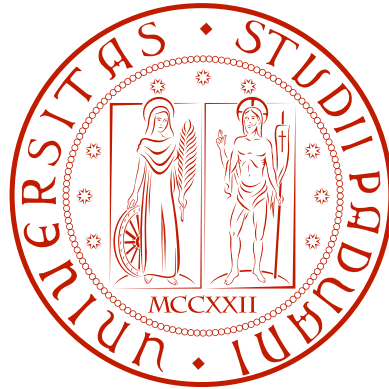


Università degli Studi di Padova
Corso di Laurea Magistrale in Scienze Statistiche



**Previsione dei picchi dei prezzi del
mercato elettrico Inglese mediante
previsione dei quantili**

Relatore: Prof. Francesco Lisi
Dipartimento di Scienze Statistiche

Laureando: Fabio Tasinato

Anno Accademico 2011/2012

Prima o poi tutto finisce, soprattutto le esperienze positive.

A volte arriva il momento di fare spazio agli altri, di prendere il posto di qualcun altro o semplicemente di cambiare.

A volte sembra che stiamo scrivendo le ultime pagine di un ciclo, non rendendoci conto che stiamo scrivendo solo le prime parole di qualcosa di nuovo.

Ringrazio tutto ciò e tutti quelli che mi hanno fatto crescere in questi anni, con esperienze positive o meno. In particolar modo a coloro i quali mi hanno maggiormente aiutato e capito.

A coloro i quali sono passati e a chi non passer mai.

INDICE

Indice	i
1 I Mercati Elettrici	5
1.1 Introduzione ai Mercati Elettrici	5
1.2 Fatti Stilizzati dei Prezzi Elettrici	9
2 Il Mercato Inglese UKPX	21
2.1 Stazionarietà	29
2.2 La Decomposizione delle Serie	35
2.2.1 Regressione mediante Spline	36
2.2.2 Componente di lungo periodo e stagionalità annuale	38
2.2.3 Stagionalità settimanale giornaliera ed effetto giorni festivi	40
2.2.4 Componente stocastica	45
2.3 I Picchi	50
2.4 Modelli di previsione dei picchi di prezzo	55
3 Modelli di Previsione dei prezzi futuri e dei picchi di prezzo	61
3.1 Modello di Regressione Lineare Classico	62
3.2 Applicazione del modello di Regressione Lineare ai Dati	64
3.3 Limiti del modello di Regressione Lineare Classico	81
4 La Regressione del quantile	87
4.1 Quantili e strumenti	88
4.2 La regressione del quantile	92
4.2.1 Proprietà	93
4.2.2 Inferenza e risultati asintotici	98
4.3 Regressione del quantile non parametrica	103
5 Applicazioni della Regressione del Quantile ai prezzi elettrici del mercato UK	105
5.1 Modello di Regressione del quantile lineare	105
5.2 Modello di regressione del quantile non parametrico	122
5.3 Scelta e valutazioni dei modelli	126
A Immagini	135
B Tabelle	149
Bibliografia	165

INTRODUZIONE

In un contesto di liberalizzazioni che hanno portato alla creazione dei mercati elettrici nei quali la contrattazione e lo scambio dell'energia elettrica tra operatori e soggetti sono nettamente cambiati, diviene interessante studiare il comportamento di questi mercati. I mercati elettrici hanno una storia relativamente recente e possiedono caratteristiche specifiche e difficilmente riscontrabili in altri contesti, alcuni aspetti sono ancora oggi argomento di studio. Le borse elettriche hanno un ruolo fondamentale nel garantire la copertura di un servizio ai nostri giorni fondamentale.

In questo elaborato si presenteranno, nel capitolo 1, le caratteristiche generali dei mercati dell'energia elettrica e se ne evidenzieranno i principali fatti stilizzati concentrandoci particolarmente sui mercati *day-ahead*, ovvero i mercati dove le operazioni di scambio dell'energia elettrica avvengono per la giornata successiva a quella delle contrattazioni. Evidenziate le principali caratteristiche della serie dei prezzi *spot*, si possono notare alcune caratteristiche pressoché uniche rispetto ai prezzi di qualsiasi altro mercato con un sottostante reale, legata appunto alle caratteristiche dell'energia elettrica. Tra le varie caratteristiche emerge una componente fondamentale dei prezzi elettrici, la presenza dei picchi di prezzo. I picchi sono salti visibili nella serie dei prezzi e sono presenti pressoché in tutti i mercati elettrici. Si può pensare ad un picco come ad un prezzo che superi una certa soglia. Riuscire ad anticipare i picchi di prezzo può essere utile ad esempio per migliorare le operazioni di *trading* ma anche per altre ragioni legate alla modellazione di alcune caratteristiche dei prezzi stessi.

L'obiettivo di tale tesi è quello di testare e proporre alcuni modelli di previsione dei picchi dei prezzi dell'energia elettrica nel mercato Inglese per il giorno

successivo alle contrattazioni. In letteratura, anche negli studi più recenti, emerge la difficoltà e la mancanza di una regola fissa su come definire i picchi dei prezzi elettrici, sia in fase di individuazione ma soprattutto in fase di previsione. È inoltre dibattuta la validità di un metodo di previsione rispetto ad un altro e non si è ancora stabilito se un metodo, in termini di capacità previsionali rispetto ad altri, sia migliore. Nel capitolo 2 vengono presentate le caratteristiche fondamentali del mercato elettrico Inglese e i dati utilizzati. I 5 anni di osservazioni giornaliere sono divise in fasce semi orarie. In tale capitolo si definisce uno studio delle serie trattate e una loro depurazione necessaria per definire opportuni modelli per la previsione dei picchi per il giorno successivo.

La proposta sostenuta in questo elaborato, per prevedere i picchi di prezzo che si realizzeranno nel giorno successivo, è quella di prevedere opportuni quantili del prezzo dell'energia elettrica con modelli di regressione. Si potrebbe pensare ad una previsione di un picco se il quantile dei prezzi previsto supera una certa soglia opportuna.

Nel capitolo 3 si espone come una previsione di quantili dei prezzi mediante opportuni modelli di regressione possa essere impiegata per prevedere i picchi dei prezzi elettrici. Questo concetto è tutt'altro che affermato in letteratura. Inizialmente, viene impiegato un modello di regressione lineare classico per la previsione dei quantili in modi differenti. I limiti teorici ed applicativi e il fatto che si stiano prevedendo quantili dei prezzi (e non più una media condizionata) sposta l'attenzione verso un altro metodo di regressione. La stima e la previsione di quantili condizionati è ottenibile mediante: la regressione del quantile; un metodo di regressione proposto proprio per questo scopo. Nel capitolo 4 vengono riassunti i risultati principali e le caratteristiche fondamentali della regressione del quantile introdotta da Koenker e Bassett (1978). Questo metodo di regressione risulta lo strumento naturale per la previsione di quantili condizionati. Nel capitolo 5 si applica la regressione del quantile ai dati del mercato elettrico Inglese. Si propongono e applicano quindi modelli di regressione del quantile parametrici e non parametrici. Verranno discussi i risultati sulla capacità previsionale dei

picchi e confrontati con quelli derivanti da altri approcci differenti da quello della regressione del quantile. Si traggono le conclusioni finali sull'approccio legato alla previsioni dei picchi di prezzo sfruttando previsioni di quantili di prezzo per il giorno successivo evidenziando eventuali pregi e limiti di questo approccio.

Le analisi riportate nel capitolo 2 e le applicazioni discusse nei capitoli 3 e 5 sono state realizzate utilizzando il software statistico R (R Development Core Team, 2012) sfruttando principalmente il pacchetto "*quantreg*" sviluppato da Koenker (2010) e molte altre funzione da me ideate. Questa tesi è stata scritta in L^AT_EX.

Capitolo 1

I MERCATI ELETTRICI

In questo capitolo verranno descritte le principali caratteristiche dei mercati elettrici. Si presenteranno fatti stilizzati solitamente comuni anche se in mercati differenti.

1.1 Introduzione ai Mercati Elettrici

Il mercato elettrico, meglio noto anche come borsa elettrica, vede la sua comparsa nei paesi industrializzati in date differenti dalla fine degli anni '80 – inizio degli anni '90, grazie al processo di liberalizzazione dell'energia elettrica decretato dai vari governi per la creazione di mercati elettrici unici per ogni paese. La nascita dei mercati elettrici ha dato la possibilità di creare un sistema elettrico nazionale concorrenziale che consenta quindi un teorico miglioramento dell'efficienza in tutta la filiera elettrica e una riduzione del prezzo dell'energia. La filiera elettrica è scomponibile in:

- *generazione*, la vera e propria fase di produzione dell'energia con vari tipi di tecnologia esistente (stazioni idroelettriche, impianti nucleari, centrali termoelettriche a carbone, gas o petrolio);
- *trasmissione*, la fase di passaggio dell'energia dalle centrali alla rete elettrica (alto voltaggio);
- *dispacciamento*, la fase complessa di bilanciamento, monitoraggio e sincronizzazione della rete per evitare *black out* all'intero sistema;
- *commerciale e vendita*, ovvero la gestione dei rapporti con i clienti finali;

La presenza della borsa elettrica ha dato la possibilità, in Italia come in ogni altro paese, di far sorgere prezzi di equilibrio e la possibilità di acquistare l'energia in modo trasparente, sicuro e al prezzo più conveniente tra produttori e grossisti.

I principali vantaggi della liberalizzazione delle contrattazioni elettriche, che in precedenza venivano effettuate solo tramite contratti bilaterali, sono:

- migliorare il processo di formazione dei prezzi;
- stimolare la concorrenza, permettendo l'entrata e l'uscita dal mercato di nuove aziende nelle varie posizioni della filiera elettrica e soddisfare la domanda ai prezzi più convenienti;
- operare in trasparenza, mancanza di asimmetrie informative e una regolamentazione nei negoziati omogenea;
- favorire la flessibilità della produzione e migliorare la risposta ad una domanda di energia che per sua natura può fluttuare e cambiare anche velocemente;
- sopperire alla scarsità di offerta, solitamente legate ad una certa area geografica di un paese;
- stabilizzare il mercato, incentivando al raggiungimento di un opportuno numero di centrali, linee elettriche e presenza di operatori;
- separare l'attività di produzione e di vendita creando due attività concorrenziali all'interno della filiera. La liberalizzazione separa principalmente le fasi di produzione e dispacciamento in prima battuta, fasi dove solitamente sarebbe presente un monopolio naturale che le unifica;

Il mercato elettrico si presenta come un mercato telematico nel quale il prezzo di equilibrio (per una data ora o semi ora) deriva dall'incontro tra domanda ed offerta di energia elettrica degli operatori partecipanti (produttori e distributori di grandi dimensioni). L'intersezione tra la domanda e l'offerta di energia non solo stabilisce un prezzo di equilibrio (*market clearing price* – MCP), ma stabilisce

un volume di equilibrio complessivamente scambiato (*market cleaning volume* – MCP). In figura 1.1 si mostra il *market cleaning price* (MCP) e il *market cleaning volume* (MCV) ottenuto dall'intersezione della curva di offerta (costruita da offerte di fornitura aggregata) e la curva di domanda (costruito da offerte di acquisto aggregato). Questo deriva da un sistema d'asta bilaterale dove l'offerente (che verrà considerato nell'offerta aggregata) si propone di vendere una certa quantità di energia ad un prezzo minimo e il richiedente (che parteciperà alla domanda aggregata) si propone di acquistare una certa quantità di energia ad un prezzo massimo. Le offerte accettate, i piani che regolano immissione e prelievo vengono

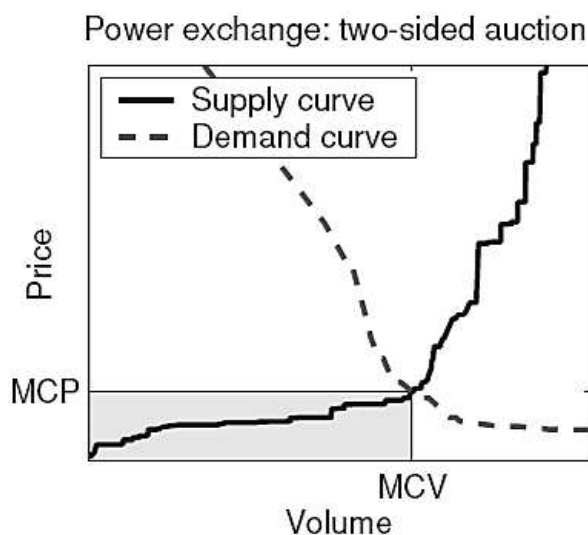


Figura 1.1: Meccanismo di formazione del prezzo in un mercato liberalizzato lasciato all'iniziativa privata di produttori, distributori e *traders* (Weron, 2006).

no sempre definite sulla base dell'incontro di domanda e offerta da un sistema d'asta bilaterale (Weron 2006).

I mercati dell'energia si dividono solitamente in mercati a pronti (Mercato del Giorno Prima (MGP) detto anche mercato *day-ahead*, Mercato Infra giornaliero (MI) e Mercato del Servizio di Dispacciamento (MSD)) e mercati a termine, con l'obbligo della consegna fisica dell'energia e la possibilità di negoziare l'energia elettrica su orizzonti temporali più estesi di quelli giornalieri tipici del mercato *day-ahead*.

La caratteristica dei mercati *day-ahead*, o mercati del giorno prima, è quella che l'energia elettrica viene contrattata e scambiata per il giorno successivo attraverso

un meccanismo d'asta che dà il diritto d'immettere e prelevare energia dalla rete. Le contrattazioni avvengono su base oraria o semi oraria, a seconda del mercato in cui ci si trova; viene stabilito un prezzo, una quantità di energia da scambiare e altri servizi secondari per ogni singola ora (esempio per la fascia 12.00–13.00) oppure per ogni singola fascia semi oraria (esempio per 12.30–13.00) per il giorno successivo, solitamente ciò avviene in un'unica asta giornaliera.

Il mercato elettrico è un mercato particolare date le caratteristiche dell'*asset* fisico che viene scambiato a differenza di altri mercati finanziari con un sottostante reale. Le caratteristiche dell'energia elettrica, quali ad esempio l'impossibilità di essere immagazzinata efficientemente e in grandi quantità, gli innumerevoli vincoli tecnici molto stringenti per il funzionamento dell'intera filiera (bilanciamento, mantenimento della tensione nella rete, sincronizzazione e così via), le caratteristiche della domanda e solitamente anche la dimensione territoriale in cui si opera, si riflettono consistentemente nei prezzi elettrici.

Il livello dei prezzi elettrici che viene stabilito in un mercato liberalizzato dall'incontro tra domanda e offerta, risente principalmente di questi fattori:

- fluttuazione del livello dei prezzi dei combustibili impiegati nel mix di sorgenti produttive che ogni singolo paese sceglie (variabile di paese in paese);
- livello di efficienza della filiera elettrica e del parco elettrico;
- rapporto tra domanda ed offerta;
- grado di competizione in ciascuna fase della filiera;

L'Italia ha visto la nascita del mercato *day-ahead* nel 2004 (*Italian Power Exchange*) e del mercato a termine nel 2008, anche se il decreto legge Bersani per la liberalizzazione era risalente al 1999, attestandosi tra gli ultimi paesi Europei in questo senso alla creazione di un mercato unico, si veda G.M.E. (2009) e Weron (2006). Molti altri paesi, soprattutto quelli Scandinavi, hanno anticipato la creazione di questi mercati anche di quasi due decenni. È stato il caso anche dell'Inghilterra il cui mercato iniziò le prime operazioni di *trading* già nel 1990 (*England & Wales*

Electricity Pool) attestandosi come primo mercato elettrico. Data la collocazione storica di questi eventi si può ipotizzare che i mercati più vecchi temporalmente siano anche quelli più maturi e per questo "facili" da analizzare e su cui operare per una loro stabilità di comportamento.

Per una più approfondita trattazione sulla struttura del mercato elettrico Italiano e non, delle caratteristiche strutturali della filiera elettrica, degli enti preposti dal processo di liberalizzazione e dei limiti di tali mercati si rimanda a Weron (2006), G.M.E (2009) e Calento & altri (2006).

1.2 Fatti Stilizzati dei Prezzi Elettrici

Si presentano ora i fatti stilizzanti che maggiormente si riscontrano nelle serie dei prezzi nei mercati elettrici *day-ahead*, comuni anche per molti aspetti nella domanda e nel margine. Per domanda si intende una domanda *day-ahead*, ovvero la domanda di energia elettrica che ad oggi si prevede si realizzerà domani in una data fascia oraria o semi-oraria; per margine, intendiamo un margine *day-ahead* ovvero una previsione che ad oggi si fa sulla capacità di soddisfare un certo surplus di domanda che è in grado di gestire il sistema elettrico, per una certa fascia oraria del giorno seguente.

Fissiamo definitivamente il concetto di mercato del giorno prima attraverso la definizione derivabile da Calento & altri (2006) e G.M.E. (2009) per il mercato Italiano (valida anche per il mercato Tedesco, Spagnolo), ma non distanti come concetti da quelle che possono definire altri mercati *day-ahead*:

Il Mercato day-ahead o mercato del Giorno Prima (MGP) dell'energia ospita la maggior parte delle transazioni di compravendita di energia elettrica e si chiude alle 9.00 del giorno prima rispetto a quello cui si riferiscono le offerte (queste si possono presentare anche nelle giornate precedenti alla sessione del mercato cui si riferiscono, fino a un massimo di 9 giorni prima). Il MGP ha come oggetto la contrattazione di energia tramite offerta di vendita e di acquisto e si svolge in un'unica sessione in asta implicita relativa al giorno successivo. Il mercato è organizzato secondo un modello di asta implicita, che

ospita la maggior parte delle transazioni di compravendita di energia elettrica. Il MGP è, infatti, un mercato per lo scambio di energia all'ingrosso dove si negoziano blocchi orari di energia elettrica per il giorno successivo e nel quale si definiscono, non solo i prezzi e le quantità scambiate, ma anche programmi di immissione e prelievo per il giorno dopo.

Nella definizione si identifica la natura delle transizioni in un meccanismo di un'unica asta giornaliera, dove le operazioni di oggi sono effettuate per soddisfare l'esigenza di domani e la presenza di un rapporto di compra-vendita (domanda e offerta che si incontrano) che danno così vita ad un prezzo di equilibrio e stabiliscono le quantità di energia elettrica da immettere e prelevare dalla rete. In un'unica asta giornaliera si negoziano blocchi di energia elettrica per ciascuna fascia del giorno successivo (24 ore o 48 semi ore). In alcuni mercati si effettuano contrattazioni in prossimità delle consegne tramite aste condotte per ciascun periodo, è il caso Inglese.

Studiamo quindi il comportamento dei mercati del giorno prima in quanto questi stabiliscono cosa accadrà il giorno successivo, attraverso una domanda e un margine che ad oggi sono previsioni per una definita fascia oraria del giorno successivo e un prezzo *spot*. Tale prezzo rappresenta il valore monetario che oggi è stato stabilito per vendere ed acquistare energia in una determinata fascia oraria di domani.

Le caratteristiche dei prezzi che si riscontrano in ogni mercato sono le seguenti, si veda Weron (2006) e Bunn (2006):

- **dati ad alta frequenza:** osservazioni orarie o semi orarie (24 o 48 osservazioni giornaliere);
- **assenza di stazionarietà;**
- **altissima volatilità** che indica un elevato rischio nel trattare l'energia elettrica come sottostante sia in fase di analisi dei mercati (*risk management*) che per strumenti derivati o per operazioni di *trading*;
- **cluster volatility:** finestre temporali con alta o bassa volatilità duratura;

- **mean reversion**: fenomeno che associa ai prezzi una caratteristica di istantaneità e veloce ritorno al livello medio anche dopo *shock*;
- **picchi di prezzo** che si manifestano come salti improvvisi che tornano poi al livello del periodo in questione (definiti con il termine *spikes*). È interessante capire quando possono avvenire e perché, ma ancor di più prevederli;
- **cambi di livello e break strutturali**, si presentano come cambiamenti più o meno duraturi nel livello della serie;
- **stagionalità** su più scale, la domanda di energia è legata alle attività umane che rispondono ad esigenze lavorative e legate alle stagioni (condizioni climatiche e ambientali). Si distinguono solitamente tre componenti periodiche:
 - **stagionalità annuale**: una componente ciclica annua che ricalca l'andamento delle stagioni e così le serie risentiranno di una ciclicità nei consumi legata al periodo dell'anno (mese in cui ci si trova), si può immaginare che nel periodo invernale il consumo sia differente in media da quello estivo per esempio. Nei paesi nordici per esempio il consumo invernale è superiore a quello estivo;
 - **stagionalità settimanale**: vi è una struttura legata al giorno della settimana in cui ci si trova che presenta un *pattern* periodico, i lunedì per esempio presenteranno condizioni simili nel tempo ma differenti dalle domeniche o dagli altri giorni della settimana. La domanda di energia è solitamente più bassa nei week end;
 - **stagionalità giornaliera**: la componente è legata al momento della giornata in cui ci si trova, la domanda è per esempio sistematicamente più bassa nelle ore notturne
- **effetti di calendario**: Sono intesi come l'effetto dei giorni festivi infrasettimanali, nei quali è solitamente presente un livello inferiore della domanda di elettricità;

- **dipendenza da fattori tecnici** legati ai processi nella filiera produttiva e dal **margin**e ovvero di quanta capacità residua il sistema elettrico è in grado di produrre per far fronte a eccessi di domanda (margini bassi sono associati a livelli di domanda molto elevata e un completo impiego delle risorse produttive disponibili già in atto – domanda elevata ha un effetto di spinta positiva sui prezzi);
- **distribuzione non normale** caratterizzata solitamente da asimmetrie (solitamente) positive e leptocurtosi nella distribuzione empirica (si veda figura 1.8);

Possiamo riscontrare i precedenti fatti stilizzati nelle figure successive che riportano alcuni grafici dei prezzi *spot* ed una loro analisi di differenti mercati e fasce orarie o semi orarie.

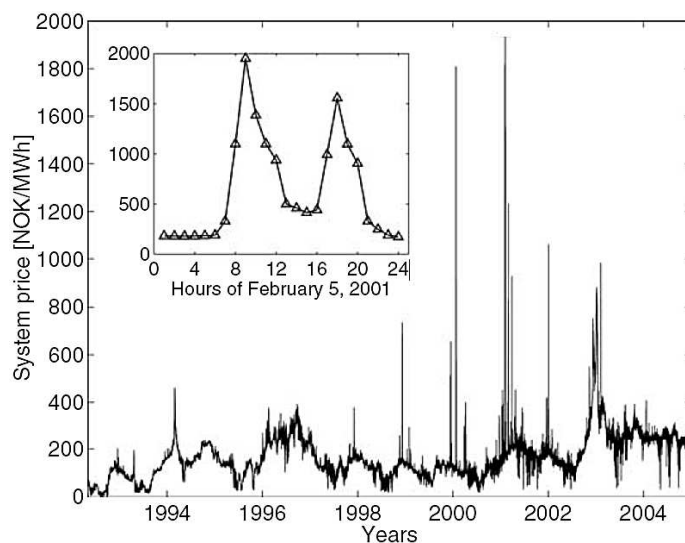


Figura 1.2: Prezzi orari per il mercato *spot* (*Elspot*) nella borsa elettrica *Nord Pool* dal 4 maggio 1992 al 31 dicembre 2004 (oltre 12 anni di dati e 111 000 osservazioni in totale), Weron (2006).

In figura 1.2 possiamo notare come la serie dei prezzi *spot* orari nel mercato *Nord Pool* evidenzia non stazionarietà, né in media né in varianza. I prezzi si presentano infatti con una sorta di andamento tendenziale e una componente di lungo periodo non ben specificabile. Si notano parecchi cambi di livello e *break* strutturali. La presenza di picchi di prezzo è evidente, in poche ore il prezzo può aumentare anche di 10 volte e portare a un picco dei prezzi simile a quello

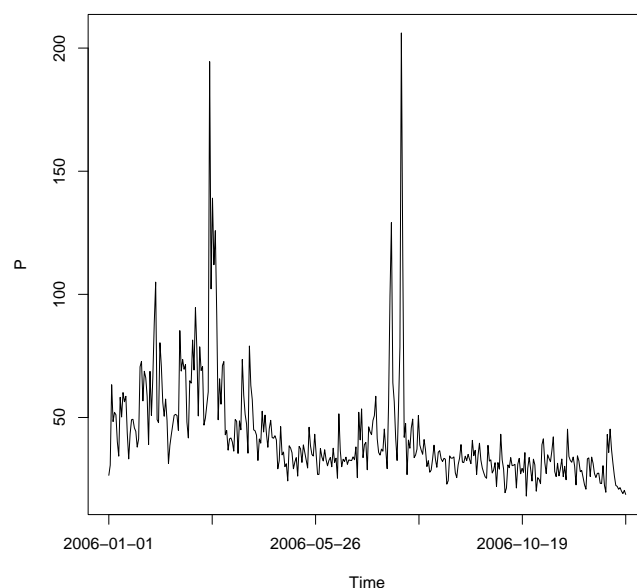


Figura 1.3: Prezzi mercato elettrico Inglese nella fascia semi oraria 9.00 – 9.30 per l’anno 2006.

del 5 febbraio 2001, quando il prezzo ha raggiunto il livello più alto di 1.951,76 NOK / MWh. Nella stessa figura 1.2 si propone uno *zoom* su una singola giornata e si riporta l’andamento dei prezzi orari che evidenziano come un picco possa manifestarsi repentinamente. Sempre in merito alla serie dei prezzi si evidenzia una forte eteroschedasticità che è particolarmente marcata in alcuni periodi e si notano anche dei *cluster* ad alta volatilità e altri a bassa volatilità, si veda figura 1.3.

Nelle figura 1.4 e 1.5 si riportano le funzioni di autocorrelazione e autocorrelazione parziale per 30 e 365 ritardi per la serie dei prezzi nel mercato UK in una data fascia semi oraria, in figura 1.6 si riporta una serie di prezzi medi giornalieri per più anni e due serie di prezzi settimanali. Da queste immagini si osserva una netta persistenza nella serie dei prezzi. Si nota la presenza di una forte componente periodica di periodo 7 (stagionalità settimanale), la presenza di una stagionalità annua (anche se non sempre chiara) e la presenza di un *trend*, indicando anche una forte autocorrelazione nei livelli dei prezzi. Anche la domanda, si veda figura 1.6, esibisce una forte componente di stagionalità settimanale e annua-

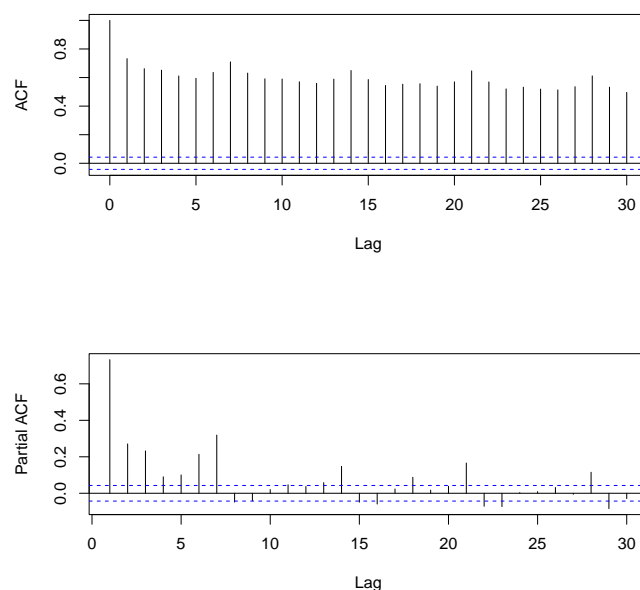


Figura 1.4: In alto la funzione di autocorrelazione (ACF), in basso la funzione di autocorrelazione parziale (partial ACF) per i primi 30 ritardi della serie dei prezzi nel mercato elettrico Inglese nella fascia semi oraria 9.00 – 9.30 per l’anno 2006.

le, rivelandosi sensibile e dipendente dalle condizioni atmosferiche e climatiche, esempio la temperatura.

Osservando i rendimenti medi giornalieri in figura 1.7 derivanti dal mercato elettrico Tedesco, per il triennio 2002–2004, si nota un’estrema volatilità, si può avere un differenziale di rendimento tra giorni consecutivi anche molto elevato.

I mercati elettrici esibiscono una volatilità nettamente superiore a quella che mostrano altri mercati. Nel mercato elettrico la volatilità giornaliera può arrivare a superare il 50%; nei mercati obbligazionari per esempio è inferiore allo 0.5%, in quelli azionari circa 1–1.5%, in mercati di *commodities* come petrolio o gas naturale rimane tra l’1.5–4% e in mercati azionari considerati ad altissima volatilità non supera mai il 4% (Weron, 2006). L’aspetto legato ad un’alta variabilità si può inoltre osservare in figura 1.9 dove si mostrano i vari diagrammi a scatola per i prezzi di ogni fascia semi oraria nel mercato Inglese. Questi mostrano che la volatilità è strettamente connessa anche alla fascia oraria (componente di stagionalità giornaliera); la fascia notturna e serale è più stabile e meno variabile che quella legata ad un orario che approssimativamente può essere dalle 9.00 alle

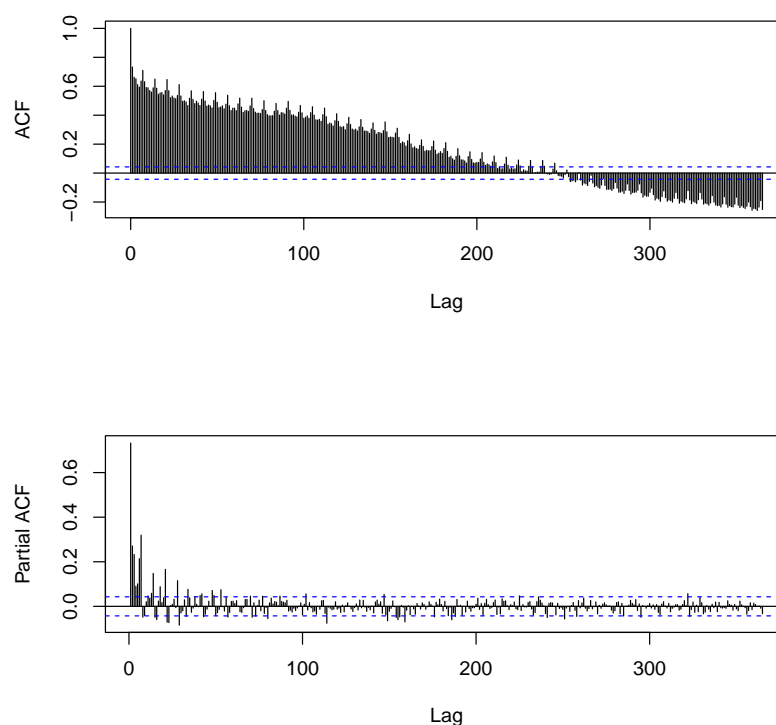


Figura 1.5: In alto la funzione di autocorrelazione (ACF), in basso la funzione di autocorrelazione parziale (partial ACF) per i primi 365 ritardi della serie dei prezzi nel mercato elettrico Inglese nella fascia semi oraria 9.00 – 9.30 per l'anno 2006.

19.00. Questi diagrammi a scatola presentano quattro caratteristiche: assenza di normalità, forte componente giornaliera, alcune fasce semi orarie caratterizzate da una maggiore volatilità e livello medio di prezzo, presenza di picchi di prezzo visibili come una sorta di *outliers* indicati da punti fuori dal baffo superiore della scatola.

Sia riferendosi alla figura 1.9, sia nel caso specifico di una singola fascia semi oraria (quella dalle 9.00–9.30) figura 1.8, si nota una chiara assenza di normalità nella serie dei prezzi, caratterizzate solitamente da un'asimmetria positiva e da leptocurtosi nella distribuzione marginale.

In ogni mercato comunque si evidenziano effetti più o meno marcati dei fatti stilizzati enunciati, ma tra loro differenti anche a seconda del paese al quale il mercato fa riferimento. In alcuni casi, alcuni mercati presentano caratteristiche strettamente proprie e non riscontrabili in altri mercati, è il caso del mercato te-

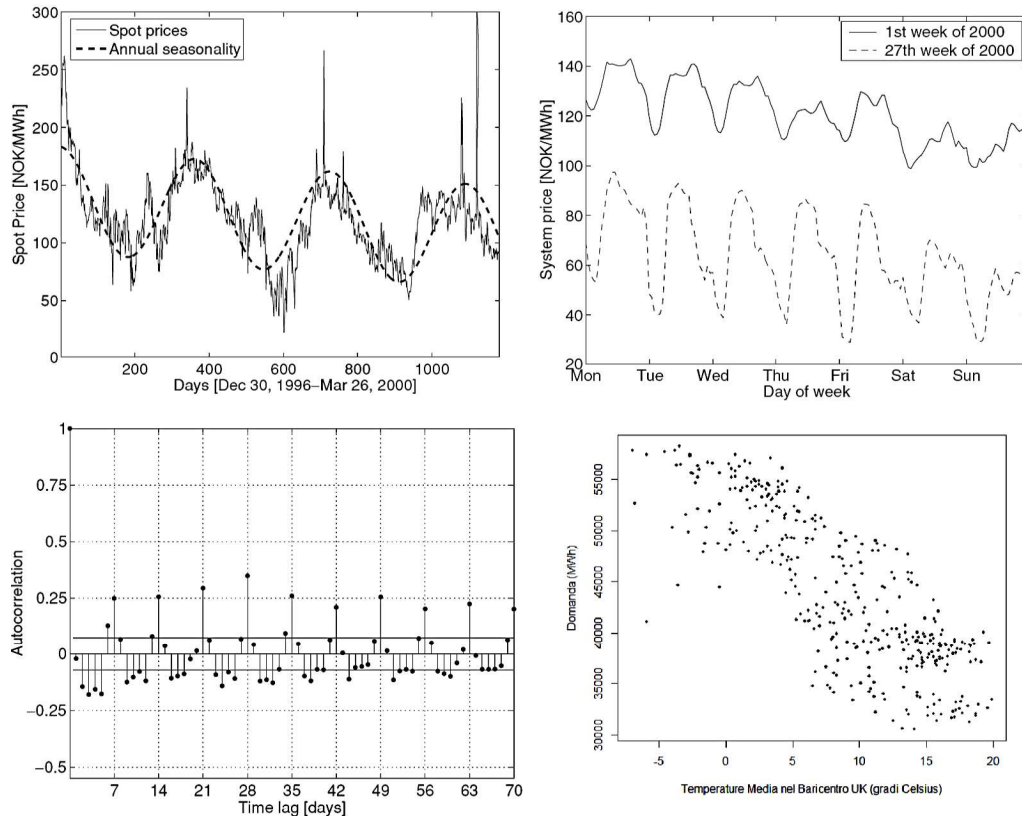


Figura 1.6: Il peso della componente stagionale e dipendenza dai fattori atmosferici su prezzi e domanda in vari mercati elettrici. In alto a sinistra serie dei prezzi medi giornalieri con marcata presenza di stagionalità annuale nel mercato *North Pool* dal 30/12/1996 al 26/03/2000. Viene sovrapposta un'approssimazione della stagionalità annua con una componente trigonometrica e un trend lineare. Fonte: Weron (2006). In alto a destra serie dei prezzi evidenzia la periodicità settimanale nel mercato *North Pool* in due settimane scelte arbitrariamente. La prima settimana (3–9 Gennaio) e la ventisettesima settimana (3–9 Luglio) del 2000, che rappresentano una tipica settimana invernale e una tipica estiva. Fonte: Weron (2006). In basso a sinistra funzione di autocorrelazione della domanda media giornaliera nel mercato CalPX (*California Power Exchange*) da Maggio 1998 ad Aprile 2000. Si evidenzia una forte stagionalità settimanale. Fonte: Weron (2006). In basso a destra l'andamento della domanda al variare della temperatura media rilevata nel mercato UK nella fascia semi oraria 18–18.30 nell'anno 2010.

desco (EPEX) che esibisce in alcuni momenti prezzi negativi, dovuti ad eccessi di produzione non modulabili nel breve periodo (riconducibili alle forti di produzione elettrica nucleare, fotovoltaica e eolica che sono fortemente impiegate) che devono essere smaltiti per evitare *black out*. Per una trattazione approfondita di altri mercati si veda Weron (2006) e Bunn (2006).

Soffermiamoci invece ora su una caratteristica assai importante che caratterizza profondamente i prezzi elettrici e questi mercati: **i picchi di prezzo**. I picchi

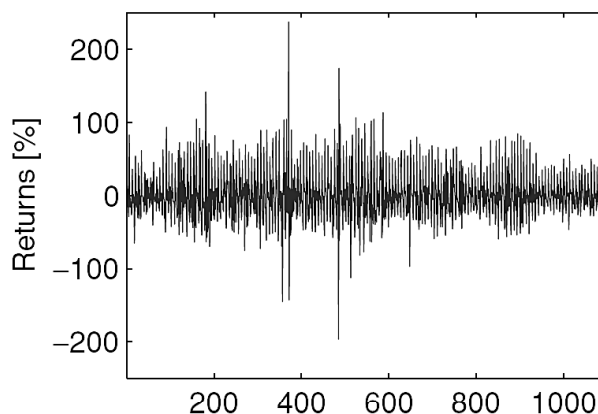


Figura 1.7: Rendimenti percentuali derivanti dai prezzi medi giornalieri nel mercato Tedesco dal 01/01/2002 al 31/12/2004. Fonte: Weron(2006).

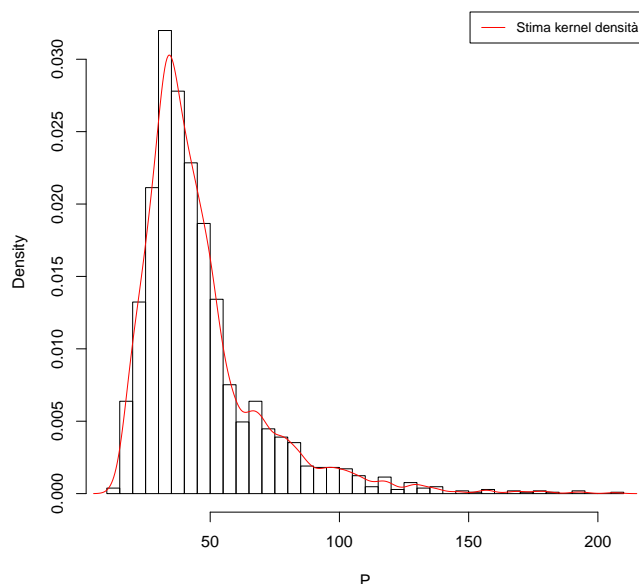


Figura 1.8: Istogramma e stima *kernel* della distribuzione dei prezzi elettrici nel mercato elettrico Inglese nella fascia semi oraria 9.00 – 9.30 per il periodo 01/04/2005 al 31/12/2009.

o salti (in letteratura: *jumps* o *spikes*) sono chiaramente visibili in ogni serie riportata, si veda figura 1.2 e 1.3. Anche nella figura 1.9 dove vengono riportati i *boxplot* dei prezzi per le varie fasce semi orarie, si possono osservare molti picchi di prezzo che si possono individuare come tante osservazioni estreme. I picchi di prezzo solitamente non sono considerabili come osservazioni ed eventi anomali, bensì sono caratteristica intrinseca dei mercati elettrici e hanno cause più o me-

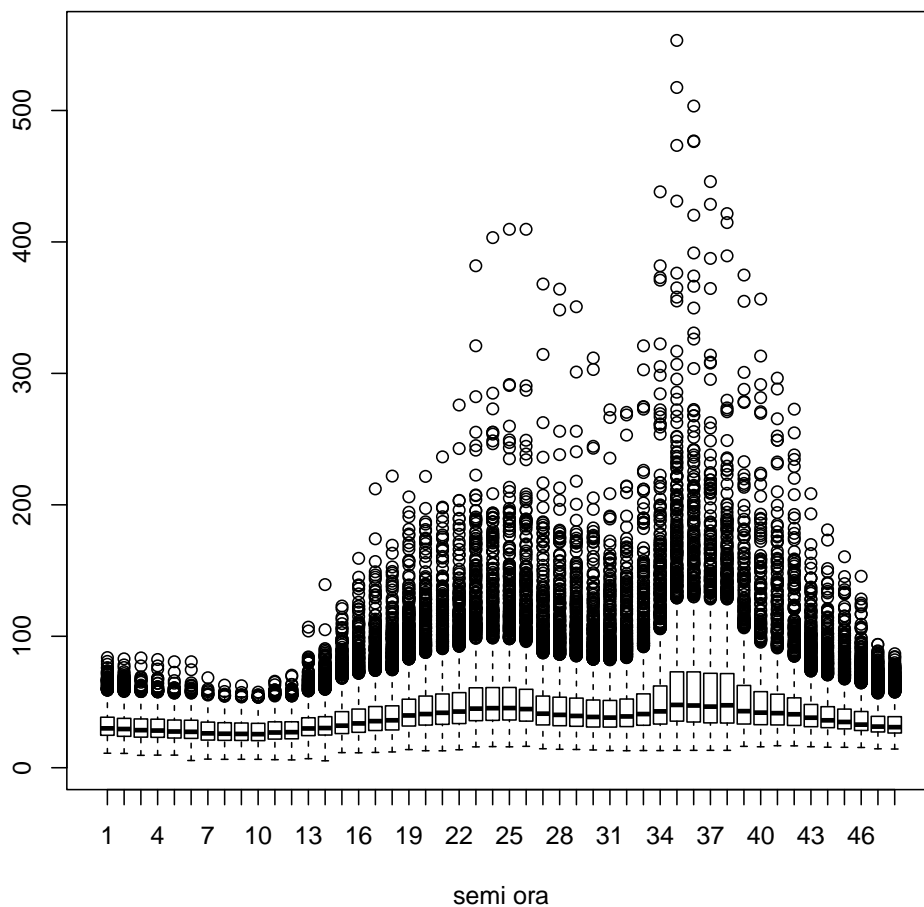


Figura 1.9: *Boxplot* delle serie dei prezzi *spot* delle 48 semi ore nel mercato Inglese da Aprile 2005 a Dicembre 2010.

no precise. Con riferimento alla figura 1.9 si può notare che vi sono alcune fasce orarie che presentano maggiore variabilità e si può immaginare una presenza di picchi più evidente in frequenza e intensità che in altri periodi.

Questi salti, che appaiono inizialmente imprevedibili e inanticipabili, possono essere anche molto consistenti e non sono omogenei, portando i prezzi fortemente al di sopra del livello medio del periodo. Dopo un picco solitamente il mercato riallinea i prezzi ad un livello medio in brevissimo tempo. Un caso particolare è quello del mercato Polacco (PolPX e POEE) che da quando ha iniziato la sua attività nel luglio 2000 non ha presentato mai un vero e proprio picco (Weron, 2006).

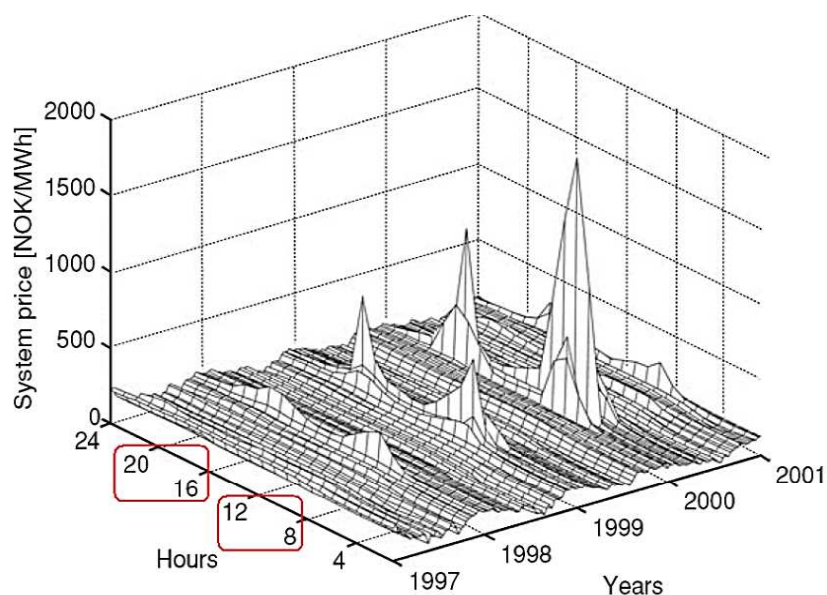


Figura 1.10: Distribuzione dei prezzi *spot* nel mercato *NorthPool* (Norvegia, Svezia, Finlandia e Danimarca) dal 1997 al 2001 (Weron 2006).

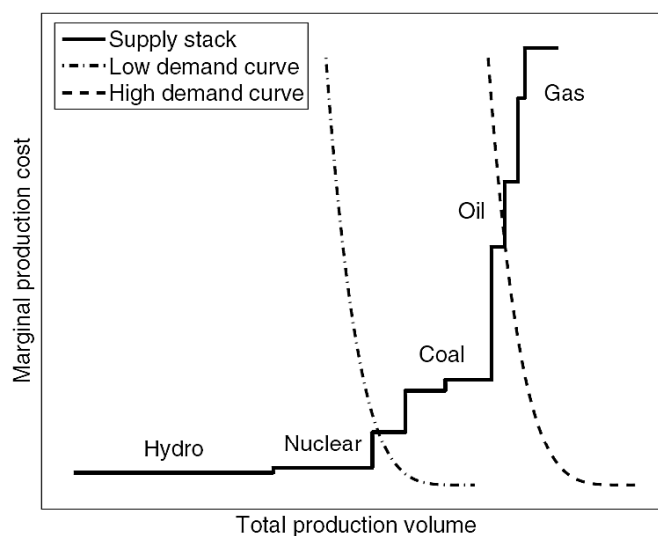


Figura 1.11: Costi marginali di produzione in funzione del volume di energia elettrica prodotta. Fonte: Weron(2006)

In figura 1.10 si mostrano i prezzi giornalieri per quattro anni del mercato *North Pool*. Si evidenzia come i picchi sono tipicamente localizzati nei periodi invernali e durante la fascia mattutina (dalle 9.00 alle 12.00) e quella serale (dalle 17.00 alle 20.00) di giorni lavorativi; questi sono considerabili come i periodi di punta dei consumi elettrici (Weron, 2006).

Le cause dei picchi, che sono anche le principali cause di molte caratteristiche

tipiche dei prezzi elettrici quali l'altissima volatilità e la *mean reversion* (Bunn, 2006), sono essenzialmente le seguenti:

- l'elettricità essenzialmente non è un bene **immagazzinabile**, il processo produttivo e quello di consumo devono essere sincronizzati e bilanciati. L'elettricità è un bene di natura istantanea;
- la presenza di resistenze e vincoli nella trasmissione della rete, interruzioni nella generazione, blocco di impianti o fallimento nella trasmissione;
- la **fluttuazione della domanda** (che a sua volta ha varie cause: clima, situazioni atmosferiche, giorni lavorativi, ...) si riversano sui prezzi causando prezzi elevati: aumenti forti della domanda vengono bilanciati usando impianti di produzione più flessibili ma più costosi, si veda a questo riguardo figura 1.11 che evidenzia come livelli sempre maggiori di domanda coinvolgano fonti produttive via via sempre più costose;
- il **margin** che definisce l'eccesso di capacità elettrica disponibile nel sistema ha effetto sui prezzi, con una domanda crescente il margine diminuisce e il prezzo sale.
- la **strategia di offerta** dei *traders* e dei vari soggetti che partecipano all'asta: essendo l'energia elettrica un bene irrinunciabile per molti partecipanti, si è disposti all'acquisto anche a prezzi molti elevati per mantenere una fornitura continua e al livello necessario;

Il motivo per cui ci si è soffermati con particolare evidenza sui picchi è perché questi sono tutt'altro che imprevedibili e una loro previsione può essere fondamentale come potrebbe essere l'attività di monitorare il rischio in un mercato. Uno dei punti cruciali rimane anche la definizione dei picchi di prezzo e la loro individuazione nelle serie osservate, che in letteratura è ancora argomento di dibattito e non ha ancora trovato un'univoca definizione.

Capitolo 2

IL MERCATO INGLESE UKPX

I dati trattati nel seguente elaborato fanno riferimento alla borsa elettrica Inglese, *England & Wales Electricity Pool*, e rispettando le contrattazioni di questo mercato, sono osservazioni giornaliere su fasce semi orarie. I dati sono stati campionati da Venerdì 1 Aprile 2005 a Venerdì 31 Dicembre 2010 per un totale di 2101 giorni di osservazione. La prima fascia semi oraria è quella che va dalle 00.00 alle 00.30, la seconda fascia semi oraria (che chiameremo anche periodo) va dalle 00.30 alle 01.00 e così via.

Le serie a disposizione sono quelle dei prezzi *spot*, della domanda *day-ahead* e del margine *day-ahead*, che indicheremo rispettivamente con P_t (*pounds*), D_t (*Megawatt*) e M_t (*Megawatt*) per ogni singola fascia semi oraria. Con il termine *day-ahead* ci riferiamo a valori determinati per il giorno successivo. Domanda e margine sono quindi previsioni per una data semi ora del giorno seguente.

Si dispone di una serie di *dummies* che fungono da indicatore dei giorni festivi infrasettimanali (1–il giorno considerato infrasettimanale è festivo, 0–il giorno considerato è un giorno tradizionale) e di una matrice di *dummies* che indicano se l’osservazione considerata fa riferimento ad un lunedì oppure a un martedì e così via; la matrice ha 2101 righe (giorni osservati) e 7 colonne (i 7 giorni settimanali), per ciascuna riga si osserva un solo 1 su una determinata colonna, che rappresenta il giorno di quell’osservazione.

Non verranno presentati graficamente tutti i dati a disposizione e oggetto di studio per le 48 semi ore per una semplice questione di spazio , ma verranno modellate tutte le 48 fasce orarie per stabilire se è possibile definire un modello

per la previsione dei picchi di prezzo.

Lo scopo di questa tesi sarà identificare e prevedere i picchi dei prezzi in questo mercato.

I fatti stilizzati e valutazioni di varia natura che vengono riscontrati negli esempi riportati per alcune fasce semi orarie, sono rappresentativi anche di quelli delle altre fasce semi orarie anche se non riportate. Alcune immagini salienti d'ora in avanti verranno riportate nell'Appendice A.

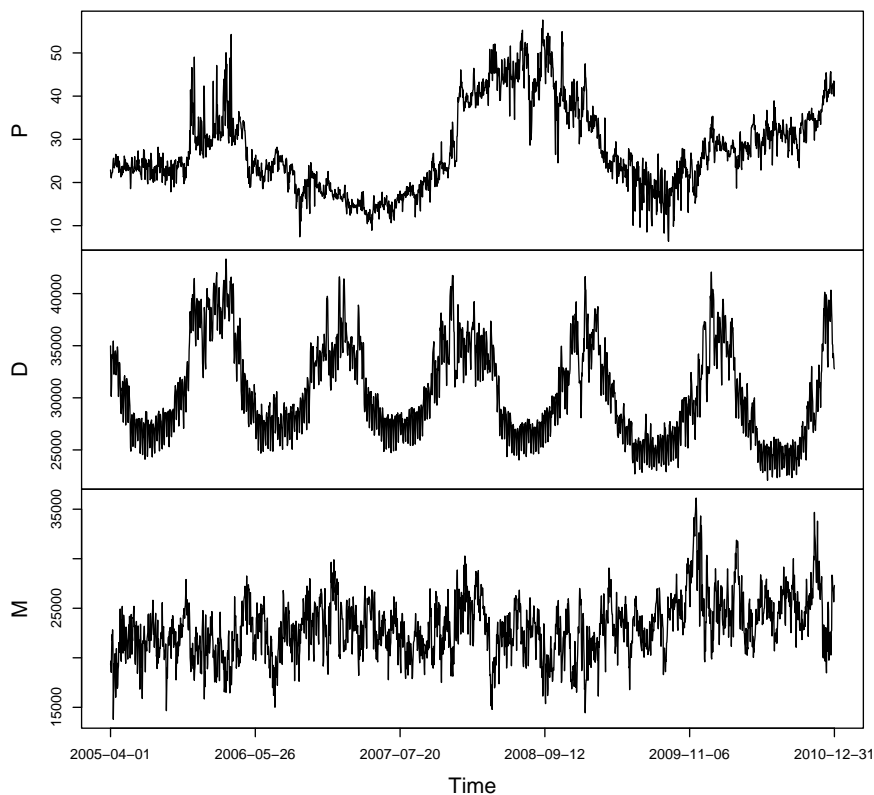


Figura 2.1: Prezzi, domanda e margine nel mercato elettrico Inglese nella fascia semi oraria 4.30–5.00 per il periodo 01/04/2005 al 31/12/2010.

