PPt[t,:] = vech(PP)

    # Passo di selezione del fattore aumentato
    w = WW[t,:].flatten()
    ww = np.diag( np.sqrt( w ) )

    # Passo di previsione delle osservazioni
    y = d + (h * w) + (ZZ @ a)

    # Passo di aggiornamento delle innovzioni
    v = Y[t,:] - y
    FF = ZZ @ PP @ ZZ.T + ww @ HH @ ww
    IFF = np.linalg.solve(FF, np.eye(dimN))
    IFF_ZZ = np.linalg.solve(FF, ZZ)

```

```

# Passo di aggiornamento del kalman gain
KK = PP @ ZZ.T @ IFF

# Passo di aggiornamento degli stati filtrati
a = a + KK @ v
PP = PP - PP @ (ZZ.T @ IFF_ZZ) @ PP

# Archiviazione in memoria
yt[t,:] = y
vt[t,:] = v
FFt[t,:] = vech(FF )
IFFt[t,:] = vech(IFF)
att[t,:] = a
PPtt[t,:] = vech(PP )
KKt[t,:] = KK.flatten()

# Passo di aggiornamento della verosimiglianza
detFF = np.linalg.det(FF)
LogFF = LogFF + np.log(detFF)
SumSQ = SumSQ + v @ IFF @ v
dimNP = dimNP + dimN

# Passo di previsione degli stati
a = c + TT @ a
PP = RR @ QQ @ RR.T + TT @ PP @ TT.T

# Valutazione della Verosimiglianza completa
LogL = - 0.5 * (dimNP * np.log(2 * np.pi) + LogFF + SumSQ)
Sigma = SumSQ / dimNP

# Output
return({'yt' : yt , 'KKt' : KKt ,
        'vt' : vt , 'FFt' : FFt , 'IFFt': IFFt,
        'at' : at , 'PPt' : PPt ,
        'att': att, 'PPtt': PPtt,
        'LogL': LogL, 'Sigma': Sigma})

def vech(mX):

```

```
# Librerie
import numpy as np

# Dimensioni
n = mX.shape[0]
k = n * (n + 1) // 2

# Allocazione
vY = np.zeros(k)

# Vettorizzazione
cnt = 0
for i in np.arange(0,n):
    vY[cnt:(cnt+n-i)] = mX[i, i:n].T
    cnt = cnt + (n - i)

# Output
return(vY)

def unvech(vP, cN):

    # Librerie
    import numpy as np

    # Allocazione
    mC = np.zeros(cN * cN).reshape(cN, cN)

    # Unvech
    cnt = 0
    for c in np.arange(0, cN):
        for r in np.arange(c, cN):
            mC[r,c] = vP[cnt]
            mC[c,r] = mC[r,c]
            cnt = cnt + 1

    # Output
    return(mC)
```

```

def SSM_DQ_DistSmoother(Y, d, ZZ, HH, c, TT, RR, QQ, a, PP, h, WW):

    # Librerie
    from My_DQ_KalmanFilter import SSM_DQ_KalmanFilter
    from My_DQ_DistSmoother import SSM_DQ_DistSmoother_RECURSION

    # Filtraggio
    Filter = SSM_DQ_KalmanFilter(Y, d, ZZ, HH, c, TT, RR, QQ, a, PP, h, WW)
    vt, FFt, KKt = Filter['vt'], Filter['FFt'], Filter['KKt']

    # Disturbance Smoothing Recursion
    rt = SSM_DQ_DistSmoother_RECURSION(Y, ZZ, TT, vt, FFt, KKt)['rt']

    # Output
    return(rt)

def SSM_DQ_DistSmoother_RECURSION(Y, ZZ, TT, vt, FFt, KKt):

    # Librerie
    import numpy as np
    from vec import vech, unvech

    # Dimensioni
    dimT = Y.shape[0]
    dimN = ZZ.shape[0]
    dimS = ZZ.shape[1]

    # Allocazione delle matrici aumentate
    Phi = np.zeros(( dimS+dimN, dimS ))
    Phi[:dimS,:dimS] = TT
    Phi[dimS:,:dimS] = ZZ

    # Allocazione
    r = np.zeros( dimS )
    NN = np.zeros(( dimS , dimS ))
    r_S = np.zeros( dimS + dimN )
    NN_S = np.zeros(( dimS + dimN, dimS + dimN ))
    rt = np.zeros(( dimT+1, dimS ))
    NNt = np.zeros(( dimT+1, dimS * ( dimS+1 ) // 2 ))
    et = np.zeros(( dimT , dimN ))

```

```

EEt = np.zeros(( dimT , dimN * ( dimN+1 ) // 2 ))

rt[ dimT,:] = r
NNt[dimT,:] = vech(NN)

# Ciclo di lisciamiento dei disturbi
for t in np.arange(dimT-1, -1, -1):

    # Passo di aggiornamento delle matrici ausiliarie
    KK    = KKt[t,:].reshape(dimS, dimN)
    FF    = unvec(FFt[t,:], dimN)
    IFF   = np.linalg.solve(FF, np.eye(dimN, dimN))
    IFF_v = np.linalg.solve(FF, vt[t,:])

    # Passo di lisciamiento dei disturbi
    e     = IFF_v - KK.T @ TT.T @ r
    EE    = IFF + KK.T @ TT.T @ NN @ TT @ KK

    et[ t,:] = e.flatten()
    EEt[t,:] = vech(EE)

    # Passo di aggiornamento dei disturbi
    r_S[:dimS] = r
    r_S[dimS:] = e

    NN_S[:dimS,:dimS] = NN
    NN_S[:dimS,dimS:] = -NN @ TT @ KK
    NN_S[dimS:,:dimS] = -KK.T @ TT @ NN
    NN_S[dimS:,dimS:] = EE

    r = Phi.T @ r_S
    NN = Phi.T @ NN_S @ Phi

    rt[ t,:] = r
    NNt[t,:] = vech(NN)

# Output
return({'rt': rt, 'NNt': NNt})

```

Codice B.3: *Disturbance simulation smoother*


```
def SSM_DQ_DistSimSmoother(Y, d, ZZ, HH, c, TT, RR, QQ, a, PP, h, WW):

    # Librerie
    import numpy as np
    import numpy.linalg as nla
    import numpy.random as nrd
    from My_DQ_DistSmoother import SSM_DQ_DistSmoother

    # Dimensioni
    dimT = Y.shape[0]
    dimN = ZZ.shape[0]
    dimS = ZZ.shape[1]
    dimQ = QQ.shape[0]

    # Matrici aumentate
    Gamma = np.zeros(( dimS+dimN, dimQ+dimN ))
    Gamma[:dimS, :dimQ] = RR
    Gamma[ dimS:, dimQ:] = np.eye( dimN )

    Omega = np.zeros(( dimQ+dimN, dimQ+dimN ))
    Omega[:dimQ, :dimQ] = QQ
    Omega[ dimQ:, dimQ:] = HH

    Phi = np.zeros(( dimS+dimN, dimS ))
    Phi[:dimS, :] = TT
    Phi[ dimS:, :] = ZZ

    Delta = np.zeros( dimS+dimN )
    Delta[:dimS] = c
    Delta[ dimS:] = d

    k = 0
    Ind = []
    for i in np.arange(0, dimS+dimN):
        if (sum(Gamma[i, :]) == 1):
            Ind.append( int(i) )
            k = k+1
```

```

# Correggo un'eventuale inizializzazione diffusa
PP_0 = PP
if (np.diag(PP) == 1000).any():
    PP_0 = np.eye( dimS )

# Allocazione
UU          = np.zeros(( dimT , dimS+dimN ))
DD          = np.zeros(( dimT+1, dimS+dimN ))
DD[0,:dimS ] = a + nla.cholesky(PP_0).T @ nrd.normal(size=dimS)
DD[0, dimS:] = np.zeros( dimN )

# Ciclo di aggiornamento delle matrici tempo-dipendenti
for t in np.arange(1,dimT+1):

    # Estrazione del fattore latente eteroschedastico
    sqrtWW = np.diag( np.sqrt( WW[t-1,:] ) )

    # Definizione delle matrici tempo dipendenti
    Delta[dimS:]      = d + ( h * WW[t-1,:] )
    Omega[dimQ:,dimQ:] = sqrtWW @ HH @ sqrtWW

    # Simulazione delle innovazioni del modello
    GG          = nla.cholesky(Omega).T
    UU[t-1,Ind] = GG @ nrd.normal(size=dimQ+dimN)

    # Estrazione del segnale e degli stati simulati
    DD[t ,: ] = Delta + Phi @ DD[t-1,:dimS] + UU[t-1,:]

# Osservazioni simulate
Y_S = DD[1:,dimS:(dimS+dimN)]

# Lisciamento dei disturbi
rt   = SSM_DQ_DistSmoother(Y , d, ZZ, HH, c, TT, RR, QQ, a, PP, h, WW)
rt_S = SSM_DQ_DistSmoother(Y_S, d, ZZ, HH, c, TT, RR, QQ, a, PP, h, WW)

# State Simulation Smoother
Simulation = SSM_DQ_DistSimSmoother_RECURSION(DD, rt, rt_S, TT, PP, RR, QQ)
at_Sim, st_S = Simulation['at_Sim'], Simulation['st_S']

# Output

```

```

return({'at_Sim': at_Sim, 'st_S': st_S, 'rt_S': rt_S})

def SSM_DQ_DistSimSmoother_RECURSION(DD, rt, rt_S, TT, PP, RR, QQ):

