Indice

Capitolo 1

Investments and Financial Markets

By *investment* we mean the commitment of current resources in economic or financial activity to achieve future benefits. If resources and benefits are expressed in terms of money, then the investment is characterized by a *cash flow stream*, which occurs at some dates from the beginning to the end of the activity. In particular, the values taken by the cash flow at the *initial* and *terminal date* of the investment are known as the investment's *initial* and *terminal cash flow*, respectively. The standard convention is that the parts of the cash flow stream representing a profit [resp. a loss] for the investor are considered with a positive [resp. negative] sign and more specifically referred to as *cash inflow* [resp. *outflow*].

Esempio 1 *We watch our bank statement in particular, the transactions: the deposits, including direct deposits, checks cashed, payments received, reimbursements, and interest earned, are shown as positive amounts; the withdrawals, including purchases, ATM withdrawals, automatic payments, checks issued, and bank fees are shown as negative amounts.*

Esempio 2 *We take a loan: the amount we receive has to be considered as our inflow. Our periodic payments to repay the loan are our outflow.*

1.0.1 Investment Decision

Comparison Principle: evaluate an investment by comparing it with other investments which are available in financial markets. Asset in financial markets constitute the reference systems for an economic evaluation of assets.

1.0.2 Arbitrage

We say economic or financial activity is an *arbitrage* if it produces a non-negative [resp. positive] cash flow with probability one [resp. with positive probability], without the commitment of initial or intermediate cash flow. We will make this concept more precise in what follows. However, we give here a simple funny example of arbitrage.

Esempio 3 *Assume that, while walking a road, we spot on the sidewalk a ticket from the current national lottery lost by a careless buyer. Assume we bend down to pick the ticket up and store it in the wallet. With this action, we get the opportunity of winning some money on the draw date at no cost. If we are lucky we will get the money. If we are not lucky we don't get the money, but, still, we lose no money out of our own pockets. Summarizing, by picking the lost ticket on the sidewalk, we get an opportunity of making money without running the risk of losing money. We have realized an arbitrage.*

1.0.3 Risk Aversion

Risk aversion principle - mean variance analysis - utility functions

1.0.4 Dynamics

The future price of an asset has to be regarded as a stochastic process, that is a time indexed sequence of random variables. An important part of the analysis of investments in financial assets is concerned with the characterization of this process.

1.1 Financial Markets

Si definiscono *mercati finanziari* i luoghi ideali nei quali vengono scambiati strumenti finanziari di varia natura. Nel contesto economico odierno, i mercati finanziari sono chiamati a svolgere due funzioni base:

- il trasferimento di risorse all'interno dell'economia tra unità in surplus e unità in deficit;
- la mitigazione mediante diversificazione e condivisione dei rischi insiti nelle attività economiche e finanziarie.

I mercati finanziari consentono infatti il trasferimento del risparmio dai soggetti che lo accumulano (prevalentemente le famiglie) ai soggetti che lo richiedono (governi, banche, imprese,...). Questi ultimi sono definiti *soggetti in disavanzo finanziario* ed emettono strumenti finanziari (titoli di stato, depositi bancari, obbligazioni, azioni,...) che cedono ai soggetti in avanzo finanziario in cambio di moneta. Lo scambio tra strumenti finanziari e moneta consente la redistribuzione dei rischi e dei profitti economici, perché vengono assunti in parte dagli acquirenti degli strumenti finanziari. Inoltre questi ultimi possono a loro volta cedere tali strumenti ad altri soggetti economici, scambiandoli in mercati appositi. Esistono quindi varie tipologie di mercati finanziari, ognuno con proprie regole e proprie caratteristiche.

La combinazione trasferimento di risorse-redistribuzione dei rischi e dei profitti costituisce la principale funzione economica dei mercati finanziari, che consentono di realizzare un'efficiente allocazione delle risorse finanziarie ai fini della formazione del capitale produttivo. In un sistema economico, domanda ed offerta di capitale devono tendere all'equilibrio e l'investimento del capitale deve essere efficiente. I mercati finanziari svolgono una funzione essenziale per il raggiungimento di questi due obiettivi.

Occupiamoci adesso di descrivere alcuni dei più comuni strumenti finanziari.

1.1.1 Bonds

I *titoli obbligazionari (bond)* sono contratti in cui l'emittente, in cambio di un prezzo alla sottoscrizione, si obbliga a remunerare l'investitore con il valore nominale del titolo sottoscritto *(principal)* alla maturità del titolo stesso, più un eventuale dividendo sotto forma di interessi pagati periodicamente in corso di maturità, noti come *cedole* (*coupons*). I titoli obbligazionari si dividono generalmente in due tipologie: i titoli a *cedola fissa (fixed coupon)*, in particolare *senza cedola (zero coupon)*, il cui dividendo è noto con certezza al momento della sottoscrizione e i titoli a *cedola variabile (variable coupon)*, il cui dividendo alla sottoscrizione è aleatorio.

In Italia, tra i titoli obbligazionari figurano numerosi Titoli di Stato, emessi dal Ministero del Tesoro per costituire risorse finanziarie da destinare agli investimenti di pubblica utilità. Tra questi ricordiamo i BTP, acronimo per Buoni del Tesoro Poliennali, a cedola fissa, i BOT, Buoni Ordinari del Tesoro, ed i CTZ, Certificati del Tesoro Zero Coupon, entrambi senza cedola. Inoltre abbiamo i CCT, Certificati di Credito del Tesoro, a cedola variabile in dipendenza dall'andamento del tasso di interesse di mercato. Sempre con la finalità di raccogliere risorse per gli investimenti, titoli obbligazionari possono essere emessi anche da imprese pubbliche e private, quali ENI, ENEL, FIAT, Telecom, ecc... Inoltre nel mercato obbligazionario italiano possono essere trattate anche obbligazioni di emittenti straniere.

1.1.2 Stocks

Con *titoli azionari* (*stock*) si intendono contratti emessi da imprese pubbliche e private, sempre con lo scopo di costituire risorse per gli investimenti, che, diversamente dai titoli obbligazionari, non impegnano l'emittente alla restituzione del debito contratto con l'investitore, ma offrono all'atto della sottoscrizione una percentuale di proprietà dell'impresa emittente stessa. Ciò comporta tuttavia il diritto dell'investitore, noto come *azionista* (*share holder*), a ricevere una remunerazione periodica costituita da una percentuale dei profitti dell'impresa, proporzionale alla percentuale di proprietà sottoscritta. Tale diritto non è però garantito, potendo essere sospeso qualora l'impresa necessiti, a giudizio della maggioranza degli azionisti, di un reinvestimento degli utili prodotti. Gli investitori che sottoscrivono titoli azionari si trovano a sopportare un rischio assai più elevato rispetto ai sottoscrittori dei titoli obbligazionari, in quanto i flussi di reddito prodotti dagli investimenti azionari sono molto più aleatori rispetto a quelli prodotti dai titoli obbligazionari. Infatti, relativamente a questi ultimi almeno il capitale inizialmente investito è garantito, a meno d'insolvenza (*default*) dell'emittente. Pertanto gli investitori in titoli azionari si attendono rendimenti molto più elevati come premio per il rischio sopportato. Il rapporto che si instaura fra l'azionista e l'impresa è un rapporto partecipativo che dipende dalle caratteristiche del titolo azionario in possesso dell'azionista. Le imprese hanno infatti la possibilità di emettere azioni di tipo diverso: oltre alle azioni ordinarie, esistono anche azioni cosiddette speciali, come le azioni privilegiate e quelle di risparmio:

- le *azioni ordinarie* attribuiscono ai loro possessori pieni diritti amministrativi, consentono quindi la partecipazione alle assemblee, sia ordinarie che straordinarie, e permettono l'esercizio del diritto di voto;

- le *azioni privilegiate* garantiscono all'azionista il diritto a una determinata quota dell'utile distribuibile prima che venga assegnato il dividendo alle azioni ordinarie. Il privilegio può anche riguardare il diritto di priorità al rimborso del capitale all'atto dello scioglimento dell'impresa. Esistono inoltre azioni privilegiate che consentono un dividendo cumulabile e quindi, entro un certo numero di anni, il recupero dei dividendi non corrisposti in precedenza per mancanza o insufficienza di utili.

- le *azioni di risparmio* possono essere emesse solo da società quotate e si differenziano dalle azioni ordinarie per la particolarità che il loro possessore non ha diritto di voto, sia in assemblea ordinaria che straordinaria, ma ha diritto ad un dividendo maggiorato rispetto all'azionista ordinario.

1.1.3 Derivatives

Gli *strumenti finanziari derivati (derivatives)*, sono così denominati perchè il loro valore deriva dal prezzo di *un'attività sottostante (underlying asset)*, che può essere costituita da *un'attività reale (commodity derivative)*, da *un'attività finanziaria (financial derivative)*, o da un *indice* sintetico dei prezzi o dei rendimenti relativo alle precedenti attività *(index derivative)*. I derivati sono distinguibili in quattro grandi famiglie: i *contratti a termine (forwards)*, i *futures*, le *opzioni (options)* e gli *swaps*. Un'ulteriore distinzione rilevante fa riferimento ai mercati nei quali tali derivati sono quotati: si distinguono derivati scambiati in mercati organizzati, *exchange traded derivatives*, e derivati negoziati fuori mercato, *over the counter derivatives*. Un vantaggio chiave delle contrattazioni over the counter è rappresentato dal fatto che le condizioni contrattuali non devono necessariamente corrispondere a quelle fissate dai mercati, ma i contraenti sono liberi di negoziare qualunque tipo di contratto risulti di reciproco interesse. Lo svantaggio maggiore è rappresentato dal rischio di credito, o più precisamente rischio d'insolvenza. C'è infatti una probabilità, per quanto piccola, che il contratto non venga onorato. Al contrario, i mercati organizzati si prefiggono lo scopo di eliminare, o quanto meno di ridurre, il rischio di credito. I contratti forwards e gli swaps sono negoziati fuori mercato, mentre i futures, proprio per le loro caratteristiche intrinseche, sono negoziati in mercati organizzati. Le opzioni sono negoziate sia nei mercati organizzati, sia *over the counter*.

Consideriamo adesso le caratteristiche principali degli strumenti derivati ed il modo in cui vengono negoziati nel mercato.

Forwards

Un *contratto a termine (forward)* è un accordo tramite il quali due contraenti, un acquirente (*buyer*) e un venditore (*seller*), si scambiano un certo sottostante (*underlying*) a una data, detta *maturità* (*maturity*), e a un prezzo, detto *prezzo di consegna (delivery price)*, che vengono concordati alla stipula dell'accordo stesso. In ciò si differenziano dai *contratti a pronti (spot)*, che hanno regolamento immediato. I forward vengono negoziati, di solito fuori mercato, tra due istituzioni finanziarie o tra un'istituzione finanziaria e uno dei suoi clienti. In questi contratti l'acquirente assume una *posizione lunga (long position)* e, alla maturità del contratto, si obbliga a comprare il sottostante dal venditore, al prezzo di consegna concordato alla stipula. Di contro, il venditore assume una *posizione corta (short position)* e, alla maturità, si obbliga a vendere il sottostante all'acquirente al prezzo di consegna. Lo scopo dei contratti forward è garantire sia all'acquirente che al venditore la copertura dal rischio derivante dalla variabilità del prezzo dell'attività sottostante dal momento della sottoscrizione alla maturità del contratto.

Esempio 4 *Supponiamo che in data* 1 *marzo* 2019*. Il tesoriere di una società statunitense sappia che tra* 6 *mesi, ossia in data* 1 *settembre* 2019*, dovrà effettuare un esborso di* £1.00 *milioni e vuole coprirsi dal rischio delle fluttuazioni del tasso di cambio che in data* 1 *marzo* 2019 *è di* 1.31559£/\$*. Il tesoriere si mette in contatto con una banca britannica e appreso che la banca è disposta a vendergli le sterline, con consegna tra* 6 *mesi, al tasso di cambio forward di* 1.32734£/\$*, cioè con la maggiorazione di* 117.5 *punti forward, accetta di entrare in un contratto per l'acquisto a termine di* £1.00 *milioni. La società si trova quindi ad avere una posizione lunga in un contratto forward sulle sterline. Si è obbligata in data* 1 *marzo* 2019 *ad acquistare in data* 1 *settembre* 2019 *la somma di* £1.00 *milioni dalla banca in cambio di* \$1.32734 *milioni. La banca si trova ad avere una posizione corta in un contratto forward sulle sterline. Si è obbligata in data* 1 *marzo* 2019 *a vendere in data* 1 *settembre* 2019 *la somma di* £1.00 *milioni alla società in cambio di* \$1.32734 *milioni. Entrambe le parti hanno assunto un* impegno vincolante (binding commitment)*. Se nel corso dei* 6 *mesi successivi alla stipula il tasso di cambio spot dovesse salire rispetto al tasso forward a, diciamo,* 1.42734£/\$*, il valore del contratto per la società sarebbe di* \$100,000*, dato che le sterline invece di essere acquistate a* \$1.42734 *milioni, verrebbero pagate* \$1.32734 *milioni. Al contrario, se nel corso dei* 6 *mesi successivi alla stipula del contratto, il tasso di cambio spot dovesse scendere rispetto al tasso forward a, diciamo,*1.22734\$/£*, il valore del contratto per la società sarebbe di* −\$100,000*, dato che il contratto forward la obbligherebbe a pagare* \$100,000 *in più rispetto al prezzo di mercato delle sterline.*

L'esempio esposto illustra un aspetto chiave della copertura mediante i contratti forward, che eliminano l'incertezza circa il costo dell'attività sottostante, o il ricavato derivante dalla vendita, ma non comportano necessariamente un risultato migliore. Dal momento che entrare in un contratto forward non comporta alcun costo, il valore finale del contratto è anche pari al profitto o alla perdita derivante dal contratto.

Futures

Un *contratto future (futures)*, al pari di un contratto forward, stabilisce tra due contraenti, un *acquirente* (*buyer*) e un *venditore* (*seller*), l'obbligo di acquistare o vendere un *titolo sottostante* (*underlying asset* o *underlying security*) a una data futura, nota come *data d'esercizio* (*exercise date*) o data di scadenza (*expiration date*) o maturità (*maturity*) e a un prezzo di consegna (*delivery price*), concordati alla stipula del contratto. Il sottostante può essere o un titolo di possesso di un bene reale (*commodity*), per esempio petrolio, oro, rame, grano, caffè, soia, e ci si riferisce a un tale future col termine *commodity futures*, o un titolo di possesso di una valuta (*currency*) denominato *currency futures*, o anche un titolo di possesso di un portafoglio di titoli (*stock portfolio*) del mercato finanziario, per esempio un indice borsistico, e in quest'ultimo caso si parla di *financial futures.* Acquistare [resp. vendere] futures significa impegnarsi ad acquistare [resp. a vendere] alla scadenza e al prezzo prefissati l'attività sottostante indipendentemente dal suo prezzo corrente di mercato (*market spot price*). Sottolineamo che per prezzo del future deve intendersi il prezzo d'esercizio. Le parti contraenti stipulano un contratto future a costo zero. Non c'è alcun esborso di denaro per entrare come acquirente in un contratto future, né tantomeno per entrarvi come venditore. Però alla maturità i contraenti sono obbligati all'acquisto o alla vendita del sottostante al prezzo d'esercizio. Il rispetto di quest'obbligo viene assicurato dall'imposizione di un *deposito di garanzia* (*futures margin*) ad ogni sottoscrittore di un contratto future proporzionale all'entità del contratto sottoscritto. Ovviamente l'acquisto di futures corrisponde ad una aspettativa da parte dell'acquirente di rialzo dell'attività sottostante; la vendita, invece, sottende un'aspettativa del venditore al ribasso. A differenza dei forward, i future sono di norma trattati in un mercato finanziario. Per rendere possibili le negoziazioni, il mercato standardizza certi aspetti del contratto. La standardizzazione consiste nella definizione del taglio unitario, della scadenza contrattuale e delle modalità che regolano i flussi finanziari tra le parti contraenti a garanzia del buon fine del contratto. Non è possibile negoziare futures che non soddisfino questi requisiti. Inoltre, dal momento che nel caso di contratti futures i due contraenti sono generalmente ignoti l'uno all'altro, viene anche fornito un meccanismo che assicura il rispetto del contratto da parte dei due contraenti. Infatti, tutti i contratti vengono stipulati di fatto con la *Cassa di Compensazione e Garanzia (Exchange Clearinghouse).* Questa è in genere una società per azioni avente un oggetto sociale esclusivo che le impone di assicurare il buon fine dei contratti future e di emanare regolamenti che disciplinano l'operatività del mercato di propria competenza. Quindi il prezzo di consegna dei future non è concordato tra le singole parti, ma è univocamente determinato sul *floor* del mercato organizzato in base alla legge della domanda e dell'offerta. Pertanto, è anche noto come *prezzo del future* (*futures price*). Se ci sono più investitori che vogliono assumere posizioni lunghe rispetto a quelli che vogliono assumere posizioni corte, il prezzo sale. Viceversa, il prezzo scende. La circostanza che i contratti futures rispettino degli standard e vengano stipulati con la Cassa di Compensazione rende possibile il loro annullamento tramite compensazione, ossia stipulando un contratto di segno opposto all'originale. In questo modo verrà evitata la consegna dell'attività sottostante il contratto. Infatti, acquistando un future con intenzioni speculative sarà essenziale effettuarne la vendita prima della scadenza contrattuale; se, invece, le intenzioni fossero di tipo assicurativo,ossia di copertura (*hedge*), per garantirsi un prezzo futuro certo di acquisto o vendita del sottostante, si aspetterà la scadenza prevista per provvedere all'acquisto o vendita del sottostante al prezzo stabilito.

Il mercato dei futures offre agli speculatori un'interessante *leva finanziaria (financial leverage).* Cerchiamo di chiarire meglio questo concetto mediante l'esempio seguente.

Esempio 5 *Consideriamo uno speculatore che in data* 1 *marzo* 2019 *ritenga che nei prossimi* 6 *mesi la sterlina si rafforzerà rispetto al dollaro ed è pronto a scommettere sulla sua intuizione la somma di* $\$250,000$. *Lo speculatore potrebbe semplicemente comprare l'equivalente in sterline di* $\$250,000$ *al prezzo spot sperando di conseguire un profitto quando le riconvertirà in dollari. Le sterline, una volta acquistate, verrebbero depositate in un conto fruttifero ed eventualmente rivendute tra* 6 *mesi. Ipotizziamo che il tasso di cambio spot in data* 1 *marzo* 2019 *sia* $1.31559£/\$$. *L'acquisto al prezzo spot darebbe allo speculatore il possesso immediato di* $£190,029$, *per cui se tra* 6 *mesi il tasso di cambio risultasse pari a, diciamo,* $1.42734£/\$$ *[resp.* $1.22734\$/£$*], lo speculatore avrebbe guadagnato [resp. perso]* $\$21,236$ *[resp.* $\$16.770$*]. Un'altra possibilità è quella di assumere una posizione lunga sulla sterlina con* 4 *contratti futures standard a* 6 *mesi (ogni future standard comporta l'acquisto di* $£62,500$*) sapendo che il tasso futures a* 6 *mesi sia di* $1.32734£/\$$. *Se tra* 6 *mesi il tasso di cambio risultasse pari a* $1.42734£/\$$, *i futures consentirebbero allo speculatore di comprare a* $\$1.32734$ *un bene che varrebbe* $\$1.42734$, *con un* *conseguente profitto di* £25,000. *Qualora però tra* 6 *mesi il tasso di cambio risultasse pari a* 1.22734£/$, *lo speculatore, costretto a comprare a* $1.32734 *un bene che varrebbe* $1.22734*, avrebbe perso* £25,000. *Le alternative sembrano quindi dare origine a profitti o perdite lievemente differenti, ma questi calcoli non tengono conto degli interessi che si incassano o si pagano. Infatti, quando si considerano gli interessi percepiti sulle sterline depositate nel conto fruttifero e quelli persi sui dollari vincolati nel deposito di garanzia, i profitti o le perdite derivanti dalle due alternative risultano approssimativamente uguali. In definitiva, la differenza tra le due alternative è rappresentata esclusivamente dal fatto che l'acquisto di sterline a pronti richiede un investimento iniziale di* $250,000*, mentre l'acquisto dei futures richiede solo che lo speculatore effettui un deposito di garanzia di circa* $30,000.

Options

Il *contratto di opzione* (*option*) è un contratto tra due contraenti, un *titolare (holder)* ed un *garante (writer)*, che sancisce l'acquisizione di un diritto e l'assunzione di un obbligo. Grazie alla stipula di un contratto di opzione il titolare, dietro la corresponsione di un *premio (prime)*, acquisisce il diritto di acquistare dal garante, nel caso di *opzione d'acquisto (call option)*, o di vendere al garante, nel caso di *opzione di vendita (put option)*, un attivo finanziario rischioso, *titolo sottostante (underlying risky asset* o *underlying security)*, entro un *scadenza (maturity* o *expiration)* e a un *prezzo d'esercizio (exercise* o *strike price)* pattuiti all'atto della stipula del contratto. Si dice che il titolare assume una *posizione lunga (long position)* sull'opzione. Il garante, in cambio del premio, si obbliga a soddisfare alla scadenza il titolare del diritto d'opzione. Si dice che il garante assume una *posizione corta (short position)* sull'opzione. Differentemente da un contratto forward o futures in cui le parti contraenti sono entrambe obbligate a onorarlo. Un contratto d'opzione garantisce al titolare il diritto di acquistare o vendere senza obbligo d'esercitarlo. Al contrario il garante è obbligato a farsi carico dell'eventuale esercizio dell'opzione da parte del titolare. Inoltre, mentre la negoziazione di contratti forward o future non implica alcun costo, fatta eccezione per il deposito di garanzia nel caso dei future, l'acquisto di un'opzione richiede un pagamento immediato. Da notare che chi acquista un'opzione call scommette che entro la scadenza il prezzo spot del sottostante vada al di sopra del prezzo d'esercizio, ossia assume una *posizione rialzista* (*bullish position*) sul sottostante, mentre chi vende l'opzione call assume una *posizione ribassista* (*bearish position*). Viceversa, chi acquista un'opzione put scommette che entro la scadenza il prezzo spot del sottostante vada al di sotto del prezzo d'esercizio, ossia assume una posizione ribassista sul sottostante, mentre chi vende l'opzione put assume una posizione rialzista. Le opzioni sono negoziabili sia nei mercati over the counter, sia nei mercati organizzati: si tratta in quest'ultimo caso delle cosiddette *listed options*. Il funzionamento dei mercati organizzati è in larga misura simile a quello dei mercati dei future. Le opzioni più comunemente trattate sono le *opzioni americane (american options)* possono essere esercitate in qualsiasi momento antecedente alla data di scadenza. Le opzioni di più semplice modellizzazione matematica sono le *opzioni europee (european options)* che possono essere esercitate solo alla data di scadenza.

Esempio 6 *Consideriamo un contratto di opzione call europea su* 100 *azioni Amazon in data di sottoscrizione* 01 *marzo* 2019*, con scadenza* 21 *giugno* 2019 *e prezzo d'esercizio di* $1,700 *(per azione). Con la titolarità di questo contratto ottieniamo il diritto di acquistare* 100 *azioni Amazon in data* 21 *giugno* 2019 *dal sottoscrittore pagandole* $1,700 *l'una. Ci sono due possibili scenari futuri, secondochè il prezzo spot del titolo Amazon alla scadenza vada sopra oppure sotto il prezzo d'esercizio:*

- *al* 21 *giugno* 2019 *le azioni Amazon hanno un prezzo spot superiore a* $1,700*, ad esempio* $1,896.17*, allora esercitiamo il diritto di opzione, compramo dal sottoscrittore del contratto le* 100 *azioni, pagandole* $170,000*, e le rivendiamo all'istante sul mercato a* $189,617*. Il guadagno è quindi di* $19,617 *al quale va tuttavia sottratto il premio pagato inizialmente per entrare nel contratto;*
- *al* 21 *giugno* 2019 *le azioni Amazon hanno un prezzo spot inferiore a* $1,700 *(o uguale), allora non esercitamo il diritto di opzione, in quanto esercitandolo ci troveremmo a comprare al prezzo*

di $1,700 *delle azioni che sul mercato valgono di meno. Quindi ci limitiamo a perdere il premio versato per entrare nel contratto.*

Come si vede dall'esempio, per ottenere un guadagno, l'acquirente dell'opzione call deve sperare che il prezzo spot alla scadenza si attesti al di sopra del prezzo d'esercizio, mentre il venditore dell'opzione call deve sperare che il prezzo spot alla scadenza si attesti al di sotto del prezzo d'esercizio. Da notare anche che mentre l'acquisto di un'opzione call consente, in linea di principio, guadagni illimitati, la vendita di un'opzione call può causare perdite illimitate. Ciò è illustrato dai due grafici seguenti

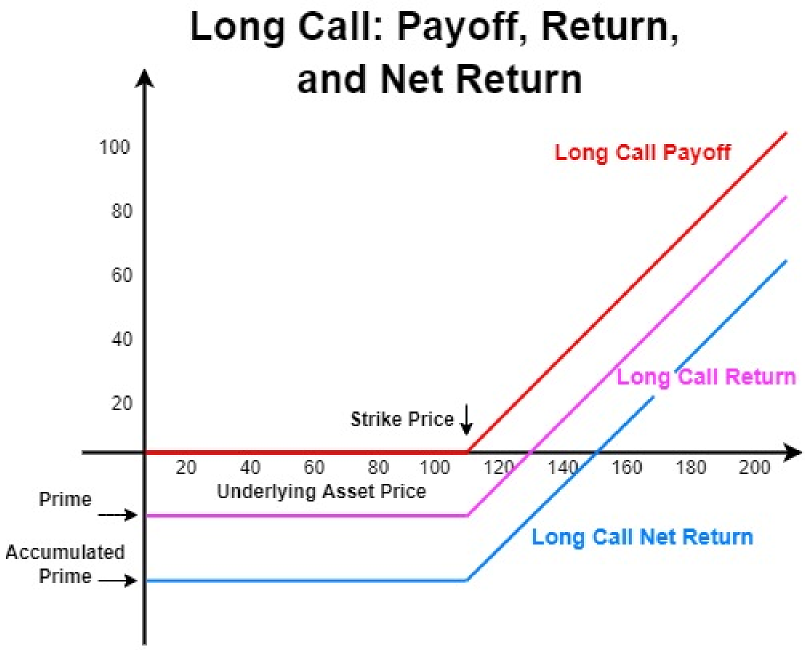

Figura 1.1: Payoff per una posizione lunga su una call option. In ascissa i possibili prezzi spot dell'asset al momento d'esercizio della call. In ordinata i possibili payoff dell'acquirente della call.

Presentiamo adesso un esempio analogo riferito alle opzioni put.

Esempio 7 *Consideriamo un contratto di opzione put europea su* 100 *azioni Amazon in data di sottoscrizione* 05 *marzo* 2019 *con scadenza* 21 *giugno* 2019 *e prezzo d'esercizio* $1,700 *(per azione). Sottoscrivendo questo contratto ottieniamo il diritto di vendere* 100 *azioni Amazon in data* 21 *giugno* 2019 *al prezzo di* $1,700 *l'una. Anche in questo caso ci sono due possibili scenari futuri, secondochè alla scadenza il prezzo spot del titolo Amazon vada sopra oppure sotto il prezzo d'esercizio:*

- *al* 21 *giugno* 2019 *le azioni Amazon hanno, un prezzo spot inferiore a* $1,700*, ad esempio* $1,496.17*, allora compriamo sul mercato le* 100 *azioni pagandole* $149,617 *e, esercitando il diritto di opzione, le rivendiamo all'istante al sottoscrittore del contratto a* $170,000*. Il guadagno è quindi di* $20.383 *ai quali va ancora sottratto il premio pagato inizialmente per entrare nel contratto.*

- *al* 21 *giugno* 2019 *le azioni Amazon hanno, un prezzo spot superiore a* $1,700 *(o uguale), allora rinunciamo a esercitare il diritto di opzione, perchè esercitandolo ci troveremmo a vendere al prezzo di* $1,700 *delle azioni che sul mercato valgono di più. Quindi ci limitiamo a perdere il premio versato per entrare nel contratto.*

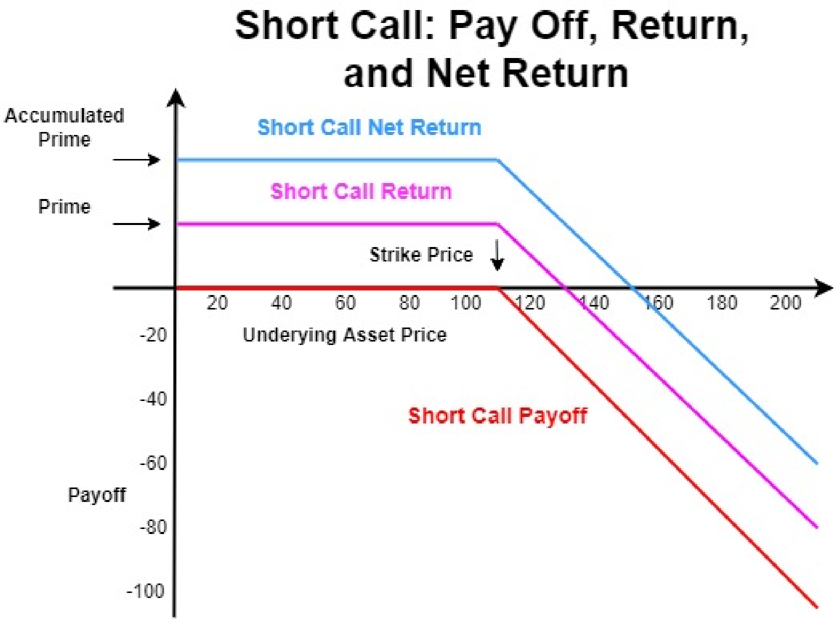

Figura 1.2: Payoff per una posizione corta su una call option. In ascissa i possibili prezzi spot dell'asset al momento d'esercizio della call. In ordinata i possibili payoff del venditore della call.

Simmetricamente al caso delle call, per ottenere un guadagno, l'acquirente dell'opzione put deve sperare che il prezzo spot alla scadenza si attesti al di sotto del prezzo d'esercizio, mentre il venditore dell'opzione put deve sperare che il prezzo spot alla scadenza si attesti al di sopra del prezzo d'esercizio. Tuttavia, differentemente dalle opzioni call, l'acquisto di un'opzione put non permette guadagni superiori alla differenza tra il prezzo d'esercizio e il premio e la vendita di un'opzione put non può causare perdite superiori alla differenza tra il premio e il prezzo d'esercizio.

Da notare che i payoff delle opzioni put e call europee o americane dipendono solo dal valore che il sottostante assume alla data d'esercizio e non dal suo andamento fino a tale data. Nei mercati reali, dove si opera prevalentemente in modalità telematica, l'acquisto di un contratto call [resp. put] è del tutto equivalente ad una scommessa: se il prezzo spot alla scadenza e maggiore [resp. minore] del prezzo d'esercizio viene pagato subito il guadagno, se è minore [resp. maggiore] o uguale al prezzo d'esercizio si perde il premio. Analogamente in caso di vendita di un contratto call [resp. put]: se il prezzo spot alla scadenza e maggiore [resp. minore] del prezzo d'esercizio si paga subito la perdita, se è minore [resp. maggiore] o uguale al prezzo d'esercizio si guadagna il premio. In genere i prezzi d'esercizio sono vicini alle quotazioni giornaliere dei sottostanti e ci danno quindi un'idea delle aspettative degli operatori sulle possibilità di rialzo o di ribasso degli stessi. In particolare un'opzione call o put è detta *at the money* [resp. *near the money*] quando il prezzo d'esercizio è uguale [resp. vicino] al prezzo corrente del sottostante. Un'opzione call [resp. put] è detta *in the money* quando il prezzo d'esercizio è minore [resp. maggiore] del prezzo corrente del sottostante. Un'opzione call [resp. put] è detta *out of the money* quando il prezzo d'esercizio è maggiore [resp. minore] del prezzo corrente del sottostante.

In base a queste caratteristiche, mentre le opzioni forniscono agli speculatori una vera e propria leva finanziaria che permette di amplificare i rendimenti di un investimento nel mercato finanziario, le stesse opzioni possono realizzare una copertura assicurativa contro il rischio di mercato.

Esempio 8 *Il* 05 *marzo* 2019 *uno speculatore vuole assumere una posizione lunga sulle azioni Amazon*

Questo esempio ci dice che invece di comprare tot azioni ad un prezzo "completo", è meglio acquisire molti più diritti di acquisto (che costano meno all'inizio, perché sono solo diritti, non azioni complete), così se mi dice bene e il loro valore aumenta, le vendo a un prezzo maggiore, di cui ne uso una parte per esercitare il diritto di acquisto. Ovviamente, se il prezzo diminuisce, perdo solo il diritto di acquisto, perché non lo esercito.

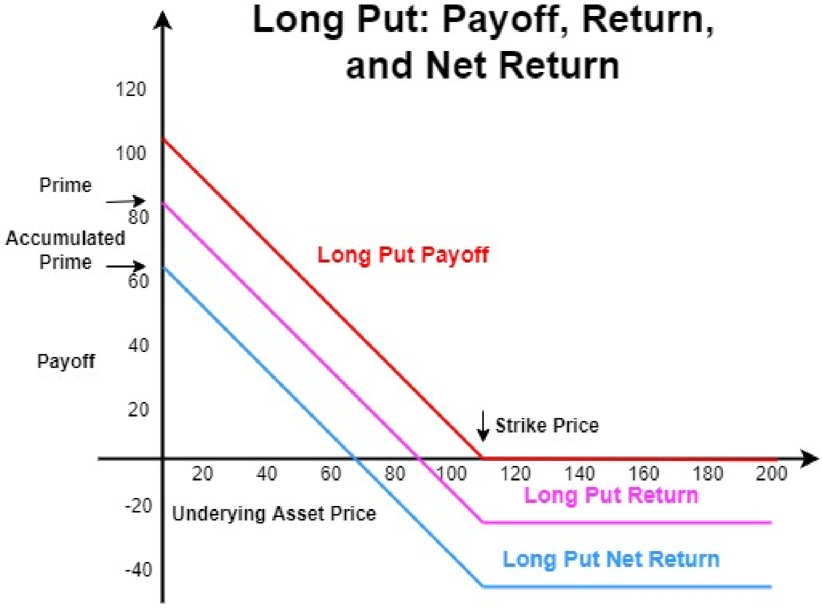

Figura 1.3: Payoff per una posizione lunga su una put option. In ascissa i possibili prezzi spot dell'asset al momento d'esercizio della put. In ordinata i possibili payoff dell'acquirente della put.

quotate al prezzo di $1,696.17, *ritenendo molto probabile che il loro prezzo salga nei successivi mesi. Il* 05 *marzo* 2019 *una call europea con scadenza* 21 *giugno* 2019 *con prezzo d'esercizio di* $1,700 *è quotata a* $102,95. *Nell'ipotesi in cui abbia una disponibilità d'investimento pari a* $169.617, *lo speculatore ha a disposizione due alternative: la prima consiste semplicemente nell'acquisto di* 100 *azioni; la seconda consiste nell'investimento di* $164.000 *per l'acquisto di* 16 *contratti di* 100 *opzioni l'uno per un totale di* 1,600 *opzioni. Supponiamo che l'intuizione dello speculatore sia corretta e che le azioni Amazon si apprezzino effettivamente, ad esempio, fino a* $1,896.17 *alla scadenza. La prima alternativa, consistente nel comprare le azioni, comporterebbe un profitto di* $100 \times \$200 = \$20,000$. *La seconda alternativa è molto più redditizia. Un'opzione call sulle Amazon, con prezzo d'esercizio di* $1,700, *comporterebbe un ricavo di* $196.17 *a opzione, consentendo di acquistare a* $1,700 *l'azione che varrebbe sul mercato* $1,896.17. *Il valore complessivo di tutte le opzioni comprate sarebbe allora pari a* $1,600 \times \$196.17 = \$313,872$. *Pertanto, sottraendo il costo originale sostenuto per l'acquisto delle opzioni, il profitto sarebbe pari a* $\$313,872 - \$164.000 = \$149.872$. *La strategia d'acquisto delle opzioni risulterebbe essere molto più redditizia della strategia consistente nell'acquisto delle azioni. Naturalmente, le opzioni comportano anche maggiori perdite potenziali. Supponiamo che il prezzo dell'azione ribassi, ad esempio sino a* $1,496.17, *alla scadenza. La prima strategia comporterebbe una perdita di* $100 \times \$200 = \$20,000$, *mentre la strategia mediante opzioni, che scadrebbero senza essere state esercitate, causerebbe una perdita di* $164.000, *ossia il premio originariamente pagato.*

Esempio 9 *Consideriamo un investitore che il* 05 *marzo* 2019, *scommettendo sul rialzo nei prossimi mesi delle azioni Amazon, decida di comprarne* 100 *al prezzo corrente di* $1,696.17 *per azione. L'investitore, più prudente dello speculatore, decide di cautelarsi dal rischio che presenta il suo investimento comprando allo stesso tempo* 100 *opzioni put europee sulle azioni Amazon a scadenza* 21 *giugno* 2019 *con prezzo d'esercizio di* $1,700. *In data il* 05 *marzo* 2019 *il premio per l'opzione put con tale prezzo d'esercizio è di* $95.94. *Nel caso in cui le azioni Amazon dovessero effettivamente apprezzarsi sul mercato, ad esempio fino a* $1,896.17, *l'investitore non eserciterà le opzioni ed incasserà il prezzo di mercato delle azioni realizzando così un profitto pari alla differenza tra l'incremento di valore di mercato delle azioni ed il premio pagato per l'acquisto delle opzioni per un totale di* $\$20,000 - \$9,594 = \$10,406$. *Invece, nel caso in cui il titolo dovesse deprezzarsi sul mercato, ad esempio sino a* $1,496.17, *l'investitore*

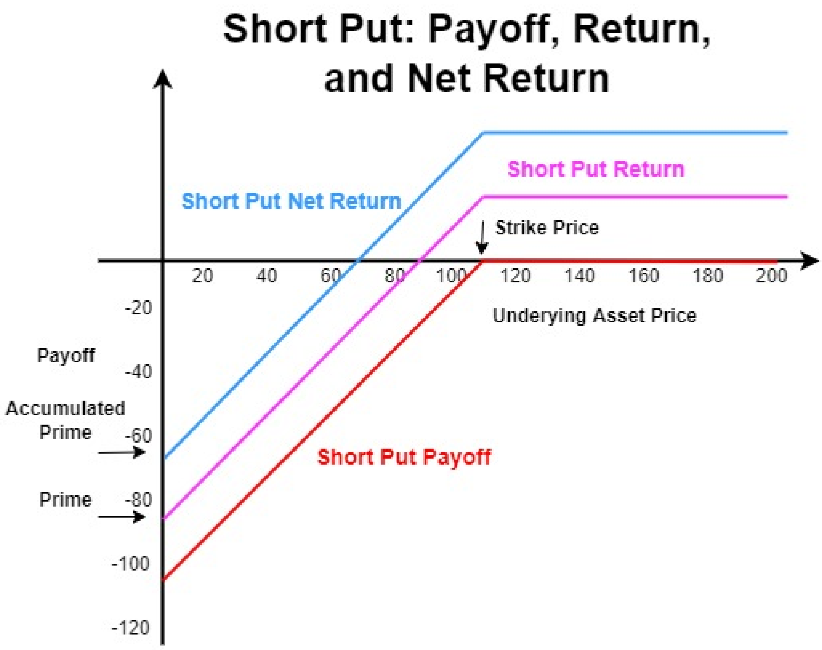

Figura 1.4: Payoff per una posizione corta su una put option. In ascissa i possibili prezzi spot dell'asset al momento d'esercizio della put. In ordinata i possibili payoff del venditore della put.

avrebbe modo di esercitare le sue opzioni put, limitando la sua perdita al costo del premio unitamente all'eventuale differenza tra il prezzo pagato per acquistare le azioni e lo strike delle opzioni per un totale di $\$9,594 - \$383 = \$9,166$*, invece della perdita di* $100 \times \$200 = \$20,000$ *che subirebbe se non avesse comprato le opzioni put.*

Le opzioni call e put sin qui descritte vengono chiamate *plain vanilla* e rappresentano la più semplice tipologia di contratto d'opzione, combinando tra di loro calls e puts si possono definire derivati *non standard* anche molto complicati.

Nelle prossime sezioni entreremo in maggior dettaglio nella presentazione dei contratti d'opzione nell'ambito di semplici modelli matematici di mercato finanziario.

Swaps

I *contratti swaps* sono accordi privati tra una società ed una banca, ma anche tra due società, per scambiarsi dei futuri pagamenti. L'accordo definisce le date in cui i pagamenti vengono scambiati ed il modo in cui devono essere calcolati. Di solito la loro determinazione viene effettuata in base al futuro valore di un tasso d'interesse, un tasso di cambio o qualche altra variabile di mercato.

Il più comune tipo di *swap sul tasso d'interesse (interest rate swap)* è chiamato *plain vanilla*. In questo contratto, una società si impegna a pagare ad un'altra, per un certo numero di anni ed in base a un capitale di riferimento detto *capitale nozionale (notional principal)*, un tasso d'interesse fisso predeterminato. A sua volta, la controparte si impegna a pagare un tasso d'interesse variabile sullo stesso capitale, per lo stesso numero di anni. I pagamenti a tasso variabile vengono calcolati in funzione dell'andamento nel tempo di un prefissato indice di riferimento, che per lo più è rappresentato dal *London InterBank Offer Rate (Libor)*, ovvero il tasso al quale le Banche Centrali offrono fondi ad altre banche nel mercato delle eurovalute.

Esempio 10 *Supponiamo che il* 5 *marzo* 2004 *Microsoft Europe si impegni a pagare per* 3 *anni alla Bank of England un tasso del* 5% *per cento annuo su un capitale nozionale di* £100 *milioni ed in cambio* *la Bank of England si impegni a pagare a Microsoft il Libor a* 6 *mesi sullo stesso capitale nozionale e per la stessa durata triennale. Supponiamo che i pagamenti vengano scambiati ogni* 6 *mesi e che il tasso d'interesse del* 5% *sia composto semestralmente. Il primo scambio di pagamenti ha luogo il* 5 *settembre* 2004, *sei mesi dopo la stipula del contratto. Microsoft paga alla Bank of England* £2, 5 *milioni. Questi sono gli interessi su un capitale di* £100 *milioni al tasso annuo del* 5% *per cento. Di contro la Bank of England paga a Microsoft gli interessi su un capitale di* \$100 *milioni al Libor a* 6 *mesi osservato sei mesi prima del* 5 *settembre* 2004, *ossia esattamente il* 5 *marzo* 2004. *Supponiamo che il* 5 *marzo* 2004 *il Libor a* 6 *mesi sia pari al* 4, 2%. *Quindi la Bank of England paga a Microsoft* £2, 1 *milioni. Si noti che non c'è incertezza circa il primo scambio di pagamenti, dato che il pagamento variabile è determinato in base al Libor osservato nel momento in cui il contratto viene stipulato. Il secondo scambio di pagamenti ha luogo il* 5 *marzo* 2005, *un anno dopo la stipula del contratto. Microsoft paga* £2, 5 *milioni alla Bank of England e la Bank of England paga a Microsoft gli interessi su un capitale di* £100 *milioni in base al Libor a* 6 *mesi osservato sei mesi prima del* 5 *marzo* 2004, *ossia il* 5 *settembre* 2004. *Supponiamo che il* 5 *settembre* 2004 *il Libor a* 6 *mesi sia pari al* 4, 8%, *allora la Bank of England paga a Microsoft un importo pari a* £2, 4 *milioni. In totale lo swap comporta sei scambi di pagamenti. I pagamenti fissi sono sempre uguali a* \$2, 5 *milioni. I pagamenti variabili vengono determinati in base al Libor a* 6 *mesi, osservato sei mesi prima di ciascuna scadenza di pagamento. Ovviamente, gli swaps su tassi d'interesse sono strutturati in modo che una delle due parti remuneri l'altra solo la differenza tra i due pagamenti. Nell'esempio in questione, Microsoft paga alla Bank of England* £0, 4 *milioni il* 5 *settembre* 2004 *e* £0, 1 *milioni il* 5 *marzo* 2005. *Si noti che il capitale viene usato solo per determinare l'importo degli interessi, esso non viene scambiato. Questo è il motivo per cui viene chiamato capitale nozionale.*

L'esempio presentato evidenzia come lo swap possa essere considerato a tutti gli effetti lo scambio di un titolo a tasso fisso contro un titolo a tasso variabile. La posizione di Microsoft Europa è lunga su un titolo a tasso variabile ed è corta su un titolo a tasso fisso. La Bank of England è lunga su un titolo a tasso fisso ed è corta su un titolo a tasso variabile. Questa caratterizzazione dei pagamenti previsti dallo swap aiuta a spiegare perché il tasso variabile dello swap venga fissato sei mesi prima del pagamento. Gli interessi pagati sui titoli a tasso variabile sono in genere fissati all'inizio del periodo al quale si riferiscono e vengono pagati alla fine dello stesso. Gli interest rate swaps più comuni vengono pertanto costruiti nel modo illustrato dall'esempio. Chiaramente gli swaps sui tassi d'interesse sono strumenti finanziari che consentono ad uno dei due contraenti di tutelarsi dall'incertezza sulla variabilità del tasso di cambio e all'altro di speculare proprio su questa variabilità.

Capitolo 2

Single-Period Investment Models

By *single-period investment* we mean any investment characterized by a cash flow stream which occurs only at two dates: an *initial* or *present* date and a *terminal* or *future* date. The difference between the terminal and the initial date of an investment is called the *maturity* of the investment.

For simplicty, write $t = 0$ [resp. $t = T$] for the initial [resp. terminal] date of a single-period investment and write X_0 [resp. X_T] for the *initial* or *present* [resp. *terminal* or *future*] cash flow of the investment. In this case the maturity is just the terminal date of the investment. The amount of the initial cash flow, also termed *principal* in a single-period investment perspective, is observed at the present date. Hence, with reference to the present date, the initial cash flow is usually considered as certain and may be represented by a real number, that is a Dirac random variable, $X_0 \sim Dir\,(X_0)$, concentrated on the amount of the cash flow. On the contrary, the value of the terminal cash flow, also termed *payoff* in a single-period investment perspective, is observed at the future date. Therefore, at the present date, the terminal cash flow should be considered as uncertain and should be represented by a real random variable X_T, defined on some probability space Ω. Depending on whether the state space $X_T\,(\Omega)$ is countable or continuous, we will speak of *countable* or *continuous space state model*. However, in some cases, it may be convenient to consider also the payoff of an investment as certain. For instance, this is the case of the payoff generated by investments in deposits of several US banks or investments in the US Treasury Bills. In fact, in these cases the uncertainty in the payoff of the investment is only due to a possible default of the depositary bank or the USA. However, the Federal Deposit Insurance Corporation provides deposit insurance that guarantees the deposit of member banks for at least $250,000 per depositor, per bank and the eventuality of defaul of the USA is considered to be rather unlikely.

Esempio 11 *Website cost and banners*

From now on, let us assume that the principal X_0 of the single period investment considered is actually certain and the payoff X_T at maturity T is a real random variable.

Definizione 12 *In a forward-looking perspective, we call the* return *or* interest *of the investment at maturity T the difference betweeen the payoff and the principal, that is the random variable*

$$R_T \overset{def}{=} X_T - X_0. \tag{2.1}$$

An interest may be either positive or negative. Accordingly, when $X_0 > 0$, it is called profit *or* loss.

Definizione 13 *Assume that $X_0 \neq 0$. We call the* rate of return *or* rate of interest *of the investment at maturity T the ratio betweeen the interest and the principal, that is the random variable*

$$r_T \overset{def}{=} \frac{R_T}{X_0}. \tag{2.2}$$

Note that the term *interest* [resp. *rate of interest*] rather than *return* [resp. *rate of return*] is more commonly used with reference to investment in loans such as bank deposit, bonds or even private loans.

Definizione 14 *Assume that $X_0 \neq 0$. We call the* accumulation factor *of the investment at maturity T the ratio betweeen the payoff and the principal, that is the random variable*

$$a_T \stackrel{def}{=} \frac{X_T}{X_0}. \tag{2.3}$$

rapporto tra valore finale (v.a.)
e valore iniziale (costante o dirac)

Osservazione 15 *Assume that $X_0 \neq 0$. We clearly have*

$$X_T = X_0 + R_T, \qquad r_T = \frac{X_T - X_0}{X_0}, \qquad R_T = r_T X_0, \qquad X_T = a_T X_0. \tag{2.4}$$

In addition,

$$a_T = 1 + r_T. \tag{2.5}$$

Proof. To prove (2.5), observe that, combinining (2.3), (2.1), (2.2), and simplifying the term X_0, we obtain

$$a_T = \frac{X_T}{X_0} = \frac{X_0 + R_T}{X_0} = \frac{X_0 + r_T X_0}{X_0} = 1 + r_T,$$

as claimed. □

Definizione 16 *In a backward-looking perspective, we call the* discount *generated by an investment at maturity T again the difference betweeen the payoff and the principal. Despite from a mathematical point of view the discount cannot be distinguished by the interest, from an economic or financial point of view distinguishing between the discount and the interest is rather useful. Therefore, for the discount is commonly used a different notation. We also follow this practice and denote the discount at the terminal date T by S_T. As a consequence,*

matematicamente è un "interesse" con
un nome diverso

$$S_T \stackrel{def}{=} X_T - X_0. \tag{2.6}$$

Definizione 17 *Assume that $\mathbf{P}(X_T = 0) = 0$. We call the* rate of discount *of an investment at maturity T the ratio betweeen the discount and the payoff, that is the random variable*

anche il tasso di sconto è un tasso
di interesse con nome diverso.

$$s_T \stackrel{def}{=} \frac{S_T}{X_T}. \tag{2.7}$$

Definizione 18 *Assume that $\mathbf{P}(X_T = 0) = 0$. We call the* discount factor *of the investment at maturity T the ratio betweeen the principal and the payoff, that is the random variable*

se faccio l'inverso del fattore di accumulazione
trovo il fattore di sconto

$$d_T \stackrel{def}{=} \frac{X_0}{X_T}. \tag{2.8}$$

sconto e rendimento sono uguali
solo se parlo di rendimenti sicuri,
cioè non metto in mezzo variabili
aleatorie.

Osservazione 19 *Assume that $\mathbf{P}(X_T = 0) = 0$. We clearly have*

$$X_0 = X_T - S_T, \qquad s_T = \frac{X_T - X_0}{X_T}, \qquad S_T = s_T X_T, \qquad X_0 = d_T X_T. \tag{2.9}$$

In addition,

$$d_T = 1 - s_T. \tag{2.10}$$

Proof. To prove (2.10), observe that

$$d_T = \frac{X_0}{X_T} = \frac{X_T - S_T}{X_T} = \frac{X_T - s_T X_T}{X_T} = 1 - s_T,$$

as claimed. □

Osservazione 20 *Assume that* $X_0 \neq 0$ *and* $\mathbf{P}(X_T = 0) = 0$*. We have*

$$a_T d_T = 1. \tag{2.11}$$

Equivalently,

$$(1 + r_T)(1 - s_T) = 1. \tag{2.12}$$

As a consequence

$$s_T = \frac{r_T}{1 + r_T}, \qquad r_T = \frac{s_T}{1 - s_T}. \tag{2.13}$$

Proof. To prove Equation (2.11), we just apply Equations (2.3) and (2.8). Thus,

$$a_T d_T = \frac{X_T}{X_0} \frac{X_0}{X_T} = 1.$$

Now, combining the latter with (2.10) and (2.10), we obtain Equation (2.12). In the end, Equation (2.13) clearly follows from (2.12). $\square$

$$R_T = X_T - X_0, \qquad r_T = \frac{R_T}{X_0} = \frac{X_T - X_0}{X_0}, \qquad a_T = \frac{X_T}{X_0} = 1 + r_T$$
$$S_T = X_T - X_0, \qquad s_T = \frac{S_T}{X_T} = \frac{X_T - X_0}{X_T}, \qquad d_T = \frac{X_0}{X_T} = 1 - s_T$$

$$R_T = S_T, \qquad a_T d_T = 1, \qquad (1 + r_T)(1 + s_T) = 1,$$
$$s_T = \frac{r_T}{1 + r_T}, \qquad r_T = \frac{s_T}{1 - s_T}$$

2.1 Cox-Ross-Rubinstein Toy Model (Single-Period Multiplicative Binomial Model)

Consideriamo un mercato finanziario uniperiodale dove al tempo $t = 0$ sia possibile investire in un titolo non rischioso, cui ci riferiremo come *bond* e denoteremo con la lettera B, in un titolo rischioso, cui ci riferiremo come *stock* e denoteremo con S, e in titoli derivati di sottostante lo stock, di volta in volta variamente nominati e denotati. A titolo d'esempio possiamo pensare che B ed S corrispondano rispettivamente a un'obbligazione e a un'azione e i derivati a opzioni call o put sull'azione. Assumiamo anche la possibilità di investire in portafogli composti in varie proporzioni mediante il bond, lo stock e i derivati. Al tempo $t = T$, raggiunta la maturità (*maturity*), l'investimento effettuato al tempo $t = 0$ viene liquidato, l'eventuale ricchezza prodotta viene consumata o l'eventuale debito contratto va ripagato. Assumiamo inoltre la validità di alcune ipotesi di natura finanziaria che nella realtà sono solo approssimativamente verificate, senza peraltro eccessivo nocumento alle risultanze del modello. Nello specifico assumiamo che:

1. i tassi di rendimento relativi alle operazioni di prestito e deposito siano gli stessi;

2. gli investimenti possano essere effettuati senza alcuna limitazione quantitativa, in altri termini sia possibile operare su una qualunque frazione di bond, stock e derivati rispetto all'unità di moneta fissata;

3. i costi di transazione siano nulli;

4. le vendite allo scoperto sul bond, sullo stock e sui derivati siano totalmente libere.

Cominciando a focalizzare l'attenzione sul bond e sullo stock, denotiamo con B_0 ed S_0 [resp. B_T ed S_T] i valori di mercato delle unità di bond e di stock al tempo $t = 0$ [resp. $t = T$] e denotiamo con r il tasso di rendimento dell'investimento sul bond alla maturità. In questo caso il valore B_T dell'investimento al tempo $t = T$ sarà semplicemente dato da

$$B_T = (1 + r)B_0. \tag{2.14}$$

Di contro il valore S_T prodotto dall'investimento nello stock al tempo $t = T$ è da ritenersi aleatorio, tratteremo quindi S_T quale variabile aleatoria. Come vedremo, la peculiarià del modello "giocattolo" di Cox, Ross e Rubinstein è proprio nella semplice, ma non semplicistica, modellizzazione della variabile aleatoria S_T.

Proposizione 21 *Stante la (2.14), una somma di denaro di valore M_0 al tempo $t = 0$ ha valore*

$$M_T = (1 + r)M_0 \tag{2.15}$$

al tempo $t = T$*. Viceversa, una somma di denaro di valore* M_T *al tempo* $t = T$ *ha valore*

$$M_0 = \frac{M_T}{1+r} \tag{2.16}$$

al tempo $t = 0$.

Proof. *La disponibilità di denaro* M_0 *al tempo* $t = 0$ *consente l'acquisto di*

$$x = \frac{M_0}{B_0}$$

unità di bond con valore di mercato B_0*. La liquidazione di questo investimento alla maturità* $t = T$ *produce un ammontare* M_T *secondo la formula*

$$M_T = xB_T = \frac{M_0}{B_0}(1+r)B_0 = M_0(1+r).$$

Viceversa, volendo produrre un ammontare M_T *al tempo* $t = T$*, è necessario liquidare*

$$x = \frac{M_T}{B_T}$$

unità bond di valore di mercato B_T*. D'altra parte l'acquisto di tali unità di bond al tempo* $t = 0$ *richiede l'investimento di una somma di denaro* M_0 *pari a*

$$M_0 = xB_0 = \frac{M_T}{B_T}B_0 = \frac{M_T}{B_0(1+r)}B_0 = \frac{M_T}{1+r}.$$

$\square$

Da notare che in conseguenza della Proposizione 21, gli acquisti e le vendite allo scoperto del bond diventano del tutto equivalenti a depositi e scoperti su un conto bancario.

Supponiamo adesso che tutta l'incertezza circa il futuro dell'investimento in stock abbia semplicemente carattere bivalente; cioè che in relazione al take investimento si possano realizzare solamente un esito positivo o un esito negativo. In questo caso l'incertezza è rappresentabile da una variabile aleatoria bernoulliana, che denotiamo con β, definita su un opportuno spazio di probabilità Ω, suscettibile di prendere al tempo $t = T$ i due soli valori u (*up*) e d (*down*), con $u > d$, secondoché per l'investimento in stock si realizzi l'esito positivo o quello negativo. In simboli,

$$\beta \overset{\text{def}}{=} \begin{cases} u, & \mathbf{P}(\beta = u) \equiv p, \\ d, & \mathbf{P}(\beta = d) \equiv q, \end{cases}$$

essendo $\mathbf{P}(\beta = u)$ [risp. $\mathbf{P}(\beta = d)$] la probabilità oggettiva che si realizzi il valore u [risp. d] di β ed essendo $q = 1 - p$. Con questa modellizzazione dell'incertezza una delle scelte naturali per la rappresentazione del valore S_T dell'investimento rischioso al tempo T è data da

$$S_T \overset{\text{def}}{=} \beta S_0. \tag{2.17}$$

Questa scelta conduce al cosiddetto *modello uniperiodale binomiale moltiplicativo* di Cox, Ross e Rubinstein, noto anche come *CRR Toy Model*, che, come vedremo più avanti, si presta anche a un semplice e ricco sviluppo multiperiodale, persino confrontabile con il celebre modello di Black & Scholes. Come immediata conseguenza della (2.17) abbiamo

$$S_T = \begin{cases} S_T^+ \equiv uS_0, & \mathbf{P}(S_T = S^+) \equiv p, \\ S_T^- \equiv dS_0, & \mathbf{P}(S_T = S^-) \equiv q, \end{cases}.$$

Proposizione 22 *L'attesa e la varianza di S_T sono rispettivamente date da*

$$\mathbf{E}[S_T] = (up + dq)S_0, \tag{2.18}$$

e

$$\mathbf{D}^2[S_T] = (u - d)^2 pqS_0^2, \tag{2.19}$$

essendo $\mathbf{E}[\cdot]$ *e* $\mathbf{D}^2[\cdot]$ *gli operatori di* speranza *e* varianza *relativi alla distribuzione di probabilità* (p, q).
Proof. *Abbiamo infatti*

$$\mathbf{E}[S_T] = S_T^+ p + S_T^- q = uS_0 p + dS_0 q = (up + dq)S_0$$

e

$$\begin{aligned}
\mathbf{D}^2[S_T] &= \mathbf{E}[S_T^2] - \mathbf{E}[S_T]^2 \\
&= u^2 S_0^2 p + d^2 S_0^2 q - (up + dq)^2 S_0^2 \\
&= (u^2 p(1 - p) + d^2 q(1 - q) - 2udpq)S_0^2 \\
&= (u^2 + d^2 - 2ud)pqS_0^2 \\
&= (u - d)^2 pqS_0^2.
\end{aligned}$$

□

Ricordiamo che la *rischiosità* di S_T è strettamente legata alla sua varianza. Minore la varianza, minore la rischiosità. La relazione tra rischio e varianza viene colta dalla celebre disuguaglianza di Tchebychev, valida per ogni variabile aleatoria X dotata di momento del secondo ordine $\mathbf{E}[X^2]$ finito. Per una tale variabile aleatoria, indipendentemente dalla sua distribuzione, risulta

$$\mathbf{P}(|X - \mathbf{E}[X]| > \varepsilon) \leqslant \frac{\mathbf{D}^2[X]}{\varepsilon^2}, \tag{2.20}$$

per ogni $\varepsilon > 0$. Stante la (2.20), minore è $\mathbf{D}^2[X]$ minore è la probabilità di scostamento di X dal suo valore atteso $\mathbf{E}[X]$. In altri termini la variabile aleatoria X è meno rischiosa. Nel caso limite $\mathbf{D}^2[X] = 0$ la probabilità che X prenda valori che si scostano dal suo valore atteso per un qualsiasi $\varepsilon > 0$ è nulla: la variabile aleatoria assume con certezza il suo valore atteso e non presenta alcun rischio. Si tratta di una cosiddetta variable aleatoria di Dirac concentrata in $\mathbf{E}[X]$. La conoscenza della specifica distribuzione di X conduce ovviamente a stime più precise della probabilità di un suo scostamento dal valore medio. L'importanza della disuguaglianza di Tchebychev è proprio nel non riferirsi alla specifica distribuzione di X che, in questo senso, la connota come universale.

Definizione 23 *Chiamamo* tasso di rendimento *dell'investimento in stock al tempo T il rapporto*

$$r_T \stackrel{def}{=} \frac{S_T - S_0}{S_0}. \tag{2.21}$$

Osservazione 24 *Il tasso di rendimento dell'investimento in stock è esso stesso una variabile aleatoria. Precisamente,*

$$r_T = \begin{cases} r_T^+ \equiv u - 1, & \mathbf{P}(r_T = r_T^+) \equiv p, \\ r_T^- \equiv d - 1, & \mathbf{P}(r_T = r_T^-) \equiv q. \end{cases} \tag{2.22}$$

Chiaramente,

$$u > d \Leftrightarrow r_T^+ > r_T^-.$$

Definizione 25 *Chiamamo* tasso di rendimento medio *(*mean rate of return*) relativo all'investimento in stock al tempo T il valor medio $\bar{r}_S$ di r_S.*

Osservazione 26 *Abbiamo*

$$\bar{r}_T \equiv \mathbf{E}[r_T] = up + dq - 1. \tag{2.23}$$

Proof. *Infatti*

$$\mathbf{E}[r_T] = r_T^+ p + r_T^- q = (u-1)\,p + (d-1)\,q = up + dq - (p+q) = up + dq - 1.$$

□

Osservazione 27 *Abbiamo*

$$S_T = (1 + r_T)\,S_0 \tag{2.24}$$

e

$$S_0 = \frac{1}{1+\bar{r}_T}\mathbf{E}[S_T]. \tag{2.25}$$

Proof. *L'Equazione (2.24) è diretta conseguenza della (2.21) e l'Equazione (2.25) si ottiene immediatamente applicando l'operatore speranza* $\mathbf{E}[\cdot]$ *alla (2.24).* □

Le Equazioni (2.24) e (2.25) andrebbero rispettivamente comparate con l'Equazione (2.14) per l'evoluzione del titolo non rischioso e con la riscrittura della stessa (2.14) in forma backward come

$$B_0 = \frac{1}{1+r}B_T. \tag{2.26}$$

Definizione 28 *Nell'ambito del CCR Toy Model, chiamiamo* portafoglio *con* posizione non rischiosa *nel bond e* posizione rischiosa *nello stock, più sinteticamente* BS-portafoglio*, una coppia* $\pi \equiv (x, y)$ *la cui componente* x *[resp.* y*] sia la quantità di bond [resp. di stock] in cui investiamo al tempo* $t = 0$*. Qualora la posizione non rischiosa sia positiva [risp. negativa] diciamo che* depositiamo *[risp.* prendiamo a prestito*] l'ammontare* $|x|\,B_0$*. Qualora la posizione rischiosa sia positiva [risp. negativa] diciamo che* acquistiamo *[risp.* vendiamo allo scoperto*] lo stock per un ammontare* $|y|\,S_0$*.*

Definizione 29 *Chiamiamo* valore *al tempo* $t = 0$ *[risp. al tempo* $t = T$*] del BS-portafoglio* $\pi \equiv (x, y)$ *l'ammontare*

$$W_0 \equiv xB_0 + yS_0, \qquad [\text{risp. } W_T \equiv xB_T + yS_T].$$

Da notare che mentre l'ammontare W_0 è un numero reale, ovvero una variabile aleatoria di Dirac concentrata in W_0, l'ammontare W_T è una variabile aleatoria bernoulliana, dal momento che S_T lo è. Infatti,

$$W_T = \begin{cases} W_T^+ \equiv (1+r)xB_0 + yS_T^+, & \mathbf{P}(W_T = W^+) \equiv p, \\ W_T^- \equiv (1+r)xB_0 + yS_T^-, & \mathbf{P}(W_T = W^-) \equiv q, \end{cases}.$$

Definizione 30 *Diciamo che un BS-portafoglio* $\pi \equiv (x, y)$ *è un* arbitraggio *se*

1. $W_0 = 0$*;*
2. $\mathbf{P}(W_T \geqslant 0) = 1$ *e* $\mathbf{P}(W_T > 0) > 0$*.*

Un BS-portafoglio é un arbitraggio se a fronte di un investimento iniziale nullo assicura con certezza un payoff non negativo e con probabilità non nulla un payoff strettamente positivo. Ciò indipendentemente dall'esito dell'investimento nella sua componente rischiosa.

Osservazione 31 *Perchè un BS-portafoglio* $\pi \equiv (x, y)$ *sia d'arbitraggio deve soddisfare la condizione*

$$xB_0 = -yS_0. \tag{2.27}$$

Alla luce della Osservazione 31 un BS-portafoglio d'arbitraggio $\pi \equiv (x, y)$ può essere costituito solo in due modi: prendendo a prestito l'ammontare $|x| B_0$ ($x < 0$) e usandolo interamente per acquistare lo stock ($y > 0$) oppure vendendo lo stock allo scoperto per un ammontare $|y| S_0$ ($y < 0$) e investendo interamente questo ammontare nell'acquisto del titolo non rischioso ($x > 0$).