Nella figure 2.1 e 2.2 vengono rappresentate serie dello stesso periodo del prezzo, della domanda e del margine in riferimento ad alcune fasce semi orarie che hanno caratteristiche diverse essendo riferite a momenti della giornata estremamente differenti. Si possono quindi notare nelle serie riportate nelle due figure andamenti assai differenti. In una fascia serale/notturna, figura 2.1, il comportamento del mercato è molto più stabile per quanto concerne l'andamento e il

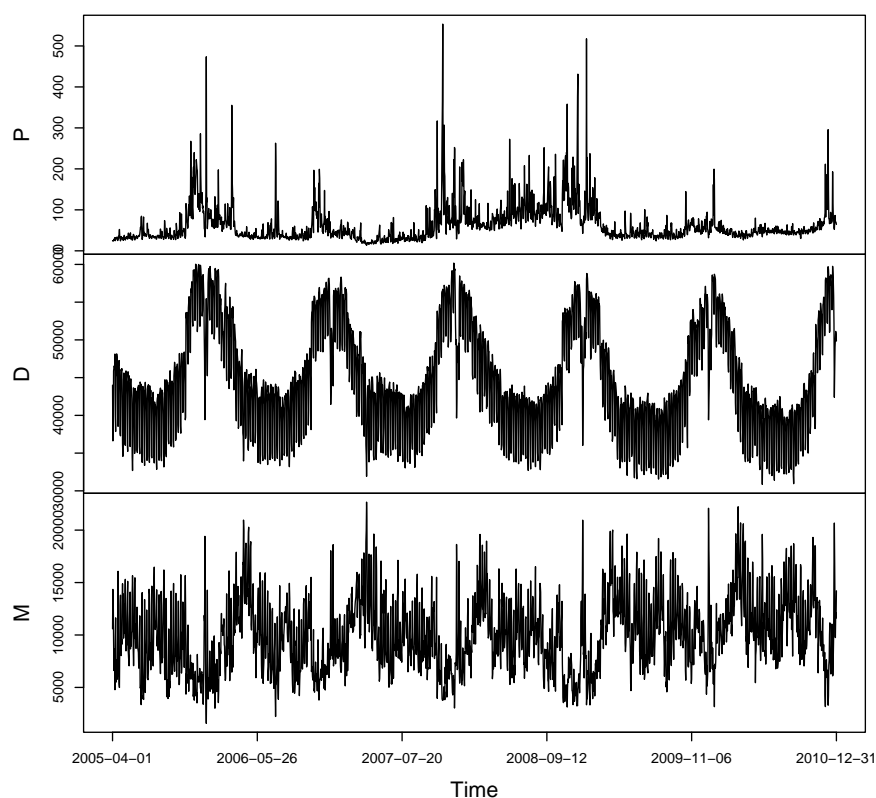


Figura 2.2: Prezzi, domanda e margine nel mercato elettrico Inglese nella fascia semi oraria 17.00–17.30 per il periodo 01/04/2005 al 31/12/2010.

comportamento dei prezzi. La stabilità del mercato in questa fascia semi oraria si osserva dal comportamento della domanda e del margine che sostanzialmente si riflette sul livello dei prezzi che hanno variabilità bassa e si muovono in un *range* limitato. In fasce che possono essere legate alle ore di maggior consumo, figura 2.2, comprese tra le 6 e le 21 si nota una grossa attività e cambiamenti forti nei livelli dei prezzi con picchi assai marcati, componenti di eteroschedasticità, cambi di livello e componenti quali *trend* e stagionalità annuale. I picchi in queste ore e i livelli di prezzi raggiunti sono anche molto alti, si può notare che nella fascia delle 4.30 del mattino il massimo livello di prezzo raggiunto nei cinque anni è intorno a 55 *pounds* mentre nella fascia oraria delle 17 arriva a superare anche i 500 *pounds*.

Queste osservazioni possono essere derivate anche osservando la figura 1.9 che rappresenta la distribuzione empirica dei prezzi in ciascuna fascia semi oraria

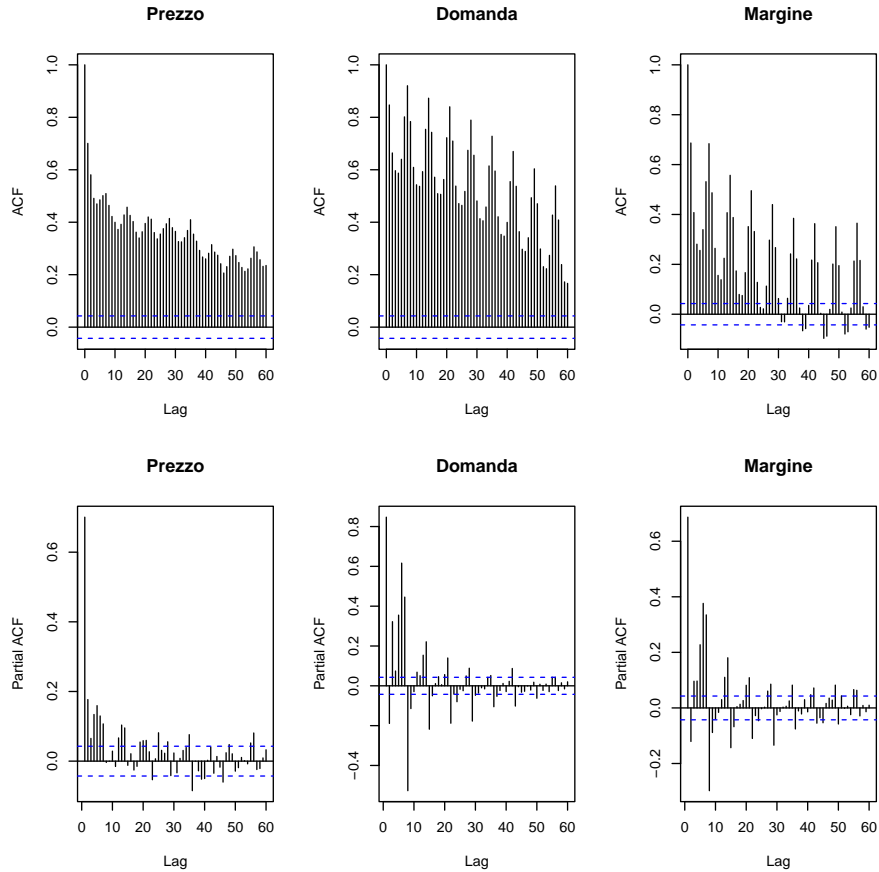


Figura 2.3: Funzione di autocorrelazione (ACF - in alto) e autocorrelazione parziale (*partial ACF* - in basso) delle serie del prezzo, domanda e margine nel mercato elettrico Inglese nella fascia semi oraria 17.00–17.30 per il periodo 01/04/2005 al 31/12/2010.

nei 5 anni in analisi. Nelle figure 2.1 e 2.2 si riscontrano i fatti stilizzati sui prezzi elettrici introdotti in sezione 1.2 e si riscontrano anche i principali elementi sulla stagionalità anche nella domanda e nel margine. Domanda e margine mostrano componenti di stagionalità settimanale e annuale visibili anche nella rappresentazione grafica della funzione di autocorrelazione e autocorrelazione parziale di figura 2.3. La domanda evidenzia una ciclicità annuale e una periodicità di periodo sette, che per le fasce semi orarie si riferisce ad una stagionalità settimanale. Sempre in figura 2.3 si può osservare come la serie dei prezzi evidenzia una forte persistenza e mancanza di stazionarietà.

Separando le serie per singola fascia semi oraria sia per prezzi, domanda e margine, si perde l'effetto della componente stagionale giornaliera. Questa scompare e non è più osservabile nelle valutazioni grafiche svolte sin ora, a tal proposi-

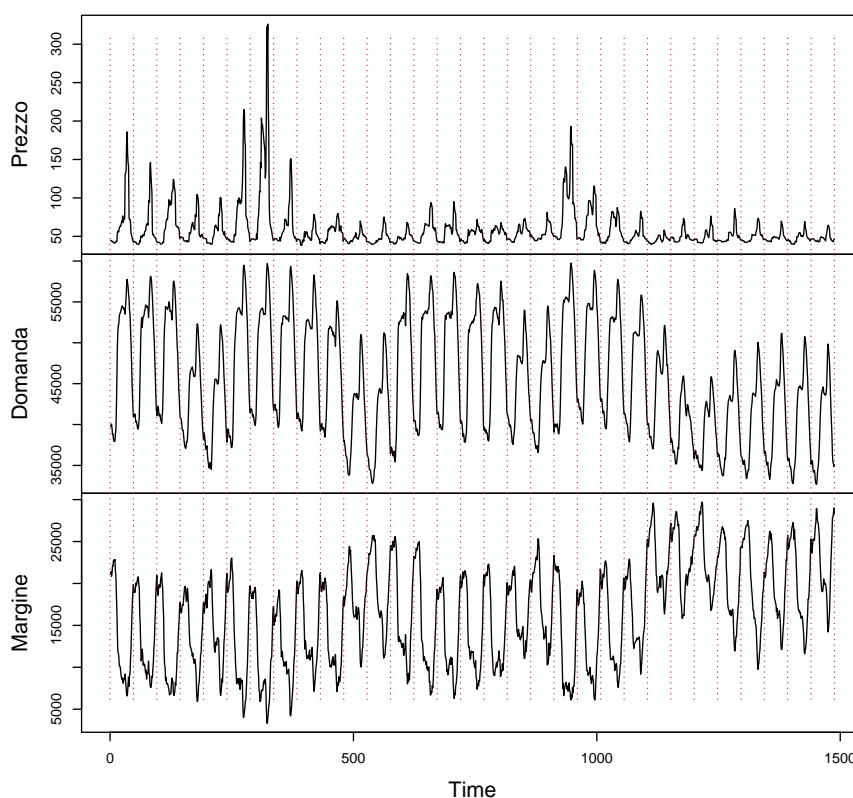


Figura 2.4: Prezzi, domanda e margine nel mercato elettrico Inglese nel dicembre 2010.

to per mostrare la sua presenza nel mercato ricreiamo le serie di prezzi, domanda e margine continue nella giornata, ovvero non semplicemente di una fascia semi oraria ma continui nel tempo. La componente giornaliera è ora visibile, in figura 2.4 si può osservare per il mese di dicembre 2010 come vi sia una stagionalità giornaliera che si ripete ogni 48 osservazioni (la giornata). Le linee rosse tratteggiate verticali di figura 2.4 servono a separare di volta in volta 48 osservazioni, cioè una giornata del mercato Inglese, e si nota un andamento ripetitivo in queste finestre. Per la serie dei prezzi si può notare come la stagionalità giornaliera comporti prezzi minori nella fasi notturne e serali e prezzi mediamente più alti nel corso della giornata con i momenti di maggior picco intorno le 18.00. Questo andamento dei prezzi risponde al comportamento della domanda e del margine che, come si può notare dalla figura 2.4, mostrano una componente di stagionalità giornaliera marcata. Queste evidenze supportano l'idea teorica: bas-

so margine/prezzi elevati e elevata domanda/prezzi elevati. Lavorare su una singola fascia semi oraria elimina il problema di dover considerare la stagionalità giornaliera.

Dalle analisi fin'ora svolte sono inoltre visibili alcune caratteristiche e componenti prima enunciate tra le quali una presenza di forte eteroschedasticità soprattutto nella serie dei prezzi.

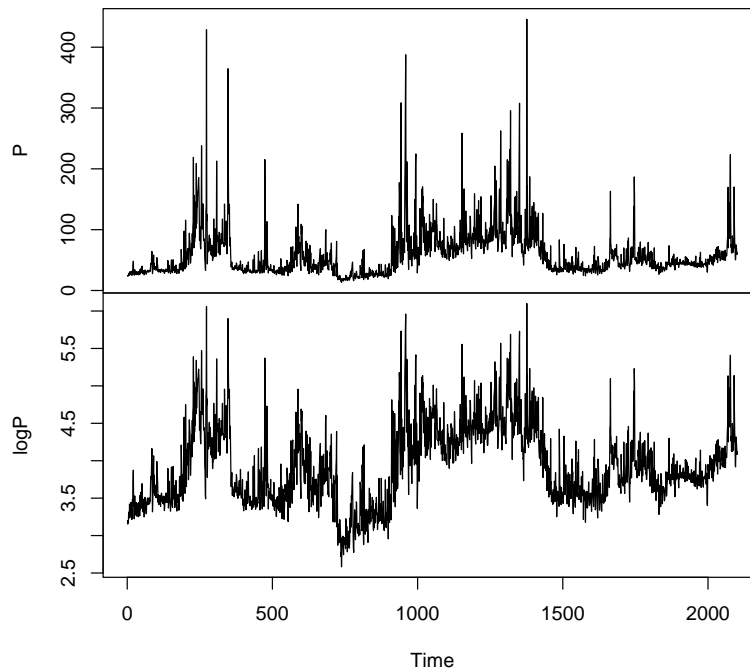


Figura 2.5: In alto i prezzi elettrici, in basso i rispettivi logaritmi nel mercato elettrico Inglese nella fascia semi oraria 18.00–18.30 per il periodo 01/04/2005 al 31/12/2010.

In figura 2.5 viene riportata una trasformazione della serie dei prezzi iniziale per la funzione logaritmo. In letteratura viene spesso considerata una trasformazione logaritmica della serie dei prezzi che tra le varie proprietà ha quella di ridurre l'impatto della presenza dei picchi. Ne può conseguire che considerare prezzi o logaritmi dei prezzi potrebbe cambiare la loro individuazione nella serie.

Se si decide di operare sui logaritmi dei prezzi per l'individuazione dei picchi e per la realizzazione dei modelli sarà necessario operare anche sui logaritmi di domanda e margine. I logaritmi di domanda e margine non presentano grandi

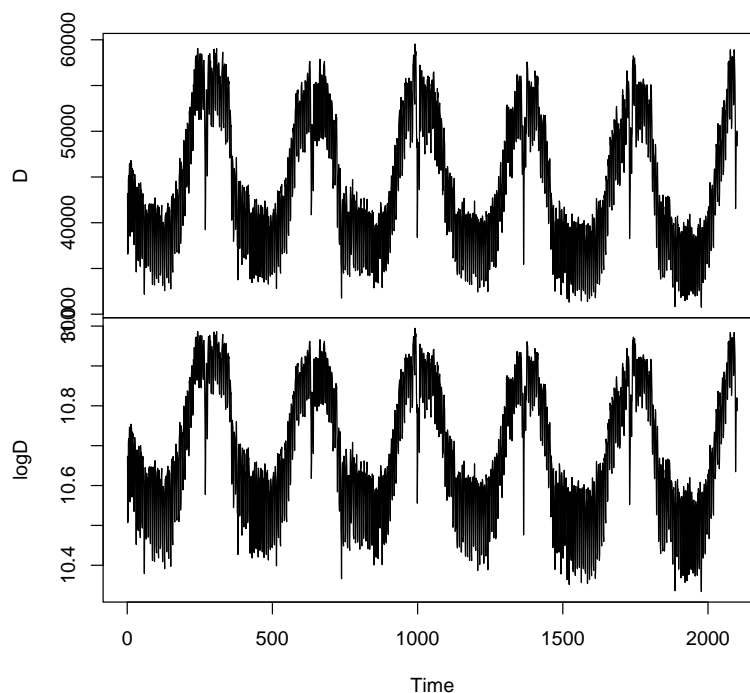


Figura 2.6: Logaritmi della domanda e domanda nel mercato elettrico Inglese nella fascia semi oraria 18.00–18.30 per il periodo 01/04/2005 al 31/12/2009.

differenze, se non nella scala, si vedano le figure 2.6 e 2.7 che riportano domanda e margine e le relative trasformazioni logaritmiche. Viene evidenziato inoltre in figura A.1 (Appendice A) come vi sia una relazione tra prezzi domanda e margine (teoricamente auspicabile) specificate già in precedenza per le serie in livello. Si nota come invece una trasformazione logaritmica delle serie, si veda figura A.2 (Appendice A), linearizzi notevolmente le relazioni tra prezzi e domanda e tra prezzi e margine sensibilmente.

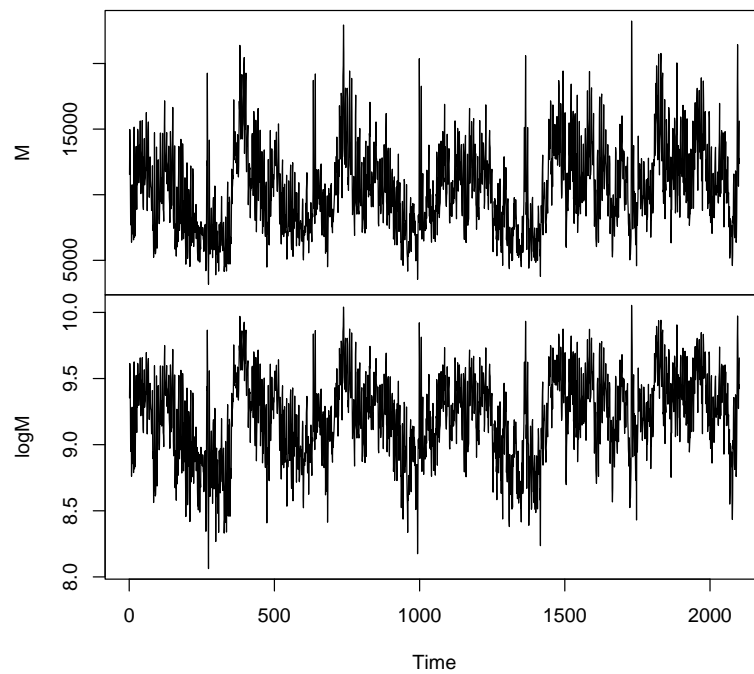


Figura 2.7: Logaritmi del margine e margine nel mercato elettrico Inglese nella fascia semi oraria 18.00–18.30 per il periodo 01/04/2005 al 31/12/2010.

2.1 Stazionarietà

Rimarcando che lo scopo della tesi rimane quello di individuare e prevedere i picchi di prezzo attraverso un opportuno modello, emerge una caratteristica fondamentale delle serie oggetto di studio relativa alla stazionarietà. Sia per l'individuazione dei picchi che per la realizzazione di modelli di previsione, diventa necessario stabilire se le serie sono stazionarie o meno, e se non stazionarie di che tipo di non stazionarietà si tratta. Le osservazioni grafiche svolte sin ora portano alla considerazione che le serie difficilmente possano essere ritenute stazionarie data per la presenza di componenti tendenziali e stagionali; dato ciò bisogna verificare con opportuni *test* di che tipo di non stazionarietà si possa trattare.

Si ha quindi la necessità di testare, in caso di non stazionarietà, se le serie derivino da un processo generatore *trend* stazionario o differenze stazionario. Limitandoci ad una serie che esibisce un *trend* lineare, una serie è *trend* stazionaria se il suo processo può essere $X_t = \gamma_0 + \gamma_1 t + \eta_t$ dove η_t è un *white noise*; una serie deriva da un processo differenze stazionario (limitiamoci ad un processo integrato di ordine uno) se il processo generatore può essere rappresentato come $X_t = \alpha_0 + X_{t-1} + \eta_t$ (*Random walk con drift*). Questa differenza è necessaria per adottare la corretta procedura di detrendizzazione che ha un impatto sulla validità dei risultati inferenziali e per capire se nel lungo periodo vi siano *trend* o componenti deterministiche o *trend* e componenti stocastiche.

Si procede quindi testando la stazionarietà (ed eventualmente di che tipo di non stazionarietà si tratti) per i processi generatori dei prezzi, della domanda e del margine attraverso *test* per radici unitarie e *test* di stazionarietà sulle serie a disposizione.

Il test per radice unitaria di *Dickey-Fuller* (aumentato) o alternativamente il test di *Philips-Perron* permettono di valutare se il processo generatore dei dati sia integrato contro l'ipotesi che questo sia stazionario (magari attorno a un *trend*),

il test fa riferimento in generale al sistema di ipotesi (2.1).

$$\begin{cases} H_0 : \text{ Presenza di Radici Unitarie} \\ H_1 : \text{ Processo Trend-Stazionario} \end{cases} \quad (2.1)$$

Si dà ora una breve introduzione al *test Dikey-Fuller* e la sua versione *Augmented*, per una trattazione più approfondita e per i dettagli tecnici si veda Bhargava (1986), Cappuccio e Orsi (2005) e Alexander (2008).

Per testare la presenza di radici unitarie si è utilizzato il *test Augmented Dikey-Fuller* (ADF) che ipotizza una dinamica dei processi stimabile da modelli AR(p). Il *test* ADF permette di valutare se il processo generatore può essere un processo integrato di ordine uno con *trend* (H_0) contro uno processo autoregressivo di ordine p con *trend* stazionario. Un processo X_t autoregressivo di ordine p è

$$X_t = D_t + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \epsilon_t, \quad (2.2)$$

dove $\epsilon_t \sim WN(0, \sigma^2)$ è un *white noise* e D_t è la componente deterministica necessaria a specificare la presenza di un *trend* lineare: $D_t = \beta_0 + \beta_1 t$ (alternativamente $D_t = 0$ e $D_t = \kappa$). Il processo AR(p) può essere scritto in forma compatta con il polinomio caratteristico

$$(1 - \phi_1 B - \dots - \phi_p B^p) X_t = D_t + \epsilon_t,$$

$$\Phi(B) X_t = D_t + \epsilon_t,$$

dove B è l'operatore ritardo ($B^p X_t = X_{t-p}$) e $\Phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$ è il polinomio caratteristico del AR(p). Vi è (almeno) una radice unitaria quando esiste (almeno) uno z tale che $\Phi(z) = 0$ (allora $\phi_1 + \dots + \phi_p = 1$). In questo modo riparametrizzando il modello AR(p) in modo da avere che $\rho = \phi_1 + \dots + \phi_p$ si può costruire un test per saggiare il sistema d'ipotesi (2.1) di radice unitaria basato sull'ipotesi nulla che $\rho = 1$. Mediante una riscrittura del polinomio caratteristico il modello (2.2) può essere riscritto come

$$X_t = D_t + \rho X_{t-1} + \sum_{j=1}^{p-1} \gamma_j X_{t-j} + \epsilon_t,$$

dove $\rho = \phi_1 + \dots + \phi_p$ e $\gamma_j = -\sum_{i=j+1}^p \phi_i$.

Testare se il processo è un processo integrato che possiede una radice unitaria contro l'ipotesi che sia un AR(p) stazionario (sistema d'ipotesi (2.1)) significa saggiare il sistema d'ipotesi

$$\begin{cases} H_0 : \rho = \phi_1 + \dots + \phi_p = 1 \\ H_1 : \rho = \phi_1 + \dots + \phi_p < 1 \end{cases}$$

dove la condizione posta in H_1 definisce la condizione necessaria affinché le soluzioni di $\Phi(z) = 0$ siano in modulo maggiori di 1, condizione di stazionarietà. La statistica test su ρ è una statistica che testa la significatività, ovvero $T = \frac{\hat{\rho} - \rho_{H_0}}{\sqrt{V_{AR}(\hat{\rho})}}$ con stime derivate con minimi quadrati ordinari, che sotto l'ipotesi nulla ha una distribuzione *non standard*. La distribuzione di T dipende dal tipo di componente deterministica D_t specificata e viene stimata in modo numerico così come i valori critici del test per differenti livelli di significatività (Bhargava, 1986). In tabella 2.1 si riportano i valori critici per differenti livelli di significatività per un processo con *trend* lineare. Per saggiare il sistema d'ipotesi unilaterale sinistro, si rifiuta l'ipotesi nulla quando il valore della statistica test cade alla sinistra del valore critico. La scelta di p , cioè del numero di ritardi da considerare nel modello $AR(p)$ viene solitamente definita mediante criteri di selezione automatica come il criterio di *Akaike* (AIC) o *Schwarz* (BIC). Per determinare la componente deterministica da includere nel modello da testare si ricorre ad un'analisi grafica delle serie, si osserva se vi è la presenza di un trend o meno. La corretta specificazione della componente deterministica è cruciale ai fini della consistenza e potenza del test.

Livello di Significatività %	Valore Critico del Test
1pct	-3.96
5pct	-3.41
10pct	-3.12

Tabella 2.1: Valori critici per la statistica *test Augmented Dikey-Fuller* con specificazione nella componente deterministica di un trend.

L'applicazione del test *Augmented Dikey Fuller* alla serie dei prezzi, della domanda e del margine per ciascuna fascia semi oraria specificando come componente deterministica la presenza di un *trend* lineare, ma anche per ciascuna spe-

cificazione alternativa delle componente deterministica, il responso del test non cambia e porta a rifiutare l'ipotesi nulla (ipotesi sulla componente deterministica che potrebbero essere riduttive per i dati elettrici, ma le uniche implementate in R). Per un livello di significatività dell'1% il test è sempre significativo e rifiuta l'ipotesi nulla del sistema d'ipotesi 2.1, il valore della statistica T è sempre inferiore al valore critico -3.96. In tabella 2.2 si riporta "l'ipotesi scelta" dal test ADF (H1: rifiuto l'ipotesi nulla, H0: non rifiuto l'ipotesi nulla). Questo risultato evidenzia la mancata presenza di radice unitaria nei processi generatori e che quindi le serie non seguono processi differenze stazionarie. Essendo i processi generatori dei dati non integrati, le componenti tendenziali che si osservano graficamente sono di tipo deterministico e quindi stimabili e non rimovibili tramite differenziazione della serie per suoi ritardi.

Test Augmented Dikey Fuller			
Fascia Semi Oraria	Prezzi	Domanda	Margine
1	H1	H1	H1
2	H1	H1	H1
...	H1	H1	H1
...	H1	H1	H1
47	H1	H1	H1
48	H1	H1	H1

Tabella 2.2: Risultati del test Augmented Dikey Fuller per un test sulla presenza di radici unitarie con specificazione "Trend" al livello di significatività dell'1%. Gli "H1" indicano che il test per quella fascia semi oraria per la serie di prezzi, domanda e margine è significativo e rifiuta quindi l'ipotesi nulla.

Per confermare tali idee, seppur sia evidente graficamente che vi sono componenti tendenziali di varia natura che rendono le serie non stazionarie, nonostante il test ADF rifiuti la presenza di radici unitarie (che è la cosa fondamentale per discriminare sul fatto di avere o meno processi con componenti stocastiche o deterministiche) si procede con un test che completa l'analisi svolta sin qui per verificare la stazionarietà delle serie attorno un livello o un *trend*. Il Test di Stazionarietà KPSS (*Kwiatkowski–Phillips–Schmidt–Shin*) verifica se il processo generatore è *trend* stazionario (oppure stazionaria attorno ad un livello) e si basa sul saggiare un sistema d'ipotesi del tipo (2.3). Avere a disposizione un test di

stazionarietà oltre che a test di radice unitaria è piuttosto utile per robustificare i risultati.

$$\begin{cases} H_0 : \text{Trend-Stazionario} \\ H_1 : \text{Presenza di Radici unitarie} \end{cases} \quad (2.3)$$

Per una trattazione approfondita e per i dettagli tecnici del test KPSS si veda Kwiatkowski e altri (1992), Cappuccio e Orsi (2005) e Bhargava (1986), qui si dà solo una breve introduzione. Prendiamo in considerazione il processo

$$X_t = \delta t + \mu_t + \eta_t \quad (2.4)$$

dove μ_t è un processo *random walk*, $\mu_t = \mu_{t-1} + \varepsilon_t$ e η_t è un processo stazionario ma non deterministico con σ_ε^2 e σ_η^2 varianze rispettivamente di ε_t e η_t . Questo processo senza restrizioni su δ , σ_ε^2 e σ_η^2 è una somma di una passeggiata casuale, di una deriva e di un processo stazionario. Se $\delta = 0$ la deriva è nulla, mentre se $\varepsilon_t = 0$ l'equazione del *random walk* diventa una costante e il processo (2.4) si riduce ad un *trend* lineare deterministico sommato ad un processo stazionario (*trend* deterministico). Il test KPSS verifica come ipotesi nulla che $\varepsilon_t = 0$ cioè la stazionarietà del processo, contro l'ipotesi che questo sia differente ovvero un processo non stazionario. La statistica test deriva dalla regressione ausiliaria delle x_t su t (o dalla regressione di x_t su una costante c se non si ritiene presente un *trend*) dalle quali si derivano i residui e_t . Si integrano i residui di regressione ovvero $S_t = \sum_{j=1}^t e_j$, infine si calcola la statistica test di KPSS come

$$KPSS = \frac{1}{n^2 \hat{\lambda}^2} \sum_{t=1}^n S_t^2$$

dove $\hat{\lambda}$ è una stima della varianza asintotica della media campionaria di η_t . La distribuzione asintotica di tale statistica non è standard ed i suoi valori sono tabulati dopo essere stati calcolati in via simulata sia per il caso con deriva sia per il caso senza *trend* deterministico, si veda Kwiatkowski e altri (1992). I valori critici (si veda tabella 2.3) e la forma distributiva della statistica test sono strettamente dipendenti dalla specificazione della componente deterministica. L'applicazione del test KPSS, definendo come componente deterministica sia un livello costante che un trend lineare (ipotesi sulla componente deterministica che

Livello di Significatività %	Specificazione della componente deterministica	
	$D_t = \kappa$	$D_t = \kappa + \delta t$
1pct	0.739	0.216
5pct	0.463	0.146
10pct	0.347	0.119

Tabella 2.3: Valori critici del test *KPSS* con componente deterministica costante oppure un *trend* lineare (Cappuccio e Orsi, 2005).

appaiono riduttive per i dati elettrici, ma le uniche implementate in R) a tutte le fasce semi orarie sui prezzi, sulla domanda e sul margine evidenzia che il test è comunque sempre significativo nel rifiutare l'ipotesi nulla. Per la serie dei prezzi in ogni fascia semi oraria il *p-value* è sempre inferiori a 0.01. Per le serie di domanda e margine il test rifiuta sempre l'ipotesi nulla ma in alcune fasce semi orarie i livelli del *p-value* sono compresi tra 0.01 e 0.06, valori comunque che non evidenziano particolari dubbi sulla stazionarietà delle serie in analisi attorno ad un trend lineare. Una tabella per il responso del test *KPSS* per tutte le fasce semi orarie e tutte le serie è analoga alla tabella 2.2 riferita al test *ADF* ed è riportata in tabella 2.4.

Fascia Semi Oraria	Test <i>KPSS</i>		
	Prezzi	Domanda	Margine
1	H1	H1	H1
2	H1	H1	H1
...	H1	H1	H1
...	H1	H1	H1
47	H1	H1	H1
48	H1	H1	H1

Tabella 2.4: Risultati del test *KPSS* per verificare la stazionarietà delle serie di prezzi, domanda e margine per ogni fascia semi oraria al livello di significatività del 5%. Gli "H1" indicano che il test è significativo e rifiuta quindi l'ipotesi nulla.

Il rifiuto delle relative ipotesi nulle nei test *ADF* e *KPSS* (con ipotesi di una componente deterministica quale un trend lineare) conduce a confermare l'ipotesi che le serie di prezzi, domanda e margine siano derivante da processi stocastici senza radici unitarie ma non stazionarie attorno ad un trend lineare. É auspicabile invece che le serie siano stazionarie attorno ad una componente deterministica variabile nel tempo.

2.2 La Decomposizione delle Serie

L'evidenza a cui hanno portato i test di radici unitarie e test di stazionarietà è quella che i prezzi, la domanda ed il margine siano serie che per ogni fascia semi oraria derivino da processi non integrati e che tali serie siano stazionarie attorno ad un trend e ad alcune componenti di natura deterministica. Ipotizziamo quindi che prezzi, domanda e margine seguano un modello additivo quale

$$P_t = DP_t + p_t \quad (2.5)$$

$$D_t = DD_t + d_t \quad (2.6)$$

$$M_t = DM_t + m_t \quad (2.7)$$

dove DP_t , DD_t e DM_t rappresentano la componente deterministica (ipotizzabili essere di uguale specificazione in tutti e tre i casi) per le rispettive serie e p_t , d_t e m_t rappresentano la componente stocastica.

Per ottenere una stima delle serie stocastiche è necessario ottenere una stima della componente deterministica D_t e sottrarla alla serie di partenza. Si è scelto un modello di decomposizione unico (per prezzo, domanda e margine per ogni fascia semi oraria) per una questione economica e per mantenere un approccio valido in generale. È quindi possibile che un modello di questo tipo non si adatti bene ad alcune fasce semi orarie per caratteristiche molto specifiche che si possono riscontrare in alcuni casi, ma in generale può essere una valida rappresentazione.

Ci si pone ora il problema di definire più specificatamente la componente deterministica e il modo in cui questa possa essere stimata. Presentiamo per brevità solo la fattorizzazione e stima della componente deterministica dei prezzi che sarà identica a quella di domanda e margine. Le possibili componenti per fattorizzare D_t sono:

- una componente tendenziale di lungo periodo (L_t) contenente: l'andamento tendenziale della serie e la componente periodica annuale (non sempre di chiara presenza), idealmente liscia;
- Una componente periodica settimanale (S_t);

- Una componente legata agli effetti di calendario (F_t);

Trattando serie per fasce semi orarie la componente giornaliera (stagionalità intragiornaliera) non è presente nei dati e quindi non va gestita nella componente deterministica.

Assumendo che la fattorizzazione della componente deterministica segua ancora un modello additivo

$$DP_t = LP_t + SP_t + FP_t \quad (2.8)$$

La stima della componente deterministica è $\widehat{DP}_t = \widehat{LP}_t + \widehat{SP}_t + \widehat{FP}_t$.

Ai fini della modellazione è stato già accennato di come la trasformazione logaritmica possa essere utile nella serie dei prezzi (riducendo l'elevata eteroschedasticità, per attenuare la presenza dei picchi di prezzo e linearizzare le relazioni). Il modello che proponiamo per le serie logaritmiche è analogo di quello proposto per le serie sui livelli (2.5), (2.6) e (2.7) ma cambia l'ipotesi di partenza sulle serie in livello. Se per i logaritmi dei prezzi il modello è

$$\log P_t = DP_t + p_t \quad (2.9)$$

si ipotizza implicitamente che il modello generatore delle varie serie dei livelli sia di tipo moltiplicativo. Trattando il logaritmo dei prezzi si tratterà il logaritmo della domanda e del margine.

Vengono ora elencati i modi in cui sono state stimate le componenti della parte deterministica, utili sia sui logaritmi che sulle serie in livello e come sia quindi possibile dare una stima della componente stocastica.

2.2.1 Regressione mediante Spline

Prendiamo a riferimento un generico modello di regressione, dove Y_t è la variabile risposta e x_t la variabile esplicativa, si ha che per dati omoschedastici

$$Y_t = f(x_t) + \varepsilon_t \quad (2.10)$$

con $\varepsilon_t \sim (0, \sigma^2)$ e $E(Y_t|X_t = x_t) = f(x_t)$. Qualora ad $f(x_t)$ non si riesca ad attribuire una forma parametrica, lineare ($f(x_t) = \beta_0 + \beta_1 x_t$) o non lineare ma

parametrico ($f(x_t) = f(x_t; \theta)$), si ricorre ad un modello di regressione non parametrico. Assumendo che $f(x_t)$ sia sufficientemente liscia è possibile stimarla con uno dei metodi non parametrici quali il metodo del nucleo (*kernel regression*), il metodo della regressione polinomiale locale (esempio il *loess*) e i metodi basati su *spline* di regressione o di lisciamento.

Le *spline* sono delle funzioni polinomiali definite a tratti che approssimano il comportamento globale di una funzione di cui sono noti solo alcuni valori. Il metodo delle *spline* stabilisce di fissare k "nodi", ovvero k punti in cui la funzione *spline*, $f(x_t)$, passerà e sarà un polinomio di grado d tra i k nodi. Affinché la funzione *spline* passi con continuità per i k nodi occorre che questo polinomio di grado d ammetta derivate fino al grado $d - 1$ in ciascun nodo. Un *spline* per un generico polinomio di grado d e k nodi $t_1 < t_2 < \dots < t_k$ è

$$f(x) = \beta_0 + \beta_1 x + \dots + \beta_d x^d + b_1(x - t_1)_+^d + b_2(x - t_2)_+^d + \dots + b_k(x - t_k)_+^d$$

dove $(x - t_j)_+^d$ assume valore x^d se $x \geq t_j$ zero altrimenti. Rappresentando tutto in modo compatto si ha che

$$f(x) = \sum_{j=0}^{d+k} \gamma_j S_j(x) \quad (2.11)$$

dove le $S_j(x)$ prendono il nome di *power basis* o basi e

$$\gamma_j = \begin{cases} \beta_j & \text{per } j = 0, 1, \dots, d \\ b_{j-d} & \text{per } j = d + 1, \dots, d + p \end{cases} \quad \text{e} \quad S_j(x) = \begin{cases} x^j & \text{per } j = 0, 1, \dots, d \\ (x - t_{j-d})_+^d & \text{per } j = d + 1, \dots, d + p \end{cases}$$

Al fine di stimare di $f(x_t)$ si sostituisce l'equazione (2.11) nella regressione di equazione (2.10), stimando i γ_j con il metodo dei minimi quadrati si ottiene una stima della funzione di regressione.

Se $d = 3$ si parla di *spline cubiche*, che sono comunemente utilizzate; ponendo opportuni vincoli quali la linearità dei polinomi nei nodi estremi, per l'identificazione in fase di stima, si parla di *spline cubiche naturali*. La scelta di k , d e la posizione dei nodi influenzano la stima di $f(x_t)$ in termini di una stima più o meno liscia, modificando il *trade-off* tra varianza e distorsione nella stima. Se il

grado del polinomio che interpola i nodi è generalmente elevato si privilegia la flessibilità e un approssimazione più precisa a discapito di un effetto lisciamiento (si parla di fenomeni di *over-fitting* accompagnati da un effetto di variabilità della stima elevato). Numero e posizione dei nodi sono definibili con vari e opportuni criteri.

Questo tipo di *spline* sono le *spline* di regressione che non sempre vengono impiegate, è possibile quindi ricorrere alle *spline* di lisciamiento.

Le *smoothing spline* risolvono il problema della selezione dei nodi, definendo $f(x)$ tra tutte le funzioni aventi derivate seconde continue fino al secondo ordine quella che minimizza la somma di quadrati di residui penalizzata $D(f; \lambda) = \sum_{t=1}^n (y_t - f(x_t))^2 + \lambda \int [f''(x)]^2 dx$. Il primo termine di $D(f; \lambda)$ definisce un grado di accostamento ai dati mentre il secondo fattore definisce una misura legata all'irregolarità. Il parametro λ attribuisce un peso più o meno elevato al fattore di irregolarità. In particolare se $\lambda = 0$ non vi è alcuna penalizzazione legata all'irregolarità, per valori di λ crescenti si privilegia una $f(x)$ sempre più liscia, fino al caso estremo in cui $\lambda = \inf$ dove la $f(x)$ diventa una funzione lineare $(\alpha + \beta x)$. Per definire il valore di λ si ricorre solitamente al metodo della *cross validation*, anche se tale metodo è valido solo per dati indipendenti. Per una trattazione approfondita sulle *spline* si veda Wang (2011).

2.2.2 Componente di lungo periodo e stagionalità annuale

La stima della componente deterministica di lungo periodo LP_t (anche per domanda e margine) viene effettuata mediante il metodo delle *spline* di lisciamiento. Nel nostro caso ipotizziamo che per estrarre LP_t non sia possibile definire una componente parametrica adatta dato l'andamento dei prezzi nel tempo, bensì si debba procedere alla seguente regressione

$$P_t = f(t) + (SP_t + FP_t + p_t)$$

dove $f(t) = E(P_t|t) = LP_t$ con $t = 1, 2, \dots, n = 2101$.

Regrediamo i prezzi in funzione del tempo ($t = 1, 2, \dots, 2101$) in modo da ottenere una stima della componente di lungo periodo che per ipotesi sarà una

componente sufficientemente liscia in linea con il concetto di un trend di lungo periodo e componente di stagionalità annuale, $\hat{f}(t) = \widehat{LP}_t$. Per ottenere $\hat{f}(t)$ adoperiamo il metodo delle *spline* di lisciamento con un parametro di banda opportunamente scelto dopo svariate prove per non subire il problema del *over-fitting*. Il criterio della *cross validation* non è affidabile in quanto i nostri dati non sono indipendenti. Applicando le *spline* di lisciamento alle serie di prezzi, doman-

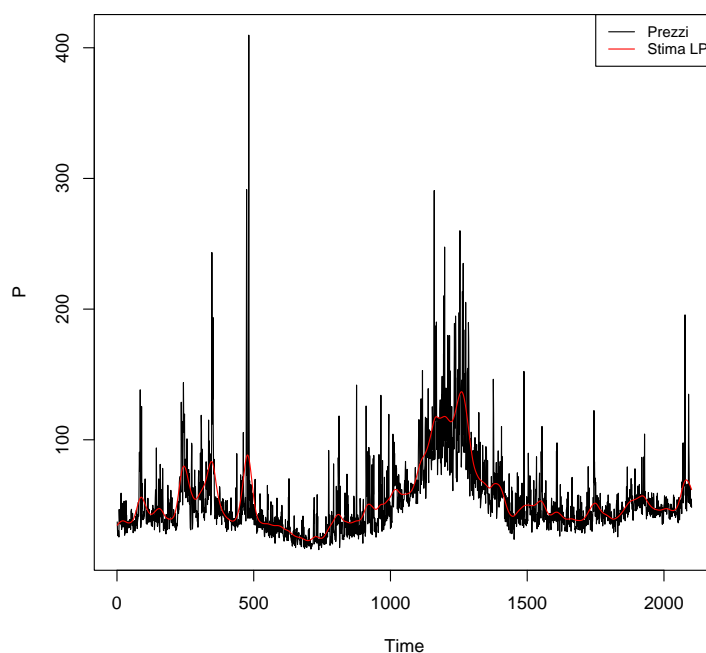


Figura 2.8: Prezzi (in nero) e sovrapposizione della stima della componente LP_t (in rosso) nel mercato elettrico Inglese nella fascia semi oraria 12.00–12.30 per il periodo 01/04/2005 al 31/12/2010.

da e margine (o ai loro logaritmi) con regressore il tempo "t" (per $t = 1, 2, \dots, 2101$) si riesce a stimare non parametricamente la componente di lungo periodo. In figura 2.8 si può osservare come la componente di lungo periodo contenente trend, stagionalità annuale e altre componenti deterministiche di lungo periodo venga sovrapposta ai prezzi per dati della fascia semi oraria 12.00-12.30. Medesima rappresentazione è data in figura 2.9 per domanda e margine. In sezione A dell'appendice si riportano in figura A.4 e A.5 l'estrazione delle componenti di lungo periodo per la stessa fascia oraria per i logaritmi di prezzi, domanda e

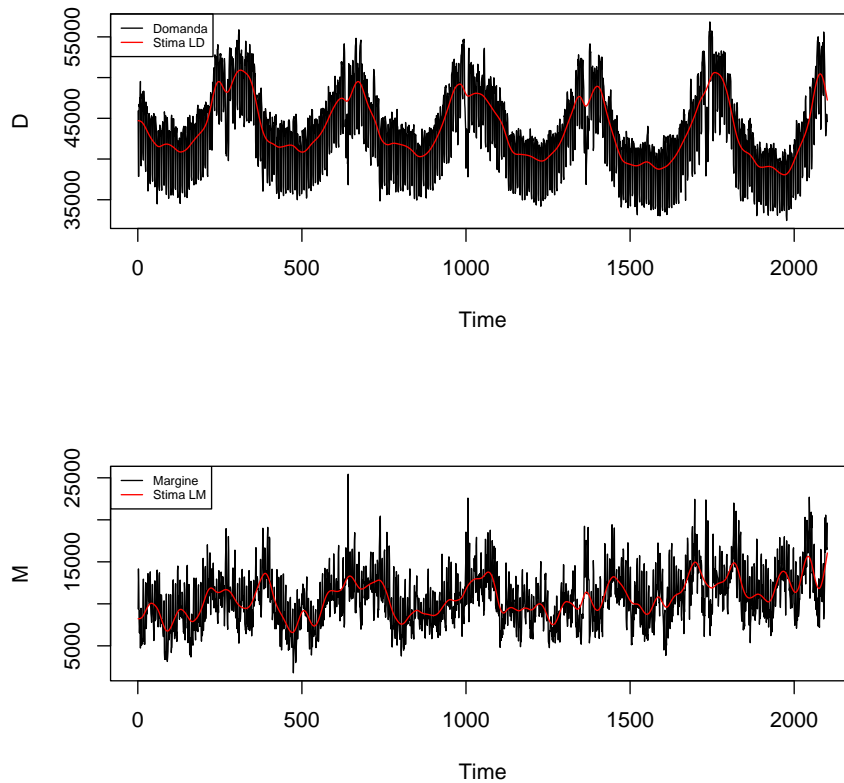


Figura 2.9: In alto Domanda e stima della componente LD_t , in basso Margine e stima della componente LM_t nel mercato elettrico Inglese nella fascia semi oraria 12.00–12.30 per il periodo 01/04/2005 al 31/12/2010.

marginale. Soprattutto per quanto riguarda la domanda si può assumere una netta componente di periodicità annuale.

2.2.3 Stagionalità settimanale giornaliera ed effetto giorni festivi

Dalla definizione di un modello per prezzi, domanda e margine (equazioni (2.5), (2.6) e (2.7)), analoghi anche per i rispettivi logaritmi ma con ipotesi di partenza sui livelli differenti, si è fornita una possibile fattorizzazione della componente deterministica come in equazione (2.8). Avendo proposto una stima delle componenti di lungo periodo (LP_t , LD_t e LM_t) mediante *spline* si è intenzionati ora a stimare gli effetti legati alla periodicità settimanale (effetto del giorno della settimana in cui ci si trova) e l'effetto della periodicità degli effetti di calendario, basandosi sul residuo che si ottiene dalla serie iniziale depurata dalle componenti di lungo periodo. Rimanendo allineati all'esposizione per i soli prezzi, che

risulta analoga negli altri casi, si ha che le componenti di periodicità settimanale e effetti di calendario sono stati ipotizzati in modo additivo nella componente deterministica. A tal proposito si ha che $SP_t + FP_t$, gli effetti periodici congiunti, sono ricavabili dalla serie dei prezzi senza la componente di lungo periodo, mediante una regressione opportuna.

Una stima della stagionalità settimanale \widehat{SP}_t e una stima dell'effetto di festività infrasettimanali \widehat{FP}_t possono essere derivate mediante una regressione sulla serie detrendizzata, ovvero regredendo con opportune *dummy* la serie:

$$P_t - \widehat{LP}_t = SP_t + FP_t + p_t \quad (2.12)$$

La componente di stagionalità settimanale contiene 7 differenti effetti, quelli legati a ciascuna giornata, si definiscono quindi 7 serie di *dummies* che indicano l'appartenenza di ciascuna osservazione ad un giorno della settimana univocamente. L'effetto della periodicità legata all'effetto di calendario viene catturato definendo una *dummy* che indica se un'osservazione è o meno un giorno festivo infrasettimanale. Un modello per la periodicità settimanale è

$$SP_t = \gamma_1 d_{LUN,t} + \gamma_2 d_{MAR,t} + \gamma_3 d_{MER,t} + \gamma_4 d_{GIOV,t} + \gamma_5 d_{VEN,t} + \gamma_6 d_{SAB,t} + \gamma_7 d_{DOM,t} \quad (2.13)$$

dove ad esempio $d_{LUN,t}$ è una variabile indicatrice che vale 1 qualora la t -esima osservazione sia un lunedì zero altrimenti e γ_1 sarà l'effetto marginale sui prezzi attribuibile ai "lunedì", e così via per le altri componenti. Un modello per stimare la periodicità legata agli effetti di calendario è

$$FP_t = \delta d_{holiday,t} \quad (2.14)$$

dove δ rappresenta l'effetto marginale di un giorno di vacanza infrasettimanale sulla serie dei prezzi e $d_{holiday,t}$ è la variabile indicatrice che assume valore 1 se la giornata t -esima rappresenta un giorno festivo infrasettimanale, 0 altrimenti. Unendo quanto specificato in equazione (2.13), (2.14) con (2.12), la stima degli effetti legati alla componente giornaliera e di calendario passa per la stima di $\gamma_1, \gamma_2, \dots, \gamma_7$ e δ che sono derivabili attraverso la stima del modello di regressione

lineare definito da sole *dummies*

$$P_t - \widehat{LP}_t = \gamma_1 d_{LUN,t} + \gamma_2 d_{MAR,t} + \gamma_3 d_{MER,t} + \\ + \gamma_4 d_{GIOV,t} + \gamma_5 d_{VEN,t} + \gamma_6 d_{SAB,t} + \gamma_7 d_{DOM,t} + \delta d_{holiday,t} + p_t$$

dove si nota la mancanza del termine costante per il problema legato all'identificabilità e p_t , il termine d'errore del modello, altro non è che la serie depurata dalle componenti deterministiche che stiamo cercando di estrarre dalle serie di partenza. Queste operazioni si riferiscono a ciascuna fascia semi oraria trattata separatamente.