    # Librerie
    import numpy as np

    # Dimensioni
    dimT = DD.shape[0] - 1
    dimS = TT.shape[0]

    # Correzione in media dei disturbi
    rt_S_ = rt - rt_S

    # Stati simulati
    st_S = np.zeros(( dimT+1, dimS ))
    st_S[0,:] = (PP @ rt_S_[0,:]).flatten()

    # Ciclo di aggiornamento degli stati simulati
    for t in np.arange(1, dimT+1):
        st_S[t,:] = (TT @ st_S[t-1,:]) + (RR @ QQ @ RR.T @ rt_S_[t,:])

    # Correzione in media degli stati simulati
    at_Sim = DD[:, :dimS] + st_S

    # Output
    return({'at_Sim': at_Sim, 'st_S': st_S})

```

B.2 Algoritmi per il modello 4.3.4

Codice B.4: Matrici della rappresentazione *State Space* per il modello 4.3.4

```

def SpatioTemporalModel( S, y, CC, KK, InvKK,
                        Sigma2_Alpha, Sigma2_Psi, Sigma2_Eps ):

    # Librerie
    import numpy as np

```

```

# Dimensioni
dimN = S.shape[ 0 ]
dimK = KK.shape[ 1 ]
dimS = dimK + 2

# Frequenza fondamentale
FreqPsi = 2.0 * np.pi / 365.0

# Equazione di misura
d          = y * np.ones( dimN )
ZZ         = np.zeros(( dimN, dimS ))
ZZ[ : , 0 ] = np.ones( dimN )
ZZ[ : , 2:] = KK
HH         = Sigma2_Eps * CC

# Equazione di transizione
c          = np.zeros( dimS )
TT         = np.eye( dimS )
TT[ 0 , 0 ] = +np.cos( FreqPsi )
TT[ 0 , 1 ] = +np.sin( FreqPsi )
TT[ 1 , 0 ] = -np.sin( FreqPsi )
TT[ 1 , 1 ] = +np.cos( FreqPsi )
TT[ 2:, 2:] = 1.0 * TT[ 2:, 2:]
RR         = np.eye( dimS )
QQ         = np.zeros(( dimS, dimS ))
QQ[0:2,0:2] = Sigma2_Psi * np.eye( 2 )
QQ[2: ,2: ] = Sigma2_Alpha * InvKK

# Distribuzione iniziale
a          = np.zeros( dimS )
PP         = np.eye( dimS )
PP[0:2,0:2] = 1000 * np.eye( 2 )
PP[2: ,2: ] = 1000 * np.eye( dimN )

# Output
return({ 'd': d, 'ZZ': ZZ, 'HH': HH,
         'c': c, 'TT': TT, 'QQ': QQ, 'RR': RR,
         'a': a, 'PP': PP, 'y' : y , 'KK': KK })

```

Codice B.5: Calcolo della matrice *kernel*

```

def KernelMatrixBuilder( mS, Span ):

    # Librerie
    import numpy          as np
    import scipy.spatial as sp

    # Vettore di posizione
    vM      = np.zeros( 2 )
    vM[ 0 ] = np.mean( mS[ : , 0 ] )
    vM[ 1 ] = np.mean( mS[ : , 1 ] )

    # Deviazioni Standard Inverse
    vV      = np.zeros( 2 )
    vV[ 0 ] = 1.0 / np.sqrt( np.var( mS[ : , 0 ] ) )
    vV[ 1 ] = 1.0 / np.sqrt( np.var( mS[ : , 1 ] ) )
    dV      = vV[ 0 ] * vV[ 1 ] / Span

    # Standardizzazione
    mS[ : , 0 ] = (mS[ : , 0 ] - vM[ 0 ]) / vV[ 0 ]
    mS[ : , 1 ] = (mS[ : , 1 ] - vM[ 1 ]) / vV[ 1 ]

    # Matrice di distanze
    mD = sp.distance_matrix( mS, mS )
    mD = np.array( mD, dtype=np.float32 ) ** 2

    # Matrice Kernel
    mK = dV * np.exp( - 0.5 * mD / Span )

    # Output
    return( mK )

```

Codice B.6: Calcolo della matrice di correlazione spaziale

```

def CorrMatrixBuilder( mS, Scale ):

    # Librerie
    import numpy          as np
    import scipy.spatial as sp

```

```

# Matrice di distanza
mD = sp.distance_matrix( mS, mS )
mD = np.array( mD, dtype=np.float32 )

# Matrice di correlazione
mC = np.exp( - Scale * mD )

# output
return( mC )

```

Codice B.7: *Full conditional* collassata per σ_{ϵ}^2 nel modello 4.3.4

```

def FC_Sigma2_Eps( Data, StSim, PrS_Eps, d, ZZ, InvCC ):

    # Librerie
    import numpy as np
    import numpy.random as nrd

    # Dimensioni
    dimT = Data.shape[ 0 ]
    dimN = Data.shape[ 1 ]

    # Somma dei quadrati
    SumSq = 0
    Pred = d + ( StSim[1:, :] @ ZZ.T )
    for t in np.arange( 0, dimT ):
        Diff[i] = Data[t, :] - Pred[t, :]
        SumSq = SumSq + ( Diff @ InvCC @ Diff )

    # Valutazione dei parametri delle full conditional
    PrPostS_Eps = np.zeros( 2 )
    PrPostS_Eps[ 0 ] = PrS_Eps[ 0 ] + 0.5 * dimT * dimN
    PrPostS_Eps[ 1 ] = PrS_Eps[ 1 ] + 0.5 * SumSq

    # Simulazione dalle distribuzioni full conditional
    Sigma2_Eps = 1.0 / nrd.gamma( PrPostS_Eps[ 0 ], 1.0 / PrPostS_Eps[ 1 ], 1 )

    # Output
    return( Sigma2_Eps )

```

Codice B.8: Riparametrizzazione *Inverse Gamma*

```
def InvGammaMoments( Mu, Sigma2 ):

    # Riparametrizzazione
    Alpha = (Mu * Mu + 2.0 * Sigma2) / Sigma2
    Beta  = Mu * (Mu * Mu + Sigma2 ) / Sigma2

    # Output
    return(Alpha, Beta)
```

Codice B.9: *Gibbs sampling* per il modello 4.3.4

```
def SSM_ST_MCMC( Y, S, NIter, NCheck, Span, Scale, PrS_Alpha, PrS_Psi, PrS_Eps ):

    # Librerie
    import numpy as np
    import numpy.linalg as nla
    import numpy.random as nrd

    # generatore delle matrici del modello
    from My_ST_8 import KernelMatrixBuilder, \
        CorrMatrixBuilder , \
        SpatioTemporalModel

    # generatori delle distribuzioni full conditional
    from My_ST_8 import FC_Sigma2_Eps

    # filtro, smoother e simulation smoother
    from My_DistSimSmoother import SSM_DistSimSmoother

    # Dimensioni
    dimT = Y.shape[ 0 ]
    dimN = Y.shape[ 1 ]
    dimS = dimN + 2

    # Allocazione della memoria
    yPredSim  = np.zeros(( NIter, dimT , dimN ))
```

```

StateSim = np.zeros(( NIter, dimT+1, dimS ))
Sigma2_Eps = np.zeros( NIter )

# Calcolo delle matrici kernel e di correlazione
y = np.mean( Y )

# matrice di correlazione
CC = CorrMatrixBuilder( S, Scale )
InvCC = nla.solve( CC, np.eye( dimN ) )

# matrice kernel
KK = KernelMatrixBuilder( S, Span )
DD = np.diag( np.diag( KK ) ** -0.5 )
InvKK = nla.solve( DD @ KK @ DD, np.eye( dimN ) )

# Simulazione dalle distribuzioni iniziali
sAlpha = PrS_Alpha
sPsi = PrS_Psi
sEps = 1.0 / nrd.gamma( PrS_Eps[ 0 ], 1.0 / PrS_Eps[ 1 ], size=1 )

# Modello State Space condizionato ai parametri iniziali
SSM = SpatioTemporalModel( S, y, CC, KK, InvKK, sAlpha, sPsi, sEps )

# Simulazione del processo latente iniziale
StSim = SSM_DistSimSmoother( Y, SSM[ 'd' ], SSM[ 'ZZ' ], SSM[ 'HH' ],
                             SSM[ 'c' ], SSM[ 'TT' ], SSM[ 'RR' ],
                             SSM[ 'QQ' ], SSM[ 'a' ], SSM[ 'PP' ])[ 'at_Sim' ]

# Allocazione dei valori iniziali
yPredSim[0,:,:] = SSM[ 'd' ] + ( StSim[1,:,:] @ SSM[ 'ZZ' ].T )
StateSim[0,:,:] = StSim
Sigma2_Eps[ 0 ] = sEps

# Gibbs Sampling
for i in np.arange( 1, NIter ):

    # STEP 1: Simulazione dalle Full Conditional
    sEps = FC_Sigma2_Eps( Y, StSim, PrS_Eps, SSM[ 'd' ], SSM[ 'ZZ' ], InvCC )

    # STEP 2: Costruzione delle matrici del modello

```



```

SSM = SpatioTemporalModel( S, y, CC, KK, InvKK, sAlpha, sPsi, sEps )

