

UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE STATISTICHE

CORSO DI LAUREA MAGISTRALE IN  
SCIENZE STATISTICHE



TESI DI LAUREA

**Previsione della domanda di energia elettrica mediante  
analisi funzionale: il caso del mercato inglese.**

RELATORE: PROF. Francesco Lisi

DIPARTIMENTO DI SCIENZE STATISTICHE

LAUREANDO: Elena Benetti

MATRICOLA: 1013898

ANNO ACCADEMICO 2013/2014



# Indice

<b>Introduzione</b>	<b>III</b>
<b>1 Il mercato elettrico</b>	<b>1</b>
1.1 Il mercato elettrico inglese . . . . .	3
1.2 Organizzazione del mercato inglese . . . . .	5
<b>2 La domanda di energia elettrica nel mercato inglese</b>	<b>9</b>
2.1 Studio della serie storica della domanda . . . . .	10
2.2 Decomposizione della serie della domanda . . . . .	16
2.2.1 La componente di lungo periodo . . . . .	16
2.2.2 La componente periodica . . . . .	17
2.2.3 La componente stocastica . . . . .	19
<b>3 Modelli univariati, multivariati e funzionali</b>	<b>23</b>
3.1 Modelli univariati . . . . .	24
3.1.1 Modelli autoregressivi AR(p) . . . . .	24
3.1.2 Modelli ARX(p) . . . . .	24
3.2 Modelli multivariati . . . . .	28
3.2.1 Modelli VAR(p) . . . . .	28
3.2.2 Modelli VARX(p) . . . . .	29
3.3 Modelli funzionali . . . . .	30
<b>4 Applicazione ai dati</b>	<b>35</b>
4.1 Previsioni in-sample . . . . .	36

---

4.1.1	Selezione dell'ordine AR . . . . .	36
4.1.2	Selezione dell'ordine ARX . . . . .	42
4.1.3	Selezione del modello VAR . . . . .	45
4.1.4	Selezione del modello VARX . . . . .	48
4.1.5	Selezione del modello funzionale . . . . .	52
4.2	Previsioni out-of-sample . . . . .	65
4.2.1	Modello AR(1) . . . . .	66
4.2.2	Modello ARX(1) . . . . .	69
4.2.3	Modello VAR(1) . . . . .	73
4.2.4	Modello VARX(1) . . . . .	76
4.2.5	Modello AR funzionale . . . . .	80
4.2.6	Modello ARX funzionale . . . . .	83
4.3	Confronto finale . . . . .	87
	<b>Conclusioni</b>	<b>91</b>
	<b>Ringraziamenti</b>	<b>93</b>

# Introduzione

Ad oggi è impossibile pensare alla nostra vita senza l'elettricità. Tutto oramai è connesso ad una presa.

Anche se si utilizza l'elettricità ormai da sempre, il mercato elettrico è relativamente più giovane e in evoluzione continua.

Nel mercato elettrico l'oggetto scambiato è l'energia, la quale viene scambiata tramite un sistema d'aste. Infatti, nelle borse elettriche, si formulano delle aste sia per la domanda sia per l'offerta; queste si incontrano in un punto stabilendo la quantità di energia da produrre e il prezzo. La particolarità di tale mercato, è che le aste vengono effettuate su base oraria o semi oraria, ovvero si registra, per ogni ora o mezz'ora, il prezzo e la quantità di energia sia domandata, sia offerta.

In questa tesi si è voluta analizzare la domanda dell'energia del mercato elettrico. L'elaborato si concentra sulla previsione della domanda in un orizzonte temporale breve; in particolare si vuole prevedere la domanda di energia del mercato elettrico del Regno Unito.

Le tecniche utilizzate e messe a confronto sono tre: l'analisi univariata, l'analisi multivariata e quella funzionale.

L'analisi funzionale, come il mercato elettrico, è un ramo della statistica abbastanza recente; è stata infatti utilizzata da Ramsey nel 1996 che analizza il movimento delle labbra. L'analisi funzionale, rispetto ai metodi più classici, ha come oggetto di studio, non più un'osservazione o una serie di osservazioni, ma una funzione.

L'elaborato è strutturato come segue. Nel primo capitolo si descrive la nascita

dei mercati elettrici così come oggi si conoscono, poi si riportano le caratteristiche del mercato elettrico oggetto di studio, quello del Regno Unito. Nel mercato inglese la domanda è formata da 48 osservazioni giornaliere le quali corrispondono ad una mezzora; ovvero le proposte d'asta che formano la domanda sono su base semi oraria. Ogni osservazione corrisponde ad una fascia semi oraria: la prima si riferisce alle ore 00.00-00.30, la seconda alle 00.30-01.00, e così via.

Per tale motivo la serie della domanda può essere analizzata con modelli univariati, se si pensa alle singole osservazioni; oppure tramite modelli multivariati, che considerano le fasce semi orarie come un'unica serie multivariata. Infine può essere studiata mediante l'analisi funzionale, che considera il profilo giornaliero della domanda come un'unica curva comprendente le diverse osservazioni.

Nel secondo capitolo viene analizzata nel dettaglio la serie della domanda di energia, studiandone le componenti principali, sia per quanto riguarda il comportamento giornaliero, sia per quanto riguarda il comportamento di una singola fascia semi oraria. Si mostra che la domanda è legata ad una componente atmosferica e legata all'attività dell'uomo; si sfruttano, quindi, le caratteristiche studiate per ottenere una serie della domanda depurata dalle componenti deterministiche.

Nel terzo capitolo, si descrivono teoricamente le tecniche utilizzate per la previsione. Si descrivono i modelli lineari funzionali, i modelli autoregressivi univariati e multivariati. Poiché la domanda di energia è legata alle condizioni atmosferiche si introducono, nei diversi modelli, le informazioni derivanti dal meteo, come la temperatura media del Regno Unito.

Il quarto capitolo è diviso sostanzialmente in due parti. Una prima parte in cui si identificano i modelli migliori mediante la previsione condotta su un sottoperiodo della serie depurata dalle componenti deterministiche. Una seconda parte, nella quale si confrontano i modelli sulla base di indici di bontà delle previsioni della domanda di energia iniziale, e si stabilisce se vi è un modello che ha migliori *performance* di tutti gli altri.

# Capitolo 1

## Il mercato elettrico

Negli ultimi trent'anni, il mercato elettrico, in molti paesi, si è avviato verso una completa liberalizzazione. Le ragioni di fondo di tale liberalizzazione sono date da idee e motivi politici comuni ai diversi paesi. Si tratta di stimolare l'innovazione tecnica e produrre degli investimenti efficienti che portino a guadagni efficienti.

Lo stato che per primo ha affrontato la liberalizzazione è il Cile, nel 1982, con una riforma che prevedeva di separare le compagnie di generazione e distribuzione dell'energia, fino alla formazione di un meccanismo di scambio all'ingrosso dell'energia. In seguito alla riforma cilena si colloca quella del mercato inglese (1990). Fino all'anno 2005, il mercato britannico includeva solo Inghilterra e Galles, successivamente si è aggregata anche la Scozia. Nell'Europa del Nord il mercato elettrico si è avviato nel 1992, inizialmente in Norvegia e poi in Svezia, Finlandia e Danimarca. In Australia, invece, i mercati hanno iniziato a operare nel 1994, ma solamente in alcune regioni; e solo nel 1998 è stato aperto l' *Australian National Electricity Market*. Negli stessi anni inizia la riforma dell'energia elettrica anche in Nuova Zelanda, che aveva aperto il mercato qualche anno prima. Nel nord America i mercati hanno iniziato a operare negli ultimi anni novanta, mentre nel 1998 in California. Qualche anno più tardi il mercato elettrico è entrato anche nel Texas e nell'Alberta.

In generale il numero di mercati elettrici liberalizzati è in continuo aumento, anche se ciò è più visibile in Europa rispetto alle altre zone del mondo. Si deve altresì constatare che la liberalizzazione ha avuto, nel lungo termine, un impatto positivo sull'economia, anche se in alcuni mercati, come quello californiano, ha aperto le porte ad una crisi più profonda.

In ogni caso, il processo di liberalizzazione è un processo lento ed impegnativo, complicato dal fatto che non vi è un unico modello di mercato migliore da seguire; ogni paese deve tener conto delle proprie caratteristiche economiche, politiche e ambientali.

Il mercato dell'energia elettrica è, infatti, un particolare mercato nel quale l'oggetto che viene scambiato è l'energia che viene prodotta. La domanda e l'offerta di energia vengono contrattate nella Borsa dell'Energia Elettrica, introdotta in diversi paesi attraverso diverse strategie. Ogni paese, infatti, ha le sue regolamentazioni e caratteristiche.

In letteratura statistica il mercato elettrico è un tema ampiamente affrontato e discusso, sia per quanto riguarda l'analisi, le caratteristiche e la previsione dei prezzi; sia per quanto riguarda lo studio e la previsione della domanda dell'energia elettrica. Molti articoli studiano e analizzano le caratteristiche del mercato elettrico, ma soprattutto si concentrano sulla costruzione di diversi modelli di previsione dei prezzi, del consumo e della domanda stessa di energia fatti generalmente sul breve periodo.

Ad esempio, Andersson e Lillestøl (2010), mostrano come si può applicare l'analisi funzionale al mercato elettrico. Essi utilizzano l'analisi funzionale della varianza per studiare i comportamenti stagionali della domanda dell'energia, introducendo come variabili dipendenti l'effetto annuale, mensile e giornaliero. Studiano, inoltre, il modello autoregressivo funzionale del primo ordine per cercare di migliorare la previsione, nel breve periodo, della domanda dell'energia.

Conejo et al (2005), studiano, invece, la previsione dei prezzi orari dell'energia per il giorno successivo. Nel loro studio decompongono la serie storica



dei prezzi fino al giorno  $t - 1$ , in un insieme di serie che corrispondono alle quattro stagioni. Per ogni insieme, applicano poi un specifico modello ARIMA per prevedere i 24 valori per il giorno  $t$ .

Blum e Riedmiller (2013) propongono un algoritmo di previsione della domanda di energia elettrica sulla base di processi gaussiani. In questo lavoro utilizzano, come variabili dipendenti, il comportamento tipico giornaliero e settimanale della domanda di energia, e i dati forniti dal meteo riguardanti il tempo, per fare delle previsioni del consumo di energia e della produzione di energia rinnovabile.

E' interessante prendere spunto da questi e molti altri articoli, per studiare e costruire un modello di previsione per la domanda di energia. L'obiettivo di questa tesi è prevedere, nel breve periodo, la domanda di energia del mercato elettrico del Regno Unito, mediante l'analisi funzionale, e confrontare i risultati con modelli univariati più semplici e modelli multivariati.

Nei paragrafi seguenti verrà descritta brevemente l'organizzazione del mercato elettrico inglese e le sue caratteristiche principali.

## 1.1 Il mercato elettrico inglese

La generazione di elettricità nel Regno Unito è data da un mix energetico, ovvero una parte di energia è fornita da centrali a gas e a carbone, una parte proviene da centrali nucleari, e infine, una parte, viene prodotta da un'enorme gamma di fonti rinnovabili, tra cui primeggia l'energia eolica. Questi sono i principali attori del mercato inglese. Vi sono poi le *utilities*, che si occupano della distribuzione, acquistando dai generatori e vendendo all'utente finale attraverso le reti al dettaglio. I rapporti tra chi genera energia e le *utilities* sono regolati dai *British Electricity Trading and Transmission Arrangements (BETTA)*, che rendono agevole lo scambio di energia agli attori fisicamente distanti, grazie all'utilizzo della rete nazionale. Questa rete è il centro del mercato ed è gestita dalle sette compagnie che hanno in carico la rete di distribuzione nelle quattordici regioni del Regno Unito, nel dettaglio troviamo EDF Energy,

Central Networks, CE Electric, Western Power Distribution, United Utilities, Scottish Power e Scottish and Southern.

L'industria elettrica inglese è stata privatizzata nel 1990, e nel decennio successivo ha sviluppato una lenta liberalizzazione, nata dal bisogno di adeguarsi alle esigenze degli operatori di mercato e delle istituzioni. Il mercato inglese è infatti il caso di liberalizzazione più noto e in continua evoluzione. Nel febbraio 1988 venne emanato il *White Paper 'Privatising Electricity'*, che contiene le principali modalità per rendere il mercato elettrico ancora più competitivo e il *framework* legislativo di regolazione del nuovo assetto organizzativo. Il *White Paper* prevedeva il cambiamento dell'assetto di mercato in termini maggiormente concorrenziali attraverso la rimozione del monopolio detenuto dalla Central Electricity Generating Board (CEGB) nelle fasi di generazione e trasmissione, agevolando così l'entrata nel mercato a nuovi operatori.

Le problematiche riscontrate dell'industria elettrica sono principalmente due: il bilanciamento della domanda e dell'offerta; e, in secondo luogo, le modalità per garantire l'innovazione di fronte al graduale esaurimento degli idrocarburi e del cambiamento climatico.

Il problema del bilanciamento risiede nel fatto che l'energia elettrica è un bene di scambio complesso, in quanto non è possibile lo stoccaggio in quantità tali da influenzare l'interazione di mercato; è quindi necessario che in ogni momento la quantità prodotta equivalga a quella richiesta, cioè è necessario che ci sia un equilibrio tra domanda e offerta. Il periodo base del bilanciamento è mezz'ora, quindi una giornata è divisa in 48 periodi. Ogni quantità scambiata ha un ammontare e un periodo preciso durante il quale sarà prodotta e consumata. L'energia elettrica così definita contrattualmente, può essere scambiata fino ad un'ora prima del periodo a cui è riferita. A questo punto il generatore deve confermare la capacità di onorare il contratto e i prezzi ai quali è, eventualmente, disposto a produrre in eccesso o in difetto, rispetto alla quantità stabilita. Grazie a queste informazioni, la società incaricata opera il bilanciamento del mercato in maniera tale che, allo scadere del

periodo, domanda e offerta si equivalgano. La modernizzazione del sistema inglese è, in parte, dipendente anche dalle problematiche del meccanismo di bilanciamento.

Il 12 Luglio 2011 il Governo inglese ha, infatti, presentato il Libro Bianco della Riforma del Mercato Elettrico, col quale il Governo si impegna a garantire un'energia più pulita, più economica, più sicura e più inglese; una riforma resa necessaria vista la grande quantità di centrali nucleari e a carbone. Quindi le nuove risorse energetiche necessarie sarebbero assicurate da un mix basato su gas, nucleare di nuova generazione, energie rinnovabili, riducendo la dipendenza dall'estero e dalla fragilità delle condizioni politiche internazionali, e sviluppando con vigore la produzione interna di energia pulita sfruttando al massimo soprattutto la risorsa eolica, settore in cui il Regno Unito è già leader mondiale. Oltre a nuove energie si introdurranno nuovi tipi di contratto a lungo termine per fornire incentivi a chi investe in queste nuove fonti di energia. Una riforma comunque che si concluderà non prima del 2020, per un ulteriore approfondimento si veda il Libro Bianco della Riforma del Mercato Elettrico (2011).

## 1.2 Organizzazione del mercato inglese

Durante il processo di trasformazione dell'industria elettrica, viene creato il Pool, ovvero un contratto aperto multilaterale stipulato tra generatori e venditori che definisce le regole e i meccanismi utilizzati per scambiare energia all'ingrosso. Il Pool è definito da:

- l'insieme di regole che definisce le modalità attraverso le quali i soggetti che scambiano energia devono contrattare;
- il sistema attraverso il quale è possibile per i generatori offrire l'elettricità e per i compratori approvvigionarsi;
- il meccanismo di definizione dei prezzi dell'elettricità all'ingrosso per ciascuna mezz'ora;

- il sistema di aggiustamento di accredito/addebito dei generatori/venditori.

Le principali caratteristiche del Pool inglese sono le seguenti:

1. il Pool è obbligatorio, cioè tutti i generatori in possesso di una licenza devono vendere la maggior parte del loro output mediante il Pool e gli acquirenti/venditori devono acquistare la maggior parte dell'elettricità attraverso tale sistema. La contrattazione al di fuori del Pool è per lo più limitata;
2. la forma per il dispacciamento dell'energia prodotta è quella di merito. I gruppi (produttori) ammessi a produrre, e il loro livello di produzione, vengono stabiliti dall'Operatore di Mercato in base a un ordine di merito risultante dalle offerte economiche presentate in borsa dai produttori per ciascuna unità di tempo e di produzione;
3. è un mercato unilaterale in quanto solamente i venditori formulano la loro offerta nel Pool, mentre la domanda degli acquirenti è stimata;
4. le offerte sono molto complesse perchè riflettono il costo dei generatori, ma non sono vincolanti e i generatori possono dichiarare nuovamente la loro disponibilità a produrre;
5. i generatori che partecipano al Pool pagano il *System Marginal Price* (SMP). Si definisce in genere, *System Marginal Price*, il prezzo dell'unità più costosa programmata in modo da incontrare la domanda prevista;
6. i costi per il bilanciamento sono distribuiti tra tutti i partecipanti al mercato;
7. il *Pool market* si basa su un contratto multilaterale stipulato tra generatori e venditori;
8. è un sistema *ex ante*, ovvero i prezzi vengono definiti il giorno prima della formulazione del contratto.

Il mercato inglese è caratterizzato da un dispacciamento di merito con borsa obbligatoria, è quindi possibile stipulare contratti bilaterali di tipo finanziario che non hanno come oggetto la consegna fisica di energia, ma solo i congruagli tra il prezzo di borsa e quello concordato, allo scopo di salvaguardare i contraenti dalle fluttuazioni dei prezzi spot dell'energia. Questo tipo di organizzazione del mercato non ha prodotto molti benefici, pertanto il Governo inglese ha chiesto all'Ofgem (*Office For Electricity Regulation*) di orientarsi verso una forma organizzativa simile a quella dei paesi nordici, con una borsa giornaliera non obbligatoria, nella quale sono ammessi contratti bilaterali tra produttori e clienti, e una borsa intragiornaliera.

Nel mercato inglese la borsa è giornaliera, quindi si definisce giorno per giorno il prezzo dell'energia per il giorno successivo, questo tipo di mercato viene detto *Day Ahead Market*. Le offerte di energia vengono effettuate su un periodo di mezz'ora. La determinazione della quantità e del prezzo di scambio dell'energia elettrica avviene attraverso la modalità *unit commitment* globale di sistema. Con questa modalità, disponendo dei prezzi di produzione, di start-up (avviamento del gruppo) e di no-load (prezzo di funzionamento a potenza nulla) offerti per i singoli gruppi partecipanti e disponendo inoltre della potenza richiesta stimata dall'Operatore di Sistema, l'operatore di mercato determina, per ciascuno dei periodi temporali previsti, i gruppi da mettere in servizio e la potenza che ciascuno dovrà erogare. Il gruppo, fra quelli chiamati a produrre, che presenta il prezzo medio di produzione del kWh più elevato, determina il prezzo dell'energia uguale per tutti.

Ricapitolando, l'operatore di sistema stabilisce il margine di potenza necessario per la regolazione successiva; la disponibilità di potenza viene remunerata a tariffa; e l'energia viene infine pagata in base alle offerte fatte alla borsa giornaliera.

Nella borsa giornaliera ogni generatore formula un'offerta per ciascuna unità di elettricità tenendo conto di diversi parametri ed elementi, quindi definiscono un'offerta che tiene conto anche dei costi di lavoro. Tutti questi elementi vengono utilizzati dall'operatore di sistema National Grid Company,

per determinare la programmazione oraria degli impianti di produzione per il giorno successivo. L'operatore di rete formula poi una previsione della domanda, in genere maggiorata di un margine di riserva, per ciascuna mezz'ora del giorno successivo e poi programma le offerte dei diversi generatori in modo tale da incontrare la domanda.

Il sistema del Pool inglese ha non pochi punti critici per quanto riguarda la formulazione delle offerte. Quest'ultime sono complicate e poco trasparenti, manca una vera competizione nella determinazione del prezzo e vi è la possibilità per alcuni operatori di manipolare il prezzo stabilito. Queste e altre criticità hanno portato il sistema inglese a revisionare il processo del Pool. Un cambiamento che prevede lo scambio di elettricità su tre mercati:

- un mercato di contratti bilaterali di lungo termine, nel quale i generatori e i produttori di elettricità all'ingrosso contrattano bilateralmente;
- un mercato bilaterale a breve termine che opera con 3 ore e mezzo di anticipo rispetto al tempo reale;
- un mercato di bilanciamento che permette all'operatore di sistema l'accettazione di offerte da generatori e acquirenti/venditori per avvicinare le discrepanze tra i livelli attesi di output o della domanda e i livelli reali delle stesse.

Come accennato inizialmente, il mercato dell'energia inglese è in continua evoluzione proprio per cercare di eliminare i limiti dell'organizzazione interna e avere un mercato più efficiente per il benessere di tutti.

L'obiettivo del governo è quello, in definitiva, di raggiungere, entro il 2020, un 30% di energia rinnovabili, mantenendo il 30% del gas e riducendo drasticamente il carbone, e producendo il restante con l'energia nucleare.

## Capitolo 2

# La domanda di energia elettrica nel mercato inglese

Come si è detto, l'obiettivo di questa tesi è quello di modellare e prevedere la domanda dell'energia del mercato inglese, attraverso l'utilizzo di analisi univariate, multivariate e funzionali. Dato che il mercato inglese è un mercato *Day Ahead*, la previsione che si vuole fare sarà anch'essa riferita al breve periodo, cioè una previsione per i due giorni successivi.

Si vogliono confrontare i risultati derivanti dall'analisi funzionale, che considera le 48 osservazioni per ogni giorno come punti di un'unica curva, con i risultati derivanti da modelli univariati, che considerano le singole fasce semi orarie come singole serie, e da modelli multivariati. In tutti questi modelli si useranno variabili dipendenti quali la temperatura e la serie della domanda ritardata di vari periodi.

In ogni caso, prima di passare all'analisi vera e propria, in questo capitolo, si mostrano le caratteristiche principali della serie della domanda di energia, e si effettuano alcune analisi preliminari.

Nel primo paragrafo si analizzano le caratteristiche principali della serie storica della domanda. Nel secondo, invece, viene applicato un modello per la decomposizione di essa nelle sue parti deterministiche.

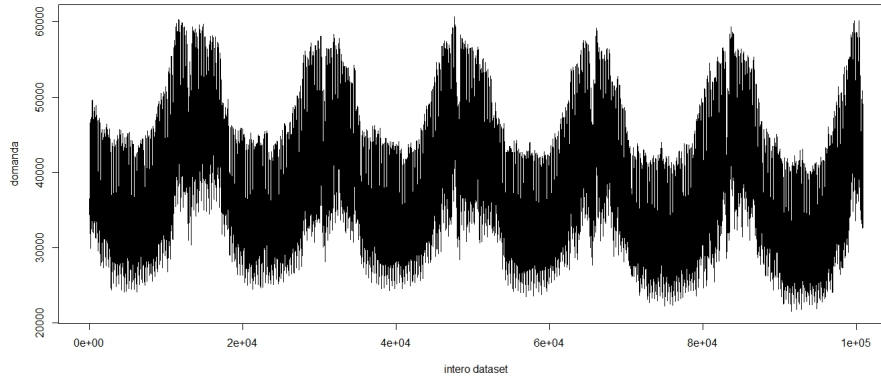


Figura 2.1: Grafico dell'intera serie, dal 1 Aprile 2005 al 31 Dicembre 2010.

## 2.1 Studio della serie storica della domanda

I dati presi in considerazione sono tratti dal mercato dell'energia inglese *England and Wales Electricity Pool*, in cui per ogni giorno si trovano 48 osservazioni che fanno riferimento alle 48 fasce semi orarie con le quali il mercato inglese è organizzato; la prima fascia semi oraria si riferisce al periodo 00.00-00.30, la seconda 00.30-01.00 e così via. I dati campionati vanno dal 1 Aprile 2005 al 31 Dicembre 2010, per un totale di 2101 osservazioni giornaliere.

Si definisce  $D_{t,i}$  la domanda di energia al giorno  $t$  ( $t = 1, 2, \dots, 2101$ ), per la  $i$ -esima fascia semi oraria ( $i = 1, 2, \dots, 48$ ).

In figura 2.1 si osserva l'andamento ciclico dell'intera domanda di energia, cioè senza considerare le fasce semi orarie. Se si prende in considerazione la serie della singola fascia semi oraria, si ha un'evidenza della presenza di componenti stagionali. Infatti, osservando la figura 2.2 che riporta l'intera serie della domanda, dal 1 Aprile 2005 al 31 Dicembre 2010, per la fascia relativa alle ore 09.30-10.00 (fascia di maggiore picco), si può notare un comportamento ciclico della serie per i diversi anni di osservazione. Si nota inoltre che sono presenti dei picchi, in corrispondenza di ogni fine anno, legati probabilmente alle vacanze natalizie. In figura 2.3 è riportata, invece, la serie della domanda nella fascia oraria relativa alle ore 3.30-4.00, di minor picco.



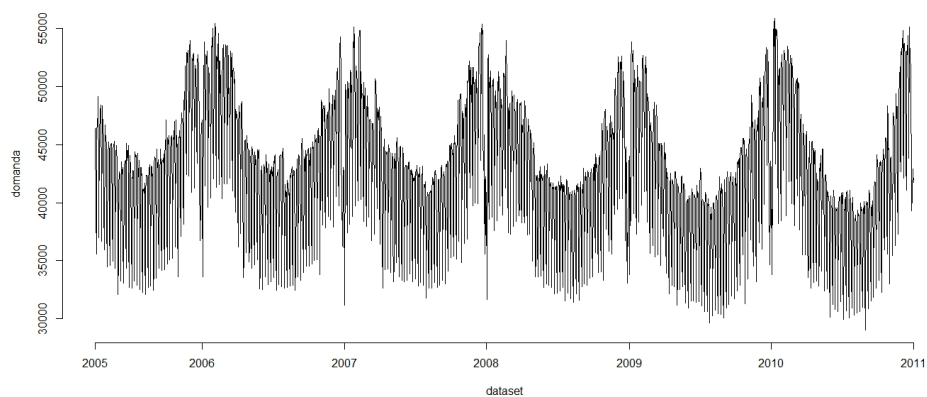


Figura 2.2: Serie della domanda, dal 1 Aprile 2005 al 31 Dicembre 2010, nella fascia oraria 09.30-10.00 (maggior picco).

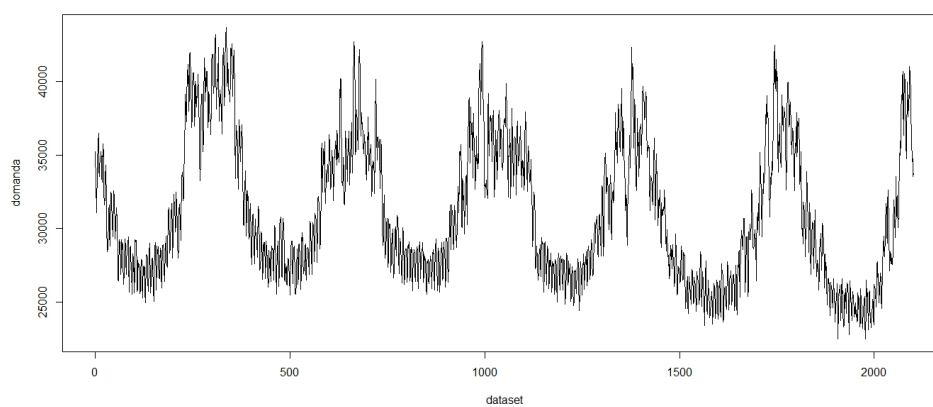


Figura 2.3: Serie della domanda, dal 1 Aprile 2005 al 31 Dicembre 2010, nella fascia oraria 3.30-4.00 (minor picco).

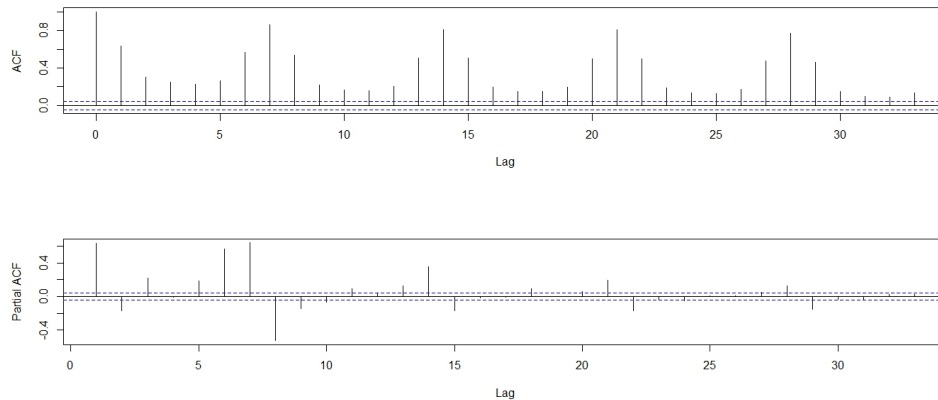


Figura 2.4: ACF e PACF della serie dal 1 Aprile 2005 al 31 Dicembre 2010 per la fascia di ore 09.30-10.00.

Queste caratteristiche sono meglio visibili attraverso l'autocorrelogramma e l'autocorrelogramma parziale della serie in una singola fascia semi oraria (figura 2.4), nell'ACF si può evidenziare il ritardo di ordine 7 legato alla componente settimanale.

Nel grafico 2.5, a pagina seguente, sono riportati i profili giornalieri della domanda per 100 giorni consecutivi. Dal grafico si evince che la domanda di energia sia relativamente più bassa nelle prime fasce semi orarie, che corrispondono alle ore notturne, e come sia più elevata nelle fasce centrali del giorno, nelle ore lavorative dalle 9 alle 20 circa. Inoltre, è interessante osservare, dalla figura 2.6, come la domanda di energia sia più elevata nei giorni lavorativi (linea continua) rispetto ai giorni non lavorativi (linea tratteggiata), come sabato e domenica, o eventuali giorni festivi infrasettimanali. La figura mostra la media dei primi 100 giorni lavorativi in linea continua, e la media dei primi 100 giorni non lavorativi in linea tratteggiata.