Una stima mediante minimi quadrati ordinari è possibile ma per robustificare i risultati di stima, ovvero non far risentire alle stime degli effetti marginali della presenza dei picchi, si procede con una regressione mediana. I valori $\hat{\gamma}_1, \hat{\gamma}_2, \dots, \hat{\gamma}_7$ e $\hat{\delta}$ si derivano quindi dalla minimizzazione di

$$\sum_{t=1}^{2101} |(P_t - \widehat{LP}_t) - (\gamma_1 d_{LUN,t} + \gamma_2 d_{MAR,t} + \gamma_3 d_{MER,t} + \dots \\ + \gamma_4 d_{GIOV,t} + \gamma_5 d_{VEN,t} + \gamma_6 d_{SAB,t} + \gamma_7 d_{DOM,t} + \delta d_{holiday,t})|$$

È quindi stata individuata una stima mediana della stagionalità settimanale $\widehat{SP}_t = \hat{\gamma}_1 d_{LUN,t} + \hat{\gamma}_2 d_{MAR,t} + \hat{\gamma}_3 d_{MER,t} + \hat{\gamma}_4 d_{GIOV,t} + \hat{\gamma}_5 d_{VEN,t} + \hat{\gamma}_6 d_{SAB,t} + \hat{\gamma}_7 d_{DOM,t}$ e dell'effetto legato ai giorni festivi (effetti di calendario) $\widehat{FP}_t = \hat{\delta} d_{holiday,t}$. In figura 2.10 riportiamo la distribuzione empirica per le 48 fasce semi orarie delle stime dell'effetto mediano che hanno sui prezzi, sulla domanda e sul margine i vari giorni della settimana (stagionalità settimanale). Una considerazione che potrebbe non apparire ovvia è che l'effetto giornaliero cambia anche nelle varie semi ore della giornata e questo è visibile data la grande variabilità per le stime in ogni giornata. Facendo sempre riferimento alla figura 2.10 si possono fare alcune considerazioni generali valide per la maggior parte delle semi ore in esame. Trovarsi in giorni compresi tra lunedì e venerdì ha un effetto solitamente positivo più o meno marcato sia su prezzi e domanda, questo non ci sorprende in quanto questi sono i giorni lavorativi. Sabati e domeniche hanno invece effetto solitamente negativo sui prezzi e domanda che tendono a scendere. Il margine

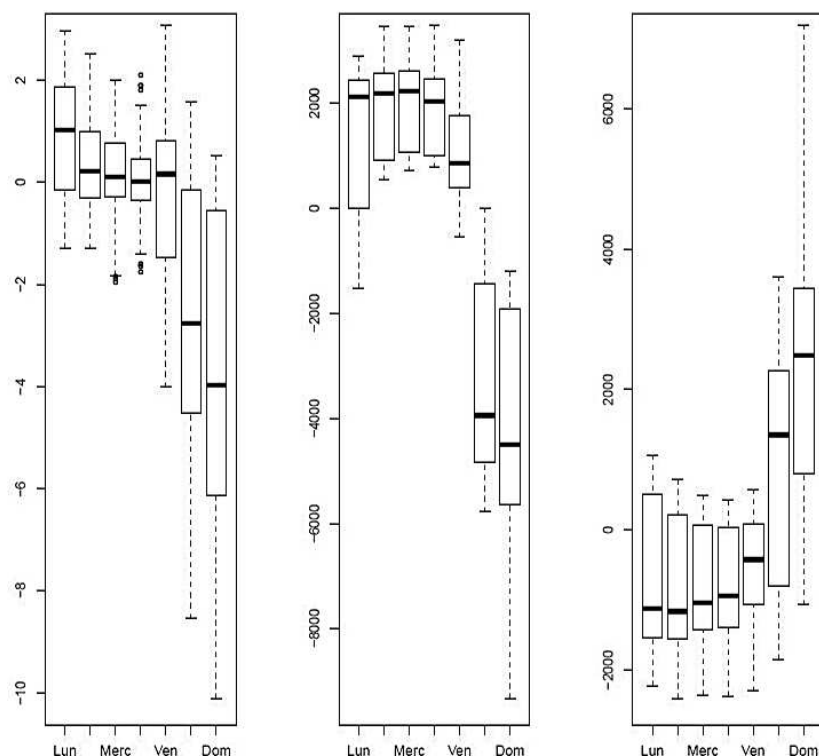


Figura 2.10: Distribuzione giornaliera (nelle 48 semi ore) dell'effetto della periodicità settimanale rispettivamente (da sinistra a destra) sui prezzi, domanda e margine nel mercato Inglese per il periodo 01/04/2005 al 31/12/2010.

subisce una riduzione nei tipici giorni lavorativi, mentre la capacità residua degli impianti tende a essere superiore dei sabati e nelle domeniche (fatto concordante con il comportamento della domanda scando ovvie anomalie nel sistema elettrico). Le semi ore notturne comprese tra le 23 e le 6 si muovono in controtendenza a quanto detto in via generale.

In figura 2.11 si possono osservare per alcune fasce semi orarie le stime dell'effetto dei singoli giorni sui prezzi. Si nota, in dipendenza alla fascia oraria, che l'effetto sui prezzi può essere più o meno forte e inoltre non è deducibile una vera sistematicità di comportamento generalizzabile alle varie fasce semi orarie. È invece deducibile una certa sistematicità per quanto concerne l'effetto sulla domanda e sul margine della componente periodica settimanale che ha una regolarità superiore di quella dei prezzi. In figura A.6 e A.7 (appendice A) si vedono infatti, per alcune fasce semi orarie, comportamenti più netti dell'effetto settimanale su domanda e margine, con alcune eccezioni per le fasce notturne.

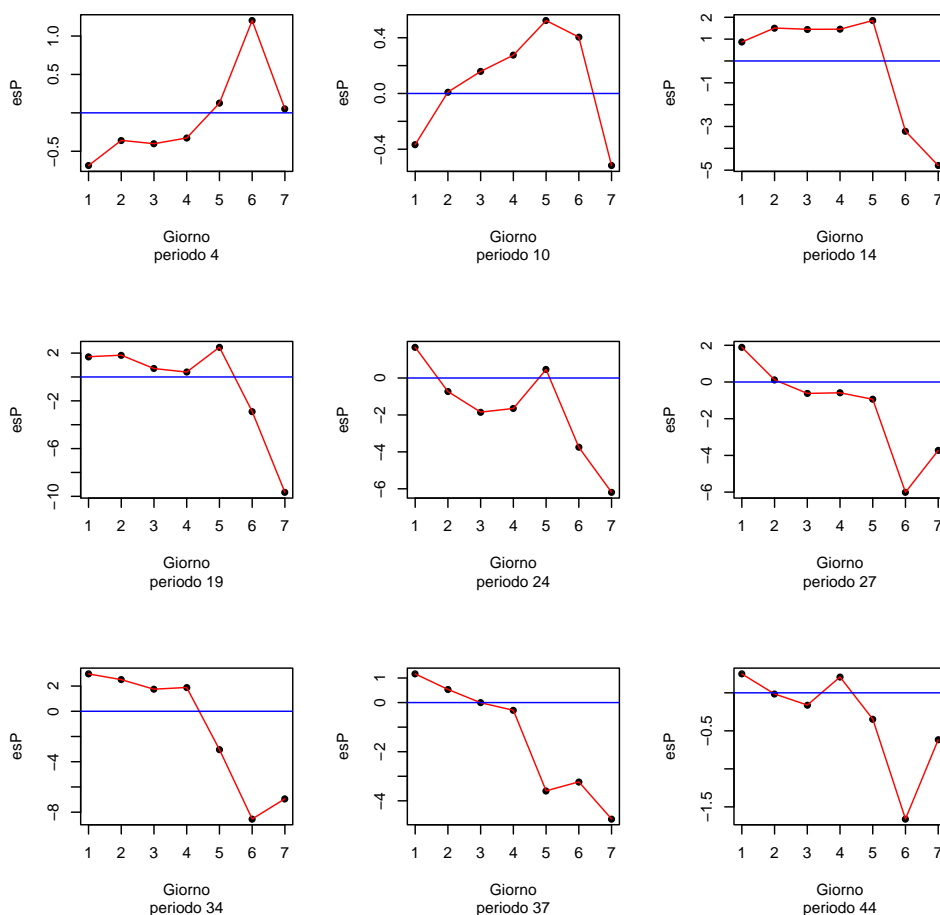


Figura 2.11: Effetto periodicità settimanale sulla serie dei prezzi per alcune fasce semi orarie nel mercato Inglese per il periodo 01/04/2005 al 31/12/2010. (1–Lunedì : $\hat{\gamma}_1$, 2–Martedì: $\hat{\gamma}_2$, ..., 7–Domenica: $\hat{\gamma}_7$). La linea blu evidenzia un effetto nullo.

Una stima degli effetti di calendario su ogni fascia semi oraria per i prezzi, domanda e margine è presentata in figura 2.12; dove si può notare che l'effetto delle festività infrasettimanali riduce i prezzi e la domanda ma aumenta il margine (il che corrobora l'idea teorica sul comportamento dell'effetto delle festività). I giorni festivi hanno effetti più marcati nelle ore comprese tra le 6.00 e le 20.00 e più limitati nelle fasce semi orarie dove non vi è particolare attività.

Per entrambe le componenti periodiche, periodicità settimanale e effetti di calendario, si osserva come le ore centrali della giornata siano solitamente più interessate a questi fenomeni. Questo è plausibilmente dovuto al fatto che nella fascia serale e notturna vi è una minore intensità nei consumi, che possono essere considerati per lo più di mantenimento di attività che non vengono comunemen-

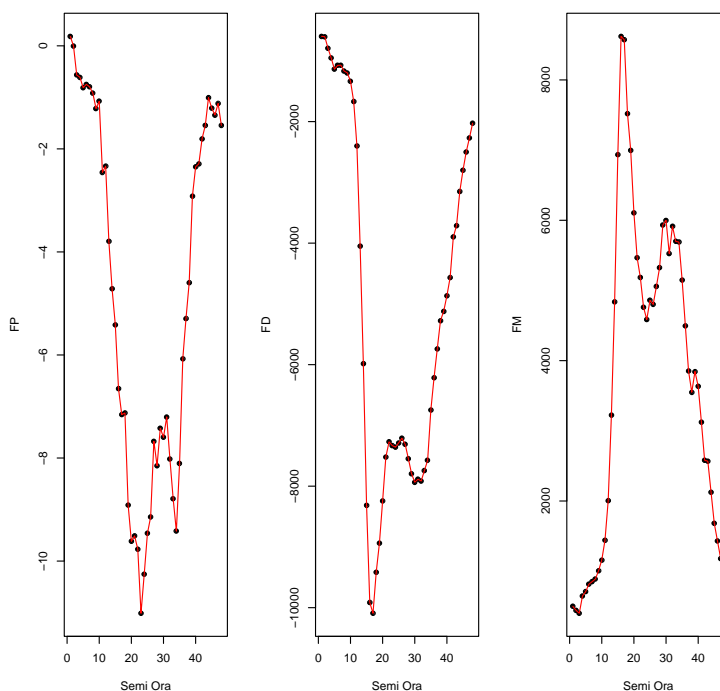


Figura 2.12: Effetti mediani della periodicità legata agli effetti di calendario su prezzi, domanda e margine per ciascuna delle 48 fasce semi orarie nel mercato Inglese per il periodo 01/04/2005 al 31/12/2010.

te interrotte o modificate nel tempo e sono quindi meno influenzate da questi effetti. Sia per i livelli che per le serie logaritmiche la stima degli effetti periodici risulta numericamente non elevata seppur solitamente statisticamente significativa.

Una stima completa della componente deterministica per prezzi, domanda e margine è riportata rispettivamente in figura A.8, A.9 e A.10 per la fascia semi oraria 12.00-12.30 (Appendice A).

2.2.4 Componente stocastica

Dato quanto detto fin'ora è possibile attribuire una stima della componente deterministica per ciascuna fascia oraria per prezzi, domanda e margine (o loro trasformazioni logaritmiche analogamente) come

$$\widehat{DP}_t = \widehat{LP}_t + \widehat{SP}_t + \widehat{FP}_t$$

dove i metodi di stima proposti precedentemente sono solo alcuni dei possibili modi per estrarre le componenti deterministiche che rendono le serie non stazionarie, alternative interessanti sono proposte in Bunn (2006) e Weron (2006).

Dalla stima degli effetti deterministici è possibile attribuire quindi una stima della componente stocastica, per prezzi è

$$\hat{p}_t = P_t - \widehat{DP}_t$$

dove in questo modo riusciamo ad estrarre delle serie plausibilmente utili per la modellazione e sulle quali è possibile adattare opportuni modelli di regressione per la previsione dei picchi di prezzo.

Può essere interessante verificare la stazionarietà di queste serie stocastiche ottenute con un'analisi grafica e dallo studio dei diagrammi di autocorrelazione e autocorrelazione parziale in modo che non vi siano evidenze di componenti deterministiche non colte precedentemente. È assumibile però che le serie stocastiche ottenute conservino una qualche forma di eteroschedasticità e dipendenza seriale che per l'appunto è quella necessaria alla modellazione. In figura 2.13, si può osservare il risultato del processo di depurazione dalle componenti deterministiche che permettono un'estrazione della componente stocastica. Vengono riportate le serie stocastiche per prezzi, domanda e margine per la fascia semi-oraria 18.00-18.30. Da una prima ispezione grafica le serie appaiono relativamente stazionarie in media. Si evidenzia nella componente stocastica dei prezzi una forte eteroschedasticità e momenti di forti picchi di carattere stocastico positivi, elemento da prendere fortemente in considerazione e sfruttabile per la previsione dei picchi. In alcuni momenti nella componente stocastica della domanda appaiono picchi ricorrenti che assumono un ruolo quasi periodico, plausibilmente perché alcuni aspetti stagionali non sono stati completamente stimati dato che la domanda ne risente particolarmente e una procedura di detrendizzazione come quella applicata è molto generica. La serie stocastica del margine solitamente appare stazionaria in media e relativamente anche in varianza. In figura 2.14 si riportano le rappresentazioni grafiche delle funzioni di autocorrelazione e autocorrelazione parziale delle serie stocastiche, si può osservare che le serie so-

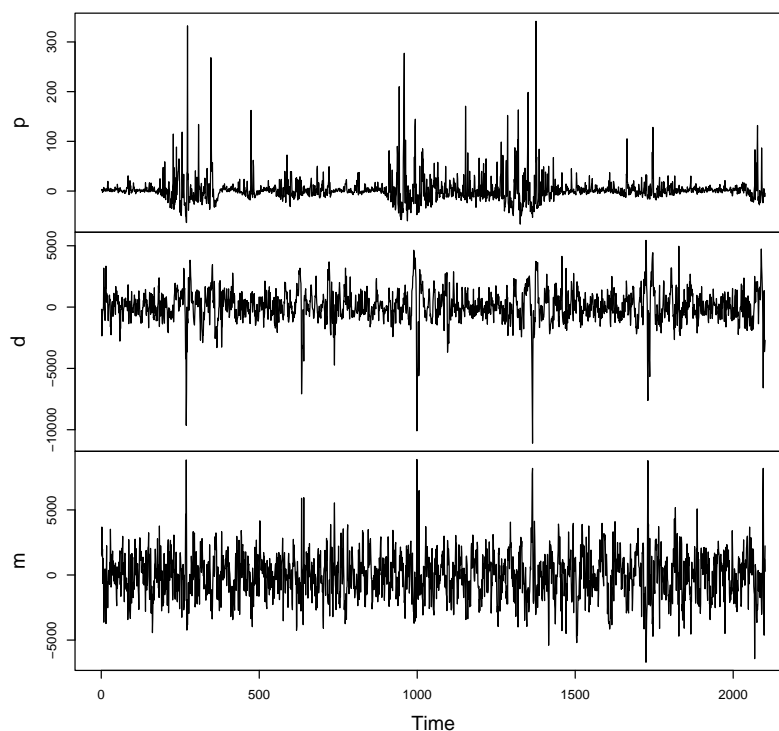


Figura 2.13: Estrazione della componente stocastica rispettivamente per prezzo, domanda e margine del mercato elettrico Inglese nella fascia semi oraria 18.00–18.30 per il periodo 01/04/2005 al 31/12/2010.

no autocorrelate. Può essere d'interesse valutare una autocorrelazione incrociata tra le serie stocastiche in modo da capire se vi siano legami lineari tra ritardi di una componente stocastica ed un'altra, in figura 2.15 ve se ne dà una rappresentazione. Escludendo dall'analisi le rappresentazioni nella diagonale che sono i diagrammi di autocorrelazione delle componenti stocastiche già rappresentati, concentrandosi sulle autocorrelazioni incrociate si può osservare che la serie stocastica dei prezzi ha una relazione lineare non trascurabile con le componenti stocastiche di domanda e margine contemporanee e a ritardo uno.

L'analogia decomposizione delle serie logaritmiche è riportata in figura 2.16 dove si propongono ancora le componenti stocastiche estratte da prezzo, domanda e margine logaritmizzate. In generale si può osservare che la trasformazione logaritmica ha un effetto evidente sulla componente stocastica dei log-prezzi. La presenza di una forte presenza di picchi positivi viene nettamente attenuata e

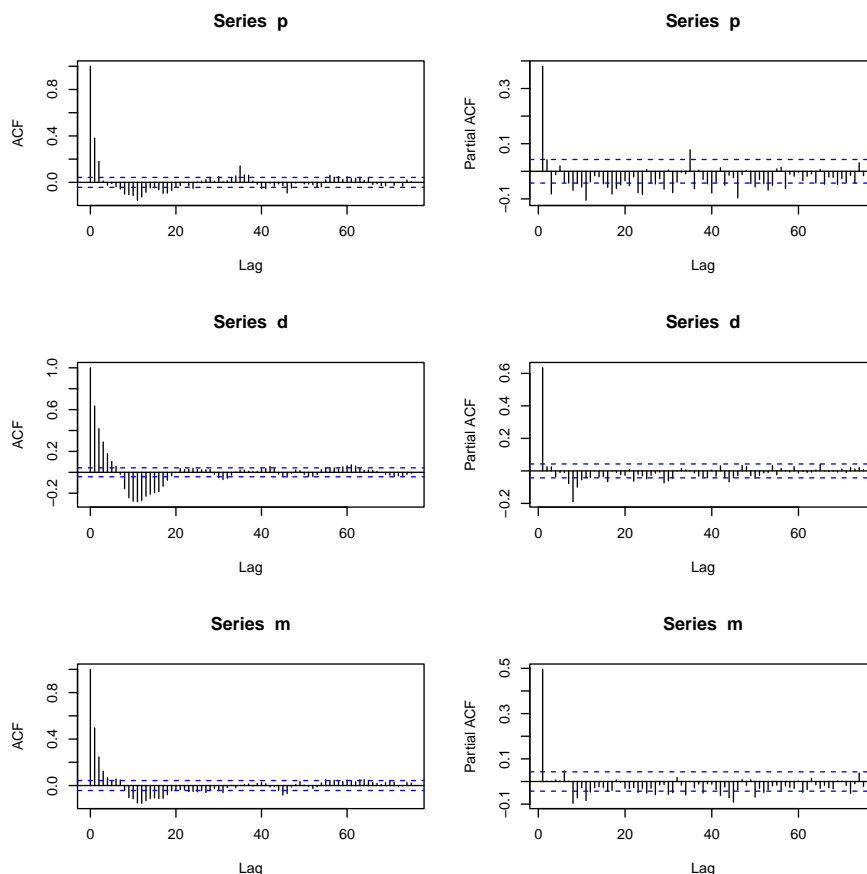


Figura 2.14: Funzione di autocorrelazione e autocorrelazione parziale per i primi 75 ritardi delle componenti stocastiche estratte dalle serie del prezzo, domanda e margine nel mercato elettrico Inglese nella fascia semi oraria 18.00–18.30 per il periodo 01/04/2005 al 31/12/2010.

la serie stocastica è ora nettamente più omoschedastica. Per domanda e margine non si evidenziano particolari cambiamenti sulle componenti stocastiche in generale. Emerge inoltre che la valutazione in generale delle funzioni di autocorrelazione e autocorrelazione parziale non si discosta molto dalle evidenze riportate precedentemente anche per le serie stocastiche derivanti dai logaritmi. Infine valutando solitamente i diagrammi di autocorrelazione parziale si può notare come più marcatamente la componente stocastica di domanda e margine contemporanee e a ritardo uno abbiano una significativa relazione lineare con la serie stocastica dei log-prezzi, si veda come esempio figura A.11 in appendice.

Una decomposizione completa delle serie logaritmiche viene riportata in figura A.12, A.13 e A.14 e rispecchia le stesse caratteristiche enunciate per la decomposizione e l'estrazione della componente deterministica dei livelli.

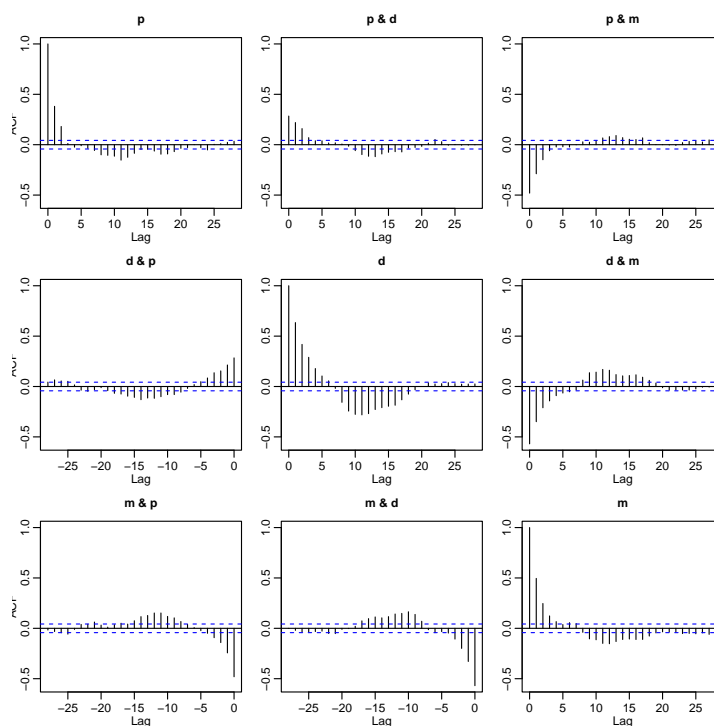


Figura 2.15: Diagramma delle funzioni di autocorrelazione (diagonale) e autocorrelazione incrociate per le componenti stocastiche di prezzo, domanda e margine nel mercato elettrico Inglese nella fascia semi oraria 18.00–18.30 per il periodo 01/04/2005 al 31/12/2010.

Si ottengono in questo modo le componenti stocastiche per tutte le 48 fasce semi orarie e per tutte le serie sia per i livelli che per i logaritmi che saranno fondamentali nella trattazione successiva.

Si osserva che, in alcune fasce semi orarie, le componenti periodiche sono particolarmente forti e le serie molto variabili. La stima della componente deterministica è meno efficace e così di conseguenza la componente stocastica ne può risentire. Questo in genere avviene in fasce semi orarie legate a grossa variabilità e soprattutto per la serie dei prezzi che, a differenza di margine e domanda, presenta caratteristiche molto più di irregolarità. Domanda e margine sono serie sufficientemente regolari nelle componenti deterministiche e sono perciò più facili da decomporre e le serie stazionarie ottenute risultano necessariamente più libere da caratteristiche non stocastiche.

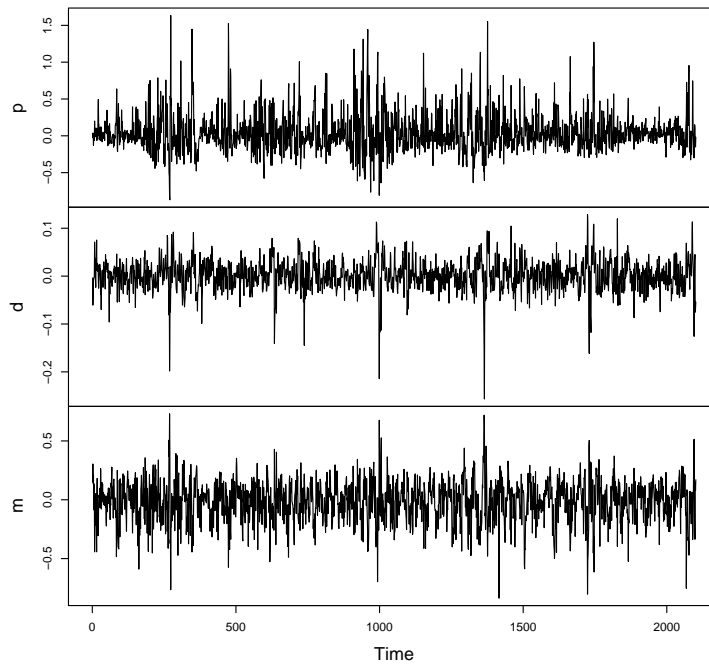


Figura 2.16: Estrazione della componente stocastica rispettivamente per i logaritmi di prezzo, domanda e margine del mercato elettrico Inglese nella fascia semi oraria 18.00–18.30 per il periodo 01/04/2005 al 31/12/2010.

2.3 I Picchi

La definizione di picco di prezzo nei mercati elettrici è una questione tutt'ora dibattuta e non unicamente definita né in modo generale né per uno specifico mercato. Uno degli scopi di questa tesi è proporre un possibile metodo per definire i picchi. L'individuazione dei picchi di prezzo e una loro previsione è un aspetto rilevante almeno per due possibili ragioni. La prima, legata più ad un aspetto predittivo, emerge dalla necessità di effettuare operazioni di *trading* e gestione del rischio di mercato che sono necessariamente influenzate dall'andamento dei prezzi (Christensen et al, 2011). La seconda, legata più ad un aspetto identificativo, emerge dalla necessità che individuare i picchi può essere rilevante per migliorare e rende più robusta la stima delle componenti deterministiche se questi vengono trattati come *outliers* nelle stime di queste componenti (Trueck et al, 2007). La nostra analisi è rivolta ad un aspetto previsionale dei picchi di prezzo che nulla ha a che fare con la robustificazione delle stime della componente

deterministica. Vi sono molti e differenti metodi per definire cos'è un picco di prezzo e per identificarlo riportati in letteratura, si veda ad esempio Bunn (2006) e Weron (2006). Metodi differenti portano a identificare solitamente picchi differenti in modo anche evidente. Stabiliamo una regola per l'identificazione dei picchi in ciascuna fascia semi oraria del mercato elettrico Inglese.

Un picco di prezzo è associabile ad un prezzo che visibilmente supera una certa soglia, o ad un prezzo che subisca un brusco aumento rispetto ai prezzi dei periodi precedenti. I prezzi però dopo un picco si riallineano velocemente al livello medio. Spesso viene fissata come soglia per definire i picchi un livello di

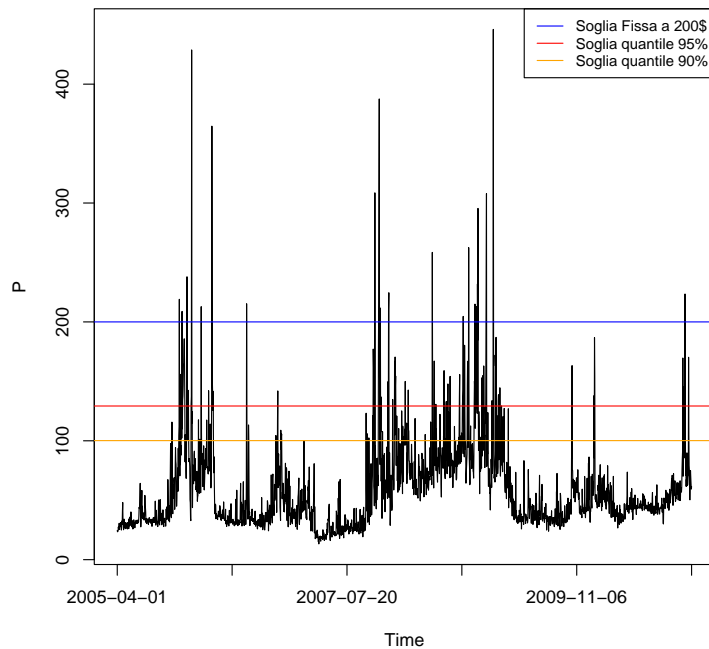


Figura 2.17: Identificazione dei picchi di prezzo mediante soglie stabilite in modo arbitrario o come quantili della distribuzione empirica dei prezzi di livello elevato nel mercato elettrico Inglese nella fascia semi oraria 18.00–18.30 per il periodo 01/04/2005 al 31/12/2010.

prezzo costante nel tempo di valore elevato, oppure vengono definiti picchi una piccola frazione delle osservazioni totali sulle quali si può derivare una soglia costante implicitamente. Nel primo caso si fissa una soglia di prezzo arbitraria legata all'esperienza degli operatori e conoscenza di informazioni che permettano di definire un *target*, esempio: sono definiti picchi tutti gli istanti in cui il prezzo

supera 200\$. Un altro metodo, quello di definire come picchi una frazione totale delle osservazioni con prezzi elevati, potrebbe essere quello di definire picchi tutti gli istanti in cui il prezzo associato supera il quantile di livello τ –esimo della distribuzione empirica dei prezzi. Per un τ sufficientemente elevato, esempio 95%, ancora questo approccio stabilisce una soglia di prezzo costante. L'esempio per i due metodi precedenti è fornito in figura 2.17, ma emerge subito una critica. Decidere di utilizzare una soglia fissa nel tempo per definire i picchi dei prezzi pone il problema di individuare effettivamente i picchi. In un orizzonte di breve periodo e con una valutazione locale dei prezzi l'idea può essere valida, ma osservando un andamento di lungo periodo questa non è più sensata. La componente deterministica, soprattutto quella di lungo periodo legata alla presenza di un *trend* e stagionalità annuale (LP_t) induce la serie dei prezzi ad assumere in alcuni periodi dell'anno valori sistematicamente più elevati o sistematicamente più bassi. Questo comporta che individuare i picchi con una soglia fissa sui prezzi può portare a non considerare effettivi picchi che non superano la soglia in quanto sottoposti alla componente deterministica (che abbassa il livello dei prezzi in alcuni momenti dell'anno). Allo stesso modo si possono individuare picchi che non sono tali, ma il cui valore di prezzo supera la soglia perché mantenuto elevato dato il livello medio di partenza del periodo. In figura 2.17 notiamo come prendendo a riferimento la soglia di livello 200\$, si vadano a escludere picchi anche evidenti perché partono da un livello medio molto basso. Questo ragionamento si riscontra in ogni fascia semi oraria, e ciò induce al problema di definire una regola più consistente per identificare un picco.

Per eliminare il problema legato alla componente deterministica, definiamo il picco di prezzo quando la componente stocastica dei prezzi p_t , ovvero il prezzo depurato dalla componente deterministica, eccede una certa soglia¹. Questa regola è più ragionevole della precedente in quanto l'idea sottostante alla nostra decomposizione è che la componente deterministica D_t sia sufficientemente liscia

¹L'idea rimane quella di continuare a lavorare sui livelli dei prezzi dove la presenza dei picchi è marcatamente visibile e intuibile e non sulle trasformazioni logaritmiche che invece riducono la presenza dei picchi omogeneizzando la serie.

e prevedibile, ma che nulla ha a che fare con i picchi che si osservano nella serie dei prezzi ma che anzi, ne distorce la valutazione. La componente che definisce e contiene il fattore che induce ad un picco è quella stocastica, che come visto risulta pressoché stazionaria in media e priva di andamenti tendenziali e stagionali. La soglia al di sopra del quale viene definito un picco osservando la componente stocastica può essere definita mediante quantili elevati della distribuzione empirica stessa di p_t , si veda Trueck et al (2007). Un picco viene identificato quando il valore della componente stocastica supera il quantile di livello 98% della sua stessa distribuzione empirica. La soglia basata sul 98 –esimo quantile è legata alla capacità di discriminare tra picchi o semplici valori elevati (cosa che non fanno soglie sotto il 95 –esimo quantile) senza però essere troppo restrittiva rischiando di non valutare la presenza di picchi (soglie superiori al 99 –esimo quantile). Un picco viene definito all’istante t quando

$$\hat{p}_t > \hat{q}_{0.98} \text{ per } t = 1, 2, \dots, 2101$$

dove $\hat{q}_{0.98}$ è tale che $\hat{F}(\hat{q}_{0.98}) = 0.98$ e $\hat{F}(q)$ è la funzione di ripartizione empirica della componente stocastica dei prezzi. Dalla regola di definizione dei picchi si può definire una soglia sui prezzi variabile nel tempo

$$TargetLevel_t = \widehat{DP}_t + \hat{q}_{0.98}. \quad (2.15)$$

ripristinando le componenti deterministiche stimate che vengono reintrodotte. Identificare i picchi con la serie stocastica o con i prezzi e la relativa soglia stabilita in (2.15) porta agli stessi risultati. Ai fini dell’esposizione però parleremo di prezzi che superano la soglia (2.15) per definire un picco.

In figura 2.18 viene mostrato graficamente il modo in cui si stabiliscono i picchi con la soglia sulla componente stocastica (il 98-esimo quantile della distribuzione in linea rossa) per la semi ora 18.00–18.30. La medesima soglia riallineata ai prezzi è ora variabilità nel tempo (linea rossa). I picchi vengono identificati come i prezzi che sfiorano la soglia rossa che può variare nel tempo.

Ai fini illustrativi in figura A.15 (Appendice A) viene riportata l’identificazione dei picchi per la fascia semi oraria 18.00–18.30 utilizzando la componente

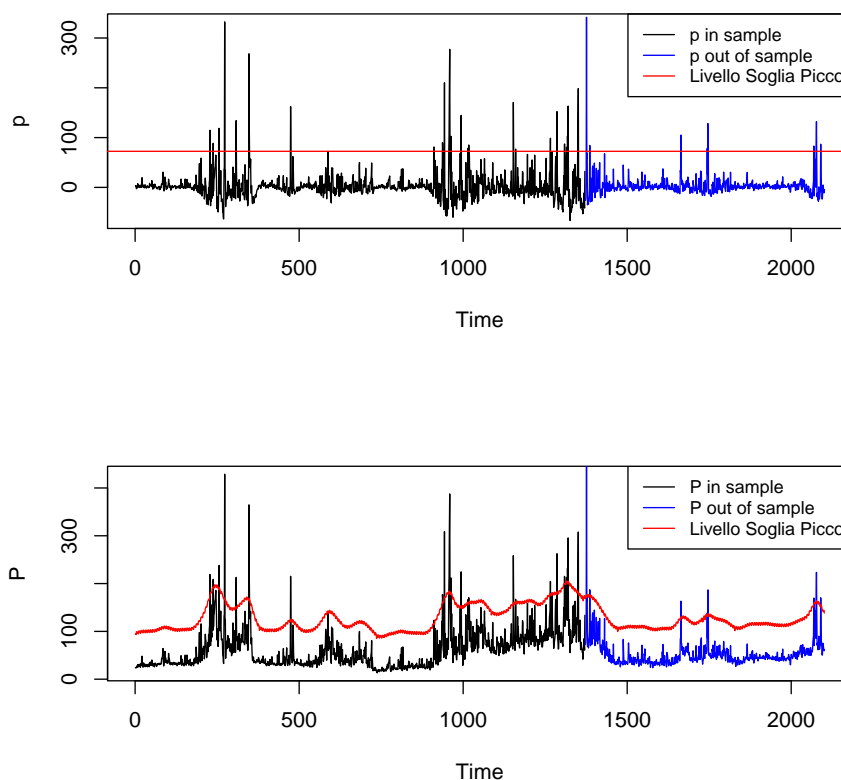


Figura 2.18: Definizione dei picchi di prezzo del mercato elettrico Inglese nella fascia semi oraria 18.00–18.30 per il periodo 01/04/2005 al 31/12/2009. In alto la componente stocastica derivante dai prezzi depurati dalla componente deterministica D_t divisa in campione *in sample* (nero) e *out of sample* (blu). La linea rossa è il quantile di livello 98% che identifica la soglia per identificare gli istanti in cui avvengono i picchi. In basso la serie dei prezzi (divisa in campione *in sample* (nero) e *out of sample* (blu)) alla quale è sovrapposta la soglia che identifica un picco di prezzo se oltrepassata.

stocastica della trasformazione logaritmica dei prezzi con medesimo approccio utilizzato per i prezzi. È possibile confrontare se i picchi identificati da logaritmi dei prezzi siano uguali o meno a quelli identificati dai prezzi; in questa fascia semi oraria il 73% dei picchi è previsto nello stesso modo. In generale anche per le alte serie i risultati sono simili, vi è quindi una chiara differenza nel fissare i picchi partendo dai prezzi o le relative trasformazioni logaritmiche anche del 20/30%.

Per ciascuna fascia semi oraria si procede all'identificazione dei picchi partendo dalla serie dei prezzi come fatto in figura 2.18, la trasformazione logaritmica non è indicabile per l'individuazione dei picchi perché ne attenua troppo la presenza come valutato e riportato in letteratura, Bunn (2006).

Si osservano in figura 2.18 e A.15 (Appendice A) che la serie dei prezzi o dei logaritmi e le relative componenti stocastiche sono rappresentata con due colori differenti, ora ne scopriremo il motivo.

2.4 Modelli di previsione dei picchi di prezzo

L'obiettivo di questa elaborato è proporre e definire un opportuno modello che fornisca previsioni ad un passo, *day-ahead*, per prevedere se nel giorno successivo si manifesterà o meno un picco di prezzo.

Per riuscire a valutare i risultati della nostre previsioni e l'effettiva capacità dei modelli di prevedere i picchi si decide di bipartire i dati campionari a disposizione in due sotto campioni. Un campione definibile come il campione *in sample* e l'altro definibile come campione *out of sample*. Il primo campione, *in sample*, rappresenta l'insieme di dati sui quali si procede ad una calibrazione e stima dei modelli. Le previsioni dei picchi di prezzo un passo in avanti, che si traducono in particolari previsioni del prezzo per il giorno successivo, avvengono invece nella seconda parte del campione, *out of sample*, che rappresenta il campione di controllo.

Data la lunghezza delle serie si opta nel fissare la dimensione del campione *in sample* in 1371 osservazioni (3 anni e 9 mesi di osservazioni circa) e quindi la dimensione del campione *out of sample* risulta di 730 osservazioni. Si definiscono quindi previsioni dei picchi di prezzo da un giorno per il successivo, per gli ultimi due anni di osservazioni che si hanno a disposizione. In figura 2.18 si distingue con colori differenti la parte del campione atta alla stima del modello (in nero) e quella da prevedere (in blu) sia nella serie dei prezzi che nella componente stocastica. Si nota come la fase di definizione dei picchi prescinde totalmente dalla previsione di questi, i picchi che si realizzeranno (stabiliti con la nostra soglia applicata alla serie dei prezzi) nel campione *out of sample* sono noti a priori della previsione. In generale per ciascuna fascia semi oraria non è possibile definire un numero univoco di picchi nella parte del campione *out of sample* in quanto questi vengono fissati globalmente.

La previsione di un picco di prezzo per il giorno successivo passa necessariamente per una previsione dei prezzi *day-ahead* (nel campione *out of sample*). Uno schema per una previsione ad un passo ragionevole è

$$\widehat{P}_{t+1} = \widehat{DP}_{t+1} + \widehat{p}_{t+1} \quad \text{per } t = 1731, \dots, 2100$$

dove $\widehat{DP}_{t+1} = \widehat{LP}_{t+1} + \widehat{SP}_{t+1} + \widehat{FP}_{t+1}$ è la previsione della componente deterministica per il giorno successivo e \widehat{p}_{t+1} è una previsione per la componente stocastica per il giorno successivo. Le serie stocastiche di prezzi, domanda e margine vengono ritenute totalmente note ed utilizzabili fino all'istante t per la previsione in $t + 1$ che avanza passo passo. La stima della componente deterministica \widehat{DP}_{t+1} , ha componenti stagionali definibili con facilità per il giorno successivo essendo componenti periodiche di periodo 7 (\widehat{SP}_{t+1}) e sono noti a priori i giorni dell'anno in cui ci saranno feste infrasettimanali (\widehat{FP}_{t+1}). Per completare la stima di DP_{t+1} occorre attribuire una stima alla componente di lungo periodo per il giorno successivo ovvero LP_{t+1} . Seppur si possa ricorrere ad una stima non parametrica di LP_{t+1} , un'ipotesi semplificatrice può essere quella di assumere che la stima della componente di lungo periodo possa assumere il valore di quella del giorno precedente. Questo è assumibile se il livello medio della serie varia nel tempo in maniera lenta e tra un giorno e il successivo il differenziale della componente di lungo periodo non è elevato (ciò accade), quindi $\widehat{LP}_{t+1} = \widehat{LP}_t$.

È invece sulla stima della componente stocastica p_t , "causa" dei picchi dei prezzi, che si possono definire differenti modelli di previsione. L'idea generale per definire una previsione della componente stocastica è quella di sfruttare l'idea di un modello di regressione che utilizzi come esplicative la dipendente ritardata, la componente stocastica di domanda e margine eventualmente anche ritardati ed un eventuale tipo d'interazione. Il modello è

$$\widehat{p}_{t+1} = \widehat{f}(p_t, p_{t-1}, p_{t-2}, \dots, p_{t-6}, d_{t+1}, d_t, m_{t+1}, m_t, dm_{t+1}, dm_t) \quad (2.16)$$

dove con p_{t-j} indichiamo la componente stocastica dei prezzi ritardata (estratta dalla serie dei prezzi senza componente deterministica), con d_t e m_t le componen-

ti stocastiche di domanda e margine e rispettivi ritardi e con dm_t un iterazione tra domanda e margine ottenuta come $dm_t = d_t m_t$ o eventualmente come $dm_t = \frac{d_t}{m_t}$.

Vengono solitamente impiegati come variabili esplicative le dipendenti ritardate $p_t, p_{t-1}, p_{t-2}, \dots, p_{t-6}$ (i primi 7 ritardi) in quanto si è dimostrata la dipendenza seriale nella componente stocastica dei prezzi. In generale, in riferimento al processo di detrendizzazione impiegato, è osservabile che alcune componenti legate alla stagionalità settimanale rimangono non colte in alcune fasce semi orarie, risulta utile inizialmente considerare un ritardo di ordine 7 come esplicativa rispetto il valore da prevedere e gli eventuali ritardi intermedi. Domanda e margine vengono inoltre impiegate perché vi è una correlazione incrociata significativa contemporanea e per il ritardo uno di queste serie con la serie stocastica del prezzo. Si nota come sono state considerate per la previsione ad un passo m_{t+1} e d_{t+1} (e quindi anche dm_{t+1}), queste non sono incognite in quanto la serie della domanda e del margine sono *day-ahead*, ovvero rappresentano previsioni per il giorno successivo. L'uso di tutte queste esplicative nei modelli di previsione per ciascuna fascia semi oraria non è necessario, si selezioneranno le sole variabili indispensabili.

Definita una previsione dei prezzi *day-ahead*, si definisce una *previsione di picco* per il giorno successivo qualora il prezzo previsto \hat{P}_{t+1} supera la soglia che per i prezzi definisce un picco reale. Si confronterà poi se i picchi effettivamente previsti siano quelli che si realizzano, controllando anche il numero di falsi allarmi, ovvero previsioni di picchi che non si realizzeranno.

Non è detto che i modelli di previsione dei prezzi *day-ahead* debbano basarsi sulla componente stocastica derivante dai livelli. È infatti opportuno, si veda figura 2.19 e 2.20, osservare che le componenti stocastiche derivanti da trasformazioni logaritmiche dei prezzi, della domanda e del margine presentano relazioni più evidenti e di natura lineare delle rispettive componenti stocastiche derivanti dalle serie non trasformate, si vedano le figure A.16 e A.3 (Appendice A). Si riportano inoltre in appendice le relazioni tra la componente stocastica e la serie ritardata di domanda e margine e le loro interazioni per le fascia semi oraria delle

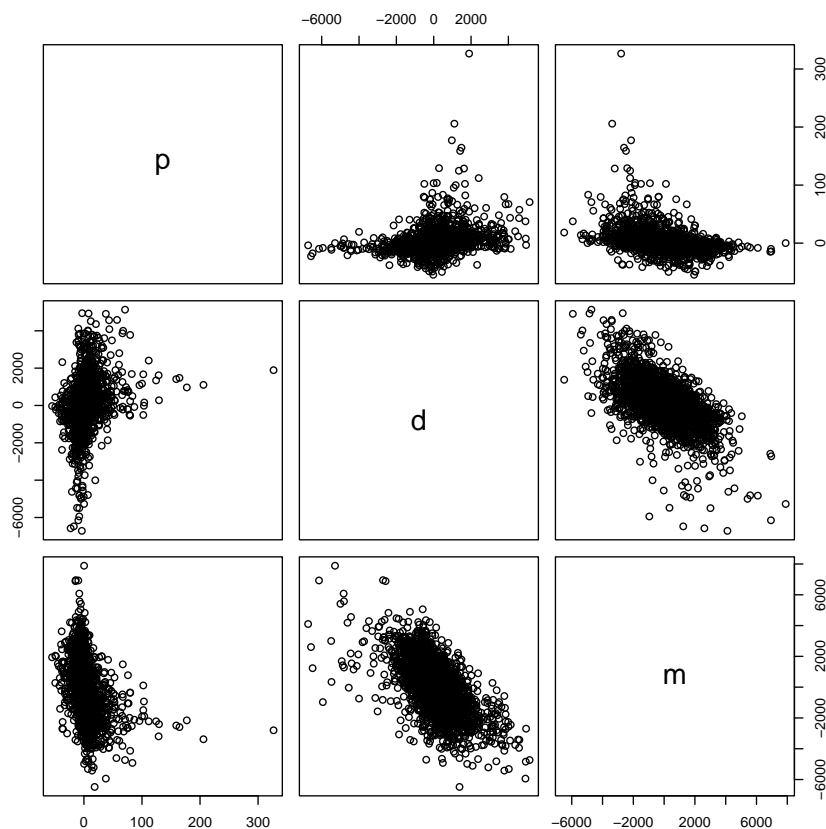


Figura 2.19: Relazioni tra le serie stocastiche di prezzi, domanda e margine trasformati logaritmicamente nel mercato elettrico Inglese nella fascia semi oraria 18.00–18.30 per il periodo 01/04/2005 al 31/12/2010.

12.00-12.30, si vedano figure A.17 e A.18. L'interazione derivante dal rapporto tra domanda e margine ed eventualmente il loro ritardo di ordine 1 non hanno alcuna capacità esplicativa della componente stocastica dei prezzi. Inoltre, operare su componenti stocastiche derivanti dai livelli pone dei problemi di "esposizione" nella fase di previsione ad un passo. Ecco perché si definiscono modelli sulle serie stocastiche dei logaritmi. Il fatto di poter utilizzare serie di partenza trasformate con il logaritmo non toglie la necessità e lo scopo di prevedere i prezzi *day-ahead*. Lo schema di previsione rimane sempre quello proposto in equazione (2.16) anche con le serie stocastiche derivate dai logaritmi delle serie originali, bisogna però operare una trasformazione di tali previsioni basata sulla funzione esponenziale in modo da ottenere previsioni dei prezzi. Da un'analisi grafica si possono osservare che le variabili esplicative che sembrano avere una relazio-

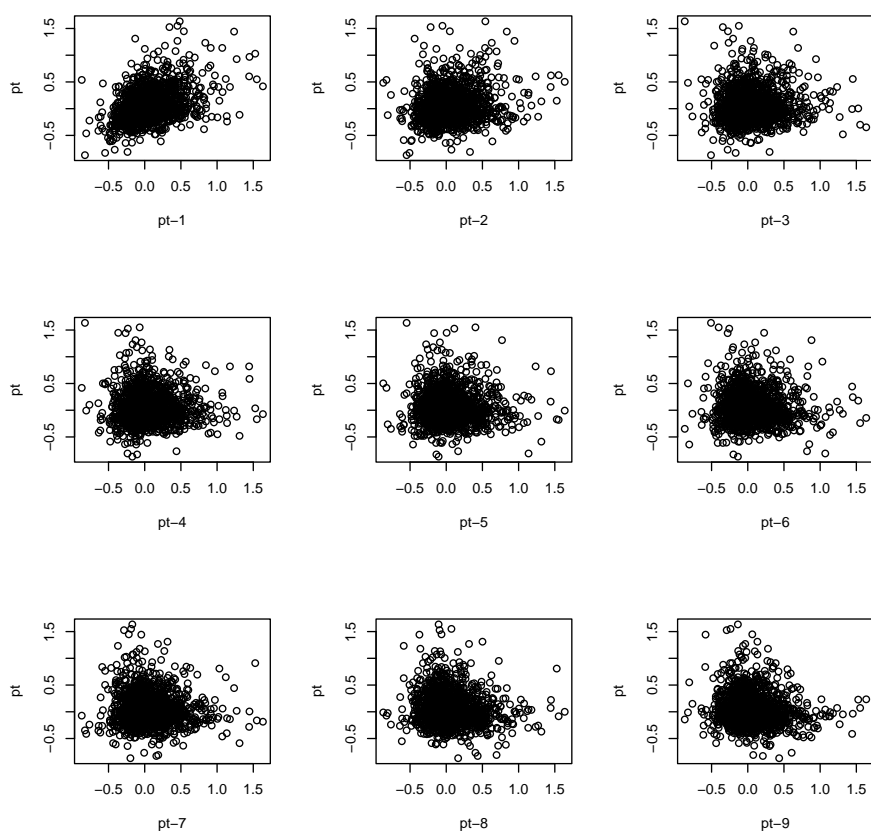


Figura 2.20: Relazioni tra la componente stocastica estratta dal logaritmo dei prezzi e i suoi 9 ritardi nel mercato elettrico Inglese nella fascia semi oraria 18.00–18.30 per il periodo 01/04/2005 al 31/12/2010.

ne più marcata con la componente stocastica dei prezzi sono il ritardo di ordine uno dei prezzi stessi, domanda e margine contemporanei; le interazioni appaiono abbastanza insensate; e non è chiara una presenza di relazione indotta da ritardi superiori dei prezzi, margine e domanda. Date le evidenze di linearità riscontrate in ogni fascia semi oraria tra p_t e le possibili esplicative, si può pensare di utilizzare modelli di regressione additivi e lineari. L'ipotesi di additività è comunque una scelta ragionevole per non dover gestire un modello troppo complesso. L'obiettivo sarà quindi inizialmente definire un opportuno modello di regressione che può essere un modello additivo lineare definito come

$$\hat{p}_{t+1} = \beta_0 + \beta_1 p_t + \beta_2 p_{t-1} + \dots + \beta_6 p_{t-6} + \beta_7 d_{t+1} + \beta_8 d_t + \beta_9 m_{t+1} + \beta_{10} m_t + \beta_{11} dm_{t+1} + \beta_{12} dm_t \quad (2.17)$$

dove la stima di p_{t+1} passa attraverso la stima dei β , oppure sempre per un mo-

dello additivo ma con componenti esplicative di natura non parametrica (o semi parametrica se alcune f_i sono definite come funzione identità) definito come

$$\hat{p}_{t+1} = \beta_0 + \beta_1 f_1(p_t) + \dots + \beta_6 f_6(p_{t-6}) + \\ + \beta_7 f_7(d_{t+1}) + \beta_8 f_8(d_t) + \beta_9 f_9(m_{t+1}) + \beta_{10} f_{10}(m_t) + \beta_{11} f_{11}(dm_{t+1}) + \beta_{12} f_{12}(dm_t)$$

La previsione dei logaritmi dei prezzi un passo in avanti come $\log \widehat{P}_{t+1} = \widehat{D}_{t+1} + \hat{p}_{t+1}$ dai quali si possono stimare i prezzi *day-ahead* come $\widehat{P}_{t+1} = e^{\log \widehat{P}_{t+1}}$.

Capitolo 3

MODELLI DI PREVISIONE DEI PREZZI FUTURI E DEI PICCHI DI PREZZO

Quanto detto nel capitolo 2 conduce la nostra analisi a proporre modelli specifici per la previsione dei prezzi per ottenere successivamente una previsione dei picchi di prezzo. Le previsioni sono svolte da un giorno al successivo (*day-ahead*), sfruttando l'intero *set* informativo passato e presente al momento della previsione; verranno effettuate 730 previsioni ad un passo (campione *out of sample*). La proposta di vari filoni di modelli è motivata dal voler selezionare un possibile modello adatto per la previsione dei picchi di prezzo, individuati in sezione 2.3, che consenta un'elevata capacità predittiva mantenendo limitato il numero di falsi allarmi (ovvero previsioni di picchi che non si realizzeranno). Data l'evidenza sulla linearità delle relazioni, utilizzando componenti stocastiche derivanti da trasformazioni logaritmiche, proponiamo inizialmente dei modelli previsionali lineari basati sul metodo di regressione lineare multivariata che sfruttano la dipendente ritardata e le serie di margine, domanda e loro interazioni anche ritardate come in equazione (2.17). Definiamo in questo modo modelli del tipo *ARX* ovvero autoregressivi con esplicative esogene. Proseguiamo presentando un'analisi dettagliata su due fasce semi orarie 12.00–12.30 e 18.00–18.30 che viene poi estesa a tutte le altre fasce semi orarie e di cui verranno presentati i risultati globalmente.

Un'anticipazione sul tipo di previsione di prezzo effettuata può essere qui già introdotta. Tutto deriva dalla considerazione che i picchi sono valori estremi nella serie dei prezzi. L'idea non sarà prevedere prezzi medi, ma quella di prevedere

quantili condizionati dei prezzi per il giorno successivo di livello opportuno. Le previsioni di picco avvengono quando il quantile del prezzo *day-ahead* previsto supera la soglia. Per la previsione di quantili dei prezzi si utilizzano principalmente tre modelli. Il primo, presentato in questo capitolo, sfrutta il modello di regressione lineare classico. Siccome vengono stimati quantili condizionati, esiste un modello di regressione idoneo e ad *hoc* per il problema: il modello di regressione del quantile. Il secondo metodo per le previsioni dei quantili è il metodo della regressione lineare parametrica del quantile. Il terzo metodo sfrutta la regressione non parametrica del quantile. Questi due metodi vengono applicati nel capitolo 5.