# STEP 3: Simulation Smoothing del processo latente
StSim = SSM_DistSimSmoother( Y, SSM[ 'd' ], SSM[ 'ZZ' ], SSM[ 'HH' ],
                             SSM[ 'c' ], SSM[ 'TT' ], SSM[ 'RR' ],
                             SSM[ 'QQ' ], SSM[ 'a' ], SSM[ 'PP' ]) [ 'at_Sim'
                             ]

# STEP 4: Allocazione in memoria
yPredSim[i,:,:] = SSM[ 'd' ] + ( StSim[1,:,:] @ SSM[ 'ZZ' ].T )
StateSim[i,:,:] = StSim
Sigma2_Eps[ i ] = sEps

# STEP 5: Stampa a schermo
if ( i % NCheck == 0 ):
    print('::::::::::::::::::::::::::::::::::::::::::::::::::\n')
    print('>>> Gibbs Sampling – Iteration: ', str(i), '\n')

# Output
return( { 'yPredSim' : yPredSim ,
          'StateSim' : StateSim ,
          'Sigma2_Eps': Sigma2_Eps } )

```

Codice B.10: Esempio di simulazione per il modello 4.3.4

```

# Librerie
import os
import pandas as pd
import numpy as np

# Directory
os.chdir("C:/Users/Cristian/Dropbox/Tesi/Code_Python/ST_Model_CorRes")

# Dati
Y = pd.read_csv('wind.txt', sep=' ').values[:,3:]
S = pd.read_csv('wind.siti.txt', sep=' ').values[:,1:]

# Funzioni di simulazione
from My_DQ_GammaPar import InvGammaInvMoments
from My_SSM_ST_MCMC import SSM_ST_MCMC

```

```

# Iperparametri
PrS_Alpha = 1e-02
PrS_Psi   = 1e-03
PrS_Eps   = InvGammaInvMoments( 4e-01, 1e+02 )

# Simulazione per Tau = 0.5
Scale = 0.17
Span  = 1.0
NIter = 1200
Burn  = 200
Sim   = SSM_ST_MCMC( Y, S, NIter, Span, Scale,
                    PrS_Alpha, PrS_Psi, PrS_Eps )

```

B.3 Algoritmi per il modello 4.4.1

Codice B.11: Matrici della rappresentazione *State Space* per il modello 4.4.1

```

def DQ_ST_ModelBuilder( S, y, KK, InvKK, Tau,
                      Sigma2_Alpha, Sigma2_Psi, Sigma2_Eps ):

    # Librerie
    import numpy as np

    # Dimensioni
    dimN = S.shape[ 0 ]
    dimK = KK.shape[ 1 ]
    dimS = dimK + 2

    # Frequenza fondamentale
    FreqPsi = 2.0 * np.pi / 365.0

    # Coefficienti quantilici
    Delta2 = 2.0 / ( Tau*(1.0-Tau) )

    # Equazione di misura
    d      = y * np.ones( dimN )
    ZZ     = np.zeros(( dimN, dimS ))

```

```

ZZ[ : , 0 ] = np.ones( dimN )
ZZ[ : , 2:] = KK
HH          = Delta2 * Sigma2_Eps * np.eye( dimN )

# Equazione di transizione
c          = np.zeros( dimS )
TT        = np.eye( dimS )
TT[ 0 , 0 ] = +np.cos( FreqPsi )
TT[ 0 , 1 ] = +np.sin( FreqPsi )
TT[ 1 , 0 ] = -np.sin( FreqPsi )
TT[ 1 , 1 ] = +np.cos( FreqPsi )
RR        = np.eye( dimS )
QQ        = np.zeros(( dimS, dimS ))
QQ[:2 ,:2 ] = np.diag( Sigma2_Psi * np.ones( 2 ) )
QQ[ 2:, 2:] = Sigma2_Alpha * InvKK

# Distribuzione iniziale
a          = np.zeros( dimS )
PP        = np.eye( dimS )
PP[:2 ,:2 ] = 1000 * np.eye( 2 )
PP[ 2:, 2:] = 1000 * np.eye( dimK )

# Output
return({ 'd': d, 'ZZ': ZZ, 'HH': HH,
        'c': c, 'TT': TT, 'QQ': QQ, 'RR': RR,
        'a': a, 'PP': PP, 'y' : y , 'KK': KK })

```

Codice B.12: Generatore di valori da *Inverse Gaussian*

```

def InvGaussRnd_MATRIX( Mu, Lambda ):

# Librerie
import numpy as np
import numpy.random as rnd

# Generazione
dT      = Mu.shape[0]
dN      = Mu.shape[1]
vC      = Mu * rnd.chisquare( 1, dT*dN ).reshape( dT, dN )
vA      = np.sqrt( 4.0 * Lambda * vC + vC * vC )

```

```

vR      = ( Mu / ( 2.0 * Lambda ) ) * ( 2.0 * Lambda + vC - vA )
vU      = rnd.uniform( 0, 1, dT*dN ).reshape( dT, dN )
vI      = ( vU * ( Mu + vR ) > Mu )
vM      = np.zeros(( dT, dN ))
vM[ vI ] = Mu[ vI ]
vR[ vI ] = ( vM * vM )[ vI ] / vR[ vI ]

# Output
return( vR )

```

Codice B.13: Funzione di perdita quantilica

```

def QuantileLossFunction( vX, Tau ):

    # Quantile Loss
    Loss = Tau * vX - ( vX * ( vX < 0.0 ) )

    # Output
    return( Loss )

```

Codice B.14: *Full conditional* collassata per σ_{ϵ}^2 nel modello 4.4.1

```

def FC_Sigma2_Eps_IG( Data, StSim, Tau, PrS_Eps, d, ZZ ):

    # Librerie
    import numpy as np
    import numpy.random as nrd

    # funzione di perdita ALD
    from My_DQ_LossFunction import QuantileLossFunction

    # Dimensioni
    dimT = Data.shape[ 0 ]
    dimN = Data.shape[ 1 ]

    # Somma delle Quantile Loss
    SumLoss = 0
    Pred    = d + ( StSim[1:,:] @ ZZ.T )
    for t in np.arange( 0, dimT ):

```

```

Diff      = Data[t,:] - Pred[t,:]
SumLoss   = SumLoss + sum( QuantileLossFunction( Diff, Tau ) )

# Valutazione dei parametri delle full conditional
PrPostS_Eps      = np.zeros( 2 )
PrPostS_Eps[ 0 ] = PrS_Eps[ 0 ] + dimT * dimN
PrPostS_Eps[ 1 ] = PrS_Eps[ 1 ] + 0.5 * SumLoss

# Simulazione dalle distribuzioni full conditional
Sigma2_Eps = 1.0 / nrd.gamma( PrPostS_Eps[ 0 ], 1.0 / PrPostS_Eps[ 1 ], 1 )

# Output
return( Sigma2_Eps )

```

Codice B.15: *Full conditional* per il fattore latente ω nel modello 4.4.1

```

def FC_LatentFactor_IN( Data, StSim, Tau, Sigma2_Eps, d, ZZ ):

# Librerie
import numpy as np
from My_DQ_InvGaussRnd import InvGaussRnd_MATRIX

# Dimensioni
dimN = Data.shape[ 1 ]

# Coefficienti quantilici
Delta2 = 2.0 / ( Tau*(1.0-Tau) )
Lambda2 = ( ( 1-2.0*Tau ) / ( Tau*(1.0-Tau) ) ) ** 2

# Calcolo degli scarti quadratici
Sq = ( Data - d - StSim[ 1:, : ] @ ZZ.T ) ** 2

# Valutazione dei parametri delle full conditional
PrPostM_Omega = np.sqrt( ( Lambda2 + 2 * Delta2 ) / Sq )
PrPostL_Omega = ( Lambda2 + 2 * Delta2 ) / ( Delta2 * Sigma2_Eps )

# Simulazione dalla distribuzione full conditional
Omega = 1.0 / InvGaussRnd_MATRIX( PrPostM_Omega, PrPostL_Omega )

# Output

```

```
return( Omega )
```

Codice B.16: *Partially collapsed Gibbs sampling with data augmentation* per il modello 4.4.1

```
def IDQ_ST_MCMC( Y, S, NIter, Span, Tau, PrS_Alpha, PrS_Psi, PrS_Eps ):

    # Librerie
    import numpy as np
    import numpy.linalg as nla
    import numpy.random as nrd

    # generatori delle matrici del modello
    from My_DQ_DGK_ModelBuilder import KernelMatrixBuilder, \
        DQ_ST_ModelBuilder

    # generatori delle distribuzioni full conditional
    from My_DQ_DGK_FullConditional import FC_LatentFactor_IN, \
        FC_Sigma2_Eps_IG

    # filtro, smoother e simulation smoother
    from My_DQ_DistSimSmoother import SSM_DQ_DistSimSmoother

    # Dimensioni
    dimT = Y.shape[ 0 ]
    dimN = Y.shape[ 1 ]
    dimS = dimN + 2

    # Coefficienti quantilici
    sLambda = ( 1-2.0*Tau ) / ( Tau*(1.0-Tau) )

    # Allocazione della memoria
    QPredSim      = np.zeros(( NIter, dimT , dimN ))
    StateSim      = np.zeros(( NIter, dimT+1, dimS ))
    Omega         = np.zeros(( NIter, dimT , dimN ))
    Sigma2_Alpha = np.zeros( NIter )
    Sigma2_Psi   = np.zeros( NIter )
    Sigma2_Eps   = np.zeros( NIter )
```