Proposizione 32 *Nell'ambito del CRR Toy Model l'assenza di un BS-portafoglio d'arbitraggio comporta che necessariamente si abbia*

$$r_T^+ > r > r_T^-. \tag{2.28}$$

Equivalentemente,

$$u > 1 + r > d. \tag{2.29}$$

Proof. *Infatti, se fosse*

$$r_T^+ > r_T^- \geqslant r, \tag{2.30}$$

potremmo costituire un BS-portafoglio $\pi \equiv (x, y)$ prendendo a prestito un ammontare $|x| B_0$ (vendendo allo scoperto $|x|$ unità di bond) e con questo ammontare acquistare $y = |x| B_0/S_0$ unità di stock. Al tempo $t = 0$ il valore di tale portafoglio sarebbe

$$W_0 = xB_0 + yS_0 = -|x| B_0 + \frac{|x| B_0}{S_0} S_0 = 0.$$

D'altra parte, con riferimento all'investimento in stock, a termine del periodo d'investimento, nel peggiore dei casi si realizzerebbe l'ammontare

$$yS_T^- = ydS_0 = \frac{|x| B_0}{S_0} dS_0 = |x| dB_0 = |x| (r_T^- + 1)B_0.$$

Pertanto, stante l'ipotesi (2.30), si potrebbe ripianare il debito, nel frattempo incrementatosi a $|x| (r + 1)B_0$ a causa degli interessi maturati dovuti, senza perdere alcunché. Se però a termine del periodo d'investimento si verificasse il migliore dei casi, si realizzerebbe un ammontare

$$yS_T^+ = yuS_0 = \frac{|x| B_0}{S_0} uS_0 = |x| uB_0 = |x| (r_T^+ + 1)B_0.$$

Quindi, sempre nell'ipotesi (2.30), si potrebbe ripianare il debito di $|x| (r+1)B_0$ e realizzare un guadagno pari a

$$|x| (r_T^+ + 1)B_0 - |x| (r + 1)B_0 = |x| (r_T^+ - r)B_0 > 0.$$

Avremmo quindi $W_T \geqslant 0$, con $\mathbf{P}(W_T > 0) = p > 0$. Se altresì fosse

$$r \geqslant r_T^+ > r_T^-, \tag{2.31}$$

allora potremmo costituire un BS-portafoglio $\pi \equiv (x, y)$ vendendo allo scoperto un'ammontare $|y| S_0$ di stock (vendendo allo scoperto $|y|$ unità di stock) e con questo ammontare acquistare $x = |y| S_0/B_0$ unità di bond. Al tempo $t = 0$ il valore di tale portafoglio sarebbe

$$W_0 = xB_0 + yS_0 = |y| \frac{S_0}{B_0} B_0 - |y| S_0 = 0.$$

A termine del periodo d'investimento, ci si ritroverebbe in ogni caso con un ammontare pari a

$$x(1 + r)B_0 = \frac{|y| S_0}{B_0} B_0(1 + r) = |y| (1 + r)S_0,$$

dovuto al maturare degli interessi prodotti dall'investimento in bond. D'altra parte, rispetto alla vendita allo scoperto dello stock, il peggiore dei casi per l'investitore è che lo stock realizzi il valore di mercato S_T^+. In questo caso per ripianare la vendita allo scoperto sarebbe necessario un'ammontare pari a

$$|y| S_T^+ = |y| uS_0 = |y| (r_T^+ + 1)S_0$$

che, stante l'ipotesi (2.31), è disponibile grazie all'investimento in bond. Se poi lo stock realizzasse il valore di mercato S_T^-, per ripianare lo scoperto sarebbe sufficiente un'ammontare pari a

$$|y|\,S_T^- = |y|\,uS_0 = |y|\,(r_T^- + 1)S_0,$$

che, sempre nell'ipotesi (2.31), l'investitore avrebbe disponibile grazie all'investimento in bond e che in più gli consentirebbe un guadagno pari a

$$|y|\,(1+r)S_0 - |y|\,(r_T^- + 1)S_0 = |y|\left(r - r_T^-\right)S_0 > 0$$

Avremmo quindi $W_T \geqslant 0$, con $\mathbf{P}\left(W_T > 0\right) = q > 0$. In definitiva, in entrambe le ipotesi (2.30) e (2.31), sarebbe possibile costituire un un BS-portafoglio d'arbitraggio. Non resta che concludere che in assenza di BS-portafogli d'arbitraggio deve valere la (2.28). □

Siano $\theta \equiv (u, v)$ e $\pi \equiv (x, y)$ BS-portafogli di valore $uB_0 + vS_0$ e $xB_0 + yS_0$ [resp. $uB_T + vS_T$ e $xB_T + yS_T$] al tempo $t = 0$ [resp. $t = T$].

Proposizione 33 *In assenza di BS-portafogli d'arbitraggio, la condizione*

$$uB_T + vS_T = xB_T + yS_T \tag{2.32}$$

comporta necessariamente che

$$u = x \quad \text{e} \quad v = y. \tag{2.33}$$

In particolare,

$$uB_0 + vS_0 = xB_0 + yS_0 \tag{2.34}$$

Proof. L'Equazione (2.32) comporta che si abbia

$$(u - x)\,B_T = (y - v)\,S_T. \tag{2.35}$$

L'Equazione (2.35), a sua volta, implica

$$(u - x)\,B_T = (y - v)\,S_T^+ \tag{2.36}$$

e

$$(u - x)\,B_T = (y - v)\,S_T^- \tag{2.37}$$

Dalle (2.36) e (2.37) si ricava

$$(u - x)\,B_T = (y - v)\,uS_0$$

e

$$(u - x)\,B_T = (y - v)\,dS_0.$$

Queste ultime combinate tra loro comportano

$$(y - v)\,uS_0 = (y - v)\,dS_0. \tag{2.38}$$

D'altra parte in assenza di BS-portafogli d'arbitraggio si ha $u > d$ (cfr. Equazione (2.29)). Pertanto, dalla (2.38), otteniamo

$$y = v. \tag{2.39}$$

Stante la (2.39), l'Equazione (2.36) implica che

$$(u - x)\,B_T = (u - x)\,(1 + r)\,B_0 = 0,$$

da cui

$$x = u. \tag{2.40}$$

Le Equazioni (2.39) e (2.40) costituiscono la (2.33). □

Definizione 34 *Chiamiamo* probabilità neutrale al rischio *(*risk neutral probability*) una probabilità* $\tilde{\mathbf{P}}$ *su* Ω *avente distribuzione* $(\tilde{p}, \tilde{q})$*, con*

$$\tilde{p} \equiv \tilde{\mathbf{P}}(S_T = S_T^+), \quad \tilde{q} \equiv \tilde{\mathbf{P}}(S_T = S_T^-), \tag{2.41}$$

per la quale risulti

$$S_0 = \frac{1}{1+r}\tilde{\mathbf{E}}[S_T], \tag{2.42}$$

essendo $\tilde{\mathbf{E}}[\cdot]$ *l'operatore speranza rispetto a* $\tilde{\mathbf{P}}$.

L'Equazione (2.42) andrebbe comparata con le Equazioni (2.25) e (2.26). Si nota allora che una probabilità neutrale al rischio consente di ottenere il prezzo del titolo rischioso al tempo $t = 0$, scontando il valor medio del prezzo al tempo $t = T$ mediante lo stesso tasso di rendimento del titolo non rischioso. In altri termini, rispetto a una probabilità neutrale al rischio il titolo rischioso ha rendimento medio pari a quello del titolo non rischioso.

Proposizione 35 *Se esiste una probabilità neutrale al rischio essa è unica.*
Proof. *Spponiamo che esistano due probabilità* $\tilde{\mathbf{P}}$ *e* $\mathring{\mathbf{P}}$ *su* Ω *aventi distribuzioni* $(\tilde{p}, \tilde{q})$ *e* $(\mathring{p}, \mathring{q})$*, dove*

$$\tilde{p} \equiv \tilde{\mathbf{P}}(S_T = S_T^+), \quad \tilde{q} \equiv \tilde{\mathbf{P}}(S_T = S_T^-)$$

e

$$\mathring{p} \equiv \mathring{\mathbf{P}}(S_T = S_T^+), \quad \mathring{q} \equiv \mathring{\mathbf{P}}(S_T = S_T^-),$$

per entrambe le quali valga la (2.42), si dovrebbe allora avere

$$\tilde{\mathbf{E}}[S_T] = \mathring{\mathbf{E}}[S_T].$$

Ossia,

$$S_T^+\tilde{p} + S_T^-\tilde{q} = S_T^+\mathring{p} + S_T^-\mathring{q}.$$

Quest'ultima, tenuto conto delle relazioni $\tilde{q} = 1 - \tilde{p}$, $\mathring{q} = 1 - \mathring{p}$, *comporterebbe chiaramente*

$$\left(S_T^+ - S_T^-\right)\tilde{p} + S_T^- = \left(S_T^+ - S_T^-\right)\mathring{p} + S_T^-.$$

Quindi

$$\tilde{p} = \mathring{p}.$$

Da quest'ultima segue immediatamente l'asserto. □

Proposizione 36 *In assenza di BS-portafogli d'arbitraggio esiste una (unica) probabilità neutrale al rischio* $\tilde{\mathbf{P}}$ *di distribuzione* $(\tilde{p}, \tilde{q})$ *caratterizzata da*

$$\tilde{p} \equiv \tilde{\mathbf{P}}(S_T = S^+) \stackrel{def}{=} \frac{1+r-d}{u-d}, \quad \tilde{q} \equiv \tilde{\mathbf{P}}(S_T = S^-) \stackrel{def}{=} \frac{u-(1+r)}{u-d}. \tag{2.43}$$

Proof. *In assenza di portafogli d'arbitraggio, la validità della (2.29) garantisce che la coppia* $(\tilde{p}, \tilde{q})$ *data dalla (2.43) soddisfa le condizioni*

$$\tilde{p}, \tilde{q} \geqslant 0 \quad e \quad \tilde{q} = 1 - \tilde{p}.$$

Pertanto $(\tilde{p}, \tilde{q})$ *è effettivamente una distribuzione di probabilità. Si ha inoltre*

$$\begin{aligned}
\tilde{\mathbf{E}}[S_T] &= S_T^+\tilde{p} + S_T^-\tilde{q} \\
&= (u\tilde{p} + d\tilde{q})S_0 \\
&= \left(u\frac{(1+r)-d}{u-d} + d\frac{u-(1+r)}{u-d}\right)S_0 \\
&= \frac{(u-d)(1+r)}{u-d}S_0 \\
&= (1+r)S_0,
\end{aligned}$$

cioè la (2.42). □

Proposizione 37 *Se esiste una (unica) probabilità neutrale al rischio* $\tilde{\mathbf{P}}$ *di distribuzione* $(\tilde{p}, \tilde{q})$*, dove* $\tilde{p} \equiv \tilde{\mathbf{P}}(S_T = S_T^+)$ *e* $\tilde{q} \equiv \tilde{\mathbf{P}}(S_T = S_T^-)$*, allora non esistono di BS-portafogli d'arbitraggio.*
Proof. *Stanti le Proposizioni 36 e 35 la distribuzione* $(\tilde{p}, \tilde{q})$ *della probabilità neutrale al rischio* $\tilde{\mathbf{P}}$ *è necessariamente data da*

$$\tilde{p} = \frac{1+r-d}{u-d} \quad e \quad \tilde{q} = \frac{u-(1+r)}{u-d}.$$

D'altra parte, se esistesse anche un BS-portafoglio d'arbitraggio $\pi \equiv (x, y)$ *dovrebbe risultare*

$$xB_0 + yS_0 = 0 \quad e \quad xB_T + yS_T \geqslant 0.$$

La prima di queste due condizioni comporterebbe

$$xB_0 = -yS_0, \tag{2.44}$$

la seconda per l'aleatorietà di S_T *si tradurrebbe nel sistema di disuguaglianze*

$$xB_T + yS_T^+ = x(1+r)B_0 + yuS_0 \geqslant 0, \tag{2.45}$$
$$xB_T + yS_T^- = x(1+r)B_0 + ydS_0 \geqslant 0. \tag{2.46}$$

Ma allora, sostituendo la (2.44) nelle (2.45) e (2.46), otterremmo

$$-y(1+r)S_0 + yuS_0 \geqslant 0$$
$$-y(1+r)S_0 + ydS_0 \geqslant 0$$

da cui

$$u \geqslant 1+r \quad e \quad d \geqslant 1+r$$

che chiaramente impedirebbero a $(\tilde{p}, \tilde{q})$ *di essere un'effettiva distribuzione di probabilità.* □

Alla luce delle Proposizioni 35 -37 possiamo enunciare il seguente fondamentale risultato

Teorema 38 (Assenza d'Arbitraggio ed Esistenza di una Probabilità Neutrale al Rischio.) *Nell'ambito del CRR Toy Model, l'assenza di BS-portafogli d'arbitraggio è equivalente all'esistenza di un'unica probabilità neutrale al rischio.*

Consideriamo adesso la possibilità d'investire in derivati di sottostante lo stock. In particolare, opzioni di acquisto e vendita e portafogli variamente composti con il bond, lo stock e i suoi derivati. In un contesto monoperiodale, non c'è chiaramente differenza tra ozioni europee e americane. Più in generale, ogni derivato è di tipo europeo in quanto il modello monoperiodale in sè prevede che ogni investimento arrivi a scadenza a termine del periodo senza che ci siano possibilità intermedie d'intervento. Ricordiamo che

Definizione 39 *Si definisce* opzione europea d'acquisto *(*european call option*) [risp.* opzione europea di vendita *(*european put option*)] un contratto di compravendita a termine tra due contraenti, con il quale l'acquirente dell'opzione,* titolare *(*holder*), dietro la corresponsione di un premio (*premium*) al venditore,* garante *(*writer*), acquisisce il diritto, senza obbligo, di acquistare dal garante[resp. vendere al garante] un attivo finanziario rischioso, per una quantità determinata, a una scadenza (*exercise date *or* maturity*) e a un prezzo d'esercizio (*exercise *or* strike price*) pattuiti all'atto della stipula del contratto.*

Nel caso di un'opzione europea d'acquisto, indicati con K il suo prezzo d'esercizio e con S_T il prezzo dell'attivo rischioso alla maturità T, il titolare sarà interessato ad esercitare il diritto d'acquisto solo se risulterà

$$S_T - K > 0.$$

Infatti, solo in questo caso, acquistando dal garante l'attivo al prezzo d'esercizio K e riscuotendo immediatamente sul mercato il valore S_T, egli potrà realizzare un utile. Al contrario, nel caso di un'opzione europea di vendita, il titolare sarà interessato al suo esercizio solo se si avrà

$$K - S_T > 0,$$

dal momento che stavolta potrà realizzare un utile acquistando sul mercato lo stock al prezzo S_T e rivendendolo immediatamente al garante al prezzo d'esercizio K. Pertanto, i payoff di un'opzione europea d'acquisto o di vendita sono rispettivamente

$$C_T = \max\{S_T - K, 0\} \equiv (S_T - K)^+ \quad \text{e} \quad P_T = \max\{K - S_T, 0\} \equiv (K - S_T)^+.$$

Considerato il differimento temporale del valore del denaro, il detentore di una opzione call [resp. put] avrà quindi un *rendimento netto (net payoff)* alla scadenza pari a

$$C_T - (1+r)\,C_0 \qquad [\text{resp. } P_T - (1+r)\,P_0].$$

Da notare che essendo S_T una variabile aleatoria anche i payoff C_T e P_T lo sono. Pertanto, alla scadenza gli stessi payoff dipenderanno dall'occorrenza dell'esito aleatorio $\omega \in \Omega$ che determina la realizzazione del valore $S_T(\omega)$. In particolare, nell'ambito del CRR Toy Model un'opzione europea d'acquisto sul titolo rischioso, di premio C_0, prezzo d'esercizio K e maturità T ha valore alla maturità dato da

$$C_T(\omega) \stackrel{\text{def}}{=} \max\{S_T(\omega) - K, 0\} = \begin{cases} \max\{S_T^+ - K, 0\} \equiv C_T^+, & \mathbf{P}\{C_T = C_T^+\} = p, \\ \max\{S_T^- - K, 0\} \equiv C_T^-, & \mathbf{P}\{C_T = C_T^-\} = q. \end{cases} \tag{2.47}$$

Un'opzione europea di vendita sul titolo rischioso, di premio P_0, prezzo d'esercizio K e maturità T ha altresì valore alla maturità dato da

$$P_T(\omega) \stackrel{\text{def}}{=} \max\{K - S_T(\omega), 0\} = \begin{cases} \max\{K - S_T^+, 0\} \equiv P_T^+, & \mathbf{P}\{P_T = P_T^+\} = p, \\ \max\{K - S_T^-, 0\} \equiv P_T^-, & \mathbf{P}\{P_T = P_T^-\} = q. \end{cases} \tag{2.48}$$

Ricordiamo che $S_T^+ \equiv uS_0$ e $S_T^- \equiv dS_0$.

Proposizione 40 *In riferimento a una call C e una put P su uno stesso sottostante rischioso S, con stesso prezzo d'esercizio K e stesso tempo d'esercizio T risulta*

$$C_T - P_T = S_T - K, \tag{2.49}$$

dove C_T [resp. P_T] è il payoff della call [resp. put] at tempo T ed S_T è il prezzo dell'attivo a T.
Proof. *Possiamo infatti scrivere*

$$\begin{aligned} C_T - P_T &= \max\{S_T - K, 0\} - \max\{K - S_T, 0\} \\ &= \frac{|S_T - K| + S_T - K}{2} - \frac{|K - S_T| + K - S_T}{2} \\ &= \frac{|S_T - K| + S_T - K - |K - S_T| - (K - S_T)}{2} \\ &= S_T - K. \end{aligned}$$

□

Da notare che l'Equazione (2.49) dipende solo dalla struttura dei payoff delle opzioni call e put di sottostante un comune titolo rischioso con stesso prezzo d'esercizio e dalla circostanza che tali opzioni siano esercitate allo stesso tempo. Si tratta pertanto di una relazione indipendente dal particolare modello di mercato considerato e anche dalla specifica che si tratti di opzioni europee o americane. Conta solo che le opzioni call e put siano esercitate simultaneamente.

Definizione 41 *Chiamiamo* portafoglio replicante *(*replicating portfolio*) dell'opzione call un BS-portafoglio* $\pi \equiv (x, y)$ *tale che a fine del periodo di contrattazione si abbia*

$$xB_T + yS_T = C_T. \tag{2.50}$$

Proposizione 42 *Esiste un unico BS-portafoglio replicante* $\pi \equiv (x, y)$ *dato da*

$$x = \frac{uC_T^- - dC_T^+}{(1+r)B_0(u-d)}, \quad y = \frac{C_T^+ - C_T^-}{(u-d)S_0}. \tag{2.51}$$

Proof. Sia $\pi \equiv (x, y)$ un ipotetico portafoglio replicante. Dalla (2.50), deve allora aversi

$$x(1+r)B_0 + yuS_0 = C_T^+, \quad \text{e} \quad x(1+r)B_0 + ydS_0 = C_T^-.$$

Questo sistema ammette un'unica soluzione (x, y) data da

$$x = \frac{\begin{vmatrix} C_T^+ & uS_0 \\ C_T^- & dS_0 \end{vmatrix}}{\begin{vmatrix} (1+r)B_0 & uS_0 \\ (1+r)B_0 & dS_0 \end{vmatrix}} = \frac{uC_T^- - dC_T^+}{(1+r)B_0(u-d)},$$

e

$$y = \frac{\begin{vmatrix} (1+r)B_0 & C_T^+ \\ (1+r)B_0 & C_T^- \end{vmatrix}}{\begin{vmatrix} (1+r)B_0 & uS_0 \\ (1+r)B_0 & dS_0 \end{vmatrix}} = \frac{C_T^+ - C_T^-}{(u-d)S_0}.$$

Da qui l'asserto. □

Definizione 43 *Chiamiamo* prezzo di non arbitraggio *(*no arbitrage price*) dell'opzione call il valore del portafoglio replicante* $\pi \equiv (x, y)$ *all'inizio del periodo di contrattazione*

$$C_0 = xB_0 + yS_0. \tag{2.52}$$

Proposizione 44 *Risulta*

$$C_0 = \frac{1}{1+r}\left(\frac{u-(1+r)}{u-d}C_T^- + \frac{(1+r)-d}{u-d}C_T^+\right) = \frac{1}{1+r}\tilde{\mathbf{E}}[C_T]. \tag{2.53}$$

Proof. *Infatti, per la (2.51) e la (2.43), risulta*

$$\begin{aligned}
C_0 &= xB_0 + yS_0 \\
&= \frac{uC_T^- - dC_T^+}{(1+r)B_0(u-d)}B_0 + \frac{C_T^+ - C_T^-}{(u-d)S_0}S_0 \\
&= \frac{1}{1+r}\left(\frac{uC_T^- - dC_T^+}{u-d} + \frac{(1+r)(C_T^+ - C_T^-)}{u-d}\right) \\
&= \frac{1}{1+r}\left(\frac{(u-(1+r))C_T^- + ((1+r)-d)C_T^+}{u-d}\right) \\
&= \frac{1}{1+r}\left(\frac{u-(1+r)}{u-d}C_T^- + \frac{(1+r)-d}{u-d}C_T^+\right) \\
&= \frac{1}{1+r}\left(\tilde{q}C_T^- + \tilde{p}C_T^+\right) \\
&= \frac{1}{1+r}\tilde{\mathbf{E}}[C_T].
\end{aligned}$$

□

Il prezzo di non arbitraggio di un'opzione call europea risponde alla seguente necessità: immaginiamo un operatore finanziario che ad inizio del periodo di contrattazione venda ad un acquirente un'opzione call europea sul titolo rischioso con prezzo d'esercizio K e maturità a fine del periodo di contrattazione T. L'agente realizza l'incasso C_0 ma si espone al rischio d'esercizio dell'opzione alla maturità per un ammontare pari a $C_T = \max\{S_T - K, 0\}$, potenzialmente illimitato. Per eliminare il rischio dovuto a tale esposizione, sfruttando la sola disponibilità C_0 prodotta dalla vendita dell'opzione, l'operatore finanziario costituisce un portafoglio depositando x unità di conto nel bond e acquistando y azioni dello stock al prezzo S_0 (con l'abituale convenzione che valori negativi di x ed y significhino vendite allo scoperto). A fine del periodo di contrattazione, con la sola ricchezza generata dal proprio portafoglio, l'operatore finanziario deve essere in grado di replicare (*hedge*) il valore dell'opzione che deve rifondere al compratore. In definitiva, si devono realizzare sia la seguente condizione di autofinanziamento del portafoglio

$$xB_0 + yS_0 = C_0 \tag{2.54}$$

sia la condizione di copertura dal rischio d'esercizio dell'opzione

$$xB_T + yS_T = \max\{S_T - K, 0\}. \tag{2.55}$$

La condizione di copertura (2.55) consente quindi di determinare x ed y, mentre la condizione di autofinanziamento (2.54) permette di ricavare il prezzo C_0 che l'operatore finanziario deve richiedere per la vendita dell'opzione.

Corollary 45 *Si ha*

$$C_0 = \begin{cases} S_0 - \frac{1}{1+r}K & \text{se } K < dS_0 \\ \frac{(1+r)-d}{(u-d)(1+r)}(uS_0 - K) & \text{se } dS_0 \leqslant K < uS_0 \\ 0 & \text{se } K \geqslant uS_0 \end{cases}. \tag{2.56}$$

Proof. *Ricordando che per ogni $x \in \mathbb{R}$ risulta*

$$\max\{x, 0\} = \frac{|x| + x}{2}$$

stante l'Equazione (2.47) possiamo scrivere

$$C_T^- = \frac{|dS_0 - K| + dS_0 - K}{2} \qquad e \qquad C_T^+ = \frac{|uS_0 - K| + uS_0 - K}{2}.$$

Combinando quest'ultima con la (2.53), otteniamo

$$\begin{aligned} C_0 &= \frac{1}{1+r}\left(\frac{u-(1+r)}{u-d}\frac{|dS_0-K|+dS_0-K}{2}+\frac{(1+r)-d}{u-d}\frac{|uS_0-K|+uS_0-K}{2}\right) \\ &= \frac{1}{(u-d)(1+r)}\left(u\frac{|dS_0-K|+dS_0-K}{2}-d\frac{|uS_0-K|+uS_0-K}{2}\right) \\ &- \frac{1}{u-d}\left(\frac{|dS_0-K|-|uS_0-K|}{2}\right)+\frac{1}{2}S_0. \end{aligned}$$

Quindi, nel caso $K < dS_0$ risulta

$$\begin{aligned} C_0 &= \frac{1}{(u-d)(1+r)}\left(u\left(dS_0-K\right)-d\left(uS_0-K\right)\right)-\frac{1}{u-d}\left(\frac{dS_0-K-(uS_0-K)}{2}\right)+\frac{1}{2}S_0 \\ &= S_0-\frac{1}{1+r}K. \end{aligned}$$

Nel caso $dS_0 \leqslant K < uS_0$ si ha invece

$$\begin{aligned} C_0 &= -\frac{1}{(u-d)(1+r)}d\left(uS_0-K\right)-\frac{1}{u-d}\left(\frac{(K-dS_0)-(uS_0-K)}{2}\right)+\frac{1}{2}S_0 \\ &= -\frac{1}{(u-d)(1+r)}d\left(uS_0-K\right)-\frac{1}{u-d}K+\frac{1}{2}\frac{u+d}{u-d}S_0+\frac{1}{2}S_0 \\ &= \frac{(1+r)-d}{(u-d)(1+r)}uS_0-\frac{d+(1+r)}{(u-d)(1+r)}K \\ &= \frac{(1+r)-d}{(u-d)(1+r)}\left(uS_0-K\right) \end{aligned}$$

Infine nel caso $uS_0 \leqslant K$ otteniamo chiaramente

$$C_0 = 0.$$

Ciò prova completamente l'Equazione (2.56). $\square$

Da notare che nel caso $K \geqslant uS_0$ il prezzo del S_T del titolo alla maturità T è certamente inferiore al prezzo strike K. Quindi un'opzione che sancisce il diritto a comprare un stock a un valore superiore al suo valore massimo alla maturità T non può che avere costo nullo.

Sempre nell'ambito del CRR Toy Model, cosideriamo due generici titoli rischiosi, Y e Z di valori Y_0 e Z_0 al tempo $t=0$ e di valori Y_T e Z_T al tempo $t=T$ tali che

$$Y_T = \begin{cases} Y_T^+, & \mathbf{P}\left(Y_T = Y_T^+\right) = p, \\ Y_T^-, & \mathbf{P}\left(Y_T = Y_T^-\right) = q, \end{cases} \quad \text{e} \quad Z_T = \begin{cases} Z_T^+, & \mathbf{P}\left(Z_T = Z_T^+\right) = p, \\ Z_T^-, & \mathbf{P}\left(Z_T = Z_T^-\right) = q, \end{cases} .$$

Definizione 46 *Chiamiamo* portafoglio *con* posizione non rischiosa *nel bond e* posizioni rischiose *nei titoli Y e Z, più sinteticamente* BSS-portafoglio, *una terna $\pi \equiv (x,y,z)$ la cui componente x [resp. y, resp. z] sia la quantità di bond [resp. di titolo rischioso Y, resp. di titolo rischioso Z] in cui investiamo al tempo $t=0$. Qualora la posizione non rischiosa sia positiva [risp. negativa] diciamo che* depositiamo *[risp.* prendiamo a prestito*] l'ammontare $|x|\,B_0$. Qualora la posizione sul titolo rischioso Y o Z sia positiva [risp. negativa] diciamo che* acquistiamo *[risp.* vendiamo allo scoperto*] il titolo rischioso Y o Z per un ammontare $|y|\,Y_0$ o $|z|\,Z_0$.*

Chiaramente un BS-portafoglio $\pi \equiv (x,y)$ può essere pensato come un BSS-portafoglio $\pi \equiv (x,y,z)$ con componente $z=0$.

Definizione 47 *Chiamiamo* valore *al tempo* $t = 0$ *[risp. al tempo* $t = T$*] del BSS-portafoglio* $\pi \equiv (x, y, z)$ *l'ammontare*

$$W_0 \equiv xB_0 + yY_0 + zZ_0, \qquad \text{[risp. } W_T \equiv xB_T + yY_T + zZ_T\text{]}.$$

Da notare che mentre l'ammontare W_0 è un numero reale, ovvero una variabile aleatoria di Dirac concentrata in W_0, l'ammontare W_T è una variabile aleatoria dal momento che Y_T e Z_T lo sono. Infatti,

$$W_T = \begin{cases} W_T^+ \equiv (1+r)xB_0 + yY_T^+ + zZ_T^+, & \mathbf{P}(W_T = W^+) \equiv p, \\ W_T^- \equiv (1+r)xB_0 + yY_T^- + zZ_T^-, & \mathbf{P}(W_T = W^-) \equiv q. \end{cases}$$

Definizione 48 *Diciamo che un BSS-portafoglio* $\pi \equiv (x, y, z)$ *è un* arbitraggio *se*

1. $W_0 = 0$;
2. $W_T \geqslant 0$ *con* $\mathbf{P}(W_T > 0) > 0$.

Analogamente al caso di un BS-portafoglio, un BSS-portafoglio $\pi \equiv (x, y, z)$ é un arbitraggio se a fronte di un investimento iniziale nullo assicura con certezza un payoff non negativo e con probabilità non nulla un payoff strettamente positivo. Ciò indipendentemente dall'esito dell'investimento nelle sue componenti rischiose.

Lemma 49 *In assenza di BSS portafogli d'arbitraggio, il sussistere dell'uguaglianza*

$$Y_T = Z_T \tag{2.57}$$

comporta necessariamente che

$$Y_0 = Z_0. \tag{2.58}$$

Proof. *Consideriamo un BSS portafoglio* $\pi \equiv (x, y, z)$ *con posizione* x *nel titolo non rischioso e posizione* y *[resp.* z*] nel titolo* Y *[resp.* Z*] e supponiamo che si realizzi l'uguaglianza*

$$Y_T = Z_T.$$

Se al tempo $t = 0$ *risultasse*

$$Y_0 > Z_0,$$

allora vendute allo scoperto y *unita di titolo* Y*, con il ricavato* $|y| Y_0$ *si potrebbero acquistare* $z \equiv |y|$ *unità di titolo* Z *con un esborso pari a* $|y| Z_0$ *ed investire il surplus* $|y|(Y_0 - Z_0)$ *nell'acquisto di* $x \equiv |y|(Y_0 - Z_0)/B_0$ *unità di titolo non rischioso con esborso pari a* $|y|(Y_0 - Z_0)$*. Infatti, si ha chiaramente*

$$|y| Y_0 = |y|(Y_0 - Z_0) + |y| Z_0.$$

Al tempo $t = T$ *la liquidazione del BSS-portafoglio* π *in merito alla posizione* z *nel titolo* Z *produrrebbe un introito aleatorio* zZ_T *tale da garantire in ogni caso la possibilità di ripianare lo scoperto della posizione su* Y*, nel frattempo maturato a* uY_T*, e al contempo la liquidazione della posizione* x *nel titolo non rischioso garantirebbe un introito certo pari a*

$$xB_T = \frac{|u|(Y_0 - Z_0)}{B_0} B_T = \frac{|u|(Y_0 - Z_0) B_0(1+r)}{B_0} = |u|(Y_0 - Z_0)(1+r) > 0.$$

Pertanto π *risulterebbe essere un BSS-portafoglio d'arbitraggio. Un BSS-portafoglio d'arbitraggio del tutto analogo si potrebbe costituire se al tempo* $t = 0$ *fosse* $Y_0 < Z_0$*. Non rimane che concludere circa la veridicità della tesi del Lemma.* $\square$

Proposizione 50 *In assenza di BSS portafogli d'arbitraggio, sussiste necessariamente la relazione di parità put-call*

$$C_0 - P_0 = S_0 - \frac{K}{1+r}. \tag{2.59}$$

Proof. *Si consideri un BSS-portafoglio $\pi \equiv (x, y, z)$ costituito al tempo $t = 0$ da un titolo rischioso Y rappresentato dall'acquisto di una call e dalla vendita di una put sullo stesso sottostante, entrambe di strike K alla maturità T e rispettivi prezzi C_0 e P_0, e si consideri un titolo Z costituito prendendo a prestito l'ammontare $K/(1+r)$ ed acquistando il titolo rischioso sottostante alle opzioni al prezzo S_0. Stante le Osservazioni 21 e 40 alla maturità T risulta*

$$Y_T = C_T - P_T = S_T - K = Z_T.$$

Allora, per il Lemma 49 deve necessariamente aversi

$$X_0 = Y_0,$$

che è l'Equazione (2.59) desiderata. □

Corollary 51 *Si ha*

$$P_0 = \begin{cases} 0 & \text{se } K < dS_0 \\ \frac{u-(1+r)}{(u-d)(1+r)}\,(K - dS_0) & \text{se } dS_0 \leqslant K < uS_0 \\ \frac{K}{1+r} - S_0 & \text{se } K \geqslant uS_0 \end{cases}. \tag{2.60}$$

Proof. *Dalla (2.59) si ha chiaramente*

$$P_0 = C_0 - S_0 + \frac{K}{1+r}.$$

Tenuto allora conto dell'Equazione (2.56) segue immediatamente la (2.60). □

Da notare che nel caso $K < dS_0$ il prezzo del S_T del titolo alla maturità T è certamente superiore al prezzo strike K. Quindi un'opzione che sancisce il diritto a vendere un titolo a un valore inferiore al suo valore minimo alla maturità T non può che avere costo nullo.

Abbiamo visto che nell'ambito del CRR Toy Model il prezzo di un'opzione call europea è replicabile mediante un unico portafoglio replicante. Chiaramente, in conseguenza della parità put-call anche il prezzo di un'opzione put europea è replicabile mediante un unico portafoglio replicante. Queste osservazioni offrono lo spunto per introdurre l'importante nozione di completezza di mercato.

Definizione 52 *Nell'ambito del CRR Toy Model, chiamiamo* derivato *una qualunque variabile aleatoria reale D_T su Ω.*

Osservazione 53 *Un derivato $D_T : \Omega \to \mathbb{R}$ dipende solo dal valore finale del sottostante S, cioè risulta*

$$D_T = F_D(S_T),$$

per un'opportuna funzione $F_D : \mathbb{R} \to \mathbb{R}$. Più specificatamente si ha

$$D_T = \begin{cases} D_T^+ = F_D(S_T^+), & \mathbf{P}\left(D_T = D_T^+\right) = p, \\ D_T^- = F_D(S_T^-), & \mathbf{P}\left(D_T = D_T^-\right) = q. \end{cases}$$

Osservazione 54 *Il titolo non rischioso e il titolo rischioso sono essi stessi derivati.*
Proof. *Per il titolo non rischioso B possiamo scrivere*

$$B_T = F_B(S_T),$$

dove $F_B : \mathbb{R} \to \mathbb{R}$ è la funzione costante data da

$$F_B\,(x) \overset{def}{=} (1 + r)\,B_0, \quad \forall x \in \mathbb{R}.$$

Per il titolo rischioso S possiamo scrivere

$$S_T = F_S(S_T),$$

dove $F_S : \mathbb{R} \to \mathbb{R}$ è la funzione identica data da

$$F_B\,(x) \overset{def}{=} x, \quad \forall x \in \mathbb{R}.$$

□

Osservazione 55 *Il titolo $B + S$ è un derivato.*
Proof. *Per il titolo $B + S$ possiamo scrivere*

$$B_T + S_T = F_{B+S}(S_T),$$

dove $F_{B+S} : \mathbb{R} \to \mathbb{R}$ è la funzione data da

$$F_{B+S}\,(x) \overset{def}{=} (1 + r)\,B_0 + x, \quad \forall x \in \mathbb{R}.$$

□

Definizione 56 *Chiamiamo* portafoglio replicante *(*replicating portfolio*) del derivato D un BS-portafoglio $\pi \equiv (x, y)$, le cui componenti x ed y siano tali che alla maturità T si abbia*

$$D_T = xB_T + yS_T. \tag{2.61}$$

Definizione 57 *Diciamo che un mercato è* completo *se ogni derivato è replicabile da un BS-portafoglio.*

E' possibile provare che nell'ambito del CRR Toy Model, l'assenza di portafogli d'arbitraggio, equivalente all'esistenza e unicità di una probabilità neutrale al rischio (cfr Teorema 38), comporta la completezza del mercato. Abbiamo infatti

Teorema 58 *Nell'ambito del CRR Toy Model, l'assenza di BS-portafogli d'arbitraggio implica la completezza di mercato. Viceversa, la completezza di mercato implica l'assenza di BS-portafogli d'arbitraggio.*

Proof. Ricordiamo che l'assenza di BS-portafogli d'arbitraggio equivale all'esistenza di un'unica probabilità neutrale al rischio $\tilde{\mathbf{P}}$ su Ω di distribuzione $(\tilde{p}, \tilde{q})$ (cfr Definizione 34 e Teorema 38). Stante l'Osservazione (53), per ogni ogni derivato D sono possibili al tempo $t = T$ solo due realizzazioni. Infatti, denotato con D_T il valore del derivato D al tempo $t = T$, possiamo scrivere

$$D_T = F_D\,(S_T) \equiv \left(D_T^+, D_T^-\right)$$

essendo D_T^+ [resp. D_T^-] la realizzazione del derivato all'occorrenza della realizzazione S_T^+ [resp. S_T^-] del titolo rischioso. Quindi, l'insieme di tutti i possibili derivati si presta a essere rappresentato come un sottoinsieme di $\mathbb{R}^2$. Denotiamo ora con Π il sottoinsieme dei BS-portafogli $\pi \equiv (x, y)$ che replicano i derivati. Notiamo che Π è un sottospazio di $\mathbb{R}^2$ e che, stante l'Osservazione (55), il portafoglio $(1, 1) \in \Pi$. D'altra parte, se qualche derivato non potesse essere replicato da un opportuno portafoglio $\pi \in \Pi$ si avrebbe necessariamente $\Pi \subset \mathbb{R}^2$. Considerato allora il prodotto scalare $\langle \cdot, \cdot \rangle_{\tilde{\mathbf{P}}} : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ dato da

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\tilde{\mathbf{P}}} \stackrel{\text{def}}{=} x_1 y_1 \tilde{p} + x_2 y_2 \tilde{q}, \quad \forall \mathbf{x} \equiv (x_1, x_2), \mathbf{y} \equiv (y_1, y_2) \in \mathbb{R}^2$$

sarebbe possibile determinare un $\xi \in \mathbb{R}^2 - \{\mathbf{0}\}$ ortogonale a Π. In termini delle componenti di $\xi \equiv (\xi_1, \xi_2)$ avremmo allora

$$\xi_1 x \tilde{p} + \xi_2 y \tilde{q} = 0,$$

per ogni $\pi \equiv (x, y) \in \Pi$. In particolare, poichè $(1, 1) \in \Pi$, si dovrebbe avere

$$\xi_1 \tilde{p} + \xi_2 \tilde{q} = 0. \tag{2.62}$$

Fissato $\lambda > 1$ definiamo

$$\mathbf{P}_\xi \stackrel{\text{def}}{=} (p_\xi, q_\xi) = \left(\left(1 + \frac{\xi_1}{\lambda \|\xi\|_\infty}\right) \tilde{p}, \left(1 + \frac{\xi_2}{\lambda \|\xi\|_\infty}\right) \tilde{q} \right)$$

dove

$$\|\xi\|_\infty = \max \{|\xi_1|, |\xi_2|\}.$$

Chiaramente,

$$\left| \frac{\xi_k}{\lambda \|\xi\|_\infty} \right| = \frac{|\xi_k|}{\lambda \|\xi\|_\infty} \leqslant \frac{1}{\lambda} < 1,$$

per $k = 1, 2$. Si ha quindi

$$p_\xi, q_\xi > 0.$$

Inoltre, stante la (2.62),

$$p_\xi + q_\xi = \left(1 + \frac{\xi_1}{\lambda \|\xi\|_\infty}\right) \tilde{p} + \left(1 + \frac{\xi_2}{\lambda \|\xi\|_\infty}\right) \tilde{q} = \tilde{p} + \tilde{q} + \frac{1}{\lambda \|\xi\|_\infty} (\xi_1 \tilde{p} + \xi_2 \tilde{q}) = 1.$$

In definitiva, la coppia (p_ξ, q_ξ) individuerebbe una distribuzione di probabilità caratterizzante una certa $\mathbf{P}_\xi \neq \tilde{\mathbf{P}}$. Ma per tale $\mathbf{P}_\xi$ avremmo

$$\begin{aligned}
\frac{1}{1+r} \mathbf{E}^{\mathbf{P}_\xi}[S_T] &= \frac{1}{1+r} \left(S_T^+ p_\xi + S_T^- q_\xi \right) \\
&= \frac{1}{1+r} \left(S_T^+ \left(1 + \frac{\xi_1}{k \|\xi\|_\infty}\right) \tilde{p} + S_T^- \left(1 + \frac{\xi_2}{k \|\xi\|_\infty}\right) \tilde{q} \right) \\
&= \frac{1}{1+r} \left(S_T^+ \tilde{p} + S_T^- \tilde{q} + \frac{1}{k \|\xi\|_\infty} (\xi_1 \tilde{p} + \xi_2 \tilde{q}) \right) \\
&= \frac{1}{1+r} \left(S_T^+ \tilde{p} + S_T^- \tilde{q} \right) \\
&= S_0
\end{aligned}$$

Pertanto $\mathbf{P}_\xi$ sarebbe essa stessa una probabilità neutrale al rischio. Ciò contraddirebbe l'unicità della probabilità neutrale al rischio.

Riguardo alla prova che la completezza di mercato implica l'assenza di BS-portafogli d'arbitraggio, osserviamo che se esistesse un BS-portafoglio d'arbitraggio $\pi^* \equiv (x^*, y^*)$ dovrebbe aversi dalla condizione $W_0 = 0$

$$x^* B_0 = -y^* S_0 \tag{2.63}$$

e dalla condizione $W_T \geqslant 0$

$$x^* B_T + y^* S_T \geqslant 0 \Leftrightarrow \begin{cases} x^* (1+r) B_0 + y^* u S_0 \geqslant 0 \\ x^* (1+r) B_0 + y^* d S_0 \geqslant 0 \end{cases}. \tag{2.64}$$

Non potendo essere $x^* = 0$, il che comporterebbe $y^* = 0$ e $W_T = 0$ con certezza, assumiamo $x^* > 0$. Combinando le (2.63) e (2.64), avremmo

$$x^* (1+r-u) B_0 \geqslant 0 \quad \text{e} \quad x^* (1+r-d) B_0 \geqslant 0,$$

quindi necessariamente

$$1+r \geqslant u \quad \text{e} \quad 1+r \geqslant d,$$

sempre con la convenzione $d < u$. Pertanto, se fosse possibile replicare una put sullo stock con strike K take che $dS_0 \leqslant K < uS_0$ con un BS portafoglio $\pi \equiv (x, y)$, dalla condizione

$$P_T = x B_T + y S_T \Leftrightarrow \begin{cases} P_T^+ = x (1+r) B_0 + y u S_0 \\ P_T^- = x (1+r) B_0 + y d S_0 \end{cases}$$

avremmo

$$x = \frac{\begin{vmatrix} P_T^+ & uS_0 \\ P_T^- & dS_0 \end{vmatrix}}{\begin{vmatrix} (1+r)B_0 & uS_0 \\ (1+r)B_0 & dS_0 \end{vmatrix}} = \frac{uP_T^- - dP_T^+}{(1+r)B_0(u-d)}$$

e

$$y = \frac{\begin{vmatrix} (1+r) B_0 & P_T^+ \\ (1+r) B_0 & P_T^- \end{vmatrix}}{\begin{vmatrix} (1+r)B_0 & uS_0 \\ (1+r)B_0 & dS_0 \end{vmatrix}} = \frac{P_T^+ - P_T^-}{(u-d)S_0}$$

D'altra parte, essendo $dS_0 \leqslant K < uS_0$, sarebbe

$$P_T = \frac{|K - S_T| + K - S_T}{2} \Rightarrow \begin{cases} P_T^+ = 0 \\ P_T^- = K - dS_0 \end{cases}.$$

Quindi

$$x = \frac{u (K - dS_0)}{(1+r)B_0(u-d)} \quad \text{e} \quad y = \frac{-(K - dS_0)}{(u-d)S_0}.$$

Otterremo allora un prezzo della put al tempo $t = 0$

$$P_0 = xB_0 + yS_0 = \frac{u (K - dS_0) B_0}{(1+r)B_0(u-d)} - \frac{-(K - dS_0) S_0}{(u-d)S_0} = \frac{u - (1+r)}{(u-d)(1+r)} (K - dS_0) < 0,$$

il che è assurdo. Similmente, assumendo $x^* < 0$ avremmo

$$y^* (u - (1+r)) S_0 \geqslant 0 \quad \text{e} \quad y^* (d - (1+r)) S_0 \geqslant 0,$$

quindi necessariamente

$$u \geqslant 1+r \quad \text{e} \quad d \geqslant 1+r.$$

Sempre con la convenzione $d < u$, se fosse possibile replicare una call sullo stock con strike K take che $dS_0 \leqslant K < uS_0$, con calcoli analoghi a quelli sopra presentati, otterremo un prezzo della put al tempo $t = 0$

$$C_0 = \frac{(1+r) - d}{(u-d)(1+r)} (uS_0 - K) < 0,$$

anche in questo caso assurdo. In definitiva, l'esistenza di un BS-portafoglio d'arbitraggio implicherebbe l'esistenza di derivati non replicabili. Vale pertanto l'asserto. □

Per concludere questa sezione consideriamo anche la possibilità d'investire in futures di sottostante lo stock. A tale proposito ricordiamo che

Definizione 59 *Si definisce* future *un contratto di compravendita a termine tra due contraenti, con il quale l'acquirente del future (*buyer*), sottoscrive l'obbligo di acquistare dal venditore (*seller*) un attivo finanziario rischioso, per una quantità determinata, a una scadenza (*exercise date *or* expiration date *or* maturity*) e a un prezzo (*exercise *or* delivery price*) pattuiti all'atto della stipula del contratto. Di contro il venditore assume l'obbligo di vendere all'acquirente il titolo finanziario in questione nelle modalità previste dal contratto.*

A differenza del contratto d'opzione nessuna delle parti contraenti paga un premio per entrare nel contratto. Tuttavia il contratto va onorato indipendentemente dal valore di mercato che possa assumere il sottostante alla data d'esercizio. Per cui denotato con F_0 [resp. F_T] il valore del future al tempo $t = 0$ [resp. $t = T$] e con K il prezzo d'esercizio di un contratto future F di sottostante lo stock S, abbiamo

$$F_0 = 0 \quad \text{e} \quad F_T = S_T - K, \tag{2.65}$$

L'aquirente del future realizzerà un profitto [resp. una perdita] se $S_T - K > 0$ [resp. $S_T - K < 0$].

Nell'ambito del CRR Toy Model, sempre a differenza del contratto d'opzione, in assenza di BSS portafogli d'arbitraggio il prezzo d'esercizio K del future è univocamente determinato. Si ha infatti

Proposizione 60 (Spot-Futures Parity Theorem) *In assenza di BSS-portafogli d'arbitraggio risulta*

$$K = \tilde{\mathbf{E}}\left[S_T\right]. \tag{2.66}$$

Equivalently

$$K = (1 + r)\, S_0. \tag{2.67}$$

Proof. In assenza di BSS-portafogli d'arbitraggio esiste un'unica probabilità neutrale al rischio $\tilde{\mathbf{P}}$ tale che, considerato l'operatore speranza $\tilde{\mathbf{E}}\left[\cdot\right]$ ad essa associato, risulti

$$F_0 = \frac{1}{1+r}\tilde{\mathbf{E}}\left[F_T\right]$$

Dalla (2.65) otteniamo allora

$$\frac{1}{1+r}\left(\tilde{\mathbf{E}}\left[S_T\right] - K\right) = 0.$$

Peranto, considerata l'Equazione (2.42) segue immediatamente la (2.66).

Possiamo dare un'altra prova mostrando concretamente come la violazione dell'Equazione (2.67) renda possibile la costruzione di un BSS-portafoglio d'arbitraggio.

Supponiamo che al tempo $t = 0$ risulti $K > (1 + r)\, S_0$. Allora prendendo a prestito una somma di denaro S_0 sarebbe possibile comprare un'unità di stock S al prezzo S_0 e vendere a costo $F_0 = 0$ un future sull'unità di stock al prezzo d'esercizio K. Costituiremmo così un BSS-portafoglio $\pi \equiv (x, y, z)$ di posizione nel bond $x \equiv -S_0/B_0$ di posizione nello stock $y \equiv 1$ e di posizione nel future $z = 1$. Un tale portafoglio ha chiaramente valore

$$W_0 = xB_0 + yS_0 + zF_0 = -S_0 + S_0 + 0 = 0.$$

Al tempo $t = T$ la posizione debitoria avrà valore incrementato a $xB_T = -\,(1 + r)\, S_0$, la posizione nello stock si porterà a valore S_T e la posizione nel future garantira l'introito $K - S_T$ derivante dall'incasso K e dalla contemporanea cessione dello stock di valore S_T. Il payoff del portafoglio sarà quindi

$$W_T = xB_T + yS_T + zF_T = -\,(1 + r)\, S_0 + S_T + K - S_T = K - (1 + r)\, S_0 > 0,$$

indipendentemente dalla realizzazione del prezzo S_T del sottostante. Avremmo pertanto costituito un portafoglio d'arbitraggio.

Con ragionamento analogo, ipotizzare che al tempo $t = 0$ risulti $K < (1 + r) S_0$ dà modo di costituire un altro portafoglio d'arbitraggio. Infatti, potremmo vendere allo scoperto un'unità di stock S al prezzo S_0, depositare la somma realizzata nel bond B, comprando S_0/B_0 unità di Bond, e comprare a costo zero un future sullo stock al prezzo d'esercizio K. Costituiremmo così un BSS-portafoglio $\pi \equiv (x, y, z)$ di posizione nel bond $x \equiv S_0/B_0$ di posizione nello stock $y_0 \equiv -1$ e di posizione nel future $z = 1$. Anche un tale portafoglio ha valore

$$W_0 = xB_0 + yS_0 + zF_0 = S_0 - S_0 + 0 = 0.$$

Al tempo $t = T$ il valore della la posizione nel bond sarà incrementato a $xB_T = S_0 (1 + r)$, la posizione nello stock darà luogo a uno scoperto pari a $-S_T$ e la posizione nel future garantira l'introito $S_T - K$ derivante dall'acquisto del titolo di valore di mercato S_T al prezzo K. Il payoff portafoglio sarà quindi

$$W_T = xB_T + yS_T + zF_T = (1 + r) S_0 - S_T + S_T - K = (1 + r) S_0 - K > 0$$

indipendentemente dalla realizzazione del prezzo S_T del sottostante. Anche in questo caso avremmo costituito un portafoglio d'arbitraggio.

In definitiva l'ipotizzata assenza di portafogli d'arbitraggio conduce ad accettare la veridicità dell'Equazione (2.67). $\square$

2.1.1 Parameter Calibration

I parametri del CRR Toy model sono:

1. il tasso di rendimento non rischioso r del bond B;
2. le realizzazioni u e d della variabile aleatoria di Bernoulli introdotta per rappresentare l'incertezza;
3. le componenti $\tilde{p}$ e $\tilde{q}$ della distribuzione di probabilità neutrale al rischio $(\tilde{p}, \tilde{q})$.

Dall'Equazione (2.43) sappiamo che

$$\tilde{p} = \frac{1 + r - d}{u - d} \quad \text{e} \quad \tilde{q} = \frac{u - (1 + r)}{u - d}. \tag{2.68}$$

Inoltre, denotata con σ la volatilità del tasso di rendimento r_T dello stock S, definito dall'Equazione (2.21), con gli stessi calcoli impiegati per l'Equazione (2.19), abbiamo

$$\sigma^2 = \tilde{\mathbf{D}}^2 [r_T] = \frac{\tilde{\mathbf{D}}^2 [S_T]}{S_0^2} = (u - d)^2 \tilde{p}\tilde{q}. \tag{2.69}$$

Le Equazioni (2.68) e (2.69) consentono di determinare i parametri interni del modello $u, d, \tilde{p}, \tilde{q}$ una volta disponibili le stime del tasso di rendimento privo di rischio r e della volatilità del tasso di rendimento dello stock σ. Tuttavia va notato che abbiamo a disposizione tre equazioni per i quattro parametri. Pertanto per poterli determinare dobbiamo introdurre un'ulteriore equazione che caratterizza una delle possibili versioni del modello. L'equazione proposta da Cox, Ross e Rubinstein è

$$d = \frac{1}{u}. \tag{2.70}$$

Un'altra equazione proposta da Jarrow e Rudd è

$$\tilde{p} = \frac{1}{2}. \tag{2.71}$$

Combinando le Equazioni (2.68) segue

$$\tilde{p}\tilde{q} = \frac{(1+r-d)\,(u-(1+r))}{(u-d)^2}.$$

Sostituendo quest'ultima nella (2.69) otteniamo

$$\sigma^2 = (u-d)^2\,\tilde{p}\tilde{q} = (1+r-d)\,(u-(1+r)) = (1+r)\,(u+d) - (1+r)^2 - du.$$

Ora, stante l'Equazione (2.70), risulta

$$\frac{(1+r)^2 + 1 + \sigma^2}{1+r} = \frac{u^2+1}{u}$$

ossia

$$u^2 - \frac{(1+r)^2 + 1 + \sigma^2}{1+r} u + 1 = 0. \tag{2.72}$$

Risolvendo la (2.72) otteniamo la determinazione di Cox, Ross e Rubinstein dei parametri u e d

$$\begin{aligned} u &= \frac{1}{2}\frac{1+\sigma^2+(1+r)^2}{1+r} + \frac{1}{2}\sqrt{\left(\frac{1+\sigma^2+(1+r)^2}{1+r}\right)^2 - 4} \\ &= \frac{1}{2(r+1)}\left(1+\sigma^2+(1+r)^2+\sqrt{\left(1+\sigma^2+(1+r)^2\right)^2 - 4(1+r)^2}\right) \\ &= \frac{1}{2(r+1)}\left(1+\sigma^2+(1+r)^2+\sqrt{(r^2+\sigma^2)(r^2+4r+\sigma^2+4)}\right) \end{aligned} \tag{2.73}$$

e

$$\begin{aligned} d &= \frac{2(r+1)}{1+\sigma^2+(1+r)^2+\sqrt{(r^2+\sigma^2)(r^2+4r+\sigma^2+4)}} \\ &= \frac{2(r+1)\left(1+\sigma^2+(1+r)^2-\sqrt{(r^2+\sigma^2)(r^2+4r+\sigma^2+4)}\right)}{\left(1+\sigma^2+(1+r)^2\right)^2 - (r^2+\sigma^2)(r^2+4r+\sigma^2+4)} \\ &= \frac{1}{2(r+1)}\left(1+\sigma^2+(1+r)^2-\sqrt{(r^2+\sigma^2)(r^2+4r+\sigma^2+4)}\right). \end{aligned} \tag{2.74}$$

Invece, stante la (2.71), dall'Equazione (2.69) otteniamo

$$4\sigma^2 = (u-d)^2$$

ossia

$$2\sigma = u - d$$

Sostituendo quest'ultima nelle Equazioni (2.68) si ottiene allora

$$\frac{1+r-d}{2\sigma} = \frac{1}{2} \quad \text{e} \quad \frac{u-(1+r)}{2\sigma} = \frac{1}{2}$$

da cui

$$d = 1+r-\sigma \quad \text{e} \quad u = 1+r+\sigma$$

Alla luce di quanto osservato ci rimangono da stimare il tasso di rendimento non rischioso r del bond B e la volatilità σ del tasso di rendimento r_T dello stock S. Per questi parametri dobbiamo ricorrere ai dati del mercato.

Per stimare la volatilità σ del tasso di rendimento r_T dello stock S, dobbiamo ricorrere ai dati storici sullo stock. Supponiamo di avere a disposizione i prezzi di chiusura giornalieri dello stock per un ampio intervallo di tempo passato, ad esempio un anno. Denotiamo con

$$S_1, S_2 \ldots, S_N, S_{N+1},$$

convenendo che

$$S_{N+1} \equiv S_0,$$

le variabili aleatorie la cui realizzazione ha dato luogo al prezzo di chiusura dello stock nell'n-simo giorno di contrattazione per $n = 1, \ldots, N+1$, dove $N+1$ è il numero di giorni di mercato del trascorso anno di riferimento. Scriviamo anche

$$\Delta t = 1/\left(N+1\right),$$

per indicare la frazione dell'anno relativa a una giornata di mercato. Formalmente il tasso di rendimento dello stock a termine dell'$n+1$-simo giorno di contrattazione, inteso come variabile aleatoria, è definito come

$$r_n \overset{\text{def}}{=} \frac{S_{n+1} - S_n}{S_n}, \quad \forall n = 1, \ldots, N,$$

avendo assunto che il prezzo di apertura dello stock nell'$n+1$-simo giorno di contrattazione coincida con il prezzo di chiusura dell'n-esimo giorno. Tuttavia, per varie ragioni che esporremo, si preferisce considerare il rendimento logaritmo dello stock a termine dell'$n+1$-simo giorno di contrattazione definito come

$$\rho_n \overset{\text{def}}{=} \log\left(\frac{S_{n+1}}{S_n}\right), \quad \forall n = 1, \ldots, N.$$

Per il momento ci limitiamo a notare che quando $S_{n+1}/S_n \approx 1$ abbiamo $S_{n+1}/S_n - 1 \approx 0$ e pertanto

$$\rho_n = \log\left(\frac{S_{n+1}}{S_n}\right) = \log\left(1 + \left(\frac{S_{n+1}}{S_n} - 1\right)\right) \approx \frac{S_{n+1}}{S_n} - 1 = \frac{S_{n+1} - S_n}{S_n} = r_n.$$

Uno degli assunti portanti della teoria classica dei mercati finanziari è che i rendimenti logaritmici siano indipendenti e normalmente distribuiti. Più precisamente, in termini di distribuzione,

$$\rho_n \sim N\left(\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t, \sigma^2 \Delta t\right),$$

dove μ [resp. σ^2] è il tasso di rendimento medio [resp. la varianza] dei rendimenti logaritmici su base annua. Ciò significa che la variabile aleatoria S_n/S_{n-1} è lognormalmente distribuita ossia

$$\frac{S_{n+1}}{S_n} = \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t + \sigma\sqrt{\Delta t} Z_n\right), \quad Z_n \sim N\left(0, 1\right).$$

per ogni $n = 1, \ldots, N$. In conseguenza, per le propretà delle variabili aleatorie lognormalmente distribuite[1], otteniamo

$$\mathbf{E}\left[\frac{S_{n+1}}{S_n}\right] = \exp\left(\mu \Delta t\right) \quad \text{e} \quad \mathbf{D}^2\left[\frac{S_{n+1}}{S_n}\right] = \exp\left(2\mu\Delta t\right)\exp\left(\sigma^2 \Delta t - 1\right),$$

[1] Se X è una variabile aleatoria log-normalmente distribuita, allora, posto $\mathbf{E}\left[\ln\left(X\right)\right] \equiv \mu$ e $\mathbf{D}^2\left[\ln\left(X\right)\right] \equiv \sigma^2$, possiamo scrivere

$$X = \exp\left(\mu + \sigma Z\right), \quad Z \sim N\left(0, 1\right).$$

In conseguenza,

$$\mathbf{E}\left[X\right] = \mathbf{E}\left[e^{\mu + \sigma Z}\right] = \mathbf{E}\left[e^{\mu} e^{\sigma Z}\right] = e^{\mu}\mathbf{E}\left[e^{\sigma Z}\right]$$

per ogni $n = 1, \dots, N$. L'ulteriore ipotesi dell'indipendenza delle variabili aleatorie $\rho_1 \equiv \ln\left(\frac{S_2}{S_1}\right), \dots, \rho_n \equiv \ln\left(\frac{S_{N+1}}{S_N}\right)$ comporta

$$\rho_1 = \left(\mu - \frac{1}{2}\sigma^2\right)\Delta t + \sigma\sqrt{\Delta t} Z_1, \dots, \rho_n = \left(\mu - \frac{1}{2}\sigma^2\right)\Delta t + \sigma\sqrt{\Delta t} Z_{N-1},$$

con $Z_1, \dots, Z_{N-1}$ indipendenti. Abbiamo allora[2]

$$\bar{\rho}_n \stackrel{\text{def}}{=} \frac{1}{N}\sum_{n=1}^{N} \rho_n \sim N\left(\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t, \frac{\sigma^2 \Delta t}{N}\right),$$

con

$$\mathbf{E}\left[\bar{\rho}_n\right] = \left(\mu - \frac{1}{2}\sigma^2\right)\Delta t \quad \text{e} \quad \mathbf{D}^2\left[\bar{\rho}_n\right] = \frac{\sigma^2 \Delta t}{N}.$$

Inoltre, posto

$$S_{\rho,N}^2 \stackrel{\text{def}}{=} \frac{1}{N-1}\sum_{n=1}^{N}\left(\rho_n - \bar{\rho}_n\right)^2,$$

abbiamo[3]

$$\frac{(N-1)\, S_{\rho,N}^2}{\sigma^2 \Delta t} \sim \chi_{N-1}^2,$$

---

e

$$\begin{aligned}\mathbf{D}^2[X] &= \mathbf{E}\left[X^2\right] - \mathbf{E}[X]^2 = \mathbf{E}\left[e^{2(\mu+\sigma Z)}\right] - e^{2\mu}\mathbf{E}\left[e^{\sigma Z}\right]^2 \\ &= e^{2\mu}\mathbf{E}\left[e^{2\sigma Z}\right] - e^{2\mu}\mathbf{E}\left[e^{\sigma Z}\right]^2 = e^{2\mu}\left(\mathbf{E}\left[e^{2\sigma Z}\right] - \mathbf{E}\left[e^{\sigma Z}\right]^2\right).\end{aligned}$$

D'altra parte,

$$\begin{aligned}\mathbf{E}\left[e^{\sigma Z}\right] &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{\sigma z} e^{-\frac{1}{2}z^2} dz = \frac{e^{\frac{1}{2}\sigma^2}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\sigma^2} e^{\sigma z - \frac{1}{2}z^2} dz \\ &= \frac{e^{\frac{1}{2}\sigma^2}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\sigma^2 + \sigma z - \frac{1}{2}z^2} dz = \frac{e^{\frac{1}{2}\sigma^2}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{(z-\sigma)^2}{2}} dz \\ &= \frac{e^{\frac{1}{2}\sigma^2}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = e^{\frac{1}{2}\sigma^2}.\end{aligned}$$

Segue chiaramente che

$$\mathbf{E}\left[e^{2\sigma Z}\right] = e^{2\sigma^2} \quad \text{and} \quad \mathbf{E}\left[e^{\sigma Z}\right]^2 = e^{\sigma^2}$$

In definitiva,

$$\mathbf{E}[X] = e^{\mu} e^{\frac{1}{2}\sigma^2} = e^{\mu + \frac{1}{2}\sigma^2}$$

e

$$\begin{aligned}\mathbf{D}^2[X] &= e^{2\mu}\left(e^{2\sigma^2} - e^{\sigma^2}\right) = e^{2\mu} e^{\sigma^2}\left(e^{\sigma^2} - 1\right) \\ &= e^{2\mu + \sigma^2}\left(e^{\sigma^2} - 1\right).\end{aligned}$$

Applicando quanto mostrato alla variabile aleatoria

$$Y = \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t, + \sigma\sqrt{\Delta t} Z\right)$$

si ottengono le equazioni desiderate.

[2] Tenere presente che $X \sim N\left(\mu, \sigma^2\right) \Rightarrow \bar{X}_N \sim N\left(\mu, \sigma^2/N\right)$.

[3] Tenere presente che $X \sim N\left(\mu, \sigma^2\right) \Rightarrow (N-1)\, S_N^2(X)/\sigma^2 \sim \chi_{N-1}^2$.

con[4]

$$\mathbf{E}\left[S_{\rho,N}^{2}\right]=\sigma^{2}\Delta t \quad \text{e} \quad \mathbf{D}^{2}\left[S_{\rho,N}^{2}\right]=\frac{2}{N-1}\sigma^{4}\Delta t^{2}.$$

Pertanto possiamo usare $\bar{\rho}_n$ come stimatore di $\left(\mu-\frac{1}{2}\sigma^{2}\right)\Delta t$ con un errore quadratico pari a $\frac{\sigma^{2}\Delta t}{N}$ e possiamo usare $S_{\rho,N}^{2}$ come stimatore di $\sigma^{2}\Delta t$ con un errore quadratico pari a $\frac{2}{N-1}\sigma^{4}\Delta t^{2}$.

Per quanto riguarda il tasso non rischioso r consideriamo i dati presenti nel sito del U.S. Deartment of the Treasure (see `https://www.treasury.gov/resource-center/data-chart-center/interest-rates/Pages/TextView.aspx?data=billRatesYear&year=2021`). Nella sezione "Data" accediamo alla sotto-sezione "Interest Rates" (see `https://home.treasury.gov/policy-issues/financing-the-government/interest-rate-statistics`). Qui abbiamo vari insiemi di dati che forniscono informazioni sul tasso di rendimento di uno zero coupon bond B con differenti maturità. In particolare la Daily Treasury Yield Curve Rates (see `https://www.treasury.gov/resource-center/data-chart-center/interest-rates/Pages/TextView.aspx?data=yield`). Un'altra rilevante fonte di dati è il sito TreasuryDirect (see `https://www.treasurydirect.gov/GA-FI/FedInvest/selectSecurityPriceDate`) in cui si può trovare una raccolta dei tassi di rendimento con maturità piuttosto dettagliate sotto forma dei prezzi di varie tipologie di bond, disponibili anche in formato csv. Inoltre non si può non menzionare il sito della Federal Reserve Bank of St. Louis (see `https://fred.stlouisfed.org/`).

L'indice Standard & Poor's 500, brevemente S&P 500 (solitamente codificato come SPX), è un indice azionario che comprende le 500 compagnie a maggior capitalizzazione con azioni ordinarie publicamente scambiate al New York Stock Exchange (NYSE). Con *capitalizzazione* si intende il valore della singola azione moltiplicato per il numero di azioni presenti sul mercato, quest'ultimo noto come *flottante*. L'SPX in quanto tale non è scambiabile sul mercato. Tuttavia esiste un fondo azionario scambiato sul NYSE Arca noto come Standard & Poor's Depositary Receipts, brevemente SPDR S&P 500 Trurst (solitamente codificato come SPY), che si pone l'obiettivo di replicare passivamente l'indice S&P 500. Il fondo SPDR S&P 500 Trurst rientra nella categoria degli exchange-traded funds (ETF) che sono fondi a partecipazione azionaria scambiabile sul mercato la cui politica è replicare l'andamento di taluni indici di riferimento. Nel Chicago Board Option Exchange (CBOE) (see `https://www.cboe.com`) è possibile scambiare opzioni di acquisto e vendita con sottostante SPX (see `https://www.cboe.com/tradable_products/sp_500/spx_options/`). Nel National Association of Security Dealers Automated Quotation (NASDAQ) (see `https://www.nasdaq.com`) è possibile scambiare opzioni di acquisto e vendita con sottostante SPY (see `https://www.nasdaq.com/market-activity/funds-and-etfs/spy/option-chain`). Le opzioni su SPX sono opzioni europee e possono essere esercitate solo alla data di scadenza (see `https://www.cboe.com/tradable_products/sp_500/spx_options/specifications/`). Le opzioni su SPY sono opzioni americane e possono essere esercitate a una qualunque data tra quella di acquisto e la scadenza. Altre differenze sono che le opzioni su SPX sono regolate in contanti dal momento che il sottostante non è scambiabile e non pagano dividendi, mentre le opzioni su SPY sono regolate in azioni dal momento che il sottostante è scambiabile e pagano potenzialmente dividendi. Come riferimento relativamente semplice per ulteriori informazioni consigliamo `https://www.thebalance.com/spx-options-vs-spy-options-2536632`.