3.1 Modello di Regressione Lineare Classico

Il modello di regressione lineare classico è impiegato per lo studio della media condizionata di una variabile oggetto di interesse dato il comportamento di alcune variabili esplicative, questo è rappresentabile come

$$y_t = \beta_0 + \beta_1 x_{t,1} + \dots + \beta_p x_{t,p} + \varepsilon_t = x_t' \beta + \varepsilon_t \quad \text{per } t=1, \dots, T \quad (3.1)$$

dove y_t rappresenta la variabile risposta o dipendente oggetto dell'analisi, $x_{t,j}$ per $j = 1, \dots, p$ sono le variabili esplicative, β_j per $j = 1, \dots, p$ sono i parametri associati a ciascun regressore che quantificano l'effetto sulla dipendente e ε_t identifica un termine stocastico o comunemente detto termine d'errore. La rappresentazione in forma compatta si ottiene definendo il vettore delle esplicative all'istante t come $x_t = (1 \ x_{t,1} \ x_{t,2} \ \dots \ x_{t,p})'$ e il vettore dei parametri $\beta = (\beta_0 \ \beta_1 \ \dots \ \beta_p)'$.

Il modello (3.1) rappresenta un classico modello di regressione lineare multivariato utile per stimare gli effetti delle esplicative sulla risposta oppure, nella previsione e stima della dipendente per opportuni valori dei regressori. Il modello è lineare per la linearità nei parametri, ma non necessariamente nella variabile risposta e nelle esplicative che possono essere anche una trasformazione di livelli (per esempio una trasformazione logaritmica), in tal caso l'interpretazione degli effetti marginali può cambiare.

Le assunzioni classiche del termine d'errore, per specificare un buon modello sono:

1. $E(\varepsilon_t | x_1, \dots, x_p) = E(\varepsilon_t) = 0$ per $t=1, \dots, T$: la media condizionata dell'errore è nulla, di conseguenza la marginale, ipotesi di esogenità stretta;
2. $Var(\varepsilon_t | x_1, \dots, x_p) = E(\varepsilon_t^2 | x_1, \dots, x_p) = E(\varepsilon_t^2) = \sigma^2$: ipotesi di omoschedasticità;
3. $E(\varepsilon_i \varepsilon_j | x_1, \dots, x_p) = E(\varepsilon_i \varepsilon_j) = 0$ per $i \neq j$: ipotesi di incorrelazione;
4. ε_t è indipendente da $x_{j,s}$ per ogni t, j e s : ipotesi di indipendenza con i regressori (vero se i regressori non sono stocastici);
Ed eventualmente in via facoltativa
5. $\varepsilon_t \sim N(0, \sigma^2)$: ipotesi di normalità;

Date T osservazioni, il metodo più frequentemente usato per stimare i coefficienti ignoti è il metodo dei minimi quadrati ordinari (*Ordinary Least Squares*: OLS). Le stime OLS dei coefficienti di regressione derivano dalla minimizzazione, per $\beta \in R^p$, della somma dei quadrati $\sum_{t=1}^T (y_t - x_t' \beta)^2$ e indicheremo con $\hat{\beta}_j$ per $j = 1, \dots, p$ la stima dei singoli effetti marginali.

Data l'ipotesi 1 sull'errore stocastico questo comporta che $E(Y_t | x_{t,1}, \dots, x_{t,p}) = \beta_0 + \beta_1 x_{t,1} + \dots + \beta_p x_{t,p}$ il quale dimostra come il modello (3.1) fornisca un previsione lineare per la media condizionata della risposta dati i regressori. Le stime ottenute dai minimi quadrati sono quelle che permettono di minimizzare potenzialmente la distanza tra le osservazioni (y_t) e il valore medio condizionato previsto ($\hat{y}_t = \hat{\beta}_1 x_{t,1} + \dots + \hat{\beta}_p x_{t,p} = \hat{E}(Y_t | x_{t,1}, \dots, x_{t,p})$).

Le proprietà statistiche degli stimatori OLS sono determinate dalle assunzioni sulla parte stocastica del modello (3.1), cioè ε_t . Quando tutte le assunzioni sono vere, lo stimatore OLS dei coefficienti di regressione gode delle seguenti proprietà: consistenza, correttezza, efficienza e normalità (è uno stimatore *BLUE*: *Best Linear Unbiased Estimator*). Quando anche i regressori sono stocastici le proprietà degli stimatori dipendono anche dalla distribuzione dei regressori, tuttavia

se ci si limita ad una analisi condizionale ai valori dei regressori osservati, l'unica fonte di variabilità torna a essere l'errore di regressione. Per una trattazione più approfondita si veda Cappuccio e Orsi (2005).

Quando si lavora su dati reali, ed in particolare su serie storiche, alcune delle ipotesi classiche possono non essere più valide. La validità delle ipotesi classiche viene verificata sugli errori di regressione stimati ovvero i residui del modello $r_t = y_t - \hat{y}_t = y_t - \hat{\beta}_1 x_{t,1} - \dots - \hat{\beta}_p x_{t,p}$. Un'ipotesi che spesso viene violata è quella della normalità, gli errori hanno distribuzione solitamente asimmetria e leptocurtica. Quando la varianza dell'errore di previsione non è costante, assenza di omoschedasticità, per tutte le osservazioni, gli stimatori OLS non sono più efficienti e gli *standard error* degli stimatori non sono consistenti. Ciò comporta che anche le statistiche T e i relativi *p-values* per l'inferenza sulle stime siano invalide. Infine può accadere che gli errori non siano serialmente indipendenti, in generale si può assumere che l'ipotesi di incorrelazione seriale venga a mancare. Quando si rileva la presenza di correlazione seriale tra gli errori di regressione gli stimatori OLS perdono la correttezza, l'efficienza e gli errori standard calcolati non sono consistenti; sono quindi possibili due strade: modellare direttamente la correlazione oppure stimare gli errori standard degli stimatori in maniera consistente.

Cerchiamo ora di adattare un modello di regressione lineare classico del tipo autoregressivo con esplicative esogene alle serie stocastiche estratte dalle trasformate logaritmiche per ottenere un modello statistico in grado di prevedere i picchi di prezzo.

3.2 Applicazione del modello di Regressione Lineare ai Dati

Inizialmente è possibile definire il modello di regressione lineare multivariato con più esplicative, da stimare nel campione *in sample*. Si selezionano solo i regressori strettamente necessari per ottenere un adattamento di buon livello e previsioni ad un passo dei prezzi. Attraverso le previsioni dei prezzi è ottenibile una previsione dei picchi per il giorno successivo. Il modello di (auto)regressione lineare

re applicato alla componente stocastica dei log-prezzi stimato sulle prime 1371 osservazioni per ciascuna fascia semioraria è

$$p_t = \beta_0 + \sum_{j=1}^7 \beta_j p_{t-j} + \beta_8 d_t + \beta_9 d_{t-1} + \beta_{10} m_t + \beta_{11} m_{t-1} + \beta_{12} d m_t + \beta_{13} d m_{t-1} + \varepsilon_t \quad (3.2)$$

dove si possono osservare i ritardati della variabile dipendente come esplicativa fino a ritardo 7 (per cogliere eventuali componenti di stagionalità non rimosse dal processo di destagionalizzazione), m_t, m_{t-1} sono la serie stocastica del margine e la sua ritardata, così come per la domanda d_t, d_{t-1} e le interazioni tra domanda e margine contemporanee e ritardate $d m_t = d_t m_t$ e $d m_{t-1} = d_{t-1} m_{t-1}$. Per una rappresentazione in forma compatta definiamo con x_t il vettore delle esplicative che al tempo t contiene tutti i regressori e con β il vettore dei rispettivi parametri, rispettivamente $x_t = (1 \ p_{t-1} \ \dots \ d m_{t-1})'$ e $\beta = (\beta_0 \ \beta_1 \ \dots \ \beta_{13})'$. La rappresentazione compatta del modello è $p_t = x_t' \beta + \varepsilon_t$ per $t = 1, \dots, 1731$. La componente deterministica del modello di regressione, date le ipotesi classiche, altro non è che la media condizionata della componente stocastica del logaritmo dei prezzi data tutta l'informazione disponibile fino al tempo t impiegata, $E(p_t | I_t) = x_t' \beta$.

In tabella 3.1 e 3.2 vengono riportate le stime puntuali, gli *standard error* stimati e i test T sulla significatività dei singoli parametri ottenuti con MQO sotto le ipotesi classiche di un modello di regressione lineare associati a ciascuna esplicativa per la fascia oraria 12.00–12.30 e 18.00–18.30. Nelle tabelle le esplicative vengono codificate con nome della serie, esempio p per la serie stocastica dei logaritmi dei prezzi, e un numero che indica il ritardo temporale della serie, ad esempio $p3 = p_{t-3}$, e così via (esempio di serie non ritardata: $m0 = m_t$).

Si osserva, già per le due fasce orarie considerate, che le esplicative hanno effetti marginali e significatività differenti per ciascuna fascia semi oraria. Si può notare inoltre che, generalmente, ben poche esplicative risultano significative. Per non mantenere un modello così complesso in termine di esplicative come quello di equazione (3.2) cerchiamo di formalizzare un modello altrettanto valido rispetto a quello proposto ma con regressori strettamente necessari. Ne deriva che ciascuna fascia semi oraria deve essere modellata singolarmente in quanto vi

saranno alcune esplicative utili in alcuni casi ma non in altri.

	Stime	Errori Standard	Statistica t	p-value
(Intercetta)	0.0172	0.0069	2.50	0.0124
p1	0.1885	0.0270	6.98	0.0000
d0	-0.4379	0.3146	-1.39	0.1642
m0	-0.9376	0.0451	-20.77	0.0000
d1	1.0830	0.3148	3.44	0.0006
m1	0.2425	0.0514	4.72	0.0000
p2	0.0049	0.0234	0.21	0.8350
p3	-0.0245	0.0233	-1.05	0.2921
p4	-0.0326	0.0232	-1.40	0.1612
p5	-0.0159	0.0232	-0.69	0.4918
p6	-0.0439	0.0232	-1.89	0.0589
p7	0.0202	0.0224	0.90	0.3681
dm0	-1.7449	1.1696	-1.49	0.1360
dm1	-1.2039	1.1708	-1.03	0.3040

Tabella 3.1: Stime dei MQO del modello (3.2) nella fascia semi oraria 12.00–12.30 per il periodo 01/04/2005 al 31/12/2008. Stime dei parametri, stima degli *standard error*, statistica T e relativo *p-value* per valutare la significatività di ogni singola esplicativa.

	Stime	Errori Standard	Statistica t	p-value
(Intercetta)	0.0243	0.0062	3.91	0.0001
p1	0.2927	0.0274	10.69	0.0000
d0	-0.0347	0.2612	-0.13	0.8944
m0	-0.8961	0.0424	-21.11	0.0000
d1	0.1899	0.2610	0.73	0.4670
m1	0.1066	0.0492	2.17	0.0305
p2	-0.0175	0.0245	-0.71	0.4750
p3	-0.0460	0.0244	-1.89	0.0595
p4	-0.0069	0.0245	-0.28	0.7772
p5	-0.0151	0.0244	-0.62	0.5375
p6	-0.0164	0.0244	-0.67	0.5020
p7	-0.0094	0.0226	-0.42	0.6767
dm0	0.4300	0.7804	0.55	0.5817
dm1	0.0606	0.7786	0.08	0.9380

Tabella 3.2: Stime dei MQO del modello (3.2) nella fascia semi oraria 18.00–18.30 per il periodo 01/04/2005 al 31/12/2008. Stime dei parametri, stima degli *standard error*, statistica T e relativo *p-value* per valutare la significatività di ogni singola esplicativa.

Per ottenere un modello con poche e significative esplicative si ricorre ad un classico approccio di eliminazione delle variabili non significative passo–passo; valutando che il modello via via ridotto sia sufficientemente valido in termine di adattamento ai dati rispetto quello di partenza. Attraverso il metodo basato sulla statistica *test* T per la significatività di ciascun regressore si elimina ad ogni

step l'esplicativa con p -value del test più elevato purché superiore a 0.01 (per la j -esima esplicativa: $T_j = \frac{\hat{\beta}_j}{\sqrt{\text{var}(\hat{\beta}_j)}}$ che converge in distribuzione ad una normale standard sotto l'ipotesi nulla di non significatività della esplicativa, $H_0 : \beta_j = 0$ contro $H_1 : \beta_j \neq 0$). Se non vi sono esplicative il cui p -value del test superiore a 0.01 tutte le variabili sono ritenute necessarie. Non basta però ridurre automaticamente il modello per soli test T, bisogna valutare ad ogni step di riduzione che il modello proposto sia valido. Questo viene verificato saggiando un sistema d'ipotesi che contrappone il modello ridotto contro quello completo mediante un test F (basato sulle devianze residue dei modelli) per un livello di significatività del 1%. Se il modello ridotto viene rifiutato dal test non ha senso accettare la riduzione dell'ultima esplicativa che ha portato al modello più sintetico e tanto meno provare con ulteriori riduzioni. Si vuole quindi garantire, non solo vi sia una riduzione del numero di variabili esplicativa ma, che l'adattamento di un modello con pochi regressori sia comunque soddisfacente.

Il modello ARX con sole variabili esplicative necessarie per la fascia semi oraria 12.00–12.30 è

$$p_t = \beta_0 + \beta_1 p_{t-1} + \beta_9 d_{t-1} + \beta_{10} m_t + \beta_{11} m_{t-1} + \varepsilon_t,$$

mentre per la fascia semi oraria 18.00–18.30 è

$$p_t = \beta_0 + \beta_1 p_{t-1} + \beta_{10} m_t + \varepsilon_t,$$

dove le stime dei modelli sono riportate rispettivamente in tabella 3.3 e 3.4.

	Stime	Errori Standard	Statistica t	p-value
(Intercetta)	0.0211	0.0058	3.63	0.0003
p1	0.1926	0.0265	7.26	0.0000
m0	-0.9152	0.0374	-24.48	0.0000
d1	0.7666	0.2648	2.89	0.0039
m1	0.2206	0.0496	4.45	0.0000

Tabella 3.3: Stime dei MQO del modello di regressione lineare con sole esplicative necessarie nella fascia semi oraria 12.00–12.30 per il periodo 01/04/2005 al 31/12/2008.

In tabella B.2 (Appendice B) vengono riportati i modelli di regressione con sole esplicative necessarie per ciascuna fascia semi oraria. L'eliminazione di variabili

	Stime	Errori Standard	Statistica t	p-value
(Intercetta)	0.0218	0.0056	3.89	0.0001
p1	0.2599	0.0215	12.12	0.0000
m0	-0.8763	0.0346	-25.33	0.0000

Tabella 3.4: Stime dei MQO del modello di regressione lineare con solo esplicative necessarie nella fascia semi oraria 18.00–18.30 per il periodo 01/04/2005 al 31/12/2008.

non utili è consistente e risultano sempre poche le esplicative che servono effettivamente per spiegare p_t . Sono pressoché sempre presenti il prezzo a ritardo uno e il margine. Sparisce quasi totalmente l'effetto dei prezzi di ritardo superiore al primo. In alcune finestre temporali sembrano utili anche il margine ritardato e la domanda contemporanea.

Dati i modelli stimati per ciascuna fascia semi oraria si può proporre una stima ad un passo, *day-ahead*, della componente stocastica nel sotto campione *out of sample*. In generale, lo schema di previsione ad un passo è

$$\hat{p}_{t+1} = z'_{t+1} \hat{\beta} \quad \text{per } t = 1371, \dots, 2100$$

dove z_{t+1} rappresenta l'insieme di variabili esplicative strettamente necessarie per definire un modello ARX specifico per ciascuna fascia semi oraria preso al tempo $t + 1$ e $\hat{\beta}$ è il vettore di parametri associati a queste esplicative stimato con MQO. Nell'eventuale comparsa di m_{t+1} , d_{t+1} e dm_{t+1} , nel vettore z_{t+1} , queste sono note in quanto domanda e margine sono previsioni per il giorno successivo. Nel caso particolare delle fasce semi orarie delle 12.00–12.30 e 18.00–18.30 la previsione ad un passo è rispettivamente

$$\hat{p}_{t+1} = \hat{\beta}_0 + \hat{\beta}_1 p_t + \hat{\beta}_9 d_t + \hat{\beta}_{10} m_{t+1} + \hat{\beta}_{11} m_t$$

e

$$\hat{p}_{t+1} = \hat{\beta}_0 + \hat{\beta}_1 p_t + \hat{\beta}_{10} m_{t+1}.$$

Il modello di regressione lineare viene ristimato ad ogni passo del processo di previsione *out of sample*, ovvero i parametri del modello stimati non rimangono fissi alla stima nel campione *in sample* bensì progressivamente aggiornati ad ogni nuovo istante dove si procede alla stima del p_{t+1} e si aggiunge idealmente un

giorno di osservazioni. Mantenere le stime dei parametri fissate alle sole stime definite nel campione *in sample* non altera di molto le previsioni dei p_{t+1} , ma può cambiare il numero di picchi previsti e falsi allarmi anche in modo consistente.

Con \hat{p}_{t+1} non si definisce altro che una previsione della media condizionata per il giorno successivo, ovvero $\hat{p}_{t+1} = E(\widehat{p}_{t+1}|I_t) = x'_{t+1}\widehat{\beta}$ dove I_t rappresenta l'insieme informativo utilizzato per la previsione ad un passo. Dalle stime di p_{t+1} per ogni giornata del campione *out of sample* è possibile ottenere una stima dei prezzi come esposto in sezione 2.3. Dai \hat{p}_{t+1} vengono quindi stimati i logaritmi dei prezzi

$$\widehat{\log P}_{t+1} = \widehat{DP}_{t+1} + \hat{p}_{t+1} \quad \text{per } t = 1731, \dots, 2100$$

dove \widehat{DP}_{t+1} è una stima che al tempo t si fa per la componente deterministica della serie dei logaritmi dei prezzi per l'istante $t + 1$. Una stima dei prezzi un passo in avanti si ottiene con una trasformazione esponenziale sui logaritmi dei prezzi stimati un passo in avanti, ovvero

$$\hat{P}_{t+1} = e^{\widehat{\log P}_{t+1}} \quad \text{per } t = 1731, \dots, 2100.$$

Dalle stime dei prezzi per il giorno successivo è possibile definire una previsione di picco di prezzo. La previsione di un picco avviene quando il prezzo previsto \hat{P}_{t+1} , *day-ahead* supera la soglia che lo definisce che è stata stabilita in equazione (2.15) (sezione 2.3). Viene successivamente valutata la concordanza delle posizioni di picchi previsti con quelli che si realizzano effettivamente, conteggiando le previsioni di picco corrette e i falsi allarmi, ovvero una previsione di picco che non avviene.

In tabella 3.5 e 3.6 sono riportate le performance in termini assoluti che si ottengono dalla previsioni dei picchi di prezzo nelle due fasce semi orarie 12.00–12.30 e 18.00–18.30. Nelle diagonali si evidenziano le interpretazioni corrette ottenute dal modello (previsione di picco–picco realizzatosi, picco non previsto–picco non realizzatosi) mentre al di fuori della diagonale si evidenziano le previsioni non corrette (previsione di un picco–picco non realizzatosi: falsi allarmi, picco non previsto–picco realizzatosi). Si può osservare come nelle 730 previsioni

che definiscono la lunghezza del campione *out of sample*, in cui si fanno previsioni ad un passo, il numero di picchi individuati non sia uguale in ciascuna fascia semi oraria, vi sono 6 picchi definiti per la semi ora delle 12.00-12.30 e 10 picchi per la semi ora delle 18; in alcune fasce semi orarie si raggiungono anche 12 picchi in altre al massimo 3 (si veda tabella B.3 – appendice B – dove viene riportato il numero di picchi *out of sample* per ciascuna fascia semi oraria). Nel caso di questi due periodi il modello realizza previsioni corrette 724 e 720 volte su 730 previsioni complessivamente, ma è interessante capire come le previsioni corrette (e non) si dividano in picchi e non correttamente previsti. Il modello lineare ridotto per ciascuna fascia semi oraria (delle 12 e delle 18) produce 2 e 4 segnali di picco di cui solo però 1 e 2 saranno picchi realizzati, gli altri sono falsi allarmi. Per la sola analisi di queste due fasce semi orarie la capacità predittiva di un modello di regressione lineare sembra alquanto scarsa. In figura 3.1 e 3.2 si fornisce un'in-

	Picchi	Non Picchi	
Previsto	1	1	2
Non Previsto	5	723	728
	6	724	730

Tabella 3.5: Frequenze di previsioni dei picchi di prezzo in riferimento a picchi veri e non nella fascia semi oraria 12.00–12.30 per il periodo *out of sample* 01/01/2009 al 31/12/2010.

	Picchi	Non Picchi	
Previsto	2	2	4
Non Previsto	8	718	726
	10	720	730

Tabella 3.6: Frequenze di previsioni dei picchi di prezzo in riferimento a picchi veri e non nella fascia semi oraria 18.00–18.30 per il periodo *out of sample* 01/01/2009 al 31/12/2010.

terpretazione grafica delle previsioni dei prezzi e dei picchi nel campione *out of sample*. I prezzi previsti *day-ahead* (linea rossa) sono sovrapposti ai prezzi che si realizzeranno (linea nera). La previsioni di picco è visibile da sforamenti della soglia (linea blu) da parte dei prezzi previsti.

Diventa interessante osservare le *performance* di questi modelli lineari nella previsione dei picchi in termini percentuali relativizzati al numero dei picchi e

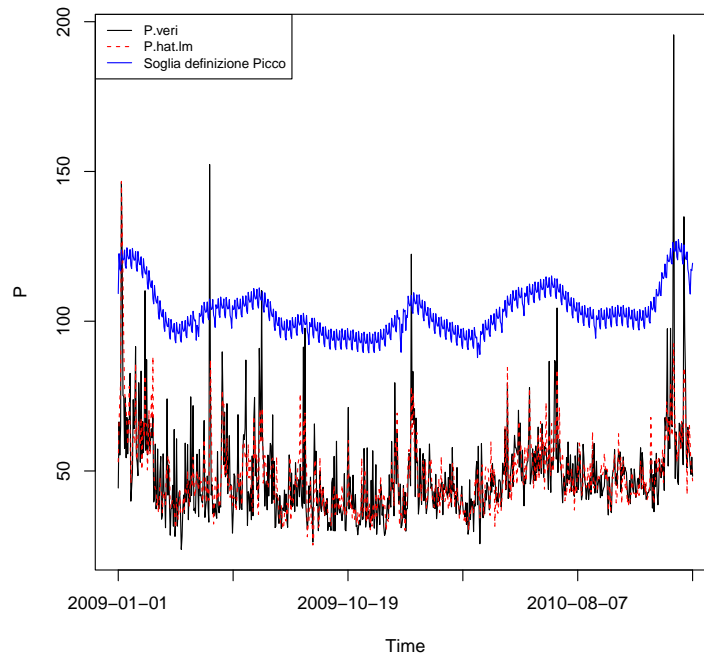


Figura 3.1: Prezzi previsti (in rosso) con relativo modello di regressione lineare, sovrapposti ai prezzi realizzati nel campione *out of sample* (in nero) e la soglia che definisce un picco di prezzo e la sua previsione nella fascia semi oraria 12.00–12.30.

non picchi che effettivamente si realizzano. In tabella 3.7 e 3.8 si possono osser-

	Picco	Non Picco
Previsti Correttamente %	16.67	99.86
Previsti non Correttamente %	83.33	0.14

Tabella 3.7: *Performance* della previsione relativizzate al vero numero di picchi e non picchi nella fascia semi oraria 12.00–12.30 per il periodo *out of sample* 01/01/2009 al 31/12/2010.

	Picchi	Non Picchi
Previsti Correttamente %	20	99.72
Previsti non Correttamente %	80	0.28

Tabella 3.8: *Performance* della previsione relativizzate al vero numero di picchi e non picchi nella fascia semi oraria 18.00–18.30 per il periodo *out of sample* 01/01/2009 al 31/12/2010.

vare la percentuale di previsioni corrette rispettivamente sia di un picco che di un non picco per le due fasce semi orarie. Una previsione corretta di un picco è nettamente più importante di una previsione di un non picco dato lo scopo della nostra analisi, un falso allarme ha quindi un peso più contenuto di un picco non

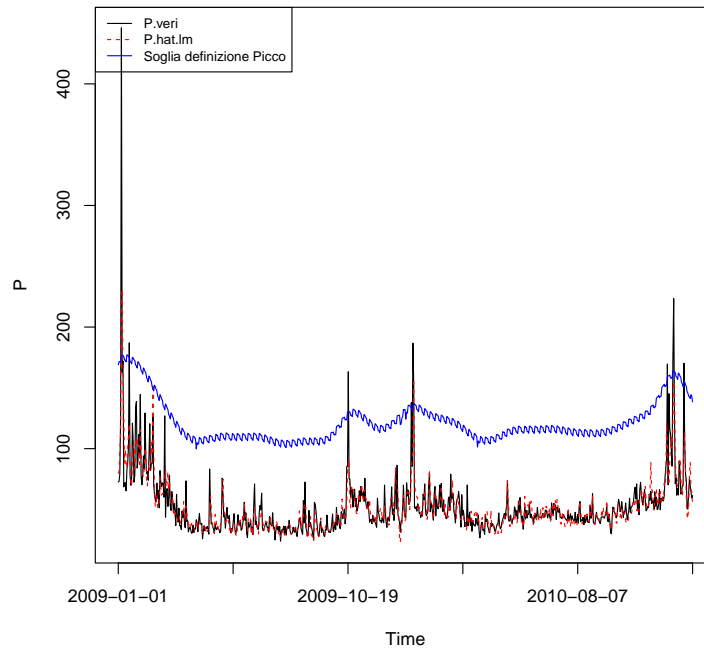


Figura 3.2: Prezzi previsti (in rosso) con relativo modello di regressione lineare, sovrapposti ai prezzi realizzati nel campione *out of sample* (in nero) e la soglia che definisce un picco di prezzo e la sua previsione nella fascia semi oraria 18.00–18.30.

previsto ma è comunque un indice importante per il buon funzionamento del modello. Emerge che sembra difficile prevedere correttamente i picchi attraverso questo modello in queste due fasce semi orarie.

Non si possono però giudicare le *performance* predittive di un modello sulla base di due sole fasce semi orarie sulle 48 presenti, si riportano allora le *performance* predittive globali del modello nella previsione per tutti i picchi identificati *out of sample*. La procedura di previsione del picco all'interno di ogni periodo è *standard* come per le fasce semi orarie delle 12 e 18 e definendo tabelle come la 3.7 e 3.8 si può definire una tabella unica globale che non separi più le fasce semi orarie in termini di risultati ma fornisce un'analisi complessiva.

In tabella 3.9 si riportano le frequenze assolute di previsioni di picco e non che si realizzano correttamente oppure no. Si può osservare che si effettuano 35040 previsioni ad un passo, derivante da 780 previsioni *day-ahead* del campione *out of sample* per ciascuna fascia semi oraria. I picchi che si realizzano effettivamente sul

	Picchi	Non Picchi	
Previsto	57	59	116 (0.33%)
Non Previsto	254	34670	34924 (99.67%)
	311 (0.89%)	34729 (99.11%)	35040

Tabella 3.9: *Performance* predittive globali dei picchi di prezzo *out of sample*.

totale sono 311 (poco meno dell'1%) mentre i non picchi sono 34729 (circa 99%) dato il criterio con cui sono stati stabiliti. I picchi effettivamente previsti sono 57

	Picco	Non Picco
Previsto come picco	18.33 %	0.17 %
Non Previsto come picco	81.67 %	99.83%

Tabella 3.10: *Performance* percentuali della previsione relativizzate al vero numero di picchi e non picchi.

dei 311 realizzatesi mentre i falsi allarmi sono 59 da confrontarsi con il totale di previsioni di picco effettuate (116); per valutare meglio tali risultati si osservi la tabella 3.10 che riporta le *performance* percentuali relativizzate al totale di picchi e non picchi. Si nota che il modello di regressione lineare è in grado di prevedere solo il 18.33% dei picchi effettivamente realizzatesi, producendo false previsioni di picco per appena 0.17% dei casi. I non picchi vengono previsti correttamente il 99.83% dei casi. Emerge che il modello ha un basso margine di errore ma non raggiunge lo scopo di prevedere i picchi dei prezzi.

Impiegare il modello di regressione lineare classico per la previsione dei picchi fornisce risultati non soddisfacenti. La spiegazione a questo problema è semplice, i picchi di prezzo sono per loro natura e definizione valori estremi, nettamente al di sopra del livello medio. Un picco di prezzo è tutt'altro che vicino alla media, idealmente possiamo associarlo ai quantili di livello elevato della distribuzione dei prezzi. Attraverso il modello di regressione la stima della componente stocastica *day-ahead* è soltanto una stima della media condizionata, ovvero

$$\hat{p}_{t+1} = E(\widehat{p}_{t+1} | I_t) = x'_{t+1} \hat{\beta} \quad \text{per } t = 1371, \dots, 2100$$

dalla quale si derivano i prezzi *day-ahead* (\widehat{P}_{t+1}) che a loro volta sono intuitivamente associabili ad una previsione media.

L'idea che può essere seguita per prevedere i picchi è quella di utilizzare previsioni di quantili condizionati dei prezzi *day-ahead* (non più una previsione media). Si prevedono i quantili condizionati di livello τ della serie dei p_t per il giorno successivo, da questi si derivano previsioni di quantili condizionati della serie dei prezzi *day-ahead*. Viene previsto un picco se il quantile dei prezzi per il giorno successivo supera la soglia che definisce il picco.

Definendo con $Q_\tau(p_t|I_t)$ il quantile condizionato della componente stocastica di livello τ , questo è

$$Pr(p_t \leq Q_\tau(p_t|I_t)) = \tau \quad \text{per } \tau \in [0, 1]$$

L'individuazione del quantile di livello τ può avvenire sfruttando l'inversa della funzione di ripartizione; una sua stima non parametrica può passare dall'uso della funzione di ripartizione empirica. È possibile però attribuire una stima parametrica di $Q_\tau(p_t|I_t)$ attraverso il modello di regressione lineare, per un qualsivoglia quantile della distribuzione condizionata di p_t (e successivamente dei prezzi). Ponendo opportune ipotesi distributive sulla componente d'errore del modello di regressione (3.2) oppure utilizzando i quantili della distribuzione empirica dei residui del modello di regressione stimato se si vuole deviare dall'ipotesi distributiva di ε_t .

La stima del quantile che interessa non è quella al tempo t , bensì quella un passo in avanti e quindi si è interessati alla stima del quantile condizionato *day ahead*. Ipotizzando che la distribuzione dell'errore del modello di regressione sia normale, sotto le assunzioni fatte in sezione 3.1, si ha che la stima del quantile condizionato è

$$\widehat{Q}_\tau(p_t|I_t) = E(\widehat{p}_{t+1}|I_t) + z_\tau \widehat{\sigma}$$

dove z_τ è il quantile di livello τ della distribuzione normale *standard* ($\Phi(z_\tau) = \tau$) e $\widehat{\sigma}$ è la stima che con i minimi quadrati ordinari si dà della radice della varianza dell'errore del modello di regressione. Dalle previsioni del quantile condizionato della componente stocastica è possibile, sfruttando la componente deterministica stimata di DP_{t+1} e la trasformazione esponenziale, fornire una stima del quantile

di livello τ dei prezzi *day-ahead* come

$$\exp\{\widehat{DP}_{t+1} + Q_{\tau}(\widehat{p}_t|I_t)\} \quad (3.3)$$

È assumibile che per prevedere i picchi meglio di una semplice media condizionata, i quantili condizionati dei prezzi da prevedere siano quelli di livello elevato. Se τ è 0.5 si stanno replicando le previsioni fatte in precedenza cioè quelle basate sul livello medio (che in termini di quantili potrebbero essere intese come una mediana condizionata stimata dalla media), è necessario quindi imporre $\tau > 0.5$.

In tabella 3.11 e B.1 (Appendice B) sono riportate le *performance* di previsione dei picchi per le fasce semi orarie delle 12.00–12.30 e 18.00–18.30, utilizzando i quantili condizionati dei prezzi derivanti dai modelli *ARX* con sole esplicative strettamente necessarie.

Tau	PR.PIC	PR.NPIC	NPR.PIC	NPR.NPIC	PicchiPrevisti %	FalsiAllarmi %
0.5	1	1	5	723	16.67	50.00
0.75	1	2	5	722	16.67	66.67
0.8	1	2	5	722	16.67	66.67
0.85	1	4	5	720	16.67	80.00
0.88	1	5	5	719	16.67	83.33
0.9	1	7	5	717	16.67	87.50
0.93	2	10	4	714	33.33	83.33
0.95	3	10	3	714	50.00	76.92
0.975	4	22	2	702	66.67	84.62

Tabella 3.11: Previsioni ad un passo modello ARX con quantili condizionati basati sulla distribuzione normale nella fascia semi oraria 12.00–12.30 per il periodo 01/01/2009 al 31/12/2010. "Tau": livello del quantile condizionato utilizzato per la previsione dei prezzi; "PR.PIC": numero di picchi previsti che si sono realizzati; "PR.NPIC": numero di picchi previsti non realizzati (falsi allarmi); "NPR.PIC": numero di picchi realizzati non previsti; "NPR.NPIC": numero di picchi non realizzati correttamente previsti; "Picchi Previsti %" percentuali di picchi previsti correttamente sul loro totale; "Falsi Allarmi %" percentuali di picchi previsti erroneamente sul totale di previsioni di picco.

Alle prime 4 colonne che riportano le frequenze di previsioni di picco e non divisibili nei casi di effettivo picco e non, si aggiunge una colonna relativa alla percentuale di picchi previsti e una colonna relativa alla percentuale di falsi allarmi. La riga con $\tau = 0.5$ è esattamente riportata in tabella 3.5 (previsione con media condizionata); si nota che se si prevede il 95 –esimo quantile condizionato vengono previsti correttamente 3 picchi su 6 (50% di picchi previsti), ma si

effettuano 13 previsioni di picco delle quali 10 sono falsi allarmi (83% di falsi allarmi). Va però sempre valutato il numero di previsioni di picco nel totale delle 730 previsioni.

I picchi di prezzo essendo in numero nettamente inferiore dei non picchi, ed essendo valori elevati, vengono via via predetti in numero maggiore per un τ sempre più vicino ad uno, ma in tal modo si aumentano anche i falsi allarmi e quindi previsioni non corrette in modo incontrollato. Emerge un importante legame tra previsioni corrette e falsi allarmi. Questo porta all'importante evidenza che per aumentare la capacità previsionale di un picco, si debba plausibilmente aumentare il livello del quantile condizionato dei prezzi da stimare, ma per converso si aumenti il numero di segnali di picco errati in modo elevato e pressoché esponenziale (anche per piccoli miglioramenti di picchi correttamente previsti).

Se si vuole utilizzare il quantile condizionato dei prezzi *day-ahead* per prevedere un picco, si deve risolvere il *trade-off* tra la capacità predittiva dei picchi e il numero di falsi allarmi. Il numero di picchi correttamente previsti e di falsi allarmi appaiono come caratteristiche che si muovono nella stessa direzione. La scelta di un opportuno modello e livello del quantile (τ) è rivolta a risolvere questo *trade-off*.

Ai fini della valutazione della qualità dei quantili previsti si può valutare la frazione di prezzi reali che oltrepassano la serie dei quantili previsti, calcolabile come $\sum_{t=1731}^{2100} \frac{I[P_{t+1} > Q_{\tau}(\widehat{p}_t | I_t)]}{730}$ nel campione *out of sample*. Se la previsione del quantile condizionato dei prezzi di livello τ è di buon livello questa dovrebbe lasciare alla propria destra una frazione di osservazioni pari a $1 - \tau$. In tabella 3.2 viene riportato tale indice per le due fasce semi orarie e si nota che la proprietà non viene garantita tanto più il livello del quantile è elevato, e così anche in generale per le rimanenti 46 fasce semi orarie. Appare da queste valutazioni che il quantile proposto sia apparentemente una sovrastima sistematica. In figura 3.3 e 3.4 vengono riportate le serie dei prezzi veri *out of sample* (in nero), la soglia che definisce un picco di prezzo (in blu) e la stima del quantile di livello 95% (in rosso) per le fasce semi orarie 12.00–12.30 e 18.00–18.30. In figura 3.1 e 3.2 si evidenzia come

Tau	0.500	0.750	0.800	0.850	0.880	0.900	0.930	0.950	0.975
Sforamenti 12	0.484	0.186	0.134	0.097	0.084	0.064	0.042	0.025	0.016
Sforamenti 18	0.481	0.158	0.114	0.081	0.059	0.049	0.038	0.030	0.018

Tabella 3.12: Frazione di sforamenti del quantile condizionato *day-ahead* previsto per vari livelli di τ ($\widehat{Q}_{P_{t+1}(\tau|I_t)}$) della serie dei prezzi reali *out of sample* per la fascia semi oraria delle 12.00-12.30 e 18.00-18.30.

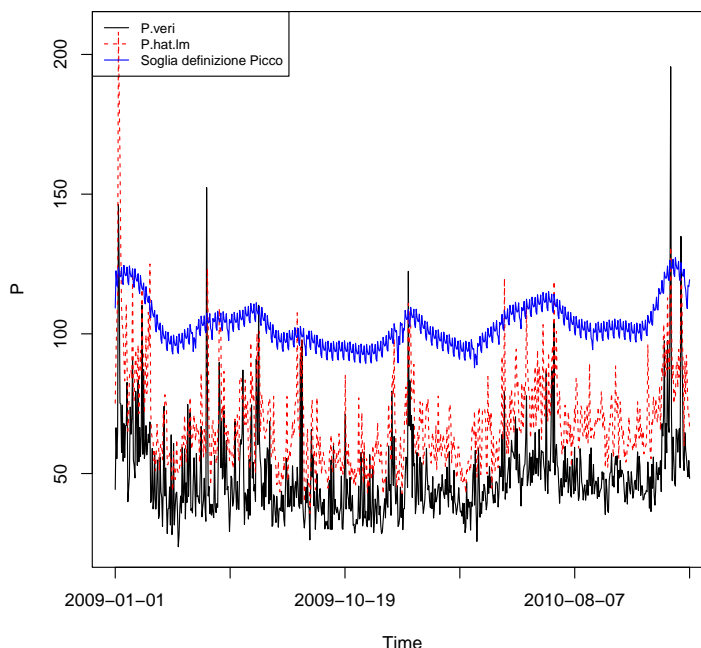


Figura 3.3: Quantile condizionato della distribuzione dei prezzi previsto di livello 95% (in rosso) con relativo modello di regressione lineare sovrapposti ai prezzi realizzati nei campioni *out of sample* (in nero) e la soglia che definisce un picco di prezzo e la sua previsione nella fascia semi oraria 12.00–12.30 per il periodo *out of sample* 01/01/2009 al 31/12/2010.

le previsioni dei quantili condizionati dei prezzi siano nettamente più elevate di un livello medio e portino a maggiori sforamenti della soglia.

Un esempio di lettura dei risultati delle tabelle 3.11 può essere il seguente. Ad un livello del quantile previsto del 97.5% vengono anticipati correttamente 4 picchi dei 6 totali, i segnali di picco non realizzati sono 22 su 26. Il modello nel complesso quindi prevede correttamente il 96.71% dei casi, indiscriminatamente tra picchi di prezzo e non. La soglia definita dalla serie del quantile condizionato previsto viene sfiorata dalla serie dei prezzi solo per un 0.016 che non approssima

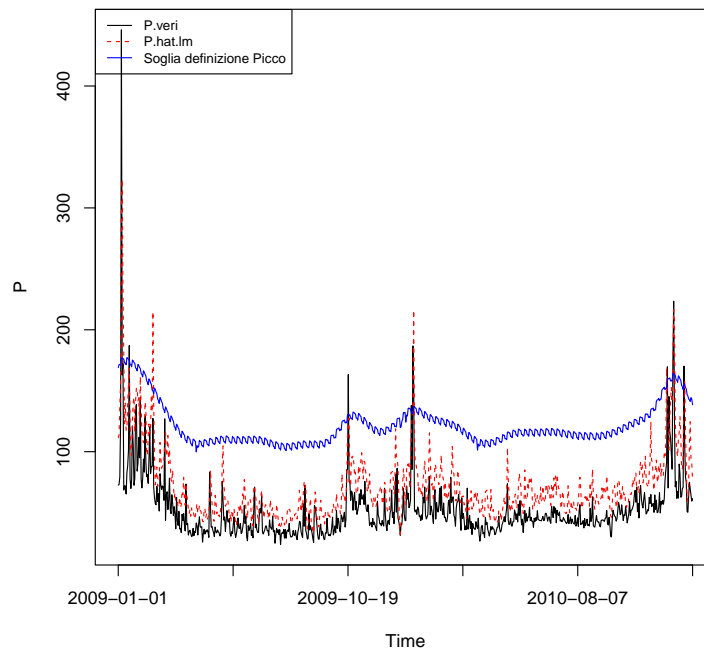


Figura 3.4: Quantile condizionato della distribuzione dei prezzi previsto di livello 95% (in rosso) con relativo modello di regressione lineare sovrapposti ai prezzi realizzati nei campioni *out of sample* (in nero) e la soglia che definisce un picco di prezzo e la sua previsione nella fascia semi oraria 18.00–18.30 per il periodo *out of sample* 01/01/2009 al 31/12/2010.

troppo bene $1 - \tau = 0.025$. Per prevedere gli ulteriori due picchi non previsti bisogna spingere τ a livelli superiori di 0.975. Il numero di falsi allarmi in relazione alle previsioni di picco è molto elevato.

È possibile quindi attribuire una previsione dei picchi complessivi nel campione *out of sample* utilizzando il quantile condizionato *day-ahead* previsto parametricamente sotto l'ipotesi di normalità della componente erratica per ciascuna fascia semi oraria; utilizzando il modello di regressione lineare con sole esplicative necessarie in ciascuna semi ora e replicando la stessa procedura effettuata sulle fasce semi orarie delle 12 e 18. Le *performance* della previsione globale è riportata in tabella 3.13 dove vengono riportate anche le percentuali dei picchi correttamente previsti e le percentuali di picchi previsti erroneamente, falsi allarmi.

Preso a riferimento il quantile condizionato di livello 97.5% previsto ad un passo ($\tau = 0.975$) si osserva come i picchi correttamente previsti siano 206 su 311,

Tau	PR.PIC	PR.NPIC	NPR.PC	NPR.NPC	PicchiPrevisti %	FalsiAllarmi %
0.5	57	59	254	34670	18.33	50.86
0.75	81	187	230	34542	26.05	69.78
0.8	89	245	222	34484	28.62	73.35
0.85	107	358	204	34371	34.41	76.99
0.88	121	477	190	34252	38.91	79.77
0.9	129	578	182	34151	41.48	81.75
0.93	147	875	164	33854	47.27	85.62
0.95	163	1261	148	33468	52.41	88.55
0.975	206	2287	105	32442	66.24	91.74

Tabella 3.13: *Performance* globale della previsione per l'intero insieme dei picchi *out of sample*, basate sulle previsioni di quantili condizionati di livello τ sotto ipotesi di normalità dell'errore del modello di regressione.

che produce un netto miglioramento rispetto al modello di regressione classico per la media che ne prevedeva soli 57 su 311. Ciò significa che i picchi previsti correttamente passano a circa il 66% rispetto al 18% nel caso della media condizionata, ma vi è anche un netto aumento di falsi allarmi, che diventano il 91% contro un 50% del modello di regressione classico. Va però evidenziato che il numero totale di previsioni di un picco (indistintamente se corrette o meno) sono 116 su 35040 utilizzando la media condizionata ($\tau = 0.5$), ma 2493 sul totale di 35040 previsioni usando il quantile di livello 97.5%. Si osserva che, globalmente, all'aumentare del livello del quantile condizionato dei prezzi *day-ahead* previsti aumenti la capacità predittiva dei picchi ma anche i falsi allarmi, evidenziando il *trade-off* già esposto. È molto importante prevedere quanti più picchi si realizzeranno, ma bilanciare opportunamente le corrette previsioni con i falsi allarmi è d'obbligo per creare un modello realmente applicabile e che non sia banale.

È assumibile una possibile deviazione dall'ipotesi sulla normalità dell'errore del modello di regressione lineare. Una caratteristica dei prezzi (e logaritmi) è la presenza di asimmetria positiva che si trasferisce alle stesse serie che vengono modellate nei modelli di regressione lineare. In genere è assai difficile ipotizzare una distribuzione adatta per l'errore del modello di regressione di facile gestione, se non una distribuzione iperbolica (Bunn, 2006). È quindi possibile prevedere il quantile condizionato di livello τ ricorrendo ai quantili della distribuzione marginale dei residui della regressione lineare. Presi i residui del modello di re-

gressione per ciascuna fascia semi oraria, $r_t = p_t - \hat{p}_t$, la stima del quantile della distribuzione marginale dei residui di livello $\tau \in [0, 1]$ può essere definito come il valore $\widehat{Q}_r(\tau)$ tale che $\sum_{t=1}^{1371} \frac{I[r_t \leq \widehat{Q}_r(\tau)]}{1371} = \tau$ dove $I[r_t \leq \widehat{Q}_r(\tau)]$ è una variabile indicatrice che assume valore 1 se $r_t \leq \widehat{Q}_r(\tau)$, 0 altrimenti. Una stima del quantile condizionato un passo in avanti di p_t è quindi

$$\widehat{Q}_{p_{t+1}}(\tau|I_t) = E(\widehat{p}_{t+1}|I_t) + \widehat{Q}_r(\tau),$$

dalla quale è possibile ottenere una stima del quantile condizionato dei prezzi *day-ahead* come in (3.3). È possibile quindi attribuire una previsione dei picchi complessivi nel campione *out of sample* utilizzando tale specificazione del quantile dei prezzi, si veda tabella 3.14. Si osserva che, rispetto al metodo basato sul-

Tau	PR.PIC	PR.NPIC	NPR.PC	NPR.NPC	PicchiPrevisti %	FalsiAllarmi %
0.5	44	36	267	34693	14.15	45.00
0.75	72	106	239	34623	23.15	59.55
0.8	80	155	231	34574	25.72	65.96
0.85	87	220	224	34509	27.97	71.66
0.88	96	297	215	34432	30.87	75.57
0.9	101	363	210	34366	32.48	78.23
0.93	124	598	187	34131	39.87	82.83
0.95	161	1049	150	33680	51.77	86.69
0.975	219	2700	92	32029	70.42	92.50

Tabella 3.14: *Performance* globale della previsione per l'intero insieme dei picchi di tutte le fasce semi orarie per il periodo *out of sample* 01/01/2009 al 31/12/2010, basate sulle previsioni di quantili condizionati di livello τ in modo semi parametrico.

l'imporre una distribuzione parametrica all'errore con riferimento a tabella 3.13, l'uso di quantili della distribuzione marginale dell'errore permette una riduzione dei falsi allarmi abbastanza evidente mantenendo un numero pressoché inalterato di corrette previsioni fino a che si utilizza un quantile di livello non superiore al 95-esimo. Utilizzando per le previsioni dei picchi un quantile di livello $\tau = 0.975$ si nota un aumento del capacità previsionale dei picchi ma con essa un aumento di falsi allarmi. In entrambi i metodi qui proposti appaiono chiare due questioni; la prima è la difficoltà di prevedere correttamente quantili di livello elevato dei prezzi, la seconda è riuscire a bilanciare il *trade-off* tra previsioni di picco corrette e falsi allarmi se si aumenta il livello del quantile per prevedere più picchi.

3.3 Limiti del modello di Regressione Lineare Classico

Il modello di regressione lineare classico utilizzato per la previsione dei picchi di prezzo, sia nella versione classica (attraverso una previsione della media condizionata) sia nella versione che sfrutta una previsione di quantili condizionati (sulla base di ipotesi distributive e non della serie dei prezzi), presenta svariati punti a sfavore che ne limitano la validità applicativa; alcune motivazioni sono qui elencate.

1. *I picchi di prezzo sono valori estremi.* Essendo i picchi di prezzo (sia da un punto di vista teorico che da un punto di vista identificativo) salti imprevedibili nella serie dei prezzi che si presentano come valori estremi, appare concettualmente improbabile che un qualsiasi modello di regressione lineare per la stima della media condizionata sia una buona proposta per prevedere i picchi di prezzo. Una previsione della media condizionata difficilmente si avvicina al concetto di picco di prezzo e difficilmente può essere utile per individuare un numero elevato in previsione. Utilizzare previsioni di quantili condizionati può essere un possibile metodo nonché l'idea che in questa tesi cerchiamo di verificare.
2. *La proprietà di Invarianza .* Un problema dell'approccio del modello lineare è legato allo stimatore del prezzo, sia come concetto di media condizionata che di quantile condizionato. Questo non gode della proprietà di invarianza per trasformazioni monotone non affini, si veda Azzalini(2001) oppure Pace e Salvan (2001). Ipotizzando che la componente deterministica dei log-prezzi (DP_t) sia indipendente dalle esplicative e che il modello di regressione su p_t (componente detrendizzata dai log-prezzi) può produrre stime un passo in avanti dei logaritmi dei prezzi come

$$\widehat{\log P_{t+1}} = E(\log \widehat{P_{t+1}} | I_t) = \widehat{DP}_{t+1} + \hat{p}_{t+1} = \widehat{DP}_{t+1} + E(\widehat{p_{t+1}} | I_t)$$

per $t = 1371, \dots, 2100$. Questa è una stima della media condizionata del logaritmo dei prezzi *day-ahead*. La stima dei prezzi per il giorno successi-

vo viene definita come \hat{P}_{t+1} , deriva dalla trasformazione esponenziale delle stime dei $\widehat{\log P}_{t+1}$ che come detto è una stima del valore atteso condizionato. La funzione esponenziale, non essendo una trasformazione lineare affine bensì, una funzione convessa non permette di fornire una stima della media condizionata dei prezzi data la diseguaglianza di *Jensen* (Azzalini 2001). Lo stimatore utilizzato in questo caso per i prezzi *day-ahead* è $\hat{P}_{t+1} = \exp\{E(\widehat{\log P}_{t+1}|I_t)\} < E(P_{t+1}|I_t)$ che crea una sottostima sistematica del prezzo medio condizionato. Questo comporta che la stima ad un passo dei prezzi non sia una stima della media condizionata, bensì sistematicamente qualcosa di inferiore; questo concetto si applica anche alla definizione dei quantili condizionati; non si prevede quindi correttamente il quantile condizionato dei prezzi.

3. *La previsione di quantili condizionati può essere un metodo efficace per la previsione dei picchi di prezzo.* Utilizzare un semplice modello di regressione lineare per derivare da questo quantili è tutt'altro che scontato. Ritenuta valida l'ipotesi di prevedere i picchi di prezzo con una previsione dei quantili condizionati, ci si può attendere che l'effetto marginale e la significatività delle esplicative nello spiegare quantili condizionati di livello τ possa cambiare con lo stesso variare del livello del quantile condizionato. La previsione di quantili condizionati viene realizzata con altri modelli di regressione differenti da quello lineare dato che questo non riesce a tenere conto di queste possibilità.
4. *L'impossibilità di accettare le ipotesi classiche del modello di regressione lineare.* Non si è presentata una diagnostica dei modelli di regressione lineari basati sullo studio dei residui del modello di regressione, su cui si dovrebbero riscontrare le ipotesi classiche del modello. Fondamentalmente se non sussistono le ipotesi di omoschedasticità e incorrelazione seriale i risultati asintotici su errori *standard* e proprietà distributive degli stimatori dei parametri cessano di esistere. L'assenza di normalità motiva solo l'incapacità di definire la distribuzione dell'errore del modello come una normale, così la stima

dei quantili condizionati basati su errori normali è difficilmente sostenibile per le caratteristiche dell'errore quale ad esempio l'asimmetria o una distribuzione che devia particolarmente dalla normalità. L'analisi dei residui è necessaria per avere una ulteriore conferma della validità del modello e per testare se sono soddisfatte le assunzioni richieste da questo tipo di analisi: distribuzione normale degli errori, ovvero dei residui, l'indipendenza degli errori dai valori della variabile indipendente, l'omogeneità delle varianze d'errore e l'assenza di valori di leva, che in questo caso è una valutazione controversa per il tipo di dati trattati e la presenza dei picchi di prezzo.