Un altro aspetto peculiare della domanda di energia, è il fatto che nei mesi invernali la domanda sia più elevata date le temperature più basse rispetto alle altre stagioni. E' stata calcolata una media dei mesi invernali per l'anno 2009-2010, una media dei mesi autunnali, primaverili e autunnali sempre nel-

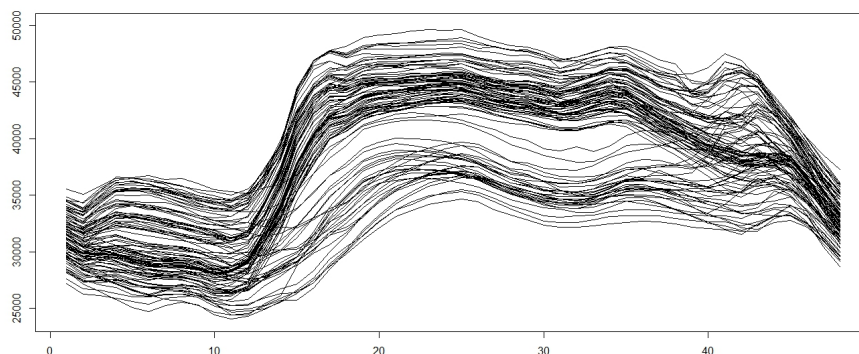


Figura 2.5: Grafico della domanda giornaliera per i primi 100 giorni. Nell'asse delle x si trovano le 48 fasce semi orarie, nell'asse delle y si trova l'ammontare della domanda di energia.

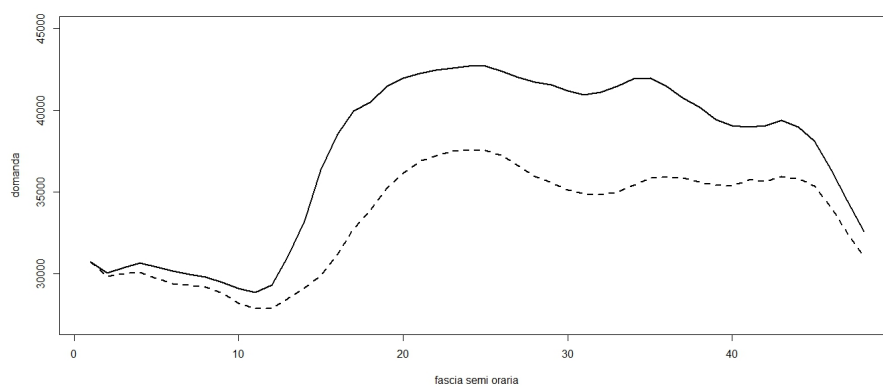


Figura 2.6: Profilo giornaliero medio della domanda di energia nei giorni lavorativi (linea continua) e nei giorni non lavorativi (linea tratteggiata).

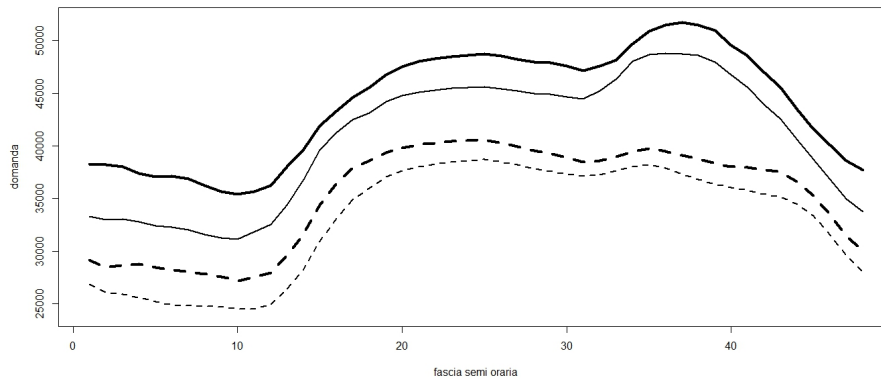


Figura 2.7: Profilo giornaliero medio della domanda di energia nelle diverse stagioni: inverno (linea continua spessa), autunno (linea continua leggera), primavera (linea tratteggiata spessa), estate (linea tratteggiata leggera).

lo stesso anno. Nella figura 2.7, si osserva la domanda media di energia nei mesi invernali (linea continua spessa), nei mesi autunnali (linea continua), nei mesi primaverili (linea tratteggiata spessa) e infine nei mesi estivi (linea tratteggiata).

Sono dunque presenti delle componenti, che andranno studiate, quali:

- una componente di lungo periodo che indica l'andamento generale della serie storica.
- Una componente settimanale, dato che la domanda di energia è legata al giorno della settimana. In particolare, ci si attende un consumo maggiore nei giorni feriali, specialmente nelle ore centrali del giorno, rispetto ai giorni festivi. Nella figura 2.8 si nota come nei giorni lavorativi (da lunedì a venerdì) la domanda abbia una mediana più elevata rispetto ai giorni non lavorativi, quali sabato e domenica.
- Una componente giornaliera, legata all'ora della giornata in cui ci si trova, generalmente nelle ore notturne la domanda è minore. Nella figura 2.9 si osserva la domanda di energia nelle diverse fasce semi orarie. Si

nota come nelle fasce centrali la mediana della domanda sia più alta rispetto alle prime fasce orarie corrispondenti alle fasce notturne.

- Una componente legata agli effetti di calendario. In questo caso sono compresi i giorni festivi che non cadono di sabato e domenica, nei quali la domanda di energia è minore.

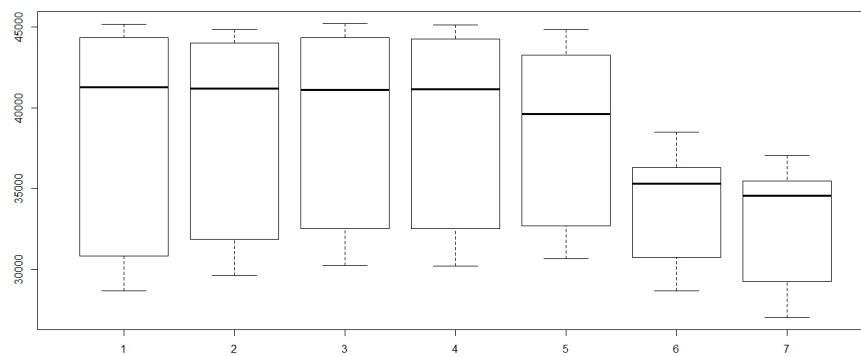


Figura 2.8: *Box-plot della domanda di energia nei giorni della settimana, da lunedì a domenica, dei primi 100 giorni di osservazione.*

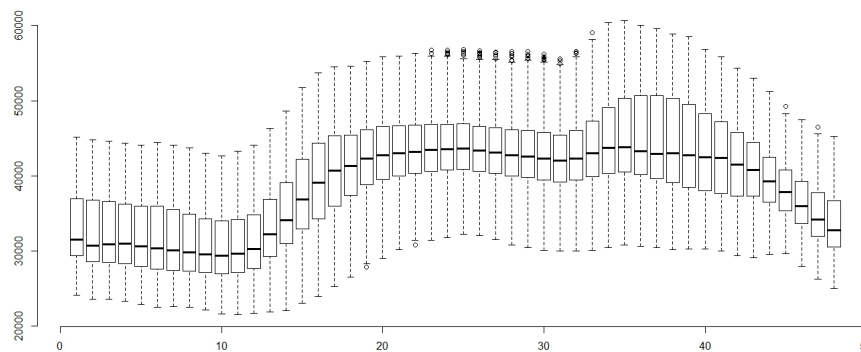


Figura 2.9: *Box-plot della domanda di energia nelle 48 fasce semi orarie.*

## 2.2 Decomposizione della serie della domanda

Vista la presenza di componenti deterministiche, è opportuno decomporre la serie della domanda in modo tale da costruire modelli di previsione su una serie depurata dai comportamenti descritti. Data, inoltre, la conformità dei dati a disposizione, si prendono in considerazione le serie semi orarie in modo separato, e ognuna verrà decomposta in ugual modo.

Si ipotizza che la serie della domanda logaritmica possa essere decomposta in un modello additivo come nel seguente modo:

$$\log D_{t,i} = L_{t,i} + S_{t,i} + F_{t,i} + d_{t,i} \quad (2.1)$$

dove  $L_{t,i}$  è la componente che contiene l'andamento di lungo periodo della serie, nel quale vi è la componente annuale,  $S_{t,i}$  è la componente periodica settimanale legata al giorno in cui ci si trova,  $F_{t,i}$  è la componente legata agli effetti di calendario e  $d_{t,i}$  è la componente stocastica. E' presente, inoltre, il logaritmo della domanda in modo tale da eliminare eventuali tracce di eteroschedasticità. Lavorando con modelli su ogni fascia semi oraria, intesa singola serie, si va a perdere la componente giornaliera, cioè quella legata all'ora del giorno in cui ci si trova, per questo motivo non viene considerata nel modello descritto sopra.

### 2.2.1 La componente di lungo periodo

Per stimare la componente di lungo periodo si è utilizzata una regressione non parametrica, in particolare le *splines* di lisciamento. Per definire le *splines* di lisciamento è necessario definire cosa sono le *splines* (per una documentazione più dettagliata si veda Azzalini e Scarpa (2004)).

Una *spline* è una funzione, costituita da un insieme di polinomi uniti tra loro, il cui scopo è interpolare in un intervallo un insieme di punti, detti nodi, in modo da essere continua, almeno fino ad un dato ordine di derivate, in ogni punto dell'intervallo. E' necessario scegliere il numero di nodi e la loro posizione; il vantaggio delle *splines* di lisciamento è quello di evitare la scelta e

il posizionamento dei nodi. Per effettuare una regressione non parametrica mediante *spline* di lisciamento si deve minimizzare la seguente funzione:

$$D(f, \lambda) = \sum_{i=1}^n [y_i - f(x_i)]^2 + \lambda \int [f''(x_i)]^2 dt \quad (2.2)$$

dove  $\lambda$  è un parametro positivo di penalizzazione del grado di irregolarità della curva  $f$ . Questo parametro è stato scelto, dopo varie prove, in modo da ottenere un giusto compromesso tra adattamento ai dati e lisciamento. Nella figura 2.10 è rappresentata la serie grezza e la stima della componente di lungo periodo (linea rossa).

Nella figura 2.11 si osserva la serie grezza del logaritmo della domanda, e il grafico della serie filtrata della componente di lungo periodo, ovvero  $\log D_{t,i} - \widehat{L}_{t,i}$ . In questa serie si notano dei valori estremi che potrebbero essere riferiti all' effetto di calendario.

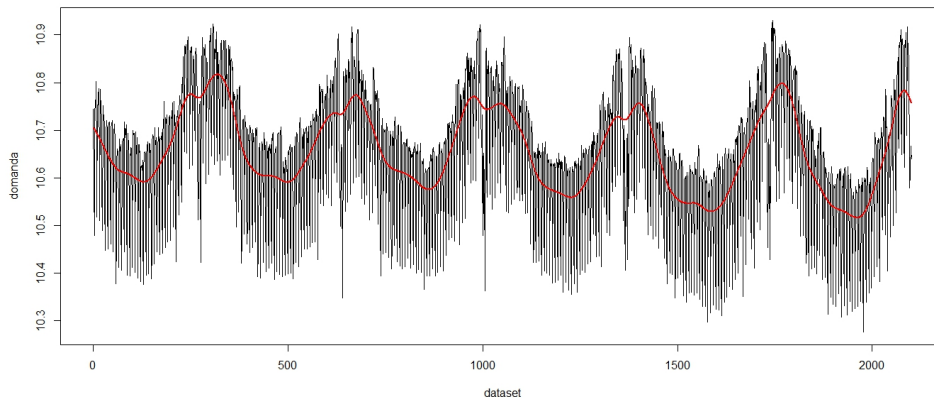


Figura 2.10: Stima della componente  $L_{t,i}$  per la fascia  $i = 20$  relativa alle ore 09.30-10.00.

## 2.2.2 La componente periodica

Dopo aver stimato la componente di lungo periodo,  $\widehat{L}_{t,i}$ , la serie residua è:

$$\log D_{t,i} - \widehat{L}_{t,i} = S_{t,i} + F_{t,i} + d_{t,i}. \quad (2.3)$$

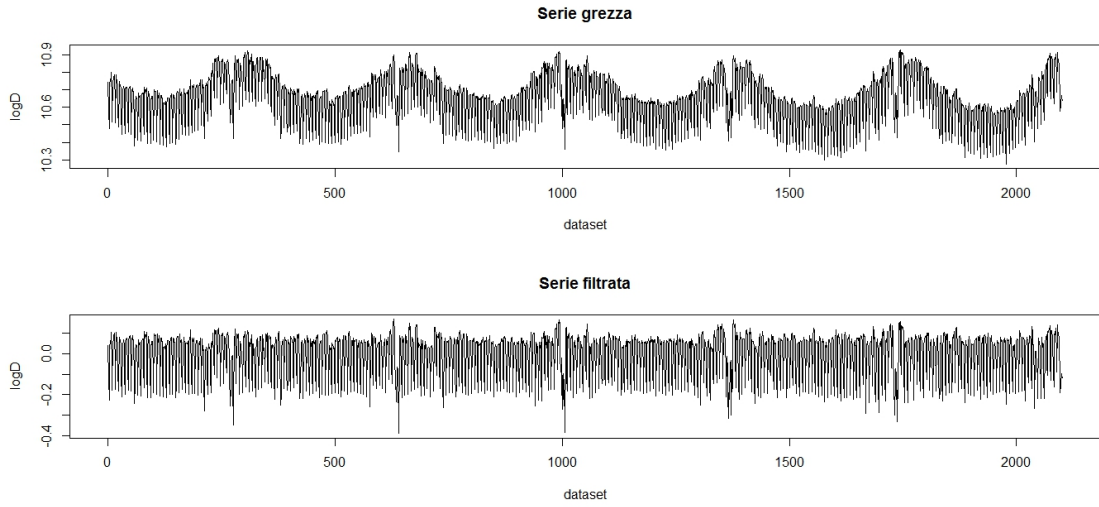


Figura 2.11: In alto serie grezza  $\log D_{t,i}$  della domanda per il periodo dal 1 Aprile 2005 al 31 Dicembre 2010; in basso la serie grezza,  $\log D_{t,i} - \hat{L}_{t,i}$  riferita allo stesso periodo.

Ora si deve depurare la serie anche dagli effetti stagionali e di calendario. Per fare questo verrà utilizzata una regressione lineare semplice con variabili *dummy* corrispondenti ai 7 giorni della settimana e una variabile *dummy* che modella gli effetti di calendario.

Il modello per la componente settimanale è il seguente:

$$S_{t,i} = \gamma_1 d_{lun,t} + \gamma_2 d_{mart,t} + \gamma_3 d_{merc,t} + \gamma_4 d_{giov,t} + \gamma_5 d_{ven,t} + \gamma_6 d_{sab,t} + \gamma_7 d_{dom,t} \quad (2.4)$$

dove  $d_{lun,t}$  vale 1 se la  $t$ -esima osservazione è lunedì e 0 altrimenti,  $d_{mart,t}$  vale 1 se la  $t$ -esima osservazione è martedì e 0 altrimenti, e così via per le restanti *dummy*. La specificazione per gli effetti di calendario è:

$$F_{t,i} = \delta_1 d_{festivo,t} + \delta_2 d_{natale,t} \quad (2.5)$$

dove  $d_{festivo,t}$  vale 1 se la  $t$ -esima osservazione è un giorno di festa infrasettimanale e 0 altrimenti,  $d_{natale,t}$  vale 1 se la  $t$ -esima osservazione è compresa



nel periodo natalizio dal 25 Dicembre all' 1 Gennaio e 0 altrimenti.

Quindi si avrà:

$$\log D_{t,i} - \hat{L}_{t,i} = \gamma_1 d_{lun,t} + \gamma_2 d_{mart,t} + \gamma_3 d_{merc,t} + \gamma_4 d_{giove,t} + \gamma_5 d_{ven,t} + \\ + \gamma_6 d_{sab,t} + \gamma_7 d_{dom,t} + \delta_1 d_{festivo,t} + \delta_2 d_{natale,t} + d_{t,i}$$

In questa equazione manca il termine costante, per evitare multicollinearità. La componente stagionale settimanale e gli effetti di calendario vengono quindi stimati con il metodo dei minimi quadrati, ottenendo così la serie depurata di tutte le componenti deterministiche.

Dal grafico 2.12 della serie destagionalizzata si notano anche dei valori estremi, che non sono stati catturati in maniera adeguata dal modello descritto. Nella figura 2.13 è riportata, invece, la serie detrendizzata  $\log D_{t,i} - \hat{L}_{t,i}$  in color nero, e la serie destagionalizzata in color verde.

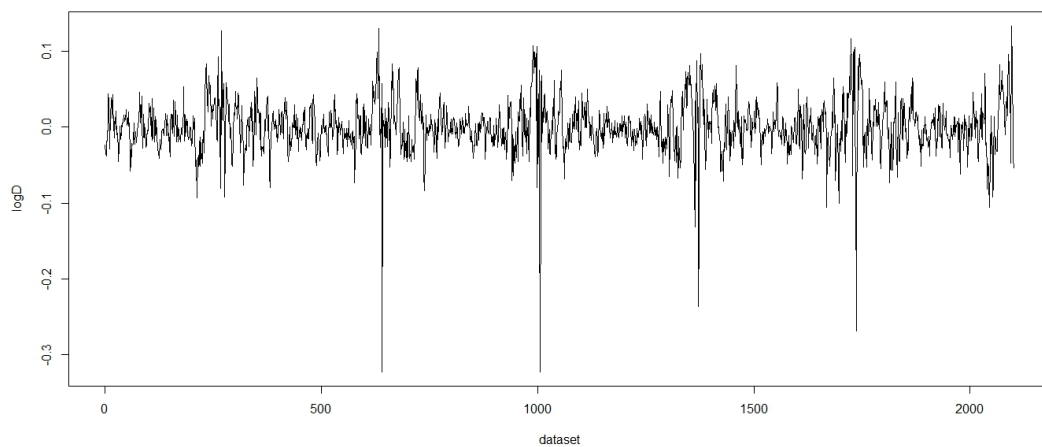


Figura 2.12: Serie della domanda destagionalizzata nel periodo dal 1 Aprile 2005 al 31 Dicembre 2010.

### 2.2.3 La componente stocastica

Si è così stimata, per ogni fascia oraria, una componente di lungo periodo, una componente stagionale settimanale e gli effetti di calendario, ottenendo

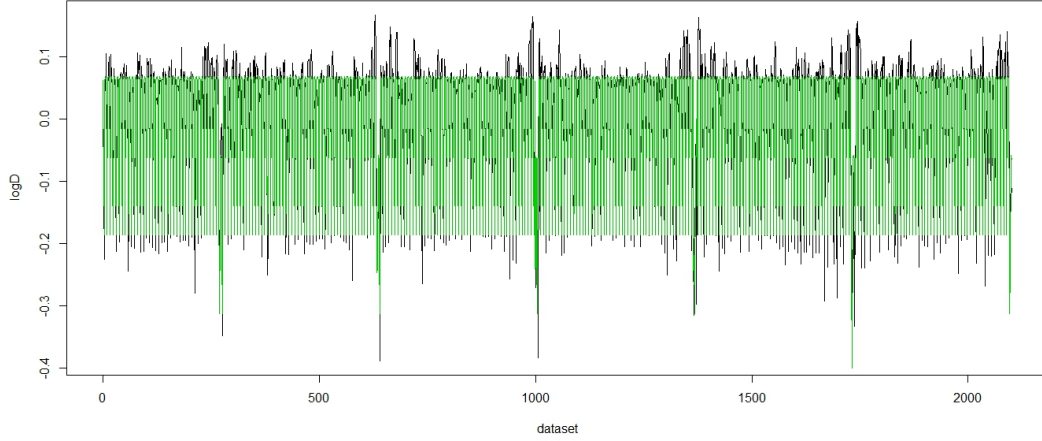


Figura 2.13: Stima della componente periodica (colore verde) e serie della domanda  $\log D_{t,i} - \widehat{L}_{t,i}$  (colore nero).

la serie depurata da tutte le componenti deterministiche:

$$d_{t,i} = \log D_{t,i} - \widehat{L}_{t,i} + \widehat{S}_{t,i} + \widehat{F}_{t,i} \quad (2.6)$$

La serie  $\widehat{d}_{t,i}$  che ne risulta è riportata in figura 2.12.

Da una prima analisi grafica le serie depurate, per ogni fascia semi oraria, sembrano avere una qualche struttura di autocorrelazione, come si vede dal grafico dell'autocorrelazione parziale in figura 2.14, a pagina seguente. Sarà quindi opportuno applicare dei modelli che catturino questa forma di autocorrelazione come, ad esempio, modelli autoregressivi.

Per prevedere la domanda  $D_{t+1}$  e  $D_{t+2}$  utilizzeremo diversi modelli, che nei capitoli seguenti verranno descritti nel dettaglio.

La previsione un passo sarà data da:

$$\widehat{D}_{t+1,i} = \exp(\widehat{L}_{t+1,i} + \widehat{S}_{t+1,i} + \widehat{F}_{t+1,i} + \widehat{d}_{t+1,i}) \quad t = 1737, \dots, 2101 \quad (2.7)$$

dove  $\widehat{d}_{t+1,i}$  è la previsione della componente stocastica, per la  $i$ -esima fascia semi oraria. La previsione delle componenti deterministiche è facilmente calcolabile. La componente settimanale,  $\widehat{S}_{t+1}$ , è nota, conoscendo a priori

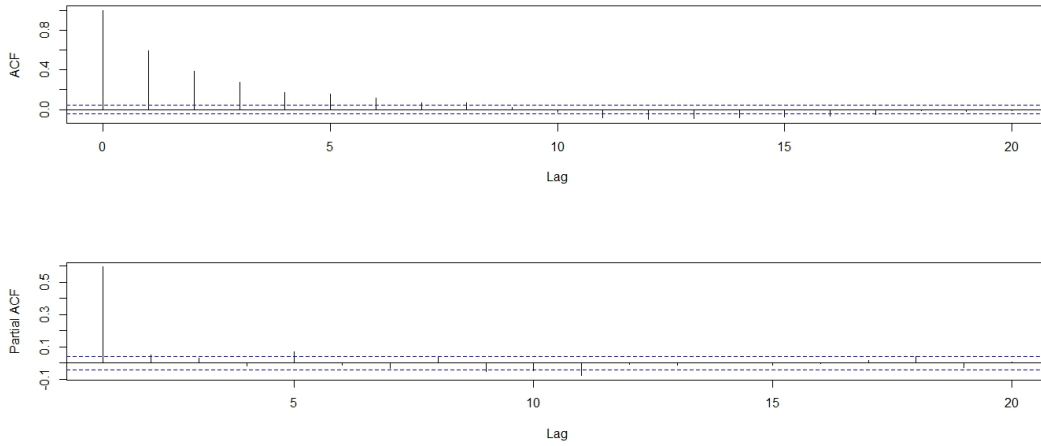


Figura 2.14: ACF e PACF della serie dei residui dal 1 Aprile 2005 al 31 Dicembre 2010 per la fascia semi oraria 09.30-10.00.

il giorno di cui si vuole fare la previsione; lo stesso vale per la componente legata agli effetti di calendario,  $\widehat{F}_{t+1}$ , visto che sono noti a priori i giorni in cui ci saranno feste infrasettimanali. Per la componente di lungo periodo  $\widehat{L}_{t+1}$ , il valore per il passo successivo  $t + 1$  si può presupporre uguale al valore del giorno prima  $t$ , cioè:  $\widehat{L}_{t+1} = \widehat{L}_t$ . E' un'ipotesi semplificatrice ma si può assumere vera in quanto il livello della serie cambia lentamente e si presume non vari di molto da un giorno a quello successivo. La previsione due passi in avanti sarà, invece:

$$\widehat{D}_{t+2,i} = \exp(\widehat{L}_{t+2,i} + \widehat{S}_{t+2,i} + \widehat{F}_{t+2,i} + \widehat{d}_{t+2,i}) \quad t = 1737, \dots, 2101 \quad (2.8)$$

dove  $\widehat{d}_{t+2,i}$  è la previsione a due passi della componente stocastica per la  $i$  -esima fascia semi oraria. Analogamente alla previsione ad un passo le componenti  $\widehat{S}_{t+2}$  e  $\widehat{F}_{t+2}$  sono note, conoscendo a priori il giorno di cui si vuole fare la previsione. Per la componente di lungo periodo si assume, invece, che sia uguale al valore di due giorni prima; sempre dato dal fatto che il livello della serie cambia lentamente, quindi si avrà  $\widehat{L}_{t+2} = \widehat{L}_t$ .

Concludendo, in questo capitolo, si è ottenuta per ogni fascia semi oraria, la serie depurata dalle componenti deterministiche, cioè  $\widehat{d}_{t,i}$ . Su tale serie, ora,

verranno costruiti dei modelli di previsione. Successivamente si confronteranno i risultati ottenuti da questi modelli e si identificheranno i modelli migliori per ottenere la previsione un passo e due passi in avanti per la serie della domanda  $D_{t,i}$ .

## Capitolo 3

# Modelli univariati, multivariati e funzionali

Nel capitolo precedente si è ottenuta la serie della componente stocastica,  $d_{t,j}$ . Lo scopo, ora, è di studiare questa serie attraverso diverse tipologie di modelli, come detto in precedenza, ed identificare quelli migliori con i quali fare le previsioni, nel breve periodo, della domanda di energia.

In questo capitolo verranno considerati alcuni modelli di regressione, univariati, multivariati e funzionali, con diverse specificazioni.

Negli articoli già citati, in particolare l'articolo di Blum e Riedmiller (2013), i dati forniti dal meteo sono spesso utilizzati per fare delle previsioni della domanda nel breve periodo; per questo motivo è interessante introdurre nelle analisi la variabile temperatura.

Nella prima parte del capitolo si prendono in considerazione modelli più semplici, come i modelli univariati. Si descrive, quindi, un modello autoregressivo di ordine  $p$  e un modello autoregressivo con l'introduzione di una variabile esogena.

Analogamente, nella seconda parte del capitolo, si descrivono due tipologie di modelli multivariati: un modello autoregressivo vettoriale di ordine  $p$  e un modello autoregressivo vettoriale con l'introduzione di una variabile esogena.

Infine, nell'ultima parte, si introduce brevemente l'analisi funzionale, e si definiscono i modelli di regressione funzionali.

### 3.1 Modelli univariati

Per identificare i modelli previsivi, ci si concentra sulla serie  $d_{t,i}$  ottenuta in precedenza.

Inizialmente si possono considerare modelli più semplici, come i modelli univariati.

Infatti, la serie della domanda inglese è composta da 48 fasce semi orarie; se si considerano queste fasce come singole serie, si può applicare ad ognuna, un modello univariato.

#### 3.1.1 Modelli autoregressivi AR(p)

Si considera un modello autoregressivo di ordine  $p$ :

$$d_{t,i} = \beta_{0,i} + \sum_{p=1}^P \beta_{p,i} d_{t-p,i} + a_{t,i} \quad (3.1)$$

in cui  $d_{t,i}$  rappresenta la domanda di energia per la  $i$  -esima fascia semi oraria al tempo  $t$ ,  $\beta_{p,i}$  sono i coefficienti relativi al  $p$  -esimo ritardo della serie  $d_{t,i}$  e  $a_{t,i}$  è l'errore a media zero.

Si può scrivere questo modello in forma più compatta, introducendo l'operatore ritardo  $L = L^p d_{t-p,i}$

$$\beta_p(L) d_{t,i} = a_{t,i} \quad (3.2)$$

dove  $\beta_p(L) = 1 - \beta_1 L - \beta_2 L^2 - \dots - \beta_p L^p$ .

#### 3.1.2 Modelli ARX(p)

E' lecito pensare che la domanda di energia elettrica sia legata alle temperature che vi sono nel paese.

Come si è detto, molti articoli che prevedono la domanda introducono nei

vari modelli, le informazioni date dalle stazioni metereologiche di un paese (si veda Blum e Riedmiller (2013), citato in precedenza).

E' consueto che, con temperature basse la domanda di energia aumenti e, con temperature più miti la domanda diminuisca. Per questo motivo è interessante verificare se le temperature influenzino o meno la domanda di energia. Avendo a disposizione la serie della domanda di energia del Regno Unito è auspicabile avere la serie delle temperature dello stesso luogo, e quindi si utilizzeranno le temperature di diverse città. Si è cercato di coprire le varie regioni dell'Inghilterra e quindi si avrà la serie delle temperature medie di Londra, Manchester, Edinburgo e Galles. Come è evidente, non si ha a disposizione la serie delle temperature per ognuna delle 48 fasce semi orarie e, infatti, è stata presa in considerazione la serie delle temperature medie, mantenuta costante per tutto il giorno.

Prima di esplicitare il modello di regressione, si analizzano le caratteristiche delle temperature.

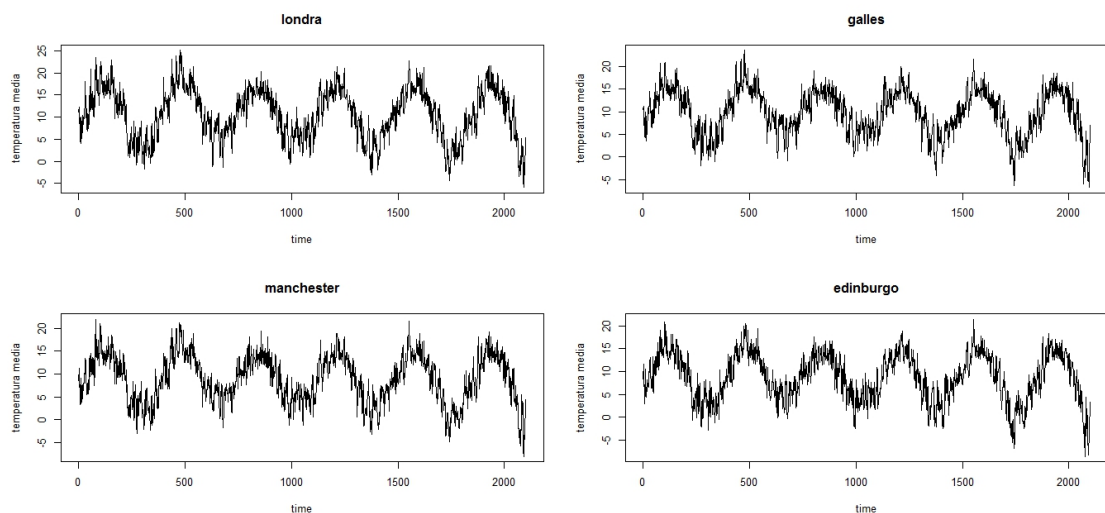


Figura 3.1: Temperature medie giornaliere dal 1 Aprile 2005 al 31 Dicembre 2010 nelle diverse città. In alto a sinistra Londra, a destra Galles, in basso a sinistra Manchester e a destra Edinburgo.