Computation of the risk free interest rate

$$P=100*\left(1-0.0024*\frac{91}{360}\right)=100*(1-0.00060667)=99.939$$

:

$$P=100*\left(1-d*\frac{91}{360}\right)$$

$$r=\frac{100-100*\left(1-d*\frac{91}{360}\right)}{100*\left(1-d*\frac{91}{360}\right)}=\frac{1-\left(1-d*\frac{91}{360}\right)}{1-d*\frac{91}{360}}=\frac{d*\frac{91}{360}}{1-d*\frac{91}{360}}=\frac{0.0024*\frac{91}{360}}{1-0.0024*\frac{91}{360}}=6.0703\times 10^{-4}$$

[4]Tenere presente che $X\sim N\left(\mu,\sigma^{2}\right)\Rightarrow\mathbf{D}^{2}\left[S_{N}^{2}\left(r\right)\right]=\frac{2\sigma^{4}}{N-1}$.

:

$$r = \frac{100 - 99.939}{99.939} = 0.00061037$$

$$r = \frac{100 - 99.932833}{99.932833} = 0.00067212$$

$$r = \frac{100 - 99.942556}{99.942556} = 0.00057477$$

$$r = \left(\frac{1}{1 - \frac{dn}{360}}\right)^{1/n} = \left(\frac{1}{1 - 0.0024 * \frac{91}{360}}\right)^{1/91}$$

: 1.0

Consider a certain period of time, e.g. one year, and fix a subperiod, e.g. one day, one week, one month,... Write Δt for the lenght of the subperiod in terms of the period, e.g. $\Delta t = 1/252$ for the subperiod of one day (there are approximately 252 trading days in one year), $\Delta t = 1/52$ for the subperiod of one week (there are approximately 52 trading weeks in one year), $\Delta t = 1/12$ for the subperiod of one month,... Let n be the natural number such that $n + 1$ is the number of subperiod constituting the period e.g. $n = 251$ for the subperiod of one day, $n = 51$ for the subperiod of one week , $n = 11$ for the subperiod of one month,... let us label the subperiods on varing of $k = 0, 1, \ldots, n$ and let s_k be the value of the stock at the end of the subperiod. Write

$$x_k \equiv \ln\left(\frac{s_k}{s_{k-1}}\right), \quad \forall k = 1, \ldots, n$$

Assume that the values $x_1, \ldots, x_n$ are the realization of a simple random rample $X_1, \ldots, X_n$ drawn from a random variable

$$X \sim \mathcal{N}\left(\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t, \sigma^2 \Delta t\right)$$

We know that the sample mean

$$\bar{X}_n \equiv \frac{1}{n}\sum_{k=1}^{n} X_k$$

is an unbiased estimator for μ_X and the unbiased sample variance

$$S_n^2 \equiv \frac{1}{n-1}\sum_{k=1}^{n}\left(X_k - \bar{X}_n\right)^2$$

is an unbiased estimator for σ_X^2. Since X is Gaussian distributed we have that

$$\frac{(n-1)\, S_n^2}{\sigma_X^2} \sim \chi_{n-1}^2$$

Hence,

$$\mathbf{D}^2\left[\frac{(n-1)\, S_n^2}{\sigma_X^2}\right] = 2\,(n-1)$$

It follows that

$$\frac{(n-1)^2}{\sigma_X^4}\mathbf{D}^2\left[S_n^2\right] = 2\,(n-1)$$

that is the standard error of S_n^2 is given by

$$\mathbf{D}\left[S_n^2\right] = \sqrt{\frac{2}{n-1}}\sigma_X^2.$$

We set

$$u = e^{\sigma\sqrt{\Delta t}}, \quad d = 1/u$$

2.1.2 Future Prices

We have

$$F_{t,T} = (1 + r_{t,T}) P_t - (\gamma_{t,T} - \kappa_{t,T})$$

where

1. S_t spot price of the asset (commodity, stock, index, currency,...) at time t;
2. $F_{t,T}$ market value (price) of the future on the asset at t with delivery date T;
3. $r_{t,T}$ risk free rate from t to T;
4. $\gamma_{t,T}$ capitalized flow of marginal convenience yield for holding the asset from t to T;
5. $\kappa_{t,T}$ per unit cost of physical storage from t to T.

Assume we have two futures at t with delivery dates T_1 and T_2, where $T_2 > T_1$. We can write

$$F_{t,T_j} = \left(1 + r_{t,T_j}\right) S_t - \left(\gamma_{t,T_j} - \kappa_{t,T_j}\right), \quad j = 1, 2.$$

Assume that

$$\kappa_{t,T_j} \equiv \kappa \left(T_j - t\right), \quad j = 1, 2,$$

for some constant $\kappa > 0$. Assume also that

$$r_{t,T_j} \equiv e^{\rho(T_j - t)} - 1, \quad j = 1, 2,$$

where

$$\rho \equiv \log\left(1 + r_a\right)$$

is the continuous time annual risk free rate equivalent to a constant annual risk free rate r_a

$$\left(1 + r_a\right)^t = e^{\rho t},$$

where t is expressed in years. In the end, assume that

$$\gamma_{t,T_j} \equiv \gamma \left(T_j - t\right), \quad j = 1, 2,$$

for some constant $\gamma > 0$. We can write

$$F_{t,T_j} = e^{\rho(T_j - t)} S_t - (\gamma - \kappa)\left(T_j - t\right), \quad j = 1, 2.$$

Therefore,

$$\begin{aligned} F_{t,T_2} - F_{t,T_1} &= \left(e^{\rho(T_2 - t)} - e^{\rho(T_1 - t)}\right) S_t - (\gamma - \kappa)\left(T_2 - T_1\right) \\ &= \left(1 - e^{-\rho(T_2 - T_1)}\right) e^{\rho(T_2 - t)} S_t - (\gamma - \kappa)\left(T_2 - T_1\right) \end{aligned}$$

Setting

$$F_{t,T_2} - F_{t,T_1} \equiv G_t,$$

we have

$$\begin{aligned} G_{t+1} - G_t &= e^{\rho(T_2 - (t+1))} \left(1 - e^{-\rho(T_2 - T_1)}\right) S_{t+1} - e^{\rho(T_2 - t)} \left(1 - e^{-\rho(T_2 - T_1)}\right) S_t \\ &= \left(1 - e^{-\rho(T_2 - T_1)}\right) \left(e^{\rho(T_2 - (t+1))} S_{t+1} - e^{\rho(T_2 - t)} S_t\right) \\ &= \left(1 - e^{-\rho(T_2 - T_1)}\right) e^{\rho(T_2 - t)} \left(e^{-\rho} S_{t+1} - S_t\right). \end{aligned}$$

As a consequence,
$$\frac{G_{t+1}-G_t}{S_{t+1}-S_t} = \left(1-e^{-\rho(T_2-T_1)}\right) e^{\rho(T_2-t)} \frac{e^{-\rho}S_{t+1}-S_t}{S_{t+1}-S_t}.$$
Assuming that
$$\frac{G_{t+1}-G_t}{S_{t+1}-S_t} > 0,$$
passing to the logarithms, we obtain
$$\log\left(\frac{G_{t+1}-G_t}{S_{t+1}-S_t}\right) = \log\left(1-e^{-\rho(T_2-T_1)}\right) + \rho\left(T_2-t\right) + \log\left(\frac{e^{-\rho}S_{t+1}-S_t}{S_{t+1}-S_t}\right)$$
and assuming that $\rho \ll 1$,
$$\log\left(\frac{e^{-\rho}S_{t+1}-S_t}{S_{t+1}-S_t}\right) \approx 0,$$
it follows
$$\log\left(\frac{G_{t+1}-G_t}{S_{t+1}-S_t}\right) \approx \log\left(1-e^{-\rho(T_2-T_1)}\right) + \rho\left(T_2-t\right).$$
In the end, we can hope to estimate ρ by regressing $\log\left(\left(G_{t+1}-G_t\right)/\left(S_{t+1}-S_t\right)\right)$ on T_2-t.

2.2 Portafogli di Titoli (Markowitz Model)

Consideriamo ancora un mercato uniperiodale in cui al tempo $t=0$ sia possibile investire in un titolo non rischioso, denominato *obbligazione* (*bond*), e in un set finito di titoli rischiosi, denominato *portafoglio azionario* (*stock portfolio*). Al tempo $t=T$, raggiunta la maturità (*maturity*), l'investimento viene liquidato, l'eventuale ricchezza prodotta viene consumata o l'eventuale debito accumulato va ripianato. Come nella Sezione 2.1, assumiamo che gli investimenti possano essere effettuati senza limitazioni quantitative e costi di transazione, che si possa prendere a prestito e depositare denaro allo stesso tasso d'interesse e che siano possibili libere vendite allo scoperto. Per omogenità di notazione, che si rivelerà conveniente in seguito, denotiamo con $S_0(0)$ [risp. $S_0(T)$] il valore di mercato di una obbligazione al tempo $t=0$ [risp. $t=T$] e denotiamo con $S_1(0),\dots,S_M(0)$ [risp. $S_1(T),\dots,S_M(T)$], con $M \geqslant 1$, i valori dei singoli titoli azionari in portafoglio al tempo $t=0$ [risp. $t=T$]. Indichiamo con r_0 il tasso di rendimento dell'investimento non rischioso alla maturità. Chiaramente vale
$$S_0(T) = (1+r_0)\,S_0(0). \tag{2.75}$$
Come nel modello binomiale, $S_1(0),\dots,S_M(0)$ [risp. $S_1(T),\dots,S_M(T)$], valori dei titoli azionari in portafoglio al tempo $t=0$ [risp. $t=T$], sono da considerarsi numeri reali [resp. variabili aleatorie] al tempo $t=0$. Infatti, assumiano che al tempo $t=0$ i valori $S_1(0),\dots,S_M(0)$ siano osservabili con certezza, ma, chiaramente, i valori $S_1(T),\dots,S_M(T)$ non sono ancora osservabili con certezza. Tuttavia, a differenza del modello binomiale non modellizziamo esplicitamente la distribuzione delle variabili aleatorie $S_1(T),\dots,S_M(T)$, salvo assumere che abbiano momento finito di ordine 2. Nondimeno, nell'adattamento del modello a dati reali di mercato sarà conveniente assumere che le variabili aleatorie $\left(S_1(T)-S_1(0)\right)/S_1(0),\dots,\left(S_M(T)-S_M(0)\right)/S_M(0)$ siano congiuntamente Gaussianamente distribuite.

Consideriamo inizialmente un investitore che al tempo $t=0$ componga il proprio portafoglio allocandovi i soli titoli rischiosi e investendo un budget $W>0$, ma scegliendo sui singoli titoli sia allocazioni positive che negative. Le prime, note come *posizioni lunghe* (*long position*), corrispondono all'acquisto di diverse unità di titoli, cosiddetti *pacchetti azionari* (*shares of stock*). Le seconde, note come *posizioni corte (short position)*, corrispondono a vendite allo scoperto di pacchetti azionari. Non escludiamo inoltre che l'investitore possa prendere una posizione nulla su alcuni titoli, semplicemente non inserendoli nel suo portafoglio. Un tale portafoglio è allora convenientemente rappresentato da una M-pla di numeri reali, $\pi(0) \equiv (y_1(0), \ldots, y_M(0)) \in \mathbb{R}^M$, tale che $y_m(0)$ rappresenti il pacchetto azionario m-esimo allocato in portafoglio al tempo $t = 0$, per $m = 1, \ldots, M$. Chiaramente un tale portafoglio viene composto alla luce dei noti valori di mercato $S_1(0), \ldots, S_M(0)$ dei singoli titoli. Al tempo $t = T$, tuttavia, i valori dei singoli titoli saranno rappresentati dalla realizzazioni delle variabili aleatorie $S_1(T), \ldots, S_M(T)$, non note al tempo $t = 0$. Ciò determinerà un variazione aleatoria del valore del portafoglio che potrà dare luogo a un ricchezza maggiore o minore del budget inizialmente investito. Comunque sia, in questo modello monoperiodale al tempo $t = T$ l'investitore liquida l'investimento, pertanto non è prevista una riconfigurazione del portafoglio, alla luce dell'osservazione delle realizzazioni delle variabili aleatorie $S_1(T), \ldots, S_M(T)$, se non quella banale $y_1(T) = \cdots = y_M(T) = 0$. Pertanto, in un modello monoperiodale possiamo trascurare la dipendenza temporale della scelta di portafoglio e semplificare le notazioni ponendo $y_1(0) \equiv y_1, \ldots, y_M(0) \equiv y_M$ e $\pi(0) \equiv \pi$. Si tenga però presente che in un modello multiperiodale le componenti del portafoglio saranno dipendenti dal tempo e saranno da considerarsi, per un tempo $t > 0$ riferito al tempo $t = 0$, quali variabili aleatorie $y_1(t) = \cdots = y_M(t)$, in quanto le successive riconfigurazioni del portafoglio da parte dell'investitore dipenderanno dai valori aleatori $S_1(t), \ldots, S_M(t)$ che assumeranno i titoli in portafoglio.

Definizione 61 *Chiamiamo* valore di mercato *(*market value*) o* prezzo (price) *del pacchetto azionario m-esimo in portafoglio al tempo $t = 0$ [risp. $t = T$] il numero [risp. la variabile aleatoria] reale*

$$W_m(y_m, 0) \overset{def}{=} y_m S_m(0) \quad \textit{[resp. } W_m(y_m, T) \overset{def}{=} y_m S_m(T) \textit{]}, \quad \forall m = 1, \ldots, M. \tag{2.76}$$

Da notare che è in effetti

$$W_m(y_m, 0) \equiv W_m(y_m, S_m(0)) \quad [\text{resp. } W_m(y_m, T) \equiv W_m(y_m, S_m(T))], \quad \forall m = 1, \ldots, M.$$

Tuttavia, mentre $S_m(0)$ e $S_m(T)$ possono essere ritenute variabili esogene del modello, per ogni $m = 1, \ldots, M$, il cui valore non dipende dalle scelte dell'investitore, nella misura in cui queste scelte non siano da sole in grado di modificare i valori di mercato dei titoli, le variabili $y_1, \ldots, y_M$ sono variabili endogene, il cui valore viene proprio scelto dall'investitore secondo le indicazioni del modello. La notazione introdotta nell'Equazione (2.76) si propone di sottolineare questi aspetti.

Osservazione 62 *Si ha $y_m = 0$ se e solo se $W_m(y_m, 0) = 0$, per ogni $m = 1, \ldots, M$.*

Osservazione 63 *Per ogni $m \in \{1, \ldots, M\}$ tale che $W_m(y_m, 0) = 0$ si ha anche $W_m(y_m, T) = 0$.*

Definizione 64 *Chiamiamo* valore di mercato *o* prezzo *del portafoglio al tempo $t = 0$ [risp. $t = T$] il numero [risp. la variabile aleatoria] reale*

$$W(y_1, \ldots, y_M, 0) \overset{def}{=} \sum_{m=1}^{M} W_m(y_m, 0) \quad \textit{[resp. } W(y_1, \ldots, y_M, T) \overset{def}{=} \sum_{m=1}^{M} W_m(y_m, T)\textit{]}$$

Osservazione 65 *L'investimento di un budget $W > 0$ nella composizione del portafoglio comporta che*

$$W(y_1, \ldots, y_M, 0) = W > 0.$$

Notare che ciò esclude che si possa avere $W_m(y_m, 0) = 0$ per ogni $m = 1, \ldots, M$.

Osservazione 66 *Si ha*

$$W(y_1, \ldots, y_M, t) = \sum_{m=1}^{M} y_m S_m(t),$$

per $t = 0, T$.

Definizione 67 *Chiamiamo* variazione del valore *(*value change*) o* rendimento *(*return*) del pacchetto azionario m-esimo in portafoglio nel periodo* $[0,T]$ *la variabile aleatoria reale*

$$W_m(y_m,T) - W_m(y_m,0) = y_m(S_m(T) - S_m(0)), \quad \forall m = 1,\dots,M. \tag{2.77}$$

Chiamiamo variazione del valore *o* rendimento *del portafoglio nel periodo* $[0,T]$ *la variabile aleatoria reale*

$$W(y_1,\dots,y_M,T) - W(y_1,\dots,y_M,0) = \sum_{m=1}^{M} W_m(y_m,T) - \sum_{m=1}^{M} W_m(y_m,0). \tag{2.78}$$

Osservazione 68 *Si ha*

$$W(y_1,\dots,y_M,T) - W(y_1,\dots,y_M,0) = \sum_{m=1}^{M} y_m(S_m(T) - S_m(0)).$$

Definizione 69 *Chiamiamo* peso *(*weight*) del titolo azionario m-esimo in portafoglio al tempo* $t = 0$ *il numero reale*

$$w_m \stackrel{def}{=} \frac{W_m(y_m,0)}{W(y_1,\dots,y_M,0)}, \quad \forall m = 1,\dots,M. \tag{2.79}$$

Osservazione 70 *Chiaramente,*

$$w_m = \frac{W_m(y_m,0)}{W} = \frac{y_m S_m(0)}{W}, \tag{2.80}$$

per ogni $m = 1,\dots,M$*. In consequenza,*

$$y_m = W\frac{w_m}{S_m(0)}, \tag{2.81}$$

per ogni $m = 1,\dots,M$*.*

Stanti le Equazioni (2.80) e (2.81), la nozione di peso di un titolo azionario in portafoglio consente di rappresentare un portafoglio, in termini della M-pla dei pesi dei titoli azionari in esso presenti, in modo perfettamente equivalente alla rappresentazione in termini dei pacchetti di titoli. La rappresentazione in termini di pesi si rivelerà peraltro cruciale nella formulazione del modello di Markowitz. Pertanto, d'ora in avanti, considereremo l'identificazione di un portafoglio di titoli con la M-pla $\pi \equiv (w_1,\dots,w_M)$, rappresentando w_m il peso del titolo azionario m-esimo in portafoglio, per ogni $m = 1,\dots,M$.

Osservazione 71 *Si ha*

$$\sum_{m=1}^{M} w_m = 1. \tag{2.82}$$

Definizione 72 *Chiamiamo* insieme dei portafogli fattibili *(*set of all feasible portfolios*) l'insieme*

$$\mathbb{H} \equiv \left\{(w_1,\dots,w_M) \in \mathbb{R}^M : \sum_{m=1}^{M} w_m = 1\right\}. \tag{2.83}$$

Proposizione 73 *L'insieme* $\mathbb{H}$ *è un iperpiano di* $\mathbb{R}^M$ *passante per i versori degli assi* $(1,0,\dots,0),\dots,$ $(0,0,\dots,1)$*. Tale iperpiano si caratterizza anche come l'iperpiano ortogonale al vettore* $(1,\dots,1)$ *e passante per il punto* $(1/M,\dots,1/M)$*.*

Proof. *Basta ricordare che l'equazione di un iperpiano ortogonale al vettore* $(v_1, \ldots, v_M) \in \mathbb{R}^M$ *e passante per il punto* $\left(x_1^0, \ldots, x_M^0\right) \in \mathbb{R}^M$ *si scrive*

$$\sum_{m=1}^{M} v_m \left(x_m - x_m^0\right) = 0$$

e osservare che l'equazione

$$\sum_{m=1}^{M} w_m = 1$$

può essere equivalentemente riscritta come

$$\sum_{m=1}^{M} \left(w_m - \frac{1}{M}\right) = 0.$$

□

Definizione 74 *Chiamiamo* tasso di rendimento *(*rate of return*) del titolo azionario m-esimo in portafoglio al tempo* $t = T$ *la variabile aleatoria reale*

$$r_m(T) \stackrel{def}{=} \frac{S_m(T) - S_m(0)}{S_m(0)}, \quad \forall m = 1, \ldots, M.$$

Notare che, per come definite, le variabili tasso di rendimento dei titoli azionari in portafoglio hanno momento finito di ordine 2. Inoltre, come già menzionato, nell'adattamento del modello ai dati reali assumeremo che $r_1(T), \ldots, r_M(T)$ siano congiuntamemente Gaussianamente distribuite.

Osservazione 75 *Chiaramente,*

$$r_m(T) = \frac{W_m(y_m, T) - W_m(y_m, 0)}{W_m(y_m, 0)},$$

per ogni $m = 1, \ldots, M$ *tale che* $W_m(y_m, 0) \neq 0$.

Definizione 76 *Chiamiamo* tasso di rendimento atteso *(*expected rate of return*) o* tasso di rendimento medio *(*mean rate of return*) del titolo azionario m-esimo in portafoglio al tempo* $t = T$ *il numero reale*

$$\bar{r}_m(T) \stackrel{def}{=} \mathbf{E}[r_m(T)], \quad \forall m = 1, \ldots, M.$$

Osservazione 77 *Si ha*

$$\bar{r}_m(T) = \frac{\mathbf{E}[S_m(T)] - S_m(0)}{S_m(0)},$$

per ogni $m = 1, \ldots, M$.

Definizione 78 *Chiamiamo* tasso di rendimento *del portafoglio al tempo* $t = T$ *la variabile aleatoria reale*

$$r(y_1, \ldots, y_M, T) \stackrel{def}{=} \frac{W(y_1, \ldots, y_M, T) - W(y_1, \ldots, y_M, 0)}{W(y_1, \ldots, y_M, 0)}. \tag{2.84}$$

Osservazione 79 *Chiaramente*

$$r(y_1, \ldots, y_M, T) = \frac{W(y_1, \ldots, y_M, T) - W}{W}.$$

Simmetricamente al caso dei pacchetti e dei pesi dei titoli azionari in portafoglio, la specifica del tempo $t = T$ nelle definizioni e notazione dei tassi di rendimento e dei tassi di rendimento atteso è anche essa ridondante, in quanto in un modello uniperiodale tale variabile non può essere definita se non in riferimento al tempo $t = T$. In questo caso la specifica è motivata per sottolineare che, per un tempo $t > 0$ riferito al tempo $t = 0$, i tassi di rendimento sono da considerarsi variabili aleatorie. Comunque, per semplificare le notazioni, questa specifica verrà eliminata a partire dalla Sottosezione 2.2.1.

Proposizione 80 *Il tasso di rendimento del portafoglio al tempo $t = T$ è la somma pesata dei tassi di rendimento dei singoli titoli azionari. Formalmente,*

$$r(y_1, \ldots, y_M, T) = \sum_{m=1}^{M} w_m r_m(T). \tag{2.85}$$

Proof. *Infatti, combinando la (2.78) con la (2.84), abbiamo*

$$\begin{aligned}
& r(y_1, \ldots, y_M, T) \\
&= \frac{1}{W(y_1, \ldots, y_M, 0)} \left(\sum_{m=1}^{M} W_m(y_m, T) - \sum_{m=1}^{M} W_m(y_m, 0) \right) \\
&= \frac{1}{W} \sum_{m=1}^{M} (W_m(y_m, T) - W_m(y_m, 0)) \\
&= \frac{1}{W} \sum_{\substack{m=1 \\ y_m \neq 0}}^{M} (W_m(y_m, T) - W_m(y_m, 0)) \\
&= \frac{1}{W} \sum_{\substack{m=1 \\ y_m \neq 0}}^{M} W_m(y_m, 0) \left(\frac{W_m(y_m, T) - W_m(y_m, 0)}{W_m(y_m, 0)} \right) \\
&= \sum_{\substack{m=1 \\ y_m \neq 0}}^{M} \frac{W_m(y_m, 0)}{W} \left(\frac{W_m(y_m, T) - W_m(y_m, 0)}{W_m(y_m, 0)} \right) \\
&= \sum_{\substack{m=1 \\ w_m \neq 0}}^{M} w_m r_m(T) \\
&= \sum_{m=1}^{M} w_m r_m(T).
\end{aligned}$$

□

Osservazione 81 *Stante l'Equazione (2.85) possiamo chiaramente scrivere*

$$r(y_1, \ldots, y_M, T) \equiv r(w_1, \ldots, w_M, T)$$

Definizione 82 *Chiamiamo* tasso di rendimento atteso *o* tasso di rendimento medio *del portafoglio al tempo $t = T$ il numero reale*

$$\bar{r}(w_1, \ldots, w_M, T) \stackrel{def}{=} \mathbf{E}[r(w_1, \ldots, w_M, T)].$$

Definizione 83 *Chiamiamo* varianza *del titolo azionario m-esimo in portafoglio la varianza del suo tasso di rendimento*

$$\sigma_m^2 \stackrel{def}{=} \mathbf{D}^2[r_m(T)], \quad \forall m = 1, \ldots, M.$$

Definizione 84 *Chiamiamo* volatilità *del titolo azionario m-esimo in portafoglio la deviazione standard del suo tasso di rendimento*

$$\sigma_m \overset{def}{=} \mathbf{D}[r_m(T)], \quad \forall m = 1, \ldots, M.$$

Definizione 85 *Chiamiamo* covarianza *dei titoli azionari ℓ-esimo ed m-esimo in portafoglio la covarianza dei loro tassi di rendimento*

$$\sigma_{\ell,m} \overset{def}{=} \mathbf{E}[(r_\ell(T) - \bar{r}_\ell(T))(r_m(T) - \bar{r}_m(T))], \quad \forall \ell, m = 1, \ldots, M.$$

Notare che tassi di rendimento atteso, varianze, volatilità e covarianze dei titoli in portafoglio vergono formalmente calcolati al tempo $t = 0$ come momenti delle variabili aleatorie $r_1(T), \ldots, r_M(T)$. Ma in effetti, come vedremo in seguito, vengono stimati come momenti campionari dei rendimenti desunti dai valori storici dei prezzi dei titoli $S_1, \ldots, S_M$, sotto opportune condizioni di ergodicità degli stessi.

Proposizione 86 *Il tasso di rendimento atteso del portafoglio è la somma pesata dei tassi di rendimento attesi dei singoli titoli azionari. Formalmente,*

$$\bar{r}(w_1, \ldots, w_M, T) = \sum_{m=1}^{M} w_m \bar{r}_m(T). \tag{2.86}$$

Proof. *Per la linearità dell'operatore speranza,*

$$\mathbf{E}[r(w_1, \ldots, w_M, T)] = \mathbf{E}\left[\sum_{m=1}^{M} w_m r_m(T)\right] = \sum_{m=1}^{M} w_m \mathbf{E}[r_m(T)] = \sum_{m=1}^{M} w_m \bar{r}_m(T).$$

□

Proposizione 87 *Si ha*

$$\sigma_{\ell,m} = \mathbf{E}[r_\ell(T) r_m(T)] - \bar{r}_\ell(T)\bar{r}_m(T)$$

e

$$|\sigma_{\ell,m}| \leqslant \sigma_\ell \sigma_m,$$

per tutti gli $\ell, m = 1, \ldots, M$.
Proof. *Ci limitiamo ad osservare che dalla disuguaglianza di Cauchy-Schwarz otteniamo*

$$\begin{aligned} |\sigma_{\ell,m}| &= |\mathbf{E}[(r_\ell(T) - \bar{r}_\ell(T))(r_m(T) - \bar{r}_m(T))]| \\ &\leqslant \mathbf{E}[(r_\ell(T) - \bar{r}_\ell(T))^2]^{1/2} \mathbf{E}[(r_m(T) - \bar{r}_m(T))^2]^{1/2} \\ &= \sigma_\ell \sigma_m, \end{aligned}$$

per tutti gli $\ell, m = 1, \ldots, M$. □

Definizione 88 *Chiamiamo* correlazione *dei titoli azionari ℓ-esimo ed m-esimo in portafoglio le correlazioni dei loro tassi di rendimento*

$$\rho_{\ell,m} \overset{def}{=} \frac{\sigma_{\ell,m}}{\sigma_\ell \sigma_m}, \quad \forall \ell, m = 1, \ldots, M.$$

Osservazione 89 *Risulta*

$$|\rho_{\ell,m}| \leqslant 1,$$

per tutti gli $\ell, m = 1, \ldots, M$.

Diciamo che i titoli azionari ℓ-esimo ed m-esimo in portafoglio sono *correlati* [risp. *anticorrelati*] se $\rho_{\ell,m} > 0$ [risp. $\rho_{\ell,m} < 0$]. Nei casi limite in cui $\rho_{\ell,m} = 1$ [risp. $\rho_{\ell,m} = -1$] o $\rho_{\ell,m} = 0$ diciamo che i titoli azionari sono *perfettamente correlati* [resp. *perfettamente anticorrelati*] o *scorrelati*.

Definizione 90 *Chiamiamo* matrice di varianza-covarianza *dei titoli azionari in portafoglio la matrice quadrata di ordine M*

$$\Sigma \equiv \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} & \cdots & \sigma_{1,M} \\ \sigma_{1,2} & \sigma_2^2 & \cdots & \sigma_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1,M} & \sigma_{2,M} & \cdots & \sigma_M^2 \end{pmatrix}.$$

Osservazione 91 *La matrice Σ è simmetrica.*

Osservazione 92 *Si ha*

$$\Sigma \equiv \begin{pmatrix} \sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 & \cdots & \rho_{1,M}\sigma_1\sigma_M \\ \rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2 & \cdots & \rho_{2,M}\sigma_2\sigma_M \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1,M}\sigma_1\sigma_M & \rho_{2,M}\sigma_2\sigma_M & \cdots & \sigma_M^2 \end{pmatrix}$$

Osservazione 93 *Si ha*

$$(w_1, \ldots, w_M)\, \Sigma\, (w_1, \ldots, w_M)^{\intercal} = \sum_{m=1}^{M} w_m^2 \sigma_m^2 + 2 \sum_{\substack{\ell,m=1 \\ \ell<m}}^{M} w_\ell w_m \sigma_{\ell,m}, \tag{2.87}$$

per ogni $(w_1, \ldots, w_M) \in \mathbb{R}^M$.

Osservazione 94 *In termini delle correlazioni dei titoli azionari in portafoglio, si ha anche*

$$(w_1, \ldots, w_M)\, \Sigma\, (w_1, \ldots, w_M)^{\intercal} = \sum_{m=1}^{M} w_m^2 \sigma_m^2 + 2 \sum_{\substack{\ell,m=1 \\ \ell<m}}^{M} w_\ell w_m \rho_{\ell,m} \sigma_\ell \sigma_\ell, \tag{2.88}$$

per ogni $(w_1, \ldots, w_M) \in \mathbb{R}^M$.

Definizione 95 *Diciamo che una matrice simmetrica Q di ordine M è* semidefinita positiva *se*

$$(w_1, \ldots, w_M)\, Q\, (w_1, \ldots, w_M)^{\intercal} \geqslant 0$$

per ogni $(w_1, \ldots, w_M) \in \mathbb{R}^M$.

Diciamo che Q è definita positiva *se*

$$(w_1, \ldots, w_M)\, Q\, (w_1, \ldots, w_M)^{\intercal} > 0$$

per ogni $(w_1, \ldots, w_M) \in \mathbb{R}^M - \{(0, \ldots, 0)\}$.

Proposizione 96 *Una matrice simmetrica Q è semidefinita [resp. definita] positiva se e solo se tutti i suoi autovalori sono positivi [resp. strettamente positivi].*

Proof. *E' ben noto (cfr. teorema spettrale) che la matrice simmetrica Q ammette una base ortonormale di autovettori, sia essa $F \equiv \{f_1, \ldots . f_M\}$ e siano $\lambda_1, \ldots, \lambda_M$ gli autovalori corrispondenti agli autovettori $f_1, \ldots . f_M$. Risulta allora*

$$M_F^E(I)\, Q M_I^F(I) = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_M \end{pmatrix} \equiv \Lambda,$$

dove $M_F^E(I)$ *[resp.* $M_I^F(I)$*] è la matrice del cambiamento di base dalla base canonica* $E \equiv \{e_1, \ldots . e_M\}$ *alla base* F *[resp. dalla base* F *alla base canonica* F*]. Si ha inoltre*

$$M_I^F(I) = M_F^E(I)^{-1} = M_I^F(I)^\intercal$$

In conseguenza,

$$\begin{aligned}(w_1,\ldots,w_M)\,Q\,(w_1,\ldots,w_M)^\intercal &= (w_1,\ldots,w_M)\,M_I^F(I)^\intercal\,\Lambda M_I^F(I)\,(w_1,\ldots,w_M)^\intercal \\ &= (\tilde{w}_1,\ldots,\tilde{w}_M)\,\Lambda\,(\tilde{w}_1,\ldots,\tilde{w}_M)^\intercal \\ &= \textstyle\sum_{m=1}^M \lambda_m \tilde{w}_m^2,\end{aligned}$$

dove $(\tilde{w}_1,\ldots,\tilde{w}_M) \equiv (w_1,\ldots,w_M)\,M_I^F(I)^\intercal$*. È allora immediato rendersi conto che* Q *è semidefinita positiva se e solo se tutti i suoi autovalori sono positivi. Per di più, avendosi*

$$(\tilde{w}_1,\ldots,\tilde{w}_M) = (0,\ldots,0) \Leftrightarrow (w_1,\ldots,w_M) = (0,\ldots,0)\,,$$

Q *è definita positiva se e solo se tutti i suoi autovalori sono strettamente positivi.* □

Proposizione 97 *La matrice* Σ *è semidefinita positiva. Inoltre* Σ *è definita positiva purchè il tasso di rendimento di nessuno dei titoli azionari in portafoglio non sia combinazione affine dei restanti, ossia purchè non esista* $m_0 \in \{1,\ldots,M\}$ *tale che*

$$r_{m_0}(T) \overset{a.s.}{=} \alpha + \sum_{\substack{m=1\\ m\neq m_0}}^M \beta_m r_m(T),$$

per opportuni $\alpha, \beta_m \in \mathbb{R}$*, con* $m = 1,\ldots,M$*,* $m \neq m_0$*.*
Proof. *Per la Osservazione (93) e per le proprietà dell'operatore speranza, possiamo scrivere*

$$\begin{aligned}&(w_1,\ldots,w_M)\,\Sigma\,(w_1,\ldots,w_M)^\intercal \\ &= \sum_{m=1}^M w_m^2\sigma_m^2 + 2\sum_{\substack{\ell,m=1\\ \ell<m}}^M w_\ell w_m \sigma_{\ell,m} \\ &= \sum_{m=1}^M w_m^2 \mathbf{E}\left[(r_m(T) - \mathbf{E}[r_m(T)])^2\right] + 2\sum_{\substack{\ell,m=1\\ \ell<m}}^M w_\ell w_m \mathbf{E}\left[(r_\ell(T) - \mathbf{E}[r_\ell(T)])\,(r_m(T) - \mathbf{E}[r_m(T)])\right] \\ &= \mathbf{E}\left[\sum_{m=1}^M w_m^2\,(r_m(T) - \mathbf{E}[r_m(T)])^2 + 2\sum_{\substack{\ell,m=1\\ \ell<m}}^M w_\ell w_m\,(r_\ell(T) - \mathbf{E}[r_\ell(T)])\,(r_m(T) - \mathbf{E}[r_m(T)])\right] \\ &= \mathbf{E}\left[\left(\sum_{m=1}^M w_m\,(r_m(T) - \mathbf{E}[r_m(T)])\right)^2\right] \\ &= \mathbf{E}\left[\left(\sum_{m=1}^M w_m r_m(T) - \sum_{m=1}^M w_m \mathbf{E}[r_m(T)]\right)^2\right] \\ &= \mathbf{E}\left[\left(\sum_{m=1}^M w_m r_m(T) - \mathbf{E}\left[\sum_{m=1}^M w_m r_m(T)\right]\right)^2\right] \\ &= \mathbf{D}^2\left[\sum_{m=1}^M w_m r_m(T)\right] \\ &= \mathbf{D}^2\left[r(w_1,\ldots,w_M,T)\right]\end{aligned}$$

per ogni $(w_1, \ldots, w_M) \in \mathbb{R}^M$*. Ciò comporta che la matrice* Σ *sia semidefinita positiva. Inoltre perchè* Σ *non sia definita positiva dovrebbe esistere* $(w_1^*, \ldots, w_M^*) \in \mathbb{R}^M - \{(0, \ldots, 0)\}$ *tale che*

$$\mathbf{D}^2 \left[r(w_1^*, \ldots, w_M^*, T) \right] = \mathbf{D}^2 \left[\sum_{m=1}^{M} w_m^* r_m(T) \right] = 0.$$

Ciò sarebbe possibile solo se la variabile aleatoria $\sum_{m=1}^{M} w_m^* r_m(T)$ *fosse una Dirac concentrata in qualche* $\alpha \in \mathbb{R}$*. Se così fosse, potremmo scrivere*

$$\sum_{m=1}^{M} w_m^* r_m(T) \stackrel{a.s.}{=} \alpha$$

ed essendo $(w_1^*, \ldots, w_M^*) \neq (0, \ldots, 0)$ *potremmo trovare* $m_0 \in \{1, \ldots, M\}$ *tale che* $w_{m_0}^* \neq 0$*. Posto allora* $\beta_m = -w_m^*/w_{m_0}^*$ *si avrebbe*

$$r_{m_0}(T) \stackrel{a.s.}{=} \alpha + \sum_{\substack{m=1 \\ m \neq m_0}}^{M} \beta_m r_m(T),$$

come asserito. $\square$

Definizione 98 *Diciamo che l'insieme dei titoli azionari in portafoglio è* non singolare *[resp.* singolare*] se la loro matrice di varianza-covarianza è [risp. non è] definita positiva.*

Definizione 99 *Diciamo che un sottoinsieme* $\mathbb{S} \subseteq \mathbb{R}^M$ *è convesso se*

$$\alpha (x_1, \ldots, x_M) + (1 - \alpha) (y_1, \ldots, y_M) \in \mathbb{S}$$

per ogni $(x_1, \ldots, x_M), (y_1, \ldots, y_M) \in \mathbb{S}$ *ed ogni* $\alpha \in [0, 1]$.

Osservazione 100 *Ogni varietà lineare di* $\mathbb{R}^M$ *è convessa. In particolare, l'insieme dei portafogli fattibili* $\mathbb{H}$ *(cfr Definition* **??***) è un convesso di* $\mathbb{R}^M$*.*

Osservazione 101 *L'intersezione di un qualsiasi numero finito di sottoinsiemi convessi è convessa.*

Definizione 102 *Diciamo che una funzione* $f : \mathbb{R}^M \to \mathbb{R}$ *è* convessa *se*

$$f(\alpha (x_1, \ldots, x_M) + (1 - \alpha) (y_1, \ldots, y_M)) \leqslant \alpha f(x_1, \ldots, x_M) + (1 - \alpha) f(y_1, \ldots, y_M)$$

per ogni $(x_1, \ldots, x_M), (y_1, \ldots, y_M) \in \mathbb{R}^M$ *ed ogni* $\alpha \in [0, 1]$*. Diciamo che* $f : \mathbb{R}^M \to \mathbb{R}$ *è* strettamente convessa *se*

$$f(\alpha (x_1, \ldots, x_M) + (1 - \alpha) (y_1, \ldots, y_M)) < \alpha f(x_1, \ldots, x_M) + (1 - \alpha) f(y_1, \ldots, y_M)$$

per ogni $(x_1, \ldots, x_M), (y_1, \ldots, y_M) \in \mathbb{R}^M$ *tali che* $(x_1, \ldots, x_M) \neq (y_1, \ldots, y_M)$ *ed ogni* $\alpha \in (0, 1)$.

Teorema 103 *Sia* $\mathbb{S} \subseteq \mathbb{R}^M$ *convesso e sia* $f : \mathbb{R}^M \to \mathbb{R}$ *convessa allora o il problema di minimizzazione vincolata*

$$\min_{(x_1, \ldots, x_M) \in \mathbb{S}} \{ f(x_1, \ldots, x_M) \}$$

non ha soluzione o una soluzione locale è anche globale. Inoltre se $f : \mathbb{R}^M \to \mathbb{R}$ *è strettamente convessa allora esiste un'unica soluzione globale.*

Proof. *L'enunciato è un teorema fondamentale di analisi convessa.* $\square$

Teorema 104 *La forma quadratica $\sigma^2 : \mathbb{R}^M \to \mathbb{R}$ definita ponendo*

$$\sigma^2(w_1, \dots, w_M) \overset{def}{=} (w_1, \dots, w_M)\,\Sigma\,(w_1, \dots, w_M)^\intercal, \quad \forall (w_1, \dots, w_M) \in \mathbb{R}^M \tag{2.89}$$

è convessa. Inoltre $\sigma^2 : \mathbb{R}^M \to \mathbb{R}$ è strettamente convessa se e solo se l'insieme dei titoli azionari in portafoglio è non singolare.
Proof. *L'enunciato è una semplice riformulazione di una ben nota proprietà delle forme quadratiche semidefinite positive e definite positive (cfr. Proposizione 97.* □

Definizione 105 *Chiamiamo* rischio *del portafoglio, e lo denotiamo con σ^2, la varianza del suo tasso di rendimento. In simboli,*

$$\sigma^2(w_1, \dots, w_M) \overset{def}{=} \mathbf{D}^2[r(w_1, \dots, w_M, T)].$$

Osservazione 106 *Si ha*

$$\sigma^2(w_1, \dots, w_M) = \sum_{\ell,m=1}^{M} w_\ell w_m \sigma_{\ell,m} = \sum_{m=1}^{M} w_m^2 \sigma_m^2 + 2 \sum_{\substack{\ell,m=1 \\ \ell<m}}^{M} w_\ell w_m \rho_{\ell,m} \sigma_\ell \sigma_m \tag{2.90}$$

per ogni $(w_1, \dots, w_M) \in \mathbb{H}$, dove $\mathbb{H}$ è l'insieme dei portafogli fattibili.
Proof. *Dalla prova della Proposizione (**??**) abbiamo che*

$$\sigma^2(w_1, \dots, w_M) = (w_1, \dots, w_M)\,\Sigma\,(w_1, \dots, w_M)^\intercal.$$

L'Equazione (2.90) è allora immediata conseguenza dell'Equazione (2.88). □

Definizione 107 *Chiamiamo* volatilità *del un portafoglio, e la denotiamo con σ, la deviazione standard del suo tasso di rendimento. In simboli*

$$\sigma(w_1, \dots, w_M) \overset{def}{=} \mathbf{D}[r(w_1, \dots, w_M, T)].$$

Osservazione 108 *Si ha*

$$\sigma = \left(\sum_{m=1}^{M} w_m^2 \sigma_m^2 + 2 \sum_{\substack{\ell,m=1 \\ \ell<m}}^{M} w_\ell w_m \rho_{\ell,m} \sigma_\ell \sigma_m \right)^{1/2}.$$

La struttura del rischio di un portafoglio suggerisce che il valore dello stesso può essere significativamente modificato mediante una scelta opportuna dei pesi dei titoli allocati in portafoglio in relazione alle loro covarianze.

Esempio 109 *Supponiamo di allocare in un portafoglio con lo stesso peso M titoli azionari aventi tutti lo stesso rendimento atteso $\bar{r}$, la stessa volatilità σ e a due a due scorrelati. Cioè tali che*

$$w_m = 1/M, \qquad \bar{r}_m(T) = \bar{r}, \qquad \sigma_m = \sigma,$$

per ogni $m = 1, \dots, M$ e

$$\rho_{\ell,m} = 0,$$

per ogni $\ell, m = 1, \dots, M$ tali che $\ell \neq m$. Abbiamo allora

$$\bar{r}(w_1, \dots, w_M, T) = \sum_{m=1}^{M} w_m \bar{r}_m(T) = \bar{r} \sum_{m=1}^{M} w_m = \bar{r},$$

ossia ***il tasso di rendimento atteso del portafoglio è pari al tasso di rendimento atteso di ciascuno dei titoli*** *in portafoglio. Risulta inoltre*

$$\sigma^2(w_1,\dots,w_M,T) = \sum_{m=1}^{M} w_m^2\sigma_m^2 + 2\sum_{\substack{\ell,m=1\\ \ell<m}}^{M} w_\ell w_m \rho_{\ell,m}\sigma_\ell\sigma_m = \sum_{m=1}^{M}\frac{\sigma_m^2}{M^2} = \frac{\sigma^2}{M},$$

ossia ***il rischio*** *[resp.* ***volatilità] del portafoglio è ridotta del fattore*** M *[resp.* $\sqrt{M}$*]* ***rispetto alla varianza [resp. volatilità] di ciascuno dei titoli*** *in portafoglio.*

Questo elementare caso limite presenta l'idea fondamentale di diversificazione nell'allocazione dei titoli in un portafoglio: **la diversificazione può ridurre il rischio di un portafoglio mantenendo inalterato il tasso di rendimento atteso**.

2.2.1 Primo problema di configurazione di portafoglio

Il primo problema configurazione di portafoglio si occupa della determinazione dei pesi dei portafogli di minima volatilità, ossia di minimo rischio, e della determinazione dei pesi dei portafogli di minima volatilità che realizzino un tasso di rendimento atteso assegnato. Come anticipato, per alleggerire le notazioni, da questa sezione scriveremo il tasso il tasso di rendimento [resp. rendimento atteso] dell'm-esimo titolo in portafoglio nella forma

$$r_m(T) \equiv r_m \qquad \text{[resp. } \bar{r}_m(T) \equiv \bar{r}_m\text{]},$$

per ogni $m = 1,\dots,M$. In conseguenza, scriveremo il tasso di rendimento [resp. rendimento atteso] di un portafoglio $\pi \equiv (w_1,\dots,w_M)$ nella forma

$$r(w_1,\dots,w_M) \equiv \sum_{m=1}^{M} w_m r_m \qquad \text{[resp. } \bar{r}(w_1,\dots,w_M) \equiv \sum_{m=1}^{M} w_m \bar{r}_m\text{]}. \tag{2.91}$$

Definizione 110 *Chiamiamo portafoglio di* minima volatilità *o di* minimo rischio *ogni portafoglio* $\pi \equiv (w_1,\dots,w_M) \in \mathbb{R}^M$ *soluzione del problema di ottimizzazione vincolata*

$$\min_{(w_1,\dots,w_M)\in\mathbb{H}} \{\sigma(w_1,\dots,w_M)\}, \tag{2.92}$$

o nell'equivalente problema, che presenta qualche vantaggio tecnico grazie alla forma della funzione obiettivo,

$$\min_{(w_1,\dots,w_M)\in\mathbb{H}} \{\frac{1}{2}\sigma^2(w_1,\dots,w_M)\}, \tag{2.93}$$

essendo $\mathbb{H}$ *l'insieme dei portafogli fattibili (cfr. Definizione 72).*

Teorema 111 *Se l'insieme dei titoli azionari in portafoglio è non singolare esiste un'unico portafoglio di minimo rischio.*
Proof. *Dall'Osservazione 100 sappiamo che* $\mathbb{H}$ *è un convesso di* $\mathbb{R}^M$*. D'altra parte il Teorema* **??** *assicura che la forma quadratica*

$$\frac{1}{2}\sigma^2(w_1,\dots,w_M) = \frac{1}{2}\sum_{m=1}^{M} w_m^2\sigma_m^2 + \sum_{\substack{\ell,m=1\\ \ell<m}}^{M} w_\ell w_m \rho_{\ell,m}\sigma_\ell\sigma_m$$

è strettamente convessa. Il risultato è allora un'applicazione del Teorema 103. □

Dal punto di vista tecnico la determinazione del portafoglio di minimo rischio assoluto si affronta con il metodo dei moltiplicatori di Lagrange. Si introduce cioè la funzione lagrangiana

$$L(w_1,\ldots,w_M,\lambda)=\frac{1}{2}\sigma^2(w_1,\ldots,w_M)-\lambda\left(\sum_{m=1}^{M}w_m-1\right),$$

e si considera il problema della ricerca dei suoi punti estremali, ossia delle soluzioni del sistema

$$\begin{aligned}\tfrac{\partial}{\partial w_1}L(w_1,\ldots,w_M,\lambda)&=0,\\ \vdots&=\vdots\\ \tfrac{\partial}{\partial w_M}L(w_1,\ldots,w_M,\lambda)&=0,\\ \tfrac{\partial}{\partial \lambda}L(w_1,\ldots,w_M,\lambda)&=0.\end{aligned}\tag{2.94}$$

Questo si esplicita come

$$\begin{aligned}\sigma_1^2w_1+\sum_{\substack{m=1\\m\neq 1}}^{M}w_m\rho_{1,m}\sigma_1\sigma_m-\lambda&=0,\\ \vdots&=\vdots\\ \sigma_M^2w_M+\sum_{\substack{m=1\\m\neq M}}^{M}w_m\rho_{M,m}\sigma_M\sigma_m-\lambda&=0,\\ \textstyle\sum_{m=1}^{M}w_m&=1.\end{aligned}\tag{2.95}$$

e costituisce un sistema non omogeneo di $M+1$ equazioni in $M+1$ incognite. L'ipotesi di non singolarità dell'insieme dei titoli azionari in portafoglio assicura sia che il Sistema (2.95) ammetta un'unica soluzione, sia che tale soluzione sia effettivamente il minimo assoluto cercato.

A titolo d'esempio, osserviamo che nel caso di un portafoglio con due titoli azionari S_1 e S_2 la matrice dei coefficienti del Sistema (2.95) è data da

$$A\equiv\begin{pmatrix}\sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 & -1\\ \rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2 & -1\\ 1 & 1 & 0\end{pmatrix},$$

con

$$\det(A)=\sigma_1^2-2\rho_{1,2}\sigma_1\sigma_2+\sigma_2^2.$$

D'altra parte, la matrice di varianza-covarianza dei titoli S_1 e S_2 è data da

$$\Sigma\equiv\begin{pmatrix}\sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2\\ \rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}$$

e si ha

$$\begin{pmatrix}1 & -1\end{pmatrix}\Sigma\begin{pmatrix}1\\ -1\end{pmatrix}=\det(A).$$

per cui se Σ è definita positiva otteniamo $\det(A)>0$. Notare che

$$\det(A)=\sum_{j,k=1}^{2}(-1)^{j+k}\left|\Sigma_{j,k}\right|,$$

dove $\Sigma_{j,k}$ è la sottomatrice di Σ ottenuta cancellando la j-esima riga e la k-esima colonna, per $j,k=1,2$, e $|\Sigma_{j,k}|\equiv\det(\Sigma_{j,k})$. Nel caso di un portafoglio con tre titoli azionari S_1, S_2 e S_3 la matrice dei coefficienti del Sistema (2.95) è data da

$$A \equiv \begin{pmatrix} \sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 & \rho_{1,3}\sigma_1\sigma_3 & -1 \\ \rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{2,3}\sigma_2\sigma_3 & -1 \\ \rho_{1,3}\sigma_1\sigma_3 & \rho_{2,3}\sigma_2\sigma_3 & \sigma_3^2 & -1 \\ 1 & 1 & 1 & 0 \end{pmatrix},$$

con

$$\begin{aligned} \det(A) &= \sigma_1^2\sigma_2^2\left(1-\rho_{1,2}^2\right) + \sigma_1^2\sigma_3^2\left(1-\rho_{1,3}^2\right) + \sigma_2^2\sigma_3^2\left(1-\rho_{2,3}^2\right) \\ &\quad + 2\sigma_1\sigma_2\sigma_3\left(\sigma_1\left(\rho_{1,2}\rho_{1,3}-\rho_{2,3}\right) + \sigma_2\left(\rho_{1,2}\rho_{2,3}-\rho_{1,3}\right) + \sigma_3\left(\rho_{1,3}\rho_{2,3}-\rho_{1,2}\right)\right). \end{aligned}$$

D'altra parte, la matrice di varianza-covarianza dei titoli S_1, S_2 e S_3 è data da

$$\Sigma \equiv \begin{pmatrix} \sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 & \rho_{1,3}\sigma_1\sigma_3 \\ \rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{2,3}\sigma_2\sigma_3 \\ \rho_{1,3}\sigma_1\sigma_3 & \rho_{2,3}\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix},$$

i minori principali di Σ sono

$$\begin{aligned} |\Sigma_{1,1}| &= \begin{vmatrix} \sigma_2^2 & \rho_{2,3}\sigma_2\sigma_3 \\ \rho_{2,3}\sigma_2\sigma_3 & \sigma_3^2 \end{vmatrix} = \sigma_2^2\sigma_3^2\left(1-\rho_{2,3}^2\right) \\ |\Sigma_{2,2}| &= \begin{vmatrix} \sigma_1^2 & \rho_{1,3}\sigma_1\sigma_3 \\ \rho_{1,3}\sigma_1\sigma_3 & \sigma_3^2 \end{vmatrix} = \sigma_1^2\sigma_3^2\left(1-\rho_{1,3}^2\right), \\ |\Sigma_{3,3}| &= \begin{vmatrix} \sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 \\ \rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2 \end{vmatrix} = \sigma_1^2\sigma_2^2\left(1-\rho_{1,2}^2\right), \end{aligned}$$

inoltre abbiamo

$$\begin{aligned} |\Sigma_{1,2}| &= \begin{matrix} \rho_{1,2}\sigma_1\sigma_2 & \rho_{2,3}\sigma_2\sigma_3 \\ \rho_{1,3}\sigma_1\sigma_3 & \sigma_3^2 \end{matrix} = \sigma_1\sigma_2\sigma_3^2\left(\rho_{1,2}-\rho_{1,3}\rho_{2,3}\right), \\ |\Sigma_{1,3}| &= \begin{matrix} \rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2 \\ \rho_{1,3}\sigma_1\sigma_3 & \rho_{2,3}\sigma_2\sigma_3 \end{matrix} = \sigma_1\sigma_2^2\sigma_3\left(\rho_{1,2}\rho_{2,3}-\rho_{1,3}\right) \\ |\Sigma_{2,3}| &= \begin{matrix} \sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 \\ \rho_{1,3}\sigma_1\sigma_3 & \rho_{2,3}\sigma_2\sigma_3 \end{matrix} = \sigma_1^2\sigma_2\sigma_3\left(\rho_{2,3}-\rho_{1,2}\rho_{1,3}\right) \end{aligned}$$

per cui possiamo ancora scrivere

$$\det(A) = |\Sigma_{1,1}| + |\Sigma_{2,2}| + |\Sigma_{2,2}| + 2\sum_{\substack{j,k=1 \\ j<k}}^{3} (-1)^{j+k}|\Sigma_{j,k}| = \sum_{j,k=1}^{3} (-1)^{j+k}|\Sigma_{j,k}|,$$

essendo sempre $\Sigma_{j,k}$ la sottomatrice di Σ ottenuta cancellando la j-esima riga e la k-esima colonna, per $j,k=1,2$, ed essendo $|\Sigma_{j,k}| \equiv \det(\Sigma_{j,k})$. Chiaramente,

$$|\Sigma_{j,k}| = |\Sigma_{k,j}|,$$

per ogni $j,k=1,2,3$. e il polinomio caratteristico è dato da

$$\det(\Sigma - \lambda I) = \lambda^3 + a\lambda^2 + b\lambda + c$$

dove

$$\begin{gathered} a = -\left(\sigma_1^2+\sigma_2^2+\sigma_3^2\right) = -\left(\lambda_1+\lambda_2+\lambda_3\right) = -\operatorname{trace}(\Sigma), \\ b = \sigma_1^2\sigma_2^2\left(1-\rho_{1,2}^2\right) + \sigma_1^2\sigma_3^2\left(1-\rho_{1,3}^2\right) + \sigma_2^2\sigma_3^2\left(1-\rho_{2,3}^2\right) = \lambda_1\lambda_2+\lambda_1\lambda_3+\lambda_2\lambda_3 \end{gathered}$$

e

$$c = -\sigma_1^2\sigma_2^2\sigma_3^2\left(1+2\rho_{1,2}\rho_{1,3}\rho_{2,3}-\rho_{1,2}^2-\rho_{1,3}^2-\rho_{2,3}^2\right) = -\lambda_1\lambda_2\lambda_3 = -\det(\Sigma).$$

...da completare...

Definizione 112 *Chiamiamo* rendimento atteso *[resp.* volatilità*] del portafoglio di minimo rischio il numero reale [risp. il numero reale positivo]*

$$\bar{r}^* \equiv \sum_{m=1}^{M} w_m^\star \bar{r}_m \qquad \textit{[resp. } \sigma^* \equiv \sigma(w_1^*, \dots, w_M^*)\textit{]},$$

essendo $(w_1^*, \dots, w_M^*)$ *la* M*-pla di pesi caratterizzante il portafoglio di minimo rischio.*

Definizione 113 *Comunque fissato* $\bar{r} \in \mathbb{R}$*, chiamiamo* insieme dei portafogli fattibili di rendimento atteso $\bar{r} \in \mathbb{R}$ *l'insieme* $\mathbb{H} \cap \mathbb{K}_{\bar{r}}$ *essendo* $\mathbb{H}$ *l'insieme dei portafogli fattibili ed essendo*

$$\mathbb{K}_{\bar{r}} \equiv \left\{ (w_1, \dots, w_M) \in \mathbb{R}^M : \sum_{m=1}^{M} w_m \bar{r}_m = \bar{r} \right\}. \tag{2.96}$$

Proposizione 114 *L'insieme* $\mathbb{K}_{\bar{r}}$ *è un iperpiano di* $\mathbb{R}^M$*. Tale iperpiano si caratterizza come l'iperpiano ortogonale al vettore* $(\bar{r}_1, \dots, \bar{r}_M)$ *e passante per il punto* $(\bar{r}/M\bar{r}_1, \dots, \bar{r}/M\bar{r}_M)$*. Conseguentemente l'insieme* $\mathbb{H} \cap \mathbb{K}_r$ *è una varietà lineare di dimensione* $M-2$*.*
Proof. *Basta osservare che l'equazione*

$$\sum_{m=1}^{M} w_m \bar{r}_m = \bar{r}$$

è equivalente all'equazione

$$\sum_{m=1}^{M} \bar{r}_m \left(w_m - \frac{\bar{r}}{M \bar{r}_m} \right) = 0.$$

□

Definizione 115 *Comunque fissato* $\bar{r} \in \mathbb{R}$*, chiamiamo portafoglio di* minima volatilità *o* minimo rischio dato il rendimento atteso $\bar{r}$ *il portafoglio* $\pi \equiv (w_1, \dots, w_M)$ *caratterizzato dalla* M*-pla di pesi soluzione del problema di ottimizzazione vincolata*

$$\min_{(w_1,\dots,w_M) \in \mathbb{H} \cap \mathbb{K}_{\bar{r}}} \{\sigma(w_1, \dots, w_M)\}, \tag{2.97}$$

o dell'equivalente problema

$$\min_{(w_1,\dots,w_M) \in \mathbb{H} \cap \mathbb{K}_{\bar{r}}} \{\frac{1}{2}\sigma^2(w_1, \dots, w_M)\}. \tag{2.98}$$

essendo $\mathbb{H} \cap \mathbb{K}_{\bar{r}}$ *l'insieme dei portafogli fattibili di assegnato tasso di rendimento atteso* $\bar{r}$ *(cfr Definizioni 72 e 113).*

Teorema 116 *Se l'insieme dei titoli azionari in portafoglio è non singolare esiste un'unico portafoglio di minimo rischio per ogni assegnato tasso di rendimento atteso* $\bar{r}$*.*
Proof. *Stante la Proposizione 114 e l'Osservazione 101 il vincolo* $\mathbb{H} \cap \mathbb{K}_{\bar{r}}$ *è convesso. La dimostrazione dell'esistenza e unicità del portafoglio di minimo rischio dato il rendimento atteso* $\bar{r}$ *si completa allora in perfetta analogia con la dimostrazione del Theorem 111* □

Similmente al caso della determinazione del portafoglio di minimo rischio assoluto, la determinazione del portafoglio di minimo rischio e avente un dato rendimento $\bar{r}$ si affronta con il metodo dei moltiplicatori di Lagrange. Si introduce la funzione lagrangiana

$$L(w_1, \dots, w_M, \lambda_1, \lambda_2) = \frac{1}{2}\sigma^2(w_1, \dots, w_M) - \lambda_1 \left(\sum_{m=1}^{M} w_m - 1 \right) - \lambda_2 \left(\sum_{m=1}^{M} w_m \bar{r}_m - \bar{r} \right),$$

e si considera il problema della ricerca dei suoi punti estremali, ossia delle soluzioni del sistema

$$
\begin{aligned}
\tfrac{\partial}{\partial w_1} L(w_1,\dots,w_M,\lambda_1,\lambda_2) &= 0,\\
\vdots &= \vdots\\
\tfrac{\partial}{\partial w_M} L(w_1,\dots,w_M,\lambda_1,\lambda_2) &= 0,\\
\tfrac{\partial}{\partial \lambda_1} L(w_1,\dots,w_M,\lambda_1,\lambda_2) &= 0,\\
\tfrac{\partial}{\partial \lambda_2} L(w_1,\dots,w_M,\lambda_1,\lambda_2) &= 0,
\end{aligned}
\tag{2.99}
$$

che si esplicita come

$$
\begin{aligned}
\sigma_1^2 w_1 + \sum_{\substack{m=1\\ m\neq 1}}^{M} w_m \rho_{1,m}\sigma_1\sigma_m - \lambda_1 - \lambda_2 \bar{r}_1 &= 0,\\
\vdots &= \vdots\\
\sigma_M^2 w_M + \sum_{\substack{m=1\\ m\neq M}}^{M} w_m \rho_{M,m}\sigma_M\sigma_m - \lambda_1 - \lambda_2 \bar{r}_M &= 0,\\
\textstyle\sum_{m=1}^{M} w_m &= 1,\\
\textstyle\sum_{m=1}^{M} w_m \bar{r}_m &= \bar{r},
\end{aligned}
\tag{2.100}
$$

e costituisce un sistema lineare non omogeneo di $M+2$ equazioni in $M+2$ incognite per il quale valgono esattamente le stesse considerazioni relative al sistema per la determinazione del portafoglio di minimo rischio assoluto.

Anche in questo caso, a titolo d'esempio, osserviamo che nel caso di un portafoglio con due titoli azionari S_1 e S_2 la matrice dei coefficienti del Sistema (2.100) è data da

$$
A \equiv \begin{pmatrix}
\sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 & -1 & -\bar{r}_1\\
\rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2 & -1 & -\bar{r}_2\\
1 & 1 & 0 & 0\\
\bar{r}_1 & \bar{r}_2 & 0 & 0
\end{pmatrix}
$$

con

$$
\det(A) = \bar{r}_1^2 - 2\bar{r}_1\bar{r}_2 + \bar{r}_2^2.
$$

Nel caso di un portafoglio con tre titoli azionari la matrice dei coefficienti del Sistema (2.100) è data da

$$
A \equiv \begin{pmatrix}
\sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 & \rho_{1,3}\sigma_1\sigma_3 & -1 & -\bar{r}_1\\
\rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{2,3}\sigma_2\sigma_3 & -1 & -\bar{r}_2\\
\rho_{1,3}\sigma_1\sigma_3 & \rho_{2,3}\sigma_2\sigma_3 & \sigma_3^2 & -1 & -\bar{r}_3\\
1 & 1 & 1 & 0 & 0\\
\bar{r}_1 & \bar{r}_2 & \bar{r}_3 & 0 & 0
\end{pmatrix}
$$

con

$$
\begin{aligned}
\det(A) = {} & \sigma_1^2(\bar{r}_2 - r_3)^2 + \sigma_2^2(\bar{r}_1 - r_3)^2 + \sigma_3^2(\bar{r}_1 - \bar{r}_2)^2\\
& - 2\sigma_1\sigma_2\rho_{1,2}(\bar{r}_1 - \bar{r}_3)(\bar{r}_2 - \bar{r}_3) + 2\sigma_1\sigma_3\rho_{1,3}(\bar{r}_1 - \bar{r}_2)(\bar{r}_2 - \bar{r}_3) - 2\sigma_2\sigma_3\rho_{2,3}(\bar{r}_1 - \bar{r}_2)(\bar{r}_1 - \bar{r}_3).
\end{aligned}
$$

Considerata la funzione $\Pi : \mathbb{R}^M \to \mathbb{R}_+ \times \mathbb{R}$ definita ponendo

$$
\Pi(w_1,\dots,w_M) \overset{\text{def}}{=} (\sigma(w_1,\dots,w_M), \bar{r}(w_1,\dots,w_M)),
$$

l'insieme $\Pi(\mathbb{H})$, immagine dell'iperpiano $\mathbb{H} \subseteq \mathbb{R}^M$ mediante Π, è un sottoinsieme di $\mathbb{R}_+ \times \mathbb{R}$ che rappresenta tutte le coppie di punti $(\sigma, \bar{r})$ che costituiscono il profilo di rischio-rendimento atteso di tutti i portafogli azionari fattibili. Se l'insieme dei titoli azionari in portafoglio è non singolare, l'esistenza di un unico portafoglio di minimo rischio, caratterizzato dal profilo di rischio-rendimento atteso $(\sigma^*, \bar{r}^*)$, assicura che l'insieme $\Pi(\mathbb{H})$ è interamente contenuto nel semipiano di $\mathbb{R}_+ \times \mathbb{R}$ definito come

$$\{(\sigma, \bar{r}) \in \mathbb{R}_+ \times \mathbb{R} : \sigma \geqslant \sigma^*\}.$$

Definizione 117 *Chiamiamo* sottoinsieme dei portafogli azionari di minimo rischio *dell'insieme dei portafogli azionari fattibili l'insieme dei portafogli $\pi \equiv (w_1, \ldots, w_M)$ soluzioni del Problema (2.97) al variare di $\bar{r} \in \mathbb{R}$.*

Da notare che, per come è costruito, l'insieme dei portafogli azionari di minimo rischio può contenere portafogli che abbiano un rendimento atteso minore del rendimento atteso del portafoglio di minimo rischio ma, paradossalmente, un rischio maggiore! Infatti anche per $\bar{r} < \bar{r}^*$ esiste, in generale, un'unica soluzione del Problema (2.97), che dovrà soddisfare la condizione $\sigma > \sigma^*$. Ovviamente, dal punto di vista di un investitore, questi portafogli non sono di alcun interesse. I restanti portafogli dell'insieme di minimo rischio possono avere interesse per un investitore e costituiscono la cosiddetta frontiera efficiente dell'insieme dei portafogli di minimo rischio. Precisamente

Definizione 118 *Chiamiamo* frontiera efficiente dell'insieme dei portafogli azionari di minimo rischio *l'insieme dei portafogli $\pi \equiv (w_1, \ldots, w_M)$ soluzioni del Problema (2.97) al variare di $\bar{r} \geqslant \bar{r}^*$, essendo $\bar{r}^*$ il rendimento atteso del portafoglio di minimo rischio.*

Definizione 119 *Chiamiamo* portafoglio azionario efficiente o*gni portafoglio azionario della frontiera efficiente.*

Proposizione 120 *Se l'insieme dei titoli azionari in portafoglio è non singolare, l'insieme*

$$\Pi(\mathbb{H} \cap \mathbb{K}_{\bar{r}}) = \{(\sigma, \bar{r}) \in \mathbb{R}_+ \times \mathbb{R} : \sigma = \sigma(w_1, \ldots, w_M),\ \bar{r} = \bar{r}(w_1, \ldots, w_M), \quad \forall (w_1, \ldots, w_M) \in \mathbb{H} \cap \mathbb{K}_{\bar{r}}\}$$

è un convesso di $\mathbb{R}_+ \times \mathbb{R}$, per ogni $\bar{r} \in \mathbb{R}$.
Proof. *Se l'insieme dei titoli azionari in portafoglio è non singolare, la forma quadratica $\sigma^2 : \mathbb{R}^M \to \mathbb{R}$ definita dall'Equazione (**??**) è definita positiva e strettamente convessa. Pertanto, l'immagine $\sigma^2(\mathbb{H} \cap \mathbb{K}_{\bar{r}})$ del convesso $\mathbb{H} \cap \mathbb{K}_{\bar{r}}$ di $\mathbb{R}^M$ è un convesso di $\mathbb{R}_+$. D'altra parte, un convesso di $\mathbb{R}_+$ è necessariamente un intervallo. Quindi $\sigma^2(\mathbb{H} \cap \mathbb{K}_{\bar{r}}) \equiv \mathbb{I}$, per un opportuno intervallo $\mathbb{I}$ di $\mathbb{R}_+$, e*

$$\sigma(\mathbb{H} \cap \mathbb{K}_{\bar{r}}) = \sqrt{\sigma^2(\mathbb{H} \cap \mathbb{K}_{\bar{r}})} = \sqrt{\mathbb{I}}.$$

Ora la funzione radice quadrata $\sqrt{\cdot} : \mathbb{R}_+ \to \mathbb{R}_+$ è una funzione continua. Ne segue che $\sqrt{\mathbb{I}} \equiv \mathbb{J}$ è anch'esso un intervallo di $\mathbb{R}_+$ e quindi un convesso. Si ha poi chiaramente che $\mathbb{J} \times \{\bar{r}\}$ è un convesso di $\mathbb{R}_+ \times \mathbb{R}$, per ogni $r \in \mathbb{R}$. In definitiva, poichè

$$\bar{r}(\mathbb{H} \cap \mathbb{K}_{\bar{r}}) = \bar{r},$$

l'insieme

$$\Pi(\mathbb{H} \cap \mathbb{K}_{\bar{r}}) = \{(\sigma, \bar{r}) \in \mathbb{R}_+ \times \mathbb{R} : \sigma \in \sigma(\mathbb{H} \cap \mathbb{K}_{\bar{r}}),\ \bar{r} \in \bar{r}(\mathbb{H} \cap \mathbb{K}_{\bar{r}}),\} = \sigma(\mathbb{H} \cap \mathbb{K}_{\bar{r}}) \times \{\bar{r}\}$$

è un convesso di $\mathbb{R}_+ \times \mathbb{R}$. □

Osservazione 121 *Se l'insieme dei titoli in portafoglio è non singolare, denotato con $(\sigma_0, \bar{r}_0)$ il profilo rischio-rendimento del portafoglio azionario di minimo rischio dato il rendimento $\bar{r}_0$ si ha*

$$\sigma \geqslant \sigma_0 \quad e \quad \bar{r} = \bar{r}_0,$$

per ogni $(\sigma, \bar{r}) \in \Pi(\mathbb{H} \cap \mathbb{K}_{\bar{r}})$.