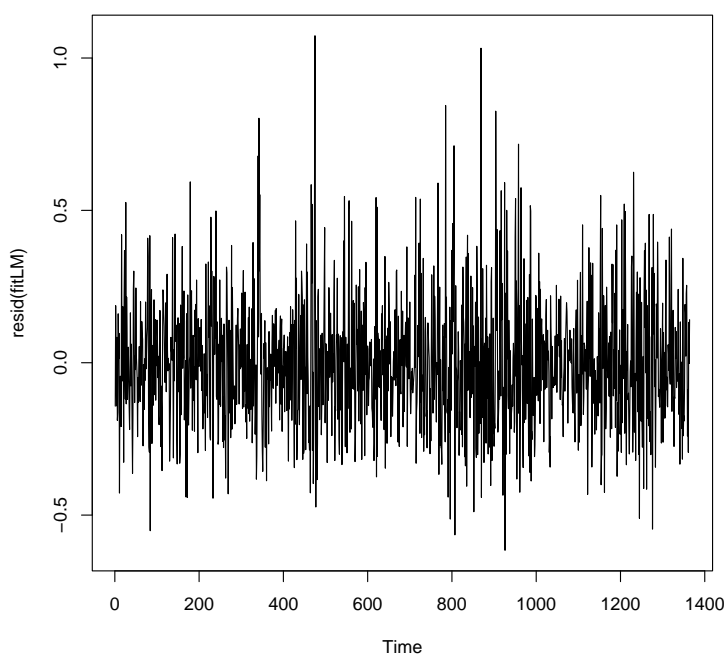


Figura 3.5: Serie dei residui del modello lineare ridotto stimato nella fascia semi oraria 12.00–12.30 per il periodo 01/04/2005 al 31/12/2008.

In figura 3.5 viene riportata la serie dei residui del modello lineare stimato *in sample* per la fascia semi oraria delle 12, si nota una eteroschedasticità connotata dalla presenza di *cluster* e valori anomali particolarmente elevati positivi. A supporto della mancata ipotesi di normalità della componente stocastica del modello vi è la marcata asimmetria positiva nella distribuzione non condizionata dei residui riportata in figura 3.8. In figura 3.6 viene

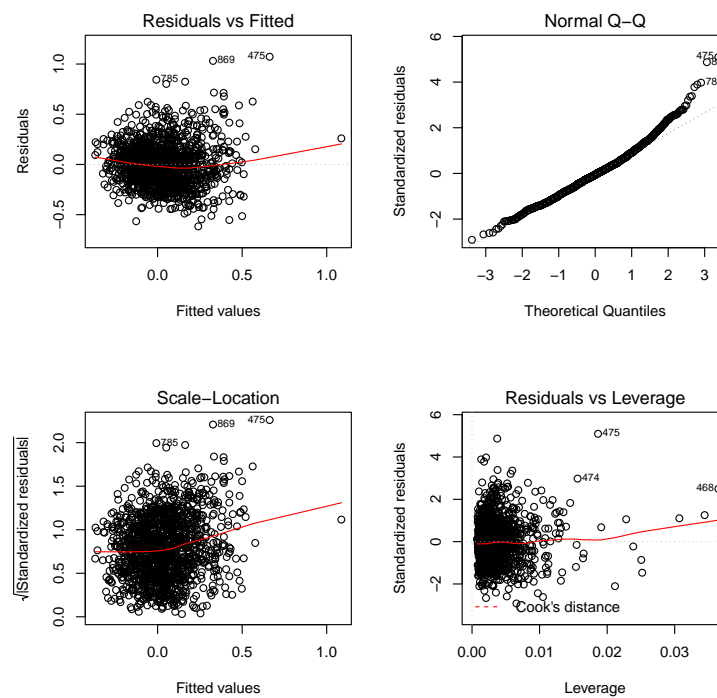


Figura 3.6: Diagnostica dei residui modello lineare ridotto stimato nella fascia semi oraria 12.00–12.30 per il periodo 01/04/2005 al 31/12/2008. In alto a sinistra grafico dei residui contro i valori teorici: può rivelare la presenza di una residua dipendenza sistematica non individuata dal modello lineare stimato. In un buon modello questo grafico dovrebbe apparire come completamente casuale; In alto a destra *normal Q-Q plot* dei residui standardizzati: verifica grafica dell'assunzione della normalità della componente stocastica del modello lineare. Quanto più i punti che rappresentano i residui ordinati giacciono in prossimità della linea Q-Q, tanto più plausibile è detta assunzione; In basso a sinistra grafico delle radici quadrate dei residui standardizzati contro i valori teorici: è utile nell'individuazione di valori outlier e per visualizzare strutture di dipendenza residue non individuate dal modello stimato; In basso a destra si mettono in evidenza i punti influenti per la regressione con le distanze di *Cook* e punti leva.

riportata una diagnostica *standard* dei residui che evidenzia l'eteroschetasticità e deviazione dall'ipotesi di normalità. In figura 3.7 si osserva come la serie dei residui non sia interpretabile come un *white noise*, ma che in realtà vi sia la presenza di una struttura tra i residui che non è stata completamente colta dal modello.

Queste valutazioni sulla componente dei residui definiscono di fatto l'impossibilità di giustificare da un punto di vista teorico i relativi risultati del modello di regressione lineari con ipotesi classiche.

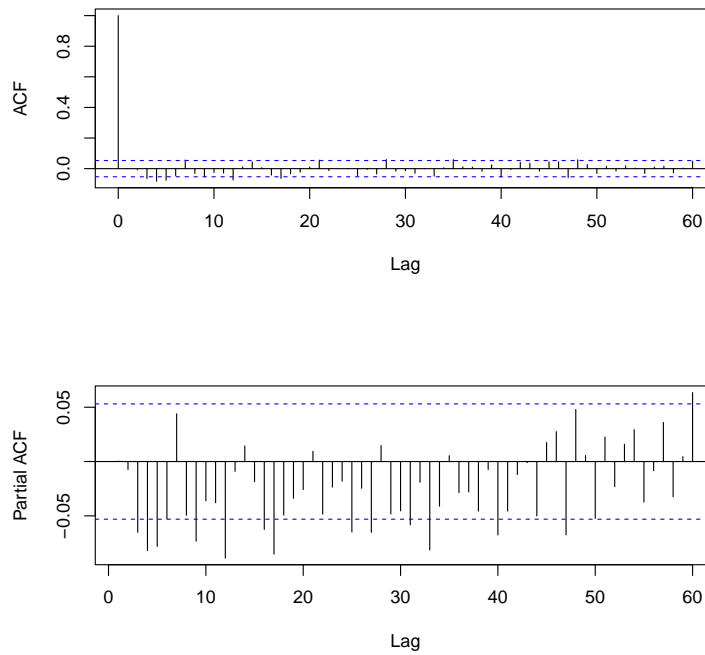


Figura 3.7: Valutazione dell'autocorrelazione e correlazione parziale nella serie dei residui del modello di regressione lineare ridotto stimato nella fascia semi oraria 12.00–12.30 per il periodo 01/04/2005 al 31/12/2008.

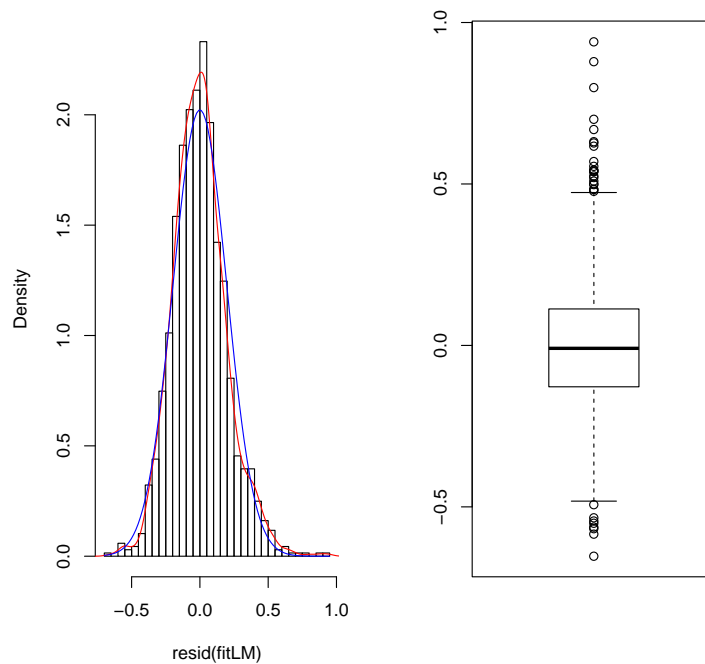


Figura 3.8: Distribuzione marginale dei residui del modello lineare ridotto stimato nella fascia semi oraria 12.00–12.30 per il periodo 01/04/2005 al 31/12/2008.

Capitolo 4

LA REGRESSIONE DEL QUANTILE

Un concetto che giustifica e motiva l'impiego della regressione del quantile semplice ma esauriente, e in generale di un'inferenza basata su quantili, può essere quello espresso da Mosteller e Tukey nel 1977 (Koenker, 2005):

"What the regression curve does is give a grand summary for the averages of the distributions corresponding to the set of x 's. We could go further and compute several different regression curves corresponding to the various percentage points of the distributions and thus get a more complete picture of the set. Ordinarily this is not done, and so regression often gives a rather incomplete picture. Just as the mean gives an incomplete picture of a single distribution, so the regression curve gives a correspondingly incomplete picture for a set of distributions."

La maggioranza degli studi empirici legati alla letteratura econometrica è incentrata sull'analisi della media (condizionata) della variabile d'interesse, tuttavia questo confinamento potrebbe fornire un quadro d'analisi incompleto e portare a conclusioni non accettabili non appena le ipotesi del modello di regressione lineare classico vengano a mancare.

Per superare questi problemi Koenker e Bassett (1978) introdussero un nuovo metodo di regressione, detto *quantile regression* (regressione del quantile), che permette di stimare interamente la distribuzione condizionata della risposta dato un insieme di esplicative (lineari). Vi è una vasta area di casi dove si vogliono sfruttare i quantili, si veda Takeuchi (2005). Il calcolo di quantili condizionati, un insieme di numeri, sostituisce il calcolo e la valutazione della media condizionata, un unico numero, e ciò permette di definire in modo più completo ed esausti-

vo le relazioni sottostanti. La regressione del quantile non è però solo utile per una descrizione più dettagliata di un fenomeno, ma gode di importanti proprietà aggiuntive rispetto la regressione lineare ad esempio l'equivarianza a trasformazioni monotone della variabile risposta, robustezza rispetto a valori anomali dei regressori ed efficienza (Koenker, 2005). L'aumento di elementi favorevoli comporta un maggior onere computazionale, di norma non è possibile attribuire una forma chiusa agli stimatori e così le stime vengono ottenute attraverso opportuni metodi numerici, che risolvendo problemi di programmazione lineari ne riescono a facilitare il calcolo. Successivamente all'introduzione del metodo della regressione del quantile è stata formulata una teoria asintotica generale che ha permesso di definire e sfruttare le classiche procedure inferenziali.

Nelle sezioni seguenti vengono presentati i principi basilari per introdurre la regressione del quantile, la sua formalizzazione, la stima e le proprietà più rilevanti; quindi alcuni cenni sulla teoria asintotica e inferenza. Infine verrà formalizzato il concetto della regressione del quantile non parametrica.

4.1 Quantili e strumenti

In questa sezione definiamo in modo basilare il concetto di funzione quantile, i quantili empirici in modo *standard* e alternativo. Si estende al concetto di regressione lo studio dei quantili definendo l'aspetto computazionale e i problemi principali. Il concetto basilare per parlare di inferenza sui quantili è quello di funzione quantile, sia X una variabile casuale (discreta o continua), per ogni $\tau \in (0, 1)$, il τ -esimo quantile di X può essere definito come $\gamma_\tau \in R$ tale che

$$Pr(X < \gamma_\tau) \leq \tau \leq Pr(X \leq \gamma_\tau) \quad (4.1)$$

dove se X è una variabile casuale continua le probabilità di equazione (4.1) coincidono e il quantile è definito con unicità, mentre se X è discreta le probabilità di equazione (4.1) definiscono un intervallo chiuso e ciò comporta che il quantile non sia unico. Per identificare con unicità il quantile si stabilisce per convenzione che questo è sempre il più piccolo elemento nell'insieme delle soluzioni possibili, tale definizione è formalizzabile con la funzione di ripartizione. Sia F_X la

funzione di ripartizione (continua a destra) di X tale che $F_X(x) = Pr(X \leq x)$, la funzione del quantile condizionato (continua a sinistra) è

$$Q_X(\tau) = F_X^{-1}(\tau) = \inf\{x | F_X(x) \geq \tau\} \quad 0 < \tau < 1 \quad (4.2)$$

che fornisce il quantile τ -esimo incondizionato di X , scegliendo il valore più piccolo nell'insieme dei valori possibili rende continua a sinistra $Q_X(\tau)$. Tra le proprietà della funzione quantile si ha che questa è monotona crescente e per una qualunque funzione h monotona crescente continua a sinistra (Koenker, 2005 e Peracchi, 1995) si ha che

$$Pr(X \leq Q_X(\tau)) = Pr(h(X) \leq h(Q_X(\tau))) = \tau. \quad (4.3)$$

Se X è una variabile casuale continua la funzione di densità di probabilità è ottenibile derivando la funzione di ripartizione ($f_X(x) = \frac{dF_X(x)}{dx}$), così l'analogo sulla funzione quantile è

$$s_X(\tau) = \frac{dQ_X(\tau)}{d\tau} = \frac{dF_X^{-1}(\tau)}{d\tau} = \frac{1}{f_X(Q_X(\tau))} \quad (4.4)$$

dove $s_X(\tau)$ è chiamata *sparsity function*, che può essere definita anche come l'inversa della funzione di densità di probabilità nel quantile e ha un ruolo importante nella regressione del quantile.

A livello campionario si possono definire i quantili empirici non condizionati. Dato un campione causale X_1, X_2, \dots, X_n , la funzione di ripartizione empirica \hat{F} è una stima non parametrica della funzione di ripartizione, così definita

$$\hat{F}_X(x) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x]}(x_i) \quad , x \in R$$

dove $I_{(-\infty, x]}(x_i)$ è la funzione indicatrice che vale uno qualora $x_i \leq x$ e zero altrimenti. È immediato verificare che $\hat{F}_X(x)$ si distribuisce come una variabile casuale Binomiale di dimensione n e parametro $F_X(x)$, in quanto $\hat{F}_X(x)$ è la somma di n variabile aleatorie indipendenti Bernoulliane, $I_{(-\infty, x]}(x_i)$, con parametro $F(x)$. Le principali proprietà di $\hat{F}_X(x)$ sono quindi:

1. $E(\hat{F}_X(x)) = F_X(x)$, ovvero è uno stimatore non distorto;

2. $Var(\hat{F}_X(x)) = \frac{1}{n}F_X(x)(1 - F_X(x))$;
3. $\hat{F}_X(x)$ è uno stimatore consistente di $F_X(x)$;
4. $\sqrt{n}[\hat{F}_X(x) - F_X(x)] \xrightarrow{d} N(0, F_X(x)(1 - F_X(x)))$ converge in distribuzione ad una variabile casuale normale.

Queste proprietà mostrano come al crescere della numerosità campionaria, la funzione di distribuzione empirica diventa sempre più simile a quella di popolazione F (Peracchi, 1995). Per analogia all'equazione (4.2), la funzione del quantile empirica è

$$\hat{Q}_X(\tau) = \hat{F}_X^{-1}(\tau) = \inf\{x | \hat{F}_X(x) \geq \tau\}, \quad 0 < \tau < 1,$$

dove ogni quantile empirico si deriva ordinando il campione e in modo tale che questo raggiunga la soglia. Sfruttando la $\hat{F}(x)$ per la stima del quantile si garantiscono buone proprietà. Alternativa a questo metodo è stata proposta da Koenker e Bassett (1978), dove il quantile viene definito risolvendo un problema di ottimizzazione, ovvero minimizzare la seguente funzione di perdita

$$\hat{Q}_X(\tau) = \arg \min_{\xi_\tau \in R} \left(\sum_{i \in \{i | X_i \geq \xi_\tau\}} \tau |X_i - \xi_\tau| + \sum_{i \in \{i | X_i < \xi_\tau\}} (1 - \tau) |X_i - \xi_\tau| \right), \quad (4.5)$$

dove le differenze assolute tra le osservazioni e il valore ottimale ignoto ξ_τ sono pesate per τ , il rimanente per $1 - \tau$ (Koenker, 2005). Se $\tau = 0.5$ si definisce come ottimo la mediana. L'equazione (4.5) può essere definita in forma compatta mediante l'introduzione della *check function*, così definita

$$\rho_\tau(u) = u(\tau - I(u < 0)) \quad 0 < \tau < 1$$

dove definiamo d'ora in poi la funzione indicatrice $I(A) = 1$ se A si realizza, zero altrimenti; a tal proposito la funzione del quantile condizionato può essere così definita

$$\hat{Q}_X(\tau) = \arg \min_{\xi_\tau \in R} \sum_i \rho_\tau(X_i - \xi_\tau) \quad (4.6)$$

che permette di ottenere i quantili non condizionati. Valutando la funzione di perdita di equazione (4.5) da minimizzare si può notare che non vi è una soluzione unica se $n\tau$ non è un numero intero (Koenker, 2005); in figura 4.1 ve ne

si riporta un esempio e si osserva che tale funzione obiettivo da minimizzare è convessa ma definita a tratti lineari con punti d'angoli in ciascuna osservazione X_i , ad ogni osservazione la pendenza della funzione obiettivo cambia. Per

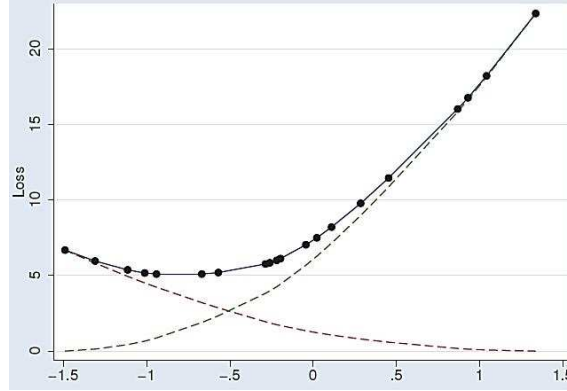


Figura 4.1: La linea continua rappresenta un esempio di una funzione di perdita come quella di equazione (4.5) per $\tau = 0.25$ al variare di ξ_τ . Le due linee tratteggiate mostrano le due somme ponderate di equazione (4.10) al variare del quantile scelto (Koenker, 2005).

dimostrare che minimizzare la funzione di perdita di equazione (4.5) permetta di individuare il quantile empirico di un qualsivoglia livello τ , passiamo al caso teorico dove X è una variabile casuale continua, da cui possiamo esprimere la funzione di perdita attesa come

$$E(\rho_\tau(X - \xi_\tau)) = \tau \int_{\xi_\tau}^{\infty} (x - \xi_\tau) dF_X(x) - (1 - \tau) \int_{-\infty}^{\xi_\tau} (x - \xi_\tau) dF_X(x)$$

per la condizione del primo ordine, la funzione di perdita viene minimizzata ponendo la derivata prima uguale a zero ($\frac{dE(\rho_\tau(X - \xi_\tau))}{d\xi_\tau} = 0$); dove la condizione del primo ordine è

$$\begin{aligned} \frac{dE(\rho_\tau(X - \xi_\tau))}{d\xi_\tau} &= \tau \frac{\int_{\xi_\tau}^{\infty} (x - \xi_\tau) dF_X(x)}{d\xi_\tau} - (1 - \tau) \frac{\int_{-\infty}^{\xi_\tau} (x - \xi_\tau) dF_X(x)}{d\xi_\tau} \\ &= -\tau \int_{\xi_\tau}^{\infty} dF_X(x) + (1 - \tau) \int_{-\infty}^{\xi_\tau} dF_X(x) \\ &= -\tau(1 - F(\xi_\tau)) + (1 - \tau)F(\xi_\tau) \\ &= F(\xi_\tau) - \tau = 0 \end{aligned} \tag{4.7}$$

Essendo la derivata seconda maggiore di zero, la funzione di perdita attesa è convessa e viene minimizzata se e solo se $F(\xi_\tau) = \tau$, ovvero che $\xi_\tau = F^{-1}(\tau)$, il quale risultato dimostra l'equazione (4.5).

4.2 La regressione del quantile

Come evidenziato nel capitolo 3 e nell'introduzione di questo capitolo, il modello di regressione lineare classico pone alcuni limiti nella spiegazione della variabile risposta in quanto definisce una valutazione solo della media condizionata, la regressione del quantile permette di dare un quadro più preciso sulla distribuzione condizionata attraverso la valutazione dei quantili condizionati. Assumiamo che il modello di regressione del quantile sia

$$Y_t = x_t' \beta_\tau + u_{t,\tau} \quad t = 1, 2, \dots, T \quad (4.8)$$

dove Y_t è la variabile risposta, x_t' un vettore di p variabili esplicative contenenti il termine costante, $u_{t,\tau}$ è il termine d'errore. L'assunzione fondamentale del modello di regressione è che, non tanto la media condizionata (come nel modello di regressione lineare) ma, il τ -esimo quantile condizionato del termine d'errore è zero $Q_\tau(u_{t,\tau}|x_t) = 0^1$ (ipotesi teorica molto forte). Data l'ipotesi sulla componente stocastica del modello di regressione di equazione (4.8) il τ -esimo quantile della variabile risposta Y_t subordinatamente alle x_t può essere scritto come

$$Q_\tau(Y_t|x_t) = x_t' \beta_\tau \quad (4.9)$$

dove si può subito osservare che il vettore dei parametri β_τ è variabile e legato al livello del quantile condizionato preso in considerazione. Ciò rende il modello di regressione del quantile un modello a parametri variabili. Mettendo insieme le equazioni (4.5), (4.6) e (4.9), per ogni $\tau \in (0, 1)$, il vettore dei parametri β_τ può essere stimato con

$$\begin{aligned} \hat{\beta}_\tau &= \arg \min_{\beta_\tau \in \mathbb{R}^p} \left\{ \sum_{i \in \{t|y_t \geq x_t' \beta_\tau\}} \tau |y_t - x_t' \beta_\tau| + \sum_{i \in \{t|y_t < x_t' \beta_\tau\}} (1 - \tau) |y_t - x_t' \beta_\tau| \right\} \\ &= \arg \min_{\beta_\tau \in \mathbb{R}^p} \sum_t \rho_\tau(y_t - x_t' \beta_\tau) \end{aligned} \quad (4.10)$$

dove tutte le osservazioni di al di sopra dell'iperpiano stimato da $X \hat{\beta}_\tau$ (precisamente le differenze in valore assoluto $y_t - x_t' \hat{\beta}_\tau$, dove $X = [x_1 \ x_2 \ \dots \ x_T]'$ è la matrice

¹Equivalentemente anche la notazione: $Q_{u_{t,\tau}}(\tau|x_t) = 0$ oppure $Q_{u_t(\tau)}(\tau|x_t) = 0$, che ha una similitudine con la condizione del modello lineare di regressione: $E(u_t|x_t) = 0$, si veda sezione 3.1

dei regressori) sono pesate per τ e le rimanenti per $1 - \tau$. Il caso speciale della regressione della mediana condizionata è un caso particolare della regressione del quantile nonché primo caso storico della regressione non basata su media condizionata (Koenker, 2005).

A dispetto delle stime dei minimi quadrati ordinari che prevedono una forma chiusa dello stimatore dei parametri, l'equazione (4.10) non permette una soluzione analitica esplicita in quanto la *check function* non è derivabile nell'origine. Tuttavia con una riscrittura del modello di regressione del quantile, il problema di ottimo di equazione (4.10) diventa un problema prettamente di programmazione lineare che può essere risolto numericamente. Barrodale e Roberts proposero un algoritmo del simplesso modificato efficiente opportuno per il problema della regressione della mediana, che fu generalizzato per un approccio per le stime numeriche per qualsivoglia quantile condizionato da Koenker e D'Orey, si veda Koenker (2005). L'elevato sforzo computazionale di questi algoritmi, che ne rendevano il principale inconveniente, venne ridotto da successive raffinazioni degli algoritmi e metodi di stima basati sempre però sull'ottimizzazione di una funzione obiettivo lineare e differenziabile.

Per ulteriori approfondimenti e delucidazioni sugli argomenti trattati si veda Koenker (2010), Buhai (2004), Cade e Noon (2003), Koenker e Xiao (2001) e Koenker e Hallock (2001).

4.2.1 Proprietà

Dopo l'introduzione basilare sulla regressione del quantile, introduciamo alcuni risultati e le proprietà più rilevanti, che si osservano e che la rendono particolarmente interessante anche rispetto il modello di regressione lineare classico.

Come già detto, il modello di regressione lineare del quantile, per qualsiasi valore di $\tau \in (0, 1)$, permette di stimare i quantili condizionati di y data una matrice di regressori X come

$$\widehat{Q}_\tau(y|X) = X\widehat{\beta}_\tau \quad \text{con} \quad \widehat{\beta}_\tau = \arg \min_{\beta_\tau \in \mathbb{R}^p} \sum_t \rho_\tau(y_t - x'_t \beta_\tau) \quad (4.11)$$

dove al variare di τ è possibile valutare l'intera distribuzione della variabile risposta condizionatamente alle esplicative in termini di quantili, fattore nettamente differente si una semplice valutazione della media condizionata. Implicitamente il metodo della regressione del quantile è in grado di rivelare forme di eteroschedasticità nei dati modellando quantili di livello differente, naturalmente per aumentare l'efficienza (come nel modello di regressione lineare) si può ricorrere ad una ponderazione delle osservazioni che tenga conto delle strutture di eteroschedasticità e correlazione. Tra le varie difficoltà di stima legate alla non continuità della funzione obiettivo di equazione (4.11), si osserva che vi possono essere errori anche grossolani sulle stime dei quantili condizionati legati all'ordinamento e alla scelta di valori iniziali da cui far partire l'algoritmo di ottimizzazione numerica per ottenere le stime. Questi problemi vengono opportunamente risolti con ripetizioni dell'algoritmo di stima e con opportuni criteri di ottimizzazione (Koenker, 2005).

È stato dimostrato che $\widehat{Q}_\tau(y_t|\bar{x})$ è sempre non decrescente in τ qualora i regressori vengano fissati alla rispettiva media.

Tra le altre proprietà a cui non si dà dimostrazione vi è quella dove, se ci sono k variabili esplicative il numero di residui nulli del modello del quantile, $y_t - \widehat{Q}_\tau(y_t|x_t)$, equivale esattamente a k .

Come per il modello di regressione lineare si può definire un indice di adattamento analogo al R^2 che identifica con una misura di adattamento ai dati del modello nonché bontà delle stime dipendente dal livello condizionato del quantile τ che si vuole stimare. Un pseudo R^2 è definibile come

$$R^2(\tau) = \frac{\min \sum_t \rho_\tau(x'_t \widehat{\beta}_\tau - Q_\tau(y))}{\min \sum_t \rho_\tau(y_t - Q_\tau(y))} = 1 - \frac{\sum_t (y_t - x'_t \widehat{\beta}_\tau)}{\min \sum_t \rho_\tau(y_t - Q_\tau(y))} \in [0, 1]$$

dove $Q_\tau(y)$ è il τ -esimo quantile della variabile y . Come R^2 , il valore di $R^2(\tau)$ è compreso tra 0 e 1. A differenza di R^2 che è una misura globale della bontà di adattamento, $R^2(\tau)$ è una misura relativa della bontà del modello utilizzato per stimare un certo livello del quantile, non ha aspetto globale (esattamente come le stime dei parametri $\widehat{\beta}_\tau$ che variano al variare del livello del quantile numericamente e in significatività).

Dopo questi brevi risultati marginali introduciamo le proprietà più importanti che furono già introdotte nell'articolo originale di Koenker e Bassett (1978) ovvero le proprietà di equivarianza dei coefficienti stimati con regressione del quantile:

$$\widehat{\beta}_\tau(\lambda y, X) = \lambda \widehat{\beta}_\tau(y, X) \quad \lambda \in [0, \infty) \quad (4.12)$$

$$\widehat{\beta}_\tau(-\lambda y, X) = \lambda \widehat{\beta}_{1-\tau}(y, X) \quad \lambda \in [0, \infty) \quad (4.13)$$

$$\widehat{\beta}_\tau(y + X\lambda, X) = \widehat{\beta}_\tau(y, X) + \lambda \quad \lambda \in \mathbb{R}^p \quad (4.14)$$

$$\widehat{\beta}_\tau(y, XA) = A^{-1} \widehat{\beta}_\tau(y, X) \quad A \text{ è una matrice invertibile} \quad (4.15)$$

dove le equazioni (4.13) e (4.14) definiscono equivarianze di scala, se la risposta y è riscalata per un fattore λ , allora il vettore dei parametri subisce medesima operazione di scala. La proprietà (4.15) è chiamata proprietà di equivarianza della regressione o di *shift*, infine la proprietà (4.15) è detta equivarianza alle riparametrazioni del disegno e significa che combinazioni lineari dei regressori si trasferiscono sul vettore dei parametri stimati con l'inversa della matrice di trasformazione.

Le proprietà (4.13) e (4.14) sono presenti anche nello stimatore dei minimi quadrati ordinari, tuttavia nella regressione del quantile vi è una proprietà ben più forte delle proprietà di equivarianza già discusse che non viene solitamente condivisa dagli altri tipi di regressione.

Preso una qualunque funzione a valori reali strettamente monotona h , segue data l'equazione (4.3) che

$$\widehat{Q}_\tau(h(y)|X) = h(\widehat{Q}_\tau(y|X)) \quad (4.16)$$

che implica che i quantili condizionali della risposta trasformata sono equivalenti ai quantili trasformati della variabile risposta. Solo se $h(\cdot)$ è affine, la media condizionata assume medesima proprietà, nel rimanente insieme di casi no data la disuguaglianza di Jensen.

L'equazione (4.16) può essere molto utile in determinate condizioni, ad esempio se la risposta viene trasformata per il logaritmo, si è sempre perfettamente

giustificati nell'interpretare $e^{x_i'\hat{\beta}_\tau}$ come una stima appropriata del quantile condizionato di livello τ della Y ; cosa non giustificabile formalmente per la media condizionata (Koenker, 2005).

La regressione del quantile è molto utile per dati censurati e appunto legati a questi vi è un'altra importante proprietà, la robustezza che si hanno nelle stime dei quantili condizionati della risposta a valori che possono essere anche *outliers* delle variabili esplicative (Koenker 2005). Questo significa che qualsiasi osservazione, anche estrema, dei regressori non può alterare la stima dei quantili condizionati che produce il modello di regressione; al contrario del modello di regressione lineare classico che non è robusto a valori estremi e perturbazioni delle esplicative.

Infine una delle ultime considerazioni può riguardare l'efficienza, per una vasta gamma di distribuzioni non normali della variabile d'interesse, la varianza asintotica della mediana campionaria è all'incirca il 50% più piccola, se non ancora nettamente più piccola, della varianza asintotica della media campionaria, che produce risultati anche disastrosi; solo nel caso di una distribuzione normale il rapporto si inverte e la media risulta più efficiente. Koenker e Bassett (1978) estendono questo concetto legato a misure campionarie a quello della regressione del quantile, affermando che la mediana condizionata è uno stimatore più efficiente di quello ottenuto con minimi quadrati ordinari per la media condizionata sotto ogni ipotesi distributiva non gaussiana. È quindi ragionevole sacrificare l'ipotesi di normalità e utilizzare la regressione del quantile per ottenere un notevole miglioramento, rispetto ai minimi quadrati ordinari, in termini di efficienza; che può essere aumentata con opportune procedure di ponderazione qualora le osservazioni presentino strutture di correlazione e eteroschedasticità (Koenker, 2010b).

L'interpretazione dei parametri nella regressione del quantile è simile a quella dei parametri del modello di regressione lineare classico qualora; per la k -esima variabile il parametro associato è β_k che altro non è che la derivata parziale del valore atteso condizionato di y dato x_k (*ceteris paribus*): $\beta_k = \frac{dE(y|X)}{dx_k}$, da interpretarsi

come l'effetto marginale che la k -esima variabile ha sulla media condizionata. Se vi è per la stessa covariata più di un coefficiente associato e questa compare in una forma polinomiale l'effetto marginale sulla media è sempre $\frac{dE(y|X)}{dx_k}$ ma non è più associato al solo parametro associato β_k . Per il modello di regressione del quantile l'interpretazione è analoga a quello di regressione lineare scanso che gli effetti marginali di una variabile (*ceteris paribus*) sono da riferirsi al quantile condizionato di livello τ . Se il quantile condizionato di livello τ viene definito come

$$Q_\tau(y_t|x_t) = \beta_{\tau,0} + \sum_{j=1}^p \beta_{\tau,j}x_{t,j}$$

l'effetto marginale della k -esima variabile sul quantile condizionato di livello τ (*ceteris paribus*) è

$$\frac{dQ_\tau(y_t|x_t)}{dx_k} = \beta_{\tau,k}$$

nel caso in cui non ci siano più parametri associati alla k -esima variabile; quando la x_k cambia, cambia anche il quantile condizionato della risposta ma la tale effetto ha impatto variabile in relazione al livello del quantile considerato. Siccome la regressione del quantile gode della proprietà di equivarianza a trasformazioni monotone della risposta, equazione (4.16), se stiamo stimando una regressione del quantile su trasformazioni logaritmiche ($Q_\tau(\log y_t|x_t) = x_t'\beta_\tau$) allora possiamo definire l'effetto marginale della k -esima variabile sul quantile τ -esimo della variabile risposta y (*ceteris paribus*) come

$$\frac{dQ_\tau(y_t|x_t)}{dx_k} = \beta_{\tau,k}e^{x_t'\beta_\tau}$$

Per esempio i modelli di regressione del quantile univariati o multivariati su come vengono rappresentate le stime dei parametri β_τ al variare di τ e su come sia possibile valutare la distribuzione condizionata tramite i quantili condizionati si veda Koenker (2010).

Per dare l'intuizione di come gli effetti marginali, e in generale i valori dei parametri, possano variare anche significativamente al variare del livello del quantile condizionato stimato consideriamo una regressione del quantile multivariata con due esplicative. Si generano 1000 valori indipendenti casuali da un unifor-

me (20,80) che prendono il ruolo del primo regressore (x_1), la seconda esplicativa viene definita dallo schema $x_2 = 100x_1 + \varepsilon$ con $\varepsilon \sim N(0, 500^2)$. É evidente come i due regressori siano strettamente correlati, infine generiamo la variabile risposta come $y = 1000 - 10x_1 + 0.2x_2 + \nu$ con $\nu \sim N(0, (200 - x_1)^2)$. Un modello di regressione lineare classico per la stima della media condizionata è

$$\widehat{E}(y|x_1, x_2) = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_2$$

mente uno per il τ -esimo quantile condizionato è

$$\widehat{Q}_\tau(y|x_1, x_2) = \widehat{\beta}_{0,\tau} + \widehat{\beta}_{1,\tau} x_1 + \widehat{\beta}_{2,\tau} x_2$$

Non ci soffermiamo sull'efficacia del modello di regressione del quantile in questo caso rispetto al modello di regressione lineare classico e sulle valutazioni applicative ma valutiamo l'idea di come i coefficienti si comportano al variare di τ . Senza riportare gli output delle stime numericamente del modello di regressione lineare classico e quello del quantile, li rappresentiamo graficamente con un tipico grafico in figura 4.2 per i soli coefficienti legati alle esplicative. Si nota come al variare del livello del quantile condizionato, i coefficienti legati a x_1 e x_2 , variano anche in modo significativo rispetto a quelli del modello di regressione lineare e a loro volta se confrontati con gli stessi coefficienti di regressione del quantile per differenti livelli di τ . Per interessanti applicazioni empiriche si veda Koenker e Hallock (2001).

4.2.2 Inferenza e risultati asintotici

In generale si ricorre alla teoria asintotica per riuscire a svolgere procedure inferenziali, e in particolar modo nella regressione del quantile questa è molto utile in quanto è difficile derivare una distribuzione opportuna per gli stimatori che non sfruttano le proprietà asintotiche. La *sparsity function* $s_y(\tau)$, introdotta in equazione (4.4), riveste un ruolo fondamentale per la definizione delle caratteristiche della distribuzione asintotica, quali il modo di specificare la forma della matrice di varianze covarianze asintotica.

Prima di estendere l'analisi alla regressione del quantile, enunciamo i risultati asintotici per lo stimatore del quantile non condizionato nel campione univariato

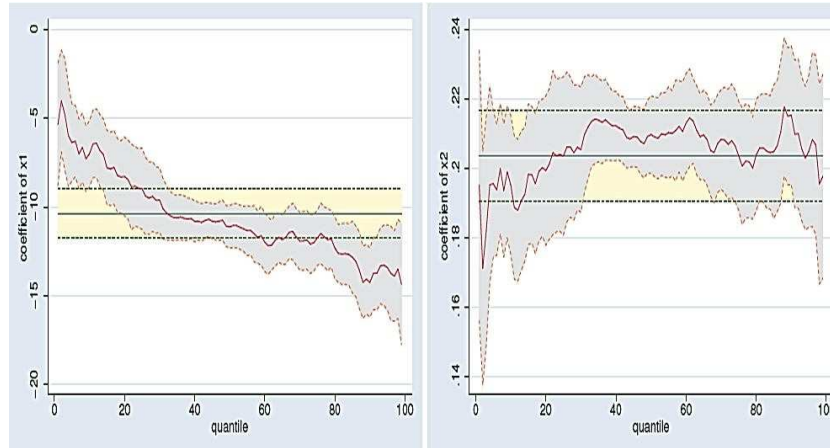


Figura 4.2: Coefficienti della regressione del quantile. La figura presenta, con la linea continua rossa, a sinistra il coefficiente del modello di regressione del quantile per x_1 e a destra quello di x_2 calcolati per 99 quantili differenti a cui viene associato un intervallo di confidenza a livello 95%. Viene sovrapposta la stima dei coefficienti dei minimi quadrati ordinari del modello di regressione con una linea blu e relativo intervallo di confidenza (Koenker, 2005).

che abbiamo già definito in equazione (4.6) come

$$\hat{Q}_Y(\tau) = \arg \min_{\xi_\tau \in R} \sum_t \rho_\tau(Y_t - \xi_\tau)$$

dove si può dimostrare sotto opportune ipotesi, soprattutto legate all'indipendenza e identica distribuzione delle osservazioni, che

$$\sqrt{T}(\hat{Q}_Y(\tau) - \xi_\tau) \mapsto N(0, w^2) \text{ con } w^2 = \frac{\tau(1-\tau)}{f_Y^2(\xi_\tau)} = \tau(1-\tau)s_Y^2(\tau) \quad (4.17)$$

dove la precisione delle stimatore del quantile non condizionato (valutata per la sua varianza w^2) è determinata dalla sparsità e dispersione dei dati attorno uno specifico quantile che si vuole stimare, dato questo il legame della varianza con la *sparsity function* e con la funzione di densità della variabile casuale. L'equazione (4.17) può essere estesa alla distribuzione congiunta multivariata dello stimatore di alcuni quantili non condizionati che chiamiamo ξ , un vettore di m quantili semplici ($\xi = (\xi_{\tau_1}, \dots, \xi_{\tau_m})$), e definiamo con $\hat{\xi}$ lo stimatore. Può essere mostrato che (Koenker e Portnoy, 1999)

$$\sqrt{T}(\hat{\xi} - \xi) \mapsto N(0, \Omega) \text{ con } \Omega = \{w_{ij}\}_{m \times m} = \frac{\min(\tau_i, \tau_j) - \tau_i \tau_j}{f(F^{-1}(\tau_i))f(F^{-1}(\tau_j))}$$

per maggiori dettagli si veda Koenker (2010b).

Passando alla teoria asintotica per gli stimatori della regressione del quantile le cose si complicano. Consideriamo che gli stimatori, $\widehat{\beta}_\tau$ convergano in probabilità ai loro veri valori, ovvero siano consistenti, sotto alcune condizioni necessarie e sufficienti (Koenker, 2010b). Per analogia a quanto detto in equazione (4.17) può essere mostrato, sotto condizioni di regolarità e senza ipotesi circa la natura delle osservazioni quali l'indipendenza e l'identica distribuzione, che

$$\sqrt{T}(\widehat{\beta}_\tau - \beta_\tau) \xrightarrow{d} N(0, \tau(1 - \tau)H_\tau^{-1}JH_\tau^{-1}) = N(0, \Gamma_\tau) \quad (4.18)$$

con

$$J = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_t x_t x_t'$$

e

$$H_\tau = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_t x_t x_t' f_t(Q_\tau(y_t|x_t))$$

dove alcuni autori motivano tali risultati asintotici sulla continuità stocastica e monotonia del gradiente delle condizioni del primo ordine, mentre altri autori motivano questi risultati sulla convessità della funzione obiettivo portata al limite, si veda Koenker (2010b).

Sotto l'ipotesi di indipendenza e identica distribuzione è possibile semplificare la matrice di varianze covarianze asintotica di equazione (4.18) come

$$\Gamma_\tau = \frac{\tau(1 - \tau)}{f^2(F^{-1}(\tau))} J^{-1} \quad \text{con} \quad J = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_t x_t x_t' \quad (4.19)$$

dove si nota una semplificazione nella forma "sandwich" dello stimatore della varianza di equazione (4.18) e l'uso di una funzione di densità indistintamente $f()$. L'ipotesi di indipendenza e identica distribuzione delle osservazioni è però molto restrittiva e difficile da assumere, se si applica la regressione in un contesto di serie storiche, o in genere in applicazioni empiriche, appare tutt'altro che perseguibile. Tuttavia la semplificazione che si ottiene sotto l'ipotesi di indipendenza ed identica distribuzione permette di calcolare matrici di varianze e covarianze che solitamente non ammetterebbero soluzione numerica, inoltre la matrice di varianze e covarianze ottenuta sotto ipotesi di i.i.d può utilizzare come *benchmark* su cui costruire test di verifica d'ipotesi.

Possiamo anche specificare la distribuzione congiunta asintotica di vettori diversi dei coefficiente di regressione per vari livelli di τ . Sia ξ un vettore di m vettori p -variati dei coefficienti di regressione del quantile per differenti livelli di τ , ($\xi = (\beta'_{\tau_1}, \dots, \beta'_{\tau_m})$), e $\widehat{\xi}$ il suo stimatore. La distribuzione asintotica di questi m stimatori vettoriali è, si veda Koenker e Portnoy (1999)

$$\sqrt{T}(\widehat{\xi} - \xi) \mapsto N(0, \Omega) \quad \text{con} \quad \Omega = (w_{ij}) = (\min(\tau_i, \tau_j) - \tau_i \tau_j) H_{\tau_1}^{-1} J H_{\tau_2}^{-1}$$

dove

$$J = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_t x_t x_t' \quad \text{e} \quad H_\tau = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_t x_t x_t' f_t(Q_\tau(y_t | x_t))$$

La procedura di stima delle matrici di varianza e covarianza asintotica è molto importante ai fini inferenziali e rispetta le ipotesi sulle osservazioni che si possono fare a livello di definizione del modello di regressione del quantile. In generale esistono metodi più o meno potenti e validi che comunque si basano sulla scelta dell'ipotesi di indipendenza e identica distribuzione delle osservazioni oppure no e al più su una linearità locale della risposta condizionatamente alle x . Le procedure di stima della matrice asintotica di varianze e covarianza di $\widehat{\beta}_\tau$ è legata ad una stima consistente della *sparsity function* che deriva dalle ipotesi precedenti e che è un elemento presente nella forma matriciale da stimare. In generale si propongono 5 metodi per ottenere una stima consistente della matrice di varianze e covarianze asintotica:

1. stima della matrice di varianze e covarianze di equazione (4.19) con opportuna stima di $s_Y(\tau)$ per l'ipotesi di osservazioni i.i.d.;
2. stima della matrice asintotica che sfrutti il metodo di *bootstrap* sui residui del modello di regressione del quantile;
3. stima della matrice asintotica sulla base dell'inversione del test del rango che può essere implementata sia sotto l'ipotesi di i.i.d. che non, metodo del rango;
4. stima della matrice di varianze e covarianze che presuppone mancata indipendenza ma identica distribuzione delle osservazioni e una linearità locale

nella funzione del quantile condizionato (nelle x), stimatore a *sandwich di Huber*;

5. stima non parametrica della matrice di varianze e covarianze asintotica che sia indipendente da ipotesi sulla struttura delle osservazioni e le relazioni possibili, metodo del nucleo (*Kernel*);

I primi 4 metodi non vengono approfonditi per brevità della trattazione e perché non sono stati impiegati in tale tesi in quanto impossibili da sostenere nelle loro ipotesi, si veda Koenker (2010b), viene invece brevemente esposto il metodo di stima del nucleo della matrice di varianze covarianze asintotica impiegato. Riferendoci alla distribuzione asintotica espressa per $\hat{\beta}_\tau$ in equazione (4.18), che non necessita di ipotesi circa l'indipendenza e l'identica distribuzione delle osservazioni, si ha che

$$\sqrt{T}(\hat{\beta}_\tau - \beta_\tau) \xrightarrow{d} N(0, \tau(1 - \tau)H_\tau^{-1}JH_\tau^{-1}) = N(0, \Gamma_\tau)$$

con

$$J = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_t x_t x_t' \quad \text{e} \quad H_\tau = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_t x_t x_t' f_t(Q_\tau(y_t|x_t))$$

che mostra come nella matrice di varianze e covarianze vi sia, via H_τ , la dipendenza da ogni differente distribuzione delle osservazioni. Dopo un opportuna stima della densità per ciascuna osservazione che deve essere sempre presa nella sua parte positiva (cosa non sempre scontata per errori di carattere computazionale legati a posizioni estreme dei quantili nel piano di proiezione della matrice di disegno) che viene definita come

$$\hat{f}_i^+(Q_\tau(y_t|x_t)) = \max \left(0; \frac{2h_T}{x_t'(\hat{\beta}_{\tau+h_T} - \hat{\beta}_{\tau-h_T}) - \varepsilon} \right)$$

dove h_T è un parametro di banda (lisciamento) definibile in vari modi e $\varepsilon > 0$ per evitare di dividere per un valore nullo (Koenker, 2005), è possibile attribuire una stima non parametrica della matrice H_τ con metodo del nucleo come

$$\hat{H}_\tau = (Th_n)^{-1} \sum_t K \left(\frac{\hat{u}_{t,\tau}}{h_n} \right) x_t x_t'$$

dove $\hat{u}_{t,\tau} = y_t - \hat{Q}_\tau(y_t|x_t)$ è il residuo della regressione del quantile di livello τ . Per ulteriori approfondimenti si veda Koenker(2010b). Procedure inferenziali quali verifiche d'ipotesi sui singoli coefficienti o blocchi di questi e costruzioni di intervalli di confidenza sono definiti in modo classico basandosi su questi risultati asintotici.

4.3 Regressione del quantile non parametrica

Anche se la regressione non parametrica è tradizionalmente utilizzata per la regressione delle media condizionata, diversi autori hanno affrontato il problema della stima non parametrica per modelli di regressione del quantile. Un esempio semplificato è quello di una regressione non parametrica con un'unica variabile esplicativa, il cui effetto viene stimato non parametricamente mediante l'uso delle *spline* di lisciamento da Portnoy, Koenker e Ng (1993). Per un $\tau \in [0, 1]$ la stima del quantile che deriva dalla stima della *spline* $g(\cdot)$ di lisciamento che minimizza

$$\sum_{t=1}^T \rho_\tau(y_t - g(x_t)) + \lambda \int (g''(x))^2 dx$$

dove $\lambda \in R^+$ controlla il grado di lisciamento imponendo alla funzione obiettivo una penalità legata alla regolarità della *spline*, la scelta di λ è critica e definisce implicitamente il numero di osservazioni che verranno interpolate dalla funzione $g(\cdot)$. Dando una forma funzionale a $g(\cdot)$ definendola ad esempio come *spline* cubica ($g(x_t) = \gamma_0 + \gamma_1 x_t + \gamma_2 x_t^2 + \gamma_3 x_t^3$) si può regolare il numero di nodi interpolati stabilendo un opportuno valore del parametro di lisciamento. Per maggiori dettagli ed esempi si veda Hendricks e Koenker (1990) e Portnoy, Koenker e Ng (1993).

In molte applicazioni pratiche però ci presenterà spesso la questione di estendere questi metodi non parametrici ad un ambito multivariato. Tra i vari approcci vi è solitamente il passaggio ad un modello forzatamente additivo a componenti non parametriche per garantire il raggiungimento di una stima del quantile condizionato che non sarebbe altrimenti ottenibile con un modello totalmente non parametrico. In generale, ipotizzando di avere p variabili esplicative, il quantile

condizionato della risposta y_t può essere definito come

$$Q_\tau(y_t|x_t) = \sum_{j=1}^p g_j(x_{j,t})$$

dove le $g_j()$ sono funzioni non parametriche a valori reali definite specificatamente per ogni esplicativa; la stima del quantile condizionato passa da una stima congiuntamente non parametrica delle funzioni g ($\widehat{Q}_\tau(y_t|x_t) = \sum_j \widehat{g}_j(x_{j,t})$) che in generale è ottenibile come

$$\min_{g_1, \dots, g_p} \sum_{t=1}^T \rho_\tau(y_t - \sum_{j=1}^p g_j(x_{j,t}))$$

dove il processo di stima è nettamente più complesso di quello della regressione del quantile e può prevedere una sequenza in più passi. Inoltre si deve scegliere un opportuna forma delle funzioni g_j ; ad esempio Hastie e Tibshirani (1990) suggeriscono come forma funzionale delle g delle *spline* di regressione o lisciamento. Per ottenere le stime dei quantili condizionati e delle relative funzioni non parametriche si può ancora adoperare un opportuno algoritmo del semplice definito opportunamente in più passi. Ad ogni modo un elevato numero di esplicative creare problemi di natura computazionale e raggiungimento delle stime delle funzioni non parametriche. Per maggiori dettagli si veda Portnoy, Koenker e Ng (1993) e Hastie e Tibshirani (1990).

Capitolo 5

APPLICAZIONI DELLA REGRESSIONE DEL QUANTILE AI PREZZI ELETTRICI DEL MERCATO UK

Proviamo ora ad applicare la regressione del quantile alla serie dei prezzi per prevederne i picchi. Il metodo di previsione di un picco di prezzo rimane legato alla realizzazione di una previsione di un prezzo estremo per il giorno successivo. Accertare che dei quantili dei prezzi *day-ahead* superino una certa soglia rappresenta la regola per definire una previsione di picco. Viene quindi impiegata la regressione del quantile per prevedere i quantili condizionati dei prezzi. I picchi previsti vengono confrontati con i picchi che si realizzeranno e si possono definire le *performance* del modello previsionale in un'ottica globale. Ogni fascia semi oraria viene trattata singolarmente adattando un modello di regressione del quantile *ad hoc*. Si presenta l'applicazione della regressione del quantile lineare parametrica e il metodo di regressione del quantile non parametrico.

5.1 Modello di Regressione del quantile lineare

Il modello di regressione del quantile pare particolarmente idoneo perché è in grado di valutare un effetto marginale non necessariamente stabile delle esplicative per ogni livello del quantile condizionato; elemento che potrebbe essere fondamentale per una corretta previsione dei picchi dei prezzi elettrici. Definiamo, per un generico livello del quantile τ , il modello di regressione del quantile lineare parametrico più completo in termini di esplicative per la stima del quantile condizionato di livello τ della componente stocastica del logaritmo dei prezzi.