```

# Calcolo della matrice kernel
y      = np.mean( Y )
KK     = KernelMatrixBuilder( S, Span )
InvKK  = nla.solve( KK, np.eye( dimN ) )

# Simulazione dalle distribuzioni iniziali e del fattore latente
sAlpha = PrS_Alpha[ 1 ] / ( PrS_Alpha[ 0 ]-1 )
sPsi   = PrS_Psi[ 1 ]   / ( PrS_Psi[ 0 ] -1 )
sEps   = 1.0 / nrd.gamma( PrS_Eps[ 0 ] , 1.0 / PrS_Eps[ 1 ] , size=1 )
sOmega = nrd.exponential( sEps, dimT*dimN ).reshape( dimT, dimN )

# Modello State Space condizionato ai parametri iniziali
SSM = DQ_ST_ModelBuilder( S, y, KK, InvKK, Tau, sAlpha, sPsi, sEps )

# Simulazione del processo latente iniziale
StSim = SSM_DQ_DistSimSmoother( Y, SSM['d' ], SSM['ZZ'], SSM['HH'],
                                SSM['c' ], SSM['TT'], SSM['RR'],
                                SSM['QQ'], SSM['a' ], SSM['PP'],
                                sLambda , sOmega )['at_Sim']

# Allocazione dei valori iniziali
QPredSim[0,:,:] = SSM['d'] + ( StSim[1,:,:] @ SSM['ZZ'].T)
StateSim[0,:,:] = StSim
Omega[0,:,:]    = sOmega
Sigma2_Eps[0]   = sEps

# Gibbs Sampling
for i in np.arange(1,NIter):

    # STEP 1: Collapsed Full Conditional per Sigma2_Eps
    sEps = FC_Sigma2_Eps_IG( Y, StSim, Tau, PrS_Eps, SSM['d'], SSM['ZZ'] )

    # STEP 2: Full Conditional per Omega
    sOmega = FC_LatentFactor_IN( Y, StSim, Tau, sEps, SSM['d'], SSM['ZZ'] )

    # STEP 3: Costruzione delle matrici del modello
    SSM = DQ_ST_ModelBuilder( S, y, KK, InvKK, Tau, sAlpha, sPsi, sEps )

    # STEP 4: Simulation Smoothing del processo latente
    StSim = SSM_DQ_DistSimSmoother( Y, SSM['d' ], SSM['ZZ'], SSM['HH'],

```

```

SSM[ 'c' ], SSM[ 'TT' ], SSM[ 'RR' ],
SSM[ 'QQ' ], SSM[ 'a' ], SSM[ 'PP' ],
sLambda , sOmega ) [ 'at_Sim' ]

# STEP 5: Allocazione in memoria
QPredSim[i,:,:] = SSM[ 'd' ] + ( StSim[1,:,:] @ SSM[ 'ZZ' ].T )
StateSim[i,:,:] = StSim
Omega[i,:,:] = sOmega
Sigma2_Eps[i] = sEps

# STEP 7: Stampa a schermo
if ( i % 50 == 0 ):
    print('::::::::::::::::::::::::::::::::::::::::::::::::::\n')
    print('>>> Gibbs Sampler with Data Augmentation')
    print('>>> Iteration : ', str(i), '\n')
    print('>>> Sigma2 : ', str(sEps), '\n')

# Output
return( { 'QPredSim' : QPredSim , 'StateSim': StateSim,
         'Sigma2_Eps': Sigma2_Eps, 'Omega' : Omega  })

```

Codice B.17: Esempio di simulazione per il modello 4.4.1

```

# Librerie
import os
import pandas as pd
import numpy as np

# Directory
os.chdir("C:/Users/Cristian/Dropbox/Tesi/Code.Python/DQ.Model.IndRes")

# Dati
Y = pd.read_csv('wind.txt', sep=' ').values[:,3:]
S = pd.read_csv('wind.siti.txt', sep=' ').values[:,1:]

# Funzioni di simulazione
from My_DQ_GammaPar import InvGammaInvMoments
from My_DQ_DGK_MCMC import IDQ_ST_MCMC

# Iperparametri

```



```
PrS_Alpha = 1e-02
PrS_Psi   = 1e-03
PrS_Eps   = InvGammaInvMoments( 4e-01, 1e+02 )

# Simulazione per Tau = 0.5
Tau       = 0.5
Span      = 1.0
NIter     = 1200
Burn      = 200
Sim       = DQ_SpatioTemporal_MCMC( Y, S, NIter, Span, Tau,
                                     PrS_Alpha, PrS_Psi, PrS_Eps )
```

B.4 Algoritmi per il modello 4.5.1

Codice B.18: Matrici della rappresentazione *State Space* per il modello 4.5.1

```
def DQ_ST_ModelBuilder( S, y, CC, KK, InvKK, Tau,
                       Sigma2_Alpha, Sigma2_Psi, Sigma2_Eps ):

    # Librerie
    import numpy as np

    # Dimensioni
    dimN = S.shape[ 0 ]
    dimK = KK.shape[ 1 ]
    dimS = dimK + 2

    # Frequenza fondamentale
    FreqPsi = 2.0 * np.pi / 365.0

    # Coefficienti quantilici
    Delta2 = 2.0 / ( Tau*(1.0-Tau) )

    # Equazione di misura
    d          = y * np.ones( dimN )
    ZZ         = np.zeros(( dimN, dimS ))
    ZZ[ : , 0 ] = np.ones( dimN )
    ZZ[ : , 2:] = KK
```

```

HH          = Delta2 * Sigma2_Eps * CC

# Equazione di transizione
c           = np.zeros( dimS )
TT         = np.eye( dimS )
TT[ 0 , 0 ] = +np.cos( FreqPsi )
TT[ 0 , 1 ] = +np.sin( FreqPsi )
TT[ 1 , 0 ] = -np.sin( FreqPsi )
TT[ 1 , 1 ] = +np.cos( FreqPsi )
RR         = np.eye( dimS )
QQ         = np.zeros(( dimS, dimS ))
QQ[:2 ,:2 ] = np.diag( Sigma2_Psi * np.ones( 2 ) )
QQ[ 2:, 2:] = Sigma2_Alpha * InvKK

# Distribuzione iniziale
a           = np.zeros( dimS )
PP         = np.eye( dimS )
PP[:2 ,:2 ] = 1000 * np.eye( 2 )
PP[ 2:, 2:] = 1000 * np.eye( dimK )

# Output
return({ 'd': d, 'ZZ': ZZ, 'HH': HH,
        'c': c, 'TT': TT, 'QQ': QQ, 'RR': RR,
        'a': a, 'PP': PP, 'y' : y , 'KK': KK })

```

Codice B.19: Generatore di valori dalla *Generalized Inverse Gaussian*

```

def GIGrnd( P, a, b, sampleSize ):

    # Librerie
    import numpy as np
    import numpy.random as rnd

    # Funzioni
    from My_DQ_GIGrnd import psi, dpsig, g

    # Riparametrizzazione della GIG a due parametri GIG(alpha,omega)
    Lambda = P
    Omega = np.sqrt( a*b )
    Alpha = np.sqrt( Omega**2 + Lambda**2 ) - Lambda

```

```
# Calcolo di t
x = - psi( 1, Alpha, Lambda )
if ( x >= 0.5 ) & ( x <= 2.0 ):
    t = 1.0
elif ( x > 2.0 ):
    t = np.sqrt( 2.0 / ( Alpha + Lambda ) )
elif ( x < 0.5 ):
    t = np.log( 4.0 / ( Alpha + 2.0*Lambda ) )

# Calcolo di s
x = - psi( -1, Alpha, Lambda )
if ( x >= 0.5 ) & ( x <= 2.0 ):
    s = 1.0
elif ( x > 2.0 ):
    s = np.sqrt( 4.0 / ( Alpha*np.cosh(1) + Lambda ) )
elif ( x < 0.5 ):
    s = np.min( 1.0/Lambda, np.log( 1.0 + 1.0/Alpha + np.sqrt( 1.0/Alpha**2 +
        2.0/Alpha ) ) )

# Generazione
Eta = - psi( t, Alpha, Lambda )
Zeta = - dpsi( t, Alpha, Lambda )
Theta = - psi(-s, Alpha, Lambda )
Xi = dpsi(-s, Alpha, Lambda )

p = 1 / Xi
r = 1 / Zeta
td = t - r*Eta
sd = s - p*Theta
q = td + sd

X = np.zeros( sampleSize )
for sample in np.arange( 0, sampleSize ):

    done = False
    while ( not done ):

        U = rnd.uniform( size=1 )
        V = rnd.uniform( size=1 )
```

```

W = rnd.uniform( size=1 )

if ( U < ( q / ( p + q + r ) ) ):
    X[ sample ] = - sd + q*V
elif ( U < ( ( q + r ) / ( p + q + r ) ) ):
    X[ sample ] = td - r*np.log(V)
else:
    X[ sample ] = - sd + p*np.log(V)

f1 = np.exp( - Eta - Zeta*( X[ sample ] - t ) )
f2 = np.exp( - Theta + Xi*( X[ sample ] + s ) )
if ( W*g( X[sample], sd, td, f1, f2 ) <=
    np.exp( psi( X[sample], Alpha, Lambda ) ) ):
    done = True

# Riparametrizzazione della GIG a tre parametri GIG(p,a,b)
X = np.exp( X ) * ( ( Lambda/Omega ) + np.sqrt( 1 + ( Lambda/Omega )**2 ) )
X = X / np.sqrt( a/b )