Qualora si debbano determinare diverse M-ple di pesi con cui comporre portafogli efficienti corrispondenti a diversi rendimenti attesi, piuttosto che affronare la risoluzione di diverse copie del Sistema (2.100), ciascuna delle quali corrispondente a uno dei rendimenti attesi cui siamo interessati, è possibile applicare una procedura che produce un significativo risparmio computazionale. Si considera il sistema di M equazioni in $M+2$ incognite ottenuto da (2.100) sopprimendo le ultime due equazioni. Ossia il sistema

$$\begin{aligned} \sigma_1^2 w_1 + \sum_{\substack{m=1 \\ m\neq 1}}^{M} w_m \rho_{1,m}\sigma_1\sigma_m - \lambda_1 - \lambda_2 \bar{r}_1 &= 0, \\ \vdots &= \vdots \\ \sigma_M^2 w_M + \sum_{\substack{m=1 \\ m\neq M}}^{M} w_m \rho_{M,m}\sigma_M\sigma_m - \lambda_1 - \lambda_2 \bar{r}_M &= 0, \end{aligned} \tag{2.101}$$

Quindi, scegliendo due coppie indipendenti di valori per λ_1 e λ_2 in (2.101), ad esempio ponendo $\lambda_1 = 0$, $\lambda_2 = 1$ e $\lambda_1 = 1$, $\lambda_2 = 0$ si eliminano 2 incognite da (2.101) e si ottengono rispettivamente i sistemi di M equazioni in M incognite

$$\begin{aligned} \sigma_1^2 w_1 + \sum_{\substack{\ell=1 \\ \ell\neq 1}}^{M} w_\ell \rho_{1,\ell}\sigma_1\sigma_\ell &= \bar{r}_1, \\ \vdots &= \vdots \\ \sigma_M^2 w_M + \sum_{\substack{\ell=1 \\ \ell\neq M}}^{M} w_\ell \rho_{M,\ell}\sigma_M\sigma_\ell &= \bar{r}_M, \end{aligned} \tag{2.102}$$

e

$$\begin{aligned} \sigma_1^2 w_1 + \sum_{\substack{\ell=1 \\ \ell\neq 1}}^{M} w_\ell \rho_{1,\ell}\sigma_1\sigma_\ell &= 1, \\ \vdots &= \vdots \\ \sigma_M^2 w_M + \sum_{\substack{\ell=1 \\ \ell\neq M}}^{M} w_\ell \rho_{M,\ell}\sigma_M\sigma_\ell &= 1. \end{aligned} \tag{2.103}$$

La risoluzione di questi due sistemi porta alla determinazione di due M-ple di pesi siano esse $(w_1^{(j)}, \dots, w_M^{(j)})$ per $j = 1, 2$, che non sono, in generale, soluzioni del Sistema (2.100), ma che possono essere trasformate in soluzioni di (2.100) con delle semplici modifiche. Infatti, considerate le M-ple $(\hat{w}_1^{(j)}, \dots, \hat{w}_M^{(j)})$ per $j = 1, 2$, definite ponendo

$$\hat{w}_m^{(j)} \stackrel{\text{def}}{=} \frac{w_m^{(j)}}{\sum_{\ell=1}^{M} w_\ell^{(j)}}, \quad m = 1, \dots, M,$$

e sostituendo tali M-ple ai pesi che figurano come incognite nel Sistema (2.100) è immediato verificare che l'equazione

$$\sum_{m=1}^{M} \hat{w}_m^{(j)} - 1 = 0$$

è soddisfatta per ogni $j = 1, 2$, mentre l'equazione

$$\sum_{m=1}^{M} \hat{w}_m^{(j)} \bar{r}_m - \bar{r} = 0$$

fornisce di fatto i rendimenti

$$\bar{r}^{(j)} \equiv \sum_{m=1}^{M} \frac{w_m^{(j)}}{\sum_{\ell=1}^{M} w_\ell^{(j)}} \bar{r}_m, \quad j = 1, 2. \tag{2.104}$$

Inoltre, i primi membri delle prime M equazioni assumono la forma

$$\begin{aligned}
&\sigma_m^2 \hat{w}_m^{(j)} + \sum_{\substack{\ell=1 \\ \ell \neq m}}^{M} \hat{w}_\ell^{(j)} \rho_{m,\ell} \sigma_m \sigma_\ell - \lambda_1 - \lambda_2 \bar{r}_m \\
&= \sigma_m^2 \frac{w_m^{(j)}}{\sum_{\ell=1}^{M} w_\ell^{(j)}} + \sum_{\substack{\ell=1 \\ \ell \neq m}}^{M} \frac{w_\ell^{(j)}}{\sum_{\ell=1}^{M} w_\ell^{(j)}} \rho_{m,\ell} \sigma_m \sigma_\ell - \lambda_1 - \lambda_2 \bar{r}_m \\
&= \frac{1}{\sum_{\ell=1}^{M} w_\ell^{(j)}} \left(\sigma_m^2 w_m^{(j)} + \sum_{\substack{\ell=1 \\ \ell \neq m}}^{M} w_\ell^{(j)} \rho_{m,\ell} \sigma_m \sigma_\ell \right) - \lambda_1 - \lambda_2 \bar{r}_m,
\end{aligned}$$

per ogni $j = 1, 2$ e ogni $m = 1, \ldots, M$. Pertanto, stanti le (2.102) e (2.103), otteniamo

$$\sigma_m^2 \hat{w}_m^{(1)} + \sum_{\substack{\ell=1 \\ \ell \neq m}}^{M} \hat{w}_\ell^{(1)} \rho_{m,\ell} \sigma_m \sigma_\ell - \lambda_1 - \lambda_2 \bar{r}_m = \frac{\bar{r}_m}{\sum_{\ell=1}^{M} w_\ell^{(1)}} - \lambda_1 - \lambda_2 \bar{r}_m \tag{2.105}$$

e

$$\sigma_m^2 \hat{w}_m^{(2)} + \sum_{\substack{\ell=1 \\ \ell \neq m}}^{M} \hat{w}_\ell^{(2)} \rho_{m,\ell} \sigma_m \sigma_\ell - \lambda_1 - \lambda_2 \bar{r}_m = \frac{1}{\sum_{\ell=1}^{M} w_\ell^{(2)}} - \lambda_1 - \lambda_2 \bar{r}_m, \tag{2.106}$$

per ogni $m = 1, \ldots, M$ e le (2.105), (2.106) si annullano scegliendo

$$\lambda_1^{(1)} = 0, \quad \lambda_2^{(1)} = \frac{1}{\sum_{\ell=1}^{M} w_\ell^{(1)}},$$

e

$$\lambda_1^{(2)} = \frac{1}{\sum_{\ell=1}^{M} w_\ell^{(2)}}, \quad \lambda_2^{(2)} = 0,$$

rispettivamente. In definitiva, questa procedura conduce alla determinazione delle soluzioni

$$(\hat{w}_1^{(1)}, \ldots, \hat{w}_M^{(1)}, \lambda_1^{(1)}, \lambda_2^{(1)}) \equiv \left(\frac{w_1^{(1)}}{\sum_{\ell=1}^{M} w_\ell^{(1)}}, \ldots, \frac{w_M^{(1)}}{\sum_{\ell=1}^{M} w_\ell^{(1)}}, 0, \frac{1}{\sum_{\ell=1}^{M} w_\ell^{(1)}} \right) \tag{2.107}$$

e

$$(\hat{w}_1^{(2)}, \ldots, \hat{w}_M^{(2)}, \lambda_1^{(2)}, \lambda_2^{(2)}) \equiv \left(\frac{w_1^{(2)}}{\sum_{\ell=1}^{M} w_\ell^{(2)}}, \ldots, \frac{w_M^{(2)}}{\sum_{\ell=1}^{M} w_\ell^{(2)}}, \frac{1}{\sum_{\ell=1}^{M} w_\ell^{(2)}}, 0 \right) \tag{2.108}$$

del Sistema (2.100) corrispondenti ai rendimenti attesi $\bar{r}^{(j)}$ dati dalla (2.104), per $j = 1, 2$. Da notare che la seconda delle soluzioni così determinate soddisfa di fatto il Sistema (2.95) e individua pertanto il portafoglio di minimo rischio assoluto. Una volta determinate le M-ple di pesi $(\hat{w}_1^{(j)}, \ldots, \hat{w}_M^{(j)})$ che individuano i portafogli efficienti di rendimenti attesi $\bar{r}^{(j)}$, per $j = 1, 2$, la costruzione della M-pla di pesi che individua il portafoglio di minimo rischio per un rendimento atteso $\bar{r}$ si effettua nel seguente modo: si determina $\alpha \in \mathbb{R}$ soluzione dell'equazione

$$\bar{r} = \alpha \bar{r}^{(1)} + (1 - \alpha) \bar{r}^{(2)}$$

ed alla luce delle precedenti considerazioni, si può facilmente provare che la M-pla $(w_1, \ldots, w_M)$ definita ponendo

$$w_m \equiv \alpha \hat{w}_m^{(1)} + (1-\alpha)\hat{w}_m^{(2)}, \quad m = 1, \ldots, M$$

caratterizza il portafoglio di minimo rischio e rendimento atteso $\bar{r}$.

Ribadiamo che la procedura considerata, applicazione del "Teorema dei Due Fondi" che esporremo di seguito, è particolarmente utile nel caso in cui sia necessario determinare i portafogli efficienti corrispondenti a diversi rendimenti. In questo caso, il risparmio computazionale è veramente significativo.

Teorema dei due fondi

Consideriamo un insieme di M titoli che dividiamo in due sottoinsiemi contenenti M_1 ed M_2 titoli rispettivamente. Denotiamo con $\bar{r}_1^{(1)}, \bar{r}_2^{(1)}, \ldots, \bar{r}_{M_1}^{(1)}$ [risp. $\sigma_{1,1}^{(1)}, \sigma_{1,2}^{(1)}, \ldots, \sigma_{M^{(1)},M^{(1)}}^{(1)}$] i rendimenti attesi [risp. le covarianze] dei titoli contenuti nel primo sottoinsieme e con $\bar{r}_1^{(2)}, \bar{r}_2^{(2)}, \ldots, \bar{r}_{M_2}^{(2)}$ [risp. $\sigma_{1,1}^{(2)}, \sigma_{1,2}^{(2)}, \ldots, \sigma_{M^{(2)},M^{(2)}}^{(2)}$] i rendimenti attesi [risp. le covarianze] dei titoli contenuti nel secondo sottoinsieme.

Proposizione 122 *Supponiamo di avere individuato un portafoglio del sottoinsieme di minimo rischio dell'insieme dei portafogli ammissibili generati dai titoli dell'insieme M_j e sia tale portafoglio $\pi_j \equiv (w_1^{(j)}, \ldots, w_{M_j}^{(j)})$, per $j = 1, 2$. Allora ogni portafoglio del sottoinsieme di minimo rischio dell'insieme dei portafogli ammissibili generati dai titoli dell'insieme M è ottenibile nella forma*

$$\alpha\pi_1 + (1-\alpha)\pi_2,$$

al variare di $\alpha \in \mathbb{R}$.
Proof. *Fissato $\bar{r} \in \mathbb{R}$ consideriamo le equazioni*

$$\begin{aligned} \sum_{m=1}^{M} w_m \sigma_{1,m} - \lambda_1 - \lambda_2 \bar{r}_1 &= 0, \\ \vdots &= \vdots \\ \sum_{m=1}^{M} w_m \sigma_{M,m} - \lambda_1 - \lambda_2 \bar{r}_M &= 0, \\ \textstyle\sum_{m=1}^{M} w_m \bar{r}_m &= \bar{r}, \\ \textstyle\sum_{m=1}^{M} w_m &= 1, \end{aligned} \tag{2.109}$$

che caratterizzano il portafoglio di minimo rischio per il rendimento $\bar{r}$ assegnato. Fissati $\bar{r}_1, \bar{r}_2 \in \mathbb{R}$, dell'insieme dei rendimenti ammissibili per i portafogli generati dall'insieme dei titoli M_1, M_2, rispettivamente, sia $\alpha \in \mathbb{R}$ tale che

$$\bar{r} = \alpha\bar{r}_1 + (1-\alpha)\bar{r}_2.$$

Quindi si considerino le equazioni che caratterizzano i portafogli di minimo rischio per il rendimento $\bar{r}_j$ assegnato nell'insieme dei portafogli generato dai titoli dell'insieme M_j, per $j = 1, 2$. Ossia

$$\begin{aligned} \sum_{m_j=1}^{M_j} w_{m_j}^{(j)} \sigma_{1,m_j}^{(j)} - \lambda_1^{(j)} - \lambda_2^{(j)} \bar{r}_1^{(j)} &= 0, \\ \vdots &= \vdots \\ \sum_{m_j=1}^{M_j} w_{m_j}^{(j)} \sigma_{M_j,m_j}^{(j)} - \lambda_1^{(j)} - \lambda_2^{(j)} \bar{r}_{M_j}^{(j)} &= 0, \\ \textstyle\sum_{m_j=1}^{M_j} w_{m_j}^{(j)} \bar{r}_{m_j}^{(j)} &= \bar{r}^{(j)}, \\ \textstyle\sum_{m_j=1}^{M_j} w_{m_j}^{(j)} &= 1, \end{aligned} \tag{2.110}$$

Se $(w_1^{(j)}, \dots, w_{M_j}^{(j)}, \lambda_1^{(j)}, \lambda_2^{(j)})$ *sono soluzioni di (2.110), per* $j = 1, 2$ *rispettivamente, allora, posto*

$$\pi_j \equiv (w_1^{(j)}, \dots, w_{M_j}^{(j)}), \quad \lambda^{(j)} \equiv (\lambda_1^{(j)}, \lambda_2^{(j)}), \quad j = 1, 2,$$

per la linearità delle equazioni di (2.109) non è difficile rendersi conto che $\pi \equiv \alpha\pi_1 + (1-\alpha)\pi_2$, $\lambda \equiv \alpha\lambda^{(1)} + (1-\alpha)\lambda^{(2)}$, *è soluzione di (2.109).* □

Come conseguenza abbiamo

Teorema 123 *Dato un insieme di* M *titoli è possibile individuare due portafogli efficienti formati da titoli di sottoinsiemi* M_j *di* M, *per* $j = 1, 2$, *tali che tutti i portafogli della frontiera efficiente dell'insieme dei portafogli generati dai titoli dell'insieme* M *siano ottenibili, in termini di tasso di rendimento atteso e deviazione standard, come combinazione lineare dei due portafogli efficienti individuati.*

Esempio 124 *Sono disponibili tre titoli con tassi di rendimento attesi e varianze-covarianze dei tassi di rendimento individuati dalle seguenti matrici*

$$(\bar{r}_m)_{m=1,2,3} \equiv \begin{pmatrix} 0.4 \\ 0.8 \\ 0.8 \end{pmatrix}, \qquad \mathbf{D}^2 \equiv \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}.$$

Assegnato un tasso di rendimento atteso, $\bar{r}$, *il portafoglio che realizza il tasso assegnato è caratterizzato dalla condizione*

$$\min_{\substack{w_1+w_2+w_3=1 \\ 0.4w_1+0.8w_2+0.8w_3=\bar{r}}} \{w_1^2 + w_2^2 + w_3^2 + w_1w_2 + w_2w_3\}. \tag{2.111}$$

La Lagrangiana associata al problema di minimizzazione è dunque

$$L(w_1, w_2, w_3, \lambda_1, \lambda_2) = w_1^2 + w_2^2 + w_3^2 + w_1w_2 + w_2w_3 - \lambda_1(w_1+w_2+w_3-1) - \lambda_2(0.4w_1+0.8w_2+0.8w_3-\bar{r}).$$

Pertanto il problema della ricerca dei punti estremali diviene

$$\begin{aligned} \frac{\partial L}{\partial w_1} &= 2w_1 + w_2 - \lambda_1 - 0.4\lambda_2 = 0, \\ \frac{\partial L}{\partial w_2} &= 2w_2 + w_1 + w_3 - \lambda_1 - 0.8\lambda_2 = 0, \\ \frac{\partial L}{\partial w_3} &= 2w_3 + w_2 - \lambda_1 - 0.8\lambda_2 = 0, \\ \frac{\partial L}{\partial \lambda_1} &= -(w_1 + w_2 + w_3 - 1) = 0, \\ \frac{\partial L}{\partial \lambda_2} &= -(0.4w_1 + 0.8w_2 + 0.8w_3 - \bar{r}) = 0. \end{aligned} \tag{2.112}$$

Non è difficile provare che la soluzione di tale sistema lineare è data da

$$w_1 = 2.0 - 2.5\bar{r}, \quad w_2 = 2.5\bar{r} - 1.5, \quad w_3 = 0.5, \quad \lambda_1 = 5.5 - 7.5\bar{r}, \quad \lambda_2 = 12.5\bar{r} - 7.5.$$

I pesi w_1, w_2, w_3 *così determinati individuano il portafoglio di minimo rischio per ogni assegnato rendimento* $\bar{r}$. *Ad esempio, considerato un rendimento atteso* $\bar{r} \equiv 1$, *il portafoglio di minimo rischio che consente di conseguirlo è caratterizzato dai pesi*

$$w_1 = -0.5, \quad w_2 = 1, \quad w_3 = 0.5.$$

Il portafoglio di minimo rischio è altresì caratterizzato dalla condizione

$$\min_{w_1+w_2+w_3=1} \{w_1^2 + w_2^2 + w_3^2 + w_1w_2 + w_2w_3\}.$$

La Lagrangiana associata a questo problema di minimizzazione è

$$L(w_1, w_2, w_3, \mu) = w_1^2 + w_2^2 + w_3^2 + w_1w_2 + w_2w_3 - \lambda_1(w_1 + w_2 + w_3 - 1).$$

Pertanto il problema della ricerca dei punti estremali diviene

$$\begin{aligned}
\frac{\partial L}{\partial w_1} &= 2w_1 + w_2 - \lambda_1 = 0, \\
\frac{\partial L}{\partial w_2} &= 2w_2 + w_1 + w_3 - \lambda_1 = 0, \\
\frac{\partial L}{\partial w_3} &= 2w_3 + w_2 - \lambda_1 = 0, \\
\frac{\partial L}{\partial \lambda_1} &= -(w_1 + w_2 + w_3 - 1) = 0.
\end{aligned}$$

La soluzione di tale sistema lineare è chiaramente

$$w_1 = 0.5, \quad w_2 = 0, \quad w_3 = 0.5, \quad \lambda_1 = 1. \tag{2.113}$$

I pesi w_1, w_2, w_3 così determinati individuano il portafoglio di minimo rischio assoluto. Questo portafoglio è caratterizzato dal tasso di rendimento

$$\bar{r}_{\min} = w_1\bar{r}_1 + w_2\bar{r}_2 + w_2\bar{r}_2 = 0.6,$$

ed ha varianza

$$\sigma^2_{\min} = 0.5.$$

Esempio 125 *Con riferimento all'Esempio 124, consideriamo il sistema delle prime tre equazioni che caratterizzano il problema della ricerca dei punti estremali*

$$\begin{aligned}
\frac{\partial L}{\partial w_1} &= 2w_1 + w_2 - \lambda_1 - 0.4\lambda_2 = 0, \\
\frac{\partial L}{\partial w_2} &= 2w_2 + w_1 + w_3 - \lambda_1 - 0.8\lambda_2 = 0, \\
\frac{\partial L}{\partial w_3} &= 2w_3 + w_2 - \lambda_1 - 0.8\lambda_2 = 0.
\end{aligned} \tag{2.114}$$

Posto $\lambda_1 = 1$, $\lambda_2 = 0$ il Sistema (2.114) diviene

$$\begin{aligned}
\frac{\partial L}{\partial w_1} &= 2w_1 + w_2 - 1 = 0, \\
\frac{\partial L}{\partial w_2} &= 2w_2 + w_1 + w_3 - 1 = 0, \\
\frac{\partial L}{\partial w_3} &= 2w_3 + w_2 - 1 = 0,
\end{aligned}$$

e ha soluzione

$$w_1^{(1)} = 0.5, \qquad w_2^{(1)} = 0, \qquad w_3^{(1)} = 0.5. \tag{2.115}$$

Posto $\lambda_1 = 0$, $\lambda_2 = 1$ il Sistema (2.114) diviene

$$\begin{aligned}
\frac{\partial L}{\partial w_1} &= 2w_1 + w_2 - 0.4 = 0, \\
\frac{\partial L}{\partial w_2} &= 2w_2 + w_1 + w_3 - 0.8 = 0, \\
\frac{\partial L}{\partial w_3} &= 2w_3 + w_2 - 0.8 = 0,
\end{aligned}$$

e ha soluzione

$$w_1^{(2)} = 0.1, \qquad w_2^{(2)} = 0.2, \qquad w_3^{(2)} = 0.3. \tag{2.116}$$

Mediante le (2.107), (2.108) alle soluzioni (2.115) e (2.116) così determinate corrispondono rispettivamente le soluzioni

$$\hat{w}_1^{(1)} = 1/2, \qquad \hat{w}_2^{(1)} = 0, \qquad \hat{w}_3^{(1)} = 1/2, \qquad \lambda_1 = 1, \qquad \lambda_2 = 0$$

e

$$\hat{w}_1^{(2)} = 1/6, \qquad \hat{w}_2^{(2)} = 1/3, \qquad \hat{w}_3^{(2)} = 1/2, \qquad \lambda_1 = 0, \qquad \lambda_2 = 1.67$$

del Sistema (2.112). Da notare che la prima di queste due coincide con la soluzione di minimo rischio assoluto (2.113) dell'insieme dei portafogli possibili. Da notare anche che i rendimenti attesi dei due portafogli $\pi \equiv \left(\hat{w}_1^{(j)}, \hat{w}_2^{(j)}, \hat{w}_3^{(j)}\right)$, *per* $j = 1, 2$, *sono dati da*

$$\bar{r}^{(1)} = \bar{r}_1 \hat{w}_1^{(1)} + \bar{r}_2 \hat{w}_2^{(1)} + \bar{r}_3 \hat{w}_3^{(1)} = 6/10 \quad e \quad \bar{r}^{(2)} = \bar{r}_1 \hat{w}_1^{(2)} + \bar{r}_2 \hat{w}_2^{(2)} + \bar{r}_3 \hat{w}_3^{(2)} = 11/15.$$

Volendo conseguire un rendimento atteso pari a $\bar{r} \equiv 1$, *dobbiamo risolvere l'equazione*

$$\alpha \bar{r}^{(1)} + (1 - \alpha)\, \bar{r}^{(2)} = \bar{r}$$

ossia

$$6/10\alpha + 11/15\,(1 - \alpha) = 1$$

da cui

$$\alpha = -2.$$

Considerati quindi i pesi

$$w_1 = \alpha \hat{w}_1^{(1)} + (1 - \alpha)\, \hat{w}_1^{(2)} = -2 \cdot 1/2 + (1 + 2) \cdot 1/6 = -0.5,$$

$$w_2 = \alpha \hat{w}_2^{(1)} + (1 - \alpha)\, \hat{w}_2^{(2)} = -2 \cdot 0 + (1 + 2) \cdot 1/3 = 1,$$

$$w_3 = \alpha \hat{w}_3^{(1)} + (1 - \alpha)\, \hat{w}_3^{(2)} = -2 \cdot 1/2 + (1 + 2) \cdot 1/2 = 0.5,$$

questi caratterizzano effettivamente il portafoglio di minimo rischio per l'assegnato rendimento atteso $\bar{r} \equiv 1$ *(cfr Esempio 124).*

L'insieme possibile, l'insieme di minima varianza, la frontiera efficiente

Consideriamo due titoli rischiosi di tasso di rendimento atteso $\bar{r}_1$ ed $\bar{r}_2$, di deviazione standard σ_1 e σ_2 e di coefficiente di correlazione $\rho_{1,2}$. L'insieme dei portafogli di dati deviazione standard σ e rendimento $\bar{r}$ è allora caratterizzato dalla condizione

$$\begin{aligned} \sigma^2 &= \sigma_1^2 w_1^2 + \sigma_2^2 w_2^2 + 2\rho_{1,2}\sigma_1\sigma_2 w_1 w_2, \\ \bar{r} &= \bar{r}_1 w_1 + \bar{r}_2 w_2, \\ 1 &= w_1 + w_2, \end{aligned} \tag{2.117}$$

per opportuni $w_1, w_2 \in \mathbb{R}$. Ponendo $w_1 \equiv w$, $w_2 \equiv 1 - w$ è possibile eliminare un parametro dalle equazioni (2.117) e riscriverle nella forma

$$\begin{aligned} \bar{r} &= (\bar{r}_1 - \bar{r}_2)w + \bar{r}_2, \\ \sigma^2 &= (\sigma_1^2 + \sigma_2^2 - 2\rho_{1,2}\sigma_1\sigma_2)w^2 - 2\sigma_2(\sigma_2 - \rho_{1,2}\sigma_1)w + \sigma_2^2, \end{aligned} \tag{2.118}$$

per un opportuno $w \in \mathbb{R}$. Dalla prima, assumendo $\bar{r}_1 - \bar{r}_2 \neq 0$, abbiamo

$$w = \frac{\bar{r} - \bar{r}_2}{\bar{r}_1 - \bar{r}_2},$$

e sostituendo quest'ultima nella seconda, otteniamo

$$\sigma^2 = (\sigma_1^2 + \sigma_2^2 - 2\rho_{1,2}\sigma_1\sigma_2)\frac{(\bar{r} - \bar{r}_2)^2}{(\bar{r}_1 - \bar{r}_2)^2} - 2\sigma_2(\sigma_2 - \rho_{1,2}\sigma_1)\frac{\bar{r} - \bar{r}_2}{\bar{r}_1 - \bar{r}_2} + \sigma_2^2.$$

Eliminando il denominatore e riordinando i termini, segue allora l'equazione cartesiana

$$(\bar{r}_1 - \bar{r}_2)^2\sigma^2 = (\bar{r} - \bar{r}_2)^2\sigma_1^2 + (\bar{r} - \bar{r}_1)^2\sigma_2^2 - 2(\bar{r} - \bar{r}_2)(\bar{r} - \bar{r}_1)\rho_{1,2}\sigma_1\sigma_2, \tag{2.119}$$

che è chiaramente l'equazione di una conica del piano $\sigma, \bar{r}$. La matrice dei coefficienti associata a tale conica è data da

$$A \equiv \begin{pmatrix} a_{1,1} & 0 & 0 \\ 0 & a_{2,2} & a_{2,3} \\ 0 & a_{2,3} & a_{3,3} \end{pmatrix},$$

dove

$$\begin{aligned} a_{1,1} &\equiv (\bar{r}_1 - \bar{r}_2)^2, \\ a_{2,2} &\equiv -(\sigma_1^2 + \sigma_2^2 - 2\rho_{1,2}\sigma_1\sigma_2), \\ a_{2,3} &\equiv \bar{r}_1\sigma_2(\sigma_2 - \rho_{1,2}\sigma_1) + \bar{r}_2\sigma_1(\sigma_1 - \rho_{1,2}\sigma_2), \\ a_{3,3} &\equiv -(\bar{r}_1^2\sigma_2^2 + \bar{r}_2^2\sigma_1^2 - 2\bar{r}_1\bar{r}_2\rho_{1,2}\sigma_1\sigma_2), \end{aligned}$$

e risulta

$$\det(A) = (\bar{r}_1 - \bar{r}_2)^4(1 - \rho_{1,2}^2)\sigma_1^2\sigma_2^2.$$

Pertanto, sempre nell'ipotesi $\bar{r}_1 - \bar{r}_2 \neq 0$ sopra considerata, la conica degenera se

$$\rho_{1,2} = \pm 1, \quad \text{o} \quad \sigma_1 = 0, \quad \text{o} \quad \sigma_2 = 0.$$

Nei casi non degeneri, risulta

$$\det(A_{3,3}) = -(\bar{r}_1 - \bar{r}_2)^2(\sigma_1^2 + \sigma_2^2 - 2\rho_{1,2}\sigma_1\sigma_2),$$

essendo $A_{3,3}$ la sottomatrice estratta da A sopprimendone la terza riga e la terza colonna, e si può provare che in questi casi si ha sicuramente

$$\sigma_1^2 + \sigma_2^2 - 2\rho_{1,2}\sigma_1\sigma_2 > 0. \tag{2.120}$$

Infatti, dalla disuguaglianza

$$(|\sigma_1| - |\sigma_2|)^2 \geqslant 0,$$

sempre vera per qualsiasi scelta di $\sigma_1, \sigma_2 \in \mathbb{R}$, segue immediatamente

$$\sigma_1^2 + \sigma_2^2 \geqslant 2\,|\sigma_1\sigma_2|\,. \tag{2.121}$$

D'altra parte, essendo nei casi di non degenerazione

$$|\rho_{1,2}| < 1,$$

risulta chiaramente

$$2\,|\sigma_1\sigma_2| > 2\,|\rho_{1,2}|\,|\sigma_1\sigma_2| \geqslant 2\rho_{1,2}\sigma_1\sigma_2. \tag{2.122}$$

Allora la (2.120) si ottiene combinando la (2.121) con la (2.122).

In definitiva, nei casi di non degenerazione, la conica risulta necessariamente essere una iperbole passante per i punti $P_1 \equiv (\sigma_1, \bar{r}_1)$ e $P_2 \equiv (\sigma_2, \bar{r}_2)$. Tale iperbole ha assi d'equazione

$$\sigma = 0, \quad \bar{r} = \frac{\bar{r}_1\sigma_2(\sigma_2 - \rho_{1,2}\sigma_1) + \bar{r}_2\sigma_1(\sigma_1 - \rho_{1,2}\sigma_2)}{\sigma_1^2 + \sigma_2^2 - 2\rho_{1,2}\sigma_1\sigma_2},$$

centro $C \equiv (c_1, c_2)$, dove

$$c_1 = 0, \quad c_2 \equiv \frac{\bar{r}_1\sigma_2(\sigma_2 - \rho_{1,2}\sigma_1) + \bar{r}_2\sigma_1(\sigma_1 - \rho_{1,2}\sigma_2)}{\sigma_1^2 + \sigma_2^2 - 2\rho_{1,2}\sigma_1\sigma_2},$$

vertice $V \equiv (v_1, v_2)$, dove

$$v_1 \equiv \frac{\sigma_1\sigma_2\sqrt{(\sigma_2 - \rho_{1,2}\sigma_1)^2 + (\sigma_1 - \rho_{1,2}\sigma_2)^2 - 2\rho_{1,2}(\sigma_2 - \rho_{1,2}\sigma_1)(\sigma_1 - \rho_{1,2}\sigma_2)}}{\sigma_1^2 + \sigma_2^2 - 2\rho_{1,2}\sigma_1\sigma_2},$$
$$v_2 \equiv \frac{\bar{r}_1\sigma_2(\sigma_2 - \rho_{1,2}\sigma_1) + \bar{r}_2\sigma_1(\sigma_1 - \rho_{1,2}\sigma_2)}{\sigma_1^2 + \sigma_2^2 - 2\rho_{1,2}\sigma_1\sigma_2},$$

ed asintoti d'equazione

$$\bar{r} = \pm\frac{\bar{r}_1 - \bar{r}_2}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho_{1,2}\sigma_1\sigma_2}}\sigma + \frac{\bar{r}_1\sigma_2(\sigma_2 - \rho_{1,2}\sigma_1) + \bar{r}_2\sigma_1(\sigma_1 - \rho_{1,2}\sigma_2)}{\sigma_1^2 + \sigma_2^2 - 2\rho_{1,2}\sigma_1\sigma_2}.$$

In definitiva, nel caso di due titoli caratterizzati dai punti $P_1 \equiv (\sigma_1, \bar{r}_1)$ e $P_2 \equiv (\sigma_2, \bar{r}_2)$ del piano $\sigma, \bar{r}$, sotto la condizione di non degenerazione $\bar{r}_1 \neq \bar{r}_2$, l'insieme dei portafogli possibili è costituito da un ramo d'iperbole passante per P_1 e P_2.

Esempio 126 *Consideriamo due titoli di tassi di rendimento*

$$r_1 \equiv \begin{cases} 3, & \mathbf{P}(r_1 = 3) = 1/4 \equiv p_1^+ \\ 1, & \mathbf{P}(r_1 = 1) = 3/4 \equiv p_1^- \end{cases} \qquad r_2 \equiv \begin{cases} 3/2, & \mathbf{P}(r_2 = 3/2) = 1/2 \equiv p_2^+ \\ 1, & \mathbf{P}(r_2 = 1) = 1/2 \equiv p_2^- \end{cases}$$

per i quali si ha

$$\bar{r}_1 \equiv \mathbf{E}[r_1] = \frac{3}{2}, \qquad \bar{r}_2 \equiv \mathbf{E}[r_2] = \frac{5}{4},$$
$$\sigma_1^2 \equiv \mathbf{D}^2[r_1] = \frac{3}{4}, \qquad \sigma_2^2 \equiv \mathbf{D}^2[r_2] = \frac{1}{16}.$$

Notiamo che dalla ben nota relazione

$$\sigma_{1,2} \equiv Cov(r_1, r_2) = \mathbf{E}[(r_1 - \mathbf{E}[r_1])(r_2 - \mathbf{E}[r_2])] = \mathbf{E}[r_1r_2] - \mathbf{E}[r_1]\mathbf{E}[r_2]$$

per calcolare $Cov(r_1, r_2)$ è necessario e sufficiente conoscere il termine $\mathbf{E}[r_1r_2]$ ed a tale scopo occorre conoscere la distribuzione di probabilità congiunta del vettore aleatorio (r_1, r_2). Infatti, supposto che sia

$$\mathbf{P}(r_1 = 3, r_2 = 3/2) \equiv p^{++}, \qquad \mathbf{P}(r_1 = 3, r_2 = 1) \equiv p^{+-},$$
$$\mathbf{P}(r_1 = 1, r_2 = 3/2) \equiv p^{-+}, \qquad \mathbf{P}(r_1 = 1, r_2 = 1) \equiv p^{--},$$

con $p^{++}, p^{+-}, p^{-+}, p^{--} \geqslant 0$ tali che $p^{++} + p^{+-} + p^{-+} + p^{--} = 1$, abbiamo

$$r_1r_2 = \begin{cases} 9/2, & \mathbf{P}(r_1r_2 = 9/2) = \mathbf{P}(r_1 = 3, r_2 = 3/2) = p^{++} \\ 3, & \mathbf{P}(r_1r_2 = 3) = \mathbf{P}(r_1 = 3, r_2 = 1) = p^{+-} \\ 3/2, & \mathbf{P}(r_1r_2 = 3/2) = \mathbf{P}(r_1 = 1, r_2 = 3/2) = p^{-+} \\ 1, & \mathbf{P}(r_1r_2 = 1) = \mathbf{P}(r_1 = 1, r_2 = 1) = p^{--} \end{cases}$$

e di conseguenza

$$\mathbf{E}[r_1r_2] = \frac{9}{2}p^{++} + 3p^{+-} + \frac{3}{2}p^{-+} + p^{--}.$$

Adesso, tenendo conto che deve aversi

$$\begin{aligned}
p^{++}+p^{+-} &= \mathbf{P}\left(r_1=3, r_2=3/2\right)+\mathbf{P}\left(r_1=3, r_2=1\right)=\mathbf{P}\left(r_1=3\right)=1/4, \\
p^{-+}+p^{--} &= \mathbf{P}\left(r_1=1, r_2=3/2\right)+\mathbf{P}\left(r_1=1, r_2=1\right)=\mathbf{P}\left(r_1=1\right)=3/4, \\
p^{++}+p^{-+} &= \mathbf{P}\left(r_1=3, r_2=3/2\right)+\mathbf{P}\left(r_1=1, r_2=3/2\right)=\mathbf{P}\left(r_2=3/2\right)=1/2, \\
p^{+-}+p^{--} &= \mathbf{P}\left(r_1=3, r_2=1\right)+\mathbf{P}\left(r_1=1, r_2=1\right)=\mathbf{P}\left(r_2=1\right)=1/2,
\end{aligned}$$

la distribuzione di probabilità assegnata deve soddisfare il sistema di equazioni

$$\begin{aligned}
p^{++}+p^{+-} &= 1/4, \\
p^{-+}+p^{--} &= 3/4, \\
p^{++}+p^{-+} &= 1/2, \\
p^{+-}+p^{--} &= 1/2,
\end{aligned}$$

essendo l'ulteriore

$$p^{++}+p^{+-}+p^{-+}+p^{--}=1$$

automaticamente soddisfatta (considerando, ad esempio, la somma delle prime due). Dalla prima e terza equazione, otteniamo

$$\begin{aligned}
p^{+-} &= 1/4-p^{++}, \\
p^{-+} &= 1/2-p^{++}.
\end{aligned}$$

Combinando la prima di queste ultime con la seconda del sistema originario, otteniamo allora

$$p^{--}=1/4+p^{++}$$

La quarta equazione del sistema originario risulta quindi ridursi a una identità. E' allora chiaro che assegnato arbitrariamente p^{++} nell'intervallo $(0,1)$, o equivalentemente uno qualsiasi tra tra $p^{++}, p^{+-}, p^{-+}, p^{--}$, la distribuzione di probabilità risulta essere conseguentemente determinati. Per esempio, assegnando $p^{++}=1/4$, otteniamo

$$p^{++}=\frac{1}{4}, \qquad p^{+-}=0, \qquad p^{-+}=\frac{1}{4}, \qquad p^{--}=\frac{1}{2}.$$

In questo caso, risulta

$$\mathbf{E}[r_1 r_2]=2$$

e di conseguenza

$$\rho_{1,2}\sigma_1\sigma_2 \equiv \sigma_{1,2}=\frac{1}{8}.$$

Notare che

$$\rho_{1,2}=\frac{\frac{1}{8}}{\frac{\sqrt{3}}{2}\frac{1}{4}}=\frac{\sqrt{3}}{3}.$$

Notare anche che se r_1 ed r_2 fossero indipendenti, avremmo

$$\begin{aligned}
\mathbf{P}\left(r_1=3, r_2=3/2\right) \equiv p^{++} &= \mathbf{P}\left(r_1=3\right)\mathbf{P}\left(r_2=3/2\right)=\frac{1}{8}, \\
\mathbf{P}\left(r_1=3, r_2=1\right) \equiv p^{+-} &= \mathbf{P}\left(r_1=3\right)\mathbf{P}\left(r_2=1\right)=\frac{1}{8}, \\
\mathbf{P}\left(r_1=1, r_2=3/2\right) \equiv p^{-+} &= \mathbf{P}\left(r_1=1\right)\mathbf{P}\left(r_2=3/2\right)=\frac{3}{8}, \\
\mathbf{P}\left(r_1=1, r_2=1\right) \equiv p^{--} &= \mathbf{P}\left(r_1=1\right)\mathbf{P}\left(r_2=1\right)=\frac{3}{8}.
\end{aligned}$$

essendo

$$p^{++} = \frac{1}{8}, \qquad p^{+-} = \frac{1}{8}, \qquad p^{-+} = \frac{3}{8}, \qquad p^{--} = \frac{3}{8},$$

un'ulteriore soluzione del sistema di equazioni sopra considerato. In questo caso, ovviamente,

$$\mathbf{E}[r_1 r_2] = \frac{15}{8}$$

ed ancora

$$\rho_{1,2}\sigma_1\sigma_2 \equiv \sigma_{1,2} = 0$$

Nel caso di titoli a rendimenti non indipendenti, consideriamo adesso la forma analitica dell'equazione rischio-rendimento associata ad un portafoglio costituito da quantità opportune di tali titoli. Ponendo $w_1 \equiv w$, $w_2 \equiv 1 - w$ le equazioni parametriche (2.118) per un tale portafoglio si riducono a

$$\begin{aligned} \bar{r} &= \frac{1}{4}w + \frac{5}{4} \\ \sigma^2 &= \frac{9}{16}w^2 + \frac{1}{8}w + \frac{1}{16}. \end{aligned}$$

Dalla prima equazione ricaviamo

$$w = 4\bar{r} - 5.$$

e sostituendo nella seconda equazione quanto ricavato dalla prima otteniamo

$$2\sigma^2 - 18\bar{r}^2 + 44\bar{r} - 27 = 0.$$

Si vede allora che l'equazione rischio-rendimento descrive un'iperbole del piano $\sigma, \bar{r}$. La matrice dei coefficienti associata è data da

$$A \equiv \begin{pmatrix} 2 & 0 & 0 \\ 0 & -18 & 22 \\ 0 & 22 & -27 \end{pmatrix}.$$

Gli assi sono le rette di equazione

$$\sigma = 0, \qquad \bar{r} = 11/9.$$

il suo centro è il punto

$$C \equiv (0, 11/9),$$

il vertice è il punto

$$V \equiv (\sqrt{2}/6, 11/9),$$

ed i suoi asintoti sono le rette di equazione

$$\bar{r} = \pm\frac{1}{3}\sigma + 11/9.$$

Notare che il portafoglio di rischio minimo, ha rischio pari a all'ascissa del vertice dell'iperbole, ossia $\sqrt{2}/6$. Corrispondentemente il suo rendimento è pari all'ordinata del vertice dell'iperbole, ossia $11/9$. Si consegue tale prortafoglio con la composizione $w_1 \equiv w$, $w_2 \equiv 1 - w$ dove

$$w = 4\bar{r} - 5 = -1/9.$$

Ciò significa che va venduto allo scoperto il titolo di rendimento atteso $\bar{r}_1$ e va acquistato il titolo di rendimento $\bar{r}_2$.

Nel caso di titoli a rendimenti indipendenti, consideriamo ancora la forma analitica dell'equazione rischio-rendimento associata ad un portafoglio costituito da quantità opportune di tali titoli. Ponendo $w_1 \equiv w$, $w_2 \equiv 1 - w$ *le equazioni parametriche (2.117) per un tale portafoglio si riducono a*

$$\bar{r} = \frac{1}{4}w + \frac{5}{4}$$
$$\sigma^2 = \frac{13}{16}w^2 - \frac{1}{8}w + \frac{1}{16}.$$

Dalla prima equazione ricaviamo ancora

$$w = 4\bar{r} - 5.$$

e sostituendo nella seconda equazione quanto ricavato dalla prima

$$\sigma^2 - 13r^2 + 33r - 21 = 0.$$

Si vede allora che l'equazione rischio-rendimento descrive un'iperbole del piano $\sigma, \bar{r}$*. La matrice dei coefficienti associata è data da*

$$A \equiv \begin{pmatrix} 1 & 0 & 0 \\ 0 & -13 & 33/2 \\ 0 & 33/2 & -21 \end{pmatrix}.$$

Gli assi sono le rette di equazione

$$\sigma = 0, \qquad \bar{r} = 33/26.$$

il suo centro è il punto

$$C \equiv (0, 33/26),$$

il vertice è il punto

$$V \equiv (13\sqrt{3}/26, 33/26),$$

ed i suoi asintoti sono le rette di equazione

$$\bar{r} = \pm\sqrt{13}/13\sigma + 33/26.$$

Notare che il portafoglio di rischio minimo, ha rischio pari a all'ascissa del vertice dell'iperbole, ossia $13\sqrt{3}/26$*. Corrispondentemente il suo rendimento è pari all'ordinata del vertice dell'iperbole, ossia* $33/26$*. Si consegue tale prortafoglio con la composizione* $w_1 \equiv w$, $w_2 \equiv 1 - w$ *dove*

$$w = 4\bar{r} - 5 = 1/13.$$

Ciò significa che vanno acquistati entrambi i titoli.

Non è difficile rendersi conto che nel caso di tre titoli caratterizzati dai punti $P_1 \equiv (\sigma_1, \bar{r}_1)$, $P_2 \equiv (\sigma_2, \bar{r}_2)$ e $P_3 \equiv (\sigma_3, \bar{r}_3)$ del piano $\sigma - \bar{r}$ l'insieme dei portafogli possibili sarà una superficie del piano $\sigma, \bar{r}$. Infatti considerati i titoli a due a due, in ciascuno dei tre casi che si possono presentare, l'insieme dei portafogli possibili è individuato da un ramo d'iperbole che congiunge i due punti rappresentanti i titoli considerati. Per esempio se consideriamo P_1 e P_2 l'insieme dei portafogli possibili è individuato dal ramo d'iperbole che congiunge P_1 e P_2. D'altra parte ogni portafoglio possibile generato dai tre titoli può essere visto come un portafoglio generato da un portafoglio intermedio di due qualsiasi di questi titoli, rappresentato quindi da un punto situato sul ramo d'iperbole da essi individuata, in combinazione con il terzo titolo. Pertanto tale portafoglio possibile è rappresentato da un punto situato sul ramo d'iperbole che congiunge il punto del piano $\sigma, \bar{r}$ rappresentante il portafoglio intermedio e dal punto rappresentante il terzo titolo. Al variare del punto rappresentante il portafoglio intermedio sul ramo d'iperbole passante per punti rappresentanti i primi due titoli, le iperboli che li congiungono vengono ad individuare un'intera superficie del piano $\sigma, \bar{r}$.

Vincoli di non negatività

Qualora non sia ammessa la vendita allo scoperto dei titoli da allocare in portafoglio bisogna imporre l'ulteriore condizione

$$w_m \geqslant 0, \qquad \forall m = 1, \dots, M.$$

e il problema della determinazione del portafoglio di minimo rischio che realizzi un tasso di rendimento atteso assegnato si traduce in termini matematici nel problema di ottimizzazione vincolata

$$\min_{(w_1,\dots,w_M)\in\mathbb{H}\cap\mathbb{K}_{\bar{r}}\cap\mathbb{R}_+^M} \{\frac{1}{2}\sigma^2(w_1,\dots,w_M)\}. \tag{2.123}$$

Anche in questo caso, la convessità dell'insieme $\mathbb{H}\cap\mathbb{K}_{\bar{r}}\cap\mathbb{R}_+^M$ e della funzione obiettivo $\frac{1}{2}\sigma^2(w_1,\dots,w_M)$ consente di affermare che se l'insieme dei titoli in portafoglio è non singolare esiste un'unica soluzione del problema (2.123). Tuttavia, tecnicamente questo problema non può essere ridotto alla soluzione di un sistema di equazioni lineari. In effetti, si tratta di un problema di **programmazione quadratica**, dato che la funzione obiettivo è quadratica ed i vincoli sono sia uguaglianze che disuguaglianze lineari. Da notare che, essendo chiaramente $\mathbb{H}\cap\mathbb{K}_{\bar{r}}\cap\mathbb{R}_+^M \subseteq \mathbb{H}\cap\mathbb{K}_{\bar{r}}$, risulta

$$\min_{(w_1,\dots,w_M)\in\mathbb{H}\cap\mathbb{K}_{\bar{r}}\cap\mathbb{R}_+^M} \{\frac{1}{2}\sigma^2(w_1,\dots,w_M)\} \geqslant \min_{(w_1,\dots,w_M)\in\mathbb{H}\cap\mathbb{K}_{\bar{r}}} \{\frac{1}{2}\sigma^2(w_1,\dots,w_M)\}. \tag{2.124}$$

Pertanto, se $(\breve{w}_1,\dots,\breve{w}_M) \equiv \arg\min\{\frac{1}{2}\sigma^2(w_1,\dots,w_M),\ (w_1,\dots,w_M)\in\mathbb{H}\cap\mathbb{K}_{\bar{r}}\}$ è tale che $(\breve{w}_1,\dots,\breve{w}_M)\in\mathbb{R}_+^M$, ossia $(\breve{w}_1,\dots,\breve{w}_M)\in\mathbb{H}\cap\mathbb{K}_{\bar{r}}\cap\mathbb{R}_+^M$, allora l'esistenza e unicità della soluzione del problema (2.123) e l'Equazione (2.124) comportano che $(\breve{w}_1,\dots,\breve{w}_M) \equiv \arg\min\{\frac{1}{2}\sigma^2(w_1,\dots,w_M),\ (w_1,\dots,w_M)\in\mathbb{H}\cap\mathbb{K}_{\bar{r}}\cap\mathbb{R}_+^M\}$.

Esempio 127 *Con riferimento all'Esempio 124, assegnato un tasso di rendimento atteso $\bar{r}$, consideriamo il problema della determinazione del portafoglio di minimo rischio che realizzi il tasso assegnato, escludendo la possibilità di vendita allo scoperto dei titoli in portafoglio. Tale portafoglio è caratterizzato dalla condizione*

$$\min_{\substack{w_1+w_2+w_3=1\\ 0.4w_1+0.8w_2+0.8w_3=\bar{r}\\ w_m\geqslant 0,\quad m=1,2,3}} \{w_1^2+w_2^2+w_3^2+w_1w_2+w_2w_3\}, \tag{2.125}$$

e dalla discussione dell'Esempio 124 si vede che il portafoglio caratterizzato dai pesi

$$w_1 = 2.0 - 2.5\bar{r}, \quad w_2 = 2.5\bar{r} - 1.5, \quad w_3 = 0.5$$

è soluzione di anche di questo problema pur di assumere

$$0.6 \leqslant \bar{r} \leqslant 0.8. \tag{2.126}$$

Infatti, in questo caso risulta

$$w_1 \geqslant 0, \qquad w_2 \geqslant 0.$$

In particolare, il portafoglio di minimo rischio relativo a un rendimento atteso $\bar{r} = 0.7$ è caratterizzato dai pesi

$$w_1 = 0.25, \quad w_2 = 0.25, \quad w_3 = 0.5.$$

Se la condizione (2.126) non è soddisfatta, per una trattazione esaustiva del problema bisogna applicare il metodo di Karush-Kuhn-Tucker.

Supponiamo che il rendimento atteso sia $\bar{r} = 0.5$. In questo caso i pesi

$$w_1 = 0.75, \quad w_2 = -0.25, \quad w_3 = 0.5$$

che costituiscono una soluzione del Problema (2.111) non sono evidentemente una soluzione del Problema (2.125). In riferimento al Teorema (283) della Sezione Constrained Optimization dell'Appendice, per trovare una soluzione del Problema (2.125) consideriamo la Lagrangiana

$$\begin{aligned} L(w_1, w_2, w_3, \lambda_1, \lambda_2, \mu_1, \mu_2, \mu_3) \equiv\, & w_1^2 + w_2^2 + w_3^2 + w_1 w_2 + w_2 w_3 \\ & - \lambda_1(w_1 + w_2 + w_3 - 1) - \lambda_2(0.4w_1 + 0.8w_2 + 0.8w_3 - 0.5) \\ & - \mu_1 w_1 - \mu_2 w_2 - \mu_3 w_3. \end{aligned}$$

Le condizioni da imporre per la ricerca degli eventuali punti di minimo sono allora

$$\begin{aligned} \frac{\partial L}{\partial w_1} &= 2w_1 + w_2 - \lambda_1 - 0.4\lambda_2 - \mu_1 = 0, \\ \frac{\partial L}{\partial w_2} &= 2w_2 + w_1 + w_3 - \lambda_1 - 0.8\lambda_2 - \mu_2 = 0, \\ \frac{\partial L}{\partial w_3} &= 2w_3 + w_2 - \lambda_1 - 0.8\lambda_2 - \mu_3 = 0, \end{aligned}$$

per quanto riguarda le condizioni d'ottimalità,

$$\begin{aligned} g_1(w_1, w_2, w_3) &\equiv w_1 + w_2 + w_3 - 1 = 0, \\ g_2(w_1, w_2, w_3) &\equiv 0.4w_1 + 0.8w_2 + 0.8w_3 - 0.5 = 0, \\ h_1(w_1, w_2, w_3) &\equiv w_1 \geqslant 0, \\ h_2(w_1, w_2, w_3) &\equiv w_2 \geqslant 0, \\ h_3(w_1, w_2, w_3) &\equiv w_3 \geqslant 0, \end{aligned}$$

per quanto riguarda i due gruppi di condizioni di fattibilità

$$\mu_1 \geqslant 0, \quad \mu_2 \geqslant 0, \quad \mu_3 \geqslant 0,$$

per quanto riguarda le condizioni di non negatività del moltiplicatori μ_k, per $k = 1, 2, 3$, e infine

$$\begin{aligned} \mu_1 h_1(w_1, w_2, w_3) &\equiv \mu_1 w_1 = 0, \\ \mu_2 h_2(w_1, w_2, w_3) &\equiv \mu_2 w_2 = 0, \\ \mu_3 h_3(w_1, w_2, w_3) &\equiv \mu_3 w_3 = 0, \end{aligned}$$

per quanto riguarda le condizioni di rilassatezza.

La risoluzione del sistema di equazioni-disequazioni che risulta dal complesso di tutte le condizioni non è particolarmente difficile anche se può risultare notevolmente lunga. Un modo sistematico di procedere può essere il seguente: in riferimento alla condizione di non negatività dei moltiplicatori μ_k, per $k = 1, 2, 3$, possono distinguersi $2^3 = 8$ casi

$$\begin{aligned} &1) \quad \mu_1 > 0, \quad \mu_2 > 0, \quad \mu_3 > 0, \\ &2) \quad \mu_1 > 0, \quad \mu_2 > 0, \quad \mu_3 = 0, \\ &3) \quad \mu_1 > 0, \quad \mu_2 = 0, \quad \mu_3 > 0, \\ &4) \quad \mu_1 > 0, \quad \mu_2 = 0, \quad \mu_3 = 0, \\ &5) \quad \mu_1 = 0, \quad \mu_2 > 0, \quad \mu_3 > 0, \\ &6) \quad \mu_1 = 0, \quad \mu_2 > 0, \quad \mu_3 = 0, \\ &7) \quad \mu_1 = 0, \quad \mu_2 = 0, \quad \mu_3 > 0, \\ &8) \quad \mu_1 = 0, \quad \mu_2 = 0, \quad \mu_3 = 0. \end{aligned}$$

In relazione al primo il sistema di tutte le altre condizioni diventa

$$\begin{aligned}
&2w_1 + w_2 - \lambda_1 - 0.4\lambda_2 - \mu_1 = 0, \\
&2w_2 + w_1 + w_3 - \lambda_1 - 0.8\lambda_2 - \mu_2 = 0, \\
&2w_3 + w_2 - \lambda_1 - 0.8\lambda_2 - \mu_3 = 0, \\
&w_1 + w_2 + w_3 - 1 = 0, \\
&0.4w_1 + 0.8w_2 + 0.8w_3 - 0.5 = 0, \\
&w_1 = 0, \\
&w_2 = 0, \\
&w_3 = 0,
\end{aligned}$$

che chiaramente non ammette soluzioni. In relazione al secondo caso il sistema di tutte le altre condizioni diventa

$$\begin{aligned}
&2w_1 + w_2 - \lambda_1 - 0.4\lambda_2 - \mu_1 = 0, \\
&2w_2 + w_1 + w_3 - \lambda_1 - 0.8\lambda_2 - \mu_2 = 0, \\
&2w_3 + w_2 - \lambda_1 - 0.8\lambda_2 = 0, \\
&w_1 + w_2 + w_3 - 1 = 0, \\
&0.4w_1 + 0.8w_2 + 0.8w_3 - 0.5 = 0, \\
&w_1 = 0, \\
&w_2 = 0,
\end{aligned}$$

e anche questo è chiaramente privo di soluzione. Procedendo sistematicamente in questo modo si vede che diversi casi non conducono ad alcuna soluzione. Tuttavia, in relazione al sesto caso il sistema di tutte le altre condizioni diventa

$$\begin{aligned}
&2w_1 + w_2 - \lambda_1 - 0.4\lambda_2 = 0, \\
&2w_2 + w_1 + w_3 - \lambda_1 - 0.8\lambda_2 - \mu_2 = 0, \\
&2w_3 + w_2 - \lambda_1 - 0.8\lambda_2 = 0, \\
&w_1 + w_2 + w_3 - 1 = 0, \\
&0.4w_1 + 0.8w_2 + 0.8w_3 - 0.5 = 0, \\
&w_2 = 0,
\end{aligned}$$

e questo ammette la soluzione

$$w_1 = 0.75, \quad w_2 = 0, \quad w_3 = 0.25, \quad \lambda_1 = 2.5, \quad \lambda_2 = -2.5, \quad \mu_2 = 0.5$$

che unitamente alla condizione

$$\mu_1 = 0, \quad \mu_2 > 0, \quad \mu_3 = 0,$$

soddisfa tutte le condizioni da imporre per la ricerca dei punti di minimo.

2.2.2 Secondo problema di configurazione di portafoglio

Il secondo problema di configurazione di portafoglio si occupa della determinazione del portafoglio di massimo tasso di rendimento che realizzi un assegnato livello di rischio σ^2. In riferimento all forma quadratica $\sigma^2 : \mathbb{R}^M \to \mathbb{R}$ (cfr. Equazione (2.89)), consideriamo l'insieme di livello

$$\mathbb{K}_{\sigma^2} \equiv \left\{ (w_1, \dots, w_M) \in \mathbb{R}^M : \sigma^2(w_1, \dots, w_M) = \sigma^2 \right\},$$

al variare di $\sigma^2 > 0$. Allora, il secondo problema di configurazione di portafoglio si traduce in termini matematici nel problema di ottimizzazione vincolata

$$\max_{(w_1,\dots,w_M)\in\mathbb{H}\cap\mathbb{K}_{\sigma^2}} \{\bar{r}(w_1, \dots, w_M)\}, \tag{2.127}$$

essendo $\bar{r} : \mathbb{R}^M \to \mathbb{R}$ la funzione tasso di rendimento atteso (cfr Equazione (2.91)). Al proposito abbiamo

Teorema 128 *Se l'insieme dei titoli in portafoglio è non singolare, l'insieme $\mathbb{K}_{\sigma^2}$ è un compatto di $\mathbb{R}^M$ per ogni $\sigma^2 > 0$.*

In conseguenza, abbiamo

Teorema 129 *Se l'insieme dei titoli azionari in portafoglio è non singolare, allora esiste almeno un portafoglio di massimo rendimento atteso per ogni assegnato livello di rischio $\sigma^2 > 0$.*

Dal punto di vista computazionale, il problema della determinazione del portafoglio di massimo tasso di rendimento che realizzi un assegnato livello di rischio può affrontarsi con la stessa tecnica di ottimizzazione vincolata adoperate per il problema simmetrico della determinazione del portafoglio di minimo rischio che realizzi un assegnato tasso di rendimento. Nondimeno, la quadraticità del vincolo non consente di ricondursi a un sistema di condizioni del primo ordine lineari e pertanto la determinazione degli zeri delle condizioni del primo ordine può risultare tecnicamente complessa.

2.2.3 Inclusione di un titolo non rischioso

Sia il problema della minimizzazione del rischio per un tasso di rendimento atteso assegnato che il problema della massimizzazione del tasso di rendimento atteso per un rischio assegnato trovano una modellizzazione molto più efficace considerando la possibilità d'introdurre in portafoglio il titolo non rischioso, già denominato *bond*. Infatti, l'introduzione del bond consente di affrontare il problema della massimizzazione del rapporto tra il tasso di rendimento atteso e il rischio di portafoglio, conducendo alla scoperta di una relazione lineare tra le posizioni ottimali di tasso di rendimento atteso e volatilità di portafoglio. Tale relazione lineare consente sia la determinazione di un portafoglio fattibile che minimizzi il rischio per un tasso di rendimento atteso assegnato che la determinazione di un portafoglio fattibile che massimizzi il tasso di rendimento atteso per un rischio assegnato.

In riferimento alla (2.75), denotiamo più brevemente con r_0 il tasso di rendimento del bond. Consideriamo quindi un portafoglio $\pi \equiv (w_1, \ldots, w_M)$ dell'insieme dei portafogli possibili caratterizzato dal tasso di rendimento atteso $\bar{r}$ e volatilità σ. L'angolo formato dalla retta condotta dal punto $(0, r_0)$ che rappresenta il titolo non rischioso al punto $(\sigma, \bar{r})$ è allora caratterizzato da

$$\tan(\theta) = \frac{\bar{r} - r_0}{\sigma}$$

Sappiamo che

$$\bar{r} = \sum_{m=1}^{M} w_m \bar{r}_m, \qquad \sigma = \left(\sum_{\ell,m=1}^{M} \sigma_{\ell,m} w_\ell w_m \right)^{1/2},$$

ma allora scrivendo, come è lecito,

$$r_0 = \sum_{m=1}^{M} w_m r_0,$$

il problema di massimizzazione può essere scritto nella forma

$$\max_{(w_1,\ldots,w_M) \in \mathbb{H} \cap \mathbb{K}_+} \tan(\theta) = \max_{(w_1,\ldots,w_M) \in \mathbb{H} \cap \mathbb{K}_+} \frac{\sum_{m=1}^{M} w_m (\bar{r}_m - r_0)}{\left(\sum_{\ell,m=1}^{M} \sigma_{\ell,m} w_\ell w_m \right)^{1/2}}, \tag{2.128}$$

dove

$$\mathbb{H} \equiv \left\{ (w_1, \ldots, w_M) \in \mathbb{R}^M : \sum_{m=1}^{M} w_m = 1 \right\} \quad \text{e} \quad \mathbb{K}_+ \equiv \left\{ (w_1, \ldots, w_M) \in \mathbb{R}^M : \sum_{m=1}^{M} w_m (\bar{r}_m - r_0) > 0 \right\}.$$

Tale problema è noto come *massimizzazione del rapporto di Sharpe*. Da notare che, mentre l'appartenenza del portafoglio π al convesso $\mathbb{H}$ di $\mathbb{R}^M$ rappresenta il vincolo tecnico di fattibilità, già introdotto, l'ulteriore appartenenza di π al convesso $\mathbb{K}_+$ di $\mathbb{R}^M$ rappresenta un vincolo di ragionevolezza finanziaria. Infatti l'assenza del vincolo $\mathbb{K}_+$ potrebbe portare alla determinazione di portafogli fattibili per i quali

$$\sum_{m=1}^{M} w_m \bar{r}_m \leqslant r_0$$

ossia aventi rendimento atteso

$$\bar{r} \leqslant r_0.$$

D'altra parte nessun investitore razionale considererebbe la possibilità d'investire in un portafoglio azionario rischioso il cui tasso di rendimento atteso sia non suoperiore al tasso di rendimento privo di rischio. Da notare inoltre che mediante l'Equazione (2.128) si cerca in effetti di determinare un portafoglio dell'insieme dei portafogli fattibili che massimizza il rapporto tra il suo rendimento atteso in eccesso, rispetto al rendimento del bond, e la sua volatilità. L'accorgimento tecnico di considerare il rendimento atteso in eccesso consente di affrontare più efficacemente il problema senza alterarlo. Valgono i seguenti Teoremi

Teorema 130 (maximization of Sharpe ratio I) *Il problema di massimizzazione (2.128) è equivalente al problema di massimizzazione*

$$\max_{(w_1,\ldots,w_M)\in\mathbb{H}_+\cap\mathbb{K}_1} \frac{1}{\left(\sum_{\ell,m=1}^{M} \sigma_{\ell,m} w_\ell w_m\right)^{1/2}}. \tag{2.129}$$

dove

$$\mathbb{H}_+ \equiv \left\{(w_1,\ldots,w_M)\in\mathbb{R}^M : \sum_{m=1}^{M} w_m > 0\right\} \quad e \quad \mathbb{K}_1 \equiv \left\{(w_1,\ldots,w_M)\in\mathbb{R}^M : \sum_{m=1}^{M} w_m(\bar{r}_m - r_0) = 1\right\}.$$

A sua volta il problema di massimizzazione (2.129) è equivalente al problema di minimizzazione

$$\min_{(w_1,\ldots,w_M)\in\mathbb{H}_+\cap\mathbb{K}_1} \sum_{\ell,m=1}^{M} \sigma_{\ell,m} w_\ell w_m. \tag{2.130}$$

Proof. Sia $(w_1^*,\ldots,w_M^*) \in \mathbb{H} \cap \mathbb{K}_+$. Allora posto

$$W^* \equiv \sum_{m=1}^{M} w_m^*(\bar{r}_m - r_0)$$

la M-pla $(w_1^\circ,\ldots,w_M^\circ)$ data da

$$w_m^\circ \stackrel{\text{def}}{=} \frac{w_m^*}{W^*}, \quad \forall m = 1,\ldots,M,$$

è tale che

$$\sum_{m=1}^{M} w_m^\circ = \frac{1}{W^*}\sum_{m=1}^{M} w_m^* > 0 \quad \text{e} \quad \sum_{m=1}^{M} w_m^\circ(\bar{r}_m - r_0) = 1.$$

Quindi

$$(w_1^\circ,\ldots,w_M^\circ) \in \mathbb{H}_+ \cap \mathbb{K}_1.$$

Inoltre,

$$\frac{1}{\left(\sum_{\ell,m=1}^{M}\sigma_{\ell,m}w_\ell^\circ w_m^\circ\right)^{1/2}} = \frac{1}{\left(\sum_{\ell,m=1}^{M}\sigma_{\ell,m}\frac{w_\ell^*}{W^*}\frac{w_m^*}{W^*}\right)^{1/2}} = \frac{\sum_{m=1}^{M}w_m^*(\bar{r}_m - r_0)}{\left(\sum_{\ell,m=1}^{M}\sigma_{\ell,m}w_\ell^* w_m^*\right)^{1/2}}.$$

Viceversa, sia $(w_1^\circ, \ldots, w_M^\circ) \in \mathbb{H}_+ \cap \mathbb{K}_1$ allora posto

$$W^\circ \equiv \sum_{m=1}^{M} w_m^\circ$$

la M-pla $(w_1^*, \ldots, w_M^*)$ data da

$$w_m^* \stackrel{\text{def}}{=} \frac{w_m^\circ}{W^\circ}, \quad \forall m = 1, \ldots, M,$$

è tale che

$$\sum_{m=1}^{M} w_m^* = 1 \quad \text{e} \quad \sum_{m=1}^{M} w_m^*(\bar{r}_m - r_0) = \frac{1}{W^\circ}\sum_{m=1}^{M} w_m^\circ(\bar{r}_m - r_0) = \frac{1}{W^\circ} > 0.$$

Quindi

$$(w_1^*, \ldots, w_M^*) \in \mathbb{H} \cap \mathbb{K}_+$$

Inoltre,

$$\frac{\sum_{m=1}^{M}w_m^*(\bar{r}_m - r_0)}{\left(\sum_{\ell,m=1}^{M}\sigma_{\ell,m}w_\ell^* w_m^*\right)^{1/2}} = \frac{\frac{1}{W^\circ}}{\left(\sum_{\ell,m=1}^{M}\sigma_{\ell,m}\frac{w_\ell^\circ}{W^\circ}\frac{w_m^\circ}{W^\circ}\right)^{1/2}} = \frac{1}{\left(\sum_{\ell,m=1}^{M}\sigma_{\ell,m}w_\ell^\circ w_m^\circ\right)^{1/2}}.$$

□

Alla luce del Teorema 130, il problema di massimizzazione (2.128) è, in ultima analisi, ricondotto al problema di minimizzazione (2.130). Quest'ultimo, nell'ipotesi che l'insieme dei titoli azionari sia non singolare, è un problema standard di minimizzazione di una funzione strettamente convessa su un convesso che ammette un'unica soluzione.