Data l'analisi sulle esplicative svolta precedentemente, il modello di regressione del quantile condizionato per ciascuna fascia semi oraria per un $\tau \in (0, 1)$ è

$$p_t = \beta_{0,\tau} + \beta_{1,\tau}p_{t-1} + \beta_{2,\tau}p_{t-2} + \beta_{3,\tau}p_{t-3} + \beta_{4,\tau}p_{t-4} + \beta_{5,\tau}p_{t-5} + \beta_{6,\tau}p_{t-6} + \beta_{7,\tau}p_{t-7} + \beta_{8,\tau}d_t + \beta_{9,\tau}d_{t-1} + \beta_{10,\tau}m_t + \beta_{11,\tau}m_{t-1} + \beta_{12,\tau}dm_t + \beta_{13,\tau}dm_{t-1} + \varepsilon_t \quad (5.1)$$

dove i $\beta_{j,\tau}$ per $j = 1, \dots, 13$ sono i parametri associati a ciascuna delle esplicative. L'assunzione del modello è che il quantile condizionato di livello τ dell'errore di regressione (e non la media condizionata come nel modello di regressione classico) sia nullo ovvero $Q_\tau(\varepsilon_t|I_t) = 0$. Questo comporta che il quantile condizionato di livello τ della componente stocastica dei prezzi p_t , $Q_\tau(p_t|I_t)$, è

$$Q_\tau(p_t|I_t) = \beta_{0,\tau} + \beta_{1,\tau}p_{t-1} + \beta_{2,\tau}p_{t-2} + \beta_{3,\tau}p_{t-3} + \beta_{4,\tau}p_{t-4} + \beta_{5,\tau}p_{t-5} + \beta_{6,\tau}p_{t-6} + \beta_{7,\tau}p_{t-7} + \dots + \beta_{8,\tau}d_t + \beta_{9,\tau}d_{t-1} + \beta_{10,\tau}m_t + \beta_{11,\tau}m_{t-1} + \beta_{12,\tau}dm_t + \beta_{13,\tau}dm_{t-1}$$

dove I_t rappresenta l'insieme informativo al quale ci si condiziona al tempo t e definisce un modello di *quantile regression* lineare autoregressivo. Si nota come i parametri $\beta_{j,\tau}$, che rappresentano gli effetti marginali di ciascuna esplicativa sul quantile condizionato della risposta, siano indicizzati a τ . Questo evidenzia che gli effetti marginali sono dipendenti dal livello del quantile stimato e vi è la possibilità che l'effetto marginale di ciascun regressore possa cambiare al variare del livello del quantile condizionato da considerare.

La rappresentazione del modello di regressione del quantile (5.1) in forma compatta è $p_t = x_t'\beta_\tau + \varepsilon_t$, dove $\beta_\tau = (\beta_0 \beta_1 \dots \beta_{13})'$ è il vettore dei parametri e $x_t' = (1 \ p_{t-1} \ p_{t-2} \ p_{t-3} \ p_{t-4} \ p_{t-5} \ p_{t-6} \ p_{t-7} \ d_t \ d_{t-1} \ m_t \ m_{t-1} \ dm_t \ dm_{t-1})'$ è il vettore delle esplicative al tempo t . La stima dei parametri, $\hat{\beta}_\tau$, dal campione *in sample* è

$$\min_{\beta_\tau \in R^{14}} \sum_{t=1}^{1731} \rho_\tau(p_t - x_t'\beta_\tau) = \min_{\beta_\tau \in R^{14}} \left(\sum_{\{t|p_t \geq x_t'\beta_\tau\}} \tau |p_t - x_t'\beta_\tau| + \sum_{\{t|p_t < x_t'\beta_\tau\}} (1 - \tau) |p_t - x_t'\beta_\tau| \right)$$

dove $\rho_\tau(u) = (\tau - I(u < 0))u$ è la *check function*. Stimati i parametri nel campione *in sample* si possono stimare i quantili condizionati di livello $\tau \in [0, 1]$ come $\widehat{Q}_\tau(p_t|x_t) = x_t'\hat{\beta}_\tau$. Per validare il modello stimato è possibile verificare la proprietà per cui l'errore abbia quantile τ -esimo nullo. Questa proprietà è comprovabile

nei residui del modello di regressione $r_t = p_t - \widehat{Q}_\tau(p_t|x_t)$, i quali dovranno esibire un quantile empirica nullo.

In tabella 5.1 e 5.2 vengono riportate le stime dei parametri del modello (5.1) per due livelli del quantile condizionato, $\tau = 0.5$ (regressione mediana) e per un livello elevato $\tau = 0.9$, per la fascia semi oraria 12.00–12.30. Vengono riportati gli *standard error* ottenuti in modo non parametrico con metodo del nucleo (unica possibilità implementata in R che eviti le ipotesi di indipendenza o identica distribuzione delle osservazioni) e il relativo test T di significatività per ciascuna esplicativa con relativo *p-value*. Si può osservare che la stima dei parametri già

	Stime	Std. Error	Statistica t	p-value statistica t
(Intercetta)	0.01	0.01	0.65	0.51
p1	0.20	0.04	5.60	0.00
d0	-0.39	0.39	-0.99	0.32
m0	-0.82	0.07	-12.59	0.00
d1	1.07	0.40	2.65	0.01
m1	0.22	0.07	3.20	0.00
p2	0.03	0.04	0.83	0.41
p3	-0.04	0.03	-1.12	0.26
p4	-0.02	0.03	-0.68	0.50
p5	-0.02	0.03	-0.77	0.44
p6	-0.05	0.03	-1.68	0.09
p7	0.04	0.03	1.24	0.22
dm0	-0.43	1.62	-0.27	0.79
dm1	-1.40	1.46	-0.96	0.34

Tabella 5.1: Stima dei parametri nel modello di regressione del quantile per livello $\tau = 0.5$ del quantile condizionato per la fascia semi oraria 12.00–12.30 campione *in sample*, *standard errors* calcolati con metodo *kernel*, statistica t per la significatività di ogni singolo parametro e relativo *p-value*.

per questi due livelli di τ risulti diversa numericamente anche in modo consistente come per altro gli *standard error* e i test di significatività. Per poter concludere che al variare del livello del quantile le stime dei parametri cambino, è possibile saggiare il sistema d'ipotesi che contrappone l'invarianza statistica dei parametri (ipotesi nulla: $H_0 : \beta_{0.5} = \beta_{0.9}$) contro parametri variabili al variare di τ (ipotesi alternativa: $H_1 : \beta_{0.5} \neq \beta_{0.9}$). La statistica test, riconducibile ad un test F, conduce a rifiutare l'ipotesi nulla presentando un *p-value* di 0.0001. Questo permette di evidenziare che gli effetti marginali cambiano significativamente se τ passano da

	Stime	Std. Error	Statistica t	p-value statistica t
(Intercetta	0.29	0.01	19.75	0.00
p1	0.26	0.06	4.38	0.00
d0	-0.15	0.79	-0.19	0.85
m0	-1.15	0.10	-11.89	0.00
d1	1.59	0.69	2.29	0.02
m1	0.30	0.09	3.27	0.00
p2	-0.02	0.04	-0.54	0.59
p3	-0.03	0.04	-0.64	0.52
p4	-0.01	0.04	-0.25	0.80
p5	0.03	0.05	0.53	0.60
p6	-0.04	0.05	-0.66	0.51
p7	-0.02	0.04	-0.45	0.65
dm0	-3.02	2.23	-1.36	0.18
dm1	-1.13	2.28	-0.50	0.62

Tabella 5.2: Stima dei parametri nel modello di regressione del quantile per livello $\tau = 0.9$ del quantile condizionato per la fascia semi oraria 12.00–12.30 campione *in sample*, *standard errors* calcolati con metodo *kernel*, statistica t per la significatività di ogni singolo parametro e relativo *p-value*.

0.5 a 0.9 e non rimangono costanti. Il test è generalizzabile anche per più livelli di τ . Al fine della nostra analisi può essere interessante valutare se per valori elevati ci sia stabilità negli effetti marginali, verificando $H_0 : \beta_{0.7} = \beta_{0.8} = \beta_{0.9}$ contro l'ipotesi che almeno un'uguaglianza non sia verificata. Il test fornisce un *p-value* di 0.017 che può condurre al rifiuto dell'ipotesi nulla. La non stabilità degli effetti marginali al variare di τ è osservabile anche in figura 5.1 dove si propongono le stime dei parametri per ciascuna esplicativa, in riferimento al modello (5.1), stimando il modello nel campione *in sample* al variare di $\tau \in [0.0025, 0.9975]$ (valori più prossimi a zero e a uno pongono problemi di stabilità nell'algoritmo di stima). Con linea nera "continua" vengono proposte le stime puntuali di β_τ per ciascuna esplicativa alle quali è associato un intervallo di confidenza basato su *standard error* calcolati in modo non parametrico con il metodo *kernel* per un livello di confidenza del 95% (area grigia); la linea rossa continua rappresenta la stima dell'effetto marginale per una regressione lineare classica della media condizionata, alla quale viene associato un intervallo fiduciario calcolato per un livello di confidenza del 95% con approccio classico dei minimi quadrati ordinari (linee tratteggiate rosse).

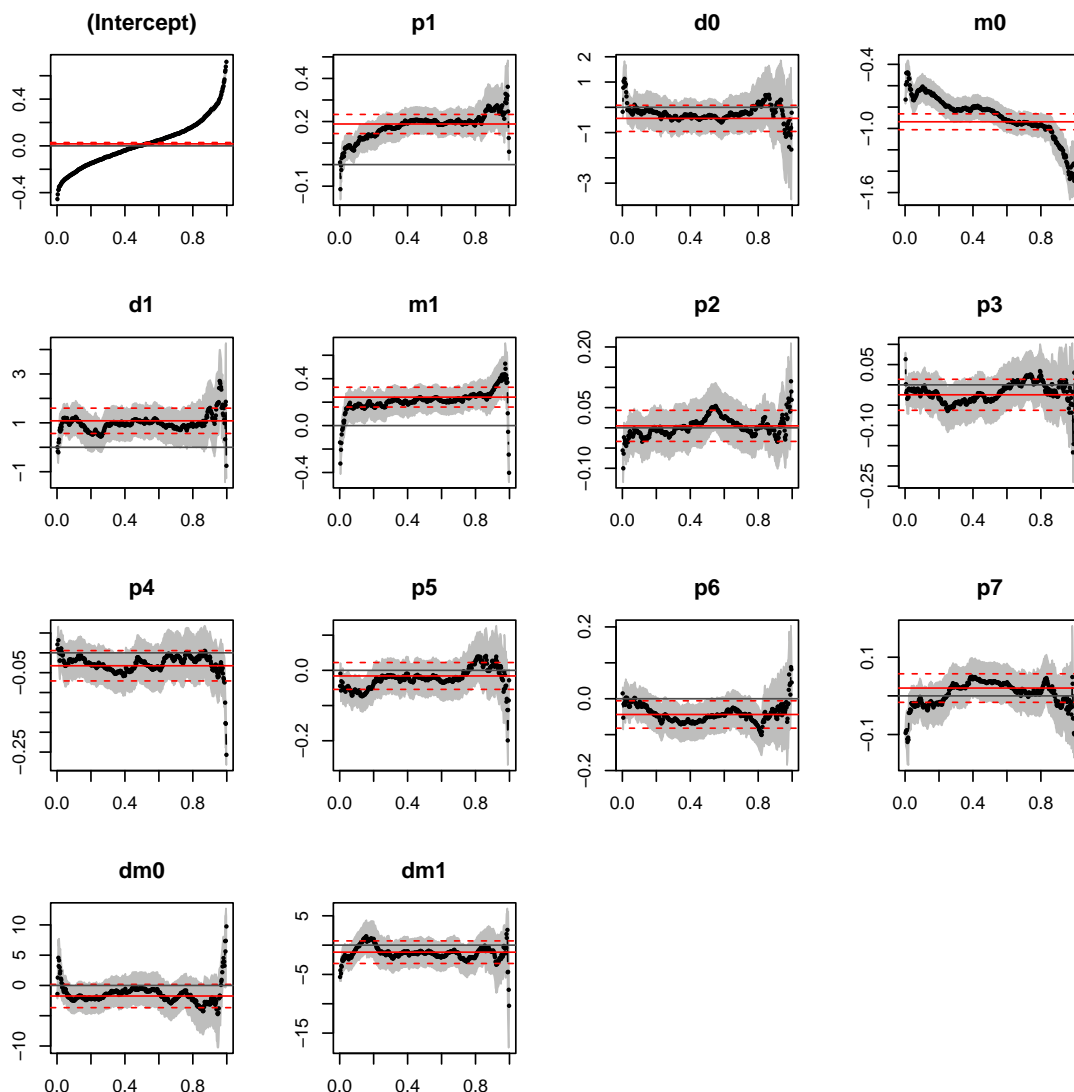


Figura 5.1: Stima dei parametri β_τ del modello di regressione del quantile con tutte le esplicative al variare di $\tau \in [0, 1]$ con intervallo di confidenza definito per un livello di copertura del 95% sotto l'ipotesi di normalità con *standard error* stimati non parametricamente con metodo *kernel* per la fascia semi oraria 12.00–12.30 nel campione *in sample*.

La figura 5.1 mostra, almeno per alcune esplicative, che l'effetto marginale è variabile e significativamente diverso da quello medio se si considera che, per alcuni livelli di τ , gli intervalli di confidenza non si intersecano nemmeno. Con riferimento alla figura 5.1 le variabili che hanno associato un effetto marginale sul quantile condizionato particolarmente variabile e che si discosta da un effetto medio costante appaiono soprattutto il prezzo con ritardo uno, il margine e il suo ritardo uno (p_{t-1} , m_t e m_{t-1}). In figura 5.2 viene riportata la stima dei

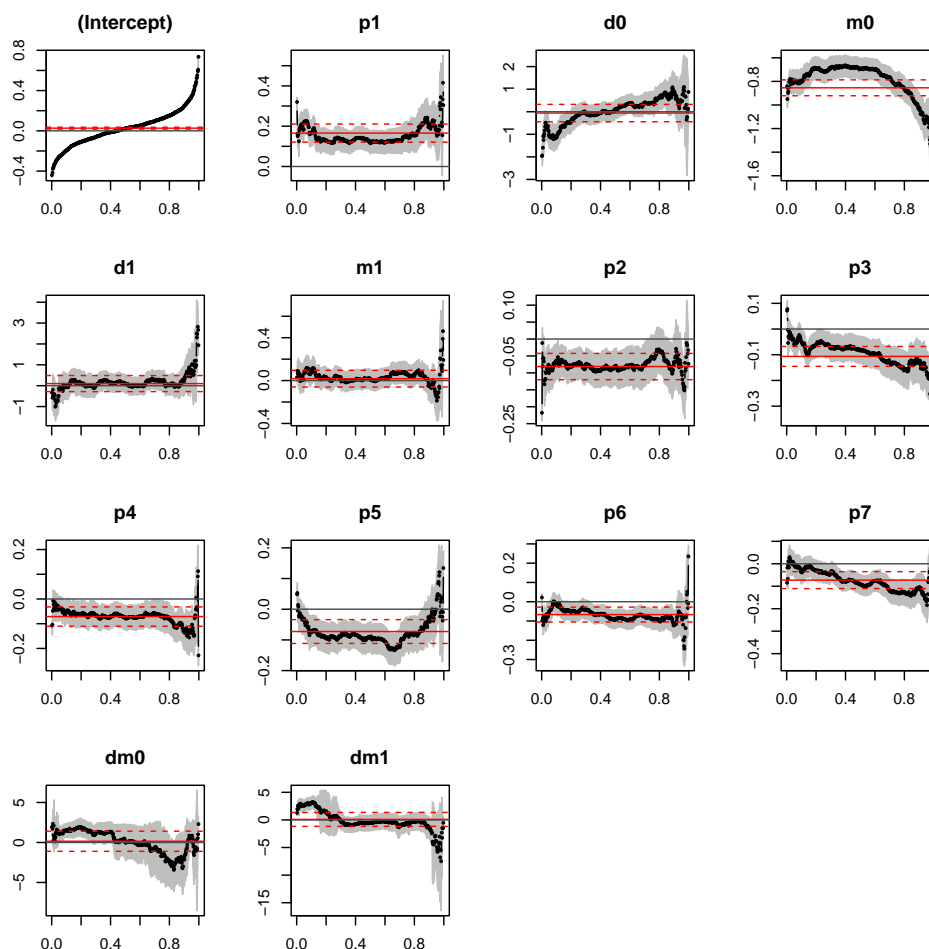


Figura 5.2: Stima dei parametri β_τ del modello di regressione del quantile con tutte le esplicative al variare di $\tau \in [0, 1]$ con intervallo di confidenza definito per un livello di copertura del 95% sotto l'ipotesi di normalità con *standard error* stimati non parametricamente con metodo *kernel* per la fascia semi oraria 18.00–18.30 nel campione *in sample*.

β_τ per ciascun esplicative del modello completo per la fascia semi oraria 18.00–18.30 al variare di τ ; si nota ancora come l'effetto marginale di alcune esplicative sia variabile e non si possa fissare ad un livello medio. Variabili particolarmente soggette a questo sono margine contemporaneo e ritardato, la domanda ritardata e alcuni ritardi della componente dei prezzi. Anche per questa fascia semi oraria le stime degli effetti marginali non appaiono stabili al variare del livello del quantile anche verificando con un test opportuno.

Quanto detto e mostrato per queste due fasce semi orarie è generalizzabile e osservabile anche per le altre fasce semi orarie che rispecchiano questi comportamenti.

In generale per dare l'idea di come l'effetto marginale congiunto delle esplicative possa variare in funzione del livello del quantile condizionato si verifica l'ipotesi $H_0 : \beta_{0.7} = \beta_{0.8} = \beta_{0.9}$ contro l'ipotesi alternativa che almeno una uguaglianza non si verifichi su ciascuna fascia semi oraria. Per 7 fasce semi orarie, per un livello di significatività fissato al 5%, il test non permette di rifiutare l'ipotesi nulla, ovvero per tali livelli del quantile condizionato l'effetto congiunto delle esplicative è da non ritenersi significativamente differente, si veda tabella 5.3. Per le rimanenti 41 fasce semi orarie l'effetto delle esplicative varia in modo significativo per i tre differenti livelli del quantile.

Fascia semi oraria	6	8	9	11	12	23	47
p-value	0.25	0.09	0.13	0.15	0.17	0.06	0.21

Tabella 5.3: Fasce semi orarie in cui il test per saggiare il sistema d'ipotesi $H_0 : \beta_{0.7} = \beta_{0.8} = \beta_{0.9}$ rifiuta l'ipotesi nulla a favore dell'alternativa, che si realizzi almeno una disuguaglianza, con un livello di significatività del 5%, si riportano i *p-value* del test per relative fasce semi orarie. Dodici sono le fasce semi orarie dove gli effetti marginali per livelli del quantile 0.7, 0.8 e 0.9 non sono ritenuti significativamente differenti.

Questa può essere una valida giustificazione per l'impiego del modello di regressione del quantile per il calcolo dei quantili condizionati, come proposta di modello a parametri variabili rispetto il livello del quantile condizionato da stimare. La cosa particolarmente interessante è soffermarsi a valori di τ elevati perché è proprio a questi particolari casi che si ricorre per la previsione dei picchi di prezzo.

Dalle figure 5.1 e 5.2 emerge però anche l'evidenza che plausibilmente alcune esplicative non siano necessarie per una regressione del quantile in quanto stabili al variare di τ e perché nei loro intervalli di confidenza compare sempre il valore nullo (sfruttiamo il concetto di dualità tra verifica d'ipotesi ed intervallo di confidenza). È necessario ottenere un modello che utilizzi soli regressori significativi per la stima e la previsione dei quantili condizionati per ciascuna fascia semi oraria. Si procede all'individuazione dei regressori necessari per ciascuna fascia semi oraria con un processo simile a quello utilizzato nei modelli *ARX*. Dovendo soffermare sulla previsione dei picchi di prezzo si è interessati a studiare il

comportamento per τ elevati e realizzare modelli per questi.

Per ciascuna fascia semi oraria vengono stimati nel campione *in sample* vari modelli come quello di equazione (5.1) per differenti livelli di τ . In ciascun caso, si procede eliminando passo-passo le variabili statisticamente non significative mediante la valutazione di un test T per la significatività di ogni singolo regressore. Si confronta se il modello con sempre meno esplicative sia sempre valido rispetto a quello di partenza per garantire un buon adattamento ai dati (un test F per modelli nidificati, Koenker(2005)). Tutti i test sono condotti con un livello di significatività dell'1%.

In tabella 5.4 vengono riportati per ogni livello di τ preso in considerazione le sole variabili esplicative necessarie per la definizione di un modello di regressione del quantile, per la fascia semi oraria 12.00–12.30. La lettura della tabella è la seguente, preso ad esempio il quantile di livello 90% ($\tau = 0.9$), le variabili necessarie per ottenere un buon modello sono: p_{t-1} , m_t , m_{t-1} , . Ogni riga contrassegna i soli regressori utili per formalizzare un modello di regressione del quantile per uno specifico livello del quantile condizionato ridotto. In tabella B.4

Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
0.5	V	-	V	V	V	-	-	-	-	V	-	-	-
0.75	V	-	V	-	V	-	-	-	-	-	-	-	-
0.8	V	-	V	-	-	-	-	-	-	-	-	-	-
0.85	V	-	V	-	-	-	-	-	-	-	-	-	-
0.88	V	-	V	-	-	-	-	-	-	-	-	-	-
0.9	V	-	V	-	V	-	-	-	-	-	-	-	-
0.93	V	-	V	V	V	-	-	-	-	-	-	-	-
0.95	V	-	V	-	V	-	-	-	-	-	-	-	-

Tabella 5.4: Scelta del miglior modello di regressione del quantile per differenti livelli di τ fascia semi oraria 12.00–12.30 stima nel campione *in sample*. Vengono contrassegnati con la "V" i regressori significativi e con "-" le esplicative non significative per test T e F condotti ad un livello di significatività del 1%.

(appendice B) viene riportata la selezione delle esplicative necessarie per definire un modello di regressione del quantile per τ differenti in ciascuna fascia semi oraria. Si può osservare che i modelli ridotti al variare di τ presentino specificazioni differenti in ciascuna fascia semi oraria e che le esplicative siano in numero nettamente ridotto. Emerge però sistematica la presenza di alcune variabili, il prezzo a

ritardato uno e il margine contemporaneo su tutte. Prendendo in considerazione la fascia semi oraria 12.00–12.30, il modello che definisce il quantile condizionato di livello $\tau = 0.5$ (regressione mediana) è

$$Q_{\tau=0.5}(p_t|I_t) = \beta_{0,\tau} + \beta_{1,\tau}p_{t-1} + \beta_{6,\tau}p_{t-6} + \beta_{9,\tau}d_{t-1} + \beta_{10,\tau}m_t + \beta_{11,\tau}m_{t-1}$$

mentre quello per $\tau = 0.9$ è

$$Q_{\tau=0.9}(p_t|I_t) = \beta_{0,\tau} + \beta_{1,\tau}p_{t-1} + \beta_{10,\tau}m_t + \beta_{11,\tau}m_{t-1}$$

È possibile notare una certa regolarità della presenza di alcune esplicative al variare di τ . Un ulteriore passo può essere quello di cercare un modello ridotto unico che si adatti necessariamente bene per stimare e prevedere quantili condizionati dei prezzi per differenti livelli del quantile di fascia elevata ($\tau > 0.75$ ad esempio). Concentrandoci sui modelli del quantile condizionato di livello elevato ($\tau > 0.7$), di tabella 5.4, potremmo notare una certa regolarità nelle esplicative presenti e definire così un modello unico che si possa adattare bene per la fascia semi oraria. Un possibile modello unico definito con questo criterio ad *hoc* è

$$Q_{\tau=0.5}(p_t|I_t) = \beta_{0,\tau} + \beta_{1,\tau}p_{t-1} + \beta_{10,\tau}m_t + \beta_{11,\tau}m_{t-1} \text{ per } \tau \in [0.75, 1]$$

dove le stime dei relativi parametri è riportata in tabella 5.5 e 5.6 per rispettivamente un livello del quantile di $\tau = 0.5$ e $\tau = 0.9$. Medesime osservazioni si

	Stime	Std. Error	Statistica t	p-value	statistica t
(Intercetta)	0.01	0.01	1.51		0.13
p1	0.19	0.04	5.05		0.00
m0	-0.82	0.05	-15.56		0.00
m1	0.14	0.06	2.44		0.01

Tabella 5.5: Stima dei parametri nel modello di QR per livello $\tau = 0.5$ del quantile condizionato, *standard errors* calcolati con metodo *kernel*, statistica t per la significatività di ogni singolo parametro e relativo *p-value* per la fascia semi oraria 12.00–12.30 campione *in sample*.

possono fare per tutte le altre fasce semi orarie osservando la tabella B.4 (Appendice B), si può definire un modello di regressione del quantile con un insieme di esplicative scelte ad *hoc* per livelli del quantile $\tau \in [0.75, 1]$. Ad esempio, per la

	Stime	Std. Error	Statistica t	p-value	statistica t
(Intercetta)	0.29	0.01	23.66		0.00
pX1	0.27	0.06	4.39		0.00
mX0	-1.13	0.07	-15.95		0.00
mX1	0.24	0.07	3.19		0.00

Tabella 5.6: Stima dei parametri nel modello di QR per livello $\tau = 0.9$ del quantile condizionato, *standard errors* calcolati con metodo *kernel*, statistica t per la significatività di ogni singolo parametro e relativo *p-value* per la fascia semi oraria 12.00–12.30 campione *in sample*.

fascia semi oraria delle 18 il modello definito è

$$Q_{\tau=0.5}(p_t|I_t) = \beta_{0,\tau} + \beta_{1,\tau}p_{t-1} + \beta_{10,\tau}m_t$$

dove in questo caso vi è una esplicitiva che differenzia il modello da quello di fascia semi oraria delle 12.00. La scelta di un modello di regressione del quantile con sole esplicative utili per τ elevati è legata ad un approccio parsimonioso e che limiti i problemi di sovradimensionamento del modello.

In figura 5.3 e 5.4 vengono riportate le stime dei parametri per i modelli scelti ad *hoc* proposti al variare di $\tau \in [0,1]$ per le fasce semi orarie rispettivamente delle 12 e delle 18. Si può notare come gli effetti marginali associati alle variabili che ora sono ritenute necessarie abbiano un comportamento molto più variabile al variare di τ . Il modello definito ad *hoc* può fornire stime del quantile condizionato nel campione *in sample* grazie ai parametri stimati, ad esempio per la fascia semi oraria 12.00–12.30, è $\widehat{Q}_\tau(p_t|I_t) = \widehat{\beta}_{0,\tau} + \widehat{\beta}_{1,\tau}p_{t-1} + \widehat{\beta}_{10,\tau}m_t + \widehat{\beta}_{11,\tau}m_{t-1}$, in modo da descrivere l'intera distribuzione condizionata. Così grazie alla proprietà di invarianza alle trasformazioni monotone garantita dal metodo della regressione del quantile, il quantile condizionato dei prezzi può essere definito come

$$\widehat{Q}_\tau(P_t|I_t) = \exp\{\widehat{DP}_t + \widehat{Q}_\tau(p_t|I_t)\}$$

È possibile valutare la validità dei modelli stimati *in sample* per i prezzi basandosi su due proprietà riscontrabili sulla base dei quantili condizionati previsti e i residui di previsione. La prima proprietà da riscontrare è legata alla frazione di P_t che nell'intero campione *in sample* superano il quantile condizionato di livello

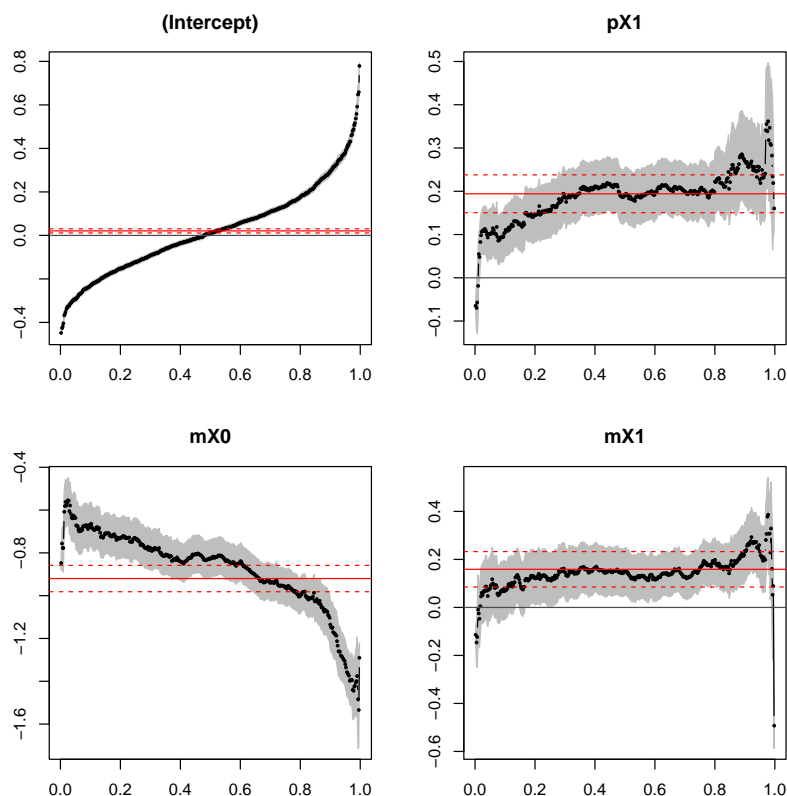


Figura 5.3: Stima dei parametri β_τ del modello di regressione del quantile scelto ad hoc per $\tau \in [0, 1]$ con intervallo di confidenza definito per un livello di copertura del 95% sotto l'ipotesi di normalità con *standard error* stimati non parametricamente con metodo *kernel* per la fascia semi oraria 12.00–12.30 nel campione *in sample*.

τ previsto, questa dovrebbe approssimare $1 - \tau$, ovvero

$$\sum_{t=1}^{1731} \frac{I(P_t > \widehat{Q}_\tau(P_t|I_t))}{1731} \approx 1 - \tau$$

dove $I(P_t > \widehat{Q}_\tau(P_t|I_t)) = 1$ se $P_t > \widehat{Q}_\tau(P_t|I_t)$ zero altrimenti, da comprovare per differenti livelli di τ . La seconda proprietà concerne la verificare che il residuo della regressione di livello τ abbia quantile non condizionato pari a zero, ciò comprova che il modello di regressione del quantile stimi effettivamente il quantile τ -esimo. Se indichiamo con $r_t = p_t - \widehat{Q}_\tau(p_t|I_t)$ il residuo occorre verificare che il quantile τ -esimo della distribuzione di r_t sia pari a zero. In tabella 5.7 e 5.8 si osserva riscontro delle due proprietà che possono garantire una diagnostica del modello di regressione del quantile proposto. Viene garantita una frazione di sforamenti del quantile condizionato previsto in linea a $1 - \tau$ e il quantile della distribuzione non condizionata di livello τ dei residui risulta sempre pressoché

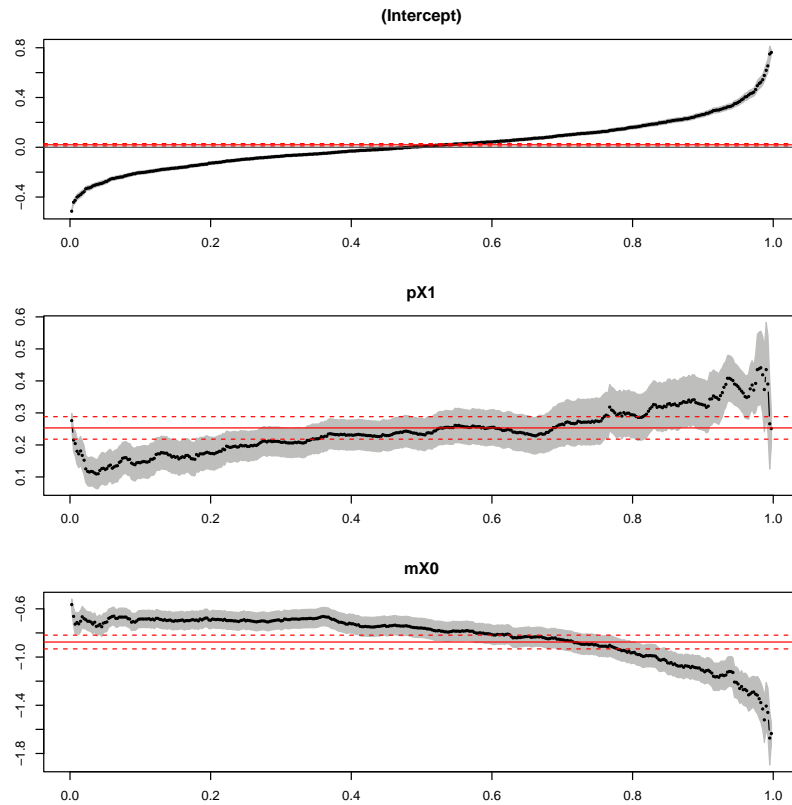


Figura 5.4: Stima dei parametri β_τ del modello di regressione del quantile scelto ad hoc per $\tau \in [0, 1]$ con intervallo di confidenza definito per un livello di copertura del 95% sotto l'ipotesi di normalità con *standard error* stimati non parametricamente con metodo *kernel* per la fascia semi oraria 18.00–18.30 nel campione *in sample*.

Tau	0.75	0.80	0.85	0.88	0.90	0.93	0.95	0.975
Sforamenti	0.500	0.298	0.249	0.200	0.148	0.100	0.050	0.026
Quantile residui	0*	0*	0*	0*	0*	0*	0*	0*

Tabella 5.7: Diagnostica sul modello lineare di QR unico per la fascia semi oraria 12.00–12.30. Per alcuni livelli di τ viene calcolato con "Sforamenti" la frazione che nel totale di P_t supera il quantile condizionato previsto e con "Quantile" si riporta il quantile della distribuzione empirica dei residui di livello τ corrispondente. Con '0*' si indicano valori minori di 0.0001

nullo se τ è elevato (nelle tabelle vengono approssimati i numeri che vedono il primo numero non nullo dopo la quarta cifra decimale). Tali proprietà verificate nel campione *in sample* costituiscono la diagnostica per il modello di regressione del quantile che deve essere applicata a ciascuna fascia semi oraria. Da valutazioni generali non emergono anomalie nella fase di calibrazione dei modelli.

Dati i modelli scelti ad *hoc* (o differenti per ogni livello del quantile) per ogni valore di τ e per ciascuna fascia semi oraria è possibile prevedere ad un passo i

Tau	0.75	0.80	0.85	0.88	0.90	0.93	0.95	0.975
Sforamenti	0.499	0.299	0.250	0.199	0.150	0.100	0.050	0.023
Quantile residui	0*	0*	0*	0*	0*	0*	0*	0*

Tabella 5.8: Diagnostica sul modello lineare di QR unico per la fascia semi oraria 18.00–18.30. Per alcuni livelli di τ viene calcolato con "Sforamenti" la frazione che nel totale di P_t supera il quantile condizionato previsto e con "Quantile" si riporta il quantile della distribuzione empirica dei residui di livello τ corrispondente. Con '0*' si indicano valori minori di 0.0001

quantili condizionati dei prezzi e con questi i picchi *out of sample*.

Per la fascia semi oraria 12.00-12.30, ad esempio, lo schema di previsione ad un passo del quantile condizionato di livello τ della componente stocastica dei logaritmi dei prezzi del giorno successivo è

$$\widehat{Q}_\tau(p_{t+1}|I_t) = \widehat{\beta}_{0,\tau} + \widehat{\beta}_{1,\tau}p_t + \widehat{\beta}_{10,\tau}m_{t+1} + \widehat{\beta}_{11,\tau}m_t \quad \text{per } t = 1731, \dots, 2100,$$

dove i parametri $\widehat{\beta}_{j,\tau}$ vengono aggiornati ad ogni passo. In generale senza escludere particolari esplicative e focalizzarsi su una fascia semi oraria la previsione ad un passo del quantile condizionato della componente stocastica dei logaritmi dei prezzi di livello τ è

$$\widehat{Q}_\tau(p_{t+1}|I_t) = z'_{t+1}\widehat{\beta}_\tau \quad \text{per } t=1731, \dots, 2100,$$

dove con z_{t+1} si indica il vettore delle esplicative, necessarie per il modello di QR, prese all'istante $t + 1$ e con $\widehat{\beta}_\tau$ il vettore dei parametri stimati associato alle esplicative significative. Dalla previsione con uno dei possibili modelli di *quantile regression* di $Q_\tau(p_{t+1}|I_t)$, procediamo alla previsione del quantile condizionato dei prezzi di livello τ per il giorno successivo come

$$\widehat{Q}_\tau(P_{t+1}|I_t) = \exp\{\widehat{DP}_{t+1} + \widehat{Q}_\tau(p_{t+1}|I_t)\} \quad \text{per } t=1731, \dots, 2100 \quad (5.2)$$

dove \widehat{DP}_{t+1} è la previsione della componente deterministica dei logaritmi dei prezzi effettuata al tempo t e $\widehat{Q}_\tau(P_{t+1}|I_t)$ è la previsione del quantile condizionato dei prezzi *day-ahead* di livello τ . L'uguaglianza di equazione (5.2) è verificata per la proprietà di invarianza alle trasformazioni monotone garantita dalla *quantile regression* (Koenker, 2005). Non si è soggetti, quindi, a possibile critica sulla

quantità che si sta stimando con il termine a destra dell'uguaglianza come accadeva invece utilizzando un modello ARX. Il modo in cui si stabilisce la previsione di un picco di prezzo rimane quello discusso precedentemente.

In tabella B.5 e B.6 (Appendice B) si riportano le *performance* predittive per le fasce semi orarie delle 12.00 e delle 18.00 ottenute con il modello scelto ad *hoc* ed unico per ogni livello del quantile sufficientemente elevato considerato. Si nota che per valori elevati al di sopra del 90 – *esimo* quantile condizionato previsto i picchi di prezzo vengano pressoché tutti previsti. Non è possibile fare però osservazioni generali basandoci su poche fasce semi orarie.

Concentriamoci sulle *performance* predittive globali che un modello del quantile scelto ad *hoc* per ciascuna fascia semi oraria può produrre. In tabella 5.9 vengono unificate le *performance* delle previsioni effettuate nelle 48 fasce semi orarie per differenti livelli del quantile condizionato. Il numero di previsioni *out of sample* è di 35040, mentre il numero di picchi che effettivamente si realizzeranno da prevedere è di 311, distribuiti nelle varie fasce semi orarie. Per un livello $\tau = 0.5$

Tau	PR.PIC	PR.NPIC	NPR.PIC	NPR.NPIC	Picchi Previsti %	Falsi Allarmi %
0.5	36	36	275	34693	11.58	50.00
0.75	72	146	239	34583	23.15	66.97
0.8	81	209	230	34520	26.05	72.07
0.85	99	361	212	34368	31.83	78.48
0.9	136	672	175	34057	43.73	83.17
0.93	171	1219	140	33510	54.98	87.70
0.95	199	2009	112	32720	63.99	90.99
0.975	242	4589	69	30140	77.81	94.99

Tabella 5.9: *Performance* globali della previsione per l'intero insieme dei picchi *out of sample*, basate su un modello di regressione del quantile unico scelto ad *hoc* per ciascuna fascia semi oraria.

si utilizza la mediana condizionata dei prezzi per la previsione dei picchi, un approccio simile al modello di regressione classico, che non porta a buoni risultati in termini di capacità predittiva. Vengono in questo modo previsti solo l'11% circa dei picchi, dei 72 segnali di picco previsto il 50% sono falsi allarmi. Si deve evidenziare che sono solo 72 i segnali di picco (0.2% delle previsioni totali).

Prendendo in considerazione i risultati che produce il quantile condizionato di

livello 90%, si può vedere che su un totale di 311 picchi realizzati, ne vengono previsti correttamente 136 (il 55% circa), ma nel totale dei 808 segnali di picco previsto l'87% sono falsi allarmi. Rispetto alla regressione mediana il numero di picchi previsti aumenta esponenzialmente, ma con esso anche il numero dei falsi allarmi. Si deve però evidenziare che su 35040 previsioni vengono segnalati 808 picchi (il 2% delle previsioni totali indistintamente tra corrette e falsi allarmi).

In riferimento alla tabella 5.9, aumentando il livello del quantile condizionato si nota come la capacità previsionale aumenti nettamente ma si presenti il già definito *trade-off* tra numero di previsioni corrette dei picchi e falsi allarmi.

Si vorrebbe stabilire per quale livello del quantile condizionato sia opportuno porre un'analisi mirata per la previsione dei picchi di prezzo, mantenendo un elevato livello di previsioni ma a un livello "ridotto" i falsi allarmi, anche se da un'ispezione semplice delle percentuali di falsi allarmi questo appare a priori complicato. È comunque importante valutare anche quanti segnali di picco produce il metodo per un dato τ sulle 35040 previsioni, per una valutazione di questo si veda l'ultima riga di tabella 5.10.

Sulla base dei risultati proposti in tabella 5.9 proviamo comunque a definire il miglior livello del quantile con il quale prevedere i picchi. Non conoscendo esplicitamente il valore in termini di costo, magari nelle operazioni di *trading* non corrette, derivanti da una previsione di picco errata e allo stesso tempo da una previsione di picco corretta, non si riesce a definire un opportuno peso per picchi correttamente previsti e falsi allarmi. Per evitare elementi di ulteriore soggettività si trattano i picchi correttamente previsti come i falsi allarmi. Si propongono quindi due indici per la valutazione globale delle *performance*. Il primo indice di valutazione delle *performance* è

$$I_1 = \frac{\text{Frazione Previsioni Correttamente}}{\text{Frazione di Falsi Allarmi}} = \frac{\frac{PR.PIC}{PR.PIC+NPR.PIC}}{\frac{PR.NPIC}{PR.NPIC+PR.PIC}}$$

viene definito come un rapporto tra la frazione di previsioni corrette di picco sulla frazione di falsi allarmi. Questo permette di legare i due criteri di valutazione che si stanno analizzando per valutare il modello. Si prendono in proporzione le frazioni di previsioni corrette e falsi allarme in virtù del fatto che queste sono

riferite a totali differenti. Se l'indice assume valore 1 la proporzione di picchi correttamente previsti è uguale alla proporzione di falsi allarmi; quando l'indice supera il valore 1 la percentuale di previsioni corrette domina la percentuale di falsi allarmi, e questo è positivo. Un indice per valutare le *performance* di previsione globali corrette indiscriminatamente per picchi o non picchi è

$$I_2 = \frac{\frac{\text{PicchiCorrettamentePrevisti}}{\text{TotalePicchiOutOfSample}} + \frac{\text{PrevisioniNonPicchiCorrette}}{\text{TotaleNonPicchiOutOfSample}}}{2} 100 \in [0, 100\%]$$

che esprime con una percentuale media la capacità di prevedere correttamente indistintamente un picco o meno; misura che può essere affiancata alla precedente. In entrambi gli indici si sono utilizzate le proporzioni rispetto i relativi totali in modo da omogenizzare la significativa differenza tra il ridotto numero di picchi sull'elevato numero di previsioni effettuate. In generale si vorrebbe scegliere il livello del quantile condizionato usato per prevedere i prezzi che massimizzi i due criteri valutando però anche i falsi allarmi. In tabella 5.10 si riporta al variare del

Tau	0.5	0.75	0.80	0.85	0.9	0.95	0.975
Criterio1	0.23	0.35	0.36	0.41	0.53	0.70	0.82
Criterio2	55.74	61.37	62.72	65.40	70.90	79.10	82.30
% Picchi Previsti	11.58	23.15	26.05	31.83	43.73	63.99	77.81
% Falsi Allarmi	50.00	66.97	72.07	78.48	83.17	90.99	94.99
Numero Segnali Picco	72	218	290	460	808	2208	4831

Tabella 5.10: Criteri di valutazione globali delle *performance* riferita al modello di regressione del quantile per la previsioni dei picchi ad un passo riportate in tabella 5.9.

livello del quantile condizionato utilizzato nelle previsioni riportate in tabella 5.9 il valore dei due criteri di valutazione delle *performance* e le percentuali di picchi previsti e falsi allarmi per discriminare sul livello di τ . Si riporta inoltre il numero assoluto dei segnali di picco previsti nel totale delle 35040, all'aumentare di τ tale numero aumenta esponenzialmente. Per entrambi i criteri proposti la scelta migliore ricade sul quantile condizionato di livello più elevato possibile. Nel caso studiato il quantile prescelto dovrebbe essere quello di livello 97.5, ma aiutandosi con la percentuali dei falsi allarmi e il numero di segnali di picco realizzati dal modello, non appare la scelta migliore. Aumentare di così tanto il livello del quantile per la previsione dei prezzi si evidenzia come i falsi allarmi non possano

essere controllati, allo stesso modo mantenere τ a livelli troppo contenuti porta a una capacità di prevedere i picchi troppo bassa. Un livello di $\tau = 0.8$ potrebbe essere più ragionevole, si prevedono poco meno del 30% dei picchi mantenendo un numero di segnali di picco discretamente controllato nell'insieme delle 35040 previsioni totali (146 falsi allarmi a fronte di 72 picchi previsti correttamente).

Un possibile tentativo per cercare di migliorare tali *performance* di previsione dei picchi è quello di utilizzare, per ciascuna fascia semi oraria, un opportuno modello di regressione specifico per ogni livello del quantile. Le variabili selezionate sono quelle riportate in tabella B.4 (Appendice B). Senza riportare un'analisi per la singola fascia semi oraria, in tabella 5.11 sono riportate le *performance* globali.

Tau	PR.PIC	PR.NPIC	NPR.PIC	NPR.NPIC	Picchi Previsti %	Falsi Allarmi %
0.75	74	148	237	34581	23.79	66.67
0.8	82	211	229	34518	26.37	72.01
0.85	101	362	210	34367	32.48	78.19
0.9	138	676	173	34053	44.37	83.05
0.93	175	1211	136	33518	56.27	87.37
0.95	197	1933	114	32796	63.34	90.75
0.975	241	4467	70	30262	77.49	94.88

Tabella 5.11: *Performance* globali della previsione per l'intero insieme dei picchi *out of sample* 01/01/2009 al 31/12/2010, basate su un modello di regressione del quantile specifico per ogni livello del quantile considerato in ciascuna fascia semi oraria.

Confrontando la tabella 5.9 (*performance* di un unico modello del quantile per ogni τ) con la tabella 5.11 (*performance* di un modello specifico del quantile per ogni τ) si osserva come non vi sia un miglioramento evidente nelle *performance* di previsione di un picco di prezzo né tanto meno una netta riduzione dei falsi allarmi. Uno sforzo nettamente maggiore, comportato dalla specificazione di un modello di regressione specifico per ogni livello τ , non è giustificabile a fronte dei risultati ottenuti che non appaiono migliorare. Date queste evidenze il modello di regressione del quantile specificato per ogni singolo livello del quantile condizionato non produce risultati migliori di quanto fatto da un modello del quantile specificato *ad hoc* per τ elevati.

5.2 Modello di regressione del quantile non parametrico

Una alternativa che potrebbe produrre differenze significative in termini di *performance* predittive dei picchi, attraverso l'uso di metodi di regressione del quantile, è rappresentata da metodi di regressione non parametrici. Valutando la possibilità che la relazione non sia semplicemente lineare e che non sia ipotizzabile una chiara relazione di tipo non lineare, tentiamo di definire opportuni modelli del quantile non parametrici. Per mantenere la massima generalità si vorrebbe definire un modello dove il quantile condizionato della componente stocastica dei logaritmi dei prezzi è

$$Q_{\tau}(p_t|I_t) = f(p_{t-1}, \dots, p_{t-7}, d_t, d_{t-1}, m_t, m_{t-1}, dm_t, dm_{t-1}) \quad (5.3)$$

deve $f()$ è la funzione del quantile condizionato di livello τ da stimare con un opportuno metodo di regressione non parametrico e le serie utilizzate come esplicative sono sempre le stesse. Questa specificazione totalmente non parametrica è difficilmente perseguibile per problemi legati alla dimensionalità, un così elevato numero di esplicative pone problemi significativi di stima. Ad un modello di regressione del quantile totalmente non parametrico viene preferito un modello additivo non parametrico che, seppur perdendo in generalità, consenta di essere stimato e ottenere un modello con sole esplicative necessarie per ciascuna fascia semi oraria in modo più semplice. Il modello di regressione del quantile non parametrico più completo in termine di esplicative applicabile sulla singola fascia semi oraria per qualsiasi livello del quantile condizionato è

$$p_t = \alpha + \sum_{j=1}^7 f_j(p_{t-j}) + f_8(d_t) + f_9(d_{t-1}) + f_{10}(m_t) + f_{11}(m_{t-1}) + f_{12}(dm_t) + f_{13}(dm_{t-1}) + \varepsilon_t \quad (5.4)$$

dove l'ipotesi del modello di regressione del quantile di livello τ sull'errore rimane $Q_{\tau}(\varepsilon_t|I_t) = 0$, ipotesi poco ragionevole ma necessaria affinché la componente deterministica del modello sia intesa come il quantile condizionato di livello τ di p_t . Questo modello è meno restrittivo del modello di regressione del quantile lineare, che assume tutte le funzioni di regressione parziale siano funzioni lineari. In questo caso per problemi di natura computazionale legati alla lunghezza

delle serie e di stima si è deciso di stimare ogni singola componente $f_j()$ attraverso il metodo delle *spline* di regressione, si veda sezione 2.2.1. In questo modo si definisce come funzione non parametrica, associata ad ogni singola esplicativa, un polinomio di grado 3 (uso di *spline* cubiche lineari) da cui si stimano i parametri associati del polinomio. Il numero di nodi da considerare viene definito opportunamente per ogni singola fascia semi oraria e per ogni singola esplicativa mediante una regressione lineare non parametrica della risposta su ciascuna delle esplicative. Per maggiori dettagli sulla formalizzazione di un modello di regressione additivo del quantile non parametrico e l'uso delle *spline* di regressione si veda Koenker (2005) e Wang (2011).

Per evitare di gestire un modello di regressione troppo complesso e generale, si ottiene per ciascuna fascia semi oraria e per ciascun livello del quantile condizionato valutato sufficientemente elevato ($\tau > 0.8$) un modello di regressione additivo non parametrico con solo regressori significativi e utili per definire un buon modello di regressione del quantile. La procedura implementata è simile a quella di eliminazione delle variabili non significative in un modello di regressione lineare del quantile. Essendo però i singoli regressori trattati in modo non parametrico mediante *spline* cubiche (o di grado superiore) non è possibile testare la significatività mediante un test T su un singolo parametro (esempio quello associato al termine di grado più elevato), ma bisogna considerare la significatività congiunta di tutti i parametri associati a ciascuna esplicativa, mediante test per modelli annidati di *quantile regression*, analoghi al test F (Koenker, 2010b). Conducendo un test al livello di significatività del 1%, si eliminano passo passo le esplicative meno significative ottenendo via via un modello con i soli regressori necessari, che a sua volta dovrà adattarsi bene ai dati.