Nella figura 3.1 è raffigurato l'andamento delle temperature delle quattro città scelte; si può notare come le quattro serie abbiano un andamento del tutto simile alla domanda di energia.

In media, la temperatura minima di Londra è attorno ai  $-6^{\circ}\text{C}$  e la massima arriva a  $25^{\circ}\text{C}$ ; quella minima media del Galles si aggira intorno ai  $-8^{\circ}\text{C}$  e la massima invece è  $24^{\circ}\text{C}$ ; per quanto riguarda la temperatura minima media di Manchester è  $-8^{\circ}\text{C}$  e la massima  $22^{\circ}\text{C}$ ; infine la minima media di Edinburgo è attorno ai  $-9^{\circ}\text{C}$  e la massima  $22^{\circ}\text{C}$ .

Si può dire, quindi, che le temperature minime e massime si aggirano intorno agli stessi gradi per tutte e quattro le città; questo si può notare meglio dalla figura 3.2, nella quale sono rappresentate, per ogni stagione, le temperature medie. In colore nero è rappresentata Londra, in blu il Galles, in rosso il Manchester e in verde Edinburgo.

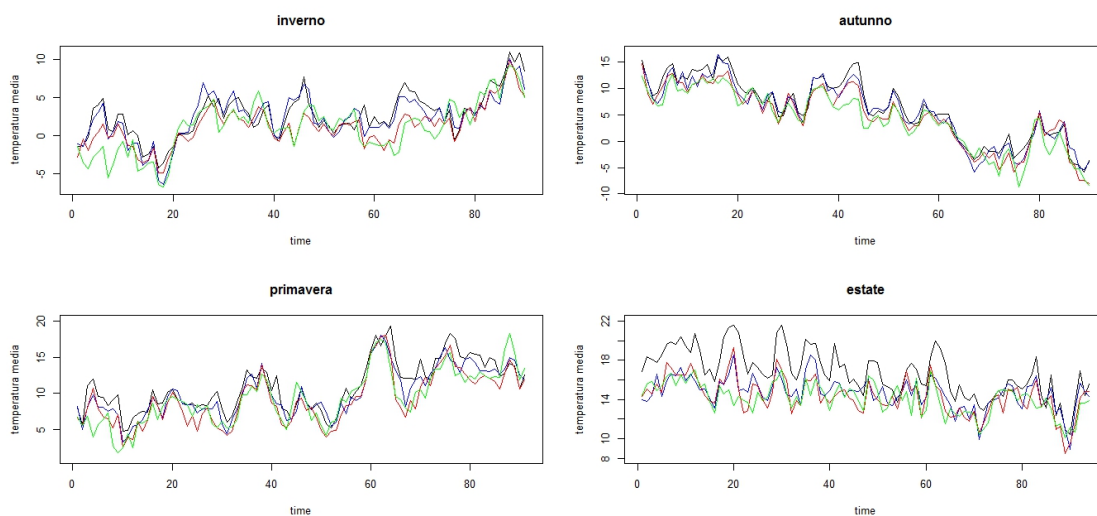


Figura 3.2: Temperature medie delle quattro stagioni (2009-2010) per Londra (nero), Galles (blu), Manchester (rosso) e Edinburgo (verde).

Nella figura 3.3 sono riportate le correlazioni tra le diverse serie delle temperature; si osserva come tra esse ci sia una relazione forte ma si ha la confer-



ma guardando i valori della correlazione (tabella 3.1).

	Londra	Edinburgo	Manchester	Galles
Londra	1	0.93	0.97	0.97
Edinburgo	0.93	1	0.96	0.94
Manchester	0.97	0.96	1	0.97
Galles	0.97	0.94	0.97	1

Tabella 3.1: *Matrice di correlazione delle temperature.*

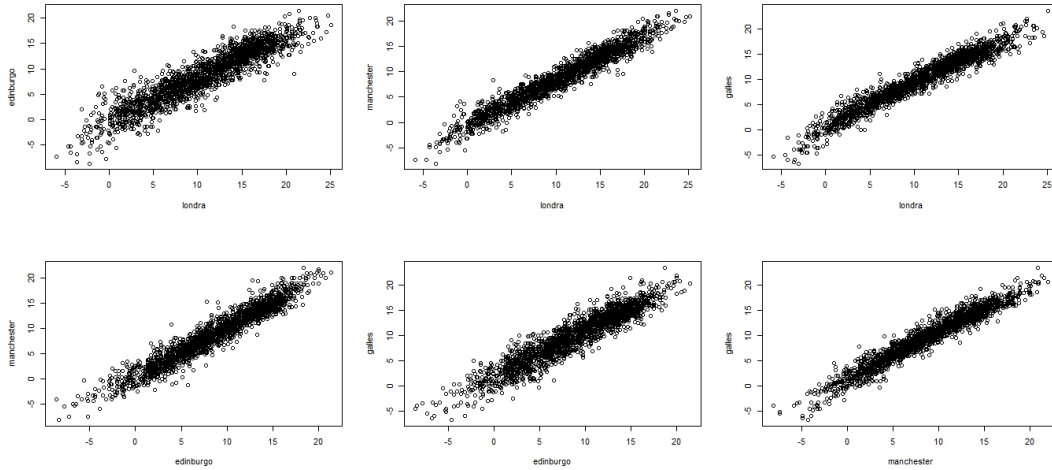


Figura 3.3: *Grafico di dispersione. Da sinistra a destra, in alto, si trova il grafico di dispersione tra Londra-Edinburgo, Londra-Manchester, Londra-Galles. In basso si trova Edinburgo-Manchester, Edinburgo-Galles e Manchester-Galles.*

Stimando i modelli autoregressivi con l'introduzione delle quattro serie delle temperature, come variabili esogene, si nota che soltanto una città risulta significativa, come era lecito attendersi data la correlazione tra esse. Si prende quindi, come variabile esogena, la media delle diverse serie delle temperature. Si stima quindi il seguente modello:

$$d_{t,i} = \beta_{0,i} + \sum_{p=1}^P \beta_{p,i} d_{t-p,i} + \sum_{j=1}^J \gamma_{j,i} temp_{t-q,i} + a_{t,i} \quad (3.3)$$

in cui  $d_{t,i}$  rappresenta la domanda di energia per la  $i$  –esima fascia semi oraria al tempo  $t$ ,  $\beta_{p,i}$  sono i coefficienti relativi al  $p$  –esimo ritardo della serie  $d_{t,i}$ ,  $\gamma_{j,i}$  sono i coefficienti relativi al  $j$  –esimo ritardo della serie delle temperatura e  $a_{t,i}$  è l'errore a media zero.

## 3.2 Modelli multivariati

Nel precedente paragrafo si sono descritte diverse specificazioni per il modello di regressione lineare per ogni singola fascia semi oraria.

In questo paragrafo si considerano le singole fasce semi orarie come una serie multivariata e si descrive un modello autoregressivo vettoriale con  $p$  ritardi e un modello autoregressivo vettoriale con l'introduzione della variabile esogena, la temperatura media.

### 3.2.1 Modelli VAR(p)

Il modello VAR(p) è:

$$d_t = \Phi_0 + \Phi_p d_{t-p} + a_t \quad (3.4)$$

dove  $d_t = (d_{1,t}, \dots, d_{K,t})'$  è un vettore ( $K \times 1$ ),  $\Phi_p$  è una matrice di coefficienti ( $K \times K$ ),  $\Phi_0$  è un vettore ( $K \times 1$ ) di termini di intercetta e  $a_t = (a_{1,t}, \dots, a_{K,t})' \sim WN(0, \Sigma)$ , con  $\Sigma$  singolare.

Esplicitando il modello nel caso  $K = 48$  e  $p = 1$ , si avrà:

$$\begin{bmatrix} d_{1,t} \\ d_{2,t} \\ \dots \\ d_{48,t} \end{bmatrix} = \begin{bmatrix} \phi_{1,0} \\ \phi_{2,0} \\ \dots \\ \phi_{48,0} \end{bmatrix} + \begin{bmatrix} \phi_{1,1} & \phi_{1,2} & \dots & \phi_{1,48} \\ \phi_{2,1} & \phi_{2,2} & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \phi_{48,1} & \dots & \dots & \phi_{48,48} \end{bmatrix} \begin{bmatrix} d_{1,t-1} \\ d_{2,t-1} \\ \dots \\ d_{48,t-1} \end{bmatrix} + \begin{bmatrix} a_{1,t} \\ a_{2,t} \\ \dots \\ a_{48,t} \end{bmatrix},$$

$$\text{con } a_t \sim WN(0, \Sigma) \text{ e } \Sigma = \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} & \dots & \sigma_{1,48} \\ \sigma_{2,1} & \sigma_{2,2} & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \sigma_{48,1} & \sigma_{48,2} & \dots & \sigma_{48,48} \end{bmatrix}.$$

Esplicitando ulteriormente si ottiene:

$$d_{i,t} = \phi_{i,0} + \phi_{i,1}d_{1,t-1} + \phi_{i,2}d_{2,t-1} + \dots + \phi_{i,48}d_{48,t-1} + a_{i,t}$$

con  $i = 1, \dots, 48$ .

### 3.2.2 Modelli VARX(p)

Si procede in parallelo alle diverse specificazioni del modello di regressione lineare, e dopo l'introduzione della variabile prezzo, si descrive un VARX con la serie della temperatura medie del Regno Unito come variabile esogena.

Il modello stimato VARX(p) è:

$$d_t = \Phi_0 + \Phi_p d_{t-1} + \Gamma_j temp_{t-1} + a_t \quad (3.5)$$

dove  $d_t = (d_{1,t}, \dots, d_{K,t})'$  è un vettore ( $K \times 1$ ),  $K = 1, \dots, 48$ ,  $\Phi_p$  è una matrice di coefficienti ( $K \times K$ ),  $\Phi_0$  è un vettore ( $K \times 1$ ) di termini di intercetta,  $temp_{t-1} = (temp_{1,t-1}, \dots, temp_{K,t-1})'$  è un vettore ( $K \times 1$ ) della variabile esogena della temperatura media,  $\Gamma_j$  è una matrice di coefficienti ( $K \times 1$ ) e  $a_t = (a_{1,t}, \dots, a_{K,t})' \sim WN(0, \Sigma)$ , con  $\Sigma$  singolare.

Esplicitando il modello nel caso  $K = 48$ ,  $p = 1$  e  $j = 1$  si avrà:

$$\begin{bmatrix} d_{1,t} \\ d_{2,t} \\ \dots \\ d_{48,t} \end{bmatrix} = \begin{bmatrix} \phi_{1,0} \\ \phi_{2,0} \\ \dots \\ \phi_{48,0} \end{bmatrix} + \begin{bmatrix} \phi_{1,1} & \phi_{1,2} & \dots & \phi_{1,48} \\ \phi_{2,1} & \phi_{2,2} & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \phi_{48,1} & \dots & \dots & \phi_{48,48} \end{bmatrix} \begin{bmatrix} d_{1,t-1} \\ d_{2,t-1} \\ \dots \\ d_{48,t-1} \end{bmatrix} +$$

$$\begin{bmatrix} \gamma_{1,1} \\ \gamma_{2,1} \\ \dots \\ \gamma_{48,1} \end{bmatrix} \begin{bmatrix} temp_{1,t-1} & temp_{2,t-1} & \dots & temp_{48,t-1} \end{bmatrix} + \begin{bmatrix} a_{1,t} \\ a_{2,t} \\ \dots \\ a_{48,t} \end{bmatrix},$$

con  $a_t \sim WN(0, \Sigma)$  e  $\Sigma = \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} & \dots & \sigma_{1,48} \\ \sigma_{2,1} & \sigma_{2,2} & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \sigma_{48,1} & \sigma_{48,2} & \dots & \sigma_{48,48} \end{bmatrix}$ .

Esplicitando ulteriormente si ottiene:

$$d_{i,t} = \phi_{i,0} + \phi_{i,1}d_{1,t-1} + \phi_{i,2}d_{2,t-1} + \dots + \phi_{i,48}d_{48,t-1} + \gamma_{i,1}t + a_{i,t}$$

con  $i = 1, \dots, 48$ .

### 3.3 Modelli funzionali

I problemi che si possono affrontare mediante l'analisi funzionale sono di vario tipo, ad esempio l'analisi della temperatura mensile rilevata in diverse stazioni meteo, oppure la variazione di altezza di un campione di persone nel tempo, o ancora la variazione di un indice economico nel tempo (si veda Ramsey e Silverman, 2002). Il fattore comune tra gli esempi include, quindi, una misurazione effettuata in diversi istanti temporali.

Diversamente dall'analisi basata sulle serie storiche, l'analisi funzionale presuppone un dominio continuo per le sue unità statistiche, nel senso che ad ogni soggetto viene associata una funzione continua nel tempo della variabile d'interesse.

Dunque, la caratterizzazione dell'approccio funzionale risiede nel fatto di considerare ogni funzione osservata come un unico oggetto e non come una serie di osservazioni singole.

Poichè il dominio di osservazione è continuo, ma l'osservazione del dato e il campionamento avviene in maniera discreta, è necessaria una qualche forma di sintetizzazione dei dati, come delle interpolazioni o dei lisciami dei valori discreti.

Una procedura consiste nell'approssimare una funzione combinando linearmente  $K$  funzioni note, dette basi:

$$d(t) = \sum_{k=1}^K c_k \phi_k(t) \quad (3.6)$$

dove  $c_k$  sono i coefficienti, determinati con i minimi quadrati ordinari, e  $\phi_k(t)$  le funzioni base.

Ci sono diversi tipi di funzione base, ed è importante scegliere quella che si

adatta meglio ai dati a disposizione. In letteratura sono state usate funzioni esponenziali, poligonali, le potenze, le *splines*, le funzioni costanti, e tante altre. La scelta del tipo di funzione base più adatta al problema non è l'obiettivo di questa tesi (per una trattazione completa dell'argomento si veda Ramsey e Silverman, 2002).

Le funzioni base di Fourier sono le più conosciute e, per la loro struttura, si adattano bene ai dati periodici, si è infatti notato come la serie della domanda di energia abbia un andamento ripetitivo nel tempo.

Le funzioni base di Fourier sono definite a partire dalla serie di Fourier:

$$x(t) = c_0 + c_1 \sin(\omega t) + c_2 \cos(\omega t) + c_3 \sin(2\omega t) + c_4 \cos(2\omega t) + \dots \quad (3.7)$$

con le basi:

$$\begin{aligned} \phi_1(t) &= 1 \\ \phi_2(t) &= \sin(\omega t) \\ \phi_3(t) &= \cos(\omega t) \\ \phi_4(t) &= \sin(2\omega t) \\ \phi_5(t) &= \cos(2\omega t) \\ &\dots \end{aligned}$$

dove  $\omega = 2\pi/T$  è una costante relativa al periodo  $T$  di osservazioni (il periodo è l'intervallo di tempo necessario a compiere un'oscillazione completa della curva).

Per definire il sistema delle funzioni base di Fourier è necessario conoscere il numero di  $K$  funzioni base, generalmente dispari, e il periodo, in questo caso  $T = 48$ . Il numero di funzioni  $K$  non dev'essere eccessivamente alto, ed è sempre scelto in maniera arbitraria in modo da avere un giusto compromesso tra lisciamiento e adattabilità ai dati.

Un altro tipo di funzioni base conosciuto sono le *B-Splines*, caso particolare delle funzioni *spline*. E' necessario spiegare in che cosa consistono quest'ultime per comprendere meglio le *B-Splines*.

Le funzioni base *splines* sono più flessibili e computazionalmente più veloci,

vengono costruite unendo funzioni polinomiali in punti  $\tau_t$  detti nodi. Il numero di nodi è scelto in modo arbitrario o con delle procedure più specifiche (si veda nuovamente Azzalini e Scarpa per una conoscenza più approfondita). Il numero di basi è definito in relazione al numero di nodi scelti attraverso la relazione numero di basi = grado del polinomio + numero di nodi interni; nel numero di nodi interni sono esclusi i nodi posizionati all'inizio e alla fine della funzione. All'interno dei nodi, la funzione polinomiale è di grado  $d$ , e in corrispondenza di un nodo interno è richiesto che i polinomi adiacenti assumano gli stessi valori per un fissato numero di derivate, in genere  $n - 1$ . Solitamente il grado scelto è pari a tre, in modo da garantire la continuità delle prime due derivate ed ottenere una curva liscia.

Le *B-Splines* sono un caso particolare perchè assumono valore 0 ovunque tranne in un intervallo definito. Anche in questo caso la scelta più comune ricade su delle *B-Splines* cubiche nell'intervallo  $(\tau_{K-2}, \tau_{K+2})$ . La selezione di dove posizionare i nodi è arbitraria e per lo più è dettata dall'esperienza.

Nell'analisi funzionale esistono tre tipologie di regressione lineare:

1. il modello funzionale-scalare in cui la variabile risposta è funzionale e le covariate sono scalari;
2. il modello scalare-funzionale in cui la variabile risposta è scalare o multivariata, e le covariate sono funzionali;
3. il modello funzionale-funzionale in cui sia la variabile risposta, sia le covariate sono funzionali.

Per l'obiettivo della tesi, si prenderà in considerazione l'ultima tipologia, detta *concurrent linear model*; in quanto la variabile risposta,  $d$ , e le variabili esplicative,  $z$ , sono entrambe funzioni di  $i$ , e l'influenza è contemporanea in quanto  $z$  influenza  $d(i)$  attraverso il suo valore  $z(i)$  nel dato  $i$ . Generalmente la variabile esplicativa potrebbe influenzare  $y(i)$  nel valore  $z(h)$  con  $i \neq h$ .

Si definisce, quindi, il seguente modello:

$$d_t(i) = \beta_0(i) + \sum_{j=1}^q z_{tj}(i)\beta_j(i) + \epsilon_t(i) \quad (3.8)$$

dove  $d_t(i)$  è la serie stocastica della domanda resa funzionale,  $i$  è l'oggetto funzionale, in questo caso  $i = 48$  corrispondente alle fasce semi orarie;  $\beta_0(t)$  è il termine d'intercetta, anch'esso funzionale,  $z_{tj}(i)$  è la variabile esplicativa funzionale,  $j$  è il numero di funzioni base scelte per trasformare la variabile in dato funzionale,  $\beta_j(i)$  sono i coefficienti funzionali e  $\epsilon_t(i)$  è l'errore.

Il parametro  $\beta = (\beta_1, \dots, \beta_q)$  viene stimato attraverso i minimi quadrati, come nell'analisi di regressione lineare classica. Si stimerà una funzione base per ogni funzione di regressione  $\beta_j$  introducendo una misura di irregolarità per controllare il grado di lisciamento e identificare in modo unico la funzione  $\beta_j$ . La misura di irregolarità è definita da:

$$PEN_j(\beta_j) = \lambda_j \int [L_j \beta_j(t)]^2 dt \quad (3.9)$$

dove  $L_j$  è un operatore differenziale lineare appropriato per il parametro funzionale. Generalmente è definito come la derivata seconda oppure l'accelerazione armonica nel caso soprattutto di funzioni base di Fourier, e il parametro  $\lambda$  definisce il *trade-off* tra variabilità e la misura di irregolarità. Anche in questo caso *lambda* può essere scelto in modo arbitrario o attraverso tecniche di convalidazione incrociata.

La stima di  $\beta$  sarà data dal valore che minimizza:

$$LMMSE(\beta) = \int r(t)'r(t)dt + \sum_j^p \lambda_j \int [L_j \beta_j(t)]^2 dt \quad (3.10)$$

dove  $r(t) = y_t(i) - (\beta_0(i) + \sum_{j=1}^q z_{tj}(i)\beta_j(i))$ .





# Capitolo 4

## Applicazione ai dati

Nel precedente capitolo si sono descritti, a livello teorico, i modelli univariati, multivariati e funzionale; ora si vuole procedere con l'applicazione degli stessi, ai dati a disposizione.

Questo capitolo si divide principalmente in due sezioni: una prima sezione, nella quale vengono stimati ed identificati i modelli per la serie  $d_{t,i}$  attraverso le previsioni *in-sample*; e una parte nella quale vengono condotte le previsioni un passo e due passi in avanti per la serie della domanda di energia,  $D_{t,i}$ , denominata previsioni *out-of-sample*.

E' utile, quindi, suddividere la serie in tre periodi di tempo. Il primo periodo è necessario per la stima dei diversi modelli; il secondo per confrontare le previsioni e identificare i modelli migliori che vengono confrontati, infine, sulla base del terzo periodo. Il primo periodo è composto da 1371 osservazioni, ovvero dal 1 Aprile 2005 al 31 Dicembre 2008; il secondo periodo ha una lunghezza pari a 365 giorni, cioè l'intero anno 2009, e infine l'ultimo periodo rappresenta l'anno 2010.

## 4.1 Previsioni in-sample

In questa parte del capitolo si riportano i risultati delle stime e delle previsioni di tutti i modelli, condotte sui primi due periodi di osservazioni, denominati *in-sample*.

Per confrontare i risultati e successivamente scegliere il modello previsivo per  $D_{t,i}$ , si definiscono degli indicatori di bontà delle previsioni:

- MAE, errore medio assoluto:  $\frac{1}{M} \sum_{t=1}^M |e_t|$ ;
- MAPE, errore medio assoluto percentuale:  $\frac{1}{M} \sum_{t=1}^M \left| \frac{e_t}{d_t} \right|$ ;
- RMSQ, radice dell'errore quadratico medio:  $\sqrt{\frac{1}{M} \sum_{t=1}^M (e_t)^2}$ ;

con  $e_t = d_t - \hat{d}_t$ , errore di previsione.

E' evidente che si prenderanno in considerazione tutti i valori predetti di tutte le fasce semiorarie, infatti, in questo caso si ha  $M = 365 * 48$ , questo per confrontare uniformemente i vari modelli con quello funzionale nel quale si prevede l'intera curva e non i singoli valori.

### 4.1.1 Selezione dell'ordine AR

In figura 4.1, nella pagina seguente, si può osservare il grafico dell'autocorrelazione e dell'autocorrelazione parziale di una delle serie su cui si vogliono costruire i modelli, in particolare la fascia relativa all'ora 09.30-10.00. Si può notare che il grafico dell'ACF tende lentamente a zero, quindi sembra avere un comportamento simile a un processo autoregressivo. Il grafico della PACF induce a stimare un modello autoregressivo del primo ordine, perchè oltre alla prima autocorrelazione significativa, le altre sono quasi tutte all'interno delle bande di confidenza. I grafici dell'autocorrelazione e dell'autocorrelazione delle serie per ogni fascia semi oraria, hanno tutte un comportamento del tutto simile a quello riportato per la fascia 20. Alcune però presentano delle autocorrelazioni parziali significative anche oltre il ritardo uno, per

questo motivo si sceglie di stimare un modello autoregressivo del primo ordine, e si sceglie di aumentare i ritardi per vedere se il modello e la previsione possano migliorare.

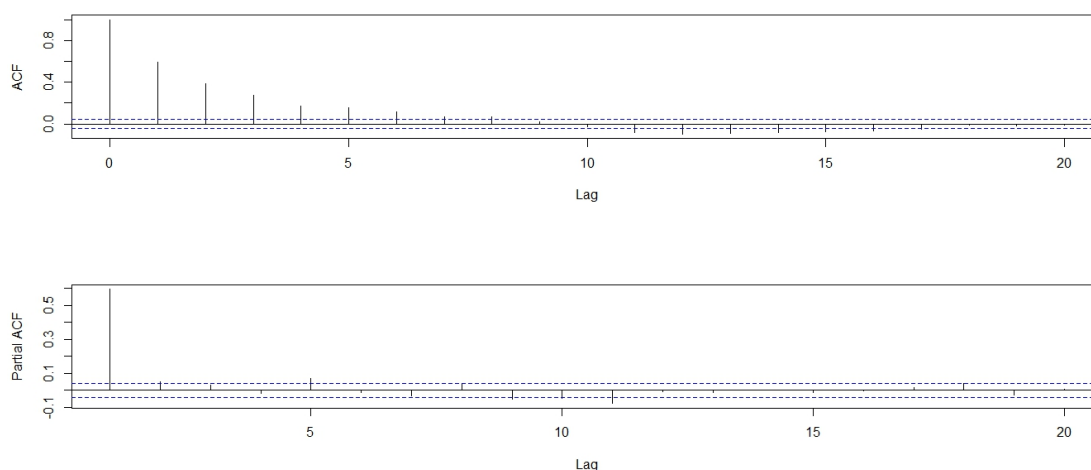


Figura 4.1: ACF e PACF della serie dei residui  $d_{t,i}$  per la fascia semi oraria relativa alle ore 9.30-10.

Nella figura 4.2 i valori previsti dal modello (colore rosso) sembrano seguire l'andamento della serie  $d_{t,i}$  (colore nero). In basso nello stesso grafico è rappresentato l'errore di stima del modello, cioè la differenza tra i valori veri e i valori stimati dal modello AR(1), si osserva come rimangano dei picchi. Questo può indurre a pensare che la decomposizione applicata in precedenza non tenga conto di altre componenti deterministiche della serie della domanda.

Dal grafico dell'autocorrelazione (figura 4.3) si evince che i residui del modello AR(1) stimato sul campione *in-sample*, siano stazionari perchè sostanzialmente le autocorrelazioni sono tutte all'interno della bande di confidenza.

La previsione della serie  $d_{t,i}$  sarà data da:

$$\hat{d}_{t+1,i} = \hat{\beta}_{0,i} + \hat{\beta}_{1,i}d_{t,i} \quad (4.1)$$

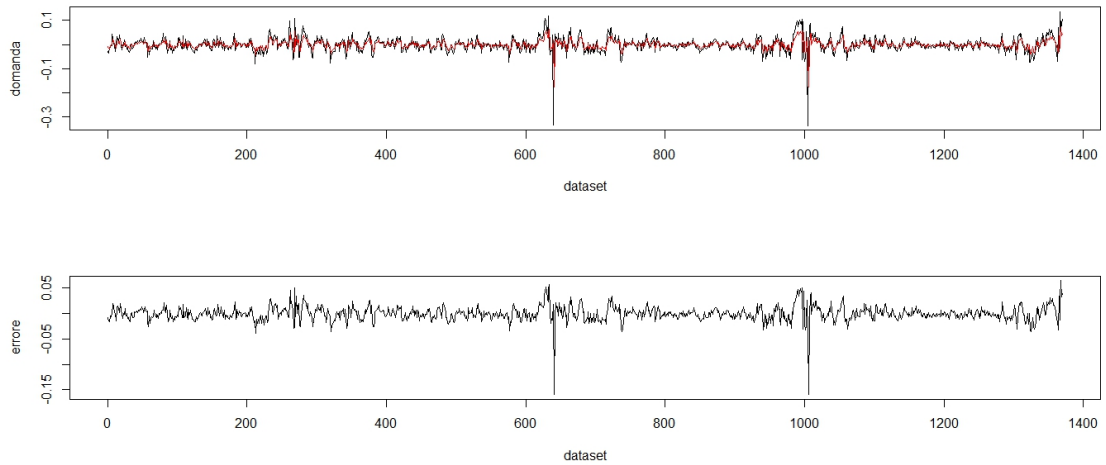


Figura 4.2: Fascia semi oraria 09.30-10.00. Valori stimati (in rosso) del modello AR(1) sul campione in-sample. In basso: serie della differenza tra valori reali e valori stimati.

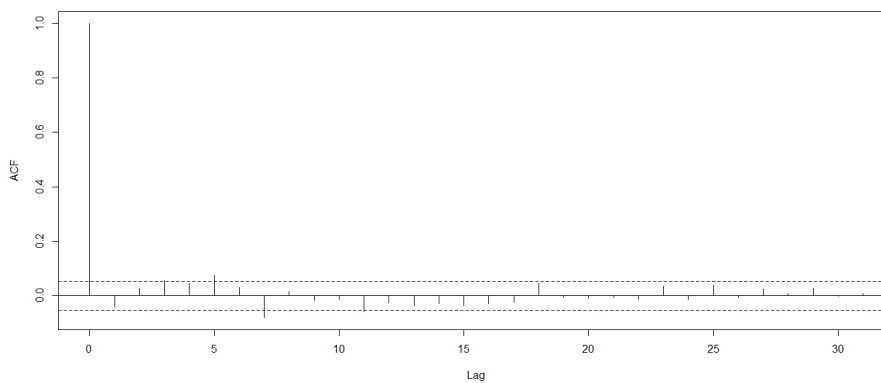


Figura 4.3: Fascia semi oraria 09.30-10.00. ACF dei residui del modello AR(1) stimato sul campione in-sample.

$\hat{\beta}_{0,i}$  e  $\hat{\beta}_{1,i}$  sono i coefficienti stimati dal AR(1).

Nella figura 4.4 è rappresentato in alto, il grafico delle 365 previsioni un passo in avanti (color rosso) e la serie reale  $d_t$  (color nero), e in basso l'errore di previsione.

Si può notare come le previsioni seguano l'andamento della serie reale ma, in alcuni punti, la vera domanda viene sottostimata. L'errore di previsione è concentrato intorno allo zero, ma nella parte finale ed iniziale della serie si discosta poco oltre lo zero.