Teorema 131 (maximization of Sharpe ratio II) *Determinata una soluzione $v_1^*, \ldots, v_M^*$ del sistema lineare*

$$\sum_{\ell=1}^{M}\sigma_{\ell,m}v_\ell = \bar{r}_m - r_0, \qquad m = 1, \ldots, M.$$

tale che $\sum_{m=1}^{M} v_m^* > 0$, il portafoglio $\pi^{(T)} \equiv \left(w_1^{(T)}, \ldots, w_M^{(T)}\right)$ definito ponendo

$$w_m^{(T)} = \frac{v_m^*}{V^*}, \qquad m = 1, \ldots, M,$$

dove $V^* \equiv \sum_{m=1}^{M} v_m^*$, è una soluzione del problema (2.128).

Proof. *Cominciamo con l'osservare che, data la particolare forma della funzione da massimizzare, in questo caso non è necessario affrontare direttamente il problema di massimizzazione vincolata. Infatti, denotata con* $(w_1^*, \ldots, w_M^*)$ *una soluzione del problema di massimizzazione*

$$\max_{(w_1,\ldots,w_M)\in\mathbb{H}_+} \frac{\sum_{m=1}^{M} w_m(\bar{r}_m - r_0)}{\left(\sum_{\ell,m=1}^{M} \sigma_{\ell,m} w_\ell w_m\right)^{1/2}} \tag{2.131}$$

dove

$$\mathbb{H}_+ \equiv \left\{(w_1, \ldots, w_M) \in \mathbb{R}^M : \sum_{m=1}^{M} w_m > 0\right\}$$

è un sottoinsieme aperto di $\mathbb{R}^M$ *e posto*

$$W^* \equiv \sum_{m=1}^{M} w_m^*$$

la M*-pla* $(\hat{w}_1, \ldots \hat{w}_M)$ *data da*

$$\hat{w}_m \equiv \frac{w_m^*}{W^*}, \quad \forall m = 1, \ldots, M, \tag{2.132}$$

è una soluzione del problema di massimizzazione (2.128). Ciò perchè, stante la posizione (2.132), risulta

$$\frac{\sum_{m=1}^{M} \hat{w}_m(\bar{r}_m - r_0)}{\left(\sum_{\ell,m=1}^{M} \sigma_{\ell,m} \hat{w}_\ell \hat{w}_m\right)^{1/2}} = \frac{\sum_{m=1}^{M} \frac{w_m^*}{W^*}(\bar{r}_m - r_0)}{\left(\sum_{\ell,m=1}^{M} \sigma_{\ell,m} \frac{w_\ell^*}{W^*}\frac{w_m^*}{W^*}\right)^{1/2}} = \frac{\sum_{m=1}^{M} w_m^*(\bar{r}_m - r_0)}{\left(\sum_{\ell,m=1}^{M} \sigma_{\ell,m} w_\ell^* w_m^*\right)^{1/2}}.$$

e chiaramente

$$\sum_{m=1}^{M} \hat{w}_m = 1.$$

Il problema di massimizzazione vincolata (2.128) si riduce allora al problema libero (2.131), le cui soluzioni producono soluzioni del problema vincolato mediante la (2.132). D'altra parte, alla luce del Teorema (130), il problema di massimizzazione vincolata (2.128) è equivalente al problema di minimizzazione (2.130), che ammette un'unica soluzione. Quindi anche il problema libero (2.131) ammetterà un'unica soluzione. Possiamo allora determinarla mediante la semplice applicazione delle condizioni del *primo ordine qualora queste ci restiuiscano un'unica soluzione. Si ha*

$$\frac{\partial \tan(\theta)}{\partial w_m} = \frac{\partial}{\partial w_m} \frac{\sum\limits_{k=1}^{M} w_k(\bar{r}_k - r_0)}{\left(\sum\limits_{\ell,k=1}^{M} \sigma_{\ell,k} w_\ell w_k\right)^{1/2}}$$

$$= \frac{(\bar{r}_m - r_0)\left(\sum\limits_{\ell,k=1}^{M} \sigma_{\ell,k} w_\ell w_k\right)^{1/2} - \frac{1}{2}\frac{2\sum\limits_{\ell=1}^{M}\sigma_{\ell,m}w_\ell}{\left(\sum\limits_{\ell,k=1}^{M}\sigma_{\ell,k}w_\ell w_k\right)^{1/2}}\sum\limits_{k=1}^{M} w_k(\bar{r}_k - r_0)}{\sum\limits_{\ell,k=1}^{M} \sigma_{\ell,k} w_\ell w_k}$$

$$= \frac{(\bar{r}_m - r_0)\left(\sum\limits_{\ell,m=1}^{M} \sigma_{\ell,m} w_\ell w_m\right)^{1/2} - \frac{1}{2}\frac{2\sum\limits_{\ell=1}^{M}\sigma_{\ell,m}w_\ell}{\left(\sum\limits_{\ell,m=1}^{M}\sigma_{\ell,m}w_\ell w_m\right)^{1/2}}\sum\limits_{m=1}^{M} w_m(\bar{r}_m - r_0)}{\sum\limits_{\ell,m=1}^{M} \sigma_{\ell,m} w_\ell w_m},$$

ponendo

$$\bar{r} = \sum_{m=1}^{M} \bar{r}_m w_m, \qquad \sigma \equiv \left(\sum_{\ell,m=1}^{M} \sigma_{\ell,m} w_\ell w_m\right)^{1/2}, \tag{2.133}$$

possiamo allora scrivere

$$\frac{\partial \tan(\theta)}{\partial w_m} = \frac{(\bar{r}_m - r_0)\sigma - \frac{\sum\limits_{\ell=1}^{M}\sigma_{\ell,m}w_\ell}{\sigma}(\bar{r}_\pi - r_0)}{\sigma^2} = \frac{\bar{r}_m - r_0}{\sigma} - \frac{\bar{r} - r_0}{\sigma^3}\sum_{\ell=1}^{M}\sigma_{\ell,m}w_\ell.$$

Pertanto, dalle condizioni del primo ordine otteniamo

$$\sum_{\ell=1}^{M}\sigma_{\ell,m}\left(\frac{\bar{r} - r_0}{\sigma^2}\right) w_\ell = \bar{r}_m - r_0, \qquad m = 1, \ldots, M. \tag{2.134}$$

Da notare che stante la (2.133) le Equazioni (2.134) costituiscono di fatto un sistema non lineare nelle incognite $w_1, \ldots, w_M$. Infatti,

$$\frac{\bar{r} - r_0}{\sigma^2} \equiv \frac{\bar{r}\,(w_1, \ldots, w_M) - r_0}{\sigma^2\,(w_1, \ldots, w_M)}.$$

Nondimeno ponendo

$$v_m \equiv \left(\frac{\bar{r} - r_0}{\sigma^2}\right) w_m, \qquad m = 1, \ldots, M, \tag{2.135}$$

ci si riconduce al sistema lineare

$$\sum_{m=1}^{M}\sigma_{\ell,m}v_m = \bar{r}_m - r_0, \qquad m = 1, \ldots, M.$$

Trovata una soluzione $v_1^*, \ldots, v_M^*$ *di quest'ultimo tale che* $\sum_{m=1}^{M} v_m^* > 0$ *e posto*

$$w_m^{(T)} = \frac{v_m^*}{V^*}, \qquad m = 1, \ldots, M,$$

dove $V^* \equiv \sum_{m=1}^{M} v_m^*$, *si ha*

$$w_m^{(T)} = \frac{\left(\frac{\bar{r}-r_0}{\sigma^2}\right) w_m^*}{\sum_{\ell=1}^{M} \left(\frac{\bar{r}-r_0}{\sigma^2}\right) w_\ell^*} = \frac{w_m^*}{W^*},$$

dove

$$w_m^* \equiv \frac{\sigma^2}{\bar{r} - r_0} v_m^* \quad e \quad W^* \equiv \sum_{m=1}^{M} w_m^*.$$

Quindi

$$\begin{aligned}
\sum_{\ell=1}^{M} \sigma_{\ell,m} \left(\frac{\bar{r}\left(w_1^{(T)}, \ldots, w_M^{(T)}\right) - r_0}{\sigma^2 \left(w_1^{(T)}, \ldots, w_M^{(T)}\right)} \right) w_\ell^{(T)} &= \sum_{\ell=1}^{M} \sigma_{\ell,m} \frac{\sum_{m=1}^{M} w_m^{(T)} \bar{r}_m - \sum_{m=1}^{M} w_m^{(T)} r_0}{\sum_{\ell,m=1}^{M} \sigma_{\ell,m} w_\ell^{(T)} w_m^{(T)}} w_\ell^{(T)} \\
&= \sum_{\ell=1}^{M} \sigma_{\ell,m} \frac{\frac{1}{W^*}\left(\sum_{m=1}^{M} w_m^* \bar{r}_m - \sum_{m=1}^{M} w_m^* r_0 \right)}{\frac{1}{(W^*)^2} \sum_{\ell,m=1}^{M} \sigma_{\ell,m} w_\ell^* w_m^*} \frac{w_\ell^*}{W^*} \\
&= \sum_{\ell=1}^{M} \sigma_{\ell,m} \frac{\sum_{m=1}^{M} w_m^* \bar{r}_m - \sum_{m=1}^{M} w_m^* r_0}{\sum_{\ell,m=1}^{M} \sigma_{\ell,m} w_\ell^* w_m^*} w_\ell^* \\
&= \sum_{\ell=1}^{M} \sigma_{\ell,m} \frac{\bar{r}\left(w_1^*, \ldots, w_M^*\right) - r_0}{\sigma^2 \left(w_1^*, \ldots, w_M^*\right)} w_\ell^* \\
&= \sum_{\ell=1}^{M} \sigma_{\ell,m} v_\ell^* \\
&= \bar{r}_m - r_0,
\end{aligned}$$

per ogni $m = 1, \ldots, M$. *Ciò prova che il portafoglio* $\pi^{(T)} \equiv \left(w_1^{(T)}, \ldots, w_M^{(T)}\right)$ *soddisfa le condizioni del primo ordine (2.134) ed è quindi una soluzione del problema (2.128).* □

2.3 Capital Asset Pricing Model

I pesi del portafoglio determinati nell'ambito del Teorema 131 individuano il portafoglio tangente nell'insieme dei portafogli ammissibili caratterizzato da rischio e rendimento atteso rispettivamente dati da

$$\sigma_T = \left(\sum_{\ell,m=1}^{M} \sigma_{\ell,m} w_\ell^{(T)} w_m^{(T)} \right)^{1/2}, \qquad \bar{r}_T = \sum_{m=1}^{M} r_m w_m^{(T)}.$$

Tale portafoglio ha la proprietà che la retta condotta dal punto $(0, r_0)$ al punto $(\sigma_T, \bar{r}_T)$ è tangente alla frontiera efficiente dei portafogli ammissibili e pertanto ogni portafoglio appartenente a tale retta presenta un profilo di rischio-rendimento migliore del corrispondente portafoglio della frontiera efficiente caratterizzato da pari rischio o da pari rendimento.

2.3.1 Capital Market Line

Proposizione 132 *Il tasso di rendimento atteso $\bar{r}$ e la deviazione standard σ di un qualsiasi titolo (portafoglio) della* capital market line *soddisfano l'equazione*

$$\frac{\bar{r} - r_0}{\bar{r}_T - r_0} = \frac{\sigma - \sigma_0}{\sigma_T - \sigma_0} = \frac{\sigma}{\sigma_T},$$

dove $\sigma_0 \equiv 0$, ossia

$$\bar{r} = r_0 + \frac{\bar{r}_T - r_0}{\sigma_T} \sigma.$$

Il coefficiente angolare della capital market line

$$\tan(\hat{\theta}) \equiv \frac{\bar{r}_T - r_0}{\sigma_T}$$

è noto come prezzo del rischio.

Determinata l'equazione della Capital Market Line il problema della determinazione di un portafoglio di minimo rischio per un rendimento atteso assegnato che la determinazione del portafoglio di massimo rendimento atteso per un rischio assegnato trovano una semplice soluzione.

Infatti siano $\bar{r}_a$ e σ_a rispettivamente il rendimento atteso e il rischio assegnati. Si avrà intanto la relazione lineare

$$\bar{r}_a = r_0 + \frac{\bar{r}_T - r_0}{\sigma_T} \sigma_a = \left(1 - \frac{\sigma_a}{\sigma_T}\right) r_0 + \frac{\sigma_a}{\sigma_T} \bar{r}_T.$$

Per determinare la composizione del portafoglio su cui investire si considerano la M-pla di pesi $(\hat{w}_1, \ldots, \hat{w}_M)$ del portafoglio tangente e la M-pla $(\breve{w}_1, \ldots, \breve{w}_M)$ definita ponendo $\breve{w}_m = \frac{1}{M}$ per ogni $m = 1, \ldots, M$. Quindi, si costituisce il portafoglio d'investimento combinando i due portafogli secondo le proporzioni

$$\frac{\sigma_a}{\sigma_T} \quad \text{e} \quad 1 - \frac{\sigma_a}{\sigma_T}$$

rispettivamente.

2.3.2 Capital Asset Pricing Model (CAPM)

La Capital Market Line pone in relazione il rendimento atteso di un portafoglio e la sua deviazione standard, ma non mostra la relazione tra il tasso di rendimento atteso di un singolo titolo ed il corrispondente rischio. Questa relazione è espressa dal Capital Asset Pricing Model.

Teorema 133 *Il tasso di rendimento atteso $\bar{r}_m$ del titolo m-esimo in portafoglio soddisfa la relazione*

$$\bar{r}_m - r_0 = \beta_m (\bar{r}_T - r_0),$$

con

$$\beta_m = \frac{\sigma_{m,T}}{\sigma_T^2} \equiv \frac{Cov\left(r_m, r_T\right)}{\mathbf{D}^2\left[r_T\right]},$$

essendo $\sigma_{m,T} \equiv Cov\left(r_m, r_T\right)$ *la covarianza del tasso di rendimento del titolo* m*-esimo in portafoglio con il tasso di rendimento del portafoglio tangente.*

Proof. *Fissato* $\alpha \in \mathbb{R}$ *consideriamo il portafoglio costituito da una percentuale* α *investita sul titolo* m*-esimo ed una percentuale* $1-\alpha$ *investita sul portafoglio tangente. Il tasso di rendimento atteso del portafoglio così costituito è*

$$\bar{r}_\alpha = \alpha\bar{r}_m + (1-\alpha)\bar{r}_T \tag{2.136}$$

e la deviazione standard del tasso di rendimento è

$$\sigma_\alpha = \left(\alpha^2\sigma_m^2 + 2\alpha(1-\alpha)\sigma_{m,T} + (1-\alpha)^2\sigma_T^2\right)^{1/2}. \tag{2.137}$$

Al variare di α *le equazioni parametriche (2.136) e (2.137) definiscono una curva nel piano* $\sigma, \bar{r}$ *che per* $\alpha = 0$ *passa per il punto* $(\sigma_T, \bar{r}_T)$ *che individua il portafoglio tangente. Questa curva non può intersecare la Capital Market Line; se lo facesse, il portafoglio corrispondente ad un punto della curva situato al di sopra della retta violerebbe la definizione stessa di Capital Market Line come confine efficiente dell'insieme dei portafogli possibili. Pertanto nel punto* $(\sigma_T, \bar{r}_T)$ *la curva di equazioni parametriche (2.136) e (2.137) deve essere tangente alla Capital Market Line. Questa tangenza è la condizione che si sfrutta per ricavare la formula. Il coefficiente angolare della tangente alla curva di equazioni parametriche (2.136) e (2.137) nel punto* $(\sigma_T, \bar{r}_T)$ *è espresso da*

$$\begin{aligned}
\left(\frac{d\bar{r}_\alpha}{d\sigma_\alpha}\right)_{\alpha=0} &= \frac{\left(\frac{d\bar{r}_\alpha}{d\alpha}\right)_{\alpha=0}}{\left(\frac{d\sigma_\alpha}{d\alpha}\right)_{\alpha=0}} \\
&= \frac{\bar{r}_m - \bar{r}_T}{\frac{1}{2}\frac{\left(2\alpha\sigma_m^2 + 2(1-\alpha)\sigma_{m,T} - 2\alpha\sigma_{m,T} - 2(1-\alpha)\sigma_T^2\right)_{\alpha=0}}{\left(\left(\alpha^2\sigma_m^2 + 2\alpha(1-\alpha)\sigma_{m,T} + (1-\alpha)^2\sigma_T^2\right)^{1/2}\right)_{\alpha=0}}} \\
&= \frac{\bar{r}_m - \bar{r}_T}{\frac{\sigma_{m,T} - \sigma_T^2}{\sigma_T}}
\end{aligned}$$

e questo valore deve essere uguale al coefficiente angolare della Capital Market Line; ossia

$$\frac{\bar{r}_m - \bar{r}_T}{\frac{\sigma_{m,T} - \sigma_T^2}{\sigma_T}} = \frac{\bar{r}_T - r_0}{\sigma_T}.$$

Ne segue immediatamente che

$$\bar{r}_m - r_0 = \frac{\sigma_{m,T}}{\sigma_T^2}(\bar{r}_T - r_0),$$

come volevasi dimostrare. □

Definizione 134 *Il valore* β_m *è chiamato* beta *del titolo* m*-esimo. Il valore* $\bar{r}_m - r_0$ *è chiamato* tasso di rendimento atteso in eccesso *(*excess expected rate of return*) del titolo* m*-esimo. Il valore* $\bar{r}_T - r_0$ *è chiamato* tasso di rendimento atteso in eccesso *del portafoglio tangente.*

2.3.3 Rischio Sistematico

Abbiamo osservato che secondo il modello CAPM sussiste la relazione

$$\bar{r}_m - r_0 = \beta_m(\bar{r}_T - r_0).$$

Consideriamo allora la differenza

$$r_m - r_0 - \beta_m(r_T - r_0).$$

Questa sarà in generale una variabile aleatoria che denotiamo con ε_m. Possiamo allora scrivere

$$r_m = r_0 + \beta_m(r_T - r_0) + \varepsilon_m.$$

In conseguenza,

$$\bar{r}_m = r_0 + \beta_m(\bar{r}_T - r_0) + \bar{\varepsilon}_m,$$

pertanto il CAPM comporta che

$$\bar{\varepsilon}_m = 0.$$

Calcoliamo quindi

$$\begin{aligned} Cov(r_m, r_T) &= Cov(r_0 + \beta_m(r_T - r_0) + \varepsilon_m, r_T) \\ &= Cov(r_0, r_T) + \beta_m Cov(r_T, r_T) - \beta_m Cov(r_0, r_T) + Cov(\varepsilon_m, r_T) \\ &= \beta_m \mathbf{D}^2 [r_T] + Cov(\varepsilon_m, r_T), \end{aligned}$$

ed ancora il CAPM comporta che

$$Cov(\varepsilon_m, r_T) = 0.$$

Pertanto,

$$\begin{aligned} \sigma_m^2 &= \mathbf{D}^2[r_0 + \beta_m(r_T - r_0) + \varepsilon_m] \\ &= \beta_m^2 \mathbf{D}^2[r_T] + \mathbf{D}^2[\varepsilon_m] \\ &= \beta_m^2 \sigma_T^2 + \mathbf{E}[\varepsilon_m^2], \end{aligned}$$

da cui si vede che σ_m^2 è la somma di due parti. La prima parte $\beta_m^2 \sigma_T^2$, è chiamata *rischio sistematico.* E' il rischio associato al mercato nel suo complesso e non può essere ridotto mediante diversificazione, perchè il rischio di mercato è associato ad ogni titolo avente beta diverso da zero. La seconda parte $\mathbf{E}[\varepsilon_m^2] = \mathbf{D}^2[\varepsilon_m]$ è chiamata rischio *specifico* o *idiosincratico* che dipende dalla natura del titolo k-esimo e può essere ridotto mediante diversificazione.

2.4 CAPM e Prezzo dei Titoli del Mercato Finanziario

Secondo il CAPM il portafoglio tangente coincide con il portafoglio di mercato, ossia con il portafoglio i cui pesi sono dati dalla capitalizzazione relativa dei titoli che compongono il mercato in rapporto alla capitalizzazione dell'intero mercato. Se denotiamo con K_m la capitalizzazione del titolo m-esimo del mercato, con S_m il suo prezzo, con n_m il numero delle azioni del titolo m-esimo circolanti nel mercato, noto come *flottante* del titolo, con K la capitalizzazione dell'intero mercato e con w_m il peso del titolo m-esimo nel portafoglio di mercato, abbiamo

$$K_m = n_m S_m, \qquad K = \sum_{m=1}^{M} K_m, \qquad w_m = \frac{K_m}{K},$$

e possiamo porre

$$\bar{r}_M = \bar{r}_T, \qquad \sigma_M = \sigma_T$$

Con queste notazioni, il tasso di rendimento atteso $\bar{r}$ di un qualsiasi titolo del mercato finanziario soddisfa l'equazione

$$\bar{r} - r_0 = \beta(\bar{r}_M - r_0) \tag{2.138}$$

essendo r_0 il tasso di rendimento privo di rischio, r_M il tasso di rendimento del portafoglio di mercato di varianza σ_M^2, ed essendo $\beta \equiv \frac{cov(r,r_M)}{\sigma_M^2}$. D'altra parte, se al tempo $t = 0$ il titolo viene acquistato al prezzo $S(0)$ e successivamente al tempo $t = T$ il titolo viene rivenduto al prezzo $S(T)$ il suo tasso di rendimento è

$$r = \frac{S(T) - S(0)}{S(0)}.$$

Sostituendo quest'ultima nella (2.138) otteniamo

$$\frac{\bar{S}(T) - S(0)}{S(0)} = r_0 + \beta(\bar{r}_M - r_0).$$

Risolvendo quest'ultima rispetto a $S(0)$ ne segue

$$S(0) = \frac{\bar{S}(T)}{1 + r_0 + \beta(\bar{r}_M - r_0)}. \tag{2.139}$$

La (2.139) fornisce un'interessante interpretazione del prezzo di un titolo del mercato finanziario: *Il prezzo corrente di un titolo del mercato finanziario è il valore atteso del suo prezzo futuro scontato per un tasso di interesse aggiustato per il rischio.* Chiaramente, con tasso di interesse aggiustato per il rischio è da intendersi il termine $r_0 + \beta(\bar{r}_M - r_0)$.

Importante conseguenza della (2.139) è la linearità dei prezzi correnti. Ossia che il prezzo corrente della combinazione lineare di due, o più, titoli sia uguale alla combinazione lineare dei loro prezzi correnti. Infatti, sostituendo

$$\beta = \frac{cov(\frac{S(T)-S(0)}{S(0)}, r_M)}{\sigma_M^2} = \frac{cov(S(T), r_M)}{\sigma_M^2 S(0)},$$

nella (2.139) abbiamo

$$S(0) = \frac{\bar{S}(T)}{1 + r_0 + \frac{cov(S(T),r_M)}{\sigma_M^2 S(0)}(\bar{r}_M - r_0)} = \frac{\bar{S}(T)\,\sigma_M^2 S(0)}{(1 + r_0)\,\sigma_M^2 S(0) + cov(S(T), r_M)(\bar{r}_M - r_0)},$$

ossia

$$(1 + r_0)\,\sigma_M^2 S(0) + cov(S(T), r_M)(\bar{r}_M - r_0) = \bar{S}(T)\,\sigma_M^2,$$

che comporta

$$S(0) = \frac{\bar{S}(T)\,\sigma_M^2 - cov(S(T), r_M)(\bar{r}_M - r_0)}{(1 + r_0)\,\sigma_M^2} = \frac{1}{1 + r_0}\left(\bar{S}(T) - \frac{cov(S(T), r_M)(\bar{r}_M - r_0)}{\sigma_M^2}\right), \tag{2.140}$$

dove il termine in parentesi è noto come *equivalente certo* di $S(T)$. La (2.140) mostra chiaramente la linearità della formula dei prezzi.

La ragione della linearità è rintracciabile nel principio dell'assenza di arbitraggio: se il prezzo corrente della combinazione lineare di due titoli non fosse uguale alla combinazione lineare dei loro prezzi correnti, si realizzerebbero possibilità di arbitraggio.

Nel contesto dei mercati perfetti la linearità dei prezzi è un principio fondamentale della teoria finanziaria.

Capitolo 3

Multi-Period Investment Model

3.1 Simple Rate of Return or Interest

Let X_0 be an investment at the time $t = 0$ with maturity $t = T$.

Definizione 135 *We call* simple return *or* simple interest *at maturity T, denoted by R_T, the return or interest which is proportional to the time to maturity of the investment and the principal by a constant factor $r > 0$, which represents the return of the investment at the maturity of the unit of time and unit of money, in which T and X_0 are expressed. In symbols,*

$$R_T = rX_0T. \tag{3.1}$$

Osservazione 136 *Assume that $X_0 \neq 0$. Under simple return, we have*

$$r_T = rT, \tag{3.2}$$

where r_T is the rate of return at maturity T (see Definition 13).

Definizione 137 *We call the factor r the* unit *of simple rate of return.*

Osservazione 138 *Assume that $X_0 \neq 0$. Under simple return, we have*

$$a_T = 1 + rT, \qquad X_T = (1 + rT)\, X_0, \tag{3.3}$$

where a_T is the accumulation factor of the investment at maturity T (see Definition 14). In addition,

$$s_T = \frac{rT}{1 + rT}, \qquad d_T = \frac{1}{1 + rT}, \tag{3.4}$$

where s_T [resp. d_T] is the rate of discount [resp. the discount factor] of the investment at maturity T (see Definitions 17 and 18)

Proof. With regard Equation (3.4), on account of (2.13), we can write

$$s_T = \frac{r_T}{1 + r_T} = \frac{rT}{1 + rT}.$$

Considering (2.8) and (3.3), we obtain

$$d_T = \frac{X_0}{X_T} = \frac{X_0}{(1 + rT)\, X_0} = \frac{1}{1 + rT}.$$

□

Definizione 139 *We call the* unit of simple rate of discount *the term*

$$s \overset{def}{=} \frac{r}{1+r}.$$

Osservazione 140 *We have*

$$s_T = \frac{sT}{1+s\left(T-1\right)}.$$

In addition,

$$S_T = X_T - X_0 = s_T X_0 = \frac{sT}{1+s\left(T-1\right)} X_0 = \frac{rT}{1+rT} X_T,$$

and

$$X_0 = X_T - S_T = \left(1-s_T\right) X_T = \frac{1-s}{1+s\left(T-1\right)} X_T = \frac{1}{1+rT} X_T.$$

Proof. In fact, from Equation (3.4),

$$s_T = \frac{rT}{1+rT} = \frac{\frac{r}{1+r}T}{\frac{1+rT}{1+r}} = \frac{\frac{r}{1+r}T}{\frac{1+r+rT-r}{1+r}} = \frac{\frac{r}{1+r}T}{\frac{1+r}{1+r} + \frac{r}{1+r}\left(T-1\right)} = \frac{sT}{1+s\left(T-1\right)}.$$

□

Osservazione 141 *Assume that $X_0 \neq 0$. Under simple return, the following table holds true.*

$$\begin{array}{l}
R_T = X_T - X_0, \qquad r_T = \dfrac{R_T}{X_0} = rT, \qquad a_T = \dfrac{X_T}{X_0} = 1 + r_T = 1 + rT \\
S_T = X_T - X_0, \qquad s_T = \dfrac{S_T}{X_T} = \dfrac{rT}{1+rT}, \qquad d_T = \dfrac{X_0}{X_T} = 1 - s_T = \dfrac{1}{1+rT} \\
R_T = S_T, \qquad a_T d_T = 1, \qquad \left(1+r_T\right)\left(1+s_T\right) = 1, \\
s = \dfrac{r}{1+r}, \qquad s_T = \dfrac{r_T}{1+r_T}, \qquad r = \dfrac{s}{1-s}, \qquad r_T = \dfrac{s_T}{1-s_T}.
\end{array}$$

Let τ_1, τ_2 be two different units of time (e.g. one month and one year) and let r_{τ_1}, [resp. r_{τ_2}] be the the unit of simple rate of return corresponding to the unit of time τ_1, [resp. τ_2].

Osservazione 142 *We have*

$$r_{\tau_2} = r_{\tau_1}\tau_{2,1}, \tag{3.5}$$

where $\tau_{2,1}$ it the time unit τ_2 expressed in terms of the time unit τ_1.

Esempio 143 *Assume that τ_1 is one month, τ_2 is one year, and $r_{\tau_1} = 0.5\%$. Then, we have*

$$r_{\tau_2} = 0.5\% \cdot 12 = 6\%. \tag{3.6}$$

Conversely, assume that $r_{\tau_2} = 4\%$. Then, we have

$$r_{\tau_1} = 3\% \cdot \frac{1}{12} = 0.25\%. \tag{3.7}$$

Let s_{τ_1}, [resp. s_{τ_2}] be the the unit of simple rate of discount corresponding to the unit of time τ_1, [resp. τ_2].

Proposizione 144 *We have*

$$s_{\tau_2} = \frac{s_{\tau_1}\tau_{2,1}}{1+s_{\tau_1}\left(\tau_{2,1}-1\right)}, \tag{3.8}$$

where $\tau_{2,1}$ it the time unit τ_2 expressed in terms of the time unit τ_1.

Proof. Thanks to Equation (3.5), we can write

$$s_{\tau_2} = \frac{r_{\tau_2}}{1+r_{\tau_2}} = \frac{r_{\tau_1}\tau_{2,1}}{1+r_{\tau_1}\tau_{2,1}} = \frac{\frac{s_{\tau_1}}{1-s_{\tau_1}}\tau_{2,1}}{1+\frac{s_{\tau_1}}{1-s_{\tau_1}}\tau_{2,1}} = \frac{s_{\tau_1}\tau_{2,1}}{1+s_{\tau_1}(\tau_{2,1}-1)},$$

as desired. $\square$

Esempio 145 *Assume that τ_1 is one month, τ_2 is one year, and $r_{\tau_1} = 0.5\%$. Then, from Example (143), we have*

$$r_{\tau_2} = 6\%.$$

It follows,

$$s_{\tau_1} = \frac{r_{\tau_1}}{1+r_{\tau_1}} = frac0.5\%1+0.5\% \approx 0.00497 \approx 0.5\%,$$

and

$$s_{\tau_2} = \frac{r_{\tau_2}}{1+r_{\tau_2}} = frac6\%1+6\% \approx 0.0566 \approx 6\%.$$

On the other hand, thanks to Equation (3.8), we can write

$$s_{\tau_2} = \frac{s_{\tau_1}\tau_{2,1}}{1+s_{\tau_1}(\tau_{2,1}-1)} \approx \frac{0.00497\cdot 12}{1+0.00497\cdot(12-1)} \approx 0.0565$$

and

$$s_{\tau_1} = \frac{s_{\tau_2}\tau_{1,2}}{1+s_{\tau_2}(\tau_{1,2}-1)} \approx \frac{0.0566\cdot 0.0833}{1+0.0566\cdot(0.0833-1)} \approx 0.0497$$

which confirm the validity of Equation (3.8).

3.1.1 Capitalizzazione degli Interessi

In un regime finanziario a tasso di interesse semplice l'operazione di capitalizzazione degli interessi maturati è vantaggiosa per l'investitore.

Il montante X_T relativo ad un capitale X_0 investito in una attività economica ad un tasso di interesse semplice r corrisposto alla maturità T è dato da

$$X_T = (1+rT)X_0.$$

Dividendo l'intervallo di tempo $[0,T]$ nei sottointervalli $[0,t]$ e $[t,T]$, con $0 < t < T$, investendo lo stesso capitale X_0 e capitalizzando gli interessi prodotti nell'intervallo di tempo $[0,t]$ allo stesso tasso di interesse semplice r (se possibile), si ottiene stavolta un montante pari a

$$X_{t,T} = (1+rt)(1+r(T-t))X_0.$$

Volendo massimizzare $X_{t,T}$ non ci resta che massimizzare la funzione

$$a(t) = (1+rt)(1+r(T-t))$$

al variare di $t \in [0,T]$. D'altra parte si ha

$$a'(t) = r^2(T-2t),$$

per cui

$$a'(t) \geqslant 0 \quad \Leftrightarrow \quad t \leqslant T/2.$$

Quindi $a(t)$ assume valore massimo per $t^* = T/2$ ed in corrispondenza si ha

$$X_{t^*,T} = \left(1 + r\frac{T}{2}\right)^2 X_0.$$

Notare che

$$\left(1 + r\frac{T}{2}\right)^2 > (1 + rT)$$

e ciò rende l'operazione di capitalizzazione degli interessi maturati vantaggiosa per l'investitore.

Dividendo l'intervallo di tempo $[0, T]$ nei sottointervalli $[0, t_1]$, $[t_1, t_2]$ e $[t_2, T]$, con $0 < t_1 < t_2 < T$, investendo lo stesso capitale X_0 e capitalizzando gli interessi prodotti nei intervalli di tempo $[0, t_1]$ e $[t_1, t_2]$ allo stesso tasso di interesse semplice r, il montante diviene

$$X_{t_1,t_2,T} = (1 + rt_1)(1 + r(t_2 - t_1))(1 + r(T - t_2))X_0.$$

Tale montante assume valore massimo per $t_1^* = T/3$, $t_2^* = 2T/3$ ed in corrispondenza si ha

$$X_{t_1^*,t_2^*,T} = \left(1 + r\frac{T}{3}\right)^3 X_0.$$

Procedendo ulteriormente, e dividendo l'intervallo di tempo $[0, T]$ nei sottointervalli $[0, t_1]$, $[t_1, t_2]$,..., $[t_{n-2}, t_{n-1}]$, $[t_{n-1}, T]$, con $0 < t_1 < t_2 < \cdots < t_{n-2} < t_{n-1} < T$, investendo lo stesso capitale X_0 e capitalizzando gli interessi prodotti nei periodi $[0, t_1]$ e $[t_1, t_2]$ allo stesso tasso di interesse semplice r_p, il montante diviene

$$X_{t_1,t_2,\ldots,t_{n-1},T} = (1 + rt_1)(1 + r(t_2 - t_1)) \cdots (1 + r(t_{n-2} - t_{n-1}))(1 + r(T - t_{n-1}))X_0.$$

Si può provare che tale montante assume valore massimo per $t_1 = T/2$, $t_2 = 2T/n$,..., $t_{n-2} = (n-2)T/n$, $t_{n-1} = (n-1)T/n$, e che in corrispondenza si ha:

$$X_{t_1^*,t_2^*,\ldots,t_{n-1}^*,T} = \left(1 + r\frac{T}{n}\right)^n X_0.$$

Al limite, se si potesse effettuare una capitalizzazione "continua" degli interessi prodotti, allo stesso tasso di interesse semplice r, si otterrebbe

$$X_{\infty,T} = \lim_{n \to +\infty} \left(1 + r\frac{T}{n}\right)^n X_0 = \exp(rT)X_0.$$

3.2 Tasso d'Interesse Composto

In un'attività finanziaria a tasso di interesse composto gli interessi vengono capitalizzati automaticamente.

L'unità di capitale investita in un dato periodo di riferimento p al tasso d'interesse r_p produce, in tale periodo, un interesse pari ad r. Il fattore di capitalizzazione è allora

$$a_p = 1 + r_p.$$

Assumendo che a termine del periodo di riferimento gli interessi vengano capitalizzati per un altro periodo di pari durata, mantenendo inalterato il tasso d'interesse, il fattore di capitalizzazione sarà

$$a_{2p} = (1 + r_p)^2 .$$

Procedendo in questo modo per una successione di m periodi, al termine dell'm-esimo periodo di riferimento il fattore di capitalizzazione sarà

$$a_{mp} = (1 + r_p)^m.$$

A tale fattore di capitalizzazione corrisponde un tasso r_{mp} che deve allore soddisfare la

$$a_{mp} = 1 + r_{mp} = (1 + r_p)^m,$$

da cui

$$r_{mp} = (1 + r_p)^m - 1.$$

Analogamente, il tasso di interesse relativo ad $1/n$ del periodo di riferimento iniziale, denotato con $r_{p/n}$, deve soddisfare la relazione

$$a_{p/n} = 1 + r_{p/n}.$$

Tale tasso sarà equivalente ad r, qualora gli interessi vengano capitalizzati ad ogni termine di $1/n$ del periodo di riferimento iniziale. Si avrà pertanto

$$(1 + r_{p/n})^n = 1 + r_p.$$

Notare che da quest'ultima otteniamo

$$r_p = (1 + r_{p/n})^n - 1,$$

e

$$r_{p/n} = (1 + r_p)^{1/n} - 1.$$

Il fattore di capitalizzazione relativo ad m/n -esimi del periodo di riferimento iniziale è allora dato da

$$a_{mp/n} = (1 + r_{p/n})^m = (1 + r_p)^{m/n}.$$

Tale formula porta a concludere che, in un regime di interesse composto, quando gli interessi vengono capitalizzati a fine di ogni periodo al tasso di interesse r il fattore di capitalizzazione è dato da

$$a_T = (1 + r_p)^T ,$$

essendo r il tasso di interesse periodale e $T > 0$ la durata dell'attività finanziaria in termini del periodo p.

Notare che in un'attività finanziaria a tasso di interesse composto eventuali operazioni intermedie di capitalizzazione degli interessi maturati non alterano il montante. Si ha infatti

$$(1 + r_p)^t (1 + r_p)^{T-t} = (1 + r_p)^T ,$$

per ogni $0 < t < T$.

La legge di formazione del tasso di interesse in un'attività finanziaria a tasso di interesse composto è data da:

$$r_t \stackrel{\text{def}}{=} (1 + r_p)^t - 1. \tag{3.9}$$

In regime di tasso d'interesse composto abbiamo:

Legge di formazione del montante

$$X_t = a_t X_0 = (1 + r_p)^t X_0$$

essendo

$$a_t = (1 + r_p)^t.$$

Legge di formazione del rendimento (interesse)

$$R_t = r_t X_0 = \left((1 + r_p)^t - 1\right) X_0.$$

Legge di formazione del fattore di attualizzazione (sconto)

$$d_t = \frac{1}{a_t} = (1 + r_p)^{-t} = (1 - s_p)^t, \tag{3.10}$$

essendo

$$s_p \equiv \frac{r_p}{1 + r_p} \qquad \text{e} \qquad r_p = \frac{s_p}{1 - s_p}.$$

il tasso di sconto e il tasso di rendimento, rispettivamente (see Equation ()) i

$$R_T = X_T - X_0, \qquad r_T = \frac{R_T}{X_0} = (1 + r)^T - 1, \qquad a_T = \frac{X_T}{X_0} = 1 + r_T = (1 + r)^T$$

$$S_T = X_T - X_0, \qquad s_T = \frac{S_T}{X_T} = \frac{(1 + r)^T - 1}{(1 + r)^T}, \qquad d_T = \frac{X_0}{X_T} = 1 - s_T = \frac{1}{(1 + r)^T}$$

$$R_T = S_T, \qquad a_T d_T = 1, \qquad (1 + r_T)(1 + s_T) = 1,$$

$$s_T = \frac{r_T}{1 + r_T}, \qquad r_T = \frac{s_T}{1 - s_T}$$

$$s = \frac{r}{1 + r}, \qquad r = \frac{s}{1 - s}$$

Legge di formazione del tasso di sconto

$$s_t = 1 - d_t = 1 - (1 + r_p)^{-t} = 1 - (1 - s_p)^t.$$

essendo

Legge di formazione del valore attuale

$$X_0 = d_T X_T = (1 + r_p)^{-T} X_T = (1 - d_p)^T X_T.$$

Legge di formazione dello sconto

$$S_0 = d_T X_T = \left(1 - (1 - d_p)^T\right) X_T = \left(1 - (1 + r_p)^{-T}\right) X_T.$$

Ricordiamo ancora le relazioni

$$r_p = (1 + r_{p/m})^m - 1,$$

e

$$r_{p/m} = (1 + r_p)^{1/m} - 1,$$

che consentono di esprimere lo stesso tasso di interesse rapportato a periodi di riferimento differenti.

Abbiamo poi le analoghe

$$d_p = 1 - (1 - d_{p/m})^m,$$

e

$$d_{p/m} = 1 - (1 - d_p)^{1/m}$$

che consentono di esprimere lo stesso tasso di sconto rapportato a periodi di riferimento differenti.

3.3 Confronto Tra i Principali Regimi Finanziari

Denotiamo con $a_t^{(s)}$, $a_t^{(c)}$ i fattori di capitalizzazione (*accumulation factors*) tipici dei due regimi di tasso d'interesse semplice, tasso d'interesse composto. Ricordiamo che, in relazione allo stesso tasso di interesse periodale r_p si ha

$$a_T^{(s)} = 1 + r_p T, \qquad a_T^{(c)} \equiv (1 + r_p)^T.$$

Abbiamo chiaramente

$$a_0^{(s)} = a_0^{(c)} = 1$$

e

$$a_1^{(s)} = a_1^{(c)} = 1 + r_p.$$

Si può inoltre provare che risulta

$$\begin{array}{ll} a_T^{(s)} \geqslant a_T^{(c)} & \text{if } 0 \leqslant T \leqslant 1 \\ a_t^{(c)} \geqslant a_t^{(s)} & \text{if } T \geqslant 1 \end{array}.$$

3.4 Capitalizzazione Mista

Nella pratica dei rapporti di conto corrente bancario trova usuale applicazione un regime finanziario che risulta una forma intermedia tra il regime ad interesse semplice ed il regime ad interesse composto, noto come regime della capitalizzazione mista.

Supponiamo che gli interessi vengano accreditati con una certa periodicità ad un tasso r_p ed a delle date prestabilite e che un capitale X_0 sia reso fruttifero per un tempo T (espresso in termini del periodo di calcolo degli interessi).

Si scomponga T nella forma

$$T = s + n + t$$

essendo s il tempo che decorre dal deposito del capitale sino alla data del primo accredito degli interessi, n il numero di periodi interi durante i quali il capitale rimane in deposito e t il tempo che decorre dalla data dell'ultimo accredito degli interessi alla data della riscossione del montante X_T. Tale montante viene allora calcolato secondo la formula

$$X_T = (1 + r_p s)(1 + r_p)^n (1 + r_p t) X_0.$$

Da notare che il regime dell'interesse composto puro porterebbe ad un montante

$$\tilde{X}_T = (1 + r_p)^T X_0 = (1 + r_p)^s (1 + r_p)^n (1 + r_p)^t X_0,$$

e risulterebbe

$$\tilde{X}_T \leqslant X_T,$$

dal momento che

$$(1 + r_p)^s \leqslant (1 + r_p s) \qquad \text{e} \qquad (1 + r_p)^t \leqslant (1 + r_p t),$$

essendo $s, t \leqslant 1$.

3.4.1 Tasso Nominale d'Interesse

Tasso nominale e tasso effettivo

Un capitale unitario, $X_0 = 1$, viene investito in regime di interesse composto, al tasso periodale r_p. L'interesse via via prodotto viene corrisposto all'investitore in m sottoperiodi del periodo di investimento. Per ciascuno dei sottoperiodi considerati l'investitore percepirà un interesse pari a $r_{p/m}$. Al termine del periodo di investimento l'investitore avrà percepito come interesse m rate di ammontare

$$r_{p/m} = (1 + r_p)^{1/m} - 1$$

ciascuna.

La somma di queste m rate, che denotiamo con $r_p(m)$, è detta tasso nominale periodale di interesse m-volte convertibile corrispondente ad r_p:

$$r_p(m) = mr_{p/m} = m((1 + r_p)^{1/m} - 1).$$

Che si possono invertire nella forma

$$r_{p/m} = \frac{r_p(m)}{m},$$

e

$$r_p = \left(1 + \frac{r_p(m)}{m}\right)^m - 1.$$

Proposizione 146 *Se in una attività finanziaria si investe un capitale X_0 per un certo periodo durante il quale viene corrisposto un tasso nominale periodale di interesse m-volte convertibile $r_p(m)$ e se gli interessi corrisposti a termine di ciascuno degli m sottoperiodi del periodo di investimento vengono via via investiti al tasso r_p per il periodo residuo di investimento, in un regime di interesse composto, allora investire nel capitale al tasso $r_p(m)$ oppure al tasso r_p in un regime di interesse composto produce lo stesso montante.*

Proof. *Assumendo per semplicità di notazioni il capitale X_0 unitario, l'investimento al tasso $r_p(m)$ produce a termine del primo sottoperiodo del periodo di investimento un interesse pari a*

$$r_p(m)/m = r_{p/m}.$$

Tale interesse investito al tasso r_p in un regime di interesse composto produce a termine dell'intervallo residuale del periodo di investimento un montante pari a

$$r_{p/m}(1 + r_p)^{1-1/m}.$$

Similmente, a termine del secondo sottoperiodo del periodo di investimento il capitale unitario investito produce un interesse ancora pari a

$$r_p(m)/m = r_{p/m}.$$

Tale interesse investito al tasso r_p in un regime di interesse composto produce stavolta a termine dell'intervallo residuale del periodo di investimento un montante pari a

$$r_{p/m}(1 + r_p)^{1-2/m}.$$

In generale, a termine del k-esimo degli m sottoperiodi del periodo di investimento viene sempre prodotto un interesse pari a

$$r_p(m)/m = r_{p/m}.$$

e tale interesse investito al tasso r_p in un regime di interesse composto produce a termine dell'intervallo residuale del periodo di investimento un montante pari a

$$r_{p/m}(1 + r_p)^{1-k/m},$$

per $k = 1, \ldots, m$. Da notare che all'ultimo dei sottoperiodi di investimento, $k = m$, il montante prodotto coincide con l'interesse stesso. Il montante complessivamente generato risulta allora

$$\begin{aligned}
\sum_{k=1}^{m} r_{p/m}(1+r_p)^{1-k/m} &= r_{p/m}(1+r_p)\sum_{k=1}^{m}(1+r_p)^{-k/m} \\
&= r_{p/m}(1+r_p)\sum_{k=1}^{m}(1+r_p)^{-k/m}(1+r_p)^{-1/m}(1+r_p)^{1/m} \\
&= r_{p/m}(1+r_p)(1+r_p)^{-1/m}\sum_{k=1}^{m}(1+r_p)^{1/m-k/m} \\
&= \left((1+r_p)^{1/m}-1\right)(1+r_p)(1+r_p)^{-1/m}\sum_{k=1}^{m}\left((1+r_p)^{-1/m}\right)^{k-1} \\
&= ((1+r_p)^{1/m}-1)(1+r_p)^{-1/m}(1+r_p)\frac{1-((1+r_p)^{-1/m})^m}{1-(1+r_p)^{-1/m}} \\
&= (1-(1+r_p)^{-1/m})(1+r_p)\frac{1-(1+r_p)^{-1}}{1-(1+r_p)^{-1/m}} \\
&= r_p.
\end{aligned}$$

□

3.4.2 Tasso Istantaneo d'Interesse

Siano r_T il tasso di interesse relativo ad un periodo T e sia $0 \equiv t_0 < t_1 < \cdots < t_n \equiv T$ una successione di istanti di tempo tale che $t_{k+1} - t_k = T/n$ per $k = 0, \ldots, n-1$. Sia quindi $v \equiv (x_0, x_1, \ldots, x_n)$ un flusso di cassa relativo a tali istanti.

Il valore futuro di v con composizione periodale degli interessi su base T/n è dato da

$$FV_n(v) = \sum_{k=0}^{n} x_k(1 + \frac{T}{n} r_T)^{n-k}$$

mentre con composizione continua degli interessi si ha

$$PV_\infty(v) = \sum_{k=0}^{n} x_k e^{r_T(T-t_k)}.$$

Da notare che, essendo $t_k = k\frac{T}{n}$,

$$\sum_{k=0}^{n} e^{r_T(T-t_k)} = \sum_{k=0}^{n} e^{r_T\frac{T}{n}(n-k)} = \sum_{k=0}^{n} (e^{r_T\frac{T}{n}})^{n-k} \approx \sum_{k=0}^{n}(1 + \frac{T}{n} r_T)^{n-k}.$$

Pertanto

$$PV_n(v) \approx PV_\infty(v)$$

e l'approssimazione è tanto migliore quanti più $n \to \infty$.

Siano r_T il tasso di interesse relativo ad un periodo T e sia $0 \equiv t_0 < t_1 < \cdots < t_n \equiv T$ una successione di istanti di tempo tale che $t_{k+1} - t_k = T/n$ per $k = 0, \ldots, n-1$. Sia quindi $v \equiv (x_0, x_1, \ldots, x_n)$ un flusso di cassa relativo a tali istanti.

Il valore attuale di v con composizione periodale degli interessi su base T/n è dato da

$$PV_n(v) = \sum_{k=0}^{n} \frac{x_k}{(1 + \frac{T}{n} r_T)^k},$$

mentre con composizione continua degli interessi si ha

$$PV_\infty(v) = \sum_{k=0}^{n} x_k e^{-r_T t_k}.$$

Da notare che

$$\sum_{k=0}^{n} e^{-r_T t_k} = \sum_{k=0}^{n} e^{-r_T \frac{T}{n} k} = \sum_{k=0}^{n} (e^{r_T \frac{T}{n}})^{-k} \approx \sum_{k=0}^{n} \frac{1}{(1 + \frac{T}{n} r_T)^k}.$$

Pertanto

$$PV_n(v) \approx PV_\infty(v)$$

e l'approssimazione è tanto migliore quanti più $n \to \infty$.

In regime finanziario ad interesse composto a tasso periodale r_p, consideriamo il tasso nominale periodale di interesse m-volte convertibile corrispondente ad r_p

$$r_p(m) = m r_{p/m} = m((1 + r_p)^{1/m} - 1).$$

Si ha

$$\lim_{m \to \infty} r_p(m) = \log(1 + r_p). \tag{3.11}$$

Infatti,

$$\begin{aligned}
\lim_{m \to \infty} r_p(m) &= \lim_{m \to \infty} \frac{(1 + r_p)^{1/m} - 1}{1/m} \\
&= \lim_{x \to 0^+} \frac{(1 + r_p)^x - 1}{x} \\
&= \log(1 + r_p).
\end{aligned}$$

La quantità

$$\rho_p \equiv \log(1 + r_p) \tag{3.12}$$

è nota come *tasso istantaneo periodale di interesse*, o *tasso nominale periodale infinite volte convertibile*, o *tasso periodale di interesse composto continuamente*. Da notare che la legge di formazione del fattore di capitalizzazione può essere scritta come:

$$r(T) = (1 + r_p)^T = \exp(\log(1 + i_p))T) = \exp(\rho_p T).$$

Legge di formazione del montante:

$$X_T = X_0 \exp(\rho_p T)$$

Legge di formazione del valore attuale

$$X_0 = X_T \exp(-\rho_p t)$$

Il tasso istantaneo relativo $\rho_{p/m}$ ad un m-esimo del periodo di riferimento iniziale p sarà correlato a ρ_p da:

$$\rho_{p/m} \equiv \log(1 + r_{p/m}) = \log((1 + r_p)^{1/m}) = \frac{1}{m} \log(1 + r_p) = \frac{1}{m} \rho_p.$$

Il tasso istantaneo di interesse ρ_p produce in una frazione dt del periodo di investimento p di un capitale unitario l'interesse

$$\rho_p dt$$

assumendo che tale somma sia disponibile all'inizio dell'intervallo $[t, t + dt]$ e capitalizzandola ad interesse composto per l'intervallo residuale del periodo otteniamo il montante

$$(1 + r_p)^{1-t} \rho_p dt = \log(1 + r_p)(1 + r_p)^{1-t} dt.$$

Il contributo complessivo dovuto a tutte le frazioni dt di periodo sarà allora dato da

$$\begin{aligned}
\int_0^1 \log(1 + i_p)(1 + i_p)^{1-t} dt &= \int_0^1 \log(1 + i_p) \exp(\log(1 + i_p)(1 - t))\, dt \\
&= \int_0^{\log(1+i_p)} \exp(u)\, du \\
&= \exp(\log(1 + i_p)) - \exp(0) \\
&= 1 + r_p - 1 \\
&= r_p.
\end{aligned}$$

Che coincide con l'interesse prodotto nel periodo al tasso periodale r_p

$$(1 + r_p)^T = \exp(\rho_p T) \qquad \rho_p = \log(1 + r_p)$$

$$(1 + r_y)^{T_y} = \exp(\rho_y T_y) \qquad \rho_y = \log(1 + r_y)$$

$$(1 + r_m)^{T_m} = \exp(\rho_m T_m) \qquad \rho_m = \log(1 + r_m)$$

$$(1 + r_m)^{12_m} = 1 + r_y$$

$$r_y = (1 + r_m)^{12_m} - 1 \simeq 1 + 12_m r_m - 1 = 12_m r_m$$

$$r_m \simeq \frac{1}{12} r_y$$

$$\rho_y = \log(1 + r_y) \simeq \log(1 + 12_m r_m)$$

$$\exp(\rho_y) = 1 + 12_m r_m \simeq (1 + r_m)^{12_m} = \exp(\rho_m 12_m)$$

$$\rho_y = \rho_m 12_m$$

$$(1 + r_m)^{T_m} = \exp(\rho_m T_m) \Rightarrow (1 + r_m)^{12_m} = \exp(\rho_m 12_m) = \exp(\rho_y 1_y) \Rightarrow \rho_y = \rho_m 12_m$$

.

3.5 Cox-Ross-Rubinstein Model (Multi-Period Multiplicative Binomial Model)

Consideriamo adesso un modello di mercato multiperiodale articolato in N periodi di contrattazione, con $N \geqslant 2$, individuati dalla successione finita di tempi $0 \equiv t_0 < t_1 < \cdots < t_{n-1} < t_N \equiv T$. Per semplicità assumiamo che tutti i periodi di contrattazione abbiano la stessa durata, cioè

$$t_{n+1} - t_n = T/N \equiv \Delta t,$$

per ogni $n = 0, \ldots, N-1$. Ovviamente ciò comporta

$$t_n = n\Delta t,$$

per ogni $n = 0, \ldots, N$. Come nel caso monoperiodale, assumiamo sia possibile investire in un titolo non rischioso, riferito come *bond* e denotato con B, in un titolo rischioso, riferito come *stock* e denotato con S, e in *derivati* di sottostante lo stock, di volta in volta specificati e denotati. Inoltre, denoteremo con B_n [risp. S_n] il valore di mercato del bond [risp. dello stock] al tempo t_n, per ogni $n = 0, \ldots, N$. Chiaramente, $B_N \equiv B_T$ e $S_N \equiv S_T$.

A partire dal tempo $t = t_0$ e a inizio di ogni periodo di contrattazione, un operatore finanziario osserva i valori di mercato del bond, dello stock e dei derivati sullo stock, e investe in un portafoglio costituito dai titoli stessi. A termine di ogni periodo di contrattazione i valori di mercato dei titoli vengono aggiornati, su base deterministica o aleatoria, secondochè si tratti del bond o dello stock e i suoi derivati, l'operatore finanziario ha quindi facoltà di riconfigurare il suo portafoglio. All'istante terminale, $t = t_N$, questo processo s'arresta, l'operatore liquida interamente il suo portafoglio, l'eventuale a ricchezza prodotta viene consumata o l'eventuale debito contratto va ripagato. Come nel caso monoperiodale, relativamente alle operazioni di compravendita, assumiamo che:

1. i tassi di rendimento relativi alle operazioni di prestito e deposito siano gli stessi;

2. gli investimenti possano essere effettuati senza alcuna limitazione quantitativa, cioè sia possibile operare su una qualunque frazione di bond, stock e derivati rispetto all'unità di moneta fissata;

3. i costi di transazione siano nulli;

4. le vendite allo scoperto sul bond, sullo stock e sui derivati siano totalmente libere;

5. lo stock non distribuisca dividendi.

Assumiamo anche che il bond frutti al possessore un tasso d'interesse $r_f > 0$, costante nel periodo di contrattazione $[t_{n-1}, t_n]$, per ogni $n = 1, \ldots, N$, con interesse pagato al termine di ogni periodo. In virtù di questa ipotesi, il valore di mercato di un investimento sul bond per un ammontare inizale B_0 segue la dinamica

$$B_0 > 0, \quad B_{n+1} \stackrel{\text{def}}{=} (1 + r_f)B_n, \quad n = 0, 1, \ldots, N-1. \tag{3.13}$$

Investendo un ammontare B_0 sul bond al tempo $t = t_0$ e denetendolo fino al tempo $t = t_n$ disporremo in portafoglio di un asset con valore di liquidazione pari a

$$B_n = (1 + r_f)^n B_0, \tag{3.14}$$

per ogni $n = 0, \ldots, N$. Analogamente a quanto osservato nel caso di un mercato monoperiodale abbiamo

Proposizione 147 *Stante la (3.14), il valore di una somma di denaro M_0 al tempo $t = 0$ ha un valore capitalizzato al tempo $t = t_n$ pari a*

$$M_n = (1 + r_f)^n M_0, \tag{3.15}$$

per ogni $n = 0, \dots, N$. *Viceversa, il valore di una somma* M_T *di denaro al tempo* $t = T$ *ha un valore scontato al tempo* $t = t_n$ *pari a*

$$M_n = \frac{M_T}{(1 + r_f)^{N-n}}, \tag{3.16}$$

per ogni $n = 0, \dots, N$.

Proof. *Come nel caso mono-periodale, la disponibilità dell'ammontare* M_0 *al tempo* $t = 0$ *consente l'acquisto di*

$$x = \frac{M_0}{B_0}$$

unità di titolo non rischioso di valore di mercato B_0. *D'altra parte, nel modello multi-periodale, la liquidazione di questo investimento al tempo* $t = t_n$ *produce un ammontare* M_n *secondo la formula*

$$M_n = xB_n = \frac{M_0}{B_0}(1 + r_f)^n B_0 = (1 + r_f)^n M_0.$$

Viceversa, volendo produrre un montante M_T *al tempo* $t = T$ *è necessario liquidare*

$$x = \frac{M_T}{B_T}$$

unità di titolo non rischioso di valore di mercato B_T. *Nel modello multi-periodale, l'acquisto di tali unità di titolo non rischioso al tempo* $t = t_n$ *richiede l'investimento di un ammontare* M_n *pari a*

$$M_n = xB_n = \frac{M_T}{B_T} B_n = \frac{M_T}{(1 + r_f)^{N-n} B_n} B_n = \frac{M_T}{(1 + r_f)^{N-n}}.$$

□

Per descrivere la dinamica rischiosa $(S_n)_{n=0}^N$ del prezzo dello stock è opportuno definire uno spazio di probabilità che supporti il modello multiperiodale. A tale scopo, assumiamo che al termine di ciascuno dei periodi di contrattazione si possano realizzare soltanto due accadimenti aleatori, uno positivo e uno negativo, in base ai quali il valore di mercato dello stock venga aggiornato a partire dal valore che aveva a inizio periodo. La dinamica dello stock è quindi tipica di un *fenomeno stocastico*, ossia un fenomeno aleatorio la cui incertezza sull'esito si disvela progressivamente nel tempo. Avendo supposto che lo stock non rilasci dividendi, il valore S_n assunto dallo stock all'istante t_n risulta allora effetto della particolare successione di accadimenti positivi o negativi che si realizzano a partire dall'istante iniziale $t_0 \equiv 0$ fino all'istante t_n, per ogni $n = 1, \dots N$. Per rappresentare la generica di tali successioni è naturale impiegare una successione di N termini $(\omega_n)_{n=1}^N$, ciascun ω_n dei quali sia denotato con uno tra due diversi simboli, come da standard 1 e 0, secondoché al tempo t_n si realizzi l'accadimento positivo o quello negativo. Lo *spazio campionario* Ω diviene allora l'insieme di tutte queste possibili successioni. Precisamente,

$$\Omega \stackrel{\text{def}}{=} \{\omega \equiv (\omega_n)_{n=1}^N \mid \omega_n = 0 \ \vee \ \omega_n = 1, \quad n = 1, \dots, N\} = \{0, 1\}^N, \tag{3.17}$$

essendo $\{0, 1\}^N \equiv \times_{n=1}^N \{0, 1\}$ il prodotto cartesiano di N copie dell'insieme $\{0, 1\}$, *potenza cartesiana* Nesima di $\{0, 1\}$. Lo spazio Ω così definito viene a essere costituito da 2^N *campioni* [1], tanti quante sono le possibili Nple le cui componenti hanno valore 0 o 1. Poichè Ω è finito, è possibile scegliere come *famiglia* $\mathcal{E}$ *degli eventi* di Ω, rappresentativa di tutta l'informazione acquisibile sul fenomeno stocastico da parte di un'osservatore, la σ-algebra discreta, ossia la famiglia $\mathcal{P}(\Omega)$ di tutti i possibili sottoinsiemi di Ω [2]. Questa scelta consente di considerare come *eventi elementari* gli eventi costituiti

[1] In simboli, $|\Omega| = 2^N$, denotando $|\Omega|$ la *cardinalità* di Ω.

[2] Poichè Ω è finito, la σ-algebra discreta $\mathcal{E} \equiv \mathcal{P}(\Omega)$ è in effetti un'algebra costituita da $2^{|\Omega|} = 2^{2^N}$ eventi.

da singole successioni di accadimenti e di definire una probabilità $\mathbf{P}$ su Ω a partire da una *distribuzione di probabilità oggettiva* sugli eventi elementari. Assumendo che gli accadimenti positivi o negativi si realizzino in successione indipendentemente gli uni dagli altri e che un singolo accadimento positivo [risp. negativo] si presenti sempre con probabilità p [risp. $q \equiv 1 - p$], la distribuzione di probabilità in questione è data da

$$\mathbf{P}(\omega) \equiv \mathbf{P}((\omega_n)_{n=1}^N) \stackrel{\text{def}}{=} p^K q^{N-K}, \quad K = |\{n : \omega_n = 1\}|, \quad \forall \omega \in \Omega, \tag{3.18}$$

dove $\mathbf{P}(\omega)$ è l'abbreviazione standard per $\mathbf{P}(\{\omega\})$ e il simbolo $|\{n : \omega_n = 1\}|$, *cardinalità* dell'insieme $\{n : \omega_n = 1\}$, rappresenta il numero degli indici n in $(\omega_n)_{n=1}^N$ per i quali $\omega_n = 1$. La probabilità oggettiva $\mathbf{P} : \mathcal{E} \to \mathbb{R}_+$ generata da tale distribuzione è allora data da

$$\mathbf{P}(E) \stackrel{\text{def}}{=} \sum_{\omega \in E} \mathbf{P}(\omega), \quad E \in \mathcal{E}. \tag{3.19}$$

In altri termini, la probabilità di ogni evento E è definita come la somma delle probabilità degli eventi elementari componenti E. Per verificare che tramite le (3.18) e (3.19) si sia effettivamente definita una probabilità è necessario e sufficiente provare che

$$\mathbf{P}(\Omega) = 1. \tag{3.20}$$

Infatti,

$$\mathbf{P}(\Omega) = \sum_{\omega \in \Omega} \mathbf{P}(\omega) = \sum_{K=0}^{N} \sum_{\omega \in \Omega : |\{n | \omega_n = 1\}| = K} p^K q^{N-K} = \sum_{K=0}^{N} \binom{N}{K} p^K q^{N-K} = (p+q)^N = 1. \tag{3.21}$$

Un'altra conseguenza della scelta $\mathcal{E} \equiv \mathcal{P}(\Omega)$ è che ogni applicazione $X : \Omega \to \mathbb{R}^M$ di Ω nello spazio euclideo reale M-dimensionale $\mathbb{R}^M$, per un qualsiasi $M \in \mathbb{N}$, soddisfa chiaramente la condizione

$$\{X \in B\} \in \mathcal{E}, \quad \forall B \in \mathcal{B}(\mathbb{R}^M), \tag{3.22}$$

essendo $\{X \in B\}$ l'abbreviazione standard per l'evento $\{\omega \in \Omega : X(\omega) \in B\}$, contro-immagine di B mediante X, costituito dagli esiti del fenomeno aleatorio sui quali la funzione X prende valori in B, ed essendo $\mathcal{B}(\mathbb{R}^M)$ la σ-algebra di Borel su $\mathbb{R}^M$. Secondo la terminologia probabilistica [risp. della teoria della misura], ogni applicazione $X : \Omega \to \mathbb{R}^M$ risulta essere una $\mathcal{E} - \mathcal{B}(\mathbb{R}^M)$ variabile aleatoria, ovvero una $\mathcal{E}$ variabile aleatoria reale M-variata [risp. una funzione $\mathcal{E} - \mathcal{B}(\mathbb{R}^M)$ misurabile], per un qualsiasi $M \in \mathbb{N}$. In un linguaggio meno formale, rappresentando la σ-algebra $\mathcal{E}$ l'informazione sul fenomeno stocastico in questione, potremmo dire che tale informazione consente di stabilire se una funzione $X : \Omega \to \mathbb{R}^M$ assuma i suoi valori in un qualsiasi insieme di $\mathcal{B}(\mathbb{R}^M)$ o meno. Per di più, per la finitezza di Ω, ogni $\mathcal{E} - \mathcal{B}(\mathbb{R}^M)$ variabile aleatoria ha momento finito di ordine K, per ogni $K \in \mathbb{N}$. Infatti, posto

$$|X(\omega)| \equiv \left(\sum_{m=1}^{M} X_m^2(\omega) \right)^{1/2},$$

per ogni $\omega \in \Omega$, risulta

$$\int_\Omega |X(\omega)|^K d\mathbf{P} = \sum_{\omega \in \Omega} |X(\omega)|^K \mathbf{P}(\omega) < \infty,$$

per ogni $K \in \mathbb{N}$.