# Output
return( X )

def psi( x, Alpha, Lambda ):

# Librerie
import numpy as np

# Psi
f = - Alpha * ( np.cosh(x) - 1 ) - Lambda * ( np.exp(x) - x - 1 )

# Output
return( f )

def dpsi( x, Alpha, Lambda ):

# Librerie
import numpy as np

# Derivata
f = - Alpha * np.sinh(x) - Lambda * ( np.exp(x) - 1 )

```

```
# Output
return( f )

def g( x, sd, td, f1, f2 ):

    a = 0.0
    b = 0.0
    c = 0.0

    if ( x >= -sd ) & ( x <= td ):
        a = 1.0
    elif ( x > td ):
        b = f1
    elif ( x < -sd ):
        c = f2

    # g
    f = a + b + c

    # Output
    return( f )
```

Codice B.20: *Full conditional* per σ_ε^2 nel modello 4.5.1

```
def FC_Sigma2_Eps_IG( Data, StSim, Omega, Tau, PrS_Eps, d, ZZ, InvCC ):

    # Librerie
    import numpy as np
    import numpy.random as nrd

    # Dimensioni
    dimT = Data.shape[ 0 ]
    dimN = Data.shape[ 1 ]

    # Parametri Quantilici
    Lambda = ( 1.0-2.0*Tau ) / ( Tau*( 1.0-Tau ) )
    Delta2 = 2.0 / ( Tau*( 1.0-Tau ) )

    # Somma dei quadrati
```

```

SumSq = 0
Pred = d + ( StSim[1,: ] @ ZZ.T )
SumXi = sum( Omega[:,0] )
for t in np.arange( 0, dimT ):
    Diff = Data[t,:] - Pred[t,:] - Lambda * Omega[t,0]
    SumSq = SumSq + ( Diff @ InvCC @ Diff ) / Omega[t,0]

# Valutazione dei parametri delle full conditional
PrPostS_Eps = np.zeros( 2 )
PrPostS_Eps[ 0 ] = PrS_Eps[ 0 ] + 0.5 * dimT * ( 2+dimN )
PrPostS_Eps[ 1 ] = PrS_Eps[ 1 ] + SumXi + 0.5 * SumSq / Delta2

# Simulazione dalle distribuzioni full conditional
Sigma2_Eps = 1.0 / nrd.gamma( PrPostS_Eps[ 0 ], 1.0 / PrPostS_Eps[ 1 ], 1 )

# Output
return( Sigma2_Eps )

```

Codice B.21: *Full conditional* per il fattore latente ω nel modello 4.5.1

```

def FC_LatentFactor_GIG( Data, StSim, Tau, Sigma2_Eps, d, ZZ, InvCC ):

# Librerie
import numpy as np
from My_DQ_GIGrnd import GIGrnd

# Dimensioni
dimT = Data.shape[ 0 ]
dimN = Data.shape[ 1 ]

# Coefficienti quantilici
Delta2 = 2.0 / ( Tau*(1.0-Tau) )
Lambda2 = ( ( 1.0-2.0*Tau ) / ( Tau*(1.0-Tau) ) ) ** 2

# Forme quadratiche
Sq = np.zeros( dimT )
Pred = d + ( StSim[1,: ] @ ZZ.T )
SumCC = np.ones( dimN ) @ InvCC @ np.ones( dimN )
for t in np.arange( 0, dimT ):
    Diff = Data[t,:] - Pred[t,:]

```

```

    Sq[ t ] = Diff @ InvCC @ Diff

# Valutazione dei parametri delle full conditional
PostNu = ( 2.0-dimN ) / 2.0
PostA  = ( SumCC * ( Lambda2 / Delta2 ) + 2.0 ) / Sigma2_Eps
PostB  = Sq / ( Delta2 * Sigma2_Eps )

# Simulazione dalla distribuzione full conditional
Omega = np.zeros(( dimT, dimN ))
for t in np.arange( 0, dimT ):
    Omega[ t, : ] = GIGrnd( PostNu, PostA, PostB[ t ], 1 ) * np.ones( dimN )

# Output
return( Omega )

```

Codice B.22: *Gibbs sampling with data augmentation* per il modello 4.5.1

```

def CDQ_ST_MCMC( Y, S, NIter, Span, Scale, Tau, PrS_Alpha, PrS_Psi, PrS_Eps ):

# Librerie
import numpy as np
import numpy.linalg as nla
import numpy.random as nrd

# generatore delle matrici del modello
from My_DQ_DGK_ModelBuilder import KernelMatrixBuilder, \
    CorrMatrixBuilder , \
    DQ_ST_ModelBuilder

# generatore delle matrici del modello
from My_DQ_DGK_ModelBuilder import DQ_ST_ModelBuilder

# generatori delle distribuzioni full conditional
from My_DQ_DGK_FullConditional import FC_LatentFactor_GIG, \
    FC_Sigma2_Eps_IG

# filtro, smoother e simulation smoother
from My_DQ_DistSimSmoother import SSM_DQ_DistSimSmoother

# Dimensioni

```

```

dimT = Y.shape[ 0 ]
dimN = Y.shape[ 1 ]
dimS = dimN + 2

# Coefficienti quantilici
sLambda = ( 1.0-2.0*Tau ) / ( Tau*(1.0-Tau) )

# Allocazione della memoria
QPredSim = np.zeros(( NIter, dimT , dimN ))
StateSim = np.zeros(( NIter, dimT+1, dimS ))
Omega = np.zeros(( NIter, dimT , dimN ))
Sigma2_Eps = np.zeros( NIter )

# Calcolo delle matrici kernel e di correlazione
y = np.mean( Y )

KK = KernelMatrixBuilder( S, Span )
InvKK = nla.solve( KK, np.eye( dimN ) )

CC = CorrMatrixBuilder( S, Scale )
InvCC = nla.solve( CC, np.eye( dimN ) )

# Simulazione dalle distribuzioni iniziali e del fattore latente
sAlpha = PrS_Alpha
sPsi = PrS_Psi
sEps = 1.0 / nrd.gamma( PrS_Eps[ 0 ] , 1.0 / PrS_Eps[ 1 ] , size=1 )
sOmega = nrd.exponential( sEps, dimT*dimN ).reshape( dimT, dimN )

# Modello State Space condizionato ai parametri iniziali
SSM = DQ_ST_ModelBuilder( S, y, CC, KK, InvKK, Tau, sAlpha, sPsi, sEps )

# Simulazione del processo latente iniziale
StSim = SSM_DQ_DistSimSmoother( Y, SSM[ 'd' ], SSM[ 'ZZ' ], SSM[ 'HH' ],
                                SSM[ 'c' ], SSM[ 'TT' ], SSM[ 'RR' ],
                                SSM[ 'QQ' ], SSM[ 'a' ], SSM[ 'PP' ],
                                sLambda , sOmega )['at_Sim']

# Allocazione dei valori iniziali
QPredSim[0,:,:] = SSM[ 'd' ] + ( StSim[1,:,:] @ SSM[ 'ZZ' ].T)
StateSim[0,:,:] = StSim

```



```

Omega[0,:,:] = sOmega
Sigma2_Eps[0] = sEps

# Metropolis-Hasting within Gibbs
for i in np.arange(1,NIter):

    # STEP 1: Full Conditional per Sigma2_Eps
    sEps = FC_Sigma2_Eps_IG( Y, StSim, sOmega, Tau, PrS_Eps,
                            SSM['d'], SSM['ZZ'], InvCC )

    # STEP 2: Full Conditional per Omega
    sOmega = FC_LatentFactor_GIG( Y, StSim, Tau, sEps,
                                  SSM['d'], SSM['ZZ'], InvCC )

    # STEP 3: Costruzione delle matrici del modello
    SSM = DQ_ST_ModelBuilder( S, y, CC, KK, InvKK, Tau, sAlpha, sPsi, sEps )

    # STEP 4: Simulation Smoothing del processo latente
    StSim = SSM_DQ_DistSimSmoother( Y, SSM['d'], SSM['ZZ'], SSM['HH'],
                                     SSM['c'], SSM['TT'], SSM['RR'],
                                     SSM['QQ'], SSM['a'], SSM['PP'],
                                     sLambda, sOmega )['at_Sim']

    # STEP 5: Allocazione in memoria
    QPredSim[i,:,:] = SSM['d'] + ( StSim[1,:,:] @ SSM['ZZ'].T )
    StateSim[i,:,:] = StSim
    Omega[i,:,:] = sOmega
    Sigma2_Eps[i] = sEps

    # STEP 6: Stampa a schermo
    if ( i % 50 == 0 ):
        print('::::::::::::::::::::::::::::::::::::::::::::\n')
        print('>>> Gibbs Sampler with Data Augmentation')
        print('>>> Iteration : ', str(i), '\n')
        print('>>> Sigma2 : ', str(sEps), '\n')