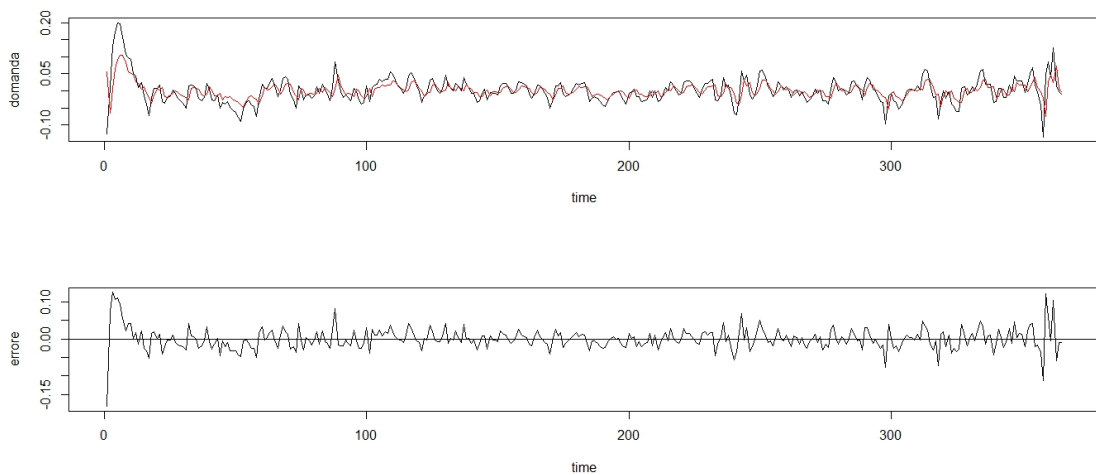


Figura 4.4: Fascia semi oraria 09.30-10.00. In alto: serie delle 365 previsioni un passo in avanti (colore rosso) nel campione in-sample, nell'asse delle  $y$  si trova la quantità della domanda di energia elettrica. In basso: serie dell'errore di previsione, nell'asse delle  $y$  si trova la variazione dell'errore.

Come si è detto già in precedenza, visto che alcune autocorrelazioni, in altre fasce semi orarie, risultano significative oltre al ritardo uno, si stima un modello autoregressivo di ordine superiore: un AR(2).

Anche in questo caso si riportano i risultati per la fascia semi oraria relativa alle ore 09.30-10.00. Risultati analoghi si ottengono per tutte le altre fasce semi orarie.

Nella figura 4.5 in alto, si osserva come la serie dei valori stimati dal modello AR(2) in colore rosso segua l'andamento della serie di  $d_{t,i}$ . Inoltre, si nota che l'errore di questa stima (nella parte in basso del grafico), è molto prossimo allo zero, tranne nei due picchi centrali della serie. Comunque dal grafico 4.6 dell'autocorrelazione parziale, si assume che i residui del modello siano stazionari, in quanto tutte le autocorrelazioni sono all'interno delle bande di confidenza.

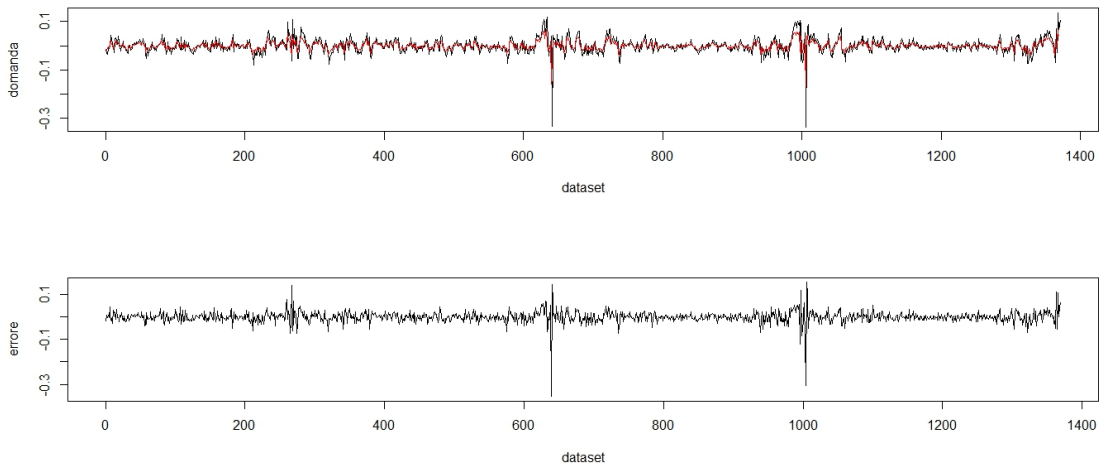


Figura 4.5: Fascia semi oraria 09.30-10.00. In alto, in rosso la serie dei valori stimati dal modello AR(2) per il campione in-sample, in nero la serie  $d_{t,i}$ , sull'asse delle y si trova la quantità della domanda. In basso il grafico della differenza tra valori della serie e valori stimati.

La previsione in questo caso sarà data da:

$$\hat{d}_{t+1,i} = \hat{\beta}_{0,i} + \hat{\beta}_{1,i}d_{t,i} + \hat{\beta}_{2,i}d_{t-1,i} \quad (4.2)$$

con  $\hat{\beta}_{0,i}$ ,  $\hat{\beta}_{1,i}$  e  $\hat{\beta}_{2,i}$  sono i coefficienti dal modello AR(2). Nella figura 4.7 in alto, è rappresentato il grafico delle 365 previsioni un passo in avanti (color rosso) e la serie reale  $d_t$  (color nero), e in basso l'errore di previsione. Anche in questo caso, le previsioni sembrano seguire l'andamento della serie reale;

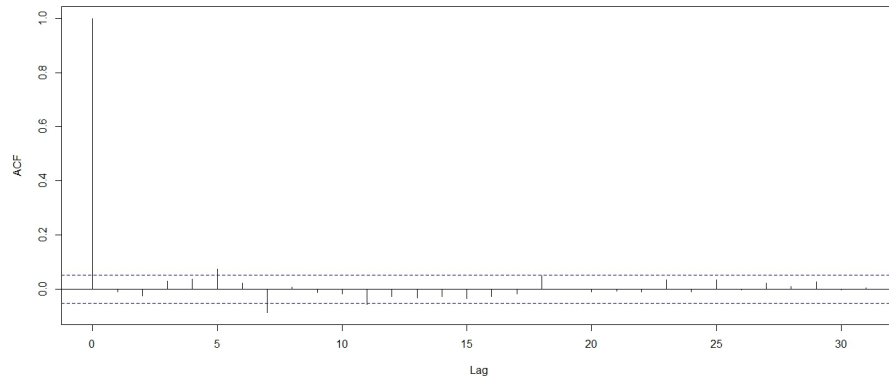


Figura 4.6: ACF dei residui del modello  $AR(2)$  stimato in-sample, fascia semi oraria 09.30-10.00.

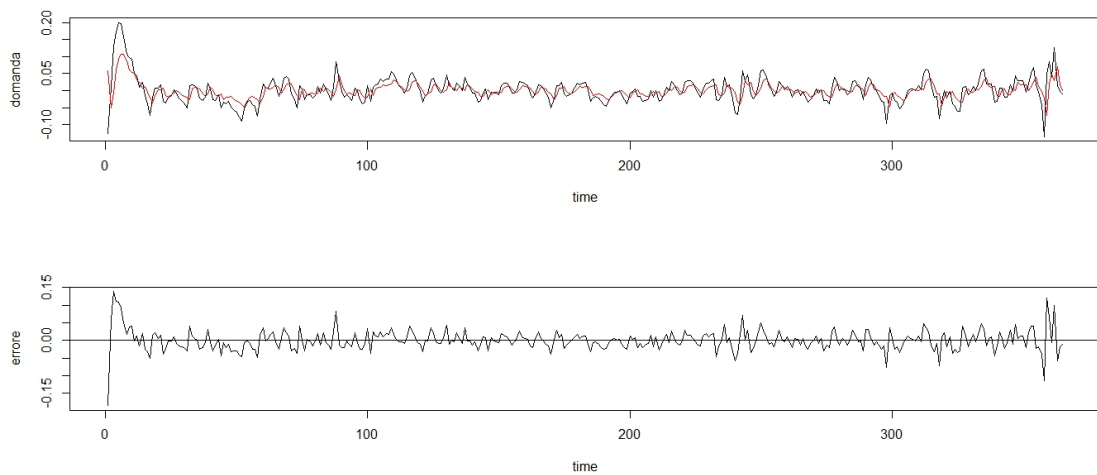


Figura 4.7: Fascia semi oraria 09.30-10.00. In alto: serie delle 365 previsioni un passo in avanti (colore rosso) nel campione in-sample dal modello  $AR(2)$ , nell'asse delle y si trova la quantità della domanda di energia. In basso serie dell'errore di previsione, nell'asse delle y si trova la variazione dell'errore.

in alcuni punti, la vera domanda viene però sottostimata, infatti l'errore di previsione ad essa associata non è zero. In generale, l'errore di previsione è prossimo allo zero.

E' possibile stimare modelli autoregressivi con ordini superiori al secondo, ma non risulta mai un modello adattabile per tutte le fasce semi orarie. Infatti, per più della metà delle fasce semi orarie, solo il primo e secondo ritardo della serie risultano significativi; aumentando i ritardi i coefficienti, risultano sostanzialmente tutti non significativi.

Si è provato a stimare un modello autoregressivo adatto per ogni fascia semi oraria, ma questo non ha portato dei miglioramenti nei risultati e quindi si è privilegiato la semplicità di adottare un unico modello uguale per tutte le fasce.

### 4.1.2 Selezione dell'ordine ARX

Si stima ora un AR(1) con l'introduzione della serie delle temperature medie come variabile esogena. Di questa variabile si considera un solo ritardo, in quanto la serie delle temperature presa in considerazione è la temperatura media di tutto il giorno. In ogni caso, aggiungendo ritardi della variabile esogena, questi risultavano non significativi per la maggior parte delle fasce semi orarie.

Nella figura 4.8 in alto, si osserva che la serie dei valori stimati dal modello ARX(1) in colore rosso sembra seguire l'andamento della serie di  $d_{t,i}$ . Inoltre, si nota che l'errore di questa stima (nella parte in basso del grafico), è molto vicina allo zero, tranne nei due picchi centrali della serie. Dal grafico 4.9 dell'autocorrelazione parziale, si assume che i residui del modello siano stazionari, in quanto tutte le autocorrelazioni sono all'interno delle bande di confidenza.

La previsione in questo caso sarà data da:

$$\hat{d}_{t+1,i} = \hat{\beta}_{0,i} + \hat{\beta}_{1,i}d_{t,i} + \hat{\gamma}_{1,i}temp_{t,i} \quad (4.3)$$



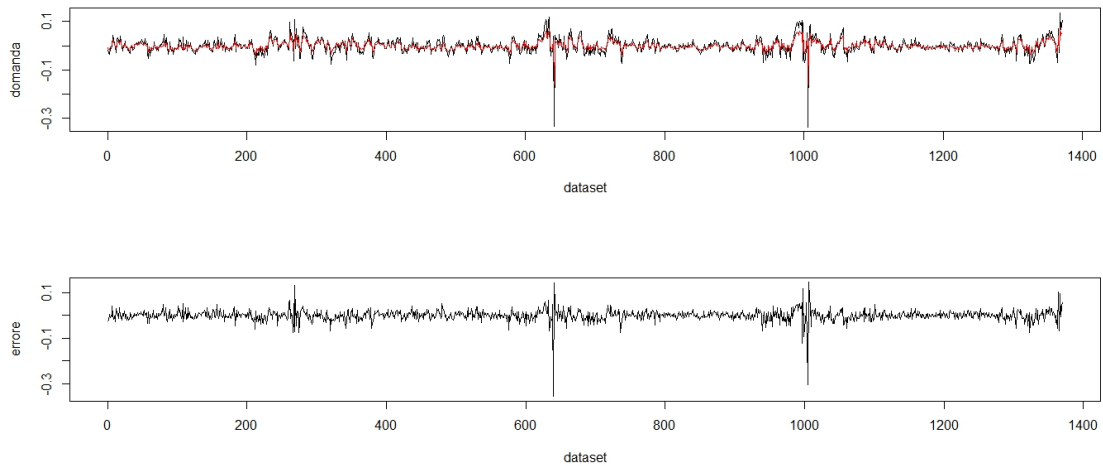


Figura 4.8: Fascia semi oraria 09.30-10.00. In alto, in rosso la serie dei valori stimati dal modello ARX(1) per il campione in-sample, in nero la serie  $d_{t,i}$ , sull'asse delle y si trova la quantità della domanda. In basso il grafico della differenza tra valori della serie e valori stimati.

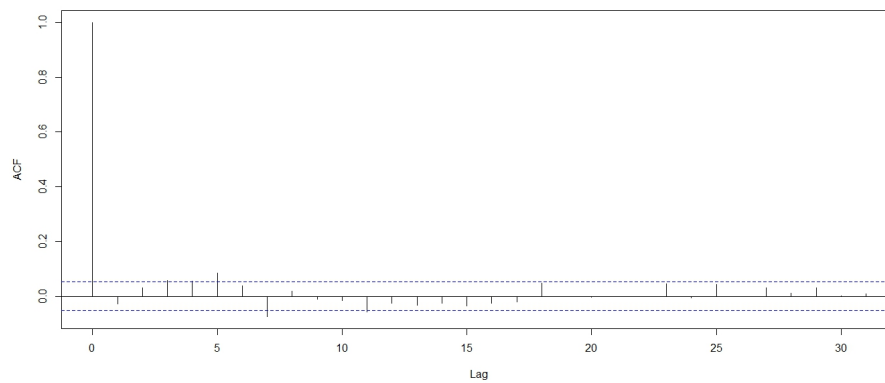


Figura 4.9: ACF dei residui del modello ARX(1) stimato in-sample, fascia semi oraria 09.30-10.00.

con  $\hat{\beta}_{0,i}$ ,  $\hat{\beta}_{1,i}$  e  $\hat{\gamma}_{1,i}$  sono i coefficienti dal modello ARX(1) per ogni fascia semi oraria.

Nella figura 4.10 in alto, è rappresentato il grafico delle 365 previsioni un passo in avanti (color rosso) e la serie reale  $d_t$  (color nero), e in basso l'errore di previsione. Si nota che le previsioni seguono fedelmente l'andamento della serie reale anche se, in alcuni punti, la vera domanda viene sottostimata; punti nei quali l'errore si discosta maggiormente dallo zero. Dato che si è sti-

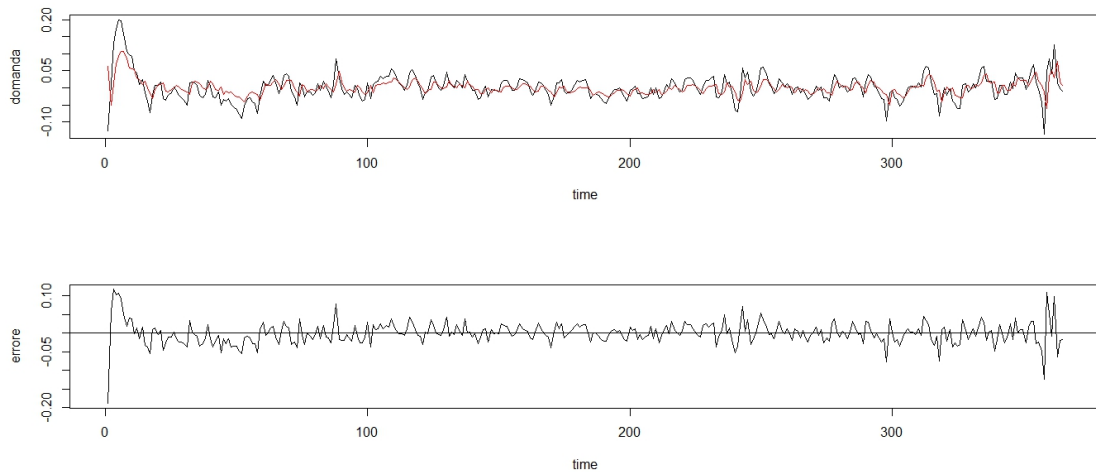


Figura 4.10: Fascia semi oraria 09.30-10.00. In alto: grafico della serie delle 365 previsioni un passo in avanti per la serie  $d_{t,i}$  dal modello ARX(1), sull'asse delle y si trova la quantità di domanda di energia. In basso la serie dell'errore di previsione.

mato un modello autoregressivo di ordine due, è interessante stimare anche il modello ARX con due ritardi nella parte autoregressiva.

Nel grafico 4.11, è raffigurata, in alto, la serie dei valori stimati dal modello (in rosso) e in basso la serie dell'errore, definito come differenza tra la serie  $d_{t,i}$  e i valori stimati. Si nota come la serie dei valori stimati segua l'andamento della serie della domanda in modo fedele, e l'errore ad esso associato risulta essere minimo, tranne nei picchi centrali della serie. Dall'autocorrelogramma di figura 4.11, si può affermare che i residui del modello ARX con 2 ritardi sembrano stazionari, essendo tutte le autocorrelazioni all'interno del-

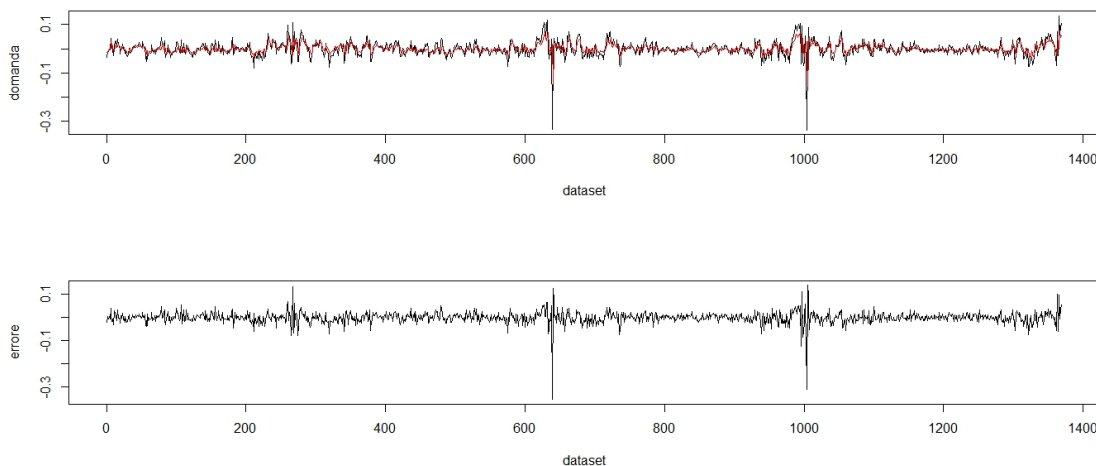


Figura 4.11: Fascia semi oraria 09.30-10.00. In alto: serie dei valori stimati dal modello ARX con  $p=2$ , colore rosso, e serie della domanda  $d_{t,i}$ . In basso: grafico della serie degli errori, cioè la differenza tra valori reali e valori stimati dal modello.

le bande di confidenza.

La previsione un passo in avanti sarà data da:

$$\hat{d}_{t+1,i} = \hat{\beta}_{0,i} + \hat{\beta}_{1,i}d_{t,i} + \hat{\beta}_{2,i}d_{t-1,i} + \hat{\gamma}_{1,i}temp_{t,i} \quad (4.4)$$

con  $\hat{\beta}_{0,i}$ ,  $\hat{\beta}_{1,i}$ ,  $\hat{\beta}_{2,i}$  e  $\hat{\gamma}_{1,i}$  sono i coefficienti dal modello ARX con due ritardi.

La serie delle 365 previsioni un passo in avanti è riportata nella figura 4.13 in alto (colore rosso), in basso invece l'errore di previsione. Anche in questo caso si evince che la serie delle previsioni segue in modo fedele la serie della domanda, e l'errore di previsione associato è sempre prossimo allo zero.

### 4.1.3 Selezione del modello VAR

In questo caso l'ordine del modello è stato scelto tramite il criterio di informazione di Akaike (AIC). Questo criterio tiene conto della bontà di adattamento e della complessità del modello, è definito come  $AIC = 2k - 2\ln(L)$  dove  $k$  è il numero di parametri del modello statistico e  $L$  è il valore massimizzato

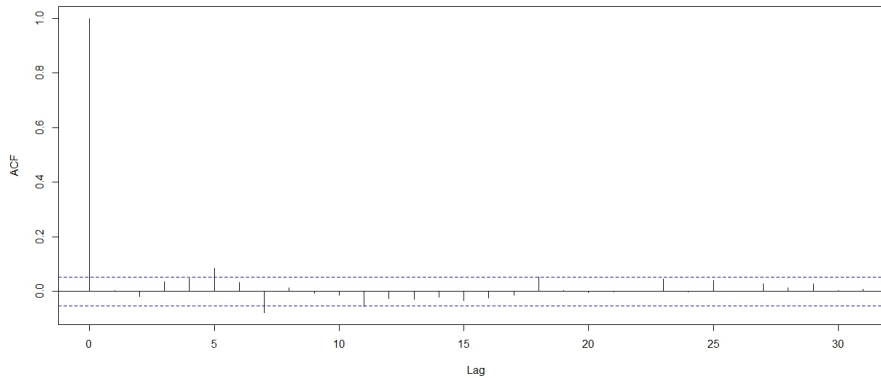


Figura 4.12: ACF dei residui del modello ARX con  $p=2$ , fascia semi oraria 09.30-10.00.

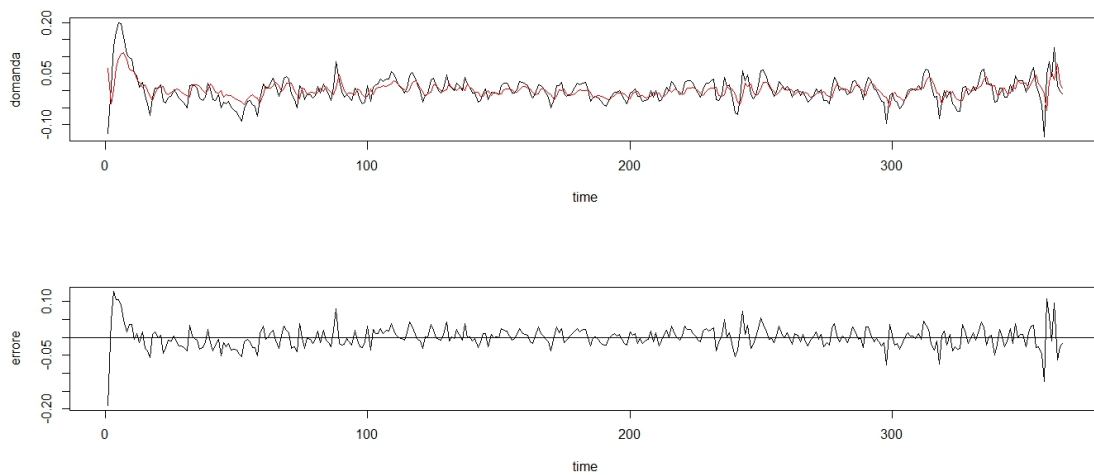


Figura 4.13: Fascia semi oraria 09.30-10.00. In alto: grafico della serie delle 365 previsioni un passo in avanti per la serie  $d_{t,i}$  dal modello ARX con due ritardi, sull'asse delle  $y$  si trova la quantità di domanda di energia. In basso la serie dell'errore di previsione.

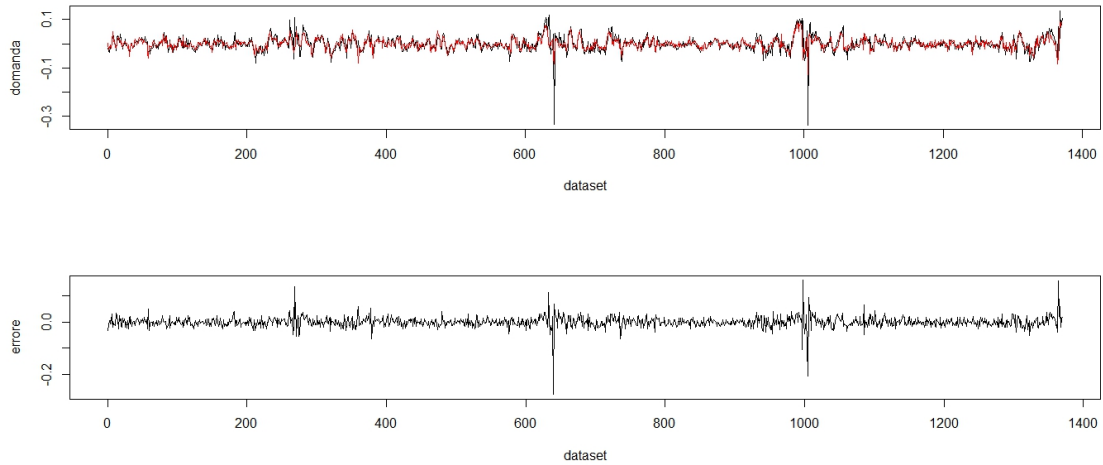


Figura 4.14: Fascia semi oraria 09.30-10.00. In alto: grafico della serie dei valori stimati dal modello VAR(1), sull'asse delle y si trova la quantità di domanda di energia. In basso la serie dell'errore come differenza tra valori veri e valori stimati dal modello.

della funzione di verosimiglianza del modello stimato. Ad ogni modo, tutti i principali criteri (BIC, HQ..) indicavano quest'ordine.

Anche in questo esempio si riportano i risultati per la fascia relativa alle ore 09.30-10.00.

Nella figura 4.14 sono rappresentati i valori stimati (color rosso) dal VAR(1), che sembrano seguire l'andamento della serie reale, tranne in alcuni picchi in cui il modello non stima bene la serie. Nella figura 4.15, invece, si valuta la bontà dei residui di questo modello, si nota che solamente un'autocorrelazione esce dalle bande di confidenza (ritardo 5), ma questo induce ad affermare che i residui del modello siano comunque stazionari e che quindi sia un modello accettabile. In questo caso la previsione un passo in avanti sarà:

$$\hat{d}_{t+1} = \hat{\Phi}_0 + \hat{\Phi}_1 d_t, \quad (4.5)$$

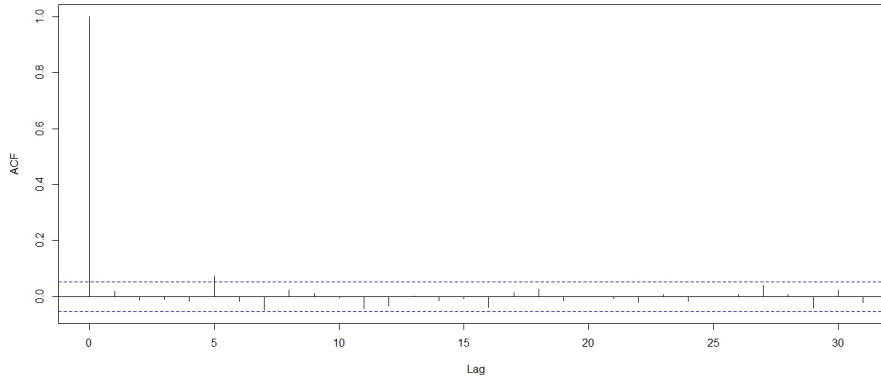


Figura 4.15: ACF residui modello VAR(1) fascia semi oraria 09.30-10.00.

$$\text{con } \hat{\Phi}_0 = \begin{bmatrix} \hat{\phi}_{1,0} \\ \hat{\phi}_{2,0} \\ \dots \\ \hat{\phi}_{48,0} \end{bmatrix} \quad e \quad \hat{\Phi}_1 = \begin{bmatrix} \hat{\phi}_{1,1} & \hat{\phi}_{1,2} & \dots & \hat{\phi}_{1,48} \\ \hat{\phi}_{2,1} & \hat{\phi}_{2,2} & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \hat{\phi}_{48,1} & \dots & \dots & \hat{\phi}_{48,48} \end{bmatrix}.$$

Esplicitando singolarmente le equazioni si avrà:

$$\hat{d}_{i,t+1} = \hat{\phi}_{i,0} + \hat{\phi}_{i,1}d_{1,t} + \hat{\phi}_{i,2}d_{2,t} + \dots + \hat{\phi}_{i,48}d_{48,t} \text{ con } i = 1, \dots, 48.$$

Nella figura 4.16 sono riportati in alto il grafico della serie delle previsioni (in rosso) in confronto alla serie reale (color nero) e in basso la serie dell'errore di previsione. L'errore di previsione è sempre prossimo allo zero, tranne, anche qui nella parte iniziale e finale della serie.

#### 4.1.4 Selezione del modello VARX

Anche in questo caso l'ordine del modello è stato scelto tramite il criterio di informazione di Akaike (AIC), ma anche tutti i principali criteri (BIC, HQ..) indicavano quest'ordine. Per quanto riguarda il ritardo della variabile temperatura si applica il ragionamento precedente e si utilizza un solo ritardo di tale variabile.

Nella figura 4.17, sono raffigurate la serie reale della domanda  $d_t$  (colore ne-

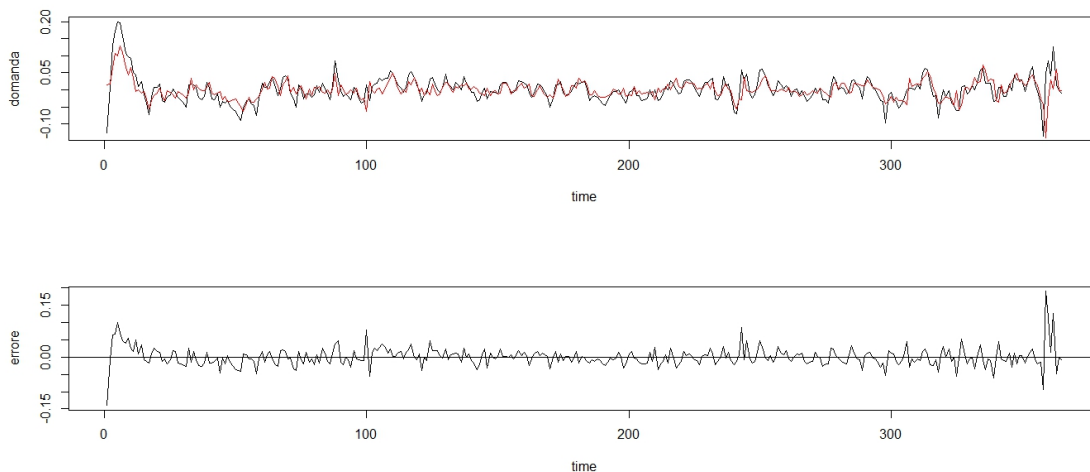


Figura 4.16: Fascia semi oraria 09.30-10.00. In alto: grafico della serie delle previsioni un passo in avanti (colore rosso) dal modello VAR(1), sull'asse delle  $y$  si trova la quantità di domanda di energia. In basso la serie dell'errore di previsione.

ro) e la serie dei valori stimati dal VARX con un ritardo (colore rosso). Come nelle specificazioni precedenti, la serie dei valori stimati segue in modo abbastanza fedele la vera serie, non cogliendo però alcuni picchi, soprattutto quelli centrali.