Riassumendo

Osservazione 148 *Ogni $X : \Omega \to \mathbb{R}^M$ di Ω in $\mathbb{R}^M$, per un qualsiasi $M \in \mathbb{N}$, è una $\mathcal{E} - \mathcal{B}(\mathbb{R}^M)$ variabile aleatoria con momento finito di ordine K, per ogni $K \in \mathbb{N}$.*

Per $K=1$ tale momento è anche noto come *speranza* di X, definito come

$$\mathbf{M}_1'(X) \equiv \mathbf{E}[X] \overset{\text{def}}{=} (\mathbf{E}[X_1],\ldots,\mathbf{E}[X_M])^{\intercal}$$

essendo, in termini di componenti, $X \equiv (X_1,\ldots,X_M)^{\intercal}$, ed essendo

$$\mathbf{E}[X_m] \overset{\text{def}}{=} \textstyle\sum_{\omega\in\Omega} X_m(\omega)\,\mathbf{P}(\omega), \quad \forall m=1,\ldots,M.$$

Per $K=2$ il momento è definito come

$$\mathbf{M}_2'(X) \equiv \mathbf{E}[X^{\intercal}X] \overset{\text{def}}{=} \begin{pmatrix} \mathbf{E}\left[X_1^2\right] & \mathbf{E}[X_1X_2] & \cdots & \mathbf{E}[X_1X_{M-1}] & \mathbf{E}[X_1X_M] \\ \mathbf{E}[X_2X_1] & \mathbf{E}\left[X_2^2\right] & \cdots & \mathbf{E}[X_2X_{M-1}] & \mathbf{E}[X_2X_M] \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{E}[X_{M-1}X_1] & \mathbf{E}[X_{M-1}X_2] & \cdots & \mathbf{E}\left[X_{M-1}^2\right] & \mathbf{E}[X_{M-1}X_M] \\ \mathbf{E}[X_MX_1] & \mathbf{E}[X_MX_2] & \cdots & \mathbf{E}[X_2X_M] & \mathbf{E}\left[X_M^2\right] \end{pmatrix}.$$

essendo

$$\mathbf{E}[X_\ell X_m] \overset{\text{def}}{=} \textstyle\sum_{\omega\in\Omega} X_\ell(\omega)\,X_m(\omega)\,\mathbf{P}(\omega), \quad \forall \ell,m=1,\ldots,M.$$

Da notare che il momento di ordine 2 di X si presta a essere concepito come un *tensore simmetrico* di ordine 2. Da notare anche che nelle applicazioni, piuttosto che il momento di ordine 2 di X, è più frequente l'uso del *momento centralizzato* di ordine 2 di X, anche noto come *varianza-covarianza* di X, definito come il momento di ordine 2 di $X-\mathbf{E}[X]$. Formalmente,

$$\begin{aligned} \mathbf{M}_2(X) &\equiv Var(X) \overset{\text{def}}{=} \mathbf{M}_2'(X-\mathbf{E}[X]) \equiv \mathbf{E}[(X-\mathbf{E}[X])^{\intercal}(X-\mathbf{E}[X])] \\ &\equiv \begin{pmatrix} \mathbf{D}^2[X_1] & Cov(X_1,X_2) & \cdots & Cov(X_1,X_{M-1}) & Cov(X_1,X_M) \\ Cov(X_2,X_1) & \mathbf{D}^2[X_2] & \cdots & Cov(X_2,X_{M-1}) & Cov(X_2,X_M) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ Cov(X_{M-1},X_1) & Cov(X_{M-1},X_2) & \cdots & \mathbf{D}^2[X_{M-1}] & Cov(X_{M-1},X_M) \\ Cov(X_M,X_1) & Cov(X_M,X_2) & \cdots & Cov(X_M,X_{M-1}) & \mathbf{D}^2[X_M] \end{pmatrix}, \end{aligned}$$

essendo

$$\mathbf{D}^2[X_m] \overset{\text{def}}{=} \textstyle\sum_{\omega\in\Omega} (X_m(\omega)-\mathbf{E}[X_m])^2\,\mathbf{P}(\omega), \quad \forall m=1,\ldots,M,$$

e

$$Cov(X_\ell,X_m) \overset{\text{def}}{=} \textstyle\sum_{\omega\in\Omega} (X_\ell(\omega)-\mathbf{E}[X_\ell])\,(X_m(\omega)-\mathbf{E}[X_m])\,\mathbf{P}(\omega), \quad \forall \ell,m=1,\ldots,M.$$

Per $K=3$ [risp. $K=4$] il momento di ordine K è un tensore di ordine 3 [risp. 4]. Questo può essere concepito come una lista 3 [risp. 4]-dimensionale di numeri reali del tipo

$$\mathbb{M}_3'(X) \overset{\text{def}}{=} (\mathbf{E}[X_kX_\ell X_m])_{k,\ell,m=1}^{M} \qquad [\text{risp. } \mathbb{M}_4'(X) \overset{\text{def}}{=} (\mathbf{E}[X_jX_kX_\ell X_m])_{j,k,\ell,m=1}^{M}\,]$$

Tra gli M^3 [risp. M^4] elementi componenti il momento di ordine 3 [risp. 4] gli elementi distinti sono individuati da una scelta di indici k,ℓ,m [risp. j,k,ℓ,m] corrispondente a una funzione non decrescente $\phi:\{1,2,3\}\to\{1,\ldots,M\}$ [risp. $\phi:\{1,2,3,4\}\to\{1,\ldots,M\}$, ovvero a una combinazione con ripetizioni di M elementi in classe 3 [risp. 4]. Queste sono in numero di

$$C_{M,3}^{(r)} \equiv \binom{M+3-1}{3} = \frac{M(M+1)(M+2)}{6}$$

$$[\text{risp. } C_{M,4}^{(r)} \equiv \binom{M+4-1}{4} = \frac{M(M+1)(M+2)(M+2)}{24}],$$

denotando $C^{(r)}_{M,K}$ il numero delle combinazioni con ripetizioni di M elementi in classe K, per $K, M \in \mathbb{N}$. Similarmente al caso $K = 2$, nel caso $K = 3$ [risp. $K = 4$], qualora il momento centralizzato di X di ordine 2 sia invertibile, nelle applicazioni è più frequente l'uso della cosiddetta *skewness* [risp. *kurtosis*] di X, denotata con $Skew(X)$ [risp. $Kurt(X)$] e definita come il momento di ordine 3 [risp. 4] della standardizzazione di X. Formalmente,

$$Skew(X) \equiv \mathbb{M}'_3(Y) \stackrel{\text{def}}{=} \left(\mathbf{E}\left[Y_k Y_\ell Y_m\right]\right)_{k,\ell,m=1}^{M} \qquad \text{[risp. } Kurt(X) \equiv \mathbb{M}'_4(Y) \stackrel{\text{def}}{=} \left(\mathbf{E}\left[Y_j Y_k Y_\ell Y_m\right]\right)_{j,k,\ell,m=1}^{M}\text{]},$$

essendo

$$Y \equiv Var(X)^{-1/2}(X - \mathbf{E}[X]).$$

Sempre grazie alla finitezza di Ω, notiamo anche che lo spazio lineare reale di tutte le applicazioni $X : \Omega \to \mathbb{R}^M$, denotato con $F(\Omega; \mathbb{R}^M)$, ha dimensione finita pari a $|\Omega| \times M = 2^N M$[3] Inoltre, essendo ogni funzione di $F(\Omega; \mathbb{R}^M)$ una $\mathcal{E} - \mathcal{B}(\mathbb{R}^M)$ variabile aleatoria con momento finito di ogni ordine, in particolare di ordine 2, lo spazio $F(\Omega; \mathbb{R}^M)$ è identificabile con lo spazio di Hilbert $L^2(\Omega; \mathbb{R}^M)$ delle $\mathcal{E} - \mathcal{B}(\mathbb{R}^M)$ variabili aleatorie che hanno momento finito di ordine 2 dotato del prodotto scalare $\langle \cdot, \cdot \rangle : L^2(\Omega; \mathbb{R}^M) \times L^2(\Omega; \mathbb{R}^M) \to \mathbb{R}$, dato da

$$\langle X, Y \rangle \stackrel{\text{def}}{=} \sum_{\omega \in \Omega} X^\intercal(\omega) Y(\omega) \mathbf{P}(\omega), \quad \forall X, Y \in F(\Omega; \mathbb{R}^M), \tag{3.23}$$

essendo $X(\omega) \equiv (X_1(\omega), \ldots, X_M(\omega))^\intercal$, $Y(\omega) \equiv (Y_1(\omega), \ldots, Y_M(\omega))^\intercal$, ed essendo

$$X^\intercal(\omega) Y(\omega) \stackrel{\text{def}}{=} \sum_{m=1}^{M} X_m(\omega) Y_m(\omega), \quad \forall \omega \in \Omega, \tag{3.24}$$

il prodotto scalare standard su $\mathbb{R}^M$. Da notare che la norma $\|\cdot\| : L^2(\Omega; \mathbb{R}^M) \to \mathbb{R}_+$ indotta dal prodotto scalare $\langle X, Y \rangle$, definita come

$$\|X\| \stackrel{\text{def}}{=} \langle X, X \rangle^{1/2}, \quad \forall X \in L^2(\Omega; \mathbb{R}^M), \tag{3.25}$$

è tale che

$$\|X\| = \left(\sum_{\omega \in \Omega} X^\intercal(\omega) X(\omega) \mathbf{P}(\omega)\right)^{1/2} = \mathbf{E}\left[X^\intercal X\right]^{1/2}, \tag{3.26}$$

per ogni $X \in L^2(\Omega; \mathbb{R}^M)$, e che la distanza $d(\cdot, \cdot) : L^2(\Omega; \mathbb{R}^M) \times L^2(\Omega; \mathbb{R}^M) \to \mathbb{R}_+$ indotta dalla norma $\|\cdot\|$, definita come

$$d(X, Y) \stackrel{\text{def}}{=} \|X - Y\| \quad \forall XY \in L^2(\Omega; \mathbb{R}^M), \tag{3.27}$$

è tale che

$$d(X, Y) = \mathbf{E}\left[(X - Y)^\intercal (X - Y)\right]^{1/2}, \tag{3.28}$$

[3] Ogni applicazione $X : \Omega \to \mathbb{R}^M$ è identificabile con il vettore delle sue componenti $(X_1, \ldots, X_M)^\intercal$ ed, essendo Ω finito, ciascuna componente X_m è esprimibile come combinazione lineare del tipo

$$X_m(\omega) = \textstyle\sum_{o \in \Omega} X_m(\omega) E_{o,m}(\omega),$$

per ogni $\omega \in \Omega$, essendo $E_{o,m} : \Omega \to \mathbb{R}$ definita ponendo

$$E_{o,m}(\omega) \stackrel{\text{def}}{=} \begin{cases} 1, & \text{se } \omega = o, \\ 0, & \text{se } \omega \neq o, \end{cases}$$

per ogni $o \in \Omega$ ed $m \in \{1, \ldots, M\}$. Quindi, una base di $F(\Omega; \mathbb{R}^M)$ è costituita dalle $2^N M$ applicazioni $E_{o_1,\ldots,o_M} : \Omega \to \mathbb{R}^M$ definite ponendo

$$E_{o_1,\ldots,o_M}(\omega) \stackrel{\text{def}}{=} (E_{o_1,1}(\omega), \ldots, E_{o_M,M}(\omega)), \quad \forall \omega \in \Omega$$

al variare di $o_1, \ldots, o_M \in \Omega$.

per ogni $X, Y \in L^2(\Omega; \mathbb{R}^M)$. In particolare, quando $M = 1$, abbiamo

$$\langle X, Y \rangle = \sum_{\omega \in \Omega} X(\omega) Y(\omega) \mathbf{P}(\omega), \qquad \|X\| = \mathbf{E}\left[X^2\right]^{1/2}, \qquad d(X, Y) = \mathbf{E}\left[(X - Y)^2\right]^{1/2},$$

per ogni $X, Y \in L^2(\Omega; \mathbb{R})$.

Una volta introdotto lo spazio di probabilità CRR, $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$, per rappresentare l'incidenza sul prezzo dello stock della realizzazione dei possibili accadimenti introduciamo anche una successione di variabili aleatorie bernoulliane $(\beta_n)_{n=1}^N$ su Ω tali che

$$\beta_n(\Omega) \equiv \beta_n((\omega_k)_{k=1}^N) \stackrel{\text{def}}{=} \begin{cases} u, & \text{se } \omega_n = 1, \\ d, & \text{se } \omega_n = 0, \end{cases} \qquad \omega \in \Omega, \quad n = 1, \dots, N, \tag{3.29}$$

essendo

$$\mathbf{P}(\beta_n = u) \equiv p, \qquad \mathbf{P}(\beta_n = d) \equiv q, \quad n = 1, \dots, N. \tag{3.30}$$

Proposizione 149 *Le variabili aleatorie $\beta_1, \dots, \beta_N$ risultano essere (totalmente) indipendenti rispetto alla probabilità oggettiva $\mathbf{P} : \mathcal{E} \to \mathbb{R}_+$.*

Proof. *Dobbiamo provare che comunque considerato un sottoinsieme di indici $\{n_1, \dots, n_K\}$ dell'insieme $\{1, \dots, N\}$, con $K \leqslant N$, si ha*

$$\mathbf{P}(\beta_{n_1} \leqslant x_1, \dots, \beta_{n_K} \leqslant x_K) = \mathbf{P}(\beta_{n_1} \leqslant x_1) \cdots \mathbf{P}(\beta_{n_K} \leqslant x_K)$$

al variare di $(x_1, \dots, x_K) \in \mathbb{R}^K$. D'altra parte, per ogni $\{n_1, \dots, n_K\}$, il vettore aleatorio $(\beta_{n_1}, \dots, \beta_{n_K})$ può prendere come valori solo una delle possibili K-ple dell'insieme finito

$$\mathbb{S} \equiv \left\{(x_1, \dots, x_K) \in \mathbb{R}^K : x_k = u \vee x_k = d\right\}.$$

Pertanto $\beta_1, \dots, \beta_N$ sono totalmente indipendenti rispetto alla probabilità oggettiva $\mathbf{P} : \mathcal{E} \to \mathbb{R}_+$ se e solo se comunque considerato $\{n_1, \dots, n_K\}$ risulta

$$\mathbf{P}(\beta_{n_1} = x_1, \dots, \beta_{n_K} = x_K) = \mathbf{P}(\beta_{n_1} = x_1) \cdots \mathbf{P}(\beta_{n_K} = x_K), \tag{3.31}$$

per ogni $(x_1, \dots, x_K) \in \mathbb{S}$. Fissato quindi un sottoinsieme di indici $\{n_1, \dots, n_K\}$ dell'insieme $\{1, \dots, N\}$ e considerato il suo complementare $\{m_1, \dots, m_{N-K}\} \equiv \{1, \dots, N\} - \{n_1, \dots, n_K\}$, per ogni $(x_1, \dots, x_K) \in \mathbb{S}$ risulta

$$\begin{aligned}
&\{\beta_{n_1} = x_1, \dots, \beta_{n_K} = x_K\} \\
&= \left\{\omega \in \Omega : \omega_{n_k} = 1 \text{ [resp. 0]} \Leftrightarrow x_{n_k} = u \text{ [resp. d]},\ k = 1, \dots, K,\ \wedge\ \left(\omega_{m_1}, \dots, \omega_{m_{N-K}}\right) \in \{0, 1\}^{N-K}\right\} \\
&= \bigcup_{H=0}^{N-K} \{\omega \in \Omega : \omega_{n_k} = 1 \text{ [resp.0]} \Leftrightarrow x_{n_k} = u \text{ [resp. d]},\ k = 1, \dots, K,\ \wedge\ |\{m_h : \omega_{m_h} = 1\}| = H\}.
\end{aligned}$$

Quindi, posto $J = |\{k \mid x_k = u,\ k = 1, \ldots, K\}|$, otteniamo

$$
\begin{aligned}
&\mathbf{P}(\beta_{n_1} = x_1, \ldots, \beta_{n_K} = x_K) \\
&= \mathbf{P}\left(\bigcup_{H=0}^{N-K} \{\omega \in \Omega : \omega_{n_k} = 1 \text{ [resp. } 0] \Leftrightarrow x_{n_k} = u \text{ [resp. } d],\ k = 1, \ldots, K,\ \wedge\ |\{m_h : \omega_{m_h} = 1\}| = H\}\right) \\
&= \sum_{H=0}^{N-K} \mathbf{P}\left(\{\omega \in \Omega : \omega_{n_k} = 1 \text{ [resp. } 0] \Leftrightarrow x_{n_k} = u \text{ [resp. } d],\ k = 1, \ldots, K,\ \wedge\ |\{m_h : \omega_{m_h} = 1\}| = H\}\right) \\
&= \sum_{H=0}^{N-K} \binom{N-K}{H} p^H q^{N-K-H} p^J q^{K-J} \\
&= p^J q^{K-J} \sum_{H=0}^{N-K} \binom{N-K}{H} p^H q^{N-K-H} \\
&= p^J q^{K-J}.
\end{aligned}
$$

D'altra parte

$$\mathbf{P}(\beta_{n_1} = x_1) \cdots \mathbf{P}(\beta_{n_K} = x_K) = p^J q^{K-J}.$$

Pertanto sussiste la (3.31), e ciò prova l'asserto. □

Sia $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ lo spazio di probabilitá CRR.

Definizione 150 *Diciamo che una famiglia $(\mathcal{F}_n)_{n=0}^N$ di sotto-sigma-algebre di $\mathcal{E}$ é una filtrazione di Ω se*

$$\mathcal{F}_{n-1} \subseteq \mathcal{F}_n, \quad \forall n = 1, \ldots, N. \tag{3.32}$$

La nozione di filtrazione viene introdotta per rappresentare il flusso dell'informazione su un fenomeno stocastico, caratterizzanto da eventi che si rivelano progressivamente nel tempo. L'equazione 3.32 vuole proprio modellare l'idea che l'informazione disponibile a un osservatore di un fenomeno stocastico si accumuli nel tempo senza dimenticanza. In particolare, il verificarsi progressivo degli eventi che caratterizzano il fenomeno stocastico presentato nel modello CRR, viene rappresentato dalla filtrazione $(\mathcal{F}_n)_{n=0}^N$ tale che

$$\mathcal{F}_0 \stackrel{\text{def}}{=} \{\varnothing, \Omega\}, \qquad \mathcal{F}_n \stackrel{\text{def}}{=} \sigma(\beta_1, \ldots, \beta_n), \quad n = 1, \ldots, N, \tag{3.33}$$

essendo $\sigma(\beta_1, \ldots, \beta_n)$ la σ-algebra generata dalle variabili aleatorie $\beta_1, \ldots, \beta_n$[4].

Definizione 151 *Chiamiamo la filtrazione $(\mathcal{F}_n)_{n=0}^N$ definita dalla Equazione 3.33* filtrazione generata dal processo di Bernoulli.

Non è difficile rendersi conto che

$$\mathcal{F}_1 \stackrel{\text{def}}{=} \sigma(\beta_1) = \{\varnothing, \Omega, E_0, E_1\},$$

dove

$$E_0 \equiv \{\omega \in \Omega : \omega_1 = 0\} \quad \text{e} \quad E_1 \equiv \{\omega \in \Omega : \omega_1 = 1\}.$$

Inoltre,

$$\mathcal{F}_2 \stackrel{\text{def}}{=} \sigma(\beta_1, \beta_2) = \sigma(E_{0,0}, E_{0,1}, E_{1,0}, E_{1,1}),$$

[4] La più piccola σ-algebra di eventi rispetto a cui tutte le funzioni $\beta_1, \ldots, \beta_n$ sono variabili aleatorie, ovvero, la minima informazione che rende osservabili i valori assunti da tutte le funzioni $\beta_1, \ldots, \beta_n$

dove

$$E_{0,0} \equiv \{\omega \in \Omega : \omega_1 = 0 \wedge \omega_2 = 0\}, \quad E_{0,1} \equiv \{\omega \in \Omega : \omega_1 = 0 \wedge \omega_2 = 1\},$$
$$E_{1,0} \equiv \{\omega \in \Omega : \omega_1 = 1 \wedge \omega_2 = 0\}, \quad E_{1,1} \equiv \{\omega \in \Omega : \omega_1 = 1 \wedge \omega_2 = 1\},$$

e $\sigma(E_{0,0}, E_{0,1}, E_{1,0}, E_{1,1})$ è la σ-algebra generata dalla famiglia di eventi $\{E_{0,0}, E_{0,1}, E_{1,0}, E_{1,1}\}$[5]. In dettaglio,

$$\sigma(E_{0,0}, E_{0,1}, E_{1,0}, E_{1,1})$$
$$= \{\varnothing, E_{0,0}, E_{0,1}, E_{1,0}, E_{1,1}, E_0, E_{0,0} \cup E_{1,0}, E_{0,0} \cup E_{1,1}, E_{0,1} \cup E_{1,0}, E_{0,1} \cup E_{1,1}, E_1, E_{1,1}^c, E_{0,1}^c, E_{1,0}^c, E_{0,0}^c, \Omega\},$$

essendo

$$E_0 = E_{0,0} \cup E_{0,1} \quad \text{e} \quad E_1 = E_{1,0} \cup E_{1,1}.$$

Infatti, poichè $\{E_{0,0}, E_{0,1}, E_{1,0}, E_{1,1}\}$ una partizione finita di Ω è noto che $\sigma(E_{0,0}, E_{0,1}, E_{1,0}, E_{1,1})$ contiene tutti e soli gli eventi E rappresentabili come

$$E = \bigcup_{h \in H} E_h,$$

al variare di H tra tutti i sottoinsiemi dell'insieme di indici $J \equiv \{(0,0), (0,1), (1,0), (1,1)\}$[6]. Ora, l'insieme delle parti $\mathcal{P}(J)$ contiene $2^4 = 16$ elementi. Più specificatamente,

$$\mathcal{P}(J) = \{H_n\}_{n=1}^{16},$$

dove

$$H_1 \equiv \varnothing, \quad H_2 \equiv \{(0,0)\}, \quad H_3 \equiv \{(0,1)\}, \quad H_4 \equiv \{(1,0)\}, \quad H_5 \equiv \{(1,1)\},$$
$$H_6 \equiv \{(0,0), (0,1)\}, \quad H_7 \equiv \{(0,0), (1,0)\}, \quad H_8 \equiv \{(0,0), (1,1)\},$$
$$H_9 \equiv \{(0,1), (1,0)\}, \quad H_{10} \equiv \{(0,1), (1,1)\}, \quad H_{11} \equiv \{(0,1), (1,1)\},$$
$$H_{12} \equiv \{(0,0), (0,1), (1,0)\}, \quad H_{13} \equiv \{(0,0), (1,0), (1,1)\},$$
$$H_{14} \equiv \{(0,0), (0,1), (1,1)\} \quad H_{15} \equiv \{(0,1), (1,0), (1,1)\},$$
$$H_{16} \equiv \{(0,0), (0,1), (1,0), (1,1)\}$$

Quindi,

$$\bigcup_{h \in H_1} E_h = \varnothing, \quad \bigcup_{h \in H_2} E_h = E_{0,0}, \quad \bigcup_{h \in H_3} E_h = E_{0,1}, \quad \bigcup_{h \in H_4} E_h = E_{1,0}, \quad \bigcup_{h \in H_5} E_h = E_{1,1},$$

$$\bigcup_{h \in H_6} E_h = E_0, \quad \bigcup_{h \in H_7} E_h = E_{0,0} \cup E_{1,0}, \quad \bigcup_{h \in H_8} E_h = E_{0,0} \cup E_{1,1},$$

$$\bigcup_{h \in H_9} E_h = E_{0,1} \cup E_{1,0}, \quad \bigcup_{h \in H_{10}} E_h = E_{0,1} \cup E_{1,1}, \quad \bigcup_{h \in H_{11}} E_h = E_1,$$

[5] La più piccola σ-algebra di eventi contenente la famiglia $\{E_{0,0}, E_{0,1}, E_{1,0}, E_{1,1}\}$.

[6] In generale, la σ-algebra dei sottoinsiemi di un insieme $\mathbb{X}$ generata da una partizione numerabile $\{\mathbb{P}_j\}_{j \in J}$ di $\mathbb{X}$ si caratterizza come la famiglia di tutti e soli i sottoinsiemi di $\mathbb{X}$ che si ottengono come unioni dei sottoinsiemi della partizione scelti in corrispondenza ad ogni possibile sottoinsieme dell'insieme J indicizzante la partizione. Formalmente,

$$\sigma\left(\{\mathbb{P}_j\}_{j \in J}\right) = \left\{\mathbb{S} \subseteq \mathbb{X} : \mathbb{S} = \bigcup_{h \in H} \mathbb{P}_h, \quad \forall H \in \mathcal{P}(J)\right\}$$

essendo $\mathcal{P}(J)$ l'insieme delle parti di J.

$$\bigcup_{h\in H_{12}} E_h = E_{1,1}^c, \quad \bigcup_{h\in H_{13}} E_h = E_{0,1}^c, \quad \bigcup_{h\in H_{14}} E_h = E_{1,0}^c, \quad \bigcup_{h\in H_{15}} E_h = E_{0,0}^c, \quad \bigcup_{h\in H_{16}} E_h = \Omega.$$

Più in generale, $\mathcal{F}_n$ rappresenta la famiglia di eventi osservabili alla luce della realizzazione delle variabili aleatorie $\beta_1, \dots, \beta_n$ e, stante la realizzazione di tali variabili aleatorie, possiamo distinguere tra successioni di accadimenti che differiscono solo sui primi n termini. Quindi, un evento $E \in \mathcal{F}_n$ se e solo se, comunque considerato un punto campionario $\hat{\omega} \equiv (\hat{\omega}_k)_{k=1}^N \in E$, ogni altro punto campionario $\omega \equiv (\omega_k)_{k=1}^N$ tale che $\omega_k = \hat{\omega}_k$ per ogni $k = 1, \dots, n$ deve essere anch'esso un elemento di E indipendentemente dai valori assunti dai restanti termini $\omega_{n+1}, \dots, \omega_N$. Ad esempio, se un evento $E \in \mathcal{F}_3$ contenesse il punto campionario $(1,0,0,0,0,0,\dots)$, allora dovrebbe contenere anche tutti gli altri punti campionari del tipo $(1,0,0,0,0,1,\dots)$, $(1,0,0,0,1,0,\dots)$, $(1,0,0,0,1,1,\dots)$, $(1,0,0,1,0,0,\dots)$, $(1,0,0,1,0,1,\dots)$, $(1,0,0,1,1,0,\dots)$, $(1,0,0,1,1,1,\dots)$, ecc... ottenuti da $(1,0,0,1,0,0,\dots)$ ripetendo i suoi primi tre termini e modificando i termini dal quarto in poi in tutti i modi possibili.

Osservazione 152 *Si ha*

$$\mathcal{E} = \mathcal{F}_N \equiv \mathcal{F}_T \tag{3.34}$$

Definizione 153 *Chiamiamo* processo stocastico *su Ω a stati in $\mathbb{R}^M$, per un qualche $M \in \mathbb{N}$, una qualsiasi successione $(X_n)_{n=0}^N$ di $\mathcal{E}$-variabili aleatorie reali M-variate su Ω.*

Definizione 154 *Diciamo che $(X_n)_{n=0}^N$, processo stocastico su Ω a stati in $\mathbb{R}^M$, è di ordine K, per $K \in \mathbb{N}$, se tutte le variabili aleatorie X_n del processo hanno momento finito di ordine K.*

Poichè nel modello CRR Ω è finito e $\mathcal{E} \equiv \mathcal{P}(\Omega)$, abbiamo

Osservazione 155 *Una qualsiasi successione $(X_n)_{n=0}^N$ di funzioni $X_n : \Omega \to \mathbb{R}^M$, al variare di $n = 0, 1, \dots, N$, è un processo stocastico su Ω a stati in $\mathbb{R}^M$ di ordine K per ogni $K \in \mathbb{N}$.*

Definizione 156 *Il processo stocastico su Ω a stati in $\mathbb{R}$ costituito dalla successione di variabili aleatorie bernoulliane indipendenti $(\beta_n)_{n=1}^N$ (cfr. Equazioni 3.29 e 3.30) è noto come* processo di Bernoulli *con parametro di successo p.*

Definizione 157 *Il processo stocastico su Ω a stati in $\mathbb{R}$ costituito dalla successione di variabili aleatorie $(N_n)_{n=1}^N$ definita ponendo*

$$N_0 \stackrel{def}{=} 0, \quad N_n \stackrel{def}{=} \sum_{k=1}^{n} \frac{\beta_k - d}{u - d}, \quad \forall n = 1, \dots, N$$

è noto come processo di conteggio *del processo di Bernoulli.*

Osservazione 158 *La variabile aleatoria N_n ha distribuzione binomiale standard con parametro numero di tentativi n e parametro di successo p, per ogni $n = 1, \dots, N$. Quindi,*

$$N_n(\omega) = k \quad e \quad \mathbf{P}(N_n = k) = \binom{n}{k} p^k q^{n-k},$$

per ogni $n = 1, \dots, N$, ogni $k = 0, 1, \dots, n$ e al variare di $\omega \in \Omega$

Osservazione 159 *La filtrazione $(\mathcal{F}_n)_{n=0}^N$ generata dal processo di Bernoulli è la più piccola filtrazione sullo spazio di probabilità $(\Omega, \mathcal{E}, \mathbf{P})$ tale che β_n sia una $\mathcal{F}_n - \mathcal{B}(\mathbb{R})$ variabile aleatoria, per ogni $n = 1, \dots, N$.*

Sia $(\mathcal{F}_n)_{n=0}^N \equiv \mathfrak{F}$ una filtrazione sullo spazo di probabilitá CRR, $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$, e sia $(X_n)_{n=0}^N$ un processo stocastico su Ω a stati in $\mathbb{R}^M$, per un qualche $M \in \mathbb{N}$.

Definizione 160 *Diciamo che* $(X_n)_{n=0}^N$ è $\mathfrak{F}$-adattato *se* X_n *è una* $\mathcal{F}_n - \mathcal{B}\left(\mathbb{R}^M\right)$ *variabile aleatoria, per ogni* $n = 0, 1, \ldots, N$.

Definizione 161 *Diciamo che* $(X_n)_{n=0}^N$ è $\mathfrak{F}$-predicibile *se* X_n *è una* $\mathcal{F}_{n-1} - \mathcal{B}\left(\mathbb{R}^M\right)$ *variabile aleatoria, per ogni* $n = 1, \ldots, N$.

Chiaramente

Osservazione 162 *Se* $(X_n)_{n=1}^N$ *è* $\mathfrak{F}$*-predicibile allora* $(X_n)_{n=1}^N$ *è* $\mathfrak{F}$*-adattato. Il viceversa non è vero.*

Sia $\mathfrak{F} \equiv (\mathcal{F}_n)_{n=0}^N$ la filtrazione su $(\Omega, \mathcal{E}, \mathbf{P})$ generata dal processo di Bernoulli.

Osservazione 163 *Il processo di Bernoulli* $(\beta_n)_{n=1}^N$ *e il processo di conteggio del processo di Bernoulli* $(N_n)_{n=0}^N$ *sono processi* $\mathfrak{F}$*-adattati.*

Supponiamo ora che la dinamica del prezzo dello stock, a partire dal prezzo iniziale $S_0 > 0$, sia rappresentata dalla successione $(S_n)_{n=0}^N$ definita ponendo

$$S_n \stackrel{\text{def}}{=} \beta_n S_{n-1}, \quad n = 1, \ldots, N. \tag{3.35}$$

Osservazione 164 *Si ha equivalentemente*

$$S_n = \beta_n \cdots \beta_1 S_0, \tag{3.36}$$

per ogni $n = 1, \ldots, N$, *e anche*

$$S_n = \beta_n \cdots \beta_{m+1} S_m, \tag{3.37}$$

per tutti gli $m, n \in \{1, \ldots, N\}$ *tali che* $m < n$.

Osservazione 165 *Si ha anche*

$$S_n = u^{N_n} d^{n-N_n} S_0. \tag{3.38}$$

per ogni $n = 1, \ldots, N$.

A titolo d'esempio, osserviamo che si ha

$$S_1 = \begin{cases} uS_0, & \mathbf{P}(S_1 = uS_0) = \mathbf{P}(\beta_1 = u) \equiv p, \\ dS_0, & \mathbf{P}(S_1 = dS_0) = \mathbf{P}(\beta_1 = d) \equiv q, \end{cases}$$

e ancora

$$S_2 = \begin{cases} uS_1, & \mathbf{P}(S_2 = uS_1) = \mathbf{P}(\beta_2 = u) \equiv p, \\ dS_1, & \mathbf{P}(S_2 = dS_1) = \mathbf{P}(\beta_2 = d) \equiv q, \end{cases} = \begin{cases} u^2 S_0, & \mathbf{P}(S_2 = u^2 S_0) = p^2, \\ udS_0, & \mathbf{P}(S_2 = udS_0) = 2pq, \\ d^2 S_0, & \mathbf{P}(S_2 = d^2 S_0) = q^2. \end{cases}$$

Infatti, per l'indipendenza di β_1 e β_2, risulta

$$uS_1 = \begin{cases} u^2 S_0, & \mathbf{P}(S_2 = u^2 S_0) = \mathbf{P}(\beta_1 = u, \beta_2 = u) = \mathbf{P}(\beta_1 = u)\mathbf{P}(\beta_2 = u) = p^2, \\ udS_0, & \mathbf{P}(S_2 = udS_0) = \mathbf{P}(\beta_1 = u, \beta_2 = d) = \mathbf{P}(\beta_1 = u)\mathbf{P}(\beta_2 = d) = pq, \end{cases}$$

e

$$dS_1 = \begin{cases} duS_0, & \mathbf{P}(S_2 = duS_0) = \mathbf{P}(\beta_1 = d, \beta_2 = u) = \mathbf{P}(\beta_1 = d)\mathbf{P}(\beta_2 = u) = qp, \\ d^2 S_0, & \mathbf{P}(S_2 = d^2 S_0) = \mathbf{P}(\beta_1 = d, \beta_2 = d) = \mathbf{P}(\beta_1 = d)\mathbf{P}(\beta_2 = d) = q^2. \end{cases}$$

Più in generale si ha

Proposizione 166 *Al variare di $\omega \in \Omega$ i possibili prezzi dello stock al tempo $t = t_n$ sono dati da*

$$S_n(\omega) = u^k d^{n-k} S_0, \quad \mathbf{P}(S_n = u^k d^{n-k} S_0) = \binom{n}{k} p^k q^{n-k}, \tag{3.39}$$

per ogni $n = 1, \ldots, N$ e ogni $k = 0, 1, \ldots, n$.
Proof. *E' sufficiente osservare che la (3.39) caratterizza l'insieme dei valori e la relativa distribuzione di probabilità di una variabile aleatoria binomiale che rappresenta il verificarsi di k successi in n prove indipendenti, essendo p [risp. q] la probabilità di successo [risp. fallimento] in una singola prova.* □

Proposizione 167 *Più in generale, al variare di $\omega \in \Omega$, si ha*

$$S_n(\omega) = u^k d^{n-m-k} S_m(\omega), \quad \mathbf{P}\left(S_n = u^k d^{n-m-k} S_m\right) = \binom{n-m}{k} p^k q^{n-m-k}, \tag{3.40}$$

per tutti gli $m, n \in \{1, \ldots, N\}$ tali che $m < n$ e ogni $k = 0, 1, \ldots, n-m$.

A titolo d'esempio, osserviamo che, al variare di $\omega \in \Omega$, abbiamo

$$S_n(\omega) = \begin{cases} uS_{n-1}(\omega), & \mathbf{P}(S_n = uS_{n-1}) = p, \\ dS_{n-1}(\omega), & \mathbf{P}(S_n = dS_{n-1}) = q, \end{cases}$$

per ogni $n = 1, \ldots, N$, e ancora

$$S_n(\omega) = \begin{cases} u^2 S_{n-2}(\omega), & \mathbf{P}(S_n = u^2 S_{n-2}) = p^2, \\ udS_{n-2}(\omega), & \mathbf{P}(S_n = udS_{n-2}) = 2pq, \\ d^2 S_{n-2}(\omega), & \mathbf{P}(S_n = d^2 S_{n-2}) = q^2, \end{cases}$$

per ogni $n = 2, \ldots, N$, e al variare di $\omega \in \Omega$.

Proposizione 168 *L'attesa e la varianza di S_n sono rispettivamente date da*

$$\mathbf{E}[S_n] = (up + dq)^n S_0, \tag{3.41}$$

e

$$\mathbf{D}^2[S_n] = ((u^2 p + d^2 q)^n - (up + dq)^{2n}) S_0^2, \tag{3.42}$$

per ogni $n = 0, 1, \ldots, N$.
Proof. *A norma di definizione, risulta*

$$\mathbf{E}[S_n] = \sum_{k=0}^{n} u^k d^{n-k} S_0 \binom{n}{k} p^k q^{n-k} = S_0 \sum_{k=0}^{n} \binom{n}{k} (up)^k (dq)^{n-k} = (up + dq)^n S_0.$$

Si ha poi

$$S_n^2(\omega) = u^{2k} d^{2(n-k)} S_0^2, \quad \mathbf{P}(S_n = u^{2k} d^{2(n-k)} S_0^2) = \binom{n}{k} p^k q^{n-k},$$

per ogni $n = 0, 1, \ldots, N$ e ogni $k = 0, 1, \ldots, N$. Quindi,

$$\mathbf{E}\left[S_n^2\right] = \sum_{k=0}^{n} u^{2k} d^{2(n-k)} S_0^2 \binom{n}{k} p^k q^{n-k} = S_0^2 \sum_{k=0}^{n} \binom{n}{k} (u^2 p)^k (d^2 q)^{n-k} = (u^2 p + d^2 q)^n S_0^2.$$

Pertanto,

$$\mathbf{D}^2[S_n] = \mathbf{E}\left[S_n^2\right] - \mathbf{E}[S_n]^2 = ((u^2 p + d^2 q)^n - (up + dq)^{2n}) S_0^2.$$

□

Come immediata conseguenza dell'Equazione 3.40 otteniamo

Corollary 169 *Risulta*

$$\mathbf{E}\left[S_n\right] = (up + dq)\mathbf{E}\left[S_{n-1}\right],$$

per ogni $n = 1, \ldots, N$.

Proposizione 170 *La successione* $(S_n)_{n=0}^N$ *è un processo stocastico* $\mathfrak{F}$*-adattato.*
Proof. *Chiaramente* S_0*, interpretabile come una variabile aleatoria di Dirac centrata in* S_0*, è una* $\mathcal{F}_0 - \mathcal{B}(\mathbb{R})$ *variabile aleatoria. Inoltre, per la (3.33) e la (3.36),* S_n *è una* $\mathcal{F}_n - \mathcal{B}(\mathbb{R})$ *variabile aleatoria, per ogni* $n = 1, \ldots, N$. □

In un contesto multiperiodale, vanno introdotti i sottospazi $L^2\left(\Omega_{\mathcal{F}_n}; \mathbb{R}^M\right) \equiv L^2\left(\Omega_n; \mathbb{R}^M\right)$ dello spazio di Hilbert $L^2\left(\Omega; \mathbb{R}^M\right)$ costituiti da tutte le $\mathcal{F}_n - \mathcal{B}\left(\mathbb{R}^M\right)$ variabili aleatorie reali M-variate su Ω che hanno momento finito di ordine 2, al variare di $n = 0, 1, \ldots, N$. Da notare che in riferimento allo spazio di probabilità CRR la condizione di momento di ordine 2 finito non è caratterizzante (cfr. Remark 148). Invece, è caratterizzante l'essere $\mathcal{F}_n - \mathcal{B}\left(\mathbb{R}^M\right)$ variabile aleatoria, per un qualche $n \in \{0, 1, \ldots, N\}$, ossia godere della proprietà

$$\{X \in B\} \in \mathcal{F}_n, \quad \forall B \in \mathcal{B}\left(\mathbb{R}^M\right). \tag{3.43}$$

Quindi, mentre in riferimento allo spazio di probabilità CRR, lo spazio di Hilbert $L^2\left(\Omega; \mathbb{R}^M\right)$ è di fatto lo spazio di tutte le funzioni $X : \Omega \to \mathbb{R}^M$, lo spazio $L^2\left(\Omega_n; \mathbb{R}^M\right)$ è lo spazio lineare di tutte le funzioni $X : \Omega \to \mathbb{R}^M$ che soddisfano la (3.43). In più, lo spazio $L^2\left(\Omega_n; \mathbb{R}^M\right)$ è un chiuso della topologia di $L^2\left(\Omega; \mathbb{R}^M\right)$[7]. Pertanto $L^2\left(\Omega_n; \mathbb{R}^M\right)$ risulta essere un sottospazio di $L^2\left(\Omega; \mathbb{R}^M\right)$. Nel contesto multiperiodale assume particolare importanza il ruolo dell'operatore $\mathbf{E}[\cdot \mid \mathcal{F}_n] : L^2\left(\Omega; \mathbb{R}^M\right) \to L^2\left(\Omega_{\mathcal{F}_n}; \mathbb{R}^M\right)$, al variare di $n = 0, 1, \ldots, N$, noto come speranza condizionata rispetto a $\mathbf{P}$ data la σ-algebra $\mathcal{F}_n$ (l'informazione $\mathcal{F}_n$ disponibile al tempo $t = t_n$) e indicato anche più brevemente con il simbolo $\mathbf{E}_n[\cdot]$, che trasforma $\mathcal{E} - \mathcal{B}\left(\mathbb{R}^M\right)$ variabili aleatorie in $\mathcal{F}_n - \mathcal{B}\left(\mathbb{R}^M\right)$ variabili aleatorie, caratterizzate come la migliore approssimazione delle prime nel senso dei minimi quadrati. Formalmente,

$$\mathbf{E}_n[X] \stackrel{\text{def}}{=} \arg\min\left\{\mathbf{E}\left[(X - Y)^\intercal (X - Y)\right] : Y \in L^2\left(\Omega_{\mathcal{F}_n}; \mathbb{R}^M\right)\right\}, \quad \forall X \in L^2\left(\Omega; \mathbb{R}^M\right).$$

Notare che

$$\mathbf{E}_0[X] = \mathbf{E}[X],$$

per ogni $X \in L^2\left(\Omega; \mathbb{R}^M\right)$.

Più in generale dato uno spazio di probabilità $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$, data una σ-algebra $\mathcal{F} \subseteq \mathcal{E}$ e considerato lo spazio di probabilità $\left(\Omega, \mathcal{F}, \mathbf{P}_{|\mathcal{F}}\right) \equiv \Omega_{\mathcal{F}}$, dove $\mathbf{P}_{|\mathcal{F}} : \mathcal{F} \to \mathbb{R}_+$ data da

$$\mathbf{P}_{|\mathcal{F}}(F) \stackrel{\text{def}}{=} \mathbf{P}(F), \quad \forall F \in \mathcal{F},$$

è la restrizione ad $\mathcal{F}$ della probabilità $\mathbf{P} : \mathcal{E} \to \mathbb{R}_+$, l'operatore $\mathbf{E}[\cdot \mid \mathcal{F}]$, speranza condizionata rispetto a $\mathbf{P}$ data $\mathcal{F}$, trasforma lo spazio di Hilbert $L^2\left(\Omega; \mathbb{R}^M\right)$ delle $\mathcal{E} - \mathcal{B}\left(\mathbb{R}^M\right)$ variabili aleatorie aventi momento finito di ordine 2 nel suo sottospazio $L^2\left(\Omega_{\mathcal{F}}; \mathbb{R}^M\right)$ delle $\mathcal{F} - \mathcal{B}\left(\mathbb{R}^M\right)$ variabili aleatorie. Tale operatore è definito ponendo

$$\mathbf{E}[X \mid \mathcal{F}] \stackrel{\text{def}}{=} \arg\min\left\{\mathbf{E}\left[(X - Y)^\intercal (X - Y)\right] : \ Y \in L^2\left(\Omega_{\mathcal{F}}; \mathbb{R}^M\right)\right\}, \quad \forall X \in L^2\left(\Omega; \mathbb{R}^M\right).$$

L'idea è che mediante l'operatore $\mathbf{E}[\cdot \mid \mathcal{F}] : L^2\left(\Omega; \mathbb{R}^M\right) \to L^2\left(\Omega_{\mathcal{F}}; \mathbb{R}^M\right)$ determiniamo la migliore approssimazione, nel senso dei minimi quadrati, di una $\mathcal{E} - \mathcal{B}\left(\mathbb{R}^M\right)$ variabile aleatoria X di cui non possiamo osservare le realizzazioni, stante una riduzione dell'informazione da $\mathcal{E}$ ad $\mathcal{F}$, mediante una $\mathcal{F} - \mathcal{B}\left(\mathbb{R}^M\right)$ variabile aleatoria $\mathbf{E}[X \mid \mathcal{F}]$ di cui possiamo osservare le realizzazioni alla luce dell'informazione ridotta.

[7] Il limite di ogni successione di $\mathcal{F}_n - \mathcal{B}\left(\mathbb{R}^M\right)$ variabili aleatorie convergente in $L^2\left(\Omega; \mathbb{R}^M\right)$ è esso stesso una $\mathcal{F}_n - \mathcal{B}\left(\mathbb{R}^M\right)$ variabile aleatoria

Proposizione 171 *In generale, l'operatore speranza condizionata* $\mathbf{E}[\cdot \mid \mathcal{F}] : L^2(\Omega; \mathbb{R}^M) \to L^2(\Omega_{\mathcal{F}}; \mathbb{R}^M)$ *gode delle seguenti proprietà:*

1. *si ha*
$$\int_F \mathbf{E}[X \mid \mathcal{F}]\, d\mathbf{P}_{|\mathcal{F}} = \int_F X d\mathbf{P},$$
per ogni $X \in L^2(\Omega; \mathbb{R}^M)$ *ed ogni* $F \in \mathcal{F}$[8], *in particolare*
$$\mathbf{E}[\mathbf{E}[X \mid \mathcal{F}]] = \mathbf{E}[X],$$
per ogni $X \in L^2(\Omega; \mathbb{R}^M)$*;*

2. *si ha*
$$\mathbf{E}[\alpha X + \beta Y \mid \mathcal{F}] = \alpha \mathbf{E}[X \mid \mathcal{F}] + \beta \mathbf{E}[Y \mid \mathcal{F}],$$
per ogni $\alpha, \beta \in \mathbb{R}$ *e ogni* $X, Y \in L^2(\Omega; \mathbb{R}^M)$*;*

3. *se* $X \in L^2(\Omega_{\mathcal{F}}; \mathbb{R}^M)$*, allora*
$$\mathbf{E}[X \mid \mathcal{F}] = X;$$

4. *se* $X \in L^2(\Omega_{\mathcal{F}}; \mathbb{R}^M)$ *e* $Y \in L^2(\Omega; \mathbb{R}^N)$*, allora*
$$\mathbf{E}[XY^\intercal \mid \mathcal{F}] = X\mathbf{E}[Y^\intercal \mid \mathcal{F}]^{9},$$
analogamente se $X \in L^2(\Omega; \mathbb{R}^M)$ *e* $Y \in L^2(\Omega_{\mathcal{F}}; \mathbb{R}^N)$ *allora*
$$\mathbf{E}[XY^\intercal \mid \mathcal{F}] = \mathbf{E}[X \mid \mathcal{F}]\, Y^\intercal;$$

5. *se* X *è indipendente da* $\mathcal{F}$*, allora*
$$\mathbf{E}[X \mid \mathcal{F}] = E[X];$$

6. *se* $\mathcal{G} \subseteq \mathcal{F}$*, allora*
$$\mathbf{E}[\mathbf{E}[X \mid \mathcal{F}] \mid \mathcal{G}] = \mathbf{E}[\mathbf{E}[X \mid \mathcal{G}] \mid \mathcal{F}] = \mathbf{E}[X \mid \mathcal{G}];$$

7. *si ha*
$$\mathbf{E}[\phi(X) \mid \mathcal{F}] \geqslant \phi(\mathbf{E}[X \mid \mathcal{F}]),$$
per ogni $X \in L^2(\Omega; \mathbb{R}^M)$ *ed ogni funzione convessa* $\phi : \mathbb{R}^M \to \mathbb{R}$ *such that* $\phi \circ X \in L^2(\Omega; \mathbb{R})$*, in particolare,*
$$\mathbf{E}[|X| \mid \mathcal{F}] \geqslant |\mathbf{E}[X \mid \mathcal{F}]|,$$
per ogni $X \in L^2(\Omega; \mathbb{R})$ *e* $\mathbf{E}[X \mid \mathcal{F}] \geqslant 0$*, per ogni* $X \in L^2(\Omega; \mathbb{R})$ *tale che* $X \geqslant 0$*.*

Corollary 172 *Risulta*
$$Var(\mathbf{E}[X \mid \mathcal{F}]) = \mathbf{E}[\mathbf{E}[X \mid \mathcal{F}]\, \mathbf{E}[X \mid \mathcal{F}]^\intercal] - \mathbf{E}[X]\, \mathbf{E}[X^\intercal]. \tag{3.44}$$
per ogni $X \in L^2(\Omega; \mathbb{R}^M)$*. In conseguenza,*
$$\mathbf{D}^2[\mathbf{E}[X \mid \mathcal{F}]] \leqslant \mathbf{D}^2[X], \tag{3.45}$$

[8] Si può provare che questa proprietà caratterizza $\mathbf{E}[X \mid \mathcal{F}]$, che risulta essere l'unica variabile aleatoria in $L^2(\Omega_{\mathcal{F}}; \mathbb{R}^M)$ che la soddisfa.

[9] Notare che nelle ipotesi considerate si ha $XY^\intercal \in L^1(\Omega; \mathbb{R}^{M \times N})$ ma, in generale $XY^\intercal \notin L^2(\Omega; \mathbb{R}^{M \times N})$. Quindi, stante la nostra definizione dell'operatore speranza condizionata, ai fini del risultato presentato sembrebbe necessario aggiungere l'ipotesi $XY^\intercal \in L^2(\Omega; \mathbb{R}^{M \times N})$. Tuttavia, è possibile provare che l'operatore speranza condizionata può essere esteso a $L^1(\Omega; \mathbb{R}^{M \times N})$ in modo tale il risultato presentato continui a essere valido senza l'ipotesi aggiuntiva.

per ogni $X \in L^2(\Omega;\mathbb{R})$ *tale che* $X^2 \in L^2(\Omega;\mathbb{R})$.
Proof. *Stante la definizione di varianza, applicando la Proprietà 1, otteniamo*

$$\begin{aligned} Var\left(\mathbf{E}\left[X \mid \mathcal{F}\right]\right) &= \mathbf{E}\left[\mathbf{E}\left[X \mid \mathcal{F}\right]\mathbf{E}\left[X \mid \mathcal{F}\right]^{\intercal}\right] - \mathbf{E}\left[\mathbf{E}\left[X \mid \mathcal{F}\right]\right]\mathbf{E}\left[\mathbf{E}\left[X \mid \mathcal{F}\right]\right]^{\intercal} \\ &= \mathbf{E}\left[\mathbf{E}\left[X \mid \mathcal{F}\right]\mathbf{E}\left[X \mid \mathcal{F}\right]^{\intercal}\right] - \mathbf{E}\left[X\right]\mathbf{E}\left[X\right]^{\intercal}, \end{aligned}$$

ossia l'Equazione (3.44). Ora, stante l'ipotesi $X^2 \in L^2(\Omega;\mathbb{R})$*, grazie alla Proprietà 7, possiamo scrivere*

$$\mathbf{E}\left[X^2 \mid \mathcal{F}\right] \geqslant \mathbf{E}\left[X \mid \mathcal{F}\right]^2.$$

Applicando la Proprietà 1 e la monotonia dell'operatore speranza, ottenamo allora

$$\mathbf{E}\left[X^2\right] = \mathbf{E}\left[\mathbf{E}\left[X^2 \mid \mathcal{F}\right]\right] \geqslant \mathbf{E}\left[\mathbf{E}\left[X \mid \mathcal{F}\right]^2\right].$$

Infine, combinando quest'ultima con la (3.44) e tenendo sempre conto della Proprietà 1, risulta

$$\mathbf{D}^2\left[X\right] = \mathbf{E}\left[X^2\right] - \mathbf{E}\left[X\right]^2 \geqslant \mathbf{E}\left[\mathbf{E}\left[X \mid \mathcal{F}\right]^2\right] - \mathbf{E}\left[X\right]^2 = \mathbf{E}\left[\mathbf{E}\left[X \mid \mathcal{F}\right]^2\right] - \mathbf{E}\left[\mathbf{E}\left[X \mid \mathcal{F}\right]\right]^2 = \mathbf{D}^2\left[\mathbf{E}\left[X \mid \mathcal{F}\right]\right],$$

e anche l'Equazione (3.45) è provata. □

Proposizione 173 *Si ha*

$$Cov\left(X - \mathbf{E}\left[X \mid \mathcal{F}\right], Y\right) = 0, \tag{3.46}$$

per ogni $X \in L^2(\Omega;\mathbb{R}^N)$ *ed ogni* $Y \in L^2(\Omega_{\mathcal{F}};\mathbb{R}^N)$*. In particolare,*

$$Cov\left(X - \mathbf{E}\left[X \mid \mathcal{F}\right], \mathbf{E}\left[X \mid \mathcal{F}\right]\right) = 0, \tag{3.47}$$

per ogni $X \in L^2(\Omega;\mathbb{R}^N)$.
Proof. *Considerando le 1 e 4 della Proposizione 171, possiamo scrivere*

$$\begin{aligned} Cov\left(X - \mathbf{E}\left[X \mid \mathcal{F}\right], Y\right) &= \mathbf{E}\left[\left(X - \mathbf{E}\left[X \mid \mathcal{F}\right]\right)Y^{\intercal}\right] - \mathbf{E}\left[X - \mathbf{E}\left[X \mid \mathcal{F}\right]\right]\mathbf{E}\left[Y^{\intercal}\right] \\ &= \mathbf{E}\left[XY^{\intercal} - \mathbf{E}\left[X \mid \mathcal{F}\right]Y^{\intercal}\right] - \left(\mathbf{E}\left[X\right] - \mathbf{E}\left[\mathbf{E}\left[X \mid \mathcal{F}\right]\right]\right)\mathbf{E}\left[Y^{\intercal}\right] \\ &= \mathbf{E}\left[XY^{\intercal}\right] - \mathbf{E}\left[\mathbf{E}\left[XY^{\intercal} \mid \mathcal{F}\right]\right] - \left(\mathbf{E}\left[X\right] - \mathbf{E}\left[X\right]\right)\mathbf{E}\left[Y^{\intercal}\right] \\ &= \mathbf{E}\left[XY^{\intercal}\right] - \mathbf{E}\left[XY^{\intercal}\right] \\ &= 0. \end{aligned}$$

□

Teorema 174 (speranza condizionata come proiezione ortogonale) *L'operatore speranza condizionata* $\mathbf{E}\left[\cdot \mid \mathcal{F}\right] : L^2(\Omega;\mathbb{R}^M) \to L^2(\Omega_{\mathcal{F}};\mathbb{R}^M)$ *è una proiezione ortogonale di* $L^2(\Omega;\mathbb{R}^M)$ *su* $L^2(\Omega_{\mathcal{F}};\mathbb{R}^M)$.

Nonostante la conoscenza delle proprietà dell'operatore $\mathbf{E}\left[\cdot \mid \mathcal{F}\right]$, il problema del calcolo concreto della speranza condizionata $\mathbf{E}\left[X \mid \mathcal{F}\right]$ di una generica variabile aleatoria $X \in L^2(\Omega;\mathbb{R}^M)$ può non essere di semplice soluzione. I seguenti risultati lo agevolano, e lo risolvono in alcuni casi particolarmente significativi.

Teorema 175 *Supponiamo che la* σ*-algebra* $\mathcal{F}$ *sia generata da una partizione numerabile di eventi* $(F_n)_{n\in N}$*, dove* $N \subseteq \mathbb{N}$*. Allora, per ogni* $X \in L^2(\Omega;\mathbb{R}^M)$*, abbiamo*

$$\mathbf{E}\left[X \mid \mathcal{F}\right] = \sum_{n\in N} \mathbf{E}\left[X \mid F_n\right] 1_{F_n}, \tag{3.48}$$

dove

$$\mathbf{E}\left[X \mid F_n\right] = \int_{F_n} X d\mathbf{P} = \int_{\Omega} X 1_{F_n} d\mathbf{P} = \mathbf{E}\left[X 1_{F_n}\right], \tag{3.49}$$

per ogni $n \in N$.

Teorema 176 *Supponiamo che la σ-algebra $\mathcal{F}$ sia generata da un insieme finito di variabili aleatorie $X_1, \ldots, X_N \in L^2\left(\Omega; \mathbb{R}^M\right)$, con $N \in \mathbb{N}$, formalmente,*

$$\mathcal{F} = \sigma\left(X_1, \ldots, X_N\right).$$

Allora per ogni $Y \in L^2\left(\Omega; \mathbb{R}^M\right)$ esiste una funzione boreliana $\varphi_Y : \mathsf{X}_{n=1}^N \mathbb{R}^M \to \mathbb{R}^M$ tale che

$$\mathbf{E}\left[Y \mid \mathcal{F}\right] \equiv \mathbf{E}\left[Y \mid \sigma\left(X_1, \ldots, X_N\right)\right] \equiv \mathbf{E}\left[Y \mid X_1, \ldots, X_N\right] = \varphi_Y\left(X_1, \ldots, X_N\right). \tag{3.50}$$

Teorema 177 *Supponiamo che $X, Y \in L^2\left(\Omega; \mathbb{R}\right)$ siano congiuntamente assolutamente continue. Sia $f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$ la densità del vettore aleatorio $(X, Y)^\intercal$ e sia $f_X : \mathbb{R} \to \mathbb{R}$ la densità di X, per la quale assumiamo*

$$f_X\left(x\right) > 0 \quad \mu_L\text{-q.o. su } \mathbb{R}.$$

Considerata una funzione boreliana $h : \mathbb{R} \to \mathbb{R}$ tale che $h\left(Y\right) \in L^2\left(\Omega; \mathbb{R}\right)$, risulta

$$\mathbf{E}\left[h\left(Y\right) \mid \sigma\left(X\right)\right] \equiv \mathbf{E}\left[h\left(Y\right) \mid X\right] = \varphi_{h(Y)}\left(X\right), \tag{3.51}$$

essendo la funzione boreliana $\varphi_{h(Y)} : \mathbb{R} \to \mathbb{R}$ data da

$$\varphi_{h(Y)}\left(x\right) = \int_{\mathbb{R}} h\left(y\right) f_{Y|X}\left(x, y\right) d\mu_L\left(y\right), \tag{3.52}$$

per ogni $x \in \mathbb{R}$, con $f_{Y|X} : \mathbb{R}^2 \to \mathbb{R}$ data da

$$f_{Y|X}\left(x, y\right) \overset{def}{=} \frac{f_{X,Y}\left(x, y\right)}{f_X\left(x\right)}, \quad \forall \left(x, y\right) \in \mathbb{R}^2. \tag{3.53}$$

Corollary 178 *Supponiamo che $X, Y \in L^2\left(\Omega; \mathbb{R}\right)$ siano congiuntamente normalmente distribuite. Risulta*

$$\mathbf{E}\left[Y \mid X\right] = \mathbf{E}\left[Y\right] + \frac{Cov(Y, X)}{\mathbf{D}^2\left[X\right]}\left(X - \mathbf{E}\left[X\right]\right) = \mathbf{E}\left[Y\right] + Corr(Y, X)\frac{\mathbf{D}\left[Y\right]}{\mathbf{D}\left[X\right]}\left(X - \mathbf{E}\left[X\right]\right), \tag{3.54}$$

e

$$\begin{aligned} \mathbf{E}\left[Y^2 \mid X\right] &= \mathbf{D}^2\left[Y\right]\left(1 - Corr(Y, X)^2\right) + \left(\mathbf{E}\left[Y\right] + Corr(Y, X)\frac{\mathbf{D}\left[Y\right]}{\mathbf{D}\left[X\right]}\left(X - \mathbf{E}\left[X\right]\right)\right)^2 \\ &= \mathbf{D}^2\left[Y\right] - \frac{Cov(Y, X)^2}{\mathbf{D}^2\left[X\right]} + \left(\mathbf{E}\left[Y\right] + \frac{Cov(Y, X)}{\mathbf{D}^2\left[X\right]}\left(X - \mathbf{E}\left[X\right]\right)\right)^2. \end{aligned} \tag{3.55}$$

Teorema 179 *Date $X, Y \in L^2\left(\Omega; \mathbb{R}\right)$, supponiamo che $\mathbf{E}\left[Y \mid X\right]$ sia lineare in X, cioè*

$$\mathbf{E}\left[Y \mid X\right] = \alpha + \beta X, \tag{3.56}$$

per qualche $\alpha, \beta \in \mathbb{R}$. Allora abbiamo

$$\mathbf{E}\left[Y \mid X\right] = \mathbf{E}\left[Y\right] + Corr(Y, X)\frac{\mathbf{D}\left[Y\right]}{\mathbf{D}\left[X\right]}\left(X - \mathbf{E}\left[X\right]\right) = \mathbf{E}\left[Y\right] + \frac{Cov(Y, X)}{\mathbf{D}^2\left[X\right]}\left(X - \mathbf{E}\left[X\right]\right). \tag{3.57}$$

Proof. *Calcolando la speranza di ambo i membri dell'Equazione (3.56), otteniamo*

$$\mathbf{E}\left[Y\right] = \mathbf{E}\left[\mathbf{E}\left[Y \mid X\right]\right] = \mathbf{E}\left[\alpha + \beta X\right] = \alpha + \beta\mathbf{E}\left[X\right]. \tag{3.58}$$

Inoltre,

$$\mathbf{E}\left[YX\right] = \mathbf{E}\left[\mathbf{E}\left[YX \mid X\right]\right] = \mathbf{E}\left[\mathbf{E}\left[Y \mid X\right]X\right] = \mathbf{E}\left[\alpha X + \beta X^2\right] = \alpha\mathbf{E}\left[X\right] + \beta\mathbf{E}\left[X^2\right]. \tag{3.59}$$

Risolvendo le Equazioni (3.58) e (3.59) in termini di α e β segue

$$\alpha = \frac{\begin{vmatrix} \mathbf{E}[Y] & \mathbf{E}[X] \\ \mathbf{E}[YX] & \mathbf{E}\left[X^2\right] \end{vmatrix}}{\begin{vmatrix} 1 & \mathbf{E}[X] \\ \mathbf{E}[X] & \mathbf{E}\left[X^2\right] \end{vmatrix}} = \frac{\mathbf{E}[Y]\,\mathbf{E}\left[X^2\right] - \mathbf{E}[X]\,\mathbf{E}[YX]}{\mathbf{D}^2[X]},$$

and

$$\beta = \frac{\begin{vmatrix} 1 & \mathbf{E}[Y] \\ \mathbf{E}[X] & \mathbf{E}[YX] \end{vmatrix}}{\begin{vmatrix} 1 & \mathbf{E}[X] \\ \mathbf{E}[X] & \mathbf{E}\left[X^2\right] \end{vmatrix}} = \frac{\mathbf{E}[YX] - \mathbf{E}[Y]\,\mathbf{E}[X]}{\mathbf{D}^2[X]} = \frac{Cov(Y,X)}{\mathbf{D}^2[X]}.$$

D'altra parte, possiamo scrivere

$$\begin{aligned}\frac{\mathbf{E}[Y]\,\mathbf{E}\left[X^2\right] - \mathbf{E}[X]\,\mathbf{E}[YX]}{\mathbf{D}^2[X]} &= \frac{\mathbf{E}[Y]\,\mathbf{E}\left[X^2\right] - \mathbf{E}[Y]\,\mathbf{E}[X]^2 + \mathbf{E}[Y]\,\mathbf{E}[X]^2 - \mathbf{E}[X]\,\mathbf{E}[YX]}{\mathbf{D}^2[X]} \\ &= \mathbf{E}[Y] - \frac{\mathbf{E}[YX] - \mathbf{E}[Y]\,\mathbf{E}[X]}{\mathbf{D}^2[X]}\mathbf{E}[X] \\ &= \mathbf{E}[Y] - \frac{Cov(Y,X)}{\mathbf{D}^2[X]}\mathbf{E}[X].\end{aligned}$$

Da queste segue il risultato desiderato. □

Tornando a riferirci allo spazio di probabilita CRR, $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$, e a un processo stocastico $(X_n)_{n=0}^N$ su Ω e a stati in $\mathbb{R}^M$ (cfr Definizione 153), introduciamo le seguenti definizioni, in forma leggermente semplificata rispetto alla loro forma più generale, in virtù dell'Osservazione 148.

Definizione 180 *Diciamo che $(X_n)_{n=0}^N$ è una $(\mathfrak{F}, \mathbf{P})$-martingala se $(X_n)_{n=0}^N$ è $\mathfrak{F}$-adattato e risulta*

$$\mathbf{E}_{n-1}[X_n] = X_{n-1}, \tag{3.60}$$

per ogni $n = 1, \ldots, N$.