# Output
return({ 'QPredSim' : QPredSim, 'StateSim': StateSim,
        'Sigma2_Eps': Sigma2_Eps, 'Omega' : Omega })

```


Bibliografia

- Anderson, B.D.O. e J.B. Moore (1979). *Optimal filtering*. Prentice-Hall information and system sciences series. Prentice-Hall, pp. xi+356. ISBN: 9-780-136-38122-8. URL: <https://books.google.it/books?id=1o0oAQAAAJ>.
- Azzalini, Adelchi e Bruno Scarpa (2012). *Data analysis and data mining*. An introduction, Translated and updated from the 2004 Italian original. Oxford University Press, Oxford, pp. x+278. ISBN: 978-0-19-976710-6.
- Bernardi, Mauro, Ghislaine Gayraud e Lea Petrella (2015). «Bayesian tail risk interdependence using quantile regression». In: *Bayesian Anal.* 10.3, pp. 553–603. ISSN: 1936-0975. DOI: [10.1214/14-BA911](https://doi.org/10.1214/14-BA911). URL: <https://doi.org/10.1214/14-BA911>.
- Bishop, Christopher M. (2006). *Pattern recognition and machine learning*. Information Science and Statistics. Springer, New York, pp. xx+738. ISBN: 978-0387-31073-2; 0-387-31073-8. DOI: [10.1007/978-0-387-45528-0](https://doi.org/10.1007/978-0-387-45528-0). URL: <https://doi.org/10.1007/978-0-387-45528-0>.
- Brockwell, Peter J. e Richard A. Davis (2016). *Introduction to time series and forecasting*. Third. Springer Texts in Statistics. Springer, [Cham], pp. xiv+425. ISBN: 978-3-319-29852-8; 978-3-319-29854-2. DOI: [10.1007/978-3-319-29854-2](https://doi.org/10.1007/978-3-319-29854-2). URL: <https://doi.org/10.1007/978-3-319-29854-2>.
- Carter, C. K. e R. Kohn (1994). «On Gibbs sampling for state space models». In: *Biometrika* 81.3, pp. 541–553. ISSN: 0006-3444. DOI: [10.1093/biomet/81.3.541](https://doi.org/10.1093/biomet/81.3.541). URL: <https://doi.org/10.1093/biomet/81.3.541>.
- (1996). «Markov chain Monte Carlo in conditionally Gaussian state space models». In: *Biometrika* 83.3, pp. 589–601. ISSN: 0006-3444. DOI: [10.1093/biomet/83.3.589](https://doi.org/10.1093/biomet/83.3.589). URL: <https://doi.org/10.1093/biomet/83.3.589>.
- Carter, Christopher Kingsley (1994). *On Markov chain Monte Carlo methods for linear state space modelling*. Thesis (Ph.D.)—University of New South Wales (Australia). ProQuest LLC, Ann Arbor, MI, (no paging). URL: http://gateway.proquest.com/openurl?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&res_dat=xri:pqdiss&rft_dat=xri:pqdiss:0574691.

- Chen, C. (2004). «An adaptive algorithm for quantile regression». In: *Theory and applications of recent robust methods*. Stat. Ind. Technol. Birkhäuser, Basel, pp. 39–48.
- Cressie, Noel e Christopher K. Wikle (2011). *Statistics for Spatio-Temporal Data*. Wiley series in Probability and Statistics. Wiley, p. 588. ISBN: 978-0-471-69274-4. URL: <https://books.google.it/books?id=-k0C6D0DiNYC>.
- Cressie, Noel A.C. (1993). *Statistics for spatial data*. Wiley series in Probability and Mathematical Statistics: Applied probability and statistics. Applied probability and statistics. J. Wiley, pp. xx+900. ISBN: 978-0-471-00255-0. URL: <https://books.google.it/books?id=4SdRAAAAMAAJ>.
- Davino, Cristina, Marilena Furno e Domenico Vistocco (2014). *Quantile regression*. Wiley Series in Probability and Statistics. Theory and applications. John Wiley & Sons, Ltd., Chichester, pp. xvi+260. ISBN: 978-1-119-97528-1. DOI: [10.1002/9781118752685](https://doi.org/10.1002/9781118752685). URL: <https://doi.org/10.1002/9781118752685>.
- De Rossi, Giuliano e Andrew Harvey (2009). «Quantiles, expectiles and splines». In: *J. Econometrics* 152.2, pp. 179–185. ISSN: 0304-4076. DOI: [10.1016/j.jeconom.2009.01.001](https://doi.org/10.1016/j.jeconom.2009.01.001). URL: <https://doi.org/10.1016/j.jeconom.2009.01.001>.
- Devroye, Luc (2014). «Random variate generation for the generalized inverse Gaussian distribution». In: *Stat. Comput.* 24.2, pp. 239–246. ISSN: 0960-3174. DOI: [10.1007/s11222-012-9367-z](https://doi.org/10.1007/s11222-012-9367-z). URL: <https://doi.org/10.1007/s11222-012-9367-z>.
- Durbin, J. e S. J. Koopman (2000). «Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives». In: *J. R. Stat. Soc. Ser. B Stat. Methodol.* 62.1. With discussion and a reply by the authors, pp. 3–56. ISSN: 1369-7412. DOI: [10.1111/1467-9868.00218](https://doi.org/10.1111/1467-9868.00218). URL: <https://doi.org/10.1111/1467-9868.00218>.
- (2002). «A simple and efficient simulation smoother for state space time series analysis». In: *Biometrika* 89.3, pp. 603–615. ISSN: 0006-3444. DOI: [10.1093/biomet/89.3.603](https://doi.org/10.1093/biomet/89.3.603). URL: <https://doi.org/10.1093/biomet/89.3.603>.
- (2012). *Time series analysis by state space methods*. Second. Vol. 38. Oxford Statistical Science Series. Oxford University Press, Oxford, pp. xxii+346. ISBN: 978-0-19-964117-8. DOI: [10.1093/acprof:oso/9780199641178.001.0001](https://doi.org/10.1093/acprof:oso/9780199641178.001.0001). URL: <https://doi.org/10.1093/acprof:oso/9780199641178.001.0001>.
- Dyk, David A. van e Taeyoung Park (2008). «Partially collapsed Gibbs samplers: theory and methods». In: *J. Amer. Statist. Assoc.* 103.482, pp. 790–796. ISSN: 0162-1459. DOI: [10.1198/016214508000000409](https://doi.org/10.1198/016214508000000409). URL: <https://doi.org/10.1198/016214508000000409>.

- Frühwirth-Schnatter, Sylvia (1994). «Data augmentation and dynamic linear models». In: *J. Time Ser. Anal.* 15.2, pp. 183–202. ISSN: 0143-9782. DOI: [10.1111/j.1467-9892.1994.tb00184.x](https://doi.org/10.1111/j.1467-9892.1994.tb00184.x). URL: <https://doi.org/10.1111/j.1467-9892.1994.tb00184.x>.
- Gaetan, Carlo e Xavier Guyon (2010). *Spatial statistics and modeling*. Springer Series in Statistics. Translated by Kevin Bleakley. Springer, New York, pp. xvi+297. ISBN: 978-0-387-92256-0. DOI: [10.1007/978-0-387-92257-7](https://doi.org/10.1007/978-0-387-92257-7). URL: <https://doi.org/10.1007/978-0-387-92257-7>.
- Geman, Stuart e Donald Geman (1984). «Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images». In: *IEEE Trans. Pattern Anal. Mach. Intell.* 6.6, pp. 721–741. ISSN: 0162-8828. DOI: [10.1109/TPAMI.1984.4767596](https://doi.org/10.1109/TPAMI.1984.4767596). URL: <https://doi.org/10.1109/TPAMI.1984.4767596>.
- Gerlach, Richard, Chris Carter e Robert Kohn (2000). «Efficient Bayesian inference for dynamic mixture models». In: *J. Amer. Statist. Assoc.* 95.451, pp. 819–828. ISSN: 0162-1459. DOI: [10.2307/2669465](https://doi.org/10.2307/2669465). URL: <https://doi.org/10.2307/2669465>.
- Gneiting, Tilmann (2002). «Nonseparable, stationary covariance functions for space-time data». In: *J. Amer. Statist. Assoc.* 97.458, pp. 590–600. ISSN: 0162-1459. DOI: [10.1198/016214502760047113](https://doi.org/10.1198/016214502760047113). URL: <https://doi.org/10.1198/016214502760047113>.
- Haslett, John e Adrian E. Raftery (1989). «Space-Time Modelling with Long-Memory Dependence: Assessing Ireland's Wind Power Resource». In: *Journal of Applied Statistics* 38.1, pp. 1–50. ISSN: 0035-9254, 1467-9876. URL: [http://onlinelibrary.wiley.com/journal/10.1111/\(ISSN\)1467-9876/issues](http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1467-9876/issues).
- Hastie, Trevor, Robert Tibshirani e Jerome Friedman (2009). *The elements of statistical learning*. Second. Springer Series in Statistics. Data mining, inference, and prediction. Springer, New York, pp. xxii+745. ISBN: 978-0-387-84857-0. DOI: [10.1007/978-0-387-84858-7](https://doi.org/10.1007/978-0-387-84858-7). URL: <https://doi.org/10.1007/978-0-387-84858-7>.
- Hunter, David R. e Kenneth Lange (2000). «Quantile regression via an MM algorithm». In: *J. Comput. Graph. Statist.* 9.1, pp. 60–77. ISSN: 1061-8600. DOI: [10.2307/1390613](https://doi.org/10.2307/1390613). URL: <https://doi.org/10.2307/1390613>.