È possibile riportare una tabella simile a quella per il modello di regressione del quantile lineare dove vengono evidenziate per ciascuna fascia semi oraria e per i livelli del quantile elevati i regressori significativi per definire un modello in grado di adattarsi bene ai dati, si veda tabella B.7 in appendice. La tabella 5.2 (che rappresenta quanto fatto per ciascuna fascia semi oraria, estratta dalla tabella B.7)

mostra le esplicative necessarie per definire uno specifico modello non parametrico del quantile per alcuni livelli del quantile condizionato per la fascia semi oraria delle 12.00–12.30. Si può osservare che al variare del livello del quantile condizionato per ogni fascia semi oraria ci siano cambiamenti nel *set* di esplicative utili, ma allo stesso tempo si nota una certa regolarità per la presenza di alcuni esplicative per i livelli del quantile elevati dentro la finestra rappresentata (e così in generale per le altre 47 fasce semi orarie). Per questa ragione è possibile

Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
0.8	V	V	V	V	V	V	-	-	-	V	V	-	-
0.85	V	-	V	-	V	V	-	-	-	-	-	-	-
0.9	V	V	V	V	V	V	-	V	V	V	V	V	V
0.95	V	-	V	V	V	V	V	V	V	-	-	-	V

Tabella 5.12: Con "V" si contrassegnano le variabili esplicative necessarie per definire un modello di regressione del quantile additivo non parametrico per alcuni τ , fascia semi oraria 12.00–12.30.

definire sulla base di una valutazione soggettiva, legata ad una valutazione ad *hoc* per ciascuna fascia semi oraria e livelli del quantile sufficientemente elevati, un modello di regressione del quantile non parametrico con un insieme unico di poche esplicative che garantisca anche una certa parsimonia. Si propone quindi un modello di regressione del quantile non parametrico unico definito ad *hoc* per ciascuna fascia semi oraria che viene stimato nel campione *in sample* (e aggiornato passo passo per le previsioni *out of sample*; esattamente come fatto nei modelli di *quantile regression* lineare), ad esempio il modello scelto per la fascia semi oraria 12.00–12.30 è

$$Q_{\tau}(p_t|I_t) = f_1(p_{t-1}) + f_2(m_t) + f_3(d_{t-1}) + f_4(m_{t-1}) + f_5(p_{t-2}) \text{ per } \tau \in (0.7, 1)$$

Una volta stimato il modello nel campione *in sample*, si possono operare opportune previsioni ad un passo nella componente stocastica *out of sample*. Si ottengono le previsioni dei quantili condizionati dei prezzi *day-ahead* esattamente come fatto nei casi precedenti. In generale, per un modello scelto ad *hoc* per ciascuna fascia semi oraria la previsione del quantile di livello τ per il giorno successivo è data

da

$$\widehat{Q}_\tau(p_t + 1|I_t) = \sum_{j=1}^K \widehat{f}_j(z_{j,t+1})$$

dove, ipotizzando che siano K le esplicative necessarie, gli $z_{j,t+1}$ rappresentano i regressori significativi del modello utilizzati al tempo $t+1$ per la previsione ad un passo non parametrica effettuata mediante le stime delle singole *spline* di regressione $\widehat{f}_j(\cdot)$. La stima del quantile condizionato dei prezzi *day-ahead* e la previsione del picco avviene esattamente come nei casi precedenti. In tabella 5.13 vengono

Tau	PR.PIC	PR.NPIC	NPR.PIC	NPR.NPIC	Picchi Previsti %	Falsi Allarmi %
0.5	47	41	264	34688	15.11	46.59
0.75	83	162	228	34567	26.69	66.12
0.8	93	240	218	34489	29.90	72.07
0.85	109	396	202	34333	35.05	78.42
0.9	132	705	179	34024	42.44	84.23
0.93	164	1115	147	33614	52.73	87.18
0.95	195	1692	116	33037	62.70	89.67
0.975	249	4006	62	30723	80.06	94.15

Tabella 5.13: *Performance* globali della previsione per l'intero insieme dei picchi *out of sample*, basate su un modello additivo di regressione non parametrico del quantile definito con una scelta delle esplicative specifica per ciascuna fascia semi oraria.

riportati le *performance* di previsione globali definendo un modello di regressione non parametrico specifico per ogni fasci semi oraria. Si riscontrano sempre i principali elementi valutati in precedenza sui picchi previsti e i falsi allarmi. Confrontando i risultati con quelli di tabella 5.9, che si riferisce alle previsioni con metodo di regressione del quantile lineare, si evidenzia un comportamento che merita una considerazione. Nonostante le percentuali di falsi allarmi nonché di previsioni corrette non siano (se non in alcuni) casi migliorate, si osserva come il modello non parametrico generi un numero sensibilmente differente di segnali di picco. In realtà confrontando le percentuali di picchi correttamente previsti tra modello di regressione lineare del quantile e modello non parametrico queste non migliorano sensibilmente anzi, hanno una leggera flessione per alcuni livelli del quantile condizionato fatta eccezione del livello $\tau = 0.975$. I falsi allarmi rimangono ancora elevati e pressoché invariati rispetto ai casi precedenti, questo è

un elemento a forte sfavore del modello, anche se questo prevede anche l'80% di picchi dei prezzi.

Si sono effettuate previsioni dei picchi ricorrendo a un modello non parametrico definito ad *hoc* per ciascun livello del quantile da utilizzare in ciascuna fascia semi oraria, ma i risultati globali non si scostano molto da quelli di tabella 5.13.

5.3 Scelta e valutazioni dei modelli

Ai fini della valutazione dei modelli per la previsione dei picchi, si riassumono 3 delle possibili proposte fatte tra il capitolo 3 e 5 che ci permetteranno di fornire le valutazioni complessive. Si mettono a confronto per due differenti livelli del quantile (quelli apparsi più ragionevoli in termini di bilanciamento tra previsioni corrette e falsi allarmi): il metodo di previsione basato sul modello di regressione lineare classico e definizione dei quantili sulla base dell'ipotesi di normalità (*ARX*), il metodo di regressione del quantile lineare (*QR*) e il metodo di regressione del quantile non parametrico (*QR.NP*). Tali modelli sono quelli già introdotti e presentati precedentemente e si riferiscono ad un'unica specificazione del modello di regressione per ciascuna fascia semi oraria (per il modello di regressione lineare classico questa è una scelta obbligata). Le seguenti tabelle sono ricavabili da quelle che presentano le performance globali dei metodi che confrontiamo, si veda tabella 3.13, 5.9 e 5.13. Alcuni ragionamenti sono ricavati dalle osservazioni fatte sulle performance globali. In tabella 5.14 e 5.15 vengono

Modello	PR.PIC	Segnali Picco	Picchi previsti %	Falsi Allarmi %
ARX	89	334	28.62	73.35
QR	81	290	26.05	72.07
QR.NP	93	333	29.90	72.07

Tabella 5.14: *Performance* di tre modelli differenti nella previsione dei picchi di prezzo utilizzando una previsione del quantile del prezzo *day ahead* di livello 80%. PR.PIC il numero nel totale dei 311 picchi previsti correttamente, Segnali Picco quanti segnali di picco il modello produce nelle 35040 previsioni ad un passo, Picchi previsti % e Falsi Allarmi % le percentuali dei picchi correttamente previsti nel totale e di falsi allarmi sul totale dei segnali di picco.

riportate le caratteristiche principali al fine di giudicare la capacità di previsione dei picchi dei tre modelli presi in esame per due differenti livelli del quantile con-

Modello	PR.PIC	Segnali Picco	Picchi previsti %	Falsi Allarmi %
ARX	107	465	34.41	76.99
QR	99	460	31.83	78.48
QR.NP	109	505	35.05	78.42

Tabella 5.15: *Performance* di tre modelli differenti nella previsione dei picchi di prezzo utilizzando una previsione del quantile del prezzo *day ahead* di livello 85%.

dizionato dei prezzi *day-ahead* previsti, rispettivamente il quantile di livello 80% e 85%. Per ogni modello di previsione, l'aumento del livello del quantile condizionato utilizzato per prevedere i picchi aumenta la capacità di prevedere un numero maggiore di picchi, ma nel complesso fa aumentare i falsi segnali, ovvero previsioni di picco che non si realizzeranno. In via più generica un aumento del livello del quantile fa aumentare i segnali di picco (indistintamente se realizzati o meno) nel totale delle previsioni.

É però possibile notare che i modelli di regressione del quantile producono, a parità di livello del quantile condizionato impiegato, risultati migliori in termini di capacità di prevedere i picchi rispetto al modello di regressione lineare.

Presa a riferimento la tabella 5.14 (modelli che sfruttano una previsione del quantile condizionato dei prezzi *day-ahead* di livello 80%) si può osservare come il modello del quantile e la sua specificazione non parametrica, siano sempre migliori di un modello ARX. Infatti un modello del quantile lineare prevede pressoché lo stesso numero di picchi di quello ARX ma con un numero di segnali di picco nel totale delle previsioni nettamente inferiori (44 falsi allarmi in meno). Il modello di regressione non parametrico del quantile prevede addirittura più picchi e meno falsi allarmi di quello ARX. Dalla tabella 5.14, nel migliore dei casi, i picchi di prezzo correttamente previsti possono essere 1 su 3 (93 su 311) se si utilizza il modello di regressione non parametrico del quantile, producendo 240 falsi segnali di picco in un totale di 35040 previsioni. Aumentando il livello del quantile dei prezzi previsto, i picchi possono essere previsti anche fino ad un 80% dai modelli di regressione non parametrici del quantile con $\tau = 0.975$, ma questo viene pagato da un aumento sproporzionato di segnali di picco che non si realizzeranno (anche fino all'12% delle 35040 previsioni totali). Il problema è

sempre emerso ed è legato alle percentuali di falsi allarmi nel totale dei segnali di picchi. Per qualsivoglia livello di τ sensatamente elevato per la previsione dei picchi, la percentuale di falsi allarmi non si pone mai al di sotto del 70% e questo rappresenta un serio limite applicativo, si veda tabella 5.14.

I falsi allarmi però tendono ad aumentare in modo esponenziale, come appunto nel complesso i segnali di picco previsto (indistintamente tra corrette previsioni e falsi allarmi che si muovono insieme), all'aumentare del livello del quantile condizionato stimato. Questa osservazione porta alla valutazione che i metodi proposti sono incapaci di discriminare in fase identificativa del picco, caratteristica che ci si aspetterebbe da ogni modello di previsione.

In definitiva si è osservato che il metodo di regressione del quantile produce sistematicamente un numero di previsioni dei picchi dei prezzi più elevato di quelli proposti con il metodo di regressione lineare ma, è sottoposto come questi ad un numero di falsi allarmi troppo elevato. Questo comporta che non sia giustificabile un opportuno impiego di questi metodi, almeno in questo mercato e in questa fascia temporale. Si deve affermare che la previsione di quantili condizionati con questi metodi non produce risultati soddisfacenti in questo particolare contesto.

L'obiettivo che si sarebbe voluto ottenere era quello di riuscire ad individuare un opportuno livello del quantile condizionato dalla serie dei prezzi previsti per il giorno successivo, capace di fornire un'elevata percentuale di previsioni di picco mantenendo però a un livello contenuto il numero di falsi allarmi. Purtroppo però la maggioranza di previsioni di picco risulta errata e diviene incontrollata per livelli del quantile sempre più elevati.

La motivazione per cui non si è riuscito ad ottenere una corretta anticipazione dei picchi è legata probabilmente alla difficoltà di cogliere il reale meccanismo che li genera. L'elevato numero di falsi allarmi non è attribuibile ad una scorretta stima della componente deterministica soprattutto in fase di previsione. Definire i picchi direttamente sulle componenti stocastiche (non tendo conto di un eventuale effetto della componente deterministica) non produce effetti differenti da

quanto riportato per le previsioni dei prezzi.

Da un'analisi approfondita per alcune serie, le previsioni dei prezzi anomale che generano falsi allarmi sono imputate ai valori pressoché sempre riconducibili al margine. Questo può essere spiegato dal fatto che il margine (ma anche la domanda) è una serie prevista e non realizzata, e questo pone un'elevata incertezza difficilmente gestibile nelle nostre analisi.

Un elemento che risulta fondamentale è comunque il modo in cui si sono individuati i picchi dei prezzi nella serie reale, attraverso un certo quantile della componente stocastica dei prezzi, che ha un duplice ruolo di definire la posizione dei picchi, ma anche stabilire la soglia utilizzata per definire la previsione di un picco di prezzo. Il fatto di definire i picchi in un altro modo potrebbe cambiare anche di molto le nostre valutazioni.

CONCLUSIONI

In questa tesi è stato considerato uno dei problemi più interessanti e dibattuti nella letteratura legata ai mercati elettrici: la previsione dei picchi dei prezzi. Lo studio è stato incentrato su un mercato in particolare: la borsa elettrica Inglese. Si è provato a verificare se la previsione dei quantili dei prezzi potesse essere uno strumento utile per la previsione dei picchi che, per loro natura, sono eventi estremi nella serie dei prezzi. Data la necessità di prevedere quantili condizionati è stata utilizzata la regressione del quantile introdotta da Koenker e Basset (1978). Dopo una presentazione sui fatti stilizzati dei mercati elettrici in generale, si sono presentate le caratteristiche della borsa elettrica Inglese su cui si è incentrata l'analisi di identificazione e previsione dei picchi. È emerso subito il problema dell'individuazione dei picchi nella stessa serie dei prezzi osservati e la definizione di una regola valida per ciascuna fascia semi oraria. Tutt'ora anche questo argomento è oggetto di dibattito nella letteratura dei mercati elettrici, e in questa stessa tesi ha avuto una grande rilevanza. È stato stabilito che un picco possa essere identificato quando il prezzo superi un opportuna soglia, soglia che tornerà utile anche nella fase di previsione dei picchi stessi.

Il problema fondamentale trattato ha riguardato la previsione di un picco di prezzo per il giorno successivo. È stata stabilita una previsione di picco per il giorno successivo quando il prezzo previsto supera la soglia che definisce lo stesso. Emerge che previsioni medie o mediane dei prezzi non sono utili a prevedere i picchi. È quindi stata applicata l'idea di prevedere quantili condizionati dei prezzi *day-ahead* e valutare se questi superino la soglia. Sono stati proposti e confrontati tre modelli differenti per la previsione di questo particolare tipo di prezzi. Il primo è un modello basato sul modello di regressione lineare classico

(adattato opportunamente per prevedere quantili), il secondo è un modello di regressione lineare del quantile e infine un modello di regressione non parametrico del quantile.

In generale la capacità di prevedere correttamente il maggiore numero di picchi appare legata alla scelta del livello del quantile dei prezzi previsti. Tanto più elevato è il livello tanto maggiore sarà il numero dei picchi correttamente previsti. Emerge però che con questo approccio aumenti esponenzialmente anche il numero di segnali di picco che sono falsi allarmi. Si evidenzia quindi il *trade off* tra picchi correttamente previsti e falsi segnali che in nessun modo è stato risolto opportunamente. Picchi previsti e falsi allarmi si muovono nella stessa direzione. Indipendentemente dal modello utilizzato non è stato possibile massimizzare il numero di picchi correttamente previsti mantenendo sotto controllo il numero di falsi segnali che domina sempre le previsioni prodotte. I risultati evidenziano come sistematicamente i modelli di regressione del quantile producano risultati migliori in un'ottica legata al *trade off* esposto, rispetto ai modelli basati sulla regressione lineare. Nonostante il numero di segnali di picco sia contenuto nel totale delle previsioni effettuate, non è però accettabile tollerare quote di falsi allarmi del 70% o più.

È apparso difficile includere nei modelli di regressione il reale meccanismo anticipatore dei picchi. Le serie di domanda e margine, che dovevano rappresentare gli elementi esogeni necessari a questo scopo, non sono state realmente utili. Una giustificazione a questo è legata al fatto che domanda e margine sono serie previste e non effettivamente realizzate. Da uno studio mirato emerge che è soprattutto imputabile all'esplicativa margine buona parte dei falsi allarmi, dovuti a previsioni di prezzi anomali. Si deve quindi concludere che il metodo proposto non è applicabile per la previsione dei picchi nel mercato elettrico Inglese nella particolare fascia temporale analizzata.

Elemento da non dimenticare è stato quello di aver dovuto gestire 48 fasce semi orarie simili tra loro, ma che presentano caratteristiche intrinseche e uniche. Le scelte svolte sono state laboriose in quanto hanno sempre richiesto un

bilanciamento tra generalità dei metodi ma valutazione di ciascun singolo caso.

Appendice A

IMMAGINI

Sono riportate alcune immagini salienti e utili che per brevità sono omesse dalla trattazione svolta.

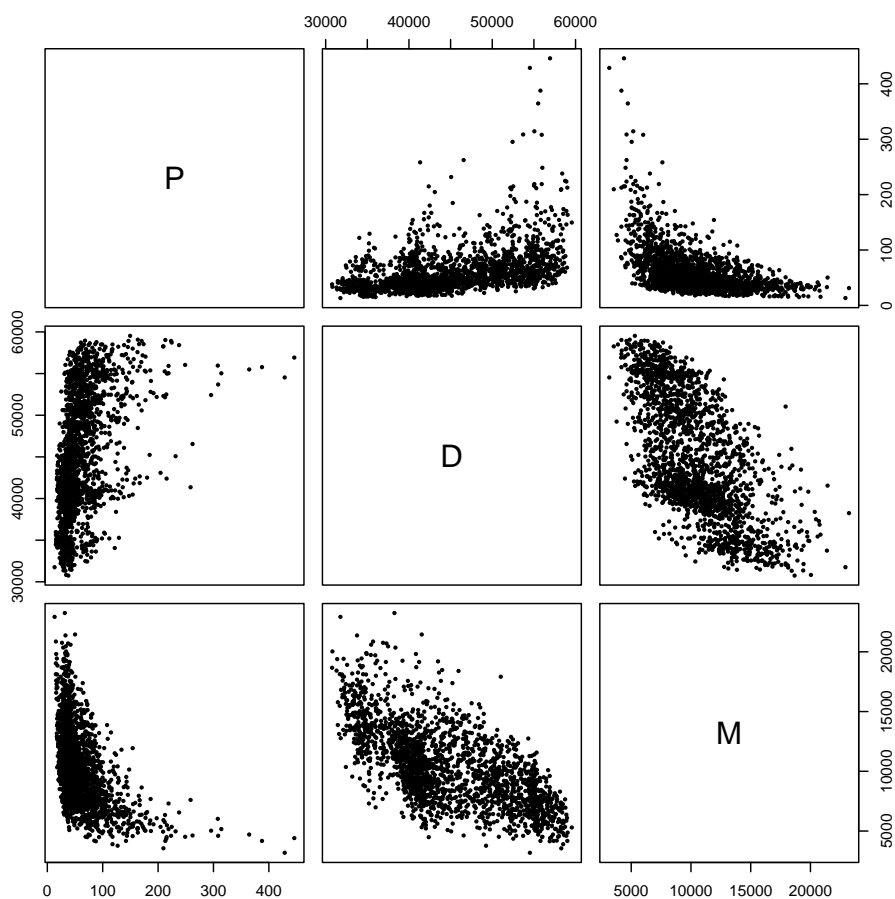


Figura A.1: Relazione tra prezzi domanda e margine del mercato elettrico Inglese nella fascia semi oraria 18.00–18.30 per il periodo 01/04/2005 al 31/12/2010.

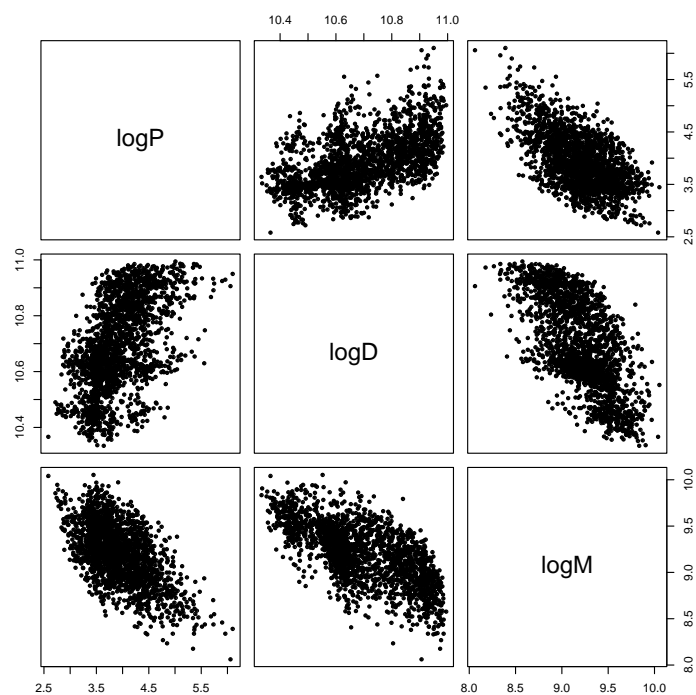


Figura A.2: Relazione tra logaritmi di prezzi domanda e margine del mercato elettrico Inglese nella fascia semi oraria 18.00–18.30 per il periodo 01/04/2005 al 31/12/2009.

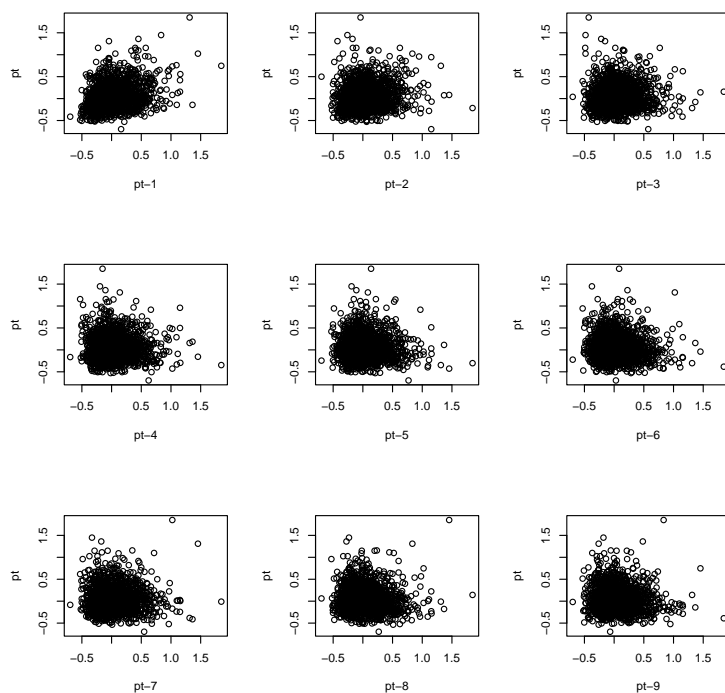


Figura A.3: Relazione tra logaritmi dei prezzi e rispettivi ritardi del mercato elettrico Inglese nella fascia semi oraria 12.00–12.30 per il periodo 01/04/2005 al 31/12/2010.

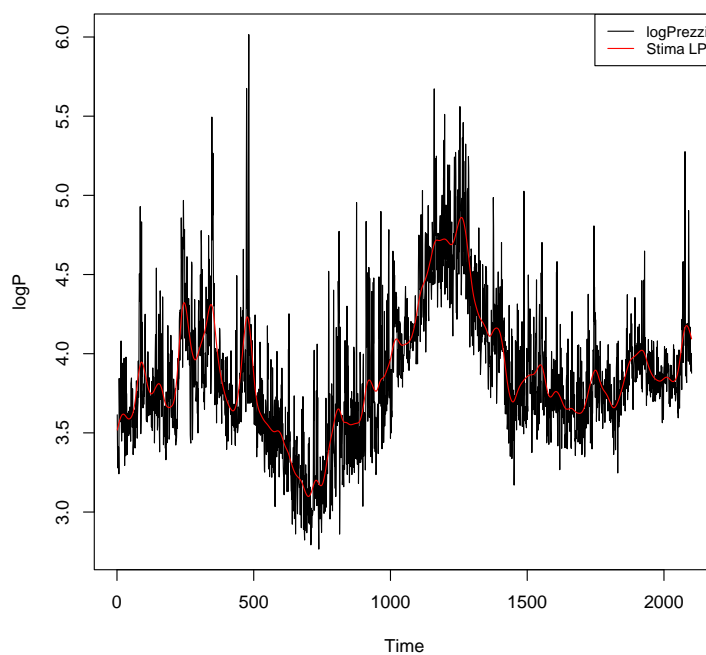


Figura A.4: Logaritmi dei Prezzi e stima della componente LP_t del mercato elettrico Inglese nella fascia semi oraria 12.00–12.30 per il periodo 01/04/2005 al 31/12/2009.

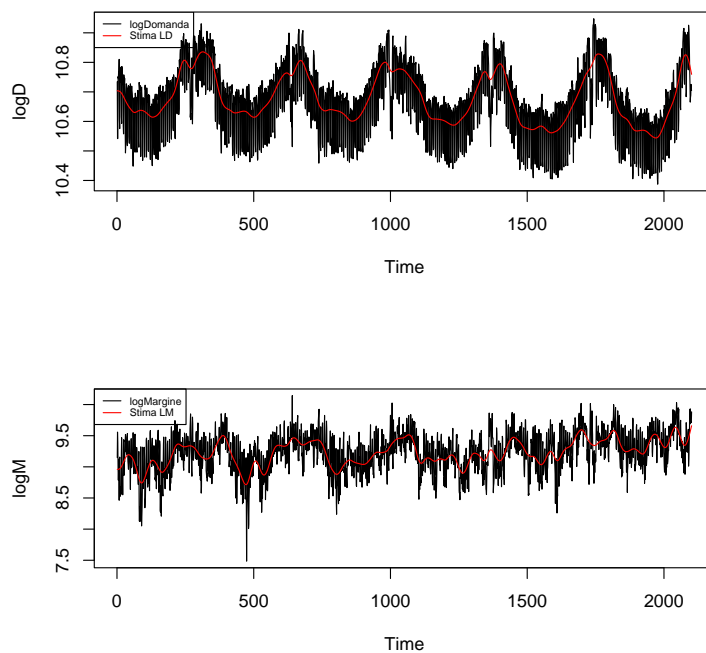


Figura A.5: In alto logaritmo della domanda e stima della componente LD_t , in basso logaritmo del margine e stima della componente LM_t nel mercato elettrico Inglese nella fascia semi oraria 12.00–12.30 per il periodo 01/04/2005 al 31/12/2010.

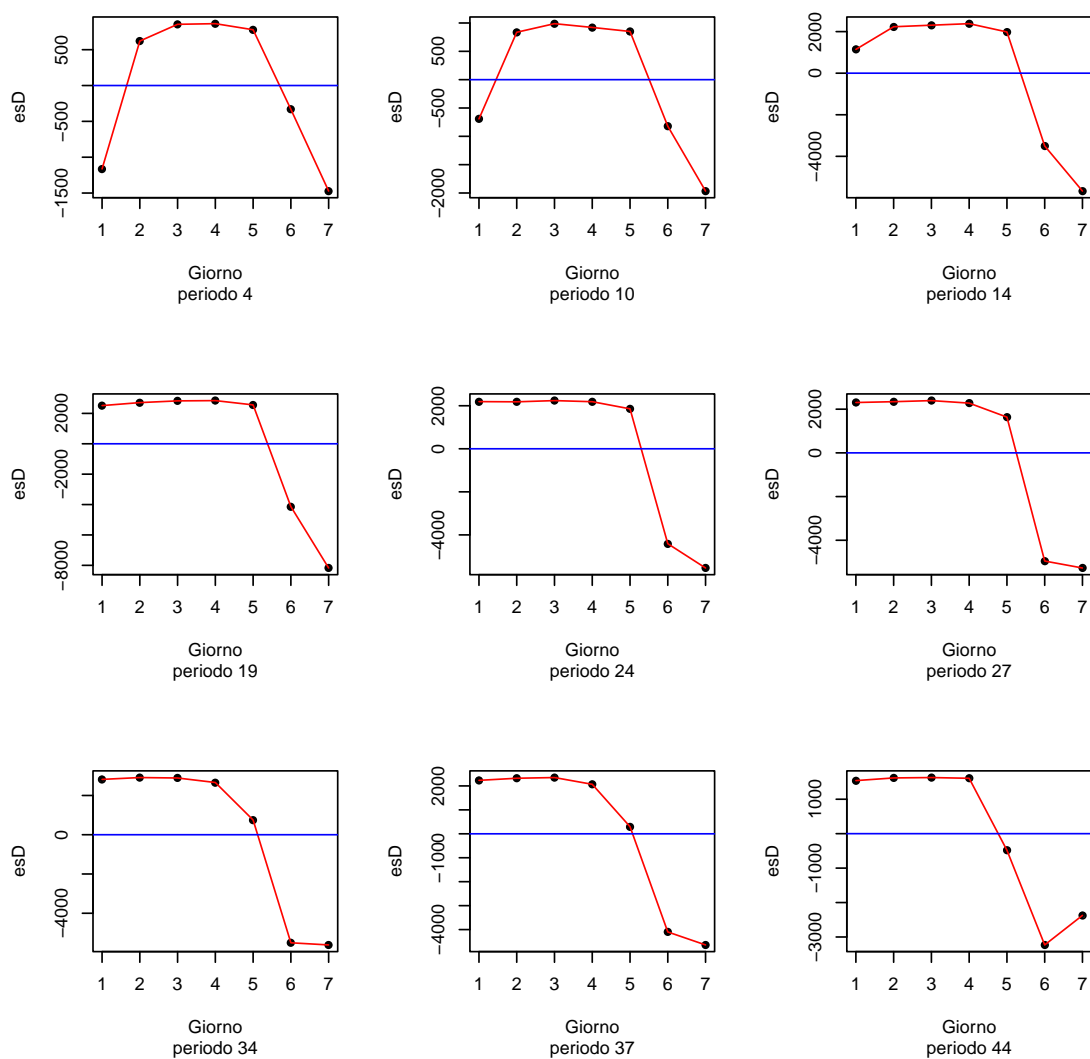


Figura A.6: Effetto stagionalità settimanale sulla Domanda per alcune fasce semi orarie del mercato Inglese nel periodo 01/04/2005 al 31/12/2010. Si mostra l'effetto stimato dei vari giorni (1-Lunedì: $\hat{\gamma}_1$, 2-Martedì: $\hat{\gamma}_2$, ..., 7-Domenica: $\hat{\gamma}_7$) su alcune fasce semi oraria per la serie dei prezzi. La linea blu evidenzia un effetto nullo.

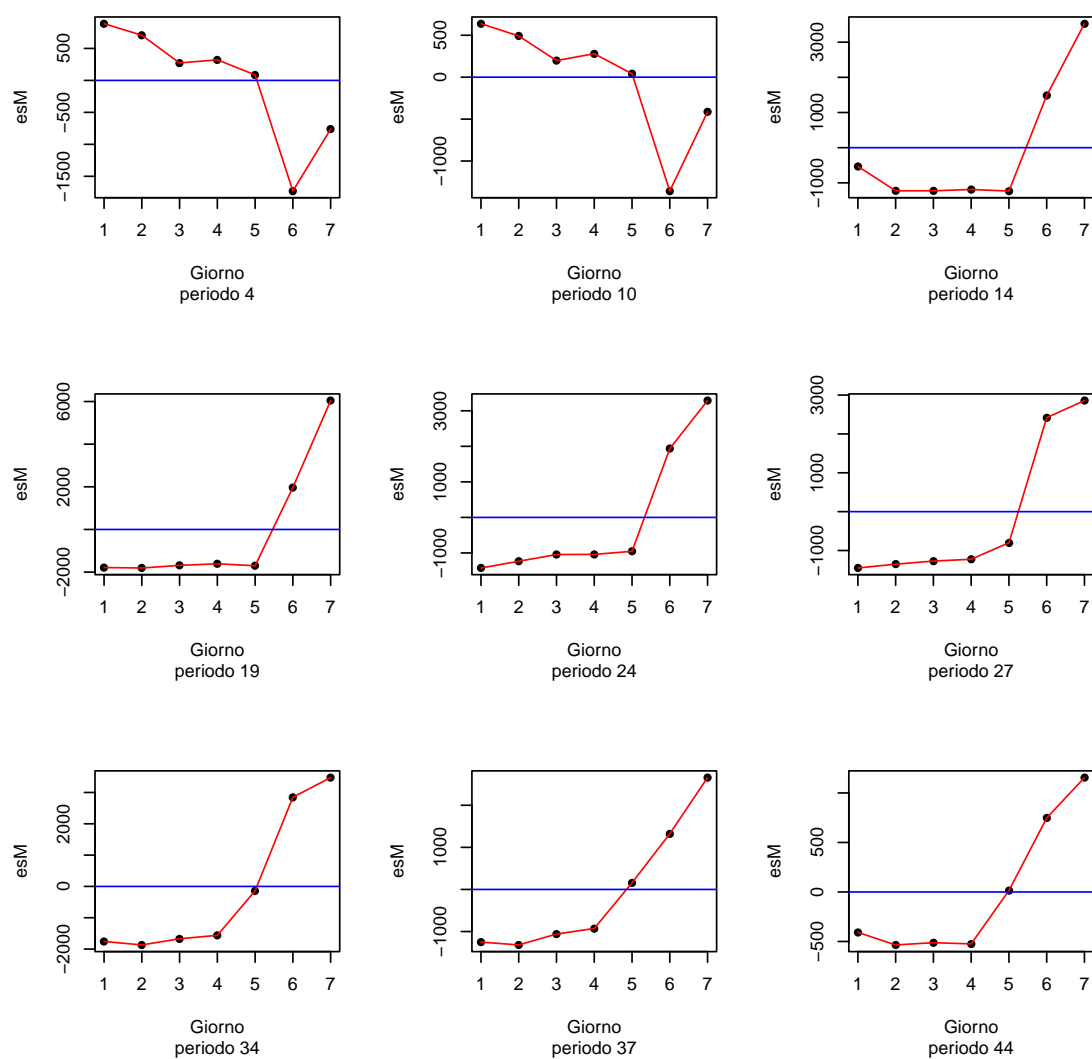


Figura A.7: Effetto stagionalità settimanale sul Margine per alcune fasce semi orarie del mercato Inglese nel periodo 01/04/2005 al 31/12/2010. Si mostra l'effetto stimato dei vari giorni (1-Lunedì : $\hat{\gamma}_1$, 2-Martedì: $\hat{\gamma}_2$, ..., 7-Domenica: $\hat{\gamma}_7$) su alcune fasce semi oraria per la serie dei prezzi. La linea blu evidenzia un effetto nullo.

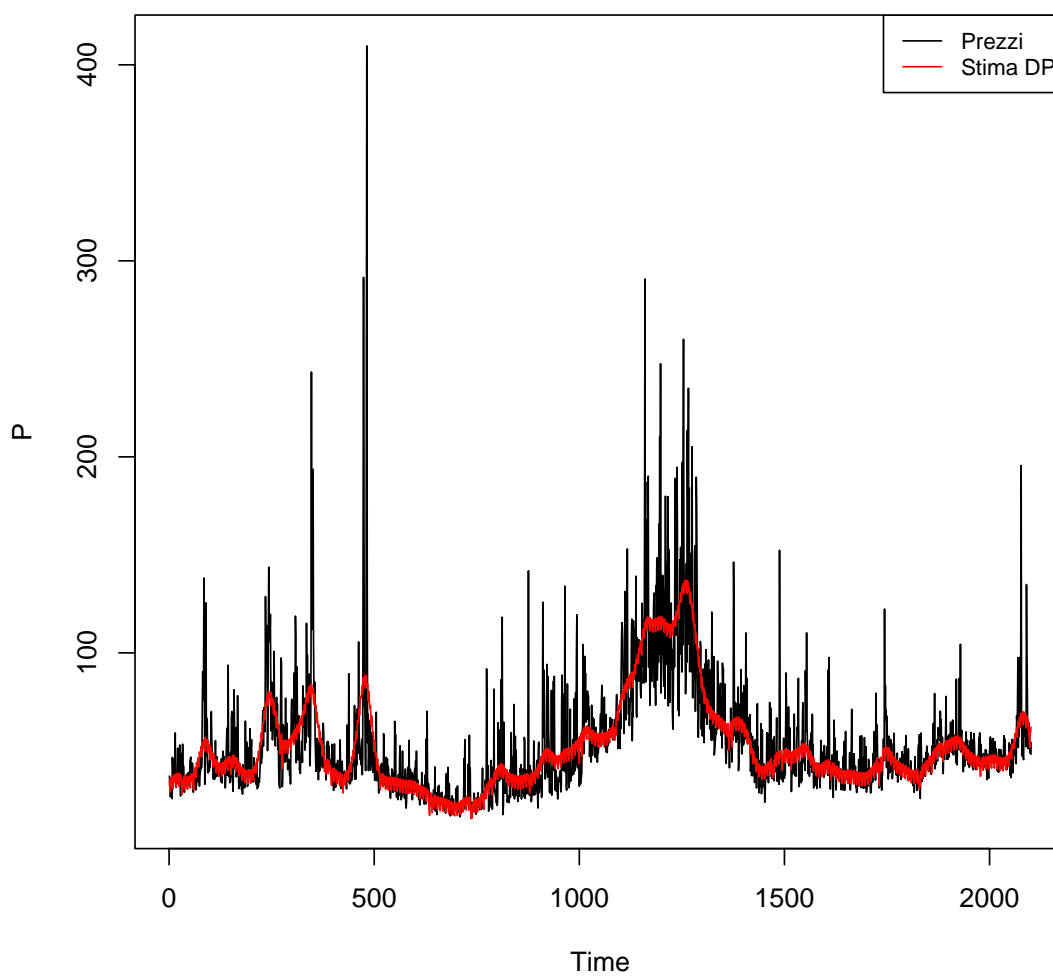


Figura A.8: Stima della componente deterministica per la serie dei prezzi sovrapposta alla relativa serie degli stessi nella fascia semi oraria 12.00–12.30 del mercato Inglese per il periodo 01/04/2005 al 31/12/2010.

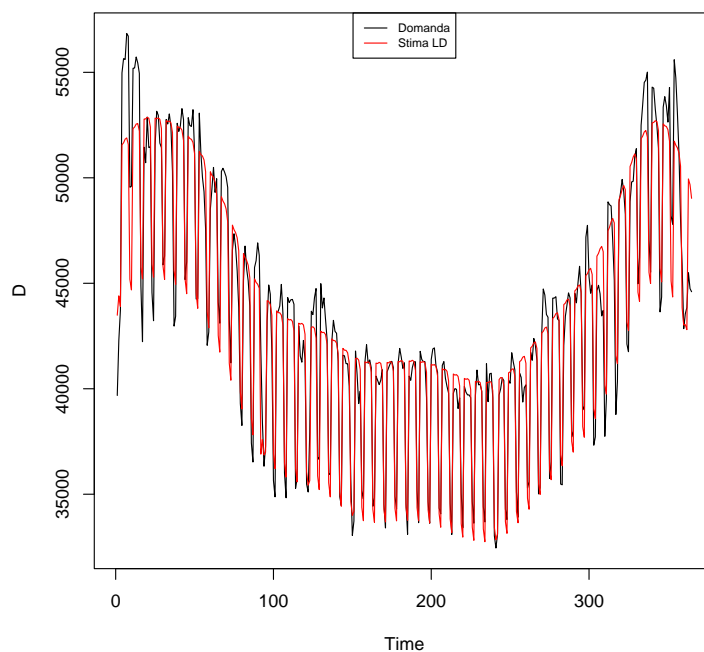


Figura A.9: Stima della componente deterministica per la serie della domanda e sovrapposizione alla serie stessa per la fascia semi oraria 12.00–12.30 del mercato Inglese per il periodo 01/01/2010 al 31/12/2010. Nota: vi è una forte componente periodica settimanale e un trend ben definito.

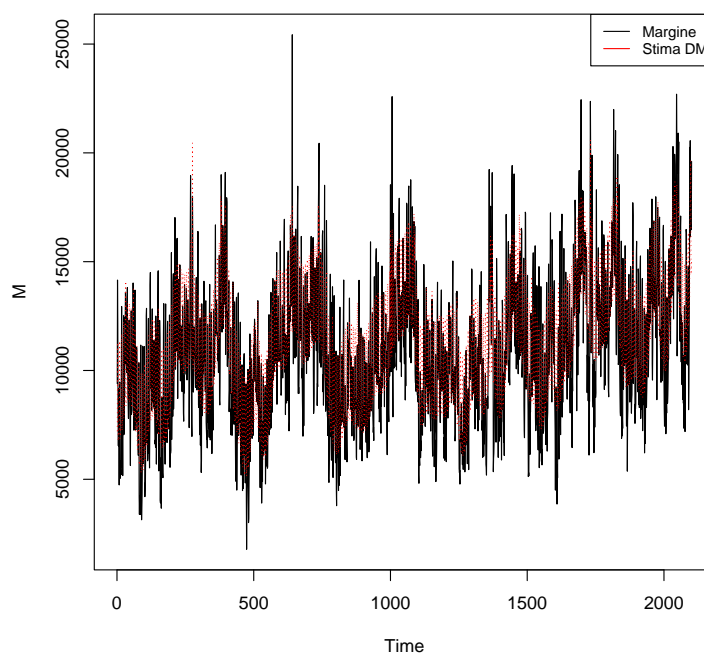


Figura A.10: Stima della componente deterministica per la serie del margine nella fascia semi oraria 12.00–12.30 del mercato Inglese per il periodo 01/04/2005 al 31/12/2010.

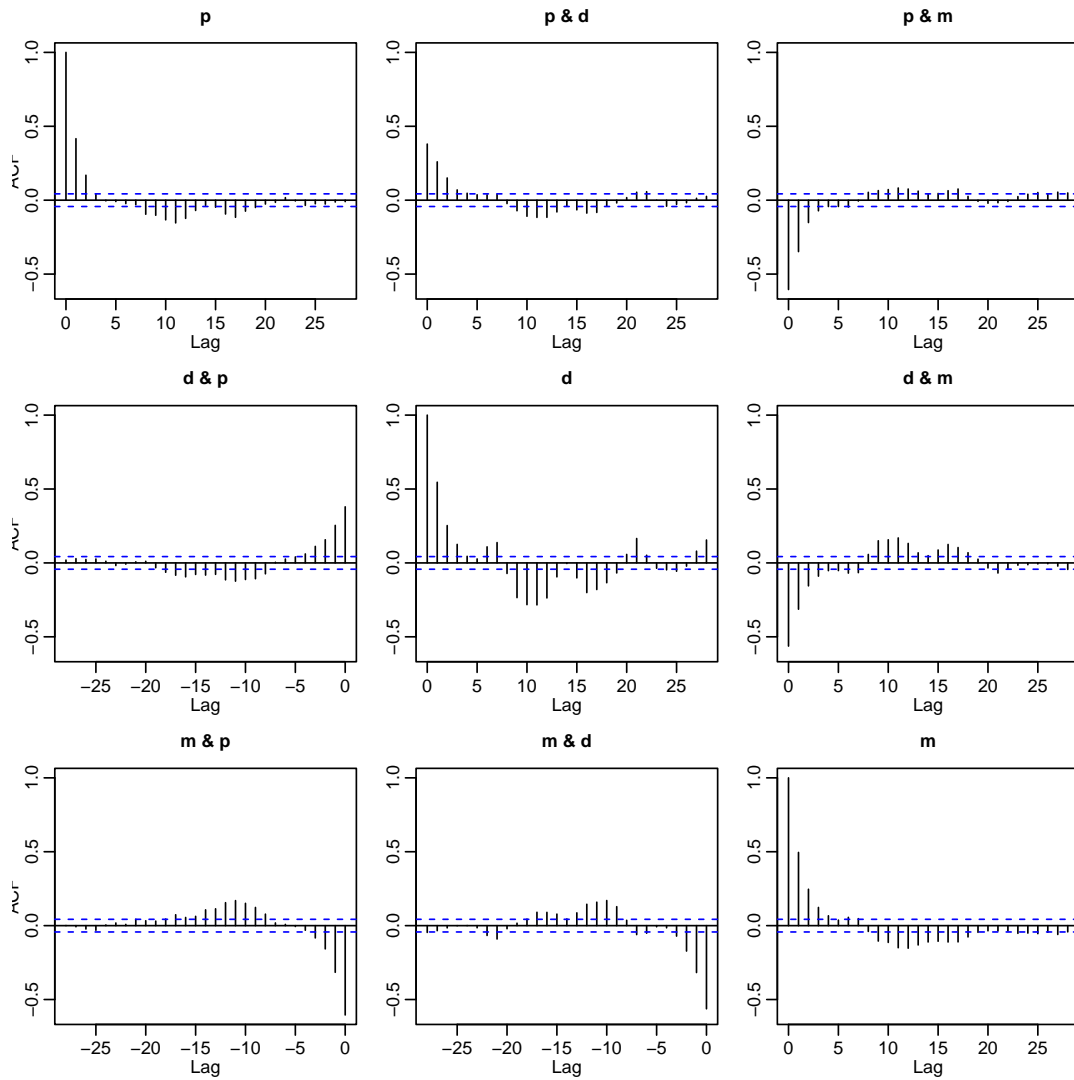


Figura A.11: Diagramma delle funzioni di autocorrelazione (diagonale) e autocorrelazione incrociate per le componenti stocastiche dei logaritmi di prezzo, domanda e margine del mercato elettrico Inglese nella fascia semi oraria 18.00–18.30 per il periodo 01/04/2005 al 31/12/2010.

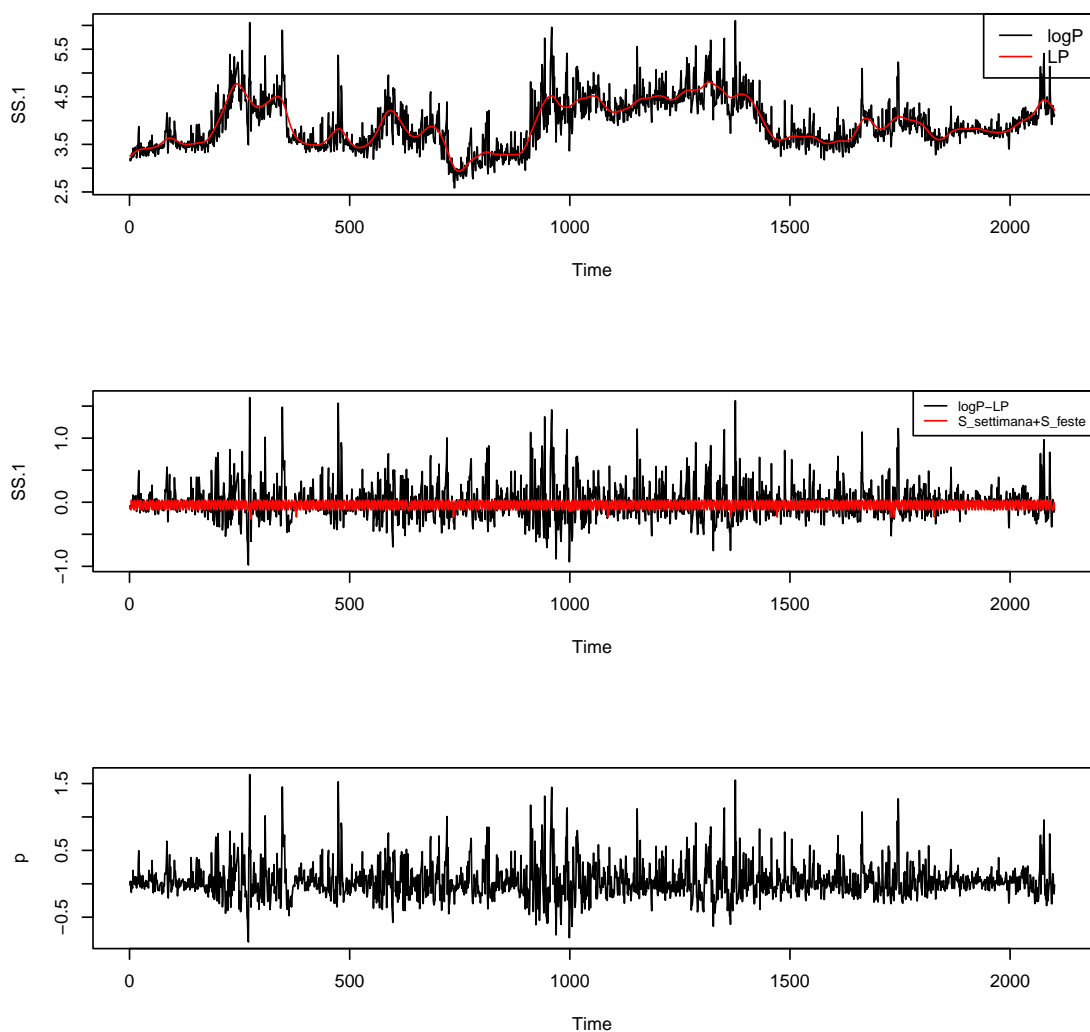


Figura A.12: Decomposizione completa della serie logaritmo dei prezzi nella fascia semi oraria 12.00–12.30 nel mercato Inglese per il periodo 01/04/2005 al 31/12/2010. In alto la serie dei $\log P_t$ con la sovrapposizione della stima della componente di lungo periodo \widehat{LP}_t . In centro i $\log P$ depurati dalla componente di lungo a cui è sovrapposta la stima della stagionalità settimanale e l'effetto delle festività. In basso la componente stocastica dei $\log P$ stimata mediante rimozione della componente deterministica: $\hat{p}_t = \log P_t - \widehat{DP}_t$.

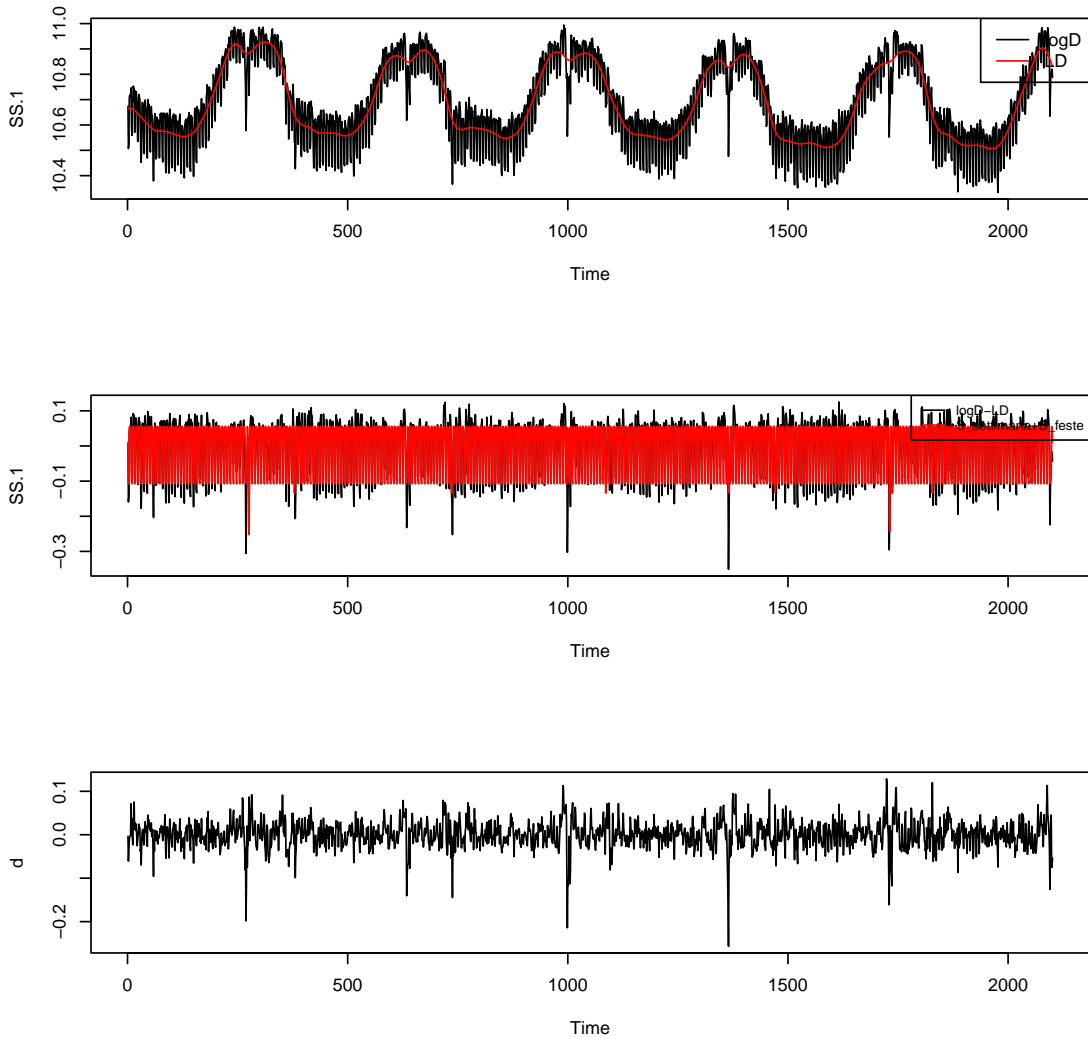


Figura A.13: Decomposizione completa della serie logaritmo della domanda nella fascia semi oraria 12.00–12.30 nel mercato Inglese per il periodo 01/04/2005 al 31/12/2010. In alto la serie dei $\log D_t$ con la sovrapposizione della stima della componente di lungo periodo \widehat{LD}_t . In centro i $\log D$ depurati dalla componente di lungo a cui è sovrapposta la stima della stagionalità settimanale e l'effetto delle festività. In basso la componente stocastica dei $\log D$ stimata mediante rimozione della componente deterministica: $\hat{d}_t = \log D_t - \widehat{DD}_t$.