Sostanzialmente tutte le autocorrelazioni dei residui, nel grafico 4.18, stanno all'interno delle bande di confidenza, quindi i residui del modello risultano essere stazionari.

La previsione un passo in avanti sarà data da:

$$\hat{d}_{t+1} = \hat{\Phi}_0 + \hat{\Phi}_1 d_t + \hat{\Gamma}_1 temp_t, \quad (4.6)$$

$$\text{con } \hat{\Phi}_0 = \begin{bmatrix} \hat{\phi}_{1,0} \\ \hat{\phi}_{2,0} \\ \dots \\ \hat{\phi}_{48,0} \end{bmatrix}, \quad \hat{\Phi}_1 = \begin{bmatrix} \hat{\phi}_{1,1} & \hat{\phi}_{1,2} & \dots & \hat{\phi}_{1,48} \\ \hat{\phi}_{2,1} & \hat{\phi}_{2,2} & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \hat{\phi}_{48,1} & \dots & \dots & \hat{\phi}_{48,48} \end{bmatrix}$$

$$\text{e } \hat{\Gamma}_1 = \begin{bmatrix} \hat{\gamma}_{1,1} & \hat{\gamma}_{2,1} & \dots & \hat{\gamma}_{48,1} \end{bmatrix}.$$

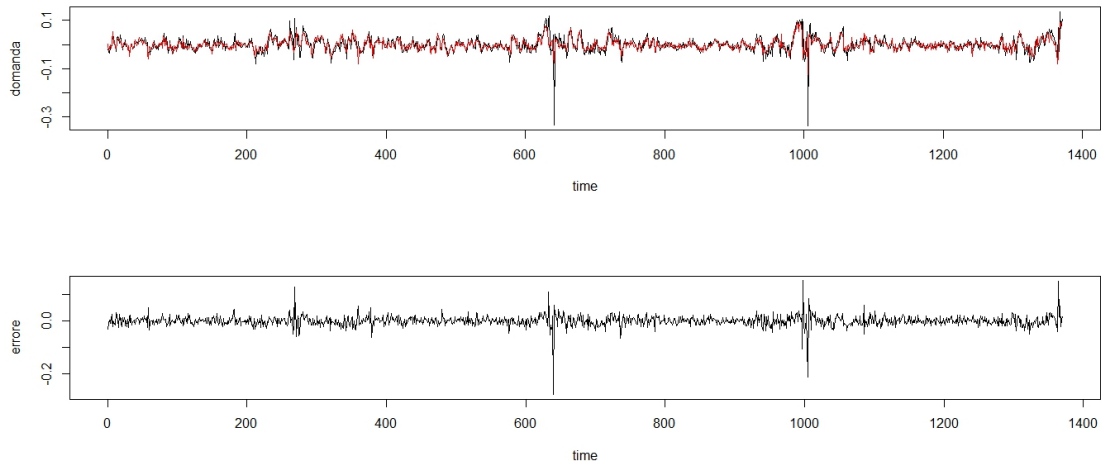


Figura 4.17: Fascia semi oraria 09.30-10.00. In alto: grafico della serie dei valori stimati dal modello VARX con 1 ritardo (colore rosso). In basso la serie della differenza tra valori veri e valori stimati.

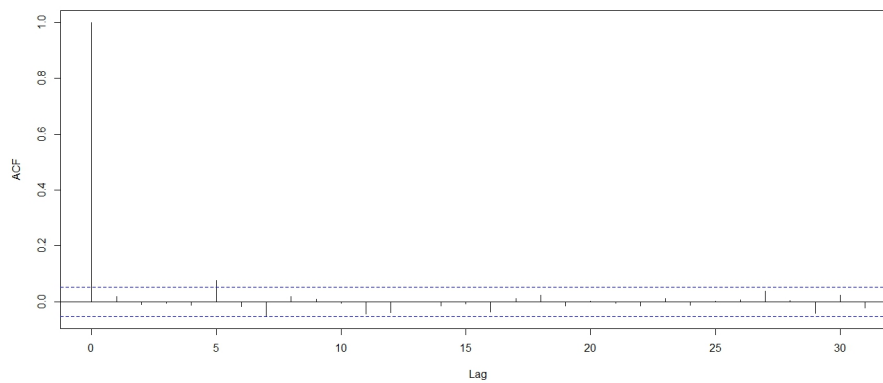


Figura 4.18: ACF dei residui del modello VARX con 1 ritardo, fascia semi oraria 09.30-10.00.



Esplicitando singolarmente le equazioni si ottiene:

$$\hat{d}_{i,t+1} = \hat{\phi}_{i,0} + \hat{\phi}_{i,1}d_{1,t} + \hat{\phi}_{i,2}d_{2,t} + \dots + \hat{\phi}_{i,48}d_{48,t} + \hat{\gamma}_{i,1}temp_{i,t}$$

con  $i = 1, \dots, 48$ .

Nella figura 4.19, in alto vi è la serie reale della domanda  $d_t$  (colore nero) e la serie delle previsioni un passo in avanti (colore rosso), in basso vi è l'errore di previsione.

Anche in questo caso è evidente come le previsioni, pur seguendo l'andamento della serie reale, spesso sottostimino o sovrastimino il vero valore. Analogamente l'errore, nella parte in basso della figura, è molto prossimo allo zero, tranne nella parte centrale e finale della serie.

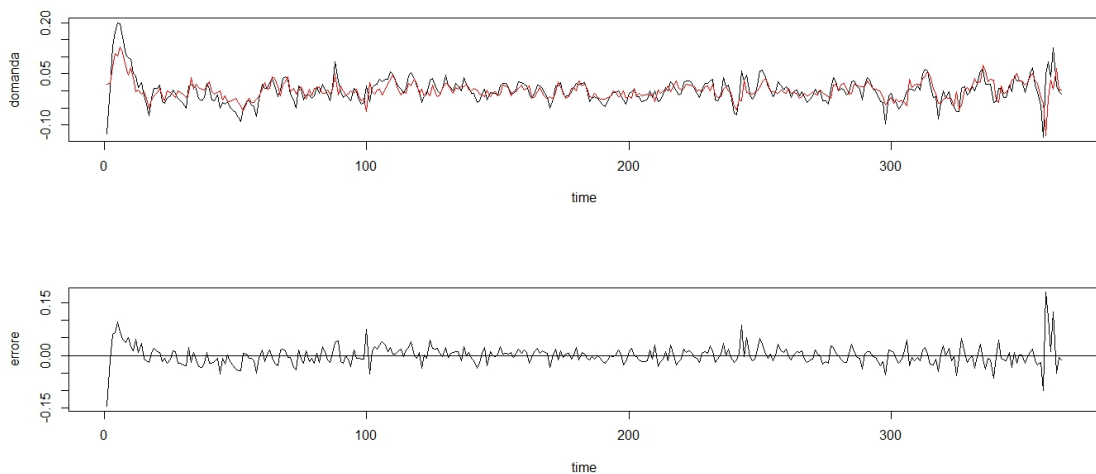


Figura 4.19: Fascia semi oraria 09.30-10.00. In alto: grafico della serie delle previsioni un passo in avanti (colore rosso) dal modello VARX con 1 ritardo, sull'asse delle  $y$  si trova la quantità di domanda di energia. In basso la serie dell'errore di previsione.

### 4.1.5 Selezione del modello funzionale

Prima di stimare il modello di regressione lineare funzionale, è necessario rendere funzionale la serie della domanda di energia, attraverso le funzioni base descritte nel capitolo 2.

Le basi scelte per i dati sono le funzioni di Fourier, visto che sono dati di tipo periodico, e le funzioni *B-Splines* che sono quelle più comuni.

E' fondamentale scegliere la funzione base e il numero di funzioni più adatte ai dati, e, non essendoci un metodo da seguire, si decide il numero  $K$  di funzioni base attraverso simulazione.

Sia per le funzioni base di Fourier, sia per le *B-Splines*, si è deciso di utilizzare i valori di  $K$  pari a 5, 9, 11, 13, 15, 17, 19, 21. Non si considerano valori troppo elevati di  $K$ , come ad esempio con  $K$  superiore di 31, perchè si otterrebbe un eccessivo adattamento ai dati osservati.

La funzione dei dati da rendere funzionale è il profilo giornaliero della domanda, si avranno, quindi, 48 osservazioni su cui stimare la funzione base. Per entrambi i tipi di funzione base scelti, si riportano in una tabella la radice gli errori quadratici medi, dove l'errore è definito come la differenza tra i valori stimati dalle funzioni base e i valori dati  $d_{t,j}$ .

In tabella 4.1 si trovano gli errori al variare del valore di  $K$  per le funzioni di Fourier e *B-Splines*. Ovviamente aumentando il valore di  $K$  l'errore diminuisce, ma si incorre nell'eccessivo adattamento ai dati.

Si sceglie, quindi, il valore di  $K$  valutando gli errori, ma anche con l'aiuto grafico. I valori di  $K = 15, 17, 19, 21$  producono un eccessivo adattamento ai dati per entrambe le funzioni base;  $K = 5$  per le *B-Splines* liscia troppo poco la serie, e  $K = 13$  risulta ancora un valore elevato. In definitiva si è scelto di procedere con i seguenti valori di  $K$ , cercando di trovare un buon compromesso tra adattabilità e lisciatura:

- per le funzioni base di Fourier è stato scelto  $K = 5, 9$ .  
Si nota come la funzione con  $K = 5$  sia meno liscia (curva blu nelle figure 4.20, 4.21) e come la funzione con  $K = 9$  a volte appaia eccessi-

Valori di $K$ , Fourier	RMSQ	Valori di $K$ , B-Splines	RMSQ
K=5	0.008076	K=5	0.008947
K=7	0.006708	K=7	0.006962
K=9	0.005662	K=9	0.005661
K=11	0.004868	K=11	0.004985
K=13	0.004304	K=13	0.004311
K=15	0.003869	K=15	0.003569
K=17	0.003482	K=17	0.003414
K=19	0.003129	K=19	0.002842
K=21	0.002856	K=21	0.002542

Tabella 4.1: Valore della radice dell'errore quadratico medio al variare di  $K$  e della funzione base.

vamente adattata (curva rossa).

Le figure si riferiscono ad alcuni giorni del campione *in-sample*.

- per le funzioni base B-Splines la scelta è  $K = 7, 9$ .

Si osserva come la funzione con  $K = 7$  sia meno liscia (curva blu nelle figure 4.22, 4.23) e come la funzione con  $K = 9$  a volte appaia eccessivamente adattata (curva rossa).

Anche in questo caso le figure si riferiscono ad alcuni giorni del campione *in-sample*.

Con questi valori di  $K$  e queste funzioni base si andrà a stimare un modello autoregressivo funzionale del primo ordine, e un modello autoregressivo funzionale con la variabile esogena temperatura media.

Per quanto riguarda il modello di regressione lineare funzionale, si farà riferimento a due specificazioni, una in cui la variabile dipendente è la serie stocastica della domanda ritardata di un periodo (modello autoregressivo funzionale), e un'altra in cui si introduce anche la serie della temperatura media del Regno Unito (ARX funzionale).

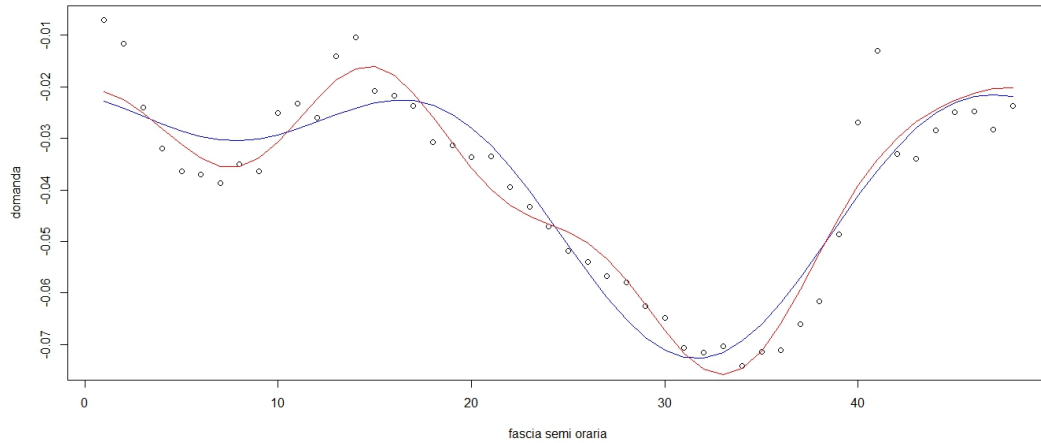


Figura 4.20: Funzioni di Fourier al variare di  $K$ .  $K = 5$  curva rossa,  $K = 7$  curva blu, per il primo giorno di osservazione.

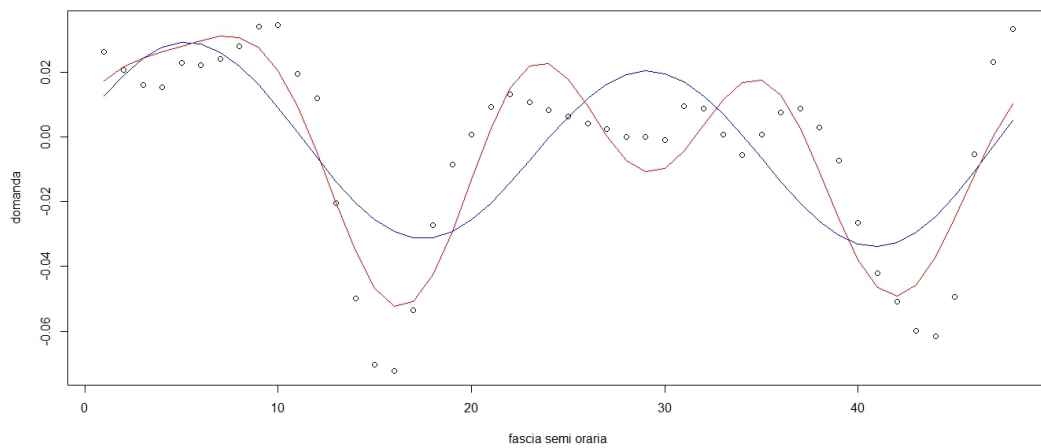


Figura 4.21: Funzioni di Fourier al variare di  $K$ .  $K = 5$  curva rossa,  $K = 7$  curva blu, per l'ultimo giorno di osservazione.

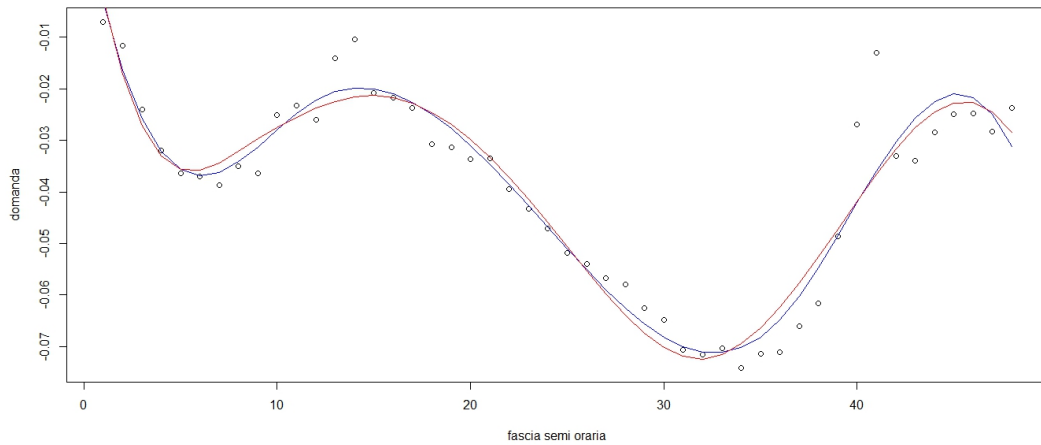


Figura 4.22: Funzioni B-Splines al variare di  $K$ .  $K = 7$  curva rossa,  $K = 7$  curva blu, per il primo giorno di osservazione.

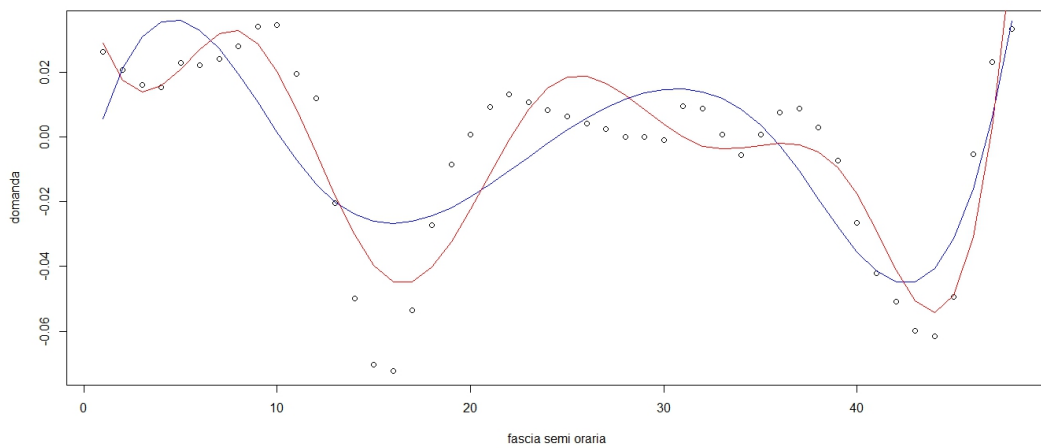


Figura 4.23: Funzioni B-Splines al variare di  $K$ .  $K = 7$  curva rossa,  $K = 7$  curva blu, per l'ultimo giorno di osservazione.

I due modelli di riferimento sono quindi:

$$d_t(i) = \beta_0(i) + d_{(t-1)}(i)\beta_1(i) + \epsilon_t(i) \quad (4.7)$$

con diverse tipologie di funzioni base, e:

$$d_t(i) = \beta_0(i) + d_{(t-1)}(i)\beta_1(i) + temp_{t-1}(i)\beta_2(i)\epsilon_t(i) \quad (4.8)$$

Le variabili risposta ed esplicative sono rese funzionali attraverso le funzioni base di Fourier e le *B-Splines*.

Di seguito vengono riportati i risultati per i diversi modelli specificati con le funzioni base di Fourier con  $K$  pari a 5 e 7, e le funzioni base *B-Spline* con  $K$  pari a 7,9.

La previsione ad un passo sarà data in modo analogo ai modelli univariati già citati.

$$d_{t+1}(i) = \hat{\beta}_0(i) + \hat{\beta}_1(i)d_{(t)}(i) \quad (4.9)$$

Per la previsione un passo in avanti dei modelli ARX funzionali si avrà:

$$d_{t+1}(i) = \hat{\beta}_0(i) + \hat{\beta}_1(i)d_{(t)}(i) + \beta_2(i)temp_{t-1}(i)\epsilon_t(i) \quad (4.10)$$

Si è stimato il modello di regressione lineare funzionale (4.7) in cui sia la variabile risposta, sia la variabile dipendente sono state trasformate in variabile funzionale tramite 5 funzioni base di Fourier.

Nella figura 4.24 si osserva, in alto, il grafico della serie dei valori stimati dal modello utilizzando 5 funzioni base di Fourier, e in basso l'errore di stima ad esso associato.

Si può notare che ci sono dei punti in cui la stima del modello non cattura in maniera adeguata i valori della serie.

Nella figura 4.25 è rappresentata, in alto, la serie delle 365 previsioni un passo in avanti svolte nel campione *in-sample*, e in basso l'errore di previsione.

Dal grafico si evince che le previsioni sembrano seguire molto fedelmente la serie della domanda  $d_{t,i}$ .

Inoltre, l'errore di previsione è molto prossimo allo zero, anche se in alcuni periodi della serie, specialmente all'inizio, si discosta di poco dallo zero.

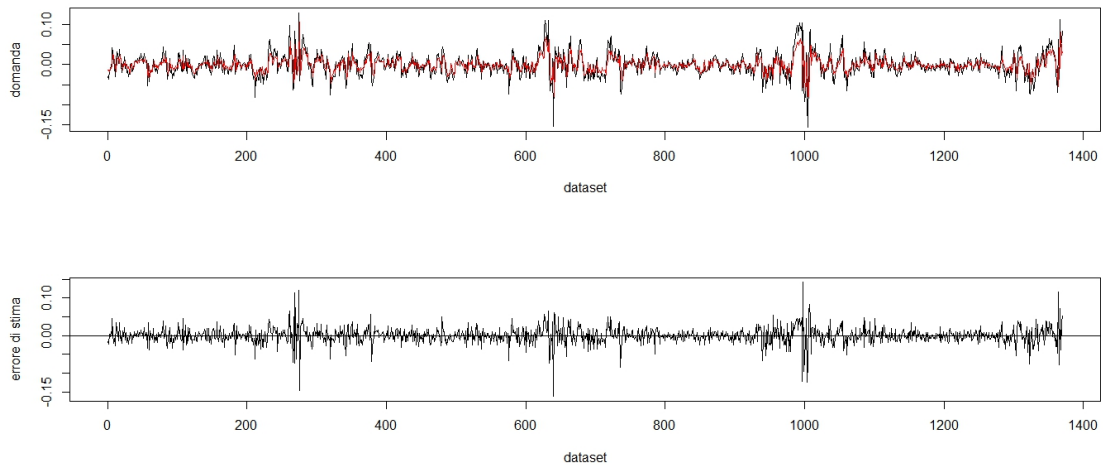


Figura 4.24: Fascia semi oraria 09.30-10.00. In alto: serie dei valori stimati dal modello funzionale con  $k = 5$  funzioni base di Fourier. In basso: l'errore di stima.

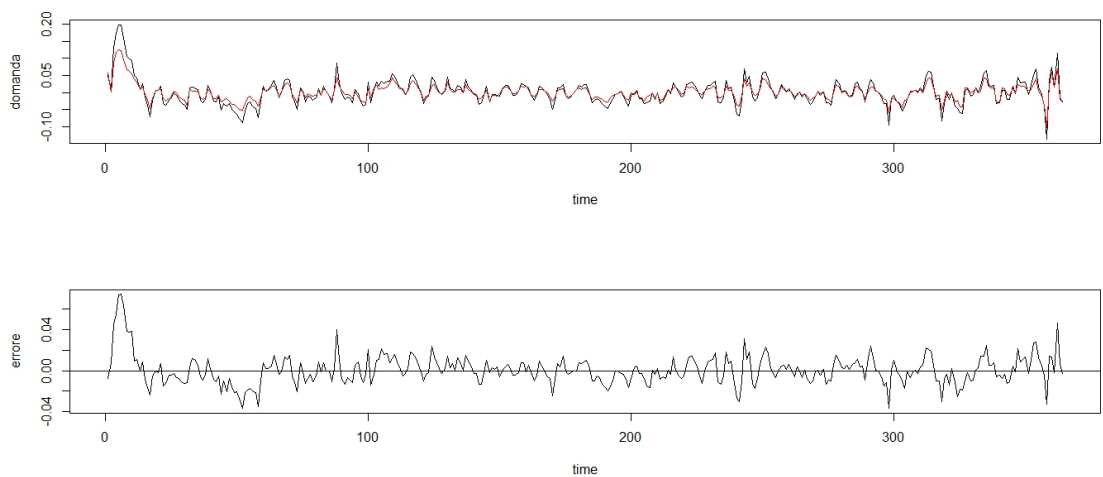


Figura 4.25: Fascia semi oraria 09.30-10.00. In alto: Serie delle 365 previsioni un passo in avanti (colore rosso) con variabili espresse con 5 funzioni base di Fourier, nel campione in-sample. In basso: errore di previsione.

Analogamente, si è stimato il modello ARX funzionale (4.8), con la temperatura media come variabile esogena. Tutte le variabile sono state rese funzionali con 5 funzioni di Fourier.

Nel grafico 4.26 è raffigurata, in alto, la serie delle 365 previsioni un passo in avanti (colore rosso) e in basso l'errore di previsione. Anche in questo caso le previsioni sembrano seguire fedelmente la serie della domanda (colore nero); l'errore di previsione appare molto contenuto e prossimo allo zero.

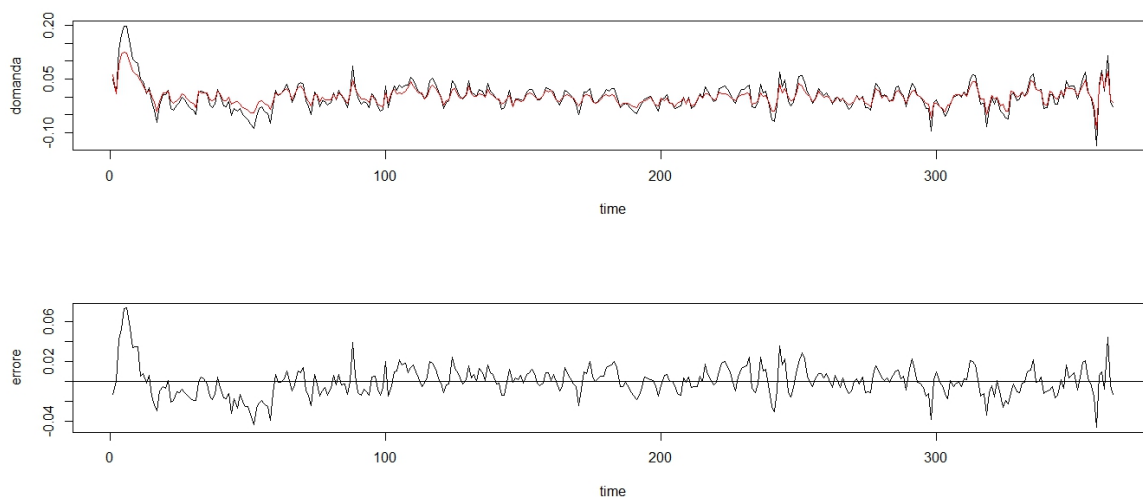


Figura 4.26: Fascia semi oraria 09.30-10.00. In alto: Serie delle 365 previsioni un passo in avanti (colore rosso). Le variabili sono espresse con 5 funzioni base di Fourier, modello stimato nel campione in-sample. In basso: serie dell'errore di previsione.



In egual modo è stato stimato il modello di regressione funzionale (4.7) sintetizzando i dati, questa volta, con 9 funzioni base di Fourier.

Graficamente i valori stimati dal modello con 9 basi di Fourier sono apparsi analoghi al precedente, per semplicità si riportano, quindi, soltanto i risultati delle previsioni un passo in avanti.

Anche in questo caso, si nota, dalla figura 4.27, che la serie delle 365 previsioni un passo in avanti (colore rosso) segue passo passo l'andamento della serie della domanda (colore nero) per il campione dei dati *in-sample*.

Anche l'errore di previsione è praticamente sempre vicino allo zero.

Solamente all'inizio della serie si nota uno scostamento dallo zero, ma comunque rimane molto piccolo.

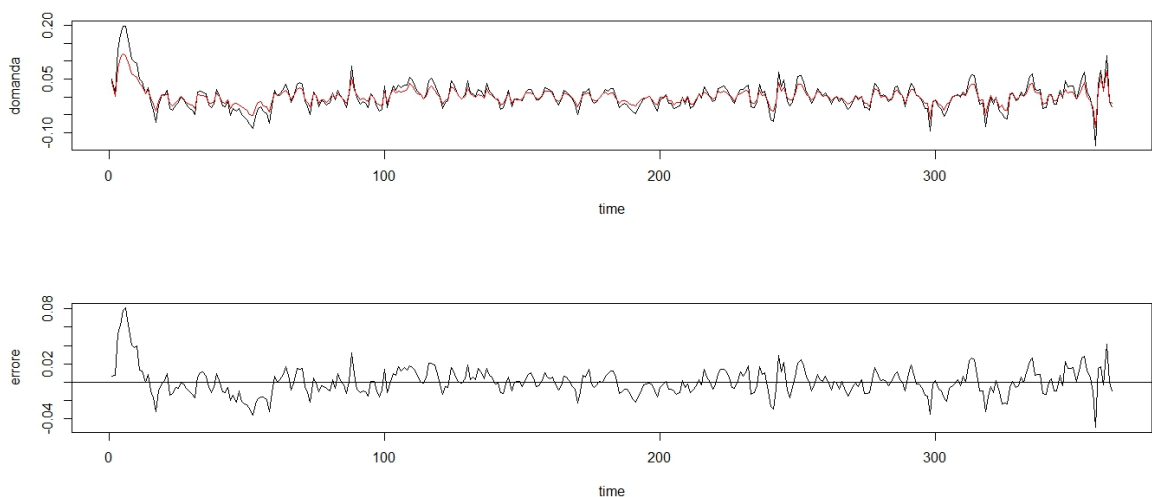


Figura 4.27: Fascia semi oraria 09.30-10.00. In alto: Serie delle 365 previsioni un passo in avanti (colore rosso). Le variabili sono espresse con 9 funzioni base di Fourier, e il modello è stato stimato nel campione *in-sample*. In basso: serie dell'errore di previsione.

Successivamente, è stato stimato il modello ARX funzionale, in cui anche la variabile temperatura è stata resa funzionale, sempre con 9 funzioni di Fourier.

Si trascrivono, quindi, i risultati per la previsione un passo in avanti.

Nella figura 4.28 si osserva, in alto, che la serie delle 365 previsioni un passo in avanti (colore rosso) ha pari andamento della serie della domanda (colore nero). In basso si nota, invece, il grafico dell'errore di previsione che appare molto contenuto e intorno allo zero.