- Jong, Piet de e Neil Shephard (1995). «The simulation smoother for time series models». In: *Biometrika* 82.2, pp. 339–350. ISSN: 0006-3444. DOI: [10.1093/biomet/82.2.339](https://doi.org/10.1093/biomet/82.2.339). URL: <https://doi.org/10.1093/biomet/82.2.339>.
- Kalman, R. E. e R. S. Bucy (1961). «New results in linear filtering and prediction theory». In: *Trans. ASME Ser. D. J. Basic Engrg.* 83, pp. 95–108.

- Koenker, Roger (1984). «A note on L -estimates for linear models». In: *Statist. Probab. Lett.* 2.6, pp. 323–325. ISSN: 0167-7152. DOI: [10.1016/0167-7152\(84\)90040-3](https://doi.org/10.1016/0167-7152(84)90040-3). URL: [https://doi.org/10.1016/0167-7152\(84\)90040-3](https://doi.org/10.1016/0167-7152(84)90040-3).
- (2005). *Quantile regression*. Vol. 38. Econometric Society Monographs. Cambridge University Press, Cambridge, pp. xvi+349. ISBN: 978-0-521-60827-5; 0-521-60827-9. DOI: [10.1017/CB09780511754098](https://doi.org/10.1017/CB09780511754098). URL: <https://doi.org/10.1017/CB09780511754098>.
- Koenker, Roger e Gilbert Bassett Jr. (1978). «Regression quantiles». In: *Econometrica* 46.1, pp. 33–50. ISSN: 0012-9682. DOI: [10.2307/1913643](https://doi.org/10.2307/1913643). URL: <https://doi.org/10.2307/1913643>.
- (1982). «Robust tests for heteroscedasticity based on regression quantiles». In: *Econometrica* 50.1, pp. 43–61. ISSN: 0012-9682. DOI: [10.2307/1912528](https://doi.org/10.2307/1912528). URL: <https://doi.org/10.2307/1912528>.
- Koopman, S. J. e J. Durbin (2000). «Fast filtering and smoothing for multivariate state space models». In: *J. Time Ser. Anal.* 21.3, pp. 281–296. ISSN: 0143-9782. DOI: [10.1111/1467-9892.00186](https://doi.org/10.1111/1467-9892.00186). URL: <https://doi.org/10.1111/1467-9892.00186>.
- (2003). «Filtering and smoothing of state vector for diffuse state-space models». In: *J. Time Ser. Anal.* 24.1, pp. 85–98. ISSN: 0143-9782. DOI: [10.1111/1467-9892.00294](https://doi.org/10.1111/1467-9892.00294). URL: <https://doi.org/10.1111/1467-9892.00294>.
- Koopman, Siem Jan (1993). «Disturbance smoother for state space models». In: *Biometrika* 80.1, pp. 117–126. ISSN: 0006-3444. DOI: [10.1093/biomet/80.1.117](https://doi.org/10.1093/biomet/80.1.117). URL: <https://doi.org/10.1093/biomet/80.1.117>.
- (1997). «Exact initial Kalman filtering and smoothing for nonstationary time series models». In: *J. Amer. Statist. Assoc.* 92.440, pp. 1630–1638. ISSN: 0162-1459. DOI: [10.2307/2965434](https://doi.org/10.2307/2965434). URL: <https://doi.org/10.2307/2965434>.
- Koopman, Siem Jan e Kai Ming Lee (2009). «Seasonality with trend and cycle interactions in unobserved components models». In: *J. R. Stat. Soc. Ser. C. Appl. Stat.* 58.4, pp. 427–448. ISSN: 0035-9254. DOI: [10.1111/j.1467-9876.2009.00661.x](https://doi.org/10.1111/j.1467-9876.2009.00661.x). URL: <https://doi.org/10.1111/j.1467-9876.2009.00661.x>.
- Kotz, Samuel, Tomasz J. Kozubowski e Krzysztof Podgórski (2001). *The Laplace distribution and generalizations*. A revisit with applications to communications, economics, engineering, and finance. Birkhäuser Boston, Inc., Boston, MA, pp. xviii+349. ISBN: 0-8176-4166-1. DOI: [10.1007/978-1-4612-0173-1](https://doi.org/10.1007/978-1-4612-0173-1). URL: <https://doi.org/10.1007/978-1-4612-0173-1>.
- Kotz, Samuel e Saralees Nadarajah (2004). *Multivariate t distributions and their applications*. Cambridge University Press, Cambridge, pp. xii+272. ISBN: 0-521-

- 82654-3. DOI: [10.1017/CB09780511550683](https://doi.org/10.1017/CB09780511550683). URL: <https://doi.org/10.1017/CB09780511550683>.
- Kozumi, Hideo e Genya Kobayashi (2011). «Gibbs sampling methods for Bayesian quantile regression». In: *J. Stat. Comput. Simul.* 81.11, pp. 1565–1578. ISSN: 0094-9655. DOI: [10.1080/00949655.2010.496117](https://doi.org/10.1080/00949655.2010.496117). URL: <https://doi.org/10.1080/00949655.2010.496117>.
- Kurose, Yuta e Yasuhiro Omori (2012). «Bayesian analysis of time-varying quantiles using a smoothing spline». In: *J. Japan Statist. Soc.* 42.1, pp. 23–46. ISSN: 1882-2754. DOI: [10.14490/jjss.42.23](https://doi.org/10.14490/jjss.42.23). URL: <https://doi.org/10.14490/jjss.42.23>.
- Liu, Jun S. (1994). «The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem». In: *J. Amer. Statist. Assoc.* 89.427, pp. 958–966. ISSN: 0162-1459. URL: [http://links.jstor.org/sici?sici=0162-1459\(199409\)89:427<958:TCGSIB>2.0.CO;2-G&origin=MSN](http://links.jstor.org/sici?sici=0162-1459(199409)89:427<958:TCGSIB>2.0.CO;2-G&origin=MSN).
- Lowe, Rachel et al. (2011). «Spatio-temporal modelling of climate-sensitive disease risk: Towards an early warning system for dengue in Brazil». In: *Computers & Geosciences* 37.3, pp. 371–381. DOI: [10.1016/j.cageo.2010.01.008](https://doi.org/10.1016/j.cageo.2010.01.008). URL: <https://doi.org/10.1016/j.cageo.2010.01.008>.
- Luna, Xavier de e Marc G. Genton (2005). «Predictive spatio-temporal models for spatially sparse environmental data». In: *Statist. Sinica* 15.2, pp. 547–568. ISSN: 1017-0405.
- Mardia, Kanti V. et al. (1998). «The Kriged Kalman filter». In: *Test* 7.2. With comments and a rejoinder by the authors, pp. 217–285. ISSN: 1133-0686. DOI: [10.1007/BF02565111](https://doi.org/10.1007/BF02565111). URL: <https://doi.org/10.1007/BF02565111>.
- Michael, John R., William R. Schucany e Roy W. Haas (1976). «Generating Random Variates Using Transformations with Multiple Roots». In: *The American Statistician* 30.2, pp. 88–90. DOI: [10.1080/00031305.1976.10479147](https://doi.org/10.1080/00031305.1976.10479147). URL: <https://www.tandfonline.com/doi/abs/10.1080/00031305.1976.10479147>.
- Ormerod, J. T. e M. P. Wand (2010). «Explaining variational approximations». In: *Amer. Statist.* 64.2, pp. 140–153. ISSN: 0003-1305. DOI: [10.1198/tast.2010.09058](https://doi.org/10.1198/tast.2010.09058). URL: <https://doi.org/10.1198/tast.2010.09058>.
- Reich, Brian J., Montserrat Fuentes e David B. Dunson (2011). «Bayesian spatial quantile regression». In: *J. Amer. Statist. Assoc.* 106.493, pp. 6–20. ISSN: 0162-1459. DOI: [10.1198/jasa.2010.ap09237](https://doi.org/10.1198/jasa.2010.ap09237). URL: <https://doi.org/10.1198/jasa.2010.ap09237>.

- Ripley, Brian D. (1981). *Spatial statistics*. Wiley series in Probability and Mathematical Statistics. Wiley New York, pp. x+252. ISBN: 0-4710-8367-4. URL: <https://nla.gov.au/nla.cat-vn1454331>.
- Rue, Havard e Leonhard Held (2005). *Gaussian Markov Random Fields: Theory And Applications*. Monographs on Statistics and Applied Probability. Chapman & Hall/CRC, pp. xix+280. ISBN: 1-5848-8432-0.
- Ruppert, David, M. P. Wand e R. J. Carroll (2003). *Semiparametric regression*. Vol. 12. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, pp. xvi+386. ISBN: 0-521-78050-0; 0-521-78516-2. DOI: [10.1017/CB09780511755453](https://doi.org/10.1017/CB09780511755453). URL: <https://doi.org/10.1017/CB09780511755453>.
- Sahu, Sujit K. e Kanti V. Mardia (2005). «A Bayesian kriged Kalman model for short-term forecasting of air pollution levels». In: *J. Roy. Statist. Soc. Ser. C* 54.1, pp. 223–244. ISSN: 0035-9254. DOI: [10.1111/j.1467-9876.2005.00480.x](https://doi.org/10.1111/j.1467-9876.2005.00480.x). URL: <https://doi.org/10.1111/j.1467-9876.2005.00480.x>.
- Scarpa, Bruno (2001). «Analisi della concentrazione di SO₂ combinando i dati raccolti da centraline fisse e mobili: un modello state space». In: *Statistica (Bologna)* 61.4, 707–730 (2002). ISSN: 0390-590X.
- Shumway, R. H. e D. S. Stoffer (1982). «An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm». In: *Journal of Time Series Analysis* 3.4, pp. 253–264. DOI: [10.1111/j.1467-9892.1982.tb00349.x](https://doi.org/10.1111/j.1467-9892.1982.tb00349.x). URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9892.1982.tb00349.x>.