Abbiamo

Proposizione 181 *Il processo stocastico $(X_n)_{n=0}^N$ è una $(\mathfrak{F}, \mathbf{P})$-martingala se e solo se $(X_n)_{n=0}^N$ è $\mathfrak{F}$-adattato e risulta*

$$\mathbf{E}_m[X_n] = X_m, \tag{3.61}$$

per ogni $n, m = 1, \ldots, N$, con $m < n$.

Proof. *Essendo la condizione sufficiente evidente, ci limitiamo a provare la condizione necessaria. Siano quindi $m, n \in \{1, \ldots, N\}$ tali che $m < n$, iterando l'applicazione la Proprietà 6 dell'operatore speranza condizionata e l'Equazione (3.60), possiamo scrivere*

$$\begin{aligned}\mathbf{E}_m[X_n] &= \mathbf{E}_m[\mathbf{E}_{n-1}[X_n]] = \mathbf{E}_m[X_{n-1}] = \mathbf{E}_m[\mathbf{E}_{n-2}[X_{n-1}]] = \mathbf{E}_m[X_{n-2}] \\ &= \cdots = \mathbf{E}_m[\mathbf{E}_{m+1}[X_{m+2}]] = \mathbf{E}_m[X_{m+1}] = X_m.\end{aligned}$$

Ciò prova l'Equazione (3.61). □

Definizione 182 *Diciamo $(X_n)_{n=0}^N$ è un $(\mathfrak{F}, \mathbf{P})$-processo di Markov se $(X_n)_{n=0}^N$ è $\mathfrak{F}$-adattato e risulta*

$$\mathbf{E}_m[f(X_n)] = \mathbf{E}[f(X_n) \mid \sigma(X_m)], \tag{3.62}$$

per ogni funzione boreliana (limitata) $f : \mathbb{R}^M \to \mathbb{R}$ e ogni $m, n \in \{0, 1, \ldots, N\}$ tali che $m < n$.

Notiamo che in termini dell'operatore probabilità condizionata $\mathbf{P}(\cdot \mid \mathcal{F}) : \mathcal{E} \to L^2(\Omega_{\mathcal{F}}; \mathbb{R}^M)$, in cui $\mathcal{F}$ è una qualsiasi sotto-σ-algebra della σ-algebra degli eventi $\mathcal{E}$ la proprietà di Markov trova un'equivalente e attraente espressione mediante l'equazione

$$\mathbf{P}(X_n \in B \mid \mathfrak{F}_n) = \mathbf{P}(X_n \in B \mid \sigma(X_m)) \tag{3.63}$$

per ogni $m, n \in \{0, 1, \ldots, N\}$ tali che $m < n$.

Proposizione 183 *Risulta*

$$\mathcal{F}_n = \sigma(S_0, S_1, \ldots, S_n), \tag{3.64}$$

per ogni $n = 0, 1, \ldots, N$, *essendo* $\sigma(S_0, S_1, \ldots, S_n)$ *la* σ*-algebra generata dalle variabili aleatorie* $S_0, S_1, \ldots, S_n$.

Proof. *Osserviamo preliminarmente che*

$$\sigma(S_0, S_1, \ldots, S_n) = \sigma(S_1, \ldots, S_n),$$

dal momento che S_0 *è interpretabile come una variabile aleatoria di Dirac, inoltre per come definite* $S_1, \ldots, S_n$ *(cfr. 3.35 e 3.36) queste sono* $\mathcal{F}_n - \mathcal{B}(\mathbb{R})$*-variabili aleatorie quindi*

$$\sigma(S_k) \subseteq \mathcal{F}_n,$$

per ogni $k = 1, \ldots, n$. *Ciò implica che*

$$\sigma(S_1, \ldots, S_n) = \bigvee_{k=1}^{n} \sigma(S_k) \subseteq \mathcal{F}_n. \tag{3.65}$$

Rinunciamo a mostrare la prova dell'inclusione inversa, ma a titolo d'esempio, ci limitiamo a mostrare che

$$\mathcal{F}_2 = \sigma(S_1, S_2). \tag{3.66}$$

Come abbiamo già visto, una partizione di Ω *in eventi di* $\mathcal{F}_2$ *è data da* $\{E_{0,0}, E_{0,1}, E_{1,0}, E_{1,1}\}$ *e abbiamo*

$$\mathcal{F}_2 = \{\varnothing, E_{0,0}, E_{0,1}, E_{1,0}, E_{1,1}, E_0, E_{0,0} \cup E_{1,0}, E_{0,0} \cup E_{1,1}, E_{0,1} \cup E_{1,0}, E_{0,1} \cup E_{1,1}, E_1, E_{1,1}^c, E_{0,1}^c, E_{1,0}^c, E_{0,0}^c, \Omega\}.$$

D'altra parte una partizione di Ω *in eventi di* $\sigma(S_1)$ *è data da* $\{E_0, E_1\}$ *e una partizione di* Ω *in eventi di* $\sigma(S_2)$ *è data da*

$$\{E_{0,0}, E_{1,1}, E_{0,1} \cup E_{1,0}\}.$$

Infatti, alla luce dei valori che può prendere S_2 *possiamo discriminare tra l'essersi realizzato uno degli eventi* $E_{0,0}$ *o* $E_{1,1}$ *o* $E_{0,1} \cup E_{1,0}$ *ma non possiamo discriminare quale tra* $E_{0,1}$ *o* $E_{1,0}$ *si sia realizzato. Tuttavia, sia gli eventi* $E_{0,0}$ *e* $E_{1,1}$ *che gli eventi*

$$E_{0,1} = E_0 \cap (E_{0,1} \cup E_{1,0}) = \quad e \quad E_{1,0} = E_1 \cap (E_{0,1} \cup E_{1,0})$$

appartengono alla σ*-algebra*

$$\sigma(S_1) \vee \sigma(S_2) = \sigma(S_1, S_2)$$

In definitiva, la stessa partizione di Ω *che genera* $\mathcal{F}_2$ *è contenuta in* $\sigma(S_1, S_2)$. *Ne segue che*

$$\mathcal{F}_2 \subseteq \sigma(S_1, S_2)$$

Per quanto preliminarmente osservato (vedi (3.66)), si ha poi

$$\sigma(S_1, S_2) \subseteq \mathcal{F}_2$$

e la (3.66) è completamente provata. $\square$

Corollary 184 *Si ha*

$$\mathcal{E} = \sigma\left(S_0, S_1, \ldots, S_n\right). \tag{3.67}$$

Corollary 185 *Per ogni* $X : \Omega \to \mathbb{R}^M$ *risulta*

$$\mathbf{E}_n[X] = \mathbf{E}\left[X \mid S_0, S_1, \ldots, S_n\right],$$

per ogni $n = 0, 1, \ldots, N$, *dove* $\mathbf{E}\left[X \mid S_0, S_1, \ldots, S_n\right] \equiv E\left[X \mid \sigma\left(S_0, S_1, \ldots, S_n\right)\right]$.

Proposizione 186 *Risulta*

$$\mathbf{E}_{n-1}\left[S_n\right] = \left(up + dq\right) S_{n-1}, \tag{3.68}$$

per ogni $n = 1, \ldots, N$. *Più in generale*

$$\mathbf{E}_m[S_n] = \left(up + dq\right)^{n-m} S_m, \tag{3.69}$$

per tutti gli $m, n \in \{0, 1, \ldots, N\}$ *tali che* $m < n$.
Proof. *Infatti,*

$$\begin{aligned}\mathbf{E}_{n-1}[S_n] &= \mathbf{E}[\beta_n S_{n-1} \mid S_0, S_1, \ldots, S_{n-1}] \\ &= S_{n-1}\mathbf{E}[\beta_n \mid S_0, S_1, \ldots, S_{n-1}] \\ &= S_{n-1}\mathbf{E}[\beta_n] = \left(up + dq\right) S_{n-1},\end{aligned}$$

per ogni $n = 1, \ldots, N$. *Inoltre*

$$\begin{aligned}\mathbf{E}_m[S_n] &= \mathbf{E}[\beta_n \cdots \beta_{m+1} S_m \mid S_0, S_1, \ldots, S_m] = S_m \mathbf{E}[\beta_n \cdots \beta_{m+1} \mid S_0, S_1, \ldots, S_m] \\ &= S_m \mathbf{E}[\beta_n \cdots \beta_{m+1}] = S_m \mathbf{E}[\beta_n] \cdots \mathbf{E}[\beta_{m+1}] = \left(up + dq\right)^{n-m} S_m,\end{aligned}$$

per tutti gli $m, n \in \{0, 1, \ldots, N\}$ *tali che* $m < n$. $\square$

Anche nel caso multiperiodale il tasso di rendimento r_n relativo all'investimento nel titolo rischioso al tempo $t = t_n$ e il tasso di rendimento $r_{m,n}$ relativo all'investimento nel titolo rischioso nell'intervallo di tempo $[t_m, t_n]$ sono variabili aleatorie. Precisamente, stante le definzioni

$$r_n \stackrel{\text{def}}{=} \frac{S_n - S_0}{S_0}, \quad n = 1, \ldots, N,$$

e

$$r_{m,n} \stackrel{\text{def}}{=} \frac{S_n - S_m}{S_m}, \quad m, n = 1, \ldots, N, \ m < n,$$

abbiamo la seguente

Proposizione 187 *Al variare di* $\omega \in \Omega$ *si ha*

$$r_n\left(\omega\right) = u^k d^{n-k} - 1, \qquad \mathbf{P}(r_n = u^k d^{n-k} - 1) = \binom{n}{k} p^k q^{n-k},$$

per ogni $n = 1, \ldots, N$, *e ogni* $k = 0, \ldots, n$. *Inoltre,*

$$r_{m,n}\left(\omega\right) = u^k d^{n-m-k} - 1, \qquad \mathbf{P}(r_{m,n} = u^k d^{n-k} - 1) = \binom{n-m}{k} p^k q^{n-m-k},$$

per ogni $m, n = 1, \ldots, N$, $m < n$, *e ogni* $k = 0, \ldots, n - m$. *In particolare,*

$$r_{n,n-1}\left(\omega\right) = \begin{cases} u - 1, & \mathbf{P}(r_{n,n-1} = u - 1) = p, \\ d - 1, & \mathbf{P}(r_{n,n-1} = u - 1) = q. \end{cases}$$

Proposizione 188 *Risulta*

$$\mathbf{E}_{n-1}[r_n] = (up + dq)\,(r_{n-1} + 1) - 1,$$

per ogni $n = 1, \ldots, N$.
Proof. *Infatti,*

$$\begin{aligned}
\mathbf{E}_{n-1}[r_n] &= \mathbf{E}\left[\frac{S_n - S_0}{S_0} \mid S_0, S_1, \ldots, S_{n-1}\right] = \mathbf{E}\left[\frac{S_n}{S_0} - 1 \mid S_0, S_1, \ldots, S_{n-1}\right] \\
&= \frac{\mathbf{E}\left[S_n \mid S_0, S_1, \ldots, S_{n-1}\right]}{S_0} - 1 = \frac{(up + qd)\, S_{n-1}}{S_0} - 1 \\
&= (up + dq)\left(\frac{S_{n-1}}{S_0} - 1\right) + up + dq - 1 \\
&= (up + dq)\, r_{n-1} + up + dq - 1 \\
&= (up + dq)\,(r_{n-1} + 1) - 1.
\end{aligned}$$

□

Definizione 189 *Analogamente alla denominazione introdotta nel caso uni-periodale, chiamiamo* portafoglio *con* posizione non rischiosa *nel bond e* posizione rischiosa *nello stock, più brevemente* BS-portafoglio*, un processo stocastico $\Pi \equiv (\Pi_n)_{n=1}^N \equiv ((X_n)_{n=1}^N, (Y_n)_{n=1}^N)$, le cui componenti $(X_n)_{n=1}^N$ e $(Y_n)_{n=1}^N$ siano successioni di variabili aleatorie reali su Ω i cui termini X_n ed Y_n rappresentino rispettivamente le quantità di bond e di stock in possesso di un operatore finanziario successivamente tempo $t = t_{n-1}$ e prima del tempo $t = t_n$, al variare di $n = 1, \ldots, N$.*

Da notare che sovente ci si riferisce a un portafoglio come a una *strategia di trading* per sottolineare la sua strutturazione ex-ante in funzione dei possibili valori futuri (aleatori) assunti dallo stock in relazione agli accadimenti di mercato.

Definizione 190 *Chiamiamo* valore *del BS-portafoglio $\Pi \equiv ((X_n)_{n=1}^N, (Y_n)_{n=1}^N)$ il processo stocastico $(W_n(\Pi))_{n=0}^N$ definito ponendo*

$$W_n(\Pi) \stackrel{def}{=} X_n B_n + Y_n S_n, \quad n = 1, \ldots, N, \tag{3.70}$$

e con $W_0 \in \mathbf{R}_+$ che rappresenta l'ammontare disponibile per l'investimento nel mercato finanziario al tempo iniziale $t = t_0 \equiv 0$.

In seguito, indicheremo più brevemente con W_n il valore di un BS-portafoglio Π al tempo $t = t_n$, per $n = 0, 1, \ldots, N$, a meno che un esplicito riferimento a Π non sia necessario.

Un operatore finanziario, volendo investire un ammontare W_0 al tempo $t = 0$ e osservati i valori B_0 del bond ed S_0 dello stock, configura prima del tempo $t = t_1$ il suo portafoglio scegliendo le posizione X_1 e Y_1 da prendere in relazione al bond e allo stock. Pertanto X_1 e Y_1 sono da considerarsi quali variabili aleatorie di Dirac concentrate in opportuni numeri reali, in definitiva numeri reali, esattamente come nel caso uniperiodale. D'altra parte, mentre è ovvio che a un tempo $t = t_n$, per un certo $n < N$, un operatore finanziario conosca con certezza la storia della composizione del suo portafoglio fino al tempo t_n stesso, non è affatto ragionevole assumere che l'operatore finanziario sia certo della composizione del suo portafoglio anche in tempi successivi a t_n. Invece è ancora ragionevole ipotizzare che l'operatore finanziario in tempi successivi a t_n configuri il suo portafoglio in risposta alle realizzazioni del prezzo dello stock, che dipendono a loro volta dagli accadimenti aleatori che influenzano il mercato. Pertanto, al tempo $t = t_n$ le sottosuccessioni $(X_k)_{k=1}^n$ e $(Y_k)_{k=1}^n$ sono interpretabili come successioni di numeri reali, ma le sottosuccessioni $(X_k)_{k=n+1}^N$ e $(Y_k)_{k=n+1}^N$ rimangono comunque successioni di variabili aleatorie. In riferimento all'istante iniziale è pertanto conveniente assumere che le successioni complete $(X_n)_{n=1}^N$ e $(Y_n)_{n=1}^N$ siano interamente costituite da variabili aleatorie.

Definizione 191 *Qualora stante l'occorrenza di un esito* $\omega \equiv (\omega_n)_{n=1}^N$ *dello spazio campionario* Ω, *il termine* $X_n(\omega)$ *sia positivo [risp. negativo], per un certo* $n = 1, \ldots, N$, *diremo che immediatamente trascorso il tempo* t_{n-1} *e in dipendenza dell'esito* ω abbiamo depositato *[risp.* preso a prestito*] l'ammontare* $X_n(\omega) B_{n-1}$. *Qualora il termine* $Y_n(\omega)$ *sia positivo [risp. negativo], per un certo* $n = 1, \ldots, N$, *diremo che immediatamente trascorso il tempo* t_{n-1} *e in dipendenza dell'esito* ω abbiamo acquistato *[risp.*venduto allo scoperto*] lo stock per un ammontare* $Y_n(\omega)S_{n-1}(\omega)$.

Ricordiamo che acquistare un bond, o equivalentemente effettuare un deposito su un conto bancario, [risp. acquistare uno stock] viene gergalmente riferito come aprire una *posizione lunga* sul bond [risp. sullo stock]. Di contro, vendere allo scoperto un bond, o equivalentemente ricevere un prestito sul conto bancario, [risp. vendere allo scoperto uno stock] viene gergalmente riferito come aprire una *posizione corta* sul bond [risp. sullo stock]. Ciò perchè mentre nulla urge alla chiusura di una posizione lunga, una posizione corta va chiusa al più presto possibile pena la corresponsione degli interessi sulla posizione, che maturano per tutto il tempo in cui la si tiene aperta.

La composizione di un portafoglio si effettua nel modo seguente: osservati i prezzi di mercato B_0 del bond ed S_0 dello stock, nell'intervallo di tempo (t_0, t_1), ossia dopo il tempo $t = t_0 \equiv 0$ e prima di $t = t_1$, un operatore finanziario costituisce un primo portafoglio che contiene X_1 unità di bond e Y_1 unità di stock attingendo all'ammontare del suo investimento iniziale. pertanto il valore di un tale portafoglio soddisferà l'equazione

$$W_0 = X_1 B_0 + Y_1 S_0.$$

Al tempo $t = t_1$, per effetto della variazione del prezzo di mercato dei titoli non rischioso e rischioso, il portafoglio assume il valore

$$W_1 \equiv X_1 B_1 + Y_1 S_1.$$

Quindi, osservati i nuovi prezzi di mercato B_1 del bond ed S_1 dello stock, nell'intervallo di tempo (t_1, t_2), ossia dopo il tempo $t = t_1$ e prima di $t = t_2$, l'operatore finanziario riconfigura il proprio portafoglio allocandovi X_2 unità di bond e Y_2 unità di stock. A seguito di tale riconfigurazione il valore del portafoglio diviene

$$X_2 B_1 + Y_2 S_1.$$

Al tempo $t = t_2$, per effetto della variazione del prezzo di mercato del bond e dello stock, il portafoglio assume il nuovo valore

$$W_2 \equiv X_2 B_2 + Y_2 S_2$$

e così via. Questo processo continua fino al tempo $t = t_{N-1}$ dopo il quale, osservati i prezzi di mercato B_{N-1} del bond ed S_{N-1} dello stock, l'operatore finanziario effettua l'ultima riconfigurazione del proprio portafoglio allocandovi X_N unità di bond e Y_N unità di stock, in modo tale che il portafoglio riconfigurato assuma il valore

$$X_N B_{N-1} + Y_N S_{N-1}$$

Al tempo $t = t_N \equiv T$, per effetto della variazione del prezzo di mercato del bond e dello stock, il portafoglio assume il valore finale

$$W_N \equiv X_N B_N + Y_N S_N$$

ed il processo s'arresta. Se grazie ad accadimenti favorevoli si realizza la circostanza $W_N \geqslant 0$ la ricchezza così prodotta viene consumata, ma se, a causa di accadimenti sfavorevoli, si realizza $W_N < 0$ l'operatore finanziario si ritrova ad avere un indebitamento che deve ripianare con fondi propri.

È naturale ipotizzare che nell'intervallo di tempo (t_{n-1}, t_n) l'operatore finanziario scelga le componenti X_n e Y_n del proprio portafoglio in funzione dei valori che si attende possano realizzarsi per il prezzo B_n del bond e il prezzo S_n dello stock al tempo $t = t_n$, osservati i valori di B_{n-1} e di S_{n-1} realizzatisi al tempo $t = t_{n-1}$. Ciò per ogni $n = 1, \ldots, N$. D'altra parte, l'evoluzione del bond è deterministica e quindi la realizzazione di B_n è sempre prevedibile con certezza. Inoltre, stante l'Equazione

(3.68), il miglior predittore del prezzo S_n è $(up + dq)\, S_{n-1}$. In definitiva, la riconfigurazione del portafoglio, ossia la scelta di X_n e Y_n, dipenderà solo dal valore di S_{n-1}. Formalmente, è naturale ipotizzare che si abbia

$$X_n = f_n\left(S_{n-1}\right) \qquad \text{e} \qquad Y_n = g_n\left(S_{n-1}\right),$$

per opportune funzioni reali f_n e g_n al variare di $n = 1, \ldots, N$. In conseguenza di queste considerazioni abbiamo

Osservazione 192 *Un BS-portafoglio* $\Pi \equiv (\Pi_n)_{n=0}^N$ *è un processo predicibile e il suo valore* $(W_n(\Pi))_{n=0}^N$ *è un processo adattato.*

Definizione 193 *Diciamo che un BS-portafoglio* $\Pi \equiv ((X_n)_{n=0}^N, (Y_n)_{n=0}^N)$ *è* autofinanziante *se*

$$W_n = X_{n+1} B_n + Y_{n+1} S_n, \tag{3.71}$$

per ogni $n = 0, 1, \ldots, N-1$.

Un portafoglio autofinanziante è un portafoglio la cui composizione cambia nel tempo, senza però che, a parte l'ammontare dell'investimento iniziale $W_0 = 0$, vi sia immissione di fondi esterni o prelievo di parte della ricchezza eventualmente prodotta in tempi antecedenti all'istante terminale $t = t_N$. Anche con una ammontare iniziale $W_0 = 0$, la costituzione di un portafoglio è possibile o prendendo a prestito un certo ammontare di bond e usando interamente questo ammontare per acquistare lo stock, oppure vendendo allo scoperto lo stock per un certo ammontare e usando interamente il ricavato per acquistare il bond. Costituito il portafoglio (X_1, Y_1) nell'intervallo di tempo (t_0, t_1), a partire da una ricchezza iniziale possibilmente nulla, ossia in modo che

$$W_0 = X_1 B_0 + Y_1 S_0,$$

al tempo $t = t_1$ l'operatore finanziaro osserva che il portafoglio prende il valore

$$W_1 = X_1 B_1 + Y_1 S_1.$$

Quindi l'operatore usa interamente ed esclusivamente W_1 per riconfigurare il proprio portafoglio. Ossia sceglie X_2 ed Y_2 in modo tale che

$$X_2 B_1 + Y_2 S_1 = W_1.$$

Ancora, al tempo $t = t_2$ l'operatore finanziaro osserva che il portafoglio riconfigurato prende il valore

$$W_2 = X_2 B_2 + Y_2 S_2$$

e usa interamente ed esclusivamente W_2 per riconfigurare il proprio portafoglio. Ossia sceglie X_3 ed Y_3 in modo tale che

$$X_3 B_2 + Y_3 S_2 = W_2.$$

Questo processo di riconfigurazione del portafoglio, senza immissione di ulteriori fondi esterni o prelievo di parte della ricchezza prodotta, continua fino al tempo $t = t_{N-1}$. Quindi nell'intervallo di tempo (t_{N-1}, t_N) l'operatore sceglie X_N e Y_N in modo da soddisfare la relazione

$$X_N B_{N-1} + Y_N S_{N-1} = W_{N-1} = X_{N-1} B_{N-1} + Y_{N-1} S_{N-1}.$$

Infine al tempo$t = t_N$ il portafoglio precedentemente riconfigurato prende il valore

$$W_N = X_N B_N + Y_N S_N$$

e come già osservato, in caso sia $W_N \geqslant 0$ [risp. $W_N < 0$] la ricchezza prodotta viene consumata [risp. l'indebitamento contratto va ripianato].

Definizione 194 *Diciamo che un BS-portafoglio autofinanziante* $\Pi \equiv ((X_n)_{n=0}^N, (Y_n)_{n=0}^N)$ *è un* arbitraggio *quando a fronte di un ammmontare nullo inizialmente investito,* $W_0 = 0$*, risulta*

$$\mathbf{P}(W_N \geqslant 0) = 1 \quad e \quad \mathbf{P}(W_N > 0) > 0. \tag{3.72}$$

Anche nel caso multiperiodale, un un portafoglio d'arbitraggio è un portafoglio che a fronte di un investimento iniziale nullo assicura con certezza, cioè indipendentementemente dagli accadimenti di mercato, che non si subiscano perdite e che sia abbia una probabilità di guadagno strettamente positiva.

Proposizione 195 *In assenza di BS-portafogli d'arbitraggo risulta*

$$r_{n-1,n}^{+} > r > r_{n-1,n}^{-} \tag{3.73}$$

per ogni $n = 1, \ldots, N$*, dove*

$$r_{n-1,n}^{+} \equiv u - 1 \quad e \quad r_{n-1,n}^{-} = d - 1.$$

Equivalentemente

$$u > 1 + r > d. \tag{3.74}$$

Proof. *Infatti, se fosse*

$$u > d \geqslant 1 + r, \tag{3.75}$$

allora, osservati sul mercato il prezzo B_0 *del bond ed* S_0 *dello stock, un operatore finanziario potrebbe prendere a prestito un'ammontare* xB_0*, con* $x < 0$ *arbitrario, e con questo ammontare acquistare* $y = -xB_0/S_0$ *unità di titolo rischioso. Al tempo* t_1 *relativamente all'acquisto del titolo rischioso, nel peggiore dei casi si realizzerebbe l'ammontare*

$$yS_1^{-} = ydS_0 = -\frac{xB_0}{S_0}dS_0 = -xdB_0.$$

Per cui, nell'ipotesi (3.75), disponendo di questo ammontare l'operatore finanziario potrebbe ripianare il debito, nel frattempo incrementato a $xB_0(r+1)$ *a causa degli interessi dovuti, senza perdere alcunchè, dal momento che in questo caso il valore del portafoglio dell'operatore finanziario sarebbe*

$$W_1^{-} = xB_1 + yS_1^{-} = xB_0(r+1) - xdB_0 = -x(d - (1+r))B_0 \geqslant 0.$$

Se però a termine del periodo d'investimento si verificasse il migliore dei casi, l'operatore finanziario realizzerebbe un ammontare

$$yS_1^{+} = yuS_0 = -\frac{xB_0}{S_0}uS_0 = -xuB_0.$$

Quindi, sempre nell'ipotesi (3.75), con questo ammontare l'operatore finanziario potrebbe ripianare il debito e realizzare un guadagno, dal momento che

$$W_1^{+} = xB_1 + yS_1^{+} = xB_0(r+1) - xuB_0 = -x(u - (1+r))B_0 > 0.$$

In definitiva avremmo

$$\mathbf{P}(W_1 \geqslant 0) = 1 \qquad e \qquad \mathbf{P}(W_1 > 0) = p > 0.$$

Se altresì fosse

$$1 + r \geqslant u > d, \tag{3.76}$$

allora un operatore finanziario potrebbe vendere allo scoperto un'ammontare yS_0*, con* $y < 0$ *arbitrario, del titolo rischioso e con questo ammontare acquistare* $x = -yS_0/B_0$ *unità di titolo non rischioso. A* *termine del periodo d'investimento, l'operatore finanziario si ritroverebbe in ogni caso con un ammontare pari a*

$$x(1+r)B_0 = -\frac{yS_0}{B_0}(1+r)B_0 = -y(1+r)S_0,$$

dovuto al maturare degli interessi prodotti dal bond. D'altra parte, rispetto alla vendita allo scoperto del titolo rischioso, il peggiore dei casi per l'operatore finanziario è che il titolo rischioso realizzi il valore di mercato S_1^+. *In questo caso per ripianare lo scoperto gli sarebbe necessario un'ammontare pari a*

$$yS_1^+ = yuS_0,$$

che, stante l'ipotesi (3.76), si sarebbe comunque reso disponibile grazie all'investimento nel titolo non rischioso. Il valore del portafoglio dell'operatore finanziario sarebbe infatti

$$W_1^- = xB_1 + yS_1^+ = x(1+r)B_0 + yS_1^+ = -y(1+r)S_0 + yuS_0 = -y\,(1+r-u) \geqslant 0.$$

Se poi il titolo rischioso realizzasse il valore di mercato S_1^-, *per ripianare lo scoperto sarebbe sufficiente un'ammontare pari a*

$$yS_1^- = ydS_0,$$

che, sempre nell'ipotesi (3.76), l'operatore finanziario avrebbe ampiamente disponibile. In questo caso il valore del portafoglio dell'operatore finanziario sarebbe

$$W_1^- = xB_1 + yS_1^- = x(1+r)B_0 + yS_1^+ = -y(1+r)S_0 + ydS_0 = -y\,(1+r-d) > 0$$

Per cui, anche nell'ipotesi (3.76) avremmo

$$\mathbf{P}\,(W_1 \geqslant 0) = 1 \geqslant 0 \qquad e \qquad \mathbf{P}(W_1 > 0) = q > 0.$$

Concludendo, sia nell'ipotesi (3.75) che nell'ipotesi (3.76), l'operatore finanziario si ritroverebbe al tempo $t = t_1$ *un portafoglio di valore non negativo e avente probabilità positiva di prendere un valore strettamente positivo. Non gli resterebbe altro da fare che investire interamente tale valore* W_1 *nel bond fino al tempo* $t = t_{N-1}$ *per ottenere un valore finale caratterizzato dalla (3.72). Ciò comporta che in assenza di portafogli autofinanzianti d'arbitraggio deve valere la (3.74).* □

Definizione 196 *Chiamiamo* probabilità neutrale al rischio *(*risk neutral probability*) una probabilità* $\tilde{\mathbf{P}} : \mathcal{E} \to \mathbb{R}_+$ *tale che*

1. *le variabili aleatorie bernoulliane* $\beta_1, \ldots, \beta_N$ *siano (totalmente) indipendenti con*

$$\tilde{\mathbf{P}}(\beta_n = u) \equiv \tilde{p} > 0, \quad \tilde{\mathbf{P}}(\beta_n = d) \equiv \tilde{q}, \tag{3.77}$$

   per ogni $n = 1, \ldots, N$, *dove* $\tilde{q} = 1 - \tilde{p}$.

2. *risulti*

$$S_{n-1} = \frac{1}{1+r}\tilde{\mathbf{E}}_{n-1}[S_n], \tag{3.78}$$

   per ogni $n = 1, \ldots, N$.

Osservazione 197 *Se* $\tilde{\mathbf{P}} : \mathcal{E} \to \mathbb{R}_+$ *è una probabilità neutrale al rischio, allora risulta*

$$S_m = \frac{1}{(1+r)^{n-m}}\tilde{\mathbf{E}}_m[S_n], \tag{3.79}$$

per ogni $m, n = 0, 1, \ldots, N$ *tali che* $m < n$. *In particolare,*

$$S_0 = \frac{1}{(1+r)^N}\tilde{\mathbf{E}}_0[S_N]. \tag{3.80}$$

Proof. Applicando la (3.78), per le proprietà dell'operatore speranza condizionata, abbiamo

$$\begin{aligned} S_m &= \frac{1}{1+r}\tilde{\mathbf{E}}_m[S_{m+1}] = \frac{1}{1+r}\tilde{\mathbf{E}}_m\left[\frac{1}{1+r}\tilde{\mathbf{E}}_{m+1}\left[S_{m+2}\right]\right] \\ &= \frac{1}{(1+r)^2}\tilde{\mathbf{E}}_m\left[\tilde{\mathbf{E}}_{m+1}\left[S_{m+2}\right]\right] = \frac{1}{(1+r)^2}\tilde{\mathbf{E}}_m\left[S_{m+2}\right]. \end{aligned}$$

Iterando questo procedimento otteniamo la (3.79). $\square$

Proposizione 198 *Se esiste una probabilità neutrale al rischio, essa unica.*
Proof. *Infatti, se esistessero due probabilità* $\tilde{\mathbf{P}}$ *e* $\mathring{\mathbf{P}}$ *su* Ω*, con* $\tilde{\mathbf{P}}$ *caratterizzata dalla (3.77) e* $\mathring{\mathbf{P}}$ *dall'analoga*

$$\mathring{\mathbf{P}}(\beta_n = u) \equiv \mathring{p} > 0, \quad \mathring{\mathbf{P}}(\beta_n = d) \equiv \mathring{q}, \tag{3.81}$$

per ogni $n = 1, \ldots, N$*, dove* $\mathring{q} = 1 - \mathring{p}$*, per entrambe delle quali valesse la (3.78), dovrebbe aversi*

$$\tilde{\mathbf{E}}_{n-1}[S_n] = \mathring{\mathbf{E}}_{n-1}[S_n].$$

Quindi, per (3.35),

$$\tilde{\mathbf{E}}_{n-1}[\beta_n S_{n-1}] = \mathring{\mathbf{E}}_{n-1}[\beta_n S_{n-1}].$$

Per le proprietà della speranza condizionata, avremmo allora

$$S_{n-1}\tilde{\mathbf{E}}_{n-1}[\beta_n] = S_{n-1}\tilde{\mathbf{E}}[\beta_n] = S_{n-1}\mathring{\mathbf{E}}[\beta_n] = S_{n-1}\mathring{\mathbf{E}}_{n-1}[\beta_n].$$

Eliminando S_{n-1} *otterremmo*

$$\tilde{\mathbf{E}}[\beta_n] = \mathring{\mathbf{E}}[\beta_n],$$

ossia

$$u\tilde{p}u + d\tilde{q} = u\mathring{p} + d\mathring{q}.$$

Da quest'ultima, considerato che $\tilde{q} = 1 - \tilde{p}$ *e* $\mathring{q} = 1 - \mathring{p}$*, chiaramente risulterebbe*

$$\tilde{p} = \mathring{p},$$

ossia $\tilde{\mathbf{P}} = \mathring{\mathbf{P}}$. $\square$

Proposizione 199 *In assenza di BS-portafogli autofinanzianti d'arbitraggio, esiste una (unica) probabilità neutrale al rischio* $\tilde{\mathbf{P}} : \mathcal{E} \to \mathbb{R}_+$ *caratterizzata come*

$$\tilde{p} \equiv \tilde{\mathbf{P}}(\beta_n = u) \overset{def}{=} \frac{1+r-d}{u-d}, \quad \tilde{q} \equiv \tilde{\mathbf{P}}(\beta_n = d) \overset{def}{=} \frac{u-(1+r)}{u-d}, \tag{3.82}$$

per ogni $n = 1, \ldots, N$*, rispetto a cui le variabili aleatorie bernoulliane* $\beta_1, \ldots, \beta_N$ *siano (totalmente) indipendenti.*
Proof. *In assenza di BS-portafogli autofinanzianti d'arbitraggio, la coppia di numeri reali* $(\tilde{p}, \tilde{q})$ *data dalla (3.82) costituisce un'effettiva distribuzione di probabilità. Possiamo allora costruire una probabilità su* Ω *ponendo*

$$\tilde{\mathbf{P}}(\omega) = \tilde{\mathbf{P}}((\omega_n)_{n=1}^N) \overset{def}{=} \tilde{p}^K \tilde{q}^{N-K}, \quad K = \left|\{n \mid \omega_n = 1\}\right|, \quad \forall \omega \in \Omega,$$

e seguendo gli stessi passi effettuati per la costruzione della probabilità $\mathbf{P} : \mathcal{E} - \mathbb{R}_+$ *(cfr. (3.18)-(3.21)). Ottenuta la probabilità* $\tilde{\mathbf{P}} : \mathcal{E} - \mathbb{R}_+$ *sulla* σ*-algebra di eventi* $\mathcal{E} \equiv \mathcal{P}(\Omega)$*, si può poi provare l'indipendenza* *delle variabili aleatorie* $\beta_1, \ldots, \beta_N$ *mediante la stessa dimostrazione della Proposizione 149. Per le proprietà dell'operatore speranza condizionata* $\tilde{\mathbf{E}}_{n-1}[\cdot]$ *rispetto a* $\tilde{\mathbf{P}}$*, si ha infine*

$$\begin{aligned}\frac{1}{1+r}\tilde{\mathbf{E}}_{n-1}[S_n] &= \frac{1}{1+r}\tilde{\mathbf{E}}_{n-1}[\beta_n S_{n-1}] = \frac{1}{1+r}\tilde{\mathbf{E}}_{n-1}[\beta_n]S_{n-1} \\ &= \frac{1}{1+r}\tilde{\mathbf{E}}[\beta_n]S_{n-1} = \frac{1}{1+r}(u\tilde{p}+d\tilde{q})S_{n-1} \\ &= \frac{1}{1+r}\left(u\frac{r+1-d}{u-d}+d\frac{u-(r+1)}{u-d}\right)S_{n-1} = \frac{1}{1+r}(r+1)S_{n-1} \\ &= S_{n-1},\end{aligned}$$

come desiderato. □

Proposizione 200 *Se esiste una (unica) probabilità neutrale al rischio* $\tilde{\mathbf{P}}$*, caratterizzata da (3.82), allora il mercato è privo di BS-portafogli autofinanzianti d'arbitraggio.*

Proof. *Supponiamo che la probabilità* $\tilde{\mathbf{P}}$ *caratterizzata dalla (3.82) sia neutrale al rischio. Deve allora valere la (3.78). D'altra parte essendo,*

$$\frac{1}{1+r}\tilde{\mathbf{E}}_{n-1}[S_n] = \frac{1}{1+r}(u\tilde{p}+d\tilde{q})S_{n-1},$$

come risulta dall'Equazione 3.68, l'Equazione (3.78) necessariamente comporta

$$u\tilde{p}+d\tilde{q} = 1+r.$$

Da quest'ultima,

$$\tilde{p} = \frac{r+1-d}{u-d}, \quad \tilde{q} = \frac{u-(r+1)}{u-d}.$$

In definitiva, se esiste una misura di probabilità neutrale al rischio questa necessariamente soddisfa l'Equazione (3.82). Supponiamo adesso che esista anche un BS-portafoglio autofinanziante d'arbitraggio $\Pi \equiv ((X_n)_{n=1}^N, (Y_n)_{n=1}^N)$ *deve aversi*

$$X_1B_0 + Y_1S_0 = 0 \quad e \quad X_NB_N + Y_NS_N \geqslant 0$$

Ha quindi senso considerare

$$\hat{n} = \min\{n \in \{1,2,\ldots,N\} \mid X_nB_n + Y_nS_n \geqslant 0\}.$$

Non può essere $\hat{n} = 1$*. Infatti, se così fosse, le condizioni*

$$X_1B_0 + Y_1S_0 = 0 \quad e \quad X_1B_1 + Y_1S_1 \geqslant 0$$

implicherebbero

$$X_1B_0 = -Y_1S_0, \quad X_1(1+r)B_0 + Y_1uS_0 \geqslant 0 \quad e \quad X_1(1+r)B_0 + Y_1dS_0 \geqslant 0.$$

Pertanto

$$(u-(1+r))Y_1S_0 \geqslant 0 \quad e \quad (d-(1+r))Y_1S_0 \geqslant 0$$

da cui

$$u \geqslant 1+r \quad e \quad d \geqslant 1+r.$$

Queste ultime impedirebbero che $\tilde{\mathbf{P}}$ *caratterizzata da* $(\tilde{p},\tilde{q})$ *possa essere un'effettiva probabilità. Sia allora* $\hat{n} > 1$*, allora la condizione di autofinanziamento*

$$X_{\hat{n}-1}B_{\hat{n}-1} + Y_{\hat{n}-1}S_{\hat{n}-1} = X_{\hat{n}}B_{\hat{n}-1} + Y_{\hat{n}}S_{\hat{n}-1} \tag{3.83}$$

esclude che si possa avere

$$X_{\hat{n}}B_{\hat{n}-1} + Y_{\hat{n}}S_{\hat{n}-1} \geqslant 0.$$

Pertanto dovremo avere

$$X_{\check{n}}B_{\check{n}-1} + Y_{\check{n}}S_{\check{n}-1} \leqslant 0. \tag{3.84}$$

oppure

$$X_{\hat{n}}B_{\hat{n}-1} + Y_{\hat{n}}S^{-}_{\hat{n}-1} < 0 \quad e \quad X_{\hat{n}}B_{\hat{n}-1} + Y_{\hat{n}}S^{+}_{\check{n}-1} > 0 \tag{3.85}$$

D'altra parte poichè

$$B_{\hat{n}} = (1+r)B_{\hat{n}-1} \quad e \quad S_{\hat{n}} = \begin{cases} uS_{\hat{n}-1} & \tilde{\mathbf{P}}(S_{\hat{n}} = uS_{\hat{n}-1}) \equiv \tilde{p} \\ dS_{\hat{n}-1} & \tilde{\mathbf{P}}(S_{\hat{n}} = dS_{\hat{n}-1}) \equiv \tilde{q} \end{cases}$$

dovrà aversi

$$X_{\hat{n}}B_{\hat{n}} + Y_{\hat{n}}S_{\hat{n}} = \begin{cases} X_{\hat{n}}(1+r)B_{\hat{n}-1} + Y_{\hat{n}}uS_{\hat{n}-1} \geqslant 0, & \tilde{\mathbf{P}}(S_{\hat{n}} = uS_{\hat{n}-1}) \equiv \tilde{p}, \\ X_{\hat{n}}(1+r)B_{\hat{n}-1} + Y_{\hat{n}}dS_{\hat{n}-1} \geqslant 0, & \tilde{\mathbf{P}}(S_{\hat{n}} = dS_{\hat{n}-1}) \equiv \tilde{q}. \end{cases} \tag{3.86}$$

Ma allora, confrontando l'Equazione (3.86) con la (3.84) oppure con la (3.85) otterremmo ancora violazioni della condizione

$$u \geqslant 1 + r \geqslant d, \tag{3.87}$$

che consente a $\tilde{\mathbf{P}}$ *di essere un'effettiva probabilità.*

Infatti, valendo la (3.84) e la (3.86) non può chiaramente essere $Y_{\hat{n}} = 0$. *D'altra parte, se fosse* $Y_{\check{n}} < 0$ *si otterrebbe*

$$X_{\check{n}}B_{\check{n}-1} \leqslant -Y_{\check{n}}S_{\check{n}-1}.$$

Quindi

$$0 \leqslant X_{\check{n}}(1+r)B_{\check{n}-1} + Y_{\check{n}}uS_{\check{n}-1} \leqslant -(1+r)Y_{\check{n}}S_{\check{n}-1} + Y_{\check{n}}uS_{\check{n}-1} = (u - (1+r))\, Y_{\check{n}}S_{\check{n}-1}$$

e

$$0 \leqslant X_{\check{n}}(1+r)B_{\check{n}-1} + Y_{\check{n}}dS_{\check{n}-1} \leqslant -(1+r)Y_{\check{n}}S_{\check{n}-1} + Y_{\check{n}}dS_{\check{n}-1} = (d - (1+r))\, Y_{\check{n}}S_{\check{n}-1}.$$

Da queste seguirebbe

$$u \leqslant 1 + r \quad e \quad d \leqslant 1 + r,$$

ossia una violazione della (3.87). Similarmente, se fosse $Y_{\check{n}} > 0$ *si otterrebbe*

$$Y_{\check{n}}S_{\check{n}-1} \leqslant -X_{\check{n}}B_{\check{n}-1}.$$

Quindi

$$0 \leqslant X_{\check{n}}(1+r)B_{\check{n}-1} + Y_{\check{n}}uS_{\check{n}-1} \leqslant X_{\check{n}}(1+r)B_{\check{n}-1} - X_{\check{n}}uB_{\check{n}-1} = (1 + r - u)\, X_{\check{n}}B_{\check{n}-1}$$

e

$$0 \leqslant X_{\check{n}}(1+r)B_{\check{n}-1} + Y_{\check{n}}dS_{\check{n}-1} \leqslant X_{\check{n}}(1+r)B_{\check{n}-1} + X_{\check{n}}dB_{\check{n}-1} = (1 + r - d)\, X_{\check{n}}B_{\check{n}-1}.$$

Seguirebbe allora

$$u \geqslant 1 + r \quad e \quad d \geqslant 1 + r,$$

ossia ancora una violazione della (3.87). Allo stesso modo, valendo la (3.85) e la (3.86) non può chiaramente essere $Y_{\hat{n}} = 0$. *Se poi fosse* $Y_{\check{n}} < 0$ *si otterrebbe*

$$X_{\hat{n}}B_{\hat{n}-1} < -Y_{\hat{n}}S^{-}_{\hat{n}-1}$$

e quindi

$$0 \leqslant X_{\hat{n}}(1+r)B_{\hat{n}-1} + Y_{\hat{n}}uS^{-}_{\hat{n}-1} < -Y_{\hat{n}}\,(1+r)\,S^{-}_{\hat{n}-1} + Y_{\hat{n}}uS^{-}_{\hat{n}-1} = (u-(1+r))\,Y_{\hat{n}}S^{-}_{\hat{n}-1}.$$

Ne seguirebbe

$$u < 1+r,$$

che viola la (3.87). Infine, se fosse $Y_{\tilde{n}} > 0$ *si otterrebbe*

$$Y_{\hat{n}}S^{-}_{\hat{n}-1} < -X_{\hat{n}}B_{\hat{n}-1}$$

e quindi

$$0 \leqslant X_{\hat{n}}(1+r)B_{\hat{n}-1} + Y_{\hat{n}}dS^{-}_{\hat{n}-1} < X_{\hat{n}}(1+r)B_{\hat{n}-1} - X_{\hat{n}}dB_{\hat{n}-1} = (1+r-d)\,X_{\hat{n}}B_{\hat{n}-1}.$$

Seguirebbe quindi

$$d > 1+r,$$

con la violazione della (3.87). $\square$

Alla luce delle Proposizioni 199 e 200 anche per un mercato multiperiodale possiamo enunciare il seguente fondamentale risultato

Teorema 201 *Nel modello CRR multiperiodale l'assenza di BS-portafogli d'arbitraggio è equivalente all'esistenza di un'unica probabilità neutrale al rischio.*

Siano $\tilde{\mathbf{P}} : \mathcal{E} \to \mathbb{R}_+$ la probabilità neutrale al rischio e sia $\tilde{\mathbf{E}}_n[\cdot] \equiv \tilde{\mathbf{E}}[\cdot \mid \mathcal{F}_n]$ l'operatore speranza condizionata rispetto a $\tilde{\mathbf{P}}$ data l'informazione $\mathfrak{F}_n$ disponibile al tempo $t = t_n$.

Proposizione 202 *La successione dei* valori di mercato scontati $(\tilde{S}_n)_{n=0}^N$ *dello stock, data da*

$$\tilde{S}_n \overset{def}{=} \frac{S_n}{B_n}, \quad n = 0, \ldots, N, \tag{3.88}$$

è una $\left(\mathfrak{F}, \tilde{\mathbf{P}}\right)$*-martingala.*
Proof. *Risulta infatti,*

$$\tilde{\mathbf{E}}_{n-1}[\tilde{S}_n] = \tilde{\mathbf{E}}_{n-1}\left[\frac{S_n}{B_n}\right] = \frac{1}{B_n}\tilde{\mathbf{E}}_{n-1}[S_n] = \frac{(1+r)S_{n-1}}{B_n} = \frac{S_{n-1}}{B_{n-1}} = \tilde{S}_{n-1}$$

per ogni $n = 1, \ldots, N$. $\square$

Proposizione 203 *Per ogni funzione boreliana (limitata)* $f : \mathbb{R} \to \mathbb{R}$ *si ha*

$$\tilde{\mathbf{E}}[f(S_n) \mid \mathcal{F}_m] = \sum_{k=0}^{n-m} f\left(u^k d^{n-m-k} S_m\right)\binom{n-m}{k}\tilde{p}^k\tilde{q}^{n-m-k}, \tag{3.89}$$

e

$$\tilde{\mathbf{E}}[f(S_n) \mid S_m] = \sum_{k=0}^{n-m} f\left(u^k d^{n-m-k} S_m\right)\binom{n-m}{k}\tilde{p}^k\tilde{q}^{n-m-k}. \tag{3.90}$$

per ogni $m, n \in \{0, 1, \ldots, N\}$ *tali che* $m < n$. *Quindi, la successione dei prezzi* $(S_n)_{n=0}^N$ *dello stock è un* $\left(\mathfrak{F}, \tilde{\mathbf{P}}\right)$*-processo di Markov.*
Proof. *Ricordato che*

$$\mathcal{F}_n = \sigma(S_1, \ldots, S_n),$$

per ogni $n = 1, \ldots, N$, *osserviamo preliminarmente che*

$$\sum_{k=0}^{n-m} f\left(u^k d^{n-m-k} S_m\right) \binom{n-m}{k} \tilde{p}^k \tilde{q}^{n-m-k}$$

è in effetti una $\sigma(S_m) - \mathcal{B}(\mathbb{R})$ *variabile aleatoria. In particolare, una* $\mathcal{F}_m - \mathcal{B}(\mathbb{R})$ *variabile aleatoria. Ora, stante la Proprietà 1 dell'operatore speranza condizionata, per ogni* $m, n \in \{0, 1, \ldots, N\}$, *tali che* $m < n$, *e ogni* $F \in \mathcal{F}_m$, *possiamo scrivere*

$$\begin{aligned}\int_F \tilde{\mathbf{E}}[f(S_n) \mid \mathcal{F}_m]\, d\tilde{\mathbf{P}}_{|\mathcal{F}_m} &= \int_F f(S_n)\, d\tilde{\mathbf{P}} = \int_F f(\beta_n \cdots \beta_{m+1} S_m)\, d\tilde{\mathbf{P}} \\ &= \sum_{(j_1,\ldots,j_{n-m}) \in \times_{j=1}^{n-m}\{u,d\}} \int_{F \cap \{\beta_{m+1}=j_1,\ldots,\beta_n=j_{n-m}\}} f(\beta_n \cdots \beta_{m+1} S_m)\, d\tilde{\mathbf{P}},\end{aligned}$$

D'altra parte, essendo $\beta_{m+1}, \ldots, \beta_{m+1}$ *indipendenti da* $\mathcal{F}_m = \sigma(\beta_1, \ldots, \beta_m)$, *abbiamo*

$$\begin{aligned}&\int_{F \cap \{\beta_m=j_1,\ldots,\beta_n=j_{n-m}\}} f(j_1 \cdots j_{n-m} S_m)\, d\tilde{\mathbf{P}} \\ &\quad = \int_\Omega f(j_1 \cdots j_{n-m} S_m)\, 1_{F \cap \{\beta_{m+1}=j_1,\ldots,\beta_n=j_{n-m}\}} d\tilde{\mathbf{P}} \\ &\quad = \int_\Omega f(j_1 \cdots j_{n-m} S_m)\, 1_F 1_{\{\beta_{m+1}=j_1,\ldots,\beta_n=j_{n-m}\}} d\tilde{\mathbf{P}} \\ &\quad = \tilde{\mathbf{E}}\left[f(j_1 \cdots j_{n-m} S_m)\, 1_F 1_{\{\beta_{m+1}=j_1,\ldots,\beta_n=j_{n-m}\}}\right] \\ &\quad = \tilde{\mathbf{E}}[f(j_1 \cdots j_{n-m} S_m)\, 1_F]\, \tilde{\mathbf{E}}\left[1_{\{\beta_{m+1}=j_1,\ldots,\beta_n=j_{n-m}\}}\right] \\ &\quad = \tilde{\mathbf{E}}[f(j_1 \cdots j_{n-m} S_m)\, 1_F]\, \tilde{\mathbf{P}}(\{\beta_{m+1}=j_1,\ldots,\beta_n=j_{n-m}\}).\end{aligned}$$

Quindi,

$$\begin{aligned}&\sum_{(j_1,\ldots,j_{n-m}) \in \times_{j=1}^{n-m}\{u,d\}} \int_{F \cap \{\beta_{m+1}=j_1,\ldots,\beta_n=j_{n-m}\}} f(\beta_n \cdots \beta_{m+1} S_m)\, d\tilde{\mathbf{P}} \\ &\quad = \sum_{(j_1,\ldots,j_{n-m}) \in \times_{j=1}^{n-m}\{u,d\}} \tilde{\mathbf{E}}[f(j_1 \cdots j_{n-m} S_m)\, 1_F]\, \tilde{\mathbf{P}}(\{\beta_{m+1}=j_1,\ldots,\beta_n=j_{n-m}\}) \\ &\quad = \sum_{k=0}^{n-m} \tilde{\mathbf{E}}\left[f\left(u^k d^{n-m-k} S_m\right) 1_F\right] \binom{n-m}{k} \tilde{p}^k \tilde{q}^{n-m-k} \\ &\quad = \tilde{\mathbf{E}}\left[\sum_{k=0}^{n-m} f\left(u^k d^{n-m-k} S_m\right) \binom{n-m}{k} \tilde{p}^k \tilde{q}^{n-m-k} 1_F\right] \\ &\quad = \int_F \sum_{k=0}^{n-m} f\left(u^k d^{n-m-k} S_m\right) \binom{n-m}{k} \tilde{p}^k \tilde{q}^{n-m-k} d\tilde{\mathbf{P}}.\end{aligned}$$

In definitiva, per ogni $m, n \in \{0, 1, \ldots, N\}$, *tali che* $m < n$, *e ogni* $F \in \mathcal{F}_m$ *abbiamo*

$$\int_F \tilde{\mathbf{E}}[f(S_n) \mid \mathcal{F}_m]\, d\tilde{\mathbf{P}}_{|\mathcal{F}_m} = \int_F \sum_{k=0}^{n-m} f\left(u^k d^{n-m-k} S_m\right) \binom{n-m}{k} \tilde{p}^k \tilde{q}^{n-m-k} d\tilde{\mathbf{P}}.$$

Sempre stante la definizione di speranza condizionata, ciò comporta la (3.89). Con lo stesso procedimento si prova poi che, per ogni $m, n \in \{0, 1, \ldots, N\}$, *tali che* $m < n$, *e ogni* $G \in \sigma(S_m)$ *vale la*

$$\int_G \tilde{\mathbf{E}}[f(S_n) \mid S_{n-1}]\, d\tilde{\mathbf{P}} = \int_G \sum_{k=0}^{n-m} f\left(u^k d^{n-m-k} S_m\right) \binom{n-m}{k} \tilde{p}^k \tilde{q}^{n-m-k} d\tilde{\mathbf{P}}.$$

Dalla Proprietà 1 segue allora la (3.90). $\square$

3.5.1 Derivatives

Sia $(\Omega, \mathcal{E}, \mathbb{P}) \equiv \Omega$ lo spazio di probabilità CRR.

Definizione 204 *Chiamiamo* derivato europeo *di sottostante S un titolo D il cui payoff alla maturità T dipende dai possibili accadimenti, positive e negativi, che occorrono al titolo S.*

La caratteristica di un derivato di tipo europeo D è che può essere esercitato solo alla scadenza. Quindi, il suo payoff è computabile solo dopo aver osservato la successione $\omega \equiv (\omega_n)_{n=1}^N$ di tutti gli accadimenti occorsi al titolo S dal tempo t_1 al tempo $t_N \equiv T$. In conseguenza, il payoff del derivato D è rappresentabile come una $\mathcal{E}-\mathcal{B}(\mathbb{R})$ variabile aleatoria $D_T : \Omega \to \mathbb{R}$. Sulla base di questa considerazione, introduciamo la seguente definizione.

Definizione 205 *Chiamiamo* payoff *di un derivato europeo D una qualunque $\mathcal{E} - \mathcal{B}(\mathbb{R})$ variabile aleatoria $D_T : \Omega \to \mathbb{R}$.*

Osservazione 206 *Stante l'Equazione 3.67 per ogni derivato di tipo europeo D esiste una funzione boreliana $F_D : \mathbb{R}^N \to \mathbb{R}$ tale che*

$$D_T(\omega) = F_D(S_1(\omega), \ldots, S_N(\omega)) \tag{3.91}$$

per ogni $\omega \in \Omega$.

Definizione 207 *Diciamo che il payoff di un derivato europeo di payoff $D_T : \Omega \to \mathbb{R}$ è* indipendente dalla storia del titolo rischioso S, *se esiste una funzione boreliana $F_D : \mathbb{R} \to \mathbb{R}$ tale che*

$$D_T(\omega) = F_D(S_N(\omega)), \tag{3.92}$$

per ogni $\omega \in \Omega$.

Le opzioni europee d'acquisto e vendita sul titolo S, con strike K alla maturità T, che anche nel modello CRR sono caratterizzate dai rispettivi payoff

$$C_T \overset{\text{def}}{=} \max\{S_T - K, 0\} \qquad \text{e} \qquad P_T \overset{\text{def}}{=} \max\{K - S_T, 0\},$$

sono derivati europei indipendenti dalla storia di S. Le opzioni asiatiche d'acquisto e vendita sul titolo S, con strike K alla maturità T, che nel modello CRR sono caratterizzate dai rispettivi payoff

$$\bar{C}_T \overset{\text{def}}{=} \max\left\{\frac{1}{N}\sum_{n=1}^{N} S_n - K, 0\right\} \qquad \text{e} \qquad \bar{P}_T \overset{\text{def}}{=} \max\left\{K - \frac{1}{N}\sum_{n=1}^{N} S_n, 0\right\},$$

sono derivati europei dipendenti dalla storia di S.

In generale, diciamo che un derivato europero è *replicabile* se è possibile costituire un portafoglio autofinanziante, con il titolo non rischioso e i titoli rischiosi sottostanti al derivato, il cui valore alla maturità del derivato replichi esattamente il payoff del derivato stesso. Un tale portafoglio è anche noto come *strategia replicante*. Diciamo che un mercato è *completo* se ogni derivato europeo è replicabile. In particolare, nell'ambito del modello CRR abbiamo.

Definizione 208 *Diciamo che un derivato europeo D del modello CRR è* replicabile *se considerato il payoff $D_T : \Omega \to \mathbb{R}$ del derivato esiste un BS-portafoglio autofinanziante $\Pi \equiv \left((X_n)_{n=1}^N, (Y_n)_{n=1}^N\right)$, tale che all'istante terminale $t_N = T$ si abbia*

$$D_T(\omega) = X_N(\omega) B_N(\omega) + Y_N(\omega) S_N(\omega), \tag{3.93}$$

per ogni $\omega \in \Omega$.

Teorema 209 *In assenza di BS-portafogli d'arbitraggio, ogni derivato europeo del modello CRR multiperiodale è replicabile mediante un BS-portafoglio autofinanziante. Pertanto, in assenza di BS-portafogli d'arbitraggio, il modello CRR multiperiodale è un modello di mercato completo.*

Proof. *Assumiamo in primo luogo che il valore al tempo $t = t_N \equiv T$ di maturità del derivato D sia indipendente dalla storia del titolo rischioso S. Quindi, esiste una funzione boreliana $F_D : \mathbb{R} \to \mathbb{R}$ tale che valga l'Equazione (3.92). Un portafoglio $\Pi \equiv \left((X_n)_{n=1}^N , (Y_n)_{n=1}^N \right)$ che alla maturità del derivato replichi il payoff del derivato stesso deve soddisfare l'equazione*

$$X_N B_N + Y_N S_N = F_D \left(S_N \right). \tag{3.94}$$

D'altra parte,

$$B_N = (1+r) B_{N-1} \qquad e \qquad S_N = \beta_N S_{N-1} = \begin{cases} uS_{N-1}, & \mathbf{P}(\beta_N = u) \equiv p, \\ dS_{N-1}, & \mathbf{P}(\beta_N = d) \equiv q. \end{cases} \tag{3.95}$$

Pertanto, le Equazioni (3.94) e (3.95) si traducono nel sistema

$$\begin{cases} (1+r) X_N B_{N-1} + uY_N S_{N-1} = F_D (uS_{N-1}), \\ (1+r) X_N B_{N-1} + dY_N S_{N-1} = F_D (dS_{N-1}). \end{cases} \tag{3.96}$$

Stanti le condizioni $B_0 > 0$, $S_0 > 0$, $u > d$ ed $r \geqslant 0$, il Sistema (3.96) ammette un'unica soluzione data da

$$X_N = \frac{\begin{vmatrix} F_D (uS_{N-1}) & uS_{N-1} \\ F_D (dS_{N-1}) & dS_{N-1} \end{vmatrix}}{\begin{vmatrix} (1+r) B_{N-1} & uS_{N-1} \\ (1+r) B_{N-1} & dS_{N-1} \end{vmatrix}} = \frac{uF_D (dS_{N-1}) - dF_D (uS_{N-1})}{(1+r)(u-d) B_{N-1}},$$

$$Y_N = \frac{\begin{vmatrix} (1+r) B_{N-1} & F_D (uS_{N-1}) \\ (1+r) B_{N-1} & F_D (dS_{N-1}) \end{vmatrix}}{\begin{vmatrix} (1+r) B_{N-1} & uS_{N-1} \\ (1+r) B_{N-1} & dS_{N-1} \end{vmatrix}} = \frac{F_D (uS_{N-1}) - F_D (dS_{N-1})}{(u-d) S_{N-1}}. \tag{3.97}$$

Da notare che l'Equazione (3.97) consente di esprimere le componenti X_N e Y_N del portafoglio al tempo $t = t_N \equiv T$ della maturità del derivato come funzioni del solo prezzo S_{N-1} dello stock S al tempo $t = t_{N-1}$, caratterizzandole quindi come $\sigma(S_{N-1}) - \mathcal{B}(\mathbb{R})$ variabili aleatorie, a fortiori $\mathcal{F}_{N-1} - \mathcal{B}(\mathbb{R})$ variabili aleatorie. In conseguenza della (3.97) e della condizione di autofinanziamento, il valore al tempo $t = t_{N-1}$ del portafoglio replicante è allora dato da

$$\begin{aligned} W_{N-1} &= X_N B_{N-1} + Y_N S_{N-1} \\ &= \frac{uF_D (dS_{N-1}) - dF_D (uS_{N-1})}{(1+r)(u-d) B_{N-1}} B_{N-1} + \frac{F_D (uS_{N-1}) - F_D (dS_{N-1})}{(u-d) S_{N-1}} S_{N-1} \\ &= \frac{uF_D (dS_{N-1}) - dF_D (uS_{N-1})}{(1+r)(u-d)} + \frac{F_D (uS_{N-1}) - F_D (dS_{N-1})}{u-d} \\ &= \frac{1}{1+r} \left(\frac{1+r-d}{u-d} F_D (uS_{N-1}) + \frac{u-(1+r)}{u-d} F_D (dS_{N-1}) \right) \end{aligned} \tag{3.98}$$

In assenza di portafogli d'arbitraggio, vale la condizione

$$u > 1 + r > d.$$

e sappiamo che

$$\tilde{p} = \frac{1+r-d}{u-d} \quad e \quad \tilde{q} = \frac{u-(1+r)}{u-d}$$

sono le componenti della misura neutrale al rischio $\tilde{\mathbf{P}}$ *per cui otteniamo*

$$W_{N-1} = \frac{1}{1+r}\left(\tilde{p}F_D\left(uS_{N-1}\right) + \tilde{q}F_D\left(dS_{N-1}\right)\right) = \frac{1}{1+r}\tilde{\mathbf{E}}\left[F_D\left(S_N\right) \mid S_{N-1}\right], \tag{3.99}$$

ed essendo $(S_n)_{n=0}^N$ *un* $\left(\mathfrak{F},\tilde{\mathbf{P}}\right)$*-processo di Markov (cfr Proposizione (203)), ne segue*

$$W_{N-1} = \frac{1}{1+r}\tilde{\mathbf{E}}_{N-1}\left[F_D\left(S_N\right)\right]. \tag{3.100}$$

Da notare che l'Equazione (3.99) stabilisce come il valore del portafoglio replicante in costruzione al tempo $t = t_{N-1}$ *coincida con il prezzo neutrale al rischio del derivato allo stesso tempo. Determinato il valore* W_{N-1} *del portafoglio replicante al tempo* $t = t_{N-1}$*, che risulta essere funzione di* S_{N-1}*, per cui possiamo scrivere*

$$W_{N-1} \equiv W_{N-1}\left(S_{N-1}\right),$$

e imponendo la condizione di replicazione, risulta che le componenti X_{N-1} *e* Y_{N-1} *del portafoglio replicante scelte al tempo immediatamente precedente il tempo* $t = t_{N-1}$ *devono soddisfare l'equazione*

$$X_{N-1}B_{N-1} + Y_{N-1}S_{N-1} = W_{N-1}\left(S_{N-1}\right).$$

Questa si traduce nel sistema

$$\begin{cases} (1+r)\,X_{N-1}B_{N-2} + uY_{N-1}S_{N-2} = W_{N-1}\left(uS_{N-2}\right), \\ (1+r)\,X_{N-1}B_{N-2} + dY_{N-1}S_{N-2} = W_{N-1}\left(dS_{N-2}\right), \end{cases} \tag{3.101}$$

da cui

$$\begin{aligned} X_{N-1} &= \frac{\begin{vmatrix} W_{N-1}\left(uS_{N-2}\right) & uS_{N-2} \\ W_{N-1}\left(dS_{N-2}\right) & dS_{N-2} \end{vmatrix}}{\begin{vmatrix} (1+r)\,B_{N-2} & uS_{N-2} \\ (1+r)\,B_{N-2} & dS_{N-2} \end{vmatrix}} = \frac{uW_{N-1}\left(dS_{N-2}\right) - dW_{N-1}\left(uS_{N-2}\right)}{(1+r)\,(u-d)\,B_{N-2}}, \\ Y_{N-1} &= \frac{\begin{vmatrix} (1+r)\,B_{N-2} & W_{N-1}\left(uS_{N-2}\right) \\ (1+r)\,B_{N-2} & W_{N-1}\left(dS_{N-2}\right) \end{vmatrix}}{\begin{vmatrix} (1+r)\,B_{N-2} & uS_{N-2} \\ (1+r)\,B_{N-2} & dS_{N-2} \end{vmatrix}} = \frac{W_{N-1}\left(uS_{N-2}\right) - W_{N-1}\left(dS_{N-2}\right)}{(u-d)\,S_{N-2}}. \end{aligned} \tag{3.102}$$

In conseguenza della (3.102), della condizione di autofinanziamento, della (3.99) e della proprietà di Markov, il valore

$$W_{N-2} \equiv W_{N-2}\left(S_{N-2}\right)$$

del portafoglio replicante al tempo $t = t_{N-2}$ è allora dato da

$$\begin{aligned}
W_{N-2} &= X_{N-1}B_{N-2} + Y_{N-1}S_{N-2} \\
&= \frac{uW_{N-1}\left(dS_{N-2}\right) - dW_{N-1}\left(uS_{N-2}\right)}{(1+r)\,(u-d)\,B_{N-2}}B_{N-2} + \frac{W_{N-1}\left(uS_{N-2}\right) - W_{N-1}\left(dS_{N-2}\right)}{(u-d)\,S_{N-2}}S_{N-2} \\
&= \frac{1}{1+r}\left(\frac{1+r-d}{u-d}W_{N-1}\left(uS_{N-2}\right) + \frac{u-(1+r)}{u-d}W_{N-1}\left(dS_{N-2}\right)\right) \\
&= \frac{1}{1+r}\left(\tilde{p}W_{N-1}\left(uS_{N-2}\right) + \tilde{q}W_{N-1}\left(dS_{N-2}\right)\right) \\
&= \frac{1}{(1+r)^2}\left(\tilde{p}^2F\left(u^2S_{N-2}\right) + 2\tilde{p}\tilde{q}F\left(udS_{N-2}\right) + \tilde{q}^2F\left(d^2S_{N-2}\right)\right) \\
&= \frac{1}{(1+r)^2}\sum_{k=0}^{2}\binom{2}{k}\tilde{p}^{2-k}\tilde{q}^kF\left(u^{2-k}d^kS_{N-2}\right) \\
&= \frac{1}{(1+r)^2}\tilde{\mathbf{E}}\left[F_D\left(S_N\right) \mid S_{N-2}\right] \\
&= \frac{1}{(1+r)^2}\tilde{\mathbf{E}}_{N-2}\left[F_D\left(S_N\right)\right].
\end{aligned} \tag{3.103}$$

Iterando questo procedimento, toviamo che le componenti X_{N-n} e Y_{N-n} del portafoglio replicante al tempo $t = t_{N-n}$, con $0 \leqslant n \leqslant N-1$, devono soddisfare l'equazione

$$\begin{aligned}
X_{N-n} &= \frac{\begin{vmatrix} W_{N-n}\left(uS_{N-(n+1)}\right) & uS_{N-(n+1)} \\ W_{N-n}\left(dS_{N-(n+1)}\right) & dS_{N-(n+1)} \end{vmatrix}}{\begin{vmatrix} (1+r)\,B_{N-(n+1)} & uS_{N-(n+1)} \\ (1+r)\,B_{N-(n+1)} & dS_{N-(n+1)} \end{vmatrix}} = \frac{uW_{N-n}\left(dS_{N-(n+1)}\right) - dW_{N-n}\left(uS_{N-(n+1)}\right)}{(1+r)\,(u-d)\,B_{N-(n+1)}}, \\
Y_{N-n} &= \frac{\begin{vmatrix} (1+r)\,B_{N-(n+1)} & W_{N-n}\left(uS_{N-(n+1)}\right) \\ (1+r)\,B_{N-(n+1)} & W_{N-n}\left(dS_{N-(n+1)}\right) \end{vmatrix}}{\begin{vmatrix} (1+r)\,B_{N-(n+1)} & uS_{N-(n+1)} \\ (1+r)\,B_{N-(n+1)} & dS_{N-(n+1)} \end{vmatrix}} = \frac{W_{N-n}\left(uS_{N-(n+1)}\right) - W_{N-n}\left(dS_{N-(n+1)}\right)}{(u-d)\,S_{N-(n+1)}}.
\end{aligned} \tag{3.104}$$

Quindi, il valore

$$W_{N-(n+1)} \equiv W_{N-(n+1)}\left(S_{N-(n+1)}\right)$$

del portafoglio replicante al tempo $t = t_{N-(n+1)}$ è dato da

$$\begin{aligned}
W_{N-(n+1)} &= X_{N-n}B_{N-(n+1)} + Y_{N-n}S_{N-(n+1)} \\
&= \frac{uW_{N-n}\left(dS_{N-(n+1)}\right) - dW_{N-n}\left(uS_{N-(n+1)}\right)}{(1+r)(u-d)B_{N-(n+1)}}B_{N-(n+1)} \\
&+ \frac{W_{N-n}\left(uS_{N-(n+1)}\right) - W_{N-n}\left(dS_{N-(n+1)}\right)}{(u-d)S_{N-(n+1)}}S_{N-(n+1)} \\
&= \frac{1}{1+r}\left(\frac{1+r-d}{u-d}W_{N-n}\left(uS_{N-(n+1)}\right) + \frac{u-(1+r)}{u-d}W_{N-n}\left(dS_{N-(n+1)}\right)\right) \\
&= \frac{1}{1+r}\left(\tilde{p}W_{N-n}\left(uS_{N-(n+1)}\right) + \tilde{q}W_{N-n}\left(dS_{N-(n+1)}\right)\right) \\
&= \frac{1}{(1+r)^{n+1}}\sum_{k=0}^{n+1}\binom{n+1}{k}\tilde{p}^{n+1-k}\tilde{q}^k F\left(u^{n+1-k}d^k S_{N-(n+1)}\right) \\
&= \frac{1}{(1+r)^{n+1}}\tilde{\mathbf{E}}\left[F_D\left(S_N\right) \mid S_{N-(n+1)}\right] \\
&= \frac{1}{(1+r)^{n+1}}\tilde{\mathbf{E}}_{N-(n+1)}\left[F_D\left(S_N\right)\right] \qquad (3.105)
\end{aligned}$$

In particolare, posto $n = N-1$, otteniamo che il valore

$$W_0 = X_1B_0 + Y_1S_0 \equiv W_0\left(S_0\right)$$

del portafoglio replicante al tempo $t = t_0 \equiv 0$ è anche dato da

$$W_0 = \frac{1}{(1+r)^N}\sum_{k=0}^{N}\binom{N}{k}\tilde{p}^{N-k}\tilde{q}^k F\left(u^{N-k}d^k S_0\right) = \frac{1}{(1+r)^N}\tilde{\mathbf{E}}\left[F_D\left(S_N\right)\right]. \qquad (3.106)$$

Quindi, uguaglia il valore neutrale al rischio del derivato stesso allo stesso tempo. Inoltre, le componenti X_1 e Y_1 del portafoglio replicante costituite tra il tempo $t = t_0$ e $t = t_1$ sono date da

$$X_1 = \frac{uW_1\left(dS_0\right) - dW_1\left(uS_0\right)}{(1+r)(u-d)B_0}, \qquad Y_1 = \frac{W_1\left(uS_0\right) - W_1\left(dS_0\right)}{(u-d)S_0}, \qquad (3.107)$$

dove il valore

$$W_1 \equiv W_1\left(S_1\right)$$

del portafoglio replicante al tempo $t = t_1$ è anche dato da

$$W_1 = \frac{1}{(1+r)^{N-1}}\sum_{k=0}^{N-1}\binom{N-1}{k}\tilde{p}^{N-1-k}\tilde{q}^k F_D\left(u^{N-1-k}d^k S_1\right). \qquad (3.108)$$

Sostituendo la (3.108) nella (3.107) otteniamo allora,

$$\begin{aligned}
X_1 &= \frac{1}{(1+r)^N(u-d)B_0}\sum_{k=0}^{N-1}\binom{N-1}{k}\tilde{p}^{N-1-k}\tilde{q}^k\left(uF_D\left(u^{N-1-k}d^{k+1}S_0\right) - dF_D\left(u^{N-k}d^kS_0\right)\right), \\
Y_1 &= \frac{1}{(1+r)^{N-1}(u-d)S_0}\sum_{k=0}^{N-1}\binom{N-1}{k}\tilde{p}^{N-1-k}\tilde{q}^k\left(F_D\left(u^{N-k}d^kS_0\right) - F_D\left(u^{N-1-k}d^{k+1}S_0\right)\right).
\end{aligned} \qquad (3.109)$$

ciò conferma la struttura deterministica, ovvero di $\mathcal{F}_0 - \mathcal{B}(\mathbb{R})$ variabili aleatorie, delle componenti X_1 e Y_1. Notare che dall'Equazione (3.109) segue

$$
\begin{aligned}
&X_1 B_1 + Y_1 S_1 \\
&= \frac{1}{(1+r)^{N-1}(u-d)} \sum_{k=0}^{N-1} \binom{N-1}{k} \tilde{p}^{N-1-k} \tilde{q}^k \left(uF_D\left(u^{N-1-k} d^{k+1} S_0\right) - dF_D\left(u^{N-k} d^k S_0\right)\right) \\
&+ \frac{\beta_1}{(1+r)^{N-1}(u-d)} \sum_{k=0}^{N-1} \binom{N-1}{k} \tilde{p}^{N-1-k} \tilde{q}^k \left(F_D\left(u^{N-k} d^k S_0\right) - dF_D\left(u^{N-1-k} d^{k+1} S_0\right)\right) \\
&= \frac{1}{(1+r)^{N-1}} \sum_{k=0}^{N-1} \binom{N-1}{k} \tilde{p}^{N-1-k} \tilde{q}^k \\
&\cdot \left(\frac{uF_D\left(u^{N-1-k} d^{k+1} S_0\right) - dF_D\left(u^{N-k} d^k S_0\right) + \beta_1\left(F_D\left(u^{N-k} d^k S_0\right) - F_D\left(u^{N-1-k} d^{k+1} S_0\right)\right)}{u-d} \right).
\end{aligned}
\tag{3.110}
$$

D'altra parte, quando $\beta_1 = d$ abbiamo

$$
\begin{aligned}
&\frac{uF_D\left(u^{N-1-k} d^{k+1} S_0\right) - dF_D\left(u^{N-k} d^k S_0\right) + \beta_1\left(F_D\left(u^{N-k} d^k S_0\right) - F_D\left(u^{N-1-k} d^{k+1} S_0\right)\right)}{u-d} \\
&= F_D\left(u^{N-1-k} d^{k+1} S_0\right) = F_D\left(u^{N-1-k} d^k S_1\right)
\end{aligned}
$$

mentre, quando $\beta_1 = u$

$$
\begin{aligned}
&\frac{uF_D\left(u^{N-1-k} d^{k+1} S_0\right) - dF_D\left(u^{N-k} d^k S_0\right) + u\left(F_D\left(u^{N-k} d^k S_0\right) - F_D\left(u^{N-1-k} d^{k+1} S_0\right)\right)}{u-d} \\
&= F_D\left(u^{N-k} d^k S_0\right) = F_D\left(u^{N-1-k} d^k S_1\right).
\end{aligned}
$$

Pertanto, possiamo scrivere

$$
\begin{aligned}
&\frac{uF_D\left(u^{N-1-k} d^{k+1} S_0\right) - dF_D\left(u^{N-k} d^k S_0\right) + \beta_1\left(F_D\left(u^{N-k} d^k S_0\right) - F_D\left(u^{N-1-k} d^{k+1} S_0\right)\right)}{u-d} \\
&= F_D\left(u^{N-1-k} d^k S_1\right).
\end{aligned}
$$

Sostituendo quest'ultima nella (3.110) e ricordando la (3.108), otteniamo allora

$$
X_1 B_1 + Y_1 S_1 = \frac{1}{(1+r)^{N-1}} \sum_{k=0}^{N-1} \binom{N-1}{k} \tilde{p}^{N-1-k} \tilde{q}^k F_D\left(u^{N-1-k} d^k S_1\right) = W_1,
$$

a ulteriore conferma che le componenti iniziali X_1 e Y_1 del portafoglio replicante sono state correttamente determinate.