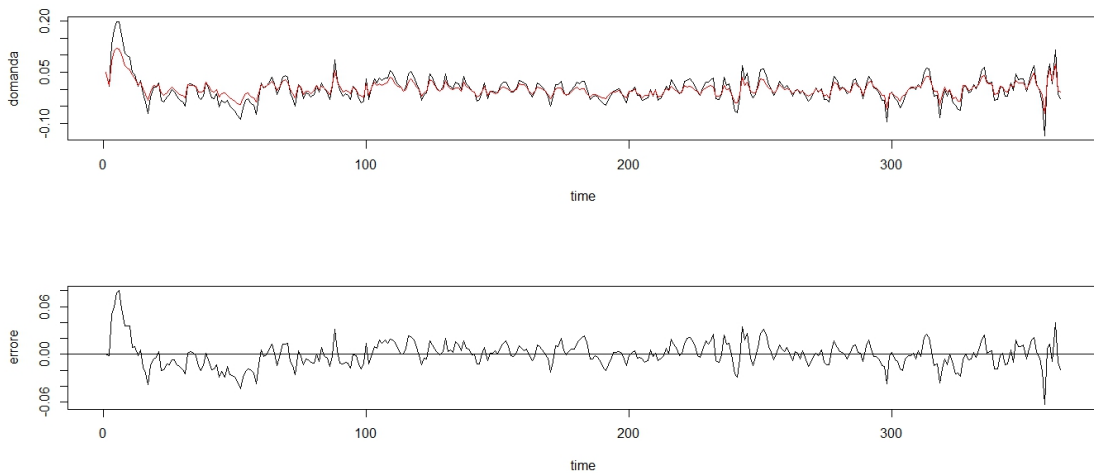


Figura 4.28: Fascia semi oraria 09.30-10.00. In alto: Serie delle 365 previsioni un passo in avanti (colore rosso) con variabili espresse con 9 funzioni base di Fourier, nel campione in-sample. In basso: errore di previsione.

È interessante verificare se, utilizzando una funzione base che appare più semplice e più usata, le previsioni del modello di regressione lineare funzionale cambino. In questo caso si utilizzano le funzioni *B-Splines* con valori di  $K$  pari a 7 e 9.

Inizialmente si stima il modello di regressione funzionale con 7 *B-Splines*.

Nella figura 4.29 a pagina seguente, si trova, come sempre, in alto la serie delle previsioni un passo in avanti e, in basso, l'errore di previsione. Si può notare che l'errore di previsione è prossimo allo zero, discostandosi solamen-

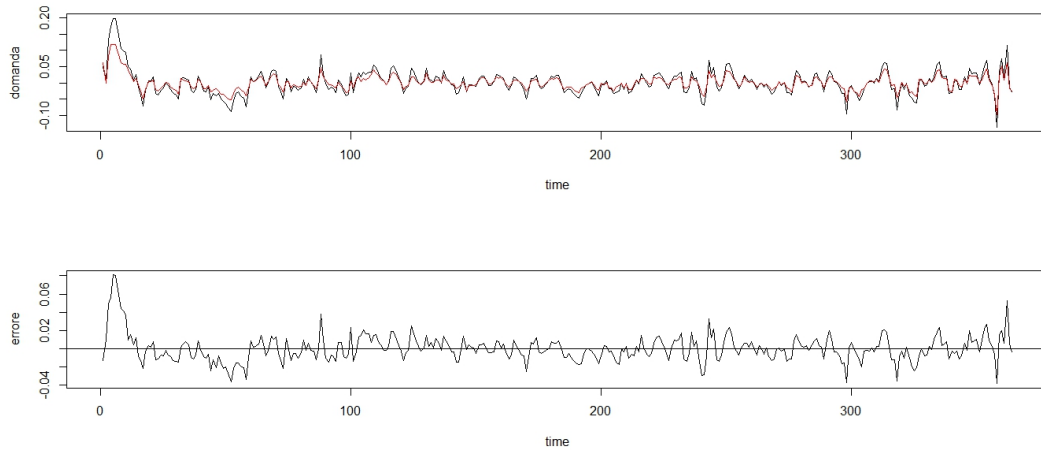


Figura 4.29: Fascia semi oraria 09.30-10.00. In alto: Serie delle 365 previsioni un passo in avanti (colore rosso) con variabili espresse con 7 funzioni base *B-Splines*, nel campione in-sample. In basso: errore di previsione.

te di poco oltre lo zero in alcuni punti della serie.

Si è proceduto in modo analogo per stimare il modello con l'introduzione della variabile esogena funzionale, anch'essa con 7 funzioni base *B-Splines*.

Nella figura 4.30 si può osservare, in alto, che la serie delle 365 previsioni un passo in avanti (colore rosso) ricalca l'andamento della serie della domanda (colore nero). In basso si trova il grafico dell'errore di previsione che appare molto contenuto e intorno allo zero, con qualche picco intorno allo 0.08.

Successivamente, si è provato a stimare un modello di regressione lineare funzionale con 9 funzioni base *B-Splines*, le quali, ricordando il grafico mostrato in precedenza, sembravano lisciare troppo i dati.

In questo caso l'errore di previsione arriva anche a 0.08, ma nella parte restante della serie è comunque prossimo allo zero.

Nel grafico 4.31 è raffigurata, in alto la serie delle previsioni un passo in avanti e, in basso, l'errore di previsione. Anche in questo caso le previsioni sono simili all'andamento della serie della domanda, e l'errore di previsione è intorno allo zero. Si stima, infine, con  $k = 9$  funzioni *B-Spline* il modello di

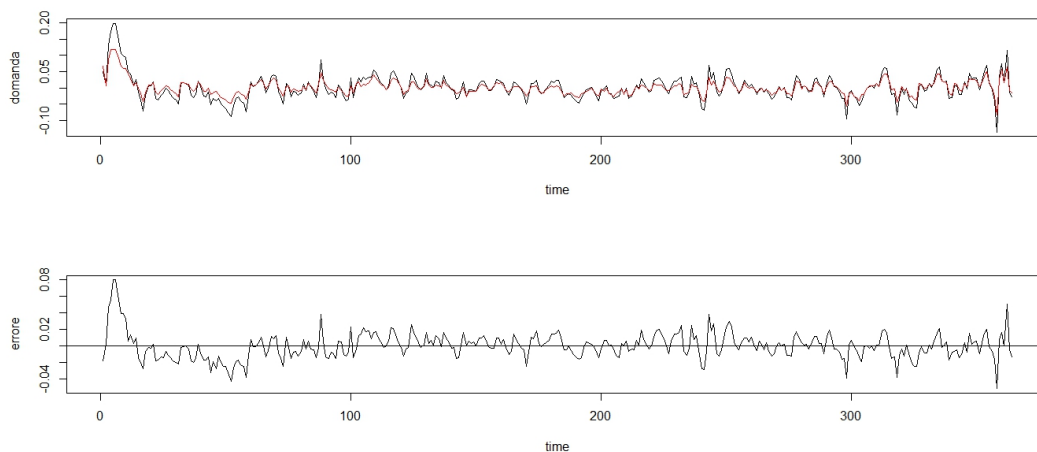


Figura 4.30: Fascia semi oraria 09.30-10.00. In alto: Serie delle 365 previsioni un passo in avanti (colore rosso) con variabili espresse con 7 funzioni base di B-Spline, nel campione in-sample. In basso: errore di previsione.

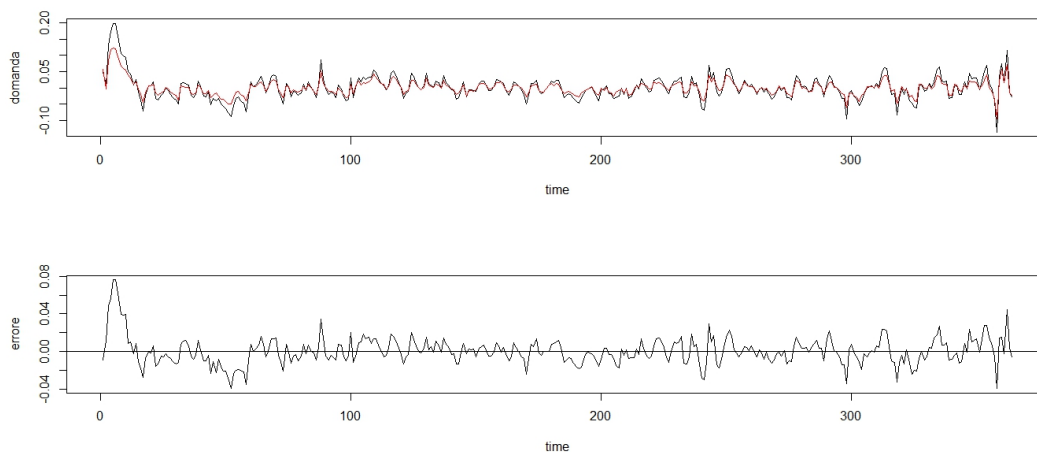


Figura 4.31: Fascia semi oraria 09.30-10.00. In alto: Serie delle 365 previsioni un passo in avanti (colore rosso) con variabili espresse con 9 funzioni base B-Splines, nel campione in-sample. In basso: errore di previsione.

regressione lineare con la temperatura media come variabile esogena. Nella figura 4.32 è rappresentata, in alto, la serie delle 365 previsioni un passo in avanti (colore rosso) e in basso il grafico dell'errore di previsione. Quest'ultimo appare molto contenuto e prossimo allo zero, infatti, la serie delle previsioni è molto vicina alla serie della domanda reale.

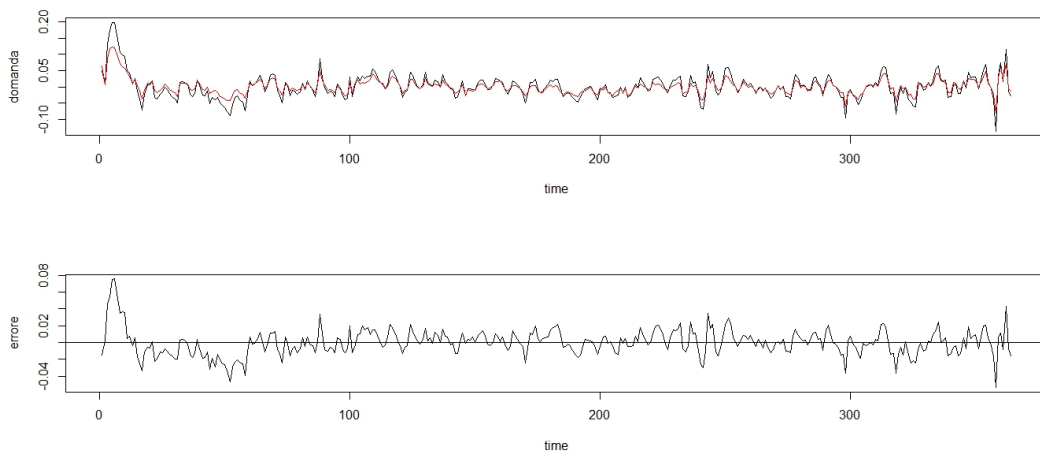


Figura 4.32: Fascia semi oraria 09.30-10.00. In alto: Serie delle 365 previsioni un passo in avanti (colore rosso) con variabili espresse con 7 funzioni base di B-Spline, nel campione in-sample. In basso: errore di previsione.

Si sono stimati modelli autoregressivi univariati, multivariati e funzionale, e si sono stimati gli stessi con l'introduzione della variabile esogena.

Si riassumono quindi in un'unica tabella (4.2) tutti i risultati ottenuti riguardo all'errore di previsione.

Per quanto riguarda l'analisi univariata si scelgono, osservando la tabella, i modelli AR(1) e ARX(1) in quanto, aumentare i ritardi non porta un miglioramento così netto alle previsioni.

Invece per l'analisi multivariata si faranno le previsioni su entrambi i modelli proposti.

Dalla tabella si evince, inoltre, che i modelli funzionali, anche con diverso

Modello	MAE	MAPE	RMSQ
AR(1)	0.02026	3.2213	0.02824
AR(2)	0.02025	3.1476	0.02822
ARX(1)	0.02054	3.1943	0.02833
ARX(2)	0.02057	3.1409	0.02840
VAR(1)	0.0177	3.4415	0.02569
VARX(1)	0.0179	3.3909	0.02574
ARF con $K = 5$ Fourier	0.01204	1.49433	0.01675
ARF con $K = 9$ Fourier	0.01137	1.09154	0.01589
ARF con $K = 7$ B-Spline	0.01169	1.14105	0.01636
ARF con $K = 9$ B-Spline	0.01187	1.11867	0.01591
ARXF con $K = 5$ Fourier	0.01275	1.60918	0.01748
ARXF con $K = 9$ Fourier	0.01268	1.28520	0.01670
ARXF con $K = 7$ B-Spline	0.01245	1.54858	0.01713
ARXF con $K = 9$ B-Spline	0.01216	1.27909	0.01672

Tabella 4.2: Indicatori sull'errore di previsione dei vari modelli in-sample. ARF: modello autoregressivo funzionale; ARXF: modello autoregressivo funzionale con l'introduzione di una variabile esogena.

numero e tipologia di funzione base, hanno gli indici con valori molto vicini tra loro.

Quindi, per quanto riguarda il modello autoregressivo funzionale, si nota che la funzione base di Fourier con  $K = 9$  risulta avere tutti gli indici minori rispetto ai restanti modelli.

Invece, per il modello ARX funzionale si nota che la funzione *B-Spline* con  $K = 9$  minimizza tutti gli errori.

## 4.2 Previsioni out-of-sample

L'obiettivo della tesi è quello di prevedere la domanda di energia  $D_{t,i}$  nel breve periodo.

Per determinare quale modello utilizzare per le previsioni, si sono stimati diversi modelli e condotto le previsioni sul campione *in-sample*.

Ora si sono determinati quali modelli utilizzare per la previsione della vera domanda di energia; queste previsioni verranno condotte sul terzo periodo di osservazioni, denominato *out-of-sample*, ovvero il periodo che va dal 1 Gennaio 2010 al 31 Dicembre 2010.

In ogni caso si hanno a disposizione i veri valori della domanda per poter effettuare un confronto tra le previsioni e il vero valore.

Si ricorda che le previsioni un passo e due passi in avanti saranno date da:

$$\begin{aligned}\widehat{D}_{t+1,i} &= \exp(\widehat{L}_{t+1,i} + \widehat{S}_{t+1,i} + \widehat{F}_{t+1,i} + \widehat{d}_{t+1,i}) & t = 1737, \dots, 2101 \\ \widehat{D}_{t+2,i} &= \exp(\widehat{L}_{t+1,i} + \widehat{S}_{t+2,i} + \widehat{F}_{t+2,i} + \widehat{d}_{t+2,i}) & t = 1737, \dots, 2100\end{aligned}$$

dove  $\widehat{d}_{t+1,i}$  e  $\widehat{d}_{t+2,i}$  sono le previsioni un passo e due passi in avanti della serie stocastica.

Di seguito verranno riportati i risultati delle previsioni ottenute dai modelli univariati, multivariati e funzionali identificati in precedenza.

### 4.2.1 Modello AR(1)

Per i modelli univariati autoregressivi è stato scelto di stimare il modello AR(1).

Nella figura 4.33 si trova in alto, il grafico della previsione un passo in avanti (colore rosso) per la serie della domanda  $D_{t,i}$ , con  $i = 20$ . Si nota che la previsione sembra seguire pari passo la serie della domanda, e che l'errore di previsione, soprattutto nella parte centrale della serie, sembra essere molto vicino a zero.

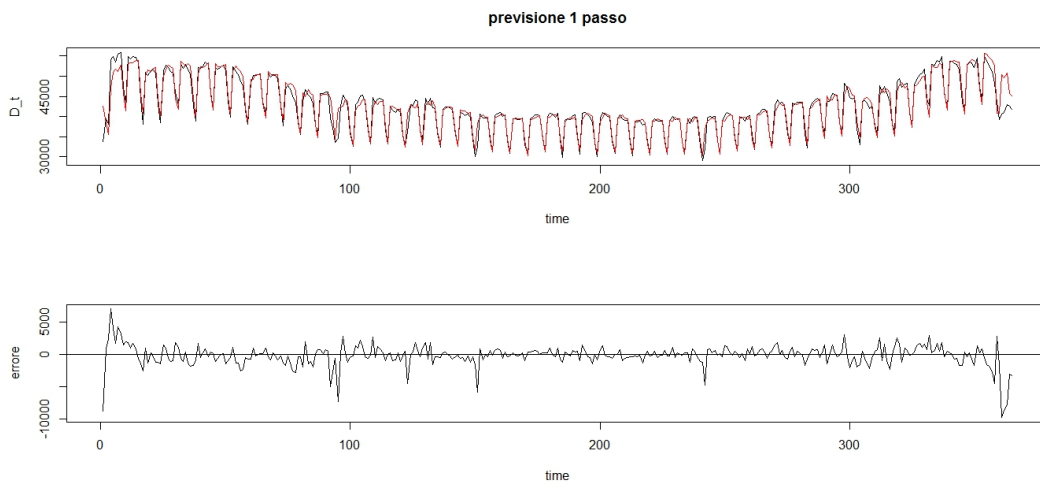


Figura 4.33: Fascia semi oraria 09.30-10.00. In alto: serie delle 365 previsioni un passo in avanti (colore rosso), in basso l'errore di previsione per il periodo 1 Gennaio-31 Dicembre 2010. Modello AR(1)

Invece, nella figura 4.34, è rappresentata la previsione due passi in avanti per la stessa fascia semi oraria ( $i = 20$ ).

Anche in questo caso si nota che la serie delle previsioni (colore rosso) segue l'andamento della serie reale fedelmente.

L'errore di previsione non sempre è vicino allo zero, nella parte iniziale e finale della serie si discosta dallo zero arrivando intorno ai  $|4000|$ .

E' ancora comunque accettabile visto che il livello medio della serie è intorno ai 40000.



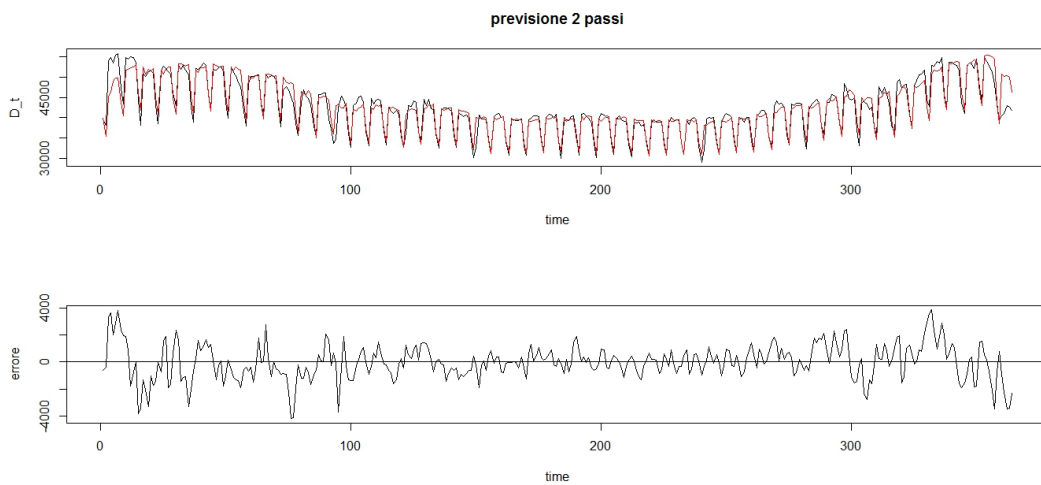


Figura 4.34: Fascia semi oraria 09.30-10.00. In alto:serie della previsione due passi in avanti (colore rosso). In basso serie dell'errore di previsione per il periodo 1 Gennaio-31 Dicembre 2010. Modello AR(1).

Nelle figure 4.35 e 4.36 a pagina seguente, sono riportate rispettivamente le previsioni e l'errore un passo e due passi in avanti per la fascia semi oraria 03.30-4.00, ovvero la fascia in cui la domanda di energia risulta minore.

Anche in questo caso si osserva che le previsioni un passo e due passi in avanti seguono fedelmente l'andamento della vera serie della domanda.

Si può notare che l'errore di previsione un passo in avanti è più variabile nelle parti iniziali e finale della serie, ma sempre molto prossimo allo zero.

Infine, nella figura 4.37, si trova la previsione giornaliera, sia un passo in avanti (colore rosso), sia due passi in avanti (colore blu), confrontate con la vera domanda giornaliera (colore nero).

Entrambe le previsioni sembrano sovrastimare la vera domanda nella prima parte del giorno, e nella seconda parte sembrano sottostimare il vero valore della domanda.

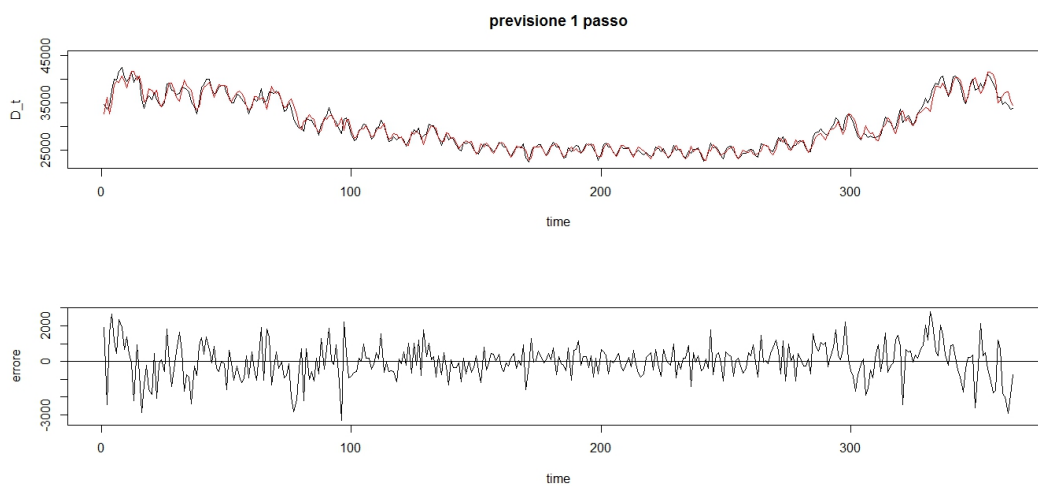


Figura 4.35: Fascia semi oraria 03.30-4.00. In alto:serie della previsione un passo in avanti (colore rosso). In basso serie dell'errore di previsione per il periodo 1 Gennaio-31 Dicembre 2010. Modello AR(1).

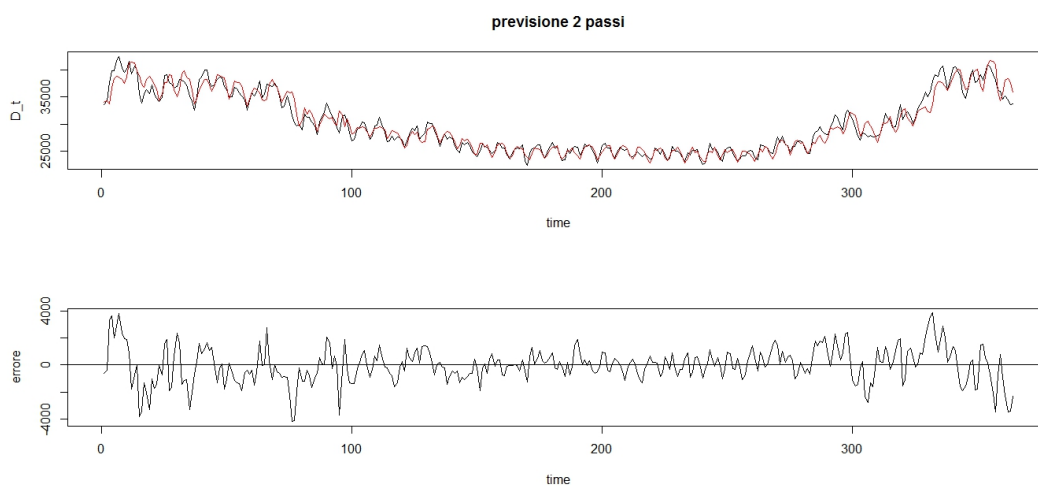


Figura 4.36: Fascia semi oraria 03.30-4.00. In alto:serie della previsione due passi in avanti (colore rosso). In basso serie dell'errore di previsione per il periodo 1 Gennaio-31 Dicembre 2010. Modello AR(1).

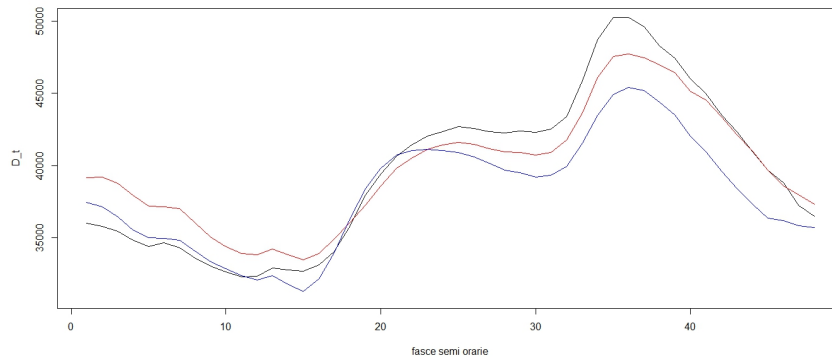


Figura 4.37: *Previsione del giorno 1738 (2 Gennaio 2010): in colore rosso la previsione un passo in avanti, in colore blu la previsione due passi in avanti e in nero vero profilo giornaliero della domanda di energia.*

#### 4.2.2 Modello ARX(1)

Tra i modelli autoregressivi univariati con variabile esogena proposti, è risultato migliore il modello ARX con un solo ritardo nella parte autoregressiva. Aumentando i ritardi nella parte autoregressiva del modello non si sono ottenuti miglioramenti netti alle previsioni.

Per quanto riguarda la previsione ad un passo e due passi in avanti si segue sempre lo stesso schema proposto. L'unica annotazione, in questo caso, vi è nella previsione due passi in avanti, nella quale, per la previsione di  $d_{t+2,i}$ , il valore della temperatura al tempo  $t + 1$  viene considerato uguale al valore al tempo  $t$ .

Questa è un'ipotesi semplificatrice, in quanto la temperatura media non cambia in maniera netta da un giorno all'altro.

Anche in questo caso, nella figura 4.38 si trova, in alto, la previsione un passo in avanti (colore rosso) e, in basso, l'errore di previsione.

Nella figura 4.39 vi è, invece, la previsione due passi in avanti (colore rosso) e l'errore di previsione ad essa associata. Si nota, come, l'errore di previsione due passi in avanti ha alcuni picchi in cui si discosta dallo zero, mentre l'errore di previsione un passo in avanti è quasi sempre attorno allo zero.

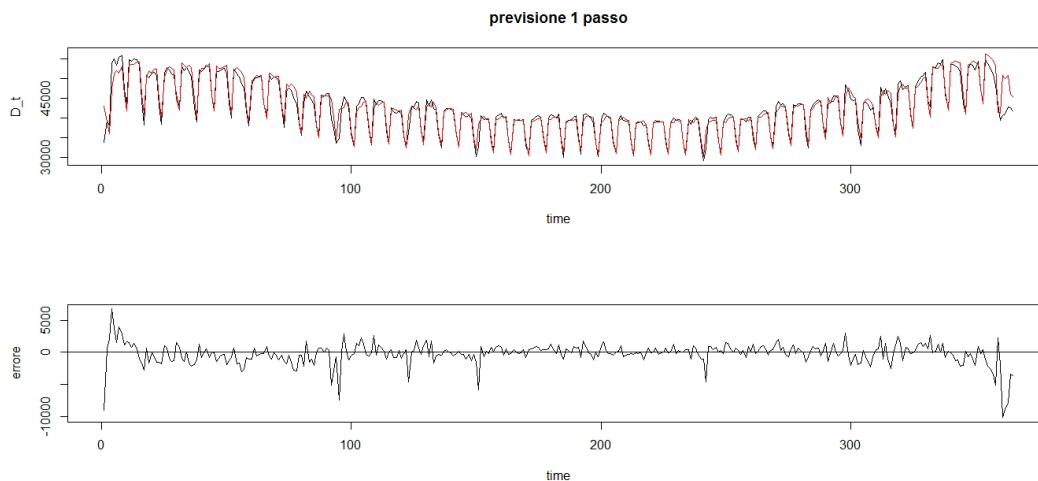


Figura 4.38: Fascia semi oraria 09.30-10.00 (maggior picco). In alto: serie delle 365 previsioni un passo in avanti (colore rosso) per la serie della domanda  $D_{t,i}$  per  $i = 20$ . In basso: l'errore di previsione per il periodo 1 Gennaio-31 Dicembre 2010. Modello ARX(1).

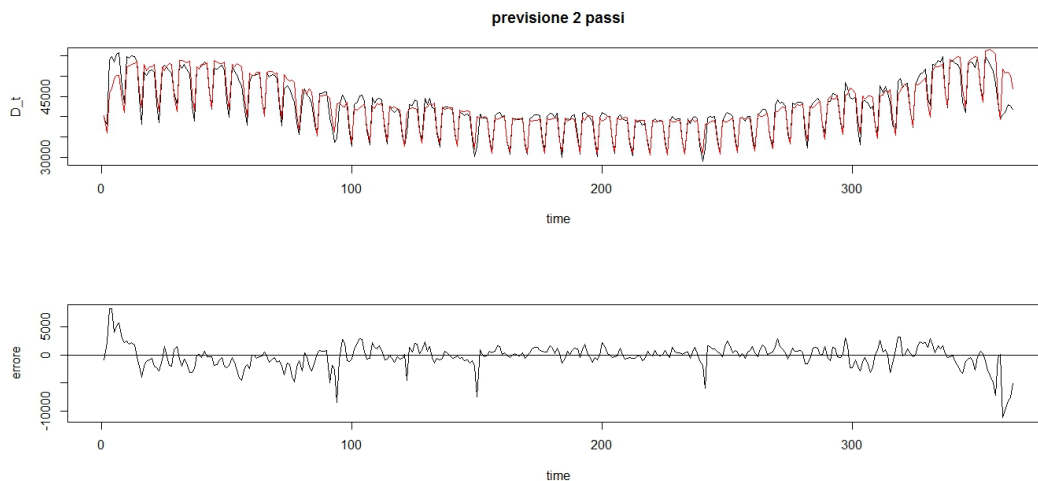


Figura 4.39: Fascia semi oraria 09.30-10.00 (maggior picco). In alto: serie delle previsioni due passi in avanti (colore rosso) per la serie della domanda. In basso l'errore di previsione per il periodo 1 Gennaio-31 Dicembre 2010. Modello ARX(1).