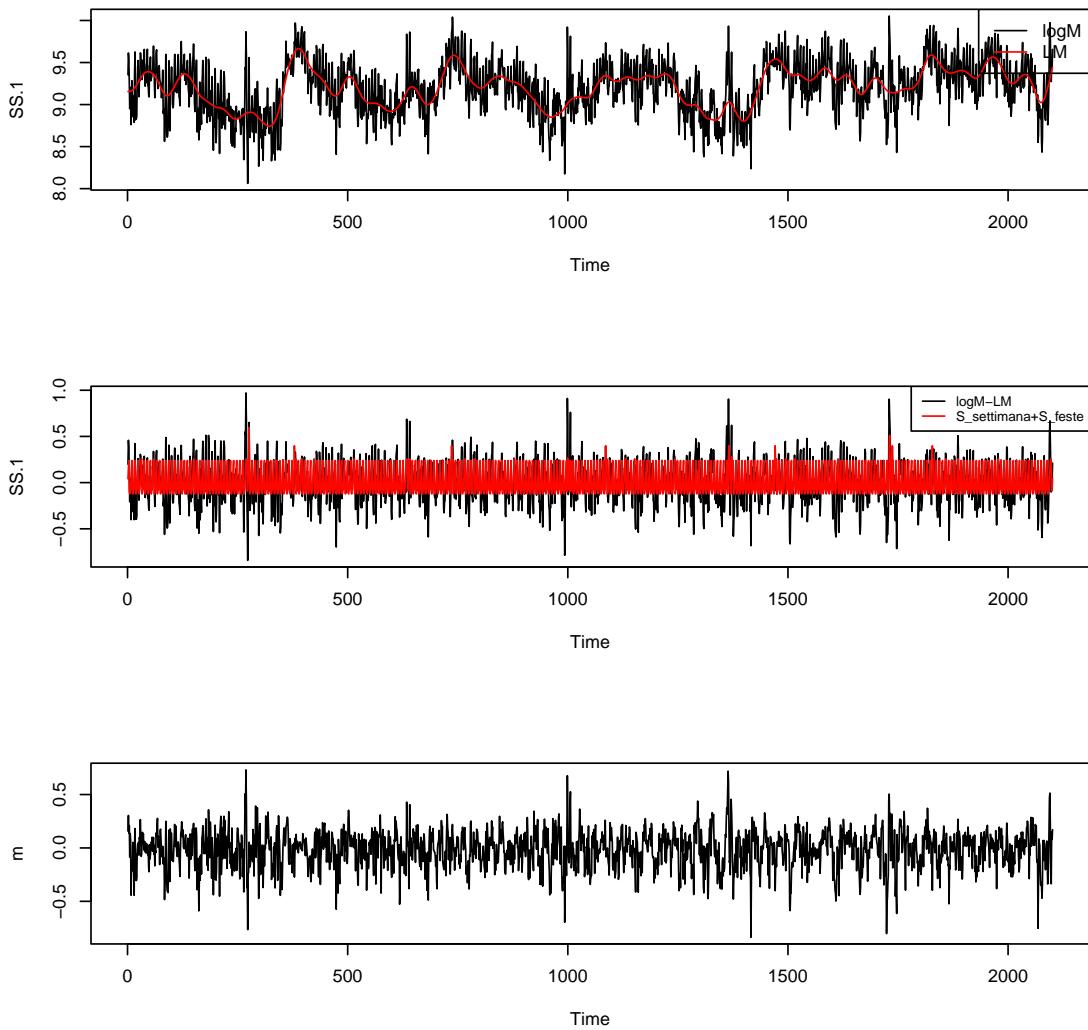


Figura A.14: Decomposizione completa della serie logaritmo del margine nella fascia semi oraria 12.00–12.30 nel mercato Inglese per il periodo 01/04/2005 al 31/12/2010. In alto la serie dei $\log M_t$ con la sovrapposizione della stima della componente di lungo periodo \widehat{LM}_t . In centro i $\log M$ depurati dalla componente di lungo a cui è sovrapposta la stima della stagionalità settimanale e l'effetto delle festività. In basso la componente stocastica dei $\log M$ stimata mediante rimozione della componente deterministica: $\hat{m}_t = \log M_t - \widehat{DM}_t$.

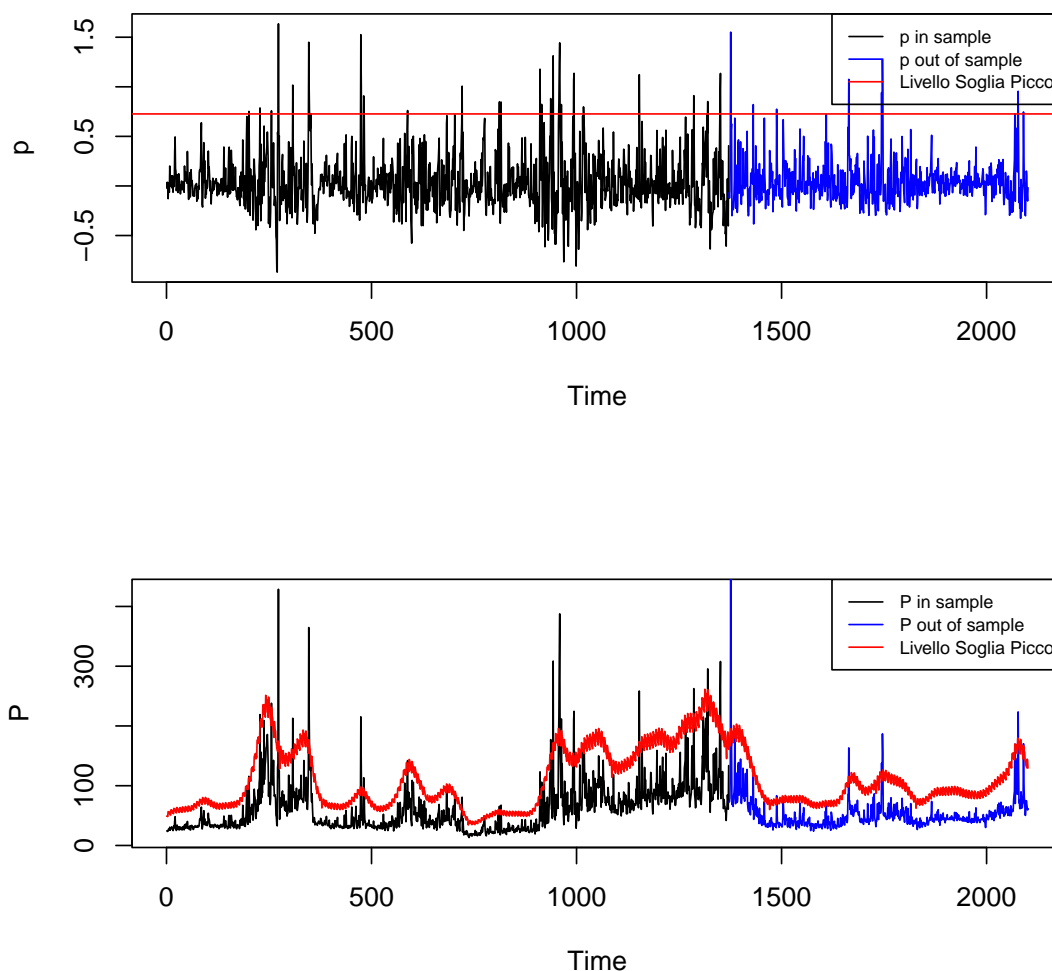


Figura A.15: Definizione dei picchi di prezzo attraverso i logaritmi dei prezzi del mercato elettrico Inglese nella fascia semi oraria 18.00–18.30 per il periodo 01/04/2005 al 31/12/2010. In alto la componente stocastica derivante dai logaritmi dei prezzi depurati dalla componente deterministica D_t divisa in campione *in sample* (nero) e *out of sample* (blu). La linea rossa è il quantile di livello 98% che identifica la soglia per identificare gli istanti in cui avvengono i picchi. In basso la serie dei prezzi (divisa in campione *in sample* (nero) e *out of sample* (blu)) alla quale è sovrapposta la soglia variabile nel tempo che identifica un picco di prezzo se oltrepassata (derivante dai $\log P$). I punti verdi sono riferimenti ottici per mostrare dove la serie dei prezzi oltrepassa la soglia che identifica un valore di picco.

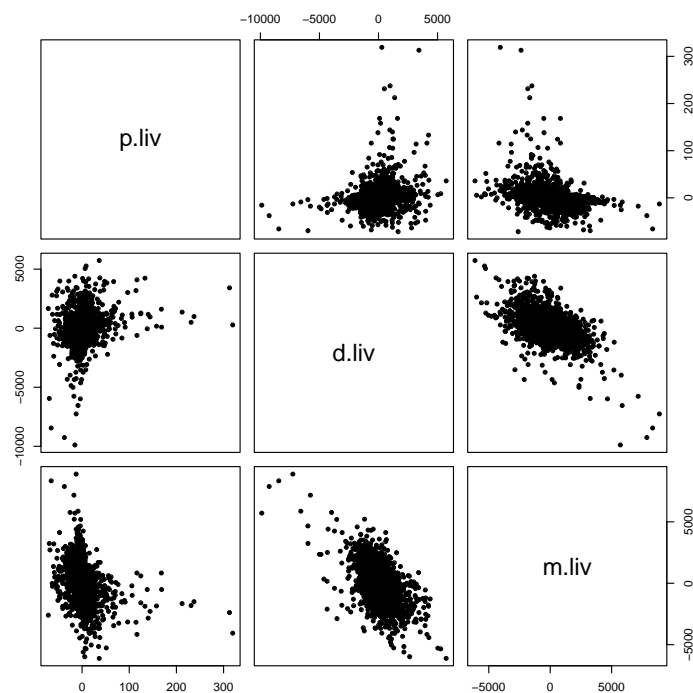


Figura A.16: Relazioni tra le serie stocastiche estratte dai prezzi, domanda e margine nel mercato elettrico Inglese nella fascia semi oraria 18.00–18.30 per il periodo 01/04/2005 al 31/12/2010.

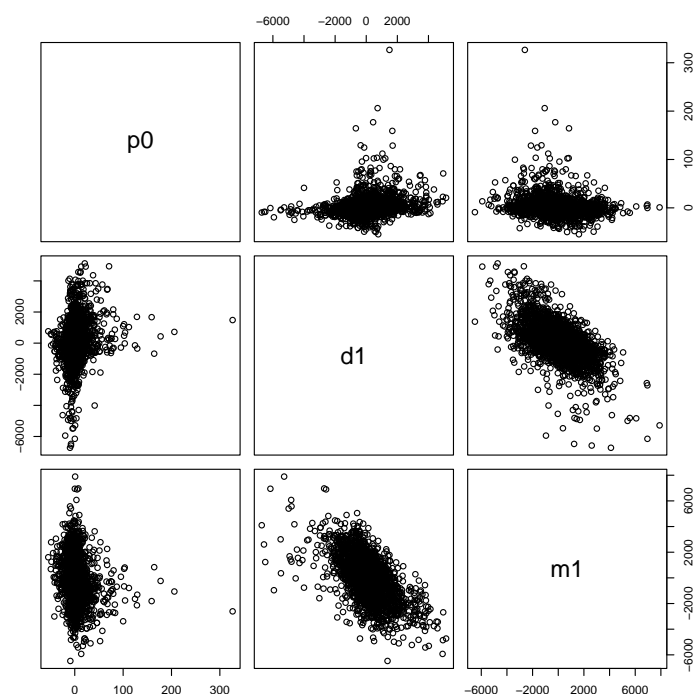


Figura A.17: Relazioni tra le serie stocastiche estratte dai prezzi (p_0) e le serie stocastiche ritardate al tempo $t-1$ di domanda (d_1) e margine (m_1) del mercato elettrico Inglese nella fascia semi oraria 12.00–12.30 per il periodo 01/04/2005 al 31/12/2010.

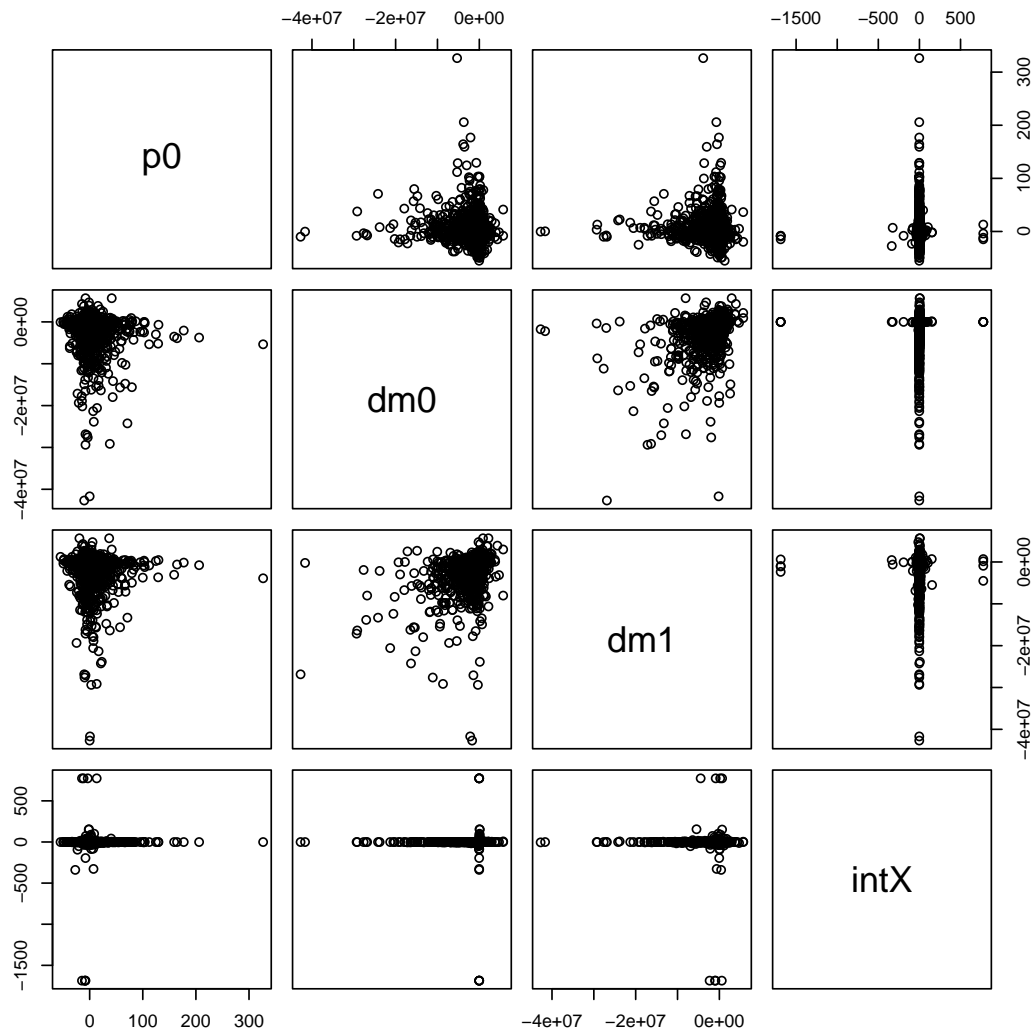


Figura A.18: Relazioni tra le serie stocastiche estratte dai prezzi ($p0$) e le serie stocastiche interazione e interazione a ritardo uno ($dm0=d_t m_t$, $dm1=d_{t-1} m_{t-1}$ e $intX=\frac{d_t}{m_t}$) del mercato elettrico Inglese nella fascia semi oraria 12.00–12.30 per il periodo 01/04/2005 al 31/12/2010.

Appendice B

TABELLE

Sono riportate alcune tabelle salienti e utili che per brevità sono omesse dalla trattazione svolta.

Tau	PR.PIC	PR.NPIC	NPR.PIC	NPR.NPIC	PicchiPrevisti %	FalsiAllarmi %
0.5	2	2	8	718	20.00	50.00
0.75	3	4	7	716	30.00	57.14
0.8	3	4	7	716	30.00	57.14
0.85	3	4	7	716	30.00	57.14
0.88	3	5	7	715	30.00	62.50
0.9	3	5	7	715	30.00	62.50
0.93	4	5	6	715	40.00	55.56
0.95	5	7	5	713	50.00	58.33
0.975	7	15	3	705	70.00	68.18

Tabella B.1: Previsioni ad un passo modello con sole esplicative significative con quantili condizionati basati sulla distribuzione normale nella fascia semi oraria 18.00–18.30 per il periodo 01/01/2009 al 31/12/2010.

Fascia Semi Oraria	Modello LM ridotto
1	$p_0 \sim p_1 + d_0 + m_0$
2	$p_0 \sim p_1 + d_0 + m_0$
3	$p_0 \sim p_1 + d_0 + m_0 + m_1$
4	$p_0 \sim p_1 + d_0 + m_0$
5	$p_0 \sim p_1 + d_0 + m_0 + p_6$
6	$p_0 \sim p_1 + d_0 + m_0 + p_6$
7	$p_0 \sim p_1 + d_0 + m_0 + m_1$
8	$p_0 \sim p_1 + d_0 + m_0$
9	$p_0 \sim p_1 + d_0 + m_0 + m_1$
10	$p_0 \sim p_1 + d_0 + m_0 + p_6$
11	$p_0 \sim p_1 + d_0 + m_0$
12	$p_0 \sim p_1 + d_0 + m_0$
13	$p_0 \sim p_1 + d_0 + m_0$
14	$p_0 \sim p_1 + d_0 + m_0 + p_3 + p_7$
15	$p_0 \sim p_1 + m_0 + p_4 + p_7$
16	$p_0 \sim p_1 + m_0 + p_4 + p_7$
17	$p_0 \sim p_1 + m_0 + p_4 + p_7$
18	$p_0 \sim p_1 + d_0 + m_0 + p_4 + p_7$
19	$p_0 \sim p_1 + m_0$
20	$p_0 \sim p_1 + m_0$
21	$p_0 \sim p_1 + m_0 + d_1 + m_1 + p_4$
22	$p_0 \sim p_1 + m_0 + d_1 + m_1 + p_4$
23	$p_0 \sim p_1 + m_0 + m_1 + p_4 + dm_0$
24	$p_0 \sim p_1 + m_0 + m_1 + p_4$
25	$p_0 \sim p_1 + m_0 + d_1 + m_1$
26	$p_0 \sim p_1 + m_0 + d_1 + m_1$
27	$p_0 \sim p_1 + m_0 + d_1 + m_1$
28	$p_0 \sim p_1 + m_0 + d_1 + m_1$
29	$p_0 \sim p_1 + m_0 + d_1 + m_1 + p_4$
30	$p_0 \sim p_1 + m_0 + d_1 + m_1 + p_4$
31	$p_0 \sim p_1 + m_0 + d_1 + m_1 + p_3$
32	$p_0 \sim p_1 + m_0 + m_1 + p_4 + d_1$
33	$p_0 \sim p_1 + m_0 + p_4$
34	$p_0 \sim p_1 + d_0 + m_0$
35	$p_0 \sim p_1 + m_0$
36	$p_0 \sim p_1 + m_0 + p_3$
37	$p_0 \sim p_1 + m_0$
38	$p_0 \sim p_1 + m_0$
39	$p_0 \sim p_1 + m_0$
40	$p_0 \sim p_1 + m_0 + m_1$
41	$p_0 \sim p_1 + m_0 + m_1$
42	$p_0 \sim p_1 + m_0 + d_1 + m_1$
43	$p_0 \sim p_1 + m_0 + d_1 + m_1 + p_2 + p_5$
44	$p_0 \sim p_1 + m_0 + d_1 + m_1 + p_5$
45	$p_0 \sim p_1 + d_0 + m_0 + m_1 + p_5$
46	$p_0 \sim p_1 + d_0 + m_0 + p_6$
47	$p_0 \sim p_1 + d_0 + m_0 + p_7$
48	$p_0 \sim p_1 + d_0 + m_0$

Tabella B.2: Miglior modello di regressione lineare stimato con minimi quadrati ordinari per ciascuna fascia semi oraria nel periodo *in sample* (01/04/2005 al 31/12/2008) utilizzato per la previsione un passo in avanti. Forma del modello di regressione lineare con esplicative significative per ciascuna fascia semi oraria. Con $p_0 \sim$ indichiamo che p_t viene regredita dall'insieme di esplicative significative.

Fascia Semi Oraria	Numero Picchi Out of Sample
1	6
2	8
3	7
4	8
5	6
6	5
7	8
8	12
9	12
10	12
11	8
12	9
13	6
14	4
15	4
16	3
17	5
18	5
19	4
20	4
21	5
22	5
23	5
24	6
25	6
26	7
27	7
28	7
29	7
30	8
31	7
32	7
33	10
34	9
35	10
36	11
37	10
38	10
39	4
40	4
41	3
42	3
43	3
44	3
45	4
46	3
47	5
48	6

Tabella B.3: Numero di picchi in ciascuna fascia semi oraria nel periodo *out of sample* fissando una soglia come 98 –esimo quantile empirico della componente stocastica dei prezzi.

SemiOra25	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
	0.85	V	-	V	-	-	-	-	-	-	-	-	-	-
	0.88	V	-	V	-	-	-	-	-	-	-	-	-	-
	0.9	V	-	V	-	V	-	-	-	-	-	-	-	-
	0.93	V	-	V	V	V	-	-	-	-	-	-	-	-
	0.95	V	-	V	-	V	-	-	-	-	-	-	-	-
SemiOra26	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
	0.85	V	-	V	-	-	-	-	-	-	-	-	-	-
	0.88	V	-	V	V	V	-	-	-	-	-	-	-	-
	0.9	V	-	V	V	V	-	-	-	-	V	-	V	-
	0.93	V	-	V	-	-	-	-	-	-	-	-	-	-
	0.95	V	V	V	V	V	-	-	-	-	-	-	-	-
SemiOra27	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
	0.85	V	-	V	-	-	-	-	-	-	-	-	-	-
	0.88	V	-	V	-	V	-	-	-	-	-	-	V	-
	0.9	V	-	V	-	-	-	-	-	-	-	-	V	-
	0.93	V	-	V	V	-	-	-	-	-	-	-	V	-
	0.95	V	-	V	V	V	-	-	-	V	-	-	V	V
SemiOra28	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
	0.85	V	V	V	-	V	-	-	-	-	-	-	V	-
	0.88	V	-	V	-	V	-	-	-	-	-	-	V	-
	0.9	V	-	V	-	V	-	-	-	-	-	-	V	-
	0.93	V	-	V	V	V	-	-	-	-	-	-	V	-
	0.95	V	-	V	V	V	-	-	-	-	-	-	V	V
SemiOra29	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
	0.85	V	-	V	-	V	-	-	-	-	-	-	-	-
	0.88	V	-	V	-	V	-	-	-	-	-	-	V	-
	0.9	V	-	V	-	V	-	V	-	-	-	-	V	-
	0.93	V	-	V	V	V	V	-	-	-	-	-	V	V
SemiOra3	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
	0.95	V	-	V	V	V	-	-	-	-	-	-	V	V
	0.85	V	-	V	-	V	-	V	-	-	-	-	-	-
	0.88	V	V	V	-	-	V	V	-	-	-	-	V	-
	0.9	V	-	V	V	V	-	-	-	-	-	-	V	-
	0.93	V	-	V	-	-	-	-	-	-	-	-	V	V
	0.95	V	V	V	V	V	-	-	-	-	-	-	V	V
SemiOra31	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
	0.85	V	-	V	-	V	-	V	-	-	-	-	V	-
	0.88	V	-	V	-	V	-	V	-	-	-	-	V	-
	0.9	V	V	V	-	-	-	-	-	-	-	-	-	-
	0.93	V	-	V	-	-	-	-	-	-	-	-	-	-
	0.95	V	-	V	V	V	V	V	-	-	-	V	-	-
SemiOra32	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
	0.85	V	-	V	-	-	V	V	-	-	-	-	-	-
	0.88	V	-	V	-	-	V	V	-	-	-	V	V	-
	0.9	V	-	V	-	-	-	-	-	-	-	V	-	-
	0.93	V	V	V	-	-	-	-	V	V	-	V	-	-
	0.95	V	V	V	-	-	V	V	-	-	-	V	-	-
SemiOra33	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
	0.85	V	-	V	-	-	-	-	-	-	-	-	-	-
	0.88	V	-	V	-	-	-	-	-	-	-	-	-	-
	0.9	V	V	V	-	-	V	-	-	-	-	-	-	-
	0.93	V	V	V	-	-	-	V	-	-	V	V	-	-
	0.95	V	V	V	V	-	-	V	-	-	V	V	-	-

SemiOra43	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
	0.85	V	-	V	-	-	-	-	-	-	-	-	-	-
	0.88	V	-	V	-	-	-	-	-	-	-	-	-	-
	0.9	V	-	V	-	-	-	-	-	-	-	-	-	-
	0.93	V	-	V	-	-	-	-	-	-	-	-	-	-
	0.95	V	-	V	-	V	-	-	-	-	-	-	-	-
SemiOra44	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
	0.85	V	V	V	-	V	-	-	-	-	V	V	-	-
	0.88	V	-	V	-	-	-	-	-	-	V	V	-	-
	0.9	V	-	V	-	V	-	-	-	-	V	V	-	-
	0.93	V	-	V	-	-	-	-	V	-	-	-	-	-
	0.95	V	-	V	-	-	-	-	-	-	-	-	-	-
SemiOra45	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
	0.85	V	V	V	-	-	-	-	-	-	-	-	-	-
	0.88	V	V	V	-	-	-	-	-	-	-	-	-	-
	0.9	V	-	V	-	-	-	-	-	-	-	-	-	-
	0.93	V	-	V	-	-	-	-	V	-	-	-	V	-
	0.95	V	V	V	-	-	-	-	V	-	-	-	-	V
SemiOra46	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
	0.85	V	V	V	-	-	-	-	-	-	-	-	-	-
	0.88	V	-	V	-	-	-	-	-	-	-	-	-	-
	0.9	V	V	V	-	-	-	-	-	V	-	-	-	-
	0.93	V	-	V	-	-	-	-	-	-	-	-	-	-
	0.95	V	-	V	-	-	-	-	-	-	-	-	-	-
SemiOra47	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
	0.85	V	-	V	-	-	-	-	-	-	V	-	-	-
	0.88	V	V	V	-	-	-	-	-	-	V	-	V	-
	0.9	V	-	V	-	-	-	-	-	-	-	-	-	-
	0.93	V	V	V	-	-	-	-	-	-	-	-	V	-
	0.95	V	-	V	-	-	-	-	-	-	-	-	-	-
SemiOra48	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
	0.85	V	V	V	-	V	-	-	-	-	-	-	-	-
	0.88	V	V	V	-	-	-	-	-	-	-	-	-	-
	0.9	V	V	V	-	V	-	-	-	-	-	-	-	-
	0.93	V	-	V	-	V	-	V	-	-	-	-	-	-
	0.95	V	-	V	-	-	-	-	-	-	-	-	-	-

Tabella B.4: Scelta del miglior modello di regressione del quantile per differenti livelli di τ stimati nel campione *in sample*, per tutte le fasce semi orarie.

Tau	PR.PIC	NPR.PIC	NPR.NPIC	PR.NPIC	Frazione Sforamenti	Quantile
0.5	1	5	724	0		0.52
0.75	1	5	722	2		0.23
0.8	1	5	721	3		0.17
0.85	1	5	719	5		0.12
0.88	1	5	717	7		0.09
0.9	2	4	715	9		0.07
0.93	3	3	706	18		0.05
0.95	3	3	696	28		0.03
0.975	5	1	668	56		0.01

Tabella B.5: Performance del modello di regressione del quantile definito ad *hoc* per differenti livelli del quantile τ per la fascia semi oraria delle 12.00–12.30.

Tau	PR.PIC	NPR.PIC	NPR.NPIC	PR.NPIC	Frazione Sforamenti	Quantile
0.5	2	8	719	1		0.53
0.75	3	7	717	3		0.24
0.8	3	7	716	4		0.17
0.85	3	7	715	5		0.13
0.88	3	7	713	7		0.11
0.9	4	6	713	7		0.08
0.93	5	5	706	14		0.06
0.95	7	3	701	19		0.04
0.975	9	1	683	37		0.02

Tabella B.6: Performance del modello di regressione del quantile definito ad *hoc* per differenti livelli del quantile τ per la fascia semi oraria delle 18.00–18.30.

Scelta delle esplicative mediante criterio di presenza per livelli elevati di τ per la definizione di un modello additivo di regressione del quantile non parametrico con soli regressori necessari per ciascuna fascia semi oraria. Selezione delle esplicative sulla base di test di significatività condotti al 1% sulla singola esplicativa e sulla tenuta del modello rispetto a quello completo. Definizione di un modello con sole variabili esplicative di QR non parametrica.

Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
1	0.85	V	-	V	V	-	V	V	-	V	V	V	-	-
1	0.9	V	-	V	-	-	-	-	-	V	V	V	-	-
1	0.95	V	-	V	V	V	V	V	-	V	V	V	-	-
1	0.975	V	V	V	V	-	V	V	-	V	-	V	-	-
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
2	0.85	V	-	V	V	-	V	V	-	V	V	V	-	-
2	0.9	V	-	V	-	-	-	-	-	V	V	V	-	-
2	0.95	V	-	V	V	V	V	V	-	V	V	V	-	-
2	0.975	V	V	V	V	-	V	V	-	V	-	V	-	-
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
3	0.85	V	-	V	V	-	V	V	-	V	V	V	-	-
3	0.9	V	-	V	-	-	-	-	-	V	V	V	-	-
3	0.95	V	-	V	V	V	V	V	-	V	V	V	-	-
3	0.975	V	V	V	V	-	V	V	-	V	-	V	-	-
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
4	0.85	V	-	V	V	-	V	V	-	V	V	V	-	-
4	0.9	V	-	V	-	-	-	-	-	V	V	V	-	-
4	0.95	V	-	V	V	V	V	V	-	V	V	V	-	-
4	0.975	V	V	V	V	-	V	V	-	V	-	V	-	-
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
5	0.85	V	V	V	-	V	-	V	-	-	V	-	-	-
5	0.9	V	V	V	-	V	-	-	V	V	-	-	-	-
5	0.95	V	V	V	-	V	-	V	V	V	-	-	V	-
5	0.975	V	-	V	-	V	-	V	-	-	V	-	V	-
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
6	0.85	V	V	V	-	V	-	V	V	V	V	V	-	-
6	0.9	V	V	V	-	-	-	V	V	V	V	-	-	-
6	0.95	V	V	V	-	V	-	-	V	-	-	-	-	-
6	0.975	V	V	V	-	V	V	V	-	-	V	-	-	V

Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
7	0.85	V	V	V	-	-	-	V	V	-	-	V	V	V
7	0.9	V	V	V	-	V	V	-	V	-	-	-	V	V
7	0.95	V	V	V	V	-	-	V	V	-	-	-	-	-
7	0.975	V	V	V	-	V	-	V	V	-	V	-	-	-
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
8	0.85	V	V	V	V	V	-	V	V	-	-	V	V	-
8	0.9	V	V	V	-	-	-	-	-	-	-	-	-	-
8	0.95	V	V	V	V	-	V	V	V	V	-	-	V	V
8	0.975	V	V	V	-	-	-	-	V	V	-	-	-	-
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
9	0.85	V	V	V	V	V	-	V	V	-	-	V	V	-
9	0.9	V	V	V	-	-	-	-	-	-	-	-	-	-
9	0.95	V	V	V	V	-	V	V	V	V	-	-	V	V
9	0.975	V	V	V	-	-	-	-	V	V	-	-	-	-
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
10	0.85	V	V	V	-	V	-	-	-	V	V	-	-	-
10	0.9	V	V	V	-	V	V	-	-	V	-	-	-	V
10	0.95	V	V	V	V	-	-	-	V	V	-	V	-	-
10	0.975	V	V	V	-	-	-	V	V	V	-	V	-	-
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
11	0.85	V	V	V	V	V	-	V	V	V	-	V	V	-
11	0.9	V	V	V	-	V	V	V	V	-	-	-	V	-
11	0.95	V	-	V	V	V	V	V	V	V	V	-	V	-
11	0.975	V	-	V	V	V	V	V	-	-	-	-	V	-
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
12	0.85	V	V	V	-	-	-	V	-	-	V	V	V	-
12	0.9	V	V	V	-	-	-	V	-	-	V	-	-	-
12	0.95	V	V	V	-	V	V	V	-	V	V	V	-	V
12	0.975	V	-	V	V	-	-	V	-	-	-	-	V	-
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
13	0.85	V	V	V	-	-	-	V	-	-	-	-	-	-
13	0.9	V	V	V	V	-	-	V	-	-	V	V	-	V
13	0.95	V	-	V	V	-	-	V	-	-	V	V	-	V
13	0.975	V	V	V	V	V	V	-	V	V	-	V	V	V
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
14	0.85	V	V	V	-	-	V	V	V	-	-	V	-	-
14	0.9	V	V	V	V	-	V	V	V	V	V	V	V	V
14	0.95	V	V	V	V	-	V	V	V	-	V	V	-	V
14	0.975	V	-	V	V	-	V	V	V	-	-	-	-	V
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
15	0.85	V	-	V	V	V	-	-	V	-	-	-	-	-
15	0.9	V	-	V	V	-	-	-	V	-	-	-	-	-
15	0.95	V	V	V	V	V	V	V	V	V	V	V	-	V
15	0.975	V	-	V	-	-	-	-	V	-	-	-	-	-
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
16	0.85	V	V	V	-	-	V	-	V	-	-	-	-	V
16	0.9	V	-	V	-	V	V	V	V	-	-	-	V	-
16	0.95	V	V	V	-	V	V	V	V	V	V	-	V	V
16	0.975	V	-	V	-	-	V	-	V	-	V	-	-	-
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
17	0.85	V	V	V	V	-	V	-	V	V	-	V	-	-
17	0.9	V	V	V	-	-	-	-	V	V	-	V	-	-
17	0.95	V	V	V	-	V	V	-	V	-	-	-	-	-
17	0.975	V	-	V	-	V	V	V	V	V	-	V	V	V

Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
18	0.85	V	V	V	-	-	V	-	V	V	-	V	-	-
18	0.9	V	V	V	-	-	V	-	-	V	-	V	V	-
18	0.95	V	V	V	-	V	V	-	V	-	V	V	V	-
18	0.975	V	V	V	-	V	V	V	V	V	-	-	V	V
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
19	0.85	V	V	V	-	V	-	-	-	-	V	V	-	-
19	0.9	V	V	V	-	V	-	-	V	V	V	V	-	-
19	0.95	V	-	V	V	-	V	-	V	-	-	-	-	-
19	0.975	V	-	V	V	V	V	V	V	-	V	V	V	-
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
20	0.85	V	V	V	-	V	-	-	-	-	V	V	-	-
20	0.9	V	V	V	-	V	-	-	V	V	V	V	-	-
20	0.95	V	-	V	V	-	V	-	V	-	-	-	-	-
20	0.975	V	-	V	V	V	V	V	V	-	V	V	V	-
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
21	0.85	V	-	V	V	V	-	-	V	-	-	V	-	V
21	0.9	V	-	V	V	V	-	-	V	V	-	-	-	-
21	0.95	-	-	V	V	V	V	V	V	V	-	-	-	-
21	0.975	-	-	V	V	-	-	V	V	V	V	V	-	V
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
22	0.85	V	-	V	V	V	-	-	V	-	-	V	-	V
22	0.9	V	V	V	V	V	V	V	V	V	V	V	-	-
22	0.95	-	-	V	-	-	V	-	-	-	-	-	-	-
22	0.975	V	-	V	V	-	V	V	V	-	-	-	-	V
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
23	0.85	V	-	V	-	V	V	-	-	-	-	V	-	-
23	0.9	V	-	V	V	V	V	-	V	V	V	-	-	-
23	0.95	V	V	V	-	V	V	-	V	-	-	-	-	-
23	0.975	V	-	V	V	V	V	V	V	V	V	V	V	-
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
24	0.85	V	V	V	V	V	V	-	V	-	V	-	V	-
24	0.9	V	V	V	V	V	V	V	V	V	V	V	V	-
24	0.95	V	V	V	V	V	V	V	V	V	-	V	-	-
24	0.975	V	-	V	V	V	V	V	V	V	-	V	V	-
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
25	0.85	V	V	V	V	V	V	-	V	-	V	-	V	-
25	0.9	V	V	V	V	V	V	V	V	V	V	V	V	-
25	0.95	V	V	V	V	V	V	V	V	V	-	V	-	-
25	0.975	V	-	V	V	V	V	V	V	V	-	V	V	-
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
26	0.85	V	V	V	V	V	V	-	V	-	V	-	V	-
26	0.9	V	V	V	V	V	V	V	V	V	V	V	V	-
26	0.95	V	V	V	V	V	V	V	V	V	-	V	-	-
26	0.975	V	-	V	V	V	V	V	V	V	-	V	V	-
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
27	0.85	V	V	V	-	V	V	-	V	-	-	-	V	-
27	0.9	V	V	V	V	V	V	-	V	-	-	-	V	V
27	0.95	V	-	V	V	V	V	-	V	V	-	V	-	-
27	0.975	V	V	V	V	V	V	-	V	-	-	-	-	-
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
28	0.85	V	-	V	-	-	V	-	-	V	-	-	-	-
28	0.9	V	-	V	V	V	V	-	-	-	-	-	-	-
28	0.95	V	V	V	V	V	V	V	V	V	V	-	-	V
28	0.975	V	V	V	-	V	V	V	-	V	V	V	-	-

Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
29	0.85	V	-	V	-	-	V	-	-	-	-	-	-	-
29	0.9	V	-	V	-	-	V	-	-	V	-	V	-	-
29	0.95	V	-	V	V	V	V	-	V	V	V	-	-	-
29	0.975	V	-	V	V	-	V	-	-	V	V	V	-	-
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
30	0.85	V	V	V	-	V	V	V	-	V	V	V	-	-
30	0.9	V	V	V	V	V	V	V	V	V	V	V	-	V
30	0.95	V	-	V	V	V	-	-	-	V	-	-	-	-
30	0.975	V	-	V	-	-	-	-	V	-	-	-	-	-
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
31	0.85	V	V	V	-	V	V	V	-	V	V	V	-	-
31	0.9	V	V	V	V	V	V	V	V	V	V	V	-	V
31	0.95	V	-	V	V	V	-	-	-	V	-	-	-	-
31	0.975	V	-	V	-	-	-	-	V	-	-	-	-	-
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
32	0.85	V	V	V	V	V	V	V	-	-	V	V	V	V
32	0.9	V	-	V	-	V	-	V	V	-	-	-	V	-
32	0.95	V	-	V	-	-	-	-	V	V	V	V	-	-
32	0.975	V	V	V	V	-	V	V	-	V	V	V	-	V
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
33	0.85	V	V	V	V	V	-	-	V	-	-	-	V	V
33	0.9	V	V	V	V	V	V	V	-	-	-	V	V	-
33	0.95	V	V	V	V	-	V	V	-	-	-	V	-	-
33	0.975	V	-	V	-	-	V	-	-	-	V	V	-	-
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
34	0.85	V	-	V	V	V	-	V	V	-	-	V	V	-
34	0.9	V	V	V	-	-	-	V	-	-	-	V	-	-
34	0.95	V	V	V	V	-	V	-	-	-	V	V	-	-
34	0.975	V	V	V	V	-	-	-	V	-	V	V	-	-
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
35	0.85	V	-	V	V	V	V	V	V	V	-	-	V	V
35	0.9	V	-	V	V	V	V	V	-	V	V	V	V	V
35	0.95	V	V	V	V	V	V	-	V	V	V	V	V	V
35	0.975	V	V	V	V	V	V	V	V	V	V	V	V	-
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
36	0.85	V	-	V	-	-	-	-	V	V	-	-	-	-
36	0.9	V	-	V	-	V	V	-	V	V	-	V	V	-
36	0.95	V	V	V	V	V	-	-	V	V	V	V	-	V
36	0.975	V	-	V	V	V	V	V	V	V	V	-	-	V
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
37	0.85	V	-	V	-	-	-	-	V	V	-	-	-	-
37	0.9	V	-	V	-	V	V	-	V	V	-	V	V	-
37	0.95	V	V	V	V	V	-	-	V	V	V	V	-	V
37	0.975	V	-	V	V	V	V	V	V	V	V	-	-	V
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
38	0.85	V	-	V	-	V	-	V	V	-	V	V	V	-
38	0.9	V	-	V	V	V	-	V	V	-	-	V	V	-
38	0.95	V	-	V	V	V	-	V	V	V	V	V	V	-
38	0.975	V	V	V	V	-	V	V	-	V	V	V	V	V
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
39	0.85	V	-	V	-	V	-	V	V	-	V	V	V	-
39	0.9	V	-	V	V	V	-	V	V	-	-	V	V	-
39	0.95	V	-	V	V	V	-	V	V	V	V	V	V	-
39	0.975	V	V	V	V	-	V	V	-	V	V	V	V	V

Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
40	0.85	V	-	V	-	V	-	V	V	-	V	V	V	-
40	0.9	V	-	V	V	V	-	V	V	-	-	V	V	-
40	0.95	V	-	V	V	V	-	V	V	V	V	V	V	-
40	0.975	V	V	V	V	-	V	V	-	V	V	V	V	V
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
41	0.85	V	-	V	-	V	-	V	-	V	V	V	-	-
41	0.9	V	-	V	V	V	-	-	V	V	V	V	-	V
41	0.95	V	V	V	-	-	-	V	-	-	-	-	-	-
41	0.975	V	V	V	-	V	V	V	V	V	V	V	V	V
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
42	0.85	V	V	V	-	-	V	-	V	V	-	-	V	V
42	0.9	V	-	V	-	V	V	-	V	V	-	V	V	V
42	0.95	V	V	V	V	V	V	-	V	-	V	V	V	V
42	0.975	V	V	V	-	V	V	V	V	-	V	V	V	V
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
43	0.85	V	-	V	V	V	-	-	-	-	-	V	-	-
43	0.9	V	-	V	-	-	-	-	-	-	-	V	-	V
43	0.95	V	-	V	V	V	-	-	V	-	-	V	V	V
43	0.975	V	-	V	-	-	V	-	V	-	-	V	-	V
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
44	0.85	V	-	V	V	V	-	-	-	-	-	V	-	-
44	0.9	V	-	V	-	-	-	-	-	-	-	V	-	V
44	0.95	V	-	V	V	V	-	-	V	-	-	V	V	V
44	0.975	V	-	V	-	-	V	-	V	-	-	V	-	V
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
45	0.85	V	-	V	V	V	-	-	-	-	-	V	-	-
45	0.9	V	-	V	-	-	-	-	-	-	-	V	-	V
45	0.95	V	-	V	V	V	-	-	V	-	-	V	V	V
45	0.975	V	-	V	-	-	V	-	V	-	-	V	-	V
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
46	0.85	V	-	V	V	V	-	-	-	-	-	V	-	-
46	0.9	V	-	V	-	-	-	-	-	-	-	V	-	V
46	0.95	V	-	V	V	V	-	-	V	-	-	V	V	V
46	0.975	V	-	V	-	-	V	-	V	-	-	V	-	V
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
47	0.85	V	V	V	-	V	V	-	-	V	-	-	V	V
47	0.9	V	V	V	-	V	V	-	-	V	V	-	V	V
47	0.95	V	V	V	V	V	V	-	V	-	V	-	V	V
47	0.975	V	-	V	V	V	V	-	V	V	-	V	V	V
Fascia	Tau	p1	d0	m0	d1	m1	p2	p3	p4	p5	p6	p7	dm0	dm1
48	0.85	V	V	V	-	-	-	V	-	V	-	-	-	-
48	0.9	V	-	V	-	-	-	-	-	V	-	-	-	-
48	0.95	V	V	V	-	V	-	-	-	-	V	V	V	V
48	0.975	V	V	V	V	V	-	V	-	V	-	V	-	V

Tabella B.7: Scelta delle esplicative per definire il miglior modello di regressione del quantile non parametrico per alcuni livelli del quantile condizionato in tutte le 48 fasce semi orarie della giornata.

BIBLIOGRAFIA

- [1] Alexander, C. (2008), *Market Risk Analysis Volume II — Practical Financial Econometrics*. Wiley (John Wiley & Sons Ltd), Chichester.
- [2] Azzalini, A. (2001), *Inferenza Statistica. Una presentazione basata sul concetto di verosomiglianza*. Springer, Milano.
- [3] Bhargava, A. (1986), *On the Theory of Testing for Unit Roots in Observed Time Series*. *The Review of Economic Studies* 53 (3): 369–384. doi:10.2307/2297634. JSTOR 2297634.
- [4] Borra, S. Di Ciaccio, A. (2008), *Statistica, metodologie per le scienze economiche e sociali*. McGraw-Hill, Milano.
- [5] Broccoli, S. Cavrini, G. Zoli, M. (2005), *Il modello di regressione quantile nell'analisi delle determinanti della qualità della vita in una popolazione anziana*. *Statistica*, anno LXV, n.4,2005.
- [6] Buhai, S. (2005), *Quantile regression: Overview and selected applications*. *Astra Journal*, Volume 4, 2005.
- [7] Bunn, D.W. (2006), *Modelling prices in competitive electricity markets*. Wiley (John Wiley & Sons Ltd), Chichester.
- [8] Cade, B. Noon, B. (2003), *A Gentle Introduction to Quantile Regression for Ecologists*. *Frontiers in Ecology and the Environment*, 1, 412-420.
- [9] Calento, R. e altri (2006), *Analisi econometrica del prezzo di borsa dell'energia elettrica*. http://www.eni.com/it_IT/attachments/lavora-con-noi/pdf/GP-Ipex.pdf.
- [10] Cappuccio, N. Orsi, R. (2005), *Econometria*. Il Mulino, Bologna.

- [11] Christensen, T. M. e altri (2011), *Forecasting spikes in electricity price*. International Journal of Forecasting. doi: 10.1016/j.ijforecast.2011.02.019
- [12] Gaglianone, W. Lima, L. Linton, O. Smith, D. (2009), *Evaluating Value-at-Risk models via Quantile Regression*. Working Paper 09-46, Economic Series (25).URL:<http://www.eco.uc3m.es/temp/09-46-25.pdf>
- [13] Gestore Mercati Elettrici (2009), *Vademecum della Borsa Elettrica*. <http://www.mercatoelettrico.org/it/MenuBiblioteca/documenti/20091028VademecumBorsaElettrica.pdf>.
- [14] Hao, L. Naiman, D.Q. (2007), *Quantile Regression*, Sage Publications.
- [15] Härdle, W. Spokoiny, V. e Wang, W. (2011), *Local Quantile Regression*. Deutsche Forschungsgemeinschaft through the SFB 649 Economic Risk. URL: <http://sfb649.wiwi.hu-berlin.de>. ISSN 1860-5664
- [16] Hastie, T. Tibshirani, R. (1990), *Generalized Additive Models*, Chapman & Hall CRC Monographs on Statistics & Applied Probability.
- [17] Hendricks, W. e Koenker, R. (1990), *Hierarchical Spline Models for Conditional Quantiles and the Demand for Electricity*. Bebr, faculty working paper NO. 90-1661, College of Commerce and Business Administration University of Illinois at Urbana-Champaign.
- [18] Koenker, R. (1994), *Confidence interval for regression quantile*. Proceedings of the fifth Prague symposium on Asymptotics statistics, Physica, pp. 349-359.
- [19] Koenker, R. (2005), *Quantile Regression*. Econometric Society Monograph Series, Cambridge.
- [20] Koenker, R (2010), *Quantile regression in R: A Vignette*. URL: <http://www.econ.uiuc.edu/~roger/research/rq/vig.pdf>
- [21] Koenker, R. (2010b), *Quantile Regression*. Econometric Society Monograph Series, Cambridge University Press.

- [22] Koenker, R. (2011), *Additive Models for quantile regression: model selection and confidence bands*. URL: <http://www.econ.uiuc.edu/~roger/research/bandaids/bandaids.pdf>.
- [23] Koenker, R. (2011b), *Package 'quantreg'*. URL:<http://cran.r-project.org/web/packages/quantreg/quantreg.pdf>
- [24] Koenker, R. Bassett, W. (1978), *Regression Quantiles*. *Econometrica*, Econometric Society, Vol. 46, No. 1. (Jan., 1978), pp. 33-50.
- [25] Koenker, R. Hallock, K. (2001), *Quantile Regression*, *Journal of Economic Perspectives*, 15, 4, pp. 143-156.
- [26] Koenker, R. e Portnoy, S. (1988), *M-Estimation of Multivariate Regressions*. Technical Report, University of Illinois.
- [27] Koenker, R. e Zhao, Q. (1996), *Conditional Quantile Estimation and Inference for Arch Models*. *Econometric Theory*, 12 , pp 793-813 doi:10.1017/S0266466600007167
- [28] Kwiatkowski, D. et al (1992), *Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root*. *Journal of Econometrics* 54, 159–178.
- [29] Kuan, C.M. (2007), *An introduction to quantile regression*. Institute of Economics Academia Sinica, June 4, 2007.
- [30] Manzan, S. e Zerom, D. (2008), *A bootstrap-based non parametric forecast density*. *International Journal of Forecasting* 24, 2008, 535-550.
- [31] Mosteller, F. e Tukey, J. (1977), *Data analysis and regression : a second course in statistics*. Addison-Wesley.
- [32] Pace, L. e Salvan, A. (2001), *Introduzione alla Statistica II Inferenza, verosomiglianza, modelli*. Cedam, Padova.
- [33] Peracchi, F. (1995), *Econometria*. McGraw-Hill, Milano.

- [34] Portnoy, S. Koenker, R. e Ng, P. (1993), *Quantile Smoothing Splines*. *Biometrika* (1994) 81 (4): 673-680. doi: 10.1093/biomet/81.4.673
- [35] R Development Core Team (2012), *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. URL: <http://www.R-project.org/>
- [36] Schulze, N. (2004), *Applied Quantile Regression: Microeconomic, Financial, and Environmental Analyses*. Inaugural-Dissertation zur Erlangung des Doktorgrades der Wirtschaftswissenschaftlichen Fakultät der Eberhard-Karls-Universität Tübingen.
- [37] Shumway, R.H. e Stoffer, D.S. (2006). *Time Series Analysis and Its Applications: With R Examples*. Springer, New York.
- [38] Takeuchi, I et al (2005), *Nonparametric Quantile Regression*. *Journal of Machine Learning Research* Nonparametric Quantile Estimation.
- [39] Trueck, S. Weron, R. Wolff, R. (2007), *Outlier Treatment and Robust Approaches for Modeling Electricity Spot Prices*. Hugo Steinhaus Center, Wroclaw University of Technology. NEP-ALL-2007-09-09, NEP-ENE-2007-09-09, Energy Economics.
- [40] Wang, Y. (2011), *Smoothing Splines Methods and Applications*. CRC Press Taylor & Francis Group.
- [41] Weron, R. (2006), *Modeling and forecasting electricity loads and prices – A Statistical Approach*. Wiley (John Wiley & Sons Ltd), Chichester.
- [42] Yu, K. Lu, Z. e Stander, J. (2003), *Quantile regression: applications and current research areas*. Royal Statistical Society, 0039–0526/03/52331.