- Simon, Dan (2006). *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. Wiley-Interscience. ISBN: 0-471-70858-5.
- Simon, Dan e Donald L. Simon (2006). «Kalman filter constraint switching for turbofan engine health estimation». In: *Eur. J. Control* 12.3. With a discussion by Daniel Viassolo, pp. 331–345. ISSN: 0947-3580. DOI: [10.3166/ejc.12.341-343](https://doi.org/10.3166/ejc.12.341-343). URL: <https://doi.org/10.3166/ejc.12.341-343>.
- (2010). «Constrained Kalman filtering via density function truncation for turbofan engine health estimation». In: *Internat. J. Systems Sci.* 41.2, pp. 159–171. ISSN: 0020-7721. DOI: [10.1080/00207720903042970](https://doi.org/10.1080/00207720903042970). URL: <https://doi.org/10.1080/00207720903042970>.
- Sriram, Karthik e R. V. Ramamoorthi (2017). «Correction to: “Posterior consistency of Bayesian quantile regression based on the misspecified asymmetric Laplace density” [MR3066950]». In: *Bayesian Anal.* 12.4, pp. 1217–1219. ISSN: 1936-0975. DOI: [10.1214/17-BA1067](https://doi.org/10.1214/17-BA1067). URL: <https://doi.org/10.1214/17-BA1067>.

- Sriram, Karthik, R. V. Ramamoorthi e Pulak Ghosh (2016). «On Bayesian Quantile Regression Using a Pseudo-joint Asymmetric Laplace Likelihood». In: *Sankhya A* 78.1, pp. 87–104. ISSN: 0976-8378. DOI: [10.1007/s13171-015-0079-2](https://doi.org/10.1007/s13171-015-0079-2). URL: <https://doi.org/10.1007/s13171-015-0079-2>.
- Stein, Michael L. (2005a). «Space-time covariance functions». In: *J. Amer. Statist. Assoc.* 100.469, pp. 310–321. ISSN: 0162-1459. DOI: [10.1198/016214504000000854](https://doi.org/10.1198/016214504000000854). URL: <https://doi.org/10.1198/016214504000000854>.
- (2005b). «Statistical methods for regular monitoring data». In: *J. R. Stat. Soc. Ser. B Stat. Methodol.* 67.5, pp. 667–687. ISSN: 1369-7412. DOI: [10.1111/j.1467-9868.2005.00520.x](https://doi.org/10.1111/j.1467-9868.2005.00520.x). URL: <https://doi.org/10.1111/j.1467-9868.2005.00520.x>.
- Stroud, Jonathan R., Peter Müller e Bruno Sansó (2001). «Dynamic models for spatiotemporal data». In: *J. R. Stat. Soc. Ser. B Stat. Methodol.* 63.4, pp. 673–689. ISSN: 1369-7412. DOI: [10.1111/1467-9868.00305](https://doi.org/10.1111/1467-9868.00305). URL: <https://doi.org/10.1111/1467-9868.00305>.
- Tan, Linda S. L. e David J. Nott (2018). «Gaussian variational approximation with sparse precision matrices». In: *Stat. Comput.* 28.2, pp. 259–275. ISSN: 0960-3174. DOI: [10.1007/s11222-017-9729-7](https://doi.org/10.1007/s11222-017-9729-7). URL: <https://doi.org/10.1007/s11222-017-9729-7>.
- Tian, Yuzhu, Maozai Tian e Qianqian Zhu (2014). «Linear quantile regression based on EM algorithm». In: *Comm. Statist. Theory Methods* 43.16, pp. 3464–3484. ISSN: 0361-0926. DOI: [10.1080/03610926.2013.766339](https://doi.org/10.1080/03610926.2013.766339). URL: <https://doi.org/10.1080/03610926.2013.766339>.
- Wahba, Grace (1990). *Spline models for observational data*. Vol. 59. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial e Applied Mathematics (SIAM), Philadelphia, PA, pp. xii+169. ISBN: 0-89871-244-0. DOI: [10.1137/1.9781611970128](https://doi.org/10.1137/1.9781611970128). URL: <https://doi.org/10.1137/1.9781611970128>.
- Wang, Yuedong (2011). *Smoothing splines*. Vol. 121. Monographs on Statistics and Applied Probability. Methods and applications. CRC Press, Boca Raton, FL, pp. xxiv+370. ISBN: 978-1-4200-7755-1. DOI: [10.1201/b10954](https://doi.org/10.1201/b10954). URL: <https://doi.org/10.1201/b10954>.
- Wikle, Christopher K., L. Mark Berliner e Noel Cressie (2004). «Hierarchical Bayesian space-time models». In: *Environmental and Ecological Statistics* 5, pp. 117–154.
- Wikle, Christopher K. e Noel Cressie (1999). «A dimension-reduced approach to space-time Kalman filtering». In: *Biometrika* 86.4, pp. 815–829. DOI: [10.1093/biomet/86.4.815](http://dx.doi.org/10.1093/biomet/86.4.815). URL: <http://dx.doi.org/10.1093/biomet/86.4.815>.

- Wilk, Mark van der, Carl Edward Rasmussen e James Hensman (2017). «Convolutional Gaussian Processes». In: *Advances in Neural Information Processing Systems 30*. A cura di I. Guyon et al. Curran Associates, Inc., pp. 2849–2858. URL: <http://papers.nips.cc/paper/6877-convolutional-gaussian-processes.pdf>.
- Yu, Keming e Rana A. Moyeed (2001). «Bayesian quantile regression». In: *Statist. Probab. Lett.* 54.4, pp. 437–447. ISSN: 0167-7152. DOI: [10.1016/S0167-7152\(01\)00124-9](https://doi.org/10.1016/S0167-7152(01)00124-9). URL: [https://doi.org/10.1016/S0167-7152\(01\)00124-9](https://doi.org/10.1016/S0167-7152(01)00124-9).
- Zou, Hui e Ming Yuan (2008). «Composite quantile regression and the oracle model selection theory». In: *Ann. Statist.* 36.3, pp. 1108–1126. ISSN: 0090-5364. DOI: [10.1214/07-AOS507](https://doi.org/10.1214/07-AOS507). URL: <https://doi.org/10.1214/07-AOS507>.

Ringraziamenti

Eccomi qui, finalmente concluso questo percorso di studi, che solo all'ultimo mi rendo conto quanto importanti siano stati per me questi anni universitari; senz'ombra di dubbio i più belli della mia vita!

Dei semplici ringraziamenti probabilmente non sono abbastanza per dimostrare quanto io sia grato alle persone che mi sono state vicine in questo periodo. Ciò nonostante vorrei esprimere la riconoscenza che ho per voi in queste poche righe, nella speranza di riuscire a comunicarvi quanto siete stati e quanto siete importanti per me.

Sopra chiunque altro voglio ringraziare i miei genitori, per il sostegno, la fiducia e l'amore incondizionato che mi hanno sempre dimostrato. Grazie mamma, grazie papà. La persona che sono oggi è il frutto della vostra educazione, dei vostri insegnamenti e di tutti i sacrifici che avete fatto per donarmi una vita piena e felice.

Non meno importanti sono i miei nonni, le colonne portanti che con passione ed entusiasmo hanno sempre fatto il tifo per me. Spero di essere riuscito a rendervi orgogliosi. Vi voglio bene.

Roberto, ai miei occhi sei un'insostituibile parte della famiglia, e questo mio traguardo lo voglio dedicare un po' anche a te, per il calore con cui mi sei sempre stato accanto.

E come non ringraziare Enrico, Lorenzo e Nicolò, il gruppo dell'Adimaria, i primi compagni di ventura di questo fantastico viaggio padovano. Santa Caterina sarebbe stato un posto più triste senza di voi, siete fantastici.

Ringrazio poi Davide (il Bomber), Davide (detto Dado), Emanuele (il nostro Presidente), Filippo (detto Pippo), Mattia (veloce come un fulmine) e Silvia (la signorina capotutor). La vostra compagnia ha reso questi ultimi due anni più divertenti e stimolanti di quanto avrei mai potuto immaginare.

Alice, Fausto, Federico, Giacomo, Marco e Mattia, spero sappiate quanto siete importanti per me, non potrei desiderare amici migliori di voi. Arricchite la mia vita in ogni suo aspetto e mi avete indubbiamente reso una persona migliore. Grazie a tutti voi.

Voglio poi dedicare un ringraziamento speciale ai professori del dipartimento di Scienze Statistiche, che sono stati per me un faro e una guida, grazie ai quali ho scoperto la mia vera passione.

Ringrazio il mio relatore, il Professor Mauro Bernardi, per la sua infinita disponibilità e per tutto il tempo che mi ha dedicato in questo periodo. Penso di non esagerare dicendo che mi ha fatto scoprire più statistica lei in questi pochi mesi, di quanta non ne abbia appresa in cinque anni.

Voglio ringraziare poi la Professoressa Brazzale, per la gentilezza e la cordialità che mi ha dimostrato quest'anno. Sebbene il nostro progetto iniziale non sia andato a buon fine, conservo dei bei ricordi del periodo in cui abbiamo lavorato insieme.

Ringrazio infine la Professoressa Menardi e il Professor Scarpa, i quali hanno tenuto i corsi che più mi hanno formato e più mi hanno entusiasmato in questo ciclo di studi.