Nelle figure 4.40 e 4.41 si trovano rispettivamente la serie delle previsioni e dell'errore un passo in avanti e due passi in avanti, per la fascia semi oraria di minor picco, ovvero le ore 03.30-4.00. Rispetto alla previsione della fascia relativa alle ore 09.30-10.00, si nota che l'errore non va oltre i  $|3000|$  per la previsione ad un passo e non va oltre i  $|4000|$  per la previsione due passi in avanti. E' però più variabile nelle parti estreme della serie.

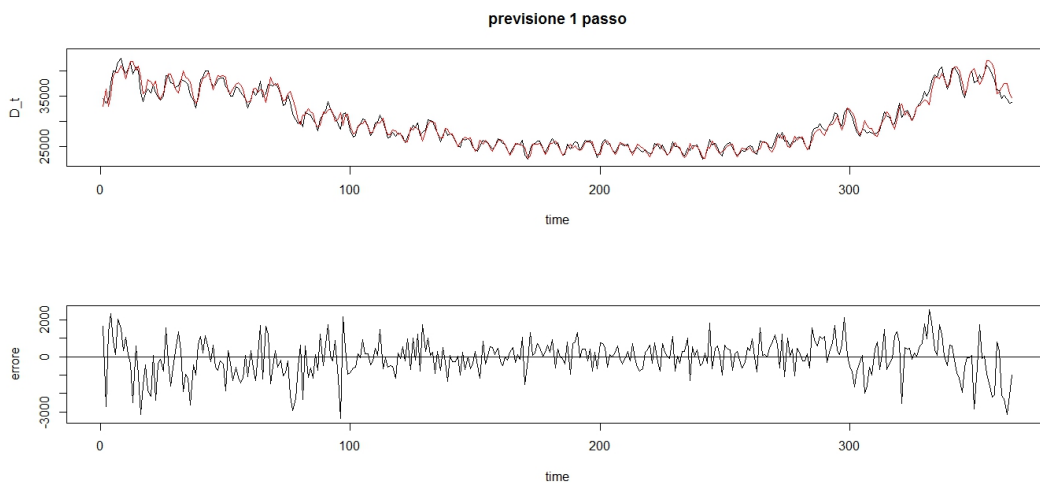


Figura 4.40: Fascia semi oraria 03.30-4.00 (minor picco). In alto: serie delle 365 previsioni un passo in avanti (colore rosso) per la serie della domanda  $D_{t,i}$  per  $i = 20$ . In basso: l'errore di previsione per il periodo 1 Gennaio-31 Dicembre 2010.. Modello ARX(1)

Nella figura 4.42 vi è la previsione di un singolo giorno (colore nero), sia un passo in avanti (colore rosso), sia due passi in avanti (colore blu). Anche in questo caso le previsioni, nella prima parte sovrastimano la domanda, e nella seconda la sottostimano.

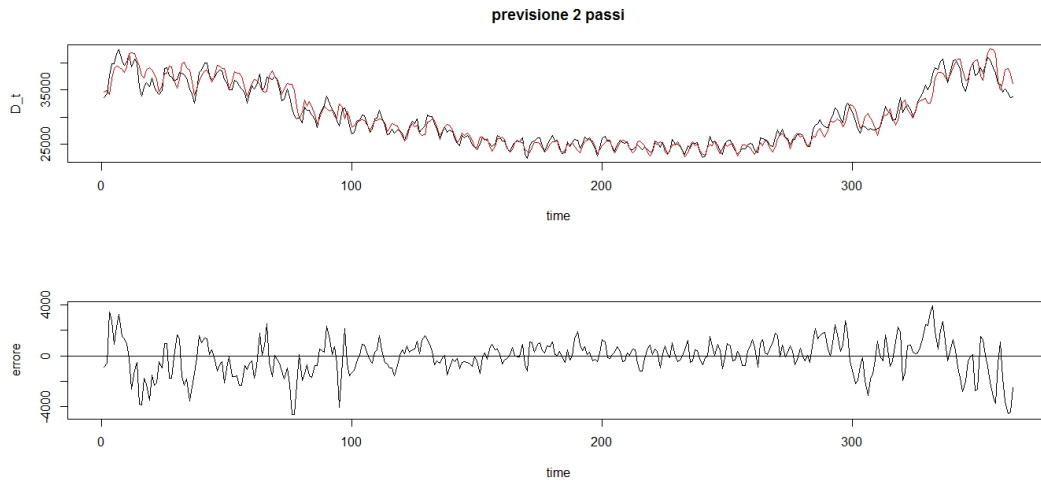


Figura 4.41: Fascia semi oraria 03.30-4.00 (minor picco). In alto: serie delle 365 previsioni un passo in avanti (colore rosso) per la serie della domanda  $D_{t,i}$  per  $i = 20$ . In basso: l'errore di previsione per il periodo 1 Gennaio-31 Dicembre 2010. Modello ARX(1).

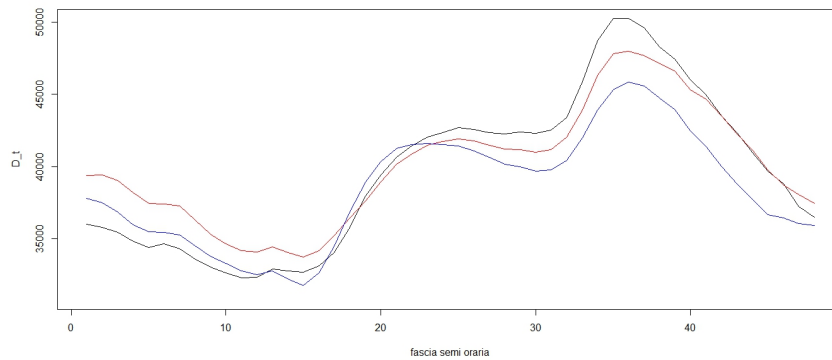


Figura 4.42: Previsione del giorno 2 Gennaio 2010: in rosso la previsione un passo in avanti, in blu la previsione a due passi e in nero il vero profilo giornaliero della domanda di energia.

### 4.2.3 Modello VAR(1)

E' interessante ora vedere, se con i modelli multivariati, i risultati sulle previsioni un passo e due passi, cambino.

Nella figura 4.43 vi è in alto la serie delle previsioni un passo in avanti per la fascia semi oraria 09.30-10.00 (colore rosso) che sembra seguire fedelmente l'andamento della domanda reale (colore nero).

In basso vi è, invece, l'errore di previsione un passo in avanti e si nota che, rispetto al caso univariato, l'errore, agli estremi, sembra essere sempre intorno allo zero, e ha dei picchi che arrivano a  $|4000|$ .

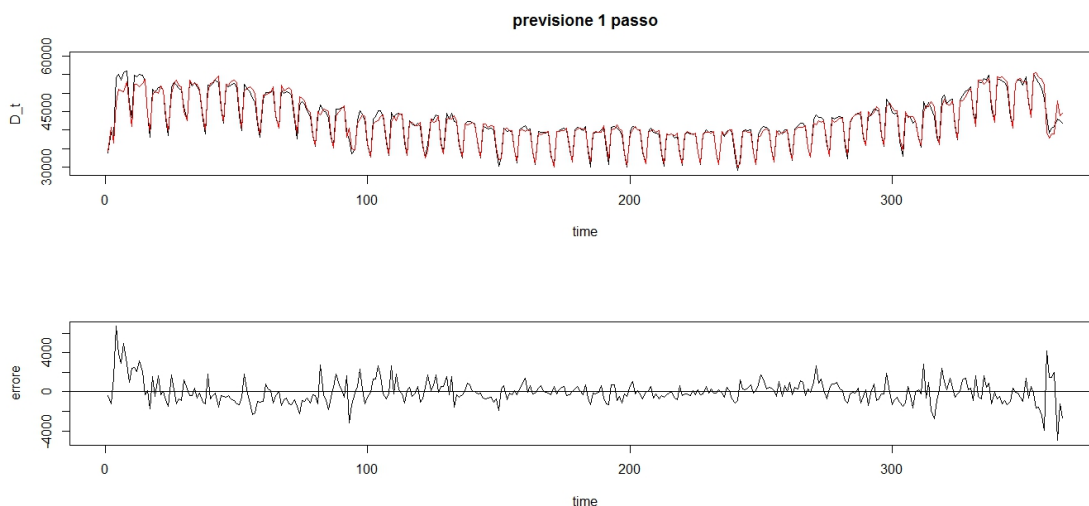


Figura 4.43: Fascia semi oraria 09.30-10.00. In alto: serie delle previsioni un passo in avanti (colore rosso) e la serie della domanda reale (colore nero). In basso serie dell'errore di previsione per il periodo 1 Gennaio-31 Dicembre 2010. Modello VAR(1).

Nella figura 4.44, è rappresentata, invece, la serie delle previsioni due passi in avanti (colore rosso) e in basso l'errore di previsione.

Si osservano dei picchi intorno ai  $|4000|$ , ma per quasi tutta la lunghezza della serie, l'errore di previsione è prossimo allo zero. Infatti, la previsione ad esso associata segue fedelmente l'andamento della serie reale (colore nero).

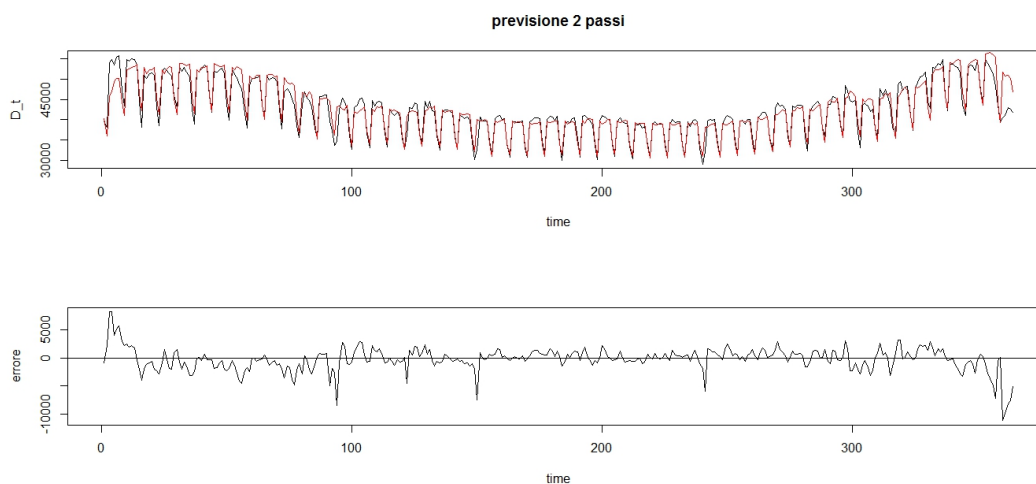


Figura 4.44: Fascia semi oraria 09.30-10.00. In alto: serie delle previsioni due passi in avanti (colore rosso) e la serie della domanda reale (colore nero). In basso serie dell'errore di previsione per il periodo 1 Gennaio-31 Dicembre 2010. Modello VAR(1)

Anche in questo caso si riportano anche i risultati per le previsioni un passo e due passi in avanti per la fascia semi oraria di minor picco, ovvero quella delle ore 3.30-4.00.

Nel grafico 4.45 vi è raffigurata la previsione e l'errore ad un passo e nella figura 4.46 quella a due passi in avanti. Si può notare che l'errore di previsione un passo in avanti è molto prossimo allo zero, e infatti la serie delle previsioni associata (in rosso) segue fedelmente l'andamento della domanda di energia. La previsione due passi in avanti si discosta un po di più dalla vera domanda, ma è comunque accettabile.

Nella figura 4.47, è rappresentata la previsione un passo in avanti (colore rosso) e due passi in avanti (colore blu) per la domanda giornaliera (colore nero). Entrambe le previsioni seguono l'andamento della domanda giornaliera, ma verso la fine sembrano sottostimare il profilo giornaliero della domanda reale.



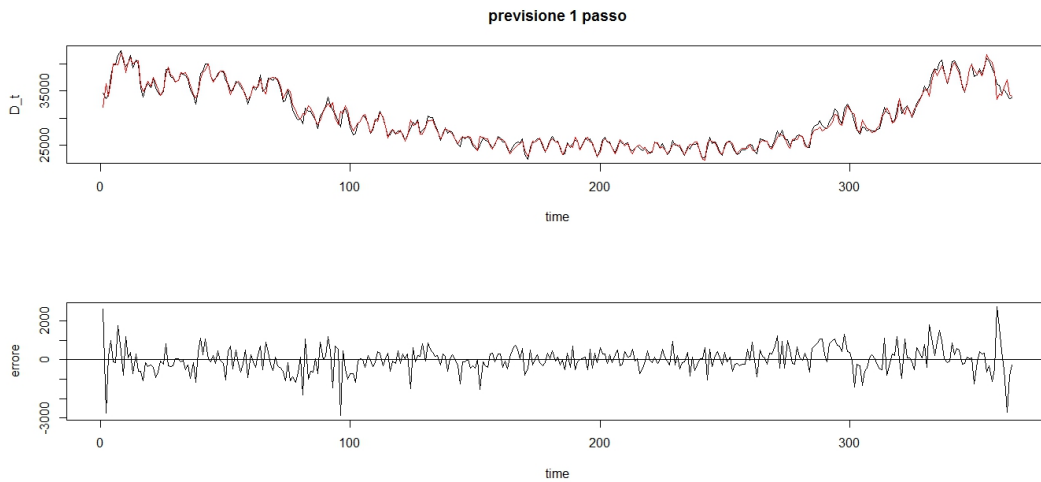


Figura 4.45: Fascia semi oraria 03.30-4.00. In alto: serie delle previsioni un passo in avanti (colore rosso) e la serie della domanda reale (colore nero). In basso serie dell'errore di previsione per il periodo 1 Gennaio-31 Dicembre 2010. Modello VAR(1)

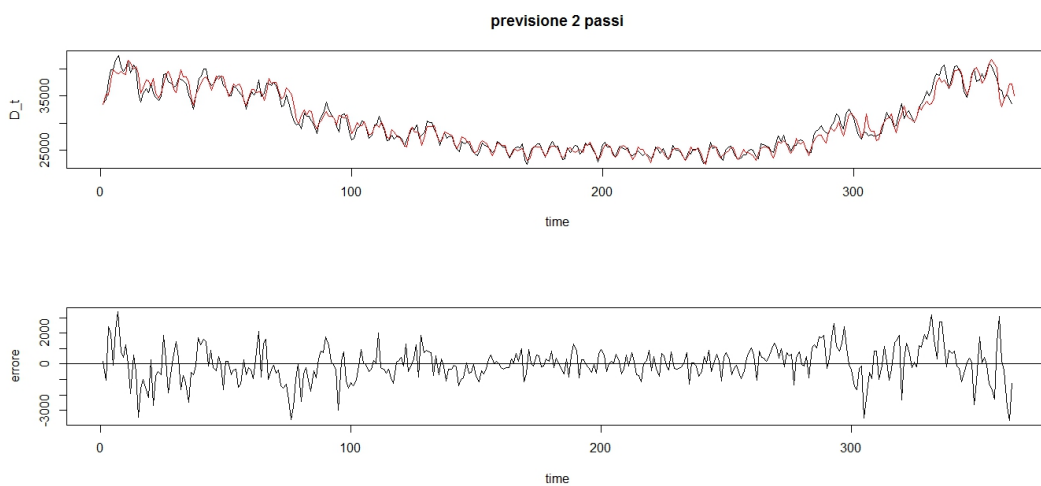


Figura 4.46: Fascia semi oraria 03.30-4.00. In alto: serie delle previsioni due passi in avanti (colore rosso) e la serie della domanda reale (colore nero). In basso serie dell'errore di previsione per il periodo 1 Gennaio-31 Dicembre 2010. Modello VAR(1)

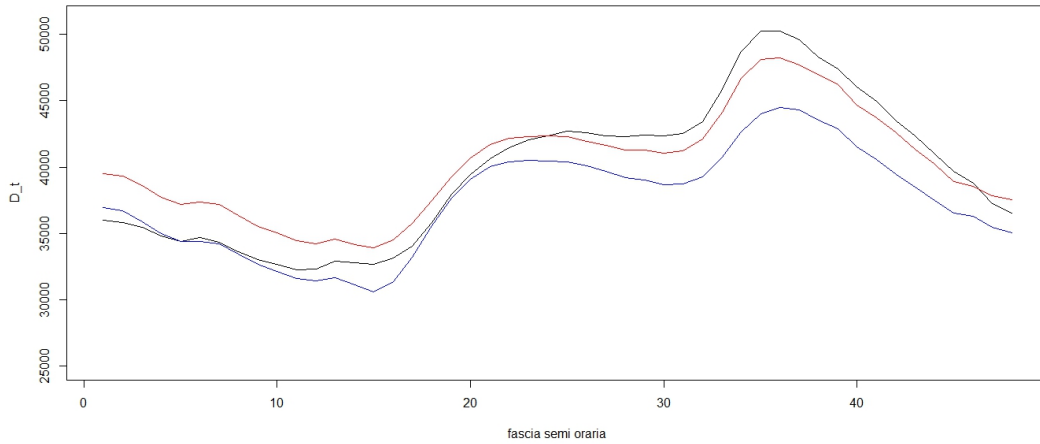


Figura 4.47: *Previsione del giorno 2 Gennaio 2010: in rosso la previsione un passo in avanti, in blu la previsione a due passi e in nero il vero valore giornaliero della domanda.*

#### 4.2.4 Modello VARX(1)

Anche in questo caso, per ottenere la previsione un passo e due passi si seguirà il consueto schema già presentato.

Come nel caso univariato, per la previsione di  $d_{t+2,i}$  il valore della temperatura al tempo  $t + 1$  viene considerato uguale al valore al tempo  $t$ , poichè il valore della temperatura non varia in modo netto da un giorno a quello successivo.

Nella figura 4.48 sono rappresentati, in alto la serie della previsione (colore rosso) e in basso l'errore di previsione per la fascia semi oraria relativa alle ore 09.30-10.00.

Si nota che l'errore di previsione, agli estremi, varia da  $-5000$  a  $5000$ , ma sostanzialmente è intorno allo zero.

Nella figura 4.49 vi è invece, rappresentato allo stesso modo, la previsione a due passi.

Qui l'errore ha dei picchi più elevati, ma nelle altre parti della serie è prossimo allo zero.

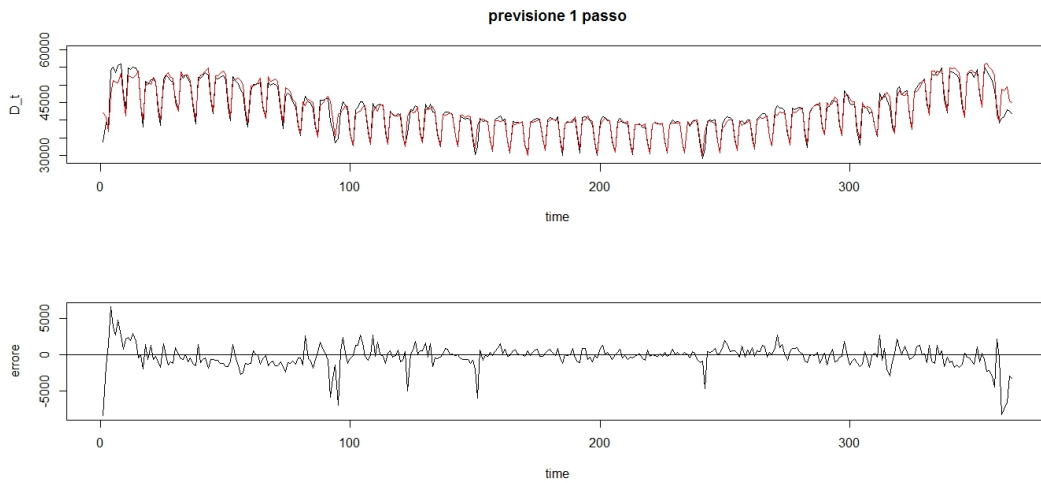


Figura 4.48: Fascia semi oraria 09.30-10.00. In alto: serie delle previsioni un passo in avanti (colore rosso) e la serie della domanda reale (colore nero). In basso serie dell'errore di previsione per il periodo 1 Gennaio-31 Dicembre 2010. Modello VARX(1)

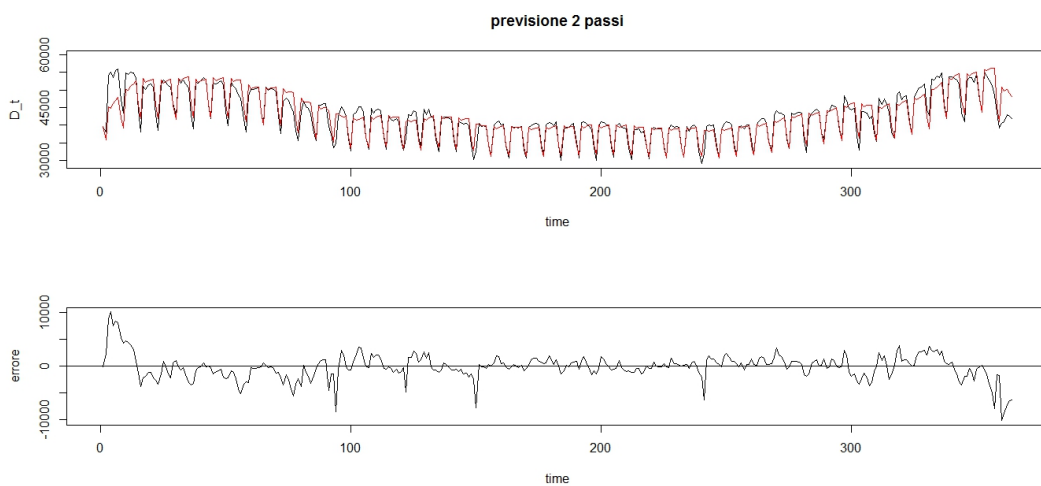


Figura 4.49: Fascia semi oraria 09.30-10.00. In alto: serie delle previsioni due passi in avanti (colore rosso) e la serie della domanda reale (colore nero). In basso serie dell'errore di previsione per il periodo 1 Gennaio-31 Dicembre 2010. Modello VARX(1)

Si riportano, inoltre, i risultati per la fascia semi oraria relativa alle ore 03.30-4.00. Nelle figure 4.50 e 4.51, vi sono rispettivamente la previsione e l'errore un passo e due passi in avanti.

Anche in questo caso si osserva che l'errore di previsione un passo in avanti è molto vicino allo zero, e infatti la previsione (colore rosso) segue in modo fedele l'andamento della serie della domanda. L'errore di previsione due passi in avanti è, invece sempre prossimo allo zero, ma più variabile all'inizio e alla fine della serie.

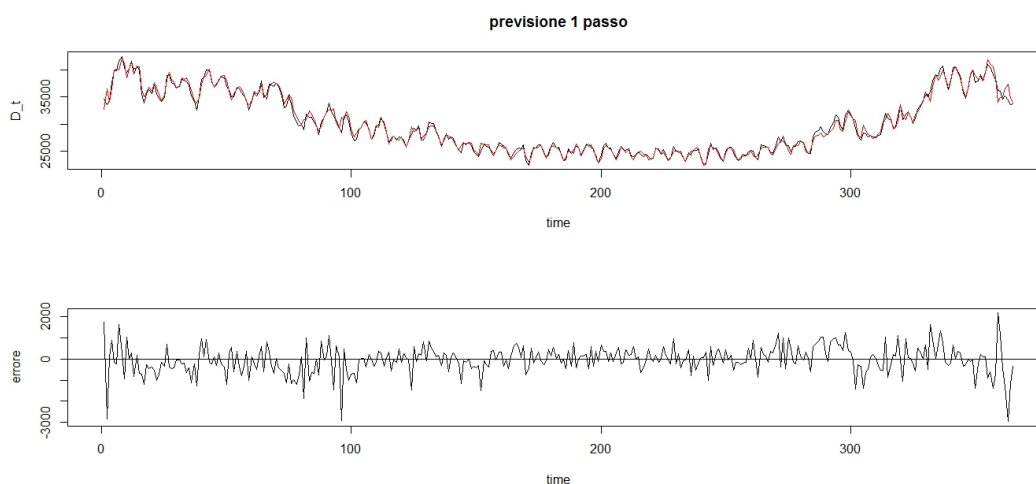


Figura 4.50: Fascia semi oraria 03.30-4.00. In alto: serie delle previsioni un passo in avanti (colore rosso) e la serie della domanda reale (colore nero). In basso serie dell'errore di previsione per il periodo 1 Gennaio-31 Dicembre 2010. Modello VARX(1)

Infine, nella figura 4.52, è rappresentata la previsione un passo (colore rosso) e due passi (colore blu) in avanti per un singolo giorno. Anche qui, nelle ultime fasce semi orarie, le previsioni sottostimano la vera domanda.

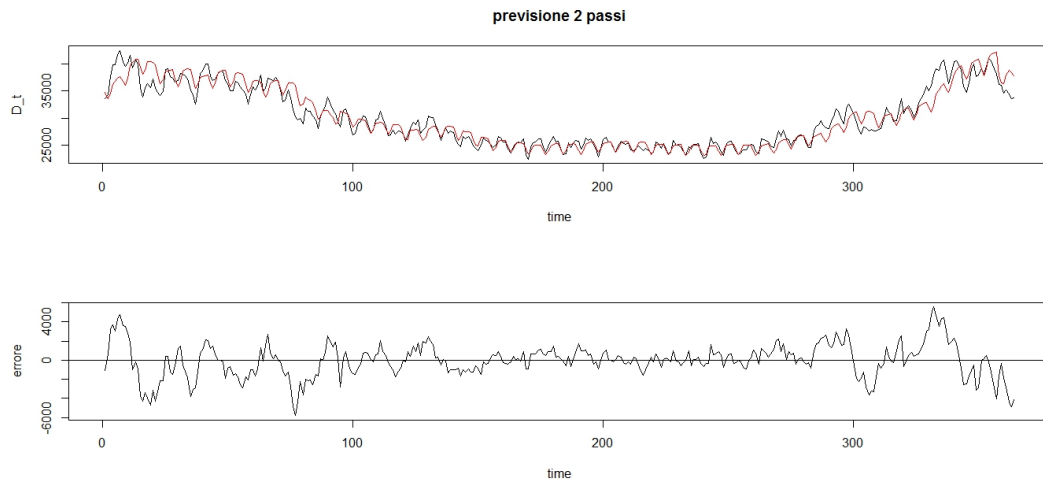


Figura 4.51: Fascia semi oraria 03.30-4.00. In alto: serie delle previsioni due passi in avanti (colore rosso) e la serie della domanda reale (colore nero). In basso serie dell'errore di previsione per il periodo 1 Gennaio-31 Dicembre 2010. Modello VARX(1)

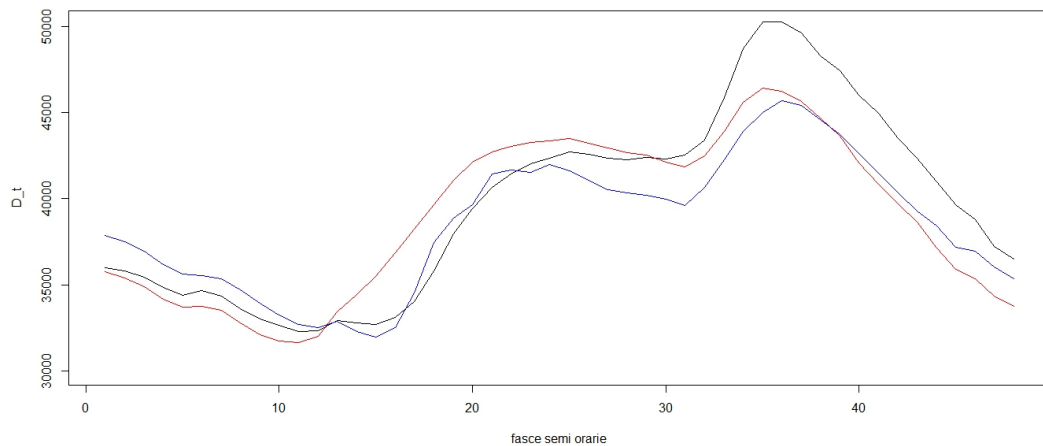


Figura 4.52: Previsione giorno 2 Gennaio 2010: in rosso la previsione un passo in avanti, in blu la previsione a due passi e in nero il vero valore giornaliero della domanda.

### 4.2.5 Modello AR funzionale

Si è stimato il modello autoregressivo funzionale con 9 funzioni base di Fourier fino al secondo periodo di osservazioni e ottenuto la previsione un passo e due passi in avanti. Successivamente, si è stimato il modello fino alla previsione un passo ottenuta precedentemente, e, in modo ricorsivo, si sono ottenute le previsioni un passo e due passi in avanti 365 volte.

In questo caso, si riportano i risultati per la fascia semi oraria relativa alle ore 09.30-10.00, e per la fascia oraria relativa alle ore 03.30-4.00, ma i risultati sono analoghi per le restanti fasce semi orarie.

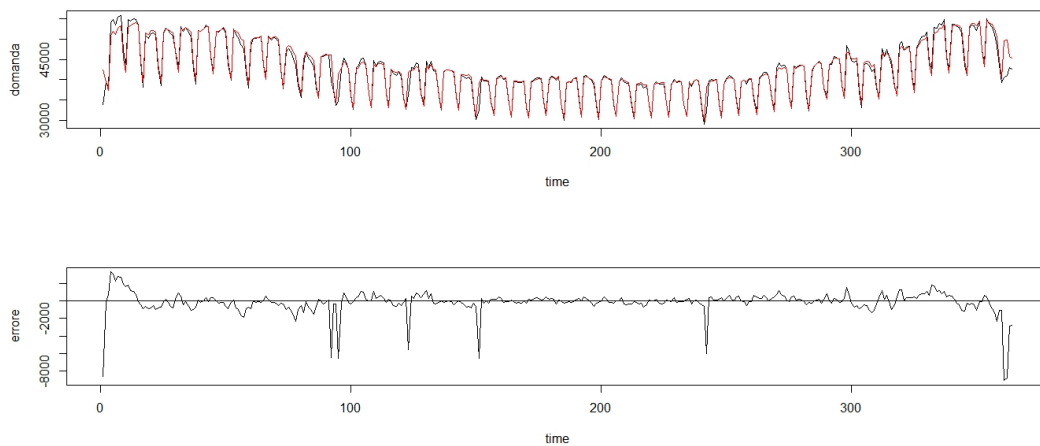


Figura 4.53: Fascia semi oraria 09.30-10.00. In alto: serie delle 365 previsioni un passo in avanti (colore rosso) e la serie della domanda reale di energia (colore nero). In basso: errore di previsione per il periodo 1 Gennaio-31 Dicembre 2010. Modello AR funzionale con 9 basi di Fourier.

Nella figura 4.53 vi è rappresentata la previsione un passo in avanti, e nella figura 4.54 la previsione due passi in avanti, per la fascia semi oraria 09.30-10.00.

Si nota che, per entrambe, la serie delle previsioni segue in modo fedele l'andamento della domanda di energia, tranne nelle prime e nelle ultime osservazioni.

L'errore di previsione un passo in avanti è quasi sempre prossimo allo zero, tranne in alcuni picchi in cui risulta negativo (quasi mai risulta positivo), e quindi la previsione risulta sottostimare la vera domanda. L'errore di previsione a due passi è anch'esso concentrato intorno allo zero, ma all'inizio e alla fine della serie si discosta dallo zero sia in positivo, sia in negativo.

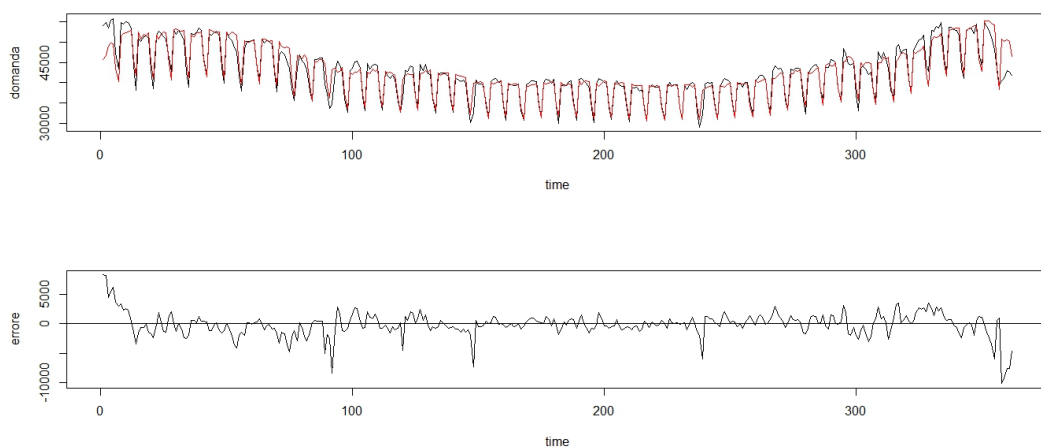


Figura 4.54: Fascia semi oraria 09.30-10.00. In alto: serie delle 365 previsioni due passi in avanti in avanti (colore rosso) e la serie della domanda reale di energia (colore nero). In basso: errore di previsione, per il periodo 1 Gennaio-31 Dicembre 2010. Modello AR funzionale con 9 funzioni di Fourier

Nelle figure 4.55 e 4.56 vi sono rispettivamente la serie delle previsioni e l'errore di previsione un passo e due passi in avanti per la fascia semi oraria relativa alle ore 03.30-4.00, cioè la fascia con la quantità di domanda minore. Si osserva che l'errore di previsione un passo in avanti è molto piccolo e prossimo allo zero, infatti la serie delle 365 previsioni un passo in avanti ad esso associata, segue fedelmente l'andamento della serie reale.

L'errore di previsione a due passi è invece un po' più variabile, ma sempre intorno allo zero.

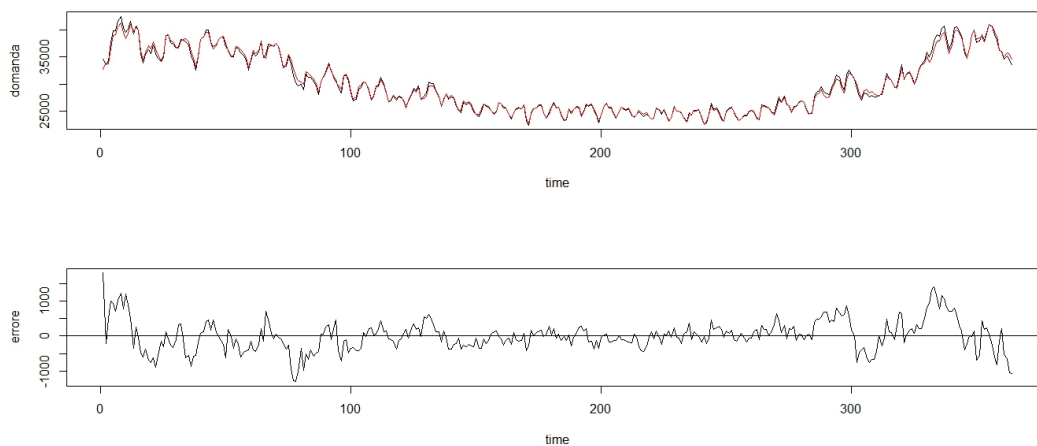


Figura 4.55: Fascia semi oraria 03.30-4.00. In alto: serie delle 365 previsioni un passo in avanti (colore rosso) e la serie della domanda reale di energia (colore nero). In basso: errore di previsione, per il periodo 1 Gennaio-31 Dicembre 2010. Modello AR funzionale con 9 basi di Fourier.

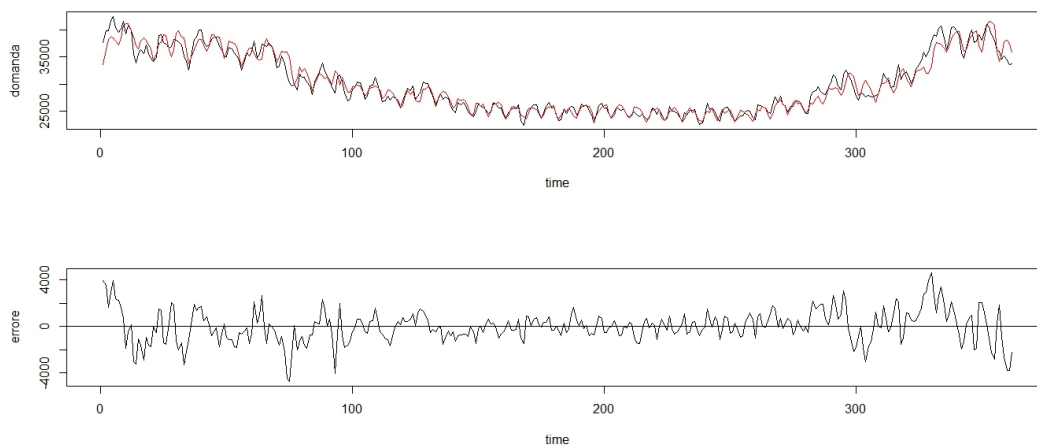


Figura 4.56: Fascia semi oraria 03.30-4.00 (fascia di minor picco). In alto: serie delle 365 previsioni un passo in avanti (colore rosso) e la serie della domanda reale di energia (colore nero). In basso: errore di previsione, per il periodo 1 Gennaio-31 Dicembre 2010. Modello AR funzionale con 9 basi di Fourier.



Nella figura 4.57 è rappresentata la previsione ad un passo (colore rosso) e la previsione a due passi (colore blu) per il giorno 5 Aprile 2005 (colore nero). E' evidente che per le prime fasce semi orarie entrambe le previsioni sottostimano la domanda reale, ma non di molto, e nella seconda parte delle fasce semi orarie sovrastimano di poco la domanda reale di energia.

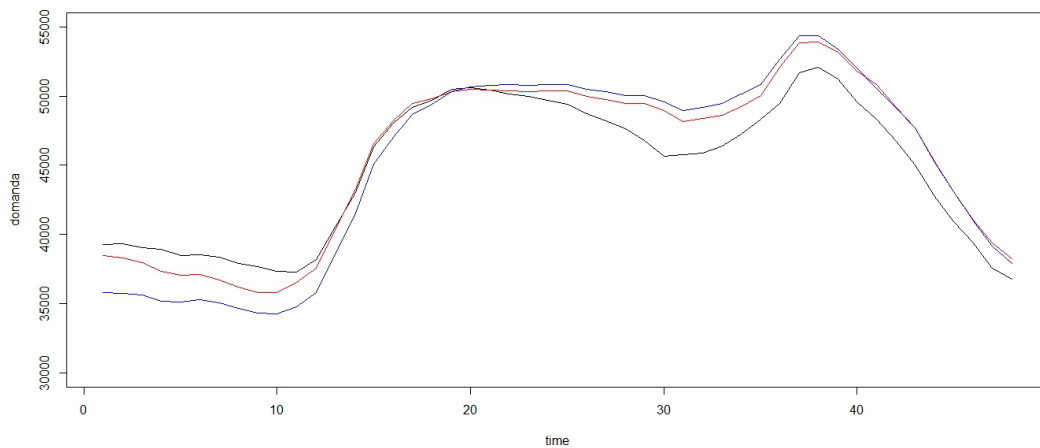


Figura 4.57: *Previsione giorno 5 Aprile 2005. In rosso la previsione un passo in avanti, in blu la previsione due passi in avanti e in nero il profilo giornaliero della vera domanda, per il modello AR funzionale con 9 funzioni di Fourier*

#### 4.2.6 Modello ARX funzionale

Si è stimato il modello ARX funzionale con 9 funzioni base *B-Spline* fino al secondo periodo di osservazioni e in modo analogo a quanto riportato precedentemente, si sono ottenute le previsioni un passo e due passi in avanti 365 volte.

E' da precisare che, per la previsione a due passi in avanti, il valore della temperatura media è stato considerato uguale al valore del periodo precedente, come fatto nell'analisi univariata e multivariata. Si riportano i risultati per la fascia semi oraria relativa alle ore 09.30-10.00, ma i risultati sono analoghi per le restanti fasce semi orarie.

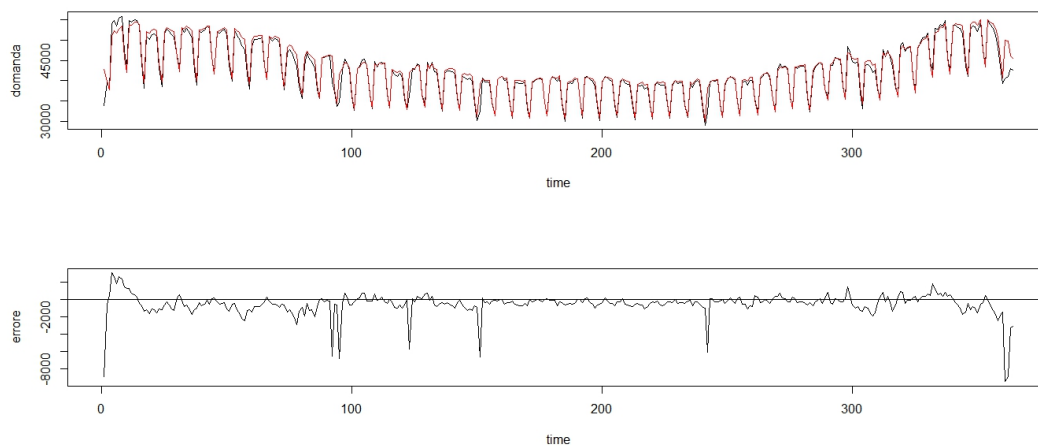


Figura 4.58: Fascia semi oraria 09.30-10.00. In alto: serie delle 365 previsioni un passo in avanti in avanti (colore rosso) e la serie della domanda reale di energia (colore nero). In basso: errore di previsione, per il periodo 1 Gennaio-31 Dicembre 2010. Modello ARX funzionale con 9 B-Spline.

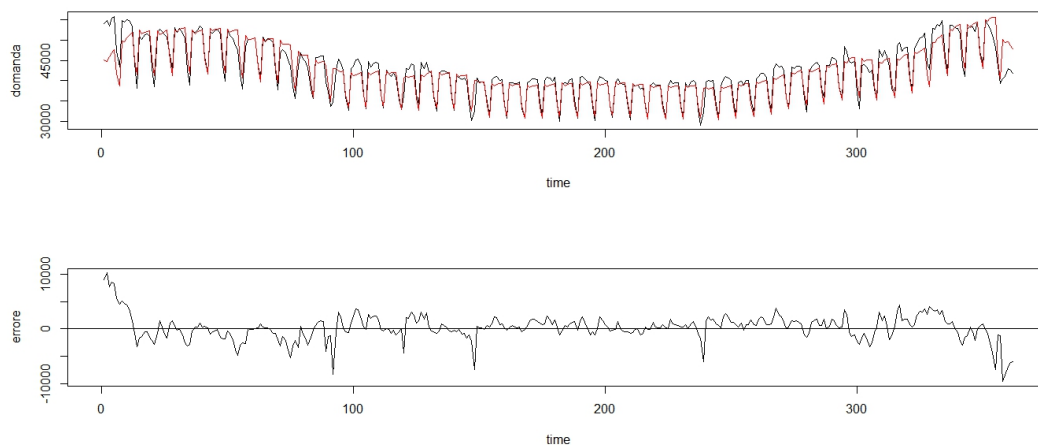


Figura 4.59: Fascia oraria 09.30-10.00. In alto: serie delle 365 previsioni due passi in avanti in avanti (colore rosso) e la serie della domanda reale di energia (colore nero). In basso: errore di previsione, per il periodo 1 Gennaio-31 Dicembre 2010. Modello ARX funzionale con 9 B-Spline.

Nella figura 4.58 e nella figura 4.59 vi sono, rispettivamente, la previsione e l'errore un passo in avanti e due passi in avanti per la fascia semi oraria 09.30-10.00.

Si osserva che la previsione un passo in avanti (colore rosso) segue l'andamento in modo fedele della domanda di energia, anche se in qualche punto sembra sovrastimare o sottostimare di poco la domanda reale.

L'errore di previsione un passo in avanti, come mostrato nel caso precedente, è sempre intorno allo zero, tranne in qualche punto d'osservazione della serie, in cui ha un picco in negativo, quasi mai in positivo.

Per quanto riguarda la previsione due passi in avanti, si nota che spesso sottostima la vera domanda di energia, ma in generale sembra essere una buona previsione. L'errore di previsione ad essa associato è sempre intorno allo zero, con qualche picco positivo e qualche negativo.

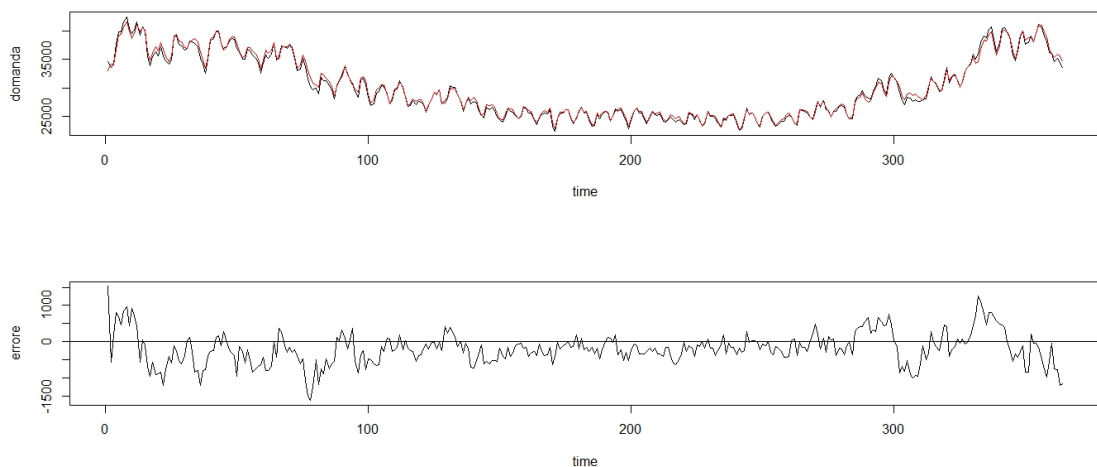


Figura 4.60: Fascia oraria 03.30-4.00. In alto: serie delle 365 previsioni un passo in avanti in avanti (colore rosso) e la serie della domanda reale di energia (colore nero). In basso: errore di previsione, per il periodo 1 Gennaio-31 Dicembre 2010. Modello ARX funzionale con 9 B-Spline.

Nelle figure 4.60 e 4.61 sono rappresentate rispettivamente la previsione e l'errore di previsione un passo e due passi in avanti per la fascia semi oraria 3.30-4.00. E' evidente che la previsione un passo in avanti segue in maniera fedele l'andamento della vera domanda e l'errore di previsione è molto contenuto e vicino allo zero. La previsione due passi in avanti è meno precisa e infatti l'errore di previsione all'inizio e alla fine della serie si discosta dallo zero e arriva fino a 4000.

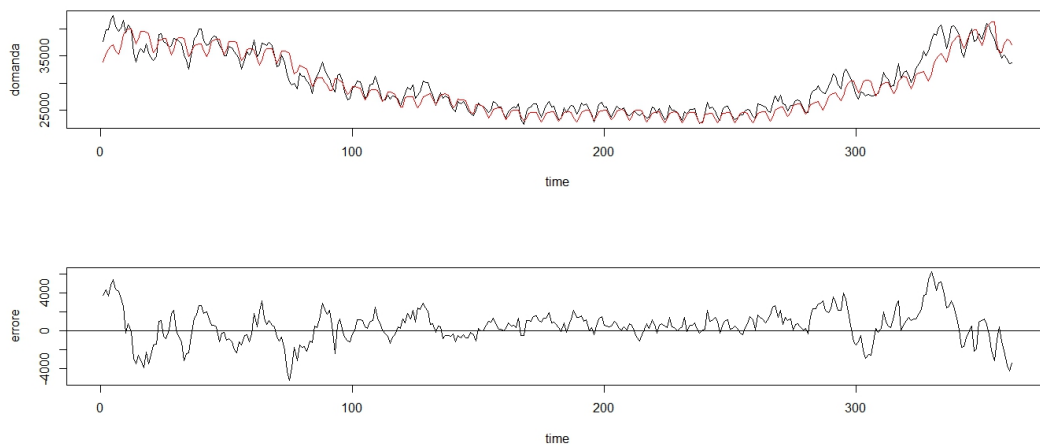


Figura 4.61: Fascia oraria 03.30-4.00. In alto: serie delle 365 previsioni due passi in avanti in avanti (colore rosso) e la serie della domanda reale di energia (colore nero). In basso: errore di previsione, per il periodo 1 Gennaio-31 Dicembre 2010. Modello ARX funzionale con 9 B-Spline.

Nella figura 4.62 si può osservare la previsione un passo in avanti (colore rosso) e due passi in avanti (colore blu) in confronto con la vera domanda di energia (colore nero), per il giorno 5 Aprile 2005. Si osserva che, anche in questo caso, nella prima parte delle fasce semi orarie le previsioni tendono a sottostimare la domanda, mentre, nella seconda parte, tendono a sovrastimarla.

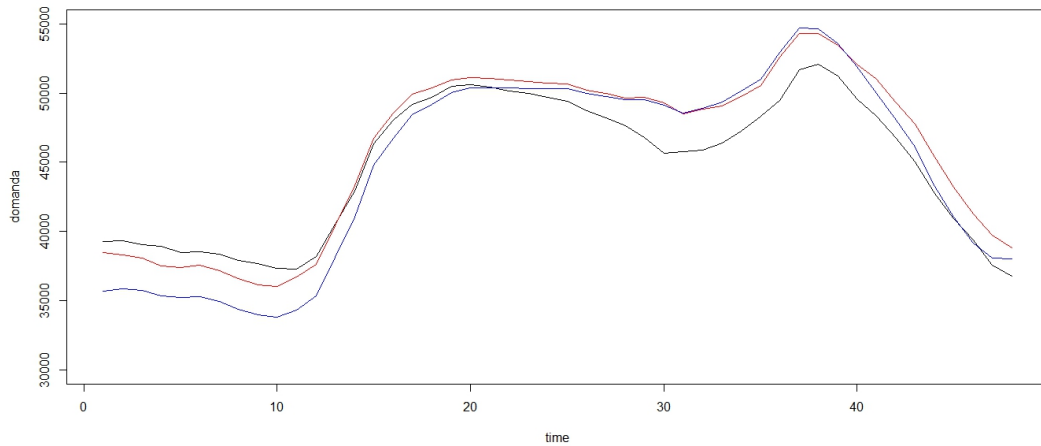


Figura 4.62: *Previsione giorno 5 Aprile 2005. In rosso la previsione un passo in avanti, in blu la previsione due passi in avanti e in nero il profilo giornaliero della vera domanda, per il modello AR funzionale con 9 funzioni B-Spline*

### 4.3 Confronto finale

Si può, in conclusione, raggruppare in due tabelle tutti i risultati ottenuti per le previsioni un passo (tabella 4.3) e due passi in avanti (tabella 4.4), dai modelli univariati, multivariati e funzionali.

Dalla tabella si evince che l'analisi univariata condotta sulle singole serie separate ha i valori degli indici di bontà delle previsioni più elevati di tutti i modelli.

Questo era intuibile in quanto considerando le fasce semi orarie come singole serie si perde la dipendenza tra la domanda di una semi ora con la domanda della semi ora successiva. Ricordando, infatti, i primi grafici, si era osservato che la domanda di energia giornaliera aveva una componente legata all'ora del giorno in cui ci si trova.

L'analisi multivariata ha risultati migliori rispetto a quella univariata, ed è comprensibile visto il fatto che si prendono in considerazione le fasce semi orarie come un'unica serie multivariata. L'analisi multivariata studia quindi una certa dipendenza tra le diverse fasce semi orarie. Plausibilmente si in-

fluenzeranno di più fasce semi orarie continue rispetto a fasce più lontane. L'analisi funzionale ha, invece, i risultati migliori rispetto a tutti gli altri tipi di analisi. Questo è accettabile dato che l'analisi funzionale considera le osservazioni di un giorno come un'unica funzione e non come punti singoli.

La differenza con l'analisi multivariata risiede nel fatto che quella funzionale cerca di sintetizzare al meglio le singole osservazioni in un'unica curva, che racchiude tutte le informazioni legando assieme la varie fasce semi orarie.

Un altro fatto è che, nell'analisi funzionale, il modello di regressione si applica su curve che sono state lisce in precedenza. Questo potrebbe influire sulla differenza di risultati tra l'analisi multivariata e funzionale.

Dai risultati si nota che esiste una differenza nell'usare una determinata funzione base rispetto ad un'altra, infatti è evidente che la funzione base *B-Spline* per il modello ARX funzionale, che considera quindi anche le informazioni fornite dal meteo, ha i valori degli indici di bontà delle previsioni minori in assoluto.

Analoghe considerazioni si possono fare guardando i valori degli indici due passi in avanti.

Si nota, anche in questo caso, come i valori degli indici sull'errore di previsione siano nettamente minori per l'analisi funzionale rispetto all'analisi univariata e multivariata. Anche per la previsione due passi in avanti, il modello ARX funzionale ha i valori degli indici più bassi in assoluto.

Previsioni un passo	MAE	MAPE	RMSQ
AR funzionale	631.91	0.01617	1036.55
ARX funzionale	551.25	0.01400	954.63
AR(1)	928.71	0.02412	1347.93
ARX(1)	952.05	0.02464	1387.25
VAR(1)	783.77	0.01993	1095.11
VARX(1)	854.50	0.02179	1262.25

Tabella 4.3: Indicatori sull'errore di previsione un passo in avanti di tutti i modelli scelti. Per il modello AR funzionale si considera il modello con 9 funzioni base di Fourier. Per il modello ARX funzionale si considera il modello con 9 funzioni B-Splines.

Previsioni due passi	MAE	MAPE	RMSQ
AR(1) funzionale	1484.14	0.03811	2044.77
ARX(1)funzionale	1212.72	0.03127	1714.85
AR(1)	2238.84	0.05976	3389.87
ARX(1)	2273.81	0.06059	3420.61
VAR(1)	2141.83	0.05704	3302.74
VARX(1)	2415.29	0.06423	3536.50

Tabella 4.4: Indicatori sull'errore di previsione due passi in avanti di tutti i modelli scelti. Per il modello AR funzionale si considera il modello con 9 funzioni base di Fourier. Per il modello ARX funzionale si considera il modello con 9 funzioni B-Splines.





# Conclusioni

In questa tesi si è voluta applicare l'analisi dei dati funzionali per stimare e prevedere la domanda di energia del mercato elettrico inglese, e confrontarla con metodi più classici, come l'analisi univariata e multivariata.

I risultati ottenuti dimostrano il fatto che l'analisi funzionale ha migliori *performance*, in termini previsivi, dell'analisi univariata e multivariata. Questo perchè l'analisi funzionale riesce a raccogliere tutte le informazioni derivanti dalla domanda di energia giornaliera, in un'unica funzione, che è l'oggetto di studio.

Una strada per migliorare ulteriormente i risultati dell'analisi funzionale potrebbe essere quella di prendere in considerazione un modello di regressione lineare, non più di tipo *concurrent* dove l'influenza tra le variabili è contemporanea, ma un modello dove la variabile risposta  $y(i)$  è influenzata da valori di  $z(h)$  con  $h$  diverso da  $i$ .

Un altro approccio potrebbe considerare di non lavorare sulla serie depurata dalle componenti deterministiche, ma applicare questa decomposizione alla domanda di energia iniziale all'interno dell'analisi funzionale. Ovvero stimare un modello di regressione funzionale in cui le covariate sono le variabili *dummy* corrispondenti ai giorni e agli effetti calendario.

Un'altra via potrebbe essere quella di prendere in considerazione l'analisi delle componenti principali funzionale sempre applicata alla serie iniziale della domanda.

Ulteriori studi futuri si possono condurre studiando il problema con approcci differenti dall'analisi funzionale o dall'analisi multivariata. Alcune alternati-

ve possono essere date dalla considerazione di altri tipi di variabili esplicative, non soltanto i dati forniti dal meteo, ma ad esempio i dati forniti da altre fonti di energia, o dall'introduzione della serie dei prezzi dell'energia.

# Ringraziamenti

Ormai lo ripeto da anni che devo scrivere un libro, beh questo è come un libro per cui ora userò queste ultime pagine per ringraziare pubblicamente chi ha contribuito al raggiungimento di questo importante traguardo.

Un grande grazie va a mamma e papà, che mi hanno lasciato studiare a tempo pieno, e che mi hanno sostenuto e sopportato sempre.

Devo ringraziare poi mia sorella, anzi, devo ringraziare Matteo che finalmente se l'è portata a casa sua. Ovviamente una sorellona ti capisce anche quando non parli.

Poi c'è Gian, che ormai da un po' d'anni mi sopporta, mi supporta, mi aiuta, ma soprattutto rende la mia vita molto più bella. Devo dire che il tentativo di fuggire lontano non ti è riuscito bene, perchè ti seguirò anche lì.

Un grazie va alla mia amica Lara, che sentirci cento volte al giorno non basta. In questi ultimi mesi è stata la mia roccia e sempre un'ottima spalla su cui disperarmi.

Devo ringraziare anche Ilaria, che mi ha risolto più volte i miei problemi al computer e che in tutti questi anni (troppi veramente) c'è sempre stata per me.

Non posso non ringraziare tutti i miei amici che hanno reso fantastici gli anni delle superiori, e con i quali ho trascorso delle vacanze indimenticabili.

E poi tutti gli amici dell'università, senza di voi sarebbe stato tutto noioso. Chi si scorderà mai i pomeriggi in aula studio, i pranzi in mensa, le feste e le vacanze.

Quindi, cercando di non dimenticare nessuno, grazie a Marta, Martina, Ali-

ce, Andrea, Federico, Ilaria, Chiara, Laretta, Lara, Laura, Valeria, Giosuè, Marco (la prima persona a cui ho parlato il primo giorno d'università), Gian, Giobi, Jo, Scogna, Dipper, Mattia, Omar, Tommy, Tommaso, Daniele. Ho sicuramente dimenticato qualcuno, ma grazie veramente a tutti.

Un ultimo grazie anche a tutte le persone pazze conosciute ai concerti, che hanno reso tutto più bello, Cristina, Lisa, Silvia, Ilaria e tutti gli altri.

La fine non è che un nuovo inizio.

# Riferimenti bibliografici

- Andersson J. e Lillestøl J- (2010). *Modeling and forecasting electricity consumption by functional data analysis*, The Journal of Energy Markets (3 – 15) Volume 3, Num. 1.
- Azzalini A. e Scarpa B. (2004). *Analisi dei dati e data mining*, Springer.
- Blum M. e Riedmiller M. (2013). *Electricity Demand Forecasting using Gaussian Processes*, The AAAI-13 Workshop on Trading Agent Design and Analysis (TADA-13)
- Cappuccio N. e Orsi R. (2005). *Econometria*, Il mulino.
- Conejo A. et al. (2005). *Day-Ahead Electricity Price Forecasting Using the Wavelet Transform and ARIMA Models*, Ieee Transactions on power system, Volume 20, No. 2.
- Febrero-Bande M. e Oviedo de la Fuente M. (2012). *Statistical Computing in Functional Data Analysis: The R Package fda.usc*, Journal of Statistical Software, Volume 51, Issue 4.
- Lütkepohl H. (2005). *New Introduction to Multiple Time Series Analysis*, Springer.
- Ramsey J. e Silverman B. (2002). *Applied Functional Data Analysis: Methods and Case Studies*, Springer.
- Ramsey J. e Silverman B. (2005). *Functional Data Analysis, Second Edition*, Springer.

Ramsey J. (1996). *Functional data analysis of lip motion*, Journal of the Acoustical Society of America, 99, 3718-3727.

Ramsey J., Hooker G., Graves S. (2009). *Functional Data Analysis with R e Matlab*, Springer.

### **Siti web**

<http://www.agienergia.it>

<http://cran.r-project.org/web/packages/fda>

<http://www.energiaspiegata.it>

<http://www.functionaldata.org>

<http://www.jstatsoft.org/v51/i04/paper>

<http://www.mercatoelettrico.org>