Assumiamo adesso che il valore al tempo $t = t_N \equiv T$ di maturità del derivato possa dipendere dalla storia del titolo rischioso S. Quindi, esiste una funzione boreliana $F_D : \mathbb{R}^N \to \mathbb{R}$ tale che valga l'Equazione (3.91). Un portafoglio $\Pi \equiv \left((X_n)_{n=1}^N, (Y_n)_{n=1}^N\right)$ che alla maturità del derivato replichi il payoff del derivato stesso deve soddisfare l'equazione

$$
X_N B_N + Y_N S_N = F_D\left(S_1(\omega), \ldots, S_N(\omega)\right).
$$

Quindi le Equazioni (3.109) e (3.95) si rducono nel sistema di equazioni

$$
\begin{cases}
(1+r) X_N B_{N-1} + uY_N S_{N-1} = F_D\left(S_1, \ldots, S_{N-1}, uS_{N-1}\right), \\
(1+r) X_N B_{N-1} + dY_N S_{N-1} = F_D\left(S_1, \ldots, S_{N-1}, dS_{N-1}\right).
\end{cases}
\tag{3.111}
$$

Sempre stanti le condizioni $B_0 > 0$, $S_0 > 0$, $u > d$ ed $r \geqslant 0$, il Sistema (3.96) ammette un'unica soluzione data da

$$\begin{aligned}
X_N &= \frac{\begin{vmatrix} F_D(S_1,\dots,S_{N-1},uS_{N-1}) & uS_{N-1} \\ F_D(S_1,\dots,S_{N-1},dS_{N-1}) & dS_{N-1} \end{vmatrix}}{\begin{vmatrix} (1+r)\,B_{N-1} & uS_{N-1} \\ (1+r)\,B_{N-1} & dS_{N-1} \end{vmatrix}} \\
&= \frac{uF_D(S_1,\dots,S_{N-1},dS_{N-1}) - dF_D(S_1,\dots,S_{N-1},dS_{N-1})}{(1+r)(u-d)\,B_{N-1}}, \\
Y_N &= \frac{\begin{vmatrix} (1+r)\,B_{N-1} & F_D(S_1,\dots,S_{N-1},uS_{N-1}) \\ (1+r)\,B_{N-1} & F_D(S_1,\dots,S_{N-1},dS_{N-1}) \end{vmatrix}}{\begin{vmatrix} (1+r)\,B_{N-1} & uS_{N-1} \\ (1+r)\,B_{N-1} & dS_{N-1} \end{vmatrix}} \\
&= \frac{F_D(S_1,\dots,S_{N-1},uS_{N-1}) - F_D(S_1,\dots,S_{N-1},dS_{N-1})}{(u-d)\,S_{N-1}}. \qquad (3.112)
\end{aligned}$$

Da notare che l'Equazione (3.112) esprimere le componenti X_N e Y_N del portafoglio al tempo $t = t_N \equiv T$ della maturità del derivato come funzioni della storia del prezzo dello stock S fino al tempo $t = t_{N-1}$, caratterizzandole quindi come $\mathcal{F}_{N-1} - \mathcal{B}(\mathbb{R})$ variabili aleatorie. In consequenza della (3.112) e della condizione di autofinanziamento, il valore al tempo $t = t_{N-1}$ del portafoglio replicante è allora dato da

$$\begin{aligned}
W_{N-1} &= X_N B_{N-1} + Y_N S_{N-1} \\
&= \frac{uF_D(S_1,\dots,S_{N-1},dS_{N-1}) - dF_D(S_1,\dots,S_{N-1},dS_{N-1})}{(1+r)(u-d)\,B_{N-1}} B_{N-1} \\
&+ \frac{F_D(S_1,\dots,S_{N-1},uS_{N-1}) - F_D(S_1,\dots,S_{N-1},dS_{N-1})}{(u-d)\,S_{N-1}} S_{N-1} \\
&= \frac{1}{1+r}\left(\frac{1+r-d}{u-d} F_D(S_1,\dots,S_{N-1},uS_{N-1}) + \frac{u-(1+r)}{u-d} F_D(S_1,\dots,S_{N-1},dS_{N-1})\right) \\
&= \frac{1}{1+r}\left(\tilde{p}F_D(S_1,\dots,S_{N-1},uS_{N-1}) + \tilde{q}F_D(S_1,\dots,S_{N-1},dS_{N-1})\right) \\
&= \frac{1}{1+r}\tilde{\mathbf{E}}\left[F_D(S_1,\dots,S_{N-1},S_N) \mid S_1,\dots,S_{N-1}\right] \\
&= \frac{1}{1+r}\tilde{\mathbf{E}}_{N-1}\left[F_D(S_1,\dots,S_{N-1},S_N)\right]. \qquad (3.113)
\end{aligned}$$

Da cui si vede che il valore del portafoglio replicante al tempo $t = t_{N-1}$ uguaglia il prezzo neutrale al rischio del derivato allo stesso tempo. Determinato il valore W_{N-1} del portafoglio replicante al tempo $t = t_{N-1}$, che risulta essere funzione di $S_1,\dots,S_{N-1}$, per cui possiamo scrivere

$$W_{N-1} \equiv W_{N-1}(S_1,\dots,S_{N-1}),$$

e imponendo la condizione di replicazione, risulta che le componenti X_{N-1} e Y_{N-1} del portafoglio replicante, scelte dopo il tempo $t = t_{N-2}$ ma prima del tempo $t = t_{N-1}$, devono soddisfare l'equazione

$$X_{N-1}B_{N-1} + Y_{N-1}S_{N-1} = W_{N-1}(S_1,\dots,S_{N-1}).$$

Questa si traduce nel sistema

$$\begin{cases} (1+r)\,X_{N-1}B_{N-2} + uY_{N-1}S_{N-2} = W_{N-1}(S_1,\dots,S_{N-2},uS_{N-2}), \\ (1+r)\,X_{N-1}B_{N-2} + dY_{N-1}S_{N-2} = W_{N-1}(S_1,\dots,S_{N-2},dS_{N-2}), \end{cases}$$

da cui

$$
\begin{aligned}
X_{N-1} &= \frac{\begin{vmatrix} W_{N-1}\left(S_1,\ldots,S_{N-2},uS_{N-2}\right) & uS_{N-2} \\ W_{N-1}\left(S_1,\ldots,S_{N-2},dS_{N-2}\right) & dS_{N-2} \end{vmatrix}}{\begin{vmatrix} (1+r)\,B_{N-2} & uS_{N-2} \\ (1+r)\,B_{N-2} & dS_{N-2} \end{vmatrix}} \\
&= \frac{uW_{N-1}\left(S_1,\ldots,S_{N-2},dS_{N-2}\right) - dW_{N-1}\left(S_1,\ldots,S_{N-2},uS_{N-2}\right)}{(1+r)\,(u-d)\,B_{N-2}}, \\
Y_{N-1} &= \frac{\begin{vmatrix} (1+r)\,B_{N-2} & W_{N-1}\left(S_1,\ldots,S_{N-2},uS_{N-2}\right) \\ (1+r)\,B_{N-2} & W_{N-1}\left(S_1,\ldots,S_{N-2},dS_{N-2}\right) \end{vmatrix}}{\begin{vmatrix} (1+r)\,B_{N-2} & uS_{N-2} \\ (1+r)\,B_{N-2} & dS_{N-2} \end{vmatrix}} \\
&= \frac{W_{N-1}\left(S_1,\ldots,S_{N-2},uS_{N-2}\right) - W_{N-1}\left(S_1,\ldots,S_{N-2},dS_{N-2}\right)}{(u-d)\,S_{N-2}}.
\end{aligned}
\tag{3.114}
$$

In consequenza delle Equazioni (3.113), (3.114) e della condizione di autofinanziamento, il valore

$$W_{N-2} \equiv W_{N-2}\left(S_1,\ldots,S_{N-2}\right)$$

del portafoglio replicante al tempo $t = t_{N-2}$ è allora dato da

$$
\begin{aligned}
& W_{N-2} \\
&= X_{N-1}B_{N-2} + Y_{N-1}S_{N-2} \\
&= \frac{uW_{N-1}\left(S_1,\ldots,S_{N-2},dS_{N-2}\right) - dW_{N-1}\left(S_1,\ldots,S_{N-2},uS_{N-2}\right)}{(1+r)\,(u-d)\,B_{N-2}} B_{N-2} \\
&\quad + \frac{W_{N-1}\left(S_1,\ldots,S_{N-2},uS_{N-2}\right) - W_{N-1}\left(S_1,\ldots,S_{N-2},dS_{N-2}\right)}{(u-d)\,S_{N-2}} S_{N-2} \\
&= \frac{1}{1+r}\left(\frac{1+r-d}{u-d} W_{N-1}\left(S_1,\ldots,S_{N-2},uS_{N-2}\right) + \frac{u-(1+r)}{u-d} W_{N-1}\left(S_1,\ldots,S_{N-2},dS_{N-2}\right)\right) \\
&= \frac{1}{1+r}\left(\tilde{p}W_{N-1}\left(S_1,\ldots,S_{N-2},uS_{N-2}\right) + \tilde{q}W_{N-1}\left(S_1,\ldots,S_{N-2},dS_{N-2}\right)\right) \\
&= \frac{1}{1+r}\tilde{p}\left(\frac{1}{1+r}\left(\tilde{p}F_D\left(S_1,\ldots,S_{N-2},uS_{N-2},uS_{N-1}\right) + \tilde{q}F_D\left(S_1,\ldots,S_{N-2},uS_{N-2},dS_{N-1}\right)\right)\right) \\
&\quad + \frac{1}{1+r}\tilde{q}\left(\frac{1}{1+r}\left(\tilde{p}F_D\left(S_1,\ldots,S_{N-2},dS_{N-2},uS_{N-1}\right) + \tilde{q}F_D\left(S_1,\ldots,S_{N-2},dS_{N-2},dS_{N-1}\right)\right)\right) \\
&= \frac{1}{(1+r)^2}\Big(\tilde{p}^2F_D\left(S_1,\ldots,S_{N-2},uS_{N-2},uS_{N-1}\right) + \tilde{p}\tilde{q}F_D\left(S_1,\ldots,S_{N-2},uS_{N-2},dS_{N-1}\right) \\
&\quad + \tilde{q}\tilde{p}F_D\left(S_1,\ldots,S_{N-2},dS_{N-2},uS_{N-1}\right) + \tilde{q}^2F_D\left(S_1,\ldots,S_{N-2},dS_{N-2},dS_{N-1}\right)\Big) \\
&= \frac{1}{(1+r)^2}\Big(\tilde{p}^2F_D\left(S_1,\ldots,S_{N-2},uS_{N-2},uS_{N-1}\right) + 2\tilde{p}\tilde{q}F_D\left(S_1,\ldots,S_{N-2},dS_{N-2},uS_{N-1}\right) \\
&\quad + \tilde{q}^2F_D\left(S_1,\ldots,S_{N-2},dS_{N-2},dS_{N-1}\right)\Big) \\
&= \frac{1}{(1+r)^2}\tilde{\mathbf{E}}_{N-2}\left[F_D\left(S_1,\ldots,S_N\right)\right].
\end{aligned}
\tag{3.115}
$$

Da quest'ultima si vede che anche il valore del portafoglio replicante al tempo $t = t_{N-2}$ uguaglia il prezzo neutrale al rischio del derivato allo stesso tempo. Iterando questo procedimento, toviamo che le componenti X_{N-n} e Y_{N-n} del portafoglio replicante al tempo $t = t_{N-n}$, con $0 \leqslant n \leqslant N-1$, devono *soddisfare l'equazione*

$$\begin{aligned}
X_{N-n} &= \frac{\begin{vmatrix} W_{N-n}\left(S_1,\ldots,S_{N-(n+1)},uS_{N-(n+1)}\right) & uS_{N-(n+1)} \\ W_{N-n}\left(S_1,\ldots,S_{N-(n+1)},dS_{N-(n+1)}\right) & dS_{N-(n+1)} \end{vmatrix}}{\begin{vmatrix} (1+r)\,B_{N-(n+1)} & uS_{N-(n+1)} \\ (1+r)\,B_{N-(n+1)} & dS_{N-(n+1)} \end{vmatrix}} \\
&= \frac{uW_{N-n}\left(S_1,\ldots,S_{N-(n+1)},dS_{N-(n+1)}\right) - dW_{N-n}\left(S_1,\ldots,S_{N-(n+1)},uS_{N-(n+1)}\right)}{(1+r)\,(u-d)\,B_{N-(n+1)}}, \\
Y_{N-n} &= \frac{\begin{vmatrix} (1+r)\,B_{N-(n+1)} & W_{N-n}\left(S_1,\ldots,S_{N-(n+1)},uS_{N-(n+1)}\right) \\ (1+r)\,B_{N-(n+1)} & W_{N-n}\left(S_1,\ldots,S_{N-(n+1)},dS_{N-(n+1)}\right) \end{vmatrix}}{\begin{vmatrix} (1+r)\,B_{N-(n+1)} & uS_{N-(n+1)} \\ (1+r)\,B_{N-(n+1)} & dS_{N-(n+1)} \end{vmatrix}} \\
&= \frac{W_{N-n}\left(S_1,\ldots,S_{N-(n+1)},uS_{N-(n+1)}\right) - W_{N-n}\left(S_1,\ldots,S_{N-(n+1)},dS_{N-(n+1)}\right)}{(u-d)\,S_{N-(n+1)}}. \qquad (3.116)
\end{aligned}$$

Quindi, il valore

$$W_{N-(n+1)} \equiv W_{N-(n+1)}\left(S_{N-(n+1)}\right)$$

del portafoglio replicante al tempo $t = t_{N-(n+1)}$ *è anche dato da*

$$\begin{aligned}
&W_{N-(n+1)} \\
&= X_{N-n}B_{N-(n+1)} + Y_{N-n}S_{N-(n+1)} \\
&= \frac{uW_{N-n}\left(S_1,\ldots,S_{N-(n+1)},dS_{N-(n+1)}\right) - dW_{N-n}\left(S_1,\ldots,S_{N-(n+1)},uS_{N-(n+1)}\right)}{(1+r)\,(u-d)\,B_{N-(n+1)}}B_{N-(n+1)} \\
&\quad + \frac{W_{N-n}\left(S_1,\ldots,S_{N-(n+1)},uS_{N-(n+1)}\right) - W_{N-n}\left(S_1,\ldots,S_{N-(n+1)},dS_{N-(n+1)}\right)}{(u-d)\,S_{N-(n+1)}}S_{N-(n+1)} \\
&= \frac{1}{1+r}\left(\frac{1+r-d}{u-d}W_{N-n}\left(S_1,\ldots,S_{N-(n+1)},uS_{N-(n+1)}\right)\right. \\
&\quad \left. +\frac{u-(1+r)}{u-d}W_{N-n}\left(S_1,\ldots,S_{N-(n+1)},dS_{N-(n+1)}\right)\right) \\
&= \frac{1}{1+r}\left(\tilde{p}W_{N-n}\left(S_1,\ldots,S_{N-(n+1)},uS_{N-(n+1)}\right) + \tilde{q}W_{N-n}\left(S_1,\ldots,S_{N-(n+1)},dS_{N-(n+1)}\right)\right) \qquad (3.117)
\end{aligned}$$

dove

$$
\begin{aligned}
&W_{N-n}\left(S_1,\ldots,S_{N-(n+1)},uS_{N-(n+1)}\right)\\
&=\frac{1}{(1+r)^n}\left(\tilde{p}^nF_D\left(S_1,\ldots,S_{N-(n+1)},uS_{N-(n+1)},uS_{N-n},\ldots,uS_{N-2},uS_{N-1}\right)\right.\\
&+\tilde{p}^{n-1}\tilde{q}F_D\left(S_1,\ldots,S_{N-(n+1)},uS_{N-(n+1)},uS_{N-n},\ldots,uS_{N-2},dS_{N-1}\right)\\
&+\tilde{p}^{n-1}\tilde{q}F_D\left(S_1,\ldots,S_{N-(n+1)},uS_{N-(n+1)},uS_{N-n},\ldots,dS_{N-2},uS_{N-1}\right)\\
&+\tilde{p}^{n-2}\tilde{q}^2F_D\left(S_1,\ldots,S_{N-(n+1)},uS_{N-(n+1)},uS_{N-n},\ldots,dS_{N-2},dS_{N-1}\right)\\
&+\cdots\\
&+\tilde{p}^{n-1}\tilde{q}F_D\left(S_1,\ldots,S_{N-(n+1)},uS_{N-(n+1)},dS_{N-n},\ldots,uS_{N-2},uS_{N-1}\right)\\
&+\tilde{p}^{n-2}\tilde{q}^2F_D\left(S_1,\ldots,S_{N-(n+1)},uS_{N-(n+1)},dS_{N-n},\ldots,uS_{N-2},dS_{N-1}\right)\\
&+\tilde{p}^{n-2}\tilde{q}^2F_D\left(S_1,\ldots,S_{N-(n+1)},uS_{N-(n+1)},dS_{N-n},\ldots,dS_{N-2},uS_{N-1}\right)\\
&+\tilde{p}^{n-3}\tilde{q}^3F_D\left(S_1,\ldots,S_{N-(n+1)},uS_{N-(n+1)},dS_{N-n},\ldots,dS_{N-2},dS_{N-1}\right)\\
&+\cdots\\
&\left.+\tilde{q}^nF_D\left(S_1,\ldots,S_{N-(n+1)},uS_{N-(n+1)},dS_{N-n},\ldots,dS_{N-2},dS_{N-1}\right)\right)\\
&=\frac{1}{(1+r)^n}\sum_{(j_1,\ldots,j_n)\in\times_{k=1}^n\{0,1\}}\tilde{p}^{\Sigma_{k=1}^n j_k}\tilde{q}^{n-\Sigma_{k=1}^n j_k}\\
&\cdot F_D\left(S_1,\ldots,S_{N-(n+1)},uS_{N-(n+1)},u^{j_1}d^{1-j_1}S_{N-n},\ldots,u^{j_{n-1}}d^{1-j_{n-1}}S_{N-2},u^{j_n}d^{1-j_n}S_{N-1}\right)
\end{aligned}
\tag{3.118}
$$

e

$$
\begin{aligned}
&W_{N-n}\left(S_1,\ldots,S_{N-(n+1)},dS_{N-(n+1)}\right)\\
&=\frac{1}{(1+r)^n}\left(\tilde{p}^nF_D\left(S_1,\ldots,S_{N-(n+1)},dS_{N-(n+1)},uS_{N-n},\ldots,uS_{N-2},uS_{N-1}\right)\right.\\
&+\tilde{p}^{n-1}\tilde{q}F_D\left(S_1,\ldots,S_{N-(n+1)},dS_{N-(n+1)},uS_{N-n},\ldots,uS_{N-2},dS_{N-1}\right)\\
&+\tilde{p}^{n-1}\tilde{q}F_D\left(S_1,\ldots,S_{N-(n+1)},dS_{N-(n+1)},uS_{N-n},\ldots,dS_{N-2},uS_{N-1}\right)\\
&+\tilde{p}^{n-2}\tilde{q}^2F_D\left(S_1,\ldots,S_{N-(n+1)},dS_{N-(n+1)},uS_{N-n},\ldots,dS_{N-2},dS_{N-1}\right)\\
&+\cdots\\
&+\tilde{p}^{n-1}\tilde{q}F_D\left(S_1,\ldots,S_{N-(n+1)},dS_{N-(n+1)},dS_{N-n},\ldots,uS_{N-2},uS_{N-1}\right)\\
&+\tilde{p}^{n-2}\tilde{q}^2F_D\left(S_1,\ldots,S_{N-(n+1)},dS_{N-(n+1)},dS_{N-n},\ldots,uS_{N-2},dS_{N-1}\right)\\
&+\tilde{p}^{n-2}\tilde{q}^2F_D\left(S_1,\ldots,S_{N-(n+1)},dS_{N-(n+1)},dS_{N-n},\ldots,dS_{N-2},uS_{N-1}\right)\\
&+\tilde{p}^{n-3}\tilde{q}^3F_D\left(S_1,\ldots,S_{N-(n+1)},dS_{N-(n+1)},dS_{N-n},\ldots,dS_{N-2},dS_{N-1}\right)\\
&+\cdots\\
&\left.+\tilde{q}^nF_D\left(S_1,\ldots,S_{N-(n+1)},dS_{N-(n+1)},dS_{N-n},\ldots,dS_{N-2},dS_{N-1}\right)\right)\\
&=\frac{1}{(1+r)^n}\sum_{(j_1,\ldots,j_n)\in\times_{k=1}^n\{0,1\}}\tilde{p}^{\Sigma_{k=1}^n j_k}\tilde{q}^{n-\Sigma_{k=1}^n j_k}\\
&\cdot F_D\left(S_1,\ldots,S_{N-(n+1)},dS_{N-(n+1)},u^{j_1}d^{1-j_1}S_{N-n},\ldots,u^{j_{n-1}}d^{1-j_{n-1}}S_{N-2},u^{j_n}d^{1-j_n}S_{N-1}\right)
\end{aligned}
\tag{3.119}
$$

Combinando le Equazioni (3.117)-(3.119), possiamo allora scrivere

$$
\begin{aligned}
&W_{N-(n+1)} \\
&= \frac{1}{(1+r)^{n+1}} \Bigg(\tilde{p} \sum_{(j_1,\ldots,j_n)\in \times_{k=1}^{n}\{0,1\}} \tilde{p}^{n-\Sigma_{k=1}^{n} j_k} \tilde{q}^{\Sigma_{k=1}^{n} j_k} \\
&\quad \cdot F_D \left(S_1,\ldots,S_{N-(n+1)}, uS_{N-(n+1)}, u^{j_1} d^{1-j_1} S_{N-n},\ldots, u^{j_{n-1}} d^{1-j_{n-1}} S_{N-2}, u^{j_n} d^{1-j_n} S_{N-1}\right) \\
&\quad + \tilde{q} \sum_{(j_1,\ldots,j_n)\in \times_{k=1}^{n}\{0,1\}} \tilde{p}^{n-\Sigma_{k=1}^{n} j_k} \tilde{q}^{\Sigma_{k=1}^{n} j_k} \\
&\quad \cdot F_D \left(S_1,\ldots,S_{N-(n+1)}, dS_{N-(n+1)}, u^{j_1} d^{1-j_1} S_{N-n},\ldots, u^{j_{n-1}} d^{1-j_{n-1}} S_{N-2}, u^{j_n} d^{1-j_n} S_{N-1}\right) \Bigg) \\
&= \frac{1}{(1+r)^{n+1}} \sum_{(j_1,\ldots,j_{n+1})\in \times_{k=1}^{n+1}\{0,1\}} \tilde{p}^{n-\Sigma_{k=1}^{n} j_k} \tilde{q}^{\Sigma_{k=1}^{n} j_k} \\
&\quad \cdot F_D \left(S_1,\ldots,S_{N-(n+1)}, u^{j_1} d^{1-j_1} S_{N-(n+1)}, u^{j_2} d^{1-j_2} S_{N-n},\ldots, u^{j_{n-1}} d^{1-j_{n-1}} S_{N-2}, u^{j_{n+1}} d^{1-j_{n+1}} S_{N-1}\right) \\
&= \frac{1}{(1+r)^{n+1}} \tilde{\mathbf{E}} \left[F_D \left(S_1,\ldots,S_{N-(n+1)}, S_{N-n}, S_{N-n+1},\ldots,S_{N-1}, S_N\right) \mid S_1,\ldots,S_{N-(n+1)} \right] \\
&= \frac{1}{(1+r)^{n+1}} \tilde{\mathbf{E}}_{N-(n+1)} \left[F_D \left(S_1,\ldots,S_N\right) \right]. \qquad (3.120)
\end{aligned}
$$

Pertanto, il valore del portafoglio replicante al generico tempo $t = t_{N-n}$, con $0 \leqslant n \leqslant N-1$, uguaglia il valore neutrale al rischio del derivato allo stesso tempo. In particolare, posto $n = N-1$, otteniamo che il valore

$$W_0 = X_1 B_0 + Y_1 S_0 \equiv W_0 (S_0)$$

del portafoglio replicante al tempo $t = t_0 \equiv 0$ è anche dato da

$$
\begin{aligned}
W_0 &= \frac{1}{(1+r)^{N}} \sum_{(j_1,\ldots,j_N)\in \times_{k=1}^{N}\{0,1\}} \tilde{p}^{N-\Sigma_{k=1}^{N} j_k} \tilde{q}^{\Sigma_{k=1}^{N} j_k} \\
&\quad \cdot F_D \left(u^{j_1} d^{1-j_1} S_0, u^{j_2} d^{1-j_2} S_1 \ldots, u^{j_{N-1}} d^{1-j_{N-1}} S_{N-2}, u^{j_N} d^{1-j_N} S_{N-1}\right) \\
&= \frac{1}{(1+r)^{N}} \sum_{(j_1,\ldots,j_N)\in \times_{k=1}^{N}\{0,1\}} \tilde{p}^{N-\Sigma_{k=1}^{N} j_k} \tilde{q}^{\Sigma_{k=1}^{N} j_k} \\
&\quad \cdot F_D \left(u^{j_1} d^{1-j_1} S_0, u^{j_1+j_2} d^{1-(j_1+j_2)} S_0 \ldots, u^{\Sigma_{k=1}^{N-1} j_k} d^{N-\Sigma_{k=1}^{N-1} j_k} S_0, u^{\Sigma_{k=1}^{N} j_k} d^{N-\Sigma_{k=1}^{N} j_k} S_0\right) \\
&= \frac{1}{(1+r)^{N}} \tilde{\mathbf{E}} \left[F_D \left(S_1,\ldots,S_N\right) \right], \qquad (3.121)
\end{aligned}
$$

uguagliando il valore neutrale al rischio del derivato stesso allo stesso tempo. Inoltre, le componenti X_1 e Y_1 del portafoglio replicante costituite tra il tempo $t = t_0$ e $t = t_1$ sono date da

$$X_1 = \frac{uW_1 (dS_0) - dW_1 (uS_0)}{(1+r)(u-d) B_0}, \qquad Y_1 = \frac{W_1 (uS_0) - W_1 (dS_0)}{(u-d) S_0}, \qquad (3.122)$$

dove il valore

$$W_1 \equiv W_1 (S_1)$$

del portafoglio replicante al tempo $t = t_1$ è anche dato da

$$W_1 = \frac{1}{(1+r)^{N-1}} \sum_{(j_1,\ldots,j_{N-1})\in \times_{k=1}^{N-1}\{0,1\}} \tilde{p}^{\Sigma_{k=1}^{N-1} j_k} \tilde{q}^{(N-1)-\Sigma_{k=1}^{N-1} j_k}$$
$$\cdot F_D \left(u^{j_1} d^{1-j_1} S_1, u^{j_1+j_2} d^{1-(j_1+j_2)} S_1 \ldots, u^{\Sigma_{k=1}^{N-2} j_k} d^{(N-2)-\Sigma_{k=1}^{N-2} j_k} S_1, u^{\Sigma_{k=1}^{N-1} j_k} d^{(N-1)-\Sigma_{k=1}^{N-1} j_k} S_1 \right) \tag{3.123}$$

Anche in questo caso si può verificare che le componenti X_1 e Y_1 del portafoglio replicante scelte tra il tempo $t = t_0$ e il tempo $t = t_1$ sono $\mathcal{F}_0 - \mathcal{B}(\mathbb{R})$ variabili aleatorie e che è verificata l'eguaglianza

$$X_1 B_1 + Y_1 S_1 = W_1.$$

□

Definizione 210 *In riferimento al modello di mercato CRR multiperiodale, in assenza di BS-portafogli d'arbitraggio, chiamiamo* prezzo di non arbitraggio *di un derivato D al tempo $t = t_n$, per ogni $n = 0, 1, \ldots, N$, il valore del portafoglio replicante allo stesso tempo.*

Osservazione 211 *In riferimento al modello di mercato CRR multiperiodale, in assenza di BS-portafogli d'arbitraggio, il prezzo di non arbitraggio di un derivato D al tempo $t = t_n$, è univocamente determinato, per ogni $n = 0, 1, \ldots, N$.*
Proof. *È sufficiente osservare che, in assenza di BS-portafogli d'arbitraggio, la dimostrazione del Teorema 209 mostra come il prezzo di non arbitraggio di un derivato D al tempo $t = t_n$ coincida con il prezzo neutrale al rischio del derivato allo stesso tempo e che si ha l'unicità della probabilità neutrale al rischio.* □

Corollary 212 *Siano U e V due derivati europei del modello di mercato CRR multiperiodale privo di BS-portafogli d'arbitraggio e siano $(U_n)_{n=0}^N$ e $(V_n)_{n=0}^N$ le successioni dei prezzi di non arbitraggio di U e V rispettivamente. Il sussistere dell'uguaglianza*

$$U_N = V_N \tag{3.124}$$

comporta necessariamente che

$$U_n = V_n, \quad n = 0, 1, \ldots, N-1. \tag{3.125}$$

Proof. *In assenza di BS-portafogli d'arbitraggio il modello di mercato CRR multiperiodale è completo quindi esistono $\Pi_U \equiv \left(\left(X_n^{(U)} \right)_{n=1}^N, \left(Y_n^{(U)} \right)_{n=1}^N \right)$ e $\Pi_V \equiv \left(\left(X_n^{(V)} \right)_{n=1}^N, \left(Y_n^{(V)} \right)_{n=1}^N \right)$ portafogli replicanti dei derivati U e V, rispettivamente. Si ha allora*

$$U_0 = X_1^{(U)} B_0 + Y_1^{(U)} S_0, \qquad V_0 = X_1^{(V)} B_0 + Y_1^{(V)} S_0,$$

e

$$U_n = X_n^{(U)} B_n + Y_n^{(U)} S_n, \qquad V_n = X_n^{(V)} B_n + Y_n^{(V)} S_n,$$

per ogni $n = 1, \ldots, N$. D'altra parte dalla dimostrazione del Teorema 209 risulta

$$X_1^{(U)} B_0 + Y_1^{(U)} S_0 = \frac{1}{(1+r)^N} \mathbf{E}[U_N], \qquad X_1^{(V)} B_0 + Y_1^{(V)} S_0 = \frac{1}{(1+r)^N} \mathbf{E}[V_N], \tag{3.126}$$

e

$$X_n^{(U)} B_n + Y_n^{(U)} S_n = \frac{1}{(1+r)^{N-n}} \mathbf{E}[U_N], \qquad X_n^{(V)} B_n + Y_n^{(V)} S_n = \frac{1}{(1+r)^{N-n}} \mathbf{E}[V_N], \tag{3.127}$$

per ogni $n = 1, \ldots, N$. Combinando le Equazioni (3.126) e (3.127) con la (3.124), si ottiene chiaramente la (3.125). □

3.5.2 European Options

Consideriamo adesso un'opzione europea d'acquisto (*call*) sul titolo rischioso di maturità T, prezzo d'esercizio K e valore di mercato $(C_n)_{n=0}^N$, di cui C_0 sia il *premio*. Anche nel caso di un mercato multiperiodale il valore dell'opzione alla maturità è dato da

$$C_N \equiv C_T \overset{\text{def}}{=} \max\{S_T - K, 0\} \equiv \max\{S_N - K, 0\}.$$

D'altra parte S_N è dato dalla (3.39). Pertanto posto

$$n_K \equiv \min \left\{ n \in \{0, \dots, N\} \mid u^n d^{N-n} S_0 \geqslant K \right\},$$

ovvero

$$n_K \equiv \left\lceil \frac{\log(K) - (\log(S_0) + N \log(d))}{\log(u) - \log(d)} \right\rceil,$$

dove $\lceil \cdot \rceil : \mathbb{R} \to \mathbb{R}$ è la funzione ceiling, risulta

$$C_N = \begin{cases} 0, & \text{se } n = 0, \dots, n_K - 1, \\ u^n d^{N-n} S_0 - K, & \text{se } n = n_K, \dots, N, \end{cases} \tag{3.128}$$

con

$$\mathbf{P}(C_N = 0) = \sum_{n=0}^{n_K - 1} \binom{N}{n} p^n q^{N-n} = 1 - \sum_{n=n_K}^{N} \binom{N}{n} p^n q^{N-n},$$
$$\mathbf{P}(C_N = u^n d^{N-n} S_0 - K) = \binom{N}{n} p^n q^{N-n}, \quad \text{se } n = n_K, \dots, N.$$

e, in riferimento alla misura neutrale al rischio, con

$$\tilde{\mathbf{P}}(C_N = 0) = \sum_{n=0}^{n_K - 1} \binom{N}{n} \tilde{p}^n \tilde{q}^{N-n} = 1 - \sum_{n=n_K}^{N} \binom{N}{n} \tilde{p}^n \tilde{q}^{N-n},$$
$$\mathbf{P}(C_N = u^n d^{N-n} S_0 - K) = \binom{N}{n} \tilde{p}^n \tilde{q}^{N-n}, \quad \text{se } n = n_K, \dots, N.$$

Definizione 213 *Come caso particolare della Definizione 208, chiamiamo* portafoglio replicante *(*replicating portfolio*) o* strategia di copertura *(*hedging strategy*) dell'opzione call un BS-portafoglio autofinanziante* $\Pi \equiv ((X_n)_{n=0}^N, (Y_n)_{n=0}^N)$ *tale che all'istante terminale* $t_N = T$ *si abbia*

$$X_N(\omega) B_N(\omega) + Y_N(\omega) S_N(\omega) = C_N(\omega), \tag{3.129}$$

per ogni $\omega \in \Omega$.

Osservazione 214 *In assenza di BS-portafogli d'arbitraggio, se* $\Pi \equiv ((X_n)_{n=0}^N, (Y_n)_{n=0}^N)$ *è un portafoglio replicante dell'opzione call risulta*

$$X_n B_n + Y_n S_n = C_n,$$

per ogni $n = 0, 1 \dots, N$.

E' poi possibile dimostrare che in assenza di BSS-portafogli d'arbitraggio, il portafoglio replicante risulta essere unico.

Teorema 215 *In assenza di BS-portafogli d'arbitraggio, esiste un unico portafoglio replicante dell'opzione call,* $\Pi \equiv ((X_n)_{n=0}^N, (Y_n)_{n=0}^N)$*, dato da*

$$X_n = \frac{uC_n^- - dC_n^+}{(u-d)B_n}, \quad Y_n = \frac{C_n^+ - C_n^-}{(u-d)S_{n-1}}, \tag{3.130}$$

per ogni $n = 0, 1, \ldots, N$*, essendo* C_n^-*,* C_n^+ *le possibili realizzazioni della call al tempo* $t = t_n$ *dato il valore* S_{n-1} *del sottostante al tempo* $t = t_{n-1}$*. Inoltre*

$$W_n = C_n = \frac{1}{(1+r)^{N-n}} \tilde{\mathbf{E}}_n[C_N], \tag{3.131}$$

per ogni $n = 0, 1 \ldots, N$.

Proof. *Se* $\Pi \equiv ((X_n)_{n=0}^N, (Y_n)_{n=0}^N)$ *è un portafoglio replicante dell'opzione call deve aversi*

$$X_N B_N + Y_N S_N = \max\{S_N - K, 0\}. \tag{3.132}$$

D'altra parte, in riferimento al tempo $t = t_{N-1}$*,*

$$S_N = \beta_N S_{N-1} = \begin{cases} uS_{N-1}, & \mathbf{P}(S_N = uS_{N-1}) \equiv \tilde{p} \\ dS_{N-1}, & \mathbf{P}(S_N = dS_{N-1}) \equiv \tilde{q} \end{cases}$$

per cui la (3.132) da luogo alle condizioni

$$\begin{aligned} X_N B_N + Y_N u S_{N-1} &= \max\{uS_{N-1} - K, 0\} \equiv C_N^+, \\ X_N B_N + Y_N d S_{N-1} &= \max\{dS_{N-1} - K, 0\} \equiv C_N^-. \end{aligned} \tag{3.133}$$

Essendo C_N^+ *e* C_N^- *le possibili realizzazioni della call al tempo* $t = t_N$ *dato il valore* S_{N-1} *del sottostante al tempo* $t = t_{N-1}$*. Il sistema (3.133) ammette un'unica soluzione* (X_N, Y_N) *data da*

$$X_N = \frac{\begin{vmatrix} C_N^+ & uS_{N-1} \\ C_N^- & dS_{N-1} \end{vmatrix}}{\begin{vmatrix} B_N & uS_{N-1} \\ B_N & dS_{N-1} \end{vmatrix}} = \frac{uC_N^- - dC_N^+}{(u-d)B_N}, \tag{3.134}$$

e

$$Y_N = \frac{\begin{vmatrix} B_N & C_N^+ \\ B_N & C_N^- \end{vmatrix}}{\begin{vmatrix} B_N & uS_{N-1} \\ B_N & dS_{N-1} \end{vmatrix}} = \frac{C_N^+ - C_N^-}{(u-d)S_{N-1}}. \tag{3.135}$$

Da notare che le condizioni (3.134) e (3.135) consentono di esprimere deterministicamente le componenti X_N *ed* Y_N *del portafoglio replicante, noto che sia il valore* S_{N-1} *dello stock al tempo* $t = t_{N-1}$*. In altri temnini, al tempo* $t = t_{N-1}$*, osservata la realizzazione del valore* S_{N-1} *del titolo rischioso, ed essendo certo il valore* B_N *del bond, possiamo costruire un portafoglio che replichi il valore dell'opzione call alla maturità, qualunque possa essere il valore futuro* S_N *del titolo rischioso. D'altra parte, se valutate al tempo iniziale* $t = t_0$ *le componenti* X_N *ed* Y_N *sono variabili aleatorie in quanto funzioni (deterministiche) della variabile aleatoria* S_{N-1}*. Da notare inoltre che* X_N *ed* Y_N *non dipendono dalla probabilità* $\tilde{p}$ *di crescita del sottostante.*

Determinati X_N *ed* Y_N*, la condizione di autofinanziamento*

$$W_{N-1} = X_{N-1}B_{N-1} + Y_{N-1}S_{N-1} = X_N B_{N-1} + Y_N S_{N-1},$$

consente allora di calcolare il valore del portafoglio replicante al tempo $t = t_{N-1}$. Precisamente,

$$\begin{aligned}
W_{N-1} &\equiv X_N B_{N-1} + Y_N S_{N-1} \\
&= \frac{uC_N^- - dC_N^+}{(u-d)B_N} B_{N-1} + \frac{C_N^+ - C_N^-}{(u-d)S_{N-1}} S_{N-1} \\
&= \frac{uC_N^- - dC_N^+}{(u-d)(1+r)} + \frac{C_N^+ - C_N^-}{u-d} \\
&= \frac{1}{1+r}\left(\frac{uC_N^- - dC_N^+}{u-d} + (1+r)\frac{C_N^+ - C_N^-}{u-d}\right) \\
&= \frac{1}{1+r}\left(C_N^+\frac{(1+r)-d}{u-d} + C_N^-\frac{u-(1+r)}{u-d}\right) \\
&= \frac{1}{1+r}\left(C_N^+\tilde{p} + C_N^-\tilde{q}\right) \\
&= \frac{1}{1+r}\left(\max\{uS_{N-1} - K, 0\}\tilde{p} + \max\{dS_{N-1} - K, 0\}\tilde{q}\right).
\end{aligned}$$

Grazie alle (3.89) e (3.90), risulta allora

$$W_{N-1} = \frac{1}{1+r}\tilde{\mathbf{E}}[C_N \mid S_{N-1}] = \frac{1}{1+r}\tilde{\mathbf{E}}[C_N \mid \mathcal{F}_{N-1}].$$

Il ragionamento precedente può essere ripetuto per calcolare X_{N-1} e Y_{N-1}. Infatti, sempre grazie al Corollario **??**, *da una parte deve aversi*

$$W_{N-1} = C_{N-1}. \tag{3.136}$$

D'altra parte, in dipendenza dal valore che potrebbe assumere S_{N-2}, il primo membro della (3.136) può assumere soltanto i valori

$$X_{N-1}B_{N-1} + Y_{N-1}uS_{N-2} \quad e \quad X_{N-1}B_{N-1} + Y_{N-1}dS_{N-2}.$$

Pertanto, l'Equazione (3.136) comporta che la call al tempo $t = t_{N-1}$ possa assume soltanto i due valori

$$C_{N-1}^+ = X_{N-1}B_{N-1} + Y_{N-1}uS_{N-2} \quad e \quad C_{N-1}^- = X_{N-1}B_{N-1} + Y_{N-1}dS_{N-2}.$$

Con un calcolo del tutto simile a quello effettuato per determinare X_N e Y_N, possiamo determinare allora

$$X_{N-1} = \frac{uC_{N-1}^- - dC_{N-1}^+}{(u-d)B_{N-1}} \quad e \quad Y_{N-1} = \frac{C_{N-1}^+ - C_{N-1}^-}{(u-d)S_{N-2}}.$$

Quindi, sempre per la condizione di autofinanziamento,

$$W_{N-2} = X_{N-2}B_{N-2} + Y_{N-2}S_{N-2} = X_{N-1}B_{N-2} + Y_{N-1}S_{N-2},$$

il valore del portafoglio replicante al tempo $t = t_{N-2}$ è dato da

$$\begin{aligned}
W_{N-2} &= X_{N-1}B_{N-2} + Y_{N-1}S_{N-2} \\
&= \frac{uC_{N-1}^- - dC_{N-1}^+}{(u-d)B_{N-1}} B_{N-2} + \frac{C_{N-1}^+ - C_{N-1}^-}{(u-d)S_{N-2}} S_{N-2} \\
&= \frac{uC_{N-1}^- - dC_{N-1}^+}{(u-d)(1+r)} + \frac{C_{N-1}^+ - C_{N-1}^-}{u-d} \\
&= \frac{1}{1+r}\left(\frac{uC_{N-1}^- - dC_{N-1}^+}{u-d} + (1+r)\frac{C_{N-1}^+ - C_{N-1}^-}{u-d}\right) \\
&= \frac{1}{1+r}\left(C_{N-1}^+\frac{(1+r)-d}{u-d} + C_{N-1}^-\frac{u-(1+r)}{u-d}\right) \\
&= \frac{1}{1+r}\left(C_{N-1}^+\tilde{p} + C_{N-1}^-\tilde{q}\right).
\end{aligned}$$

Abbiamo quindi, ancora per la proprietà di Markov (cfr. (3.89) e (3.90)) e le proprietà dell'operatore speranza condizionata,

$$
\begin{aligned}
W_{N-2} &= \frac{1}{1+r}\tilde{\mathbf{E}}[C_{N-1} \mid S_{N-2}] \\
&= \frac{1}{1+r}\tilde{\mathbf{E}}[\frac{1}{1+r}\tilde{\mathbf{E}}[C_N \mid S_{N-1}] \mid S_{N-2}] \\
&= \frac{1}{(1+r)^2}\tilde{\mathbf{E}}[\tilde{\mathbf{E}}[C_N \mid S_{N-1}] \mid S_{N-2}] \\
&= \frac{1}{(1+r)^2}\tilde{\mathbf{E}}[\tilde{\mathbf{E}}[C_N \mid \mathfrak{F}_{N-1}] \mid \mathfrak{F}_{N-2}] \\
&= \frac{1}{(1+r)^2}\tilde{\mathbf{E}}[C_N \mid \mathfrak{F}_{N-2}].
\end{aligned}
$$

Non è difficile rendersi conto che iterando il ragionamento si ottiene

$$X_{N-n} = \frac{uC^-_{N-n} - dC^+_{N-n}}{(u-d)B_{N-n}} \quad e \quad X_{N-n} = \frac{C^+_{N-n} - C^-_{N-n}}{(u-d)S_{N-n-1}}, \quad n = 0, \ldots, N-1.$$

e

$$W_{N-n} = C_{N-n} = \frac{1}{(1+r)^n}\tilde{\mathbf{E}}[C_N \mid \mathcal{F}_{N-n}], \quad n = 0, \ldots, N-1.$$

Infine, otteniamo

$$W_0 = C_0 = \frac{1}{(1+r)^N}\tilde{\mathbf{E}}[C_N \mid \mathcal{F}_0]$$

La validità delle (3.130) e (3.131) segue immediatamente scambiando $N-n$ con n. □

Corollary 216 *Si ha*

$$C_n \geqslant 0 \tag{3.137}$$

per ogni $n = 0, 1, \ldots, N$.
Proof. *La (3.137) è immediata conseguenza dell'Equazione 3.131, stanti la positività di C_N e la positività dell'operatore speranza condizionata $\tilde{\mathbf{E}}_n[\cdot]$.* □

Definizione 217 *Chiamiamo* prezzo di non arbitraggio *(*no arbitrage price*) al tempo $t = t_n$ dell'opzione call il valore del portafoglio replicante $\Pi \equiv ((X_n)_{n=0}^N, (Y_n)_{n=0}^N)$ al tempo $t = t_n$*

$$C_n = X_n B_n + Y_n S_n, \quad n = 1, \ldots, N. \tag{3.138}$$

Proposizione 218 *Risulta*

$$C_0 = \frac{1}{(1+r)^N}\sum_{n=n_K}^{N}\left(u^n d^{N-n} S_0 - K\right)\binom{N}{n}\tilde{p}^n\tilde{q}^{N-n}, \tag{3.139}$$

dove

$$n_K \equiv \min\{n \in \{0, \ldots, N\} \mid u^n d^{N-n} S_0 \geqslant K\}, \tag{3.140}$$

ovvero

$$n_K \equiv \left\lceil \frac{\log(K) - (\log(S_0) + N\log(d))}{\log(u) - \log(d)} \right\rceil, \tag{3.141}$$

dove $\lceil\cdot\rceil : \mathbb{R} \to \mathbb{R}$ è la funzione ceiling, e

$$\tilde{p} = \frac{r+1-d}{u-d}, \quad \tilde{q} = \frac{u-(r+1)}{u-d}.$$

Proof. *Dalla (3.131) considerata nel caso $n = 0$ e dalla (3.128) segue*

$$\begin{aligned}
C_0 &= \frac{1}{(1+r)^N} \tilde{\mathbf{E}}[C_N \mid \mathfrak{F}_0] \\
&= \frac{1}{(1+r)^N} \tilde{\mathbf{E}}[C_N] \\
&= \frac{1}{(1+r)^N} \sum_{n=n_K}^{N} \left(u^n d^{N-n} S_0 - K\right) \binom{N}{n} \tilde{p}^n \tilde{q}^{N-n}.
\end{aligned}$$

□

Corollary 219 *Si ha*

$$C_n \geqslant \left(S_n - \frac{K}{(1+r)^{N-n}}\right)^+, \tag{3.142}$$

per ogni $n = 0, 1, \ldots, N$.

Proof. *Stanti le proprietà dell'operatore speranza condizionata, dalla (3.131), otteniamo*

$$\begin{aligned}
C_n &= \frac{1}{(1+r)^{N-n}} \tilde{\mathbf{E}}_n[C_N] \\
&= \frac{1}{(1+r)^{N-n}} \tilde{\mathbf{E}}_n[\max\{S_N - K, 0\}] \\
&= \frac{1}{(1+r)^{N-n}} \tilde{\mathbf{E}}_n[\frac{1}{2}\left(|S_N - K| + S_N - K\right)] \\
&= \frac{1}{(1+r)^{N-n}} \frac{1}{2} \left(\tilde{\mathbf{E}}_n[|S_N - K|] + \tilde{\mathbf{E}}_n[S_N - K]\right) \\
&= \frac{1}{(1+r)^{N-n}} \frac{1}{2} \left(\tilde{\mathbf{E}}_n[|S_n \beta_{n+1} \cdots \beta_N - K|] + \tilde{\mathbf{E}}_n[S_n \beta_{n+1} \cdots \beta_N - K]\right) \\
&\geqslant \frac{1}{(1+r)^{N-n}} \frac{1}{2} \left(\left|\tilde{\mathbf{E}}_n[S_n \beta_{n+1} \cdots \beta_N - K]\right| + \tilde{\mathbf{E}}_n[S_n \beta_{n+1} \cdots \beta_N - K]\right) \\
&= \frac{1}{(1+r)^{N-n}} \frac{1}{2} \left(\left|S_n \tilde{\mathbf{E}}_n[\beta_{n+1} \cdots \beta_N] - K\right| + S_n \tilde{\mathbf{E}}_n[\beta_{n+1} \cdots \beta_N] - K\right) \\
&= \frac{1}{(1+r)^{N-n}} \frac{1}{2} \left(\left|S_n \tilde{\mathbf{E}}[\beta_{n+1} \cdots \beta_N] - K\right| + S_n \tilde{\mathbf{E}}[\beta_{n+1} \cdots \beta_N] - K\right) \\
&= \frac{1}{(1+r)^{N-n}} \frac{1}{2} \left(\left|S_n \tilde{\mathbf{E}}[\beta_{n+1}] \cdots \tilde{\mathbf{E}}[\beta_N] - K\right| + S_n \tilde{\mathbf{E}}[\beta_{n+1}] \cdots \tilde{\mathbf{E}}[\beta_N] - K\right) \\
&= \frac{1}{(1+r)^{N-n}} \frac{1}{2} \left(\left|S_n \left(u\tilde{p} + d\tilde{q}\right)^{N-n} - K\right| + S_n \left(u\tilde{p} + d\tilde{q}\right)^{N-n} - K\right)
\end{aligned} \tag{3.143}$$

D'altra parte vale la

$$u\tilde{p} + d\tilde{q} = 1 + r, \tag{3.144}$$

e combinandola con la (3.143) otteniamo

$$\begin{aligned}
C_n &\geqslant \frac{1}{(1+r)^{N-n}} \frac{1}{2} \left(\left|S_n(1+r)^{N-n} - K\right| + S_n(1+r)^{N-n} - K\right) \\
&= \frac{1}{2} \left(\left|S_n - \frac{K}{(1+r)^{N-n}}\right| + S_n - \frac{K}{(1+r)^{N-n}}\right) \\
&= \max\left\{S_n - \frac{K}{(1+r)^{N-n}}, 0\right\}.
\end{aligned}$$

come desiderato. □

Il prezzo di non arbitraggio di un'opzione call europea è definito nel caso di un mercato multiperiodale binomiale sulla base di considerazioni analoghe a quelle effettuate nel caso di mercato monoperiodale. Immaginiamo che un operatore finanziario ad inizio dei periodi di contrattazione venda un'opzione call europea sul titolo rischioso con prezzo d'esercizio K e maturità T. L'agente realizza l'incasso C_0 ma si espone al rischio d'esercizio dell'opzione alla maturità da parte dell'acquirente per un ammontare pari a $C_T = \max\{S_T - K, 0\}$. Per eliminare il rischio dovuto a tale esposizione, sfruttando la sola disponibilità C_0 generata dalla vendita dell'opzione, l'operatore finanziario costituisce un portafoglio acquistando X_1 unità del titolo non rischioso al prezzo B_0 e acquistando Y_1 azioni del titolo rischioso al prezzo S_0 (valori negativi di X_1 e Y_1 significano rispettivamente un prestito sul titolo non rischioso e la vendita allo scoperto del titolo rischioso). Questo portafoglio deve essere tale da replicare il valore atteso (aleatorio) C_1 della call al tempo $t = t_1$ dato S_0. A fine periodo, il valore S_1 dello stock si realizza ed il portafoglio costituito immediatamente dopo il tempo $t = 0$ prende il nuovo valore $X_1 B_1 + Y_1 S_1 \equiv V_1$. Con tale ricchezza l'operatore finanziario riconfigura il portafoglio in modo da replicare il valore atteso C_2 della call al tempo $t = t_2$, dato il prezzo S_1 dello stock. L'iterazione di questo processo mette l'operatore finanziario in grado di replicare il valore della call alla maturità T quale che sia il valore S_T assunto dallo stock.

Supponiamo adesso che nel mercato sia trattata anche un'opzione *put* europea sul titolo rischioso, sempre con prezzo d'esercizio K alla maturità T, la cui successione dei valori di mercato denotiamo con $(P_n)_{n=0}^N$, essendo

$$P_N \equiv P_T \overset{\text{def}}{=} \max\{K - S_T, 0\} \equiv \max\{K - S_N, 0\}.$$

D'altra parte S_N è sempre dato dalla (3.39). Pertanto posto

$$\check{n}_K \equiv \max\left\{n \in \{0, \dots, N\} \mid K \geqslant u^n d^{N-n} S_0\right\},$$

ovvero

$$\check{n}_K \equiv \left\lfloor \frac{\log(K) - (\log(S_0) + N\log(d))}{\log(u) - \log(d)} \right\rfloor,$$

dove $\lfloor\cdot\rfloor : \mathbb{R} \to \mathbb{R}$ è la funzione floor, risulta

$$P_N = \begin{cases} K - u^n d^{N-n} S_0, & \text{se } n = 0, \dots, \check{n}_K, \\ 0, & \text{se } n = \check{n}_K + 1, \dots, N, \end{cases}$$

con

$$\mathbf{P}(P_N = K - u^n d^{N-n} S_0) = \binom{N}{n} p^n q^{N-n}, \quad \text{se } n = 0, \dots, \check{n}_K,$$

$$\mathbf{P}(P_N = 0) = \sum_{n=\check{n}_K+1}^{N} \binom{N}{n} p^n q^{N-n} = 1 - \sum_{n=0}^{\check{n}_K} \binom{N}{n} p^n q^{N-n}.$$

ovvero, in assenza di BS-portafogli d'arbitraggio, con

$$\tilde{\mathbf{P}}(P_N = K - u^n d^{N-n} S_0) = \binom{N}{n} \tilde{p}^n \tilde{q}^{N-n}, \quad \text{se } n = 0, \dots, \check{n}_K,$$

$$\tilde{\mathbf{P}}(P_N = 0) = \sum_{n=\check{n}_K+1}^{N} \binom{N}{n} \tilde{p}^n \tilde{q}^{N-n} = 1 - \sum_{n=0}^{\check{n}_K} \binom{N}{n} \tilde{p}^n \tilde{q}^{N-n},$$

Come nel caso delle opzioni call abbiamo

$$P_n = \frac{1}{(1+r)^{N-n}} \tilde{\mathbf{E}}_n[P_N] \tag{3.145}$$

e

$$P_n \geqslant 0 \tag{3.146}$$

per ogni $n = 0, 1, \dots, N$.

Proposizione 220 *Si ha chiaramente*

$$C_T - P_T = S_T - K. \tag{3.147}$$

Proof. *La dimostrazione è del tutto analoga a quella della Proposizione 2.49 relativa al modello CRR monoperiodale.* □

Proposizione 221 *In assenza di BSS-portafogli d'arbitraggio, sussiste necessariamente la relazione di parità put-call*

$$C_n - P_n = S_n - \frac{K}{(1+r)^{N-n}}, \tag{3.148}$$

per ogni $n = 0, \ldots, N$.

Proof. *Stante l'Equazione (3.147) possiamo scrivere*

$$C_T = P_T + S_T - K.$$

Allora, stante l'ipotesi di assenza di BS-prtafogli d'arbitraggio, il Corollario 212 implica che i derivati di pay-off C_T *e* $P_T + S_T - K$*, costituiti rispettivamente da un'opzione call* C *e da un portafoglio contenente un'opzione put* P*, uno share di stock* S *e un bond* B *venduto allo scoperto per un ammontare di* $\frac{K}{(1+r)^N}$ *debbano avere gli stessi prezzi di non arbitraggio al tempo* $t = t_n$*, per ogni* $n = 0, 1, \ldots, N$*. A loro volta, i prezzi di non arbitraggio devono coincidere con i prezzi neutrali al rischio. Pertanto, deve aversi*

$$C_n = \frac{1}{(1+r)^{N-n}} \mathbf{E}_n \left[C_T\right] = \frac{1}{(1+r)^{N-n}} \mathbf{E}_n \left[P_T + S_T - K\right]. \tag{3.149}$$

La proprietà di linearità dell'operatore speranza condizionata, comporta che

$$\mathbf{E}_n \left[P_T + S_T - K\right] = \mathbf{E}_n \left[P_T\right] + \mathbf{E}_n \left[S_T\right] - \mathbf{E}_n \left[K\right]. \tag{3.150}$$

Combinando le Equazioni (3.149) e (3.150), tenendo conto che il prezzo di non arbitraggio dell'opzione put coincide con il prezzo neutrale al rischio, otteniamo

$$C_n = P_n + \frac{1}{(1+r)^{N-n}} \mathbf{E}_n \left[S_T\right] - \frac{K}{(1+r)^{N-n}}$$

ovvero la (3.148). □

L'introduzione dell'opzione put e l'aver stabilito la relazione di parità put-call (3.148) in modo indipendente dalla struttura binomiale moltiplicativa del modello CRR consente di stabilire anche l'Equazione (3.142) in modo indipendente dalla struttura binomiale moltiplicativa del modello CRR, purchè si possa provare l'esistenza di una probabilità neutrale al rischio per la quale valgano la (3.131) e (3.145). Infatti, grazie alla (3.148) e alla (3.146) possiamo scrivere

$$C_n = S_n - \frac{K}{(1+r)^{N-n}} + P_n \geqslant S_n - \frac{K}{(1+r)^{N-n}},$$

per ogni $n = 0, \ldots, N-1$. Ma allora, tenuto conto della (3.137) segue immediatamente la (3.142). Similarmente abbiamo

$$P_n = \frac{K}{(1+r)^{N-n}} - S_n + C_n \geqslant \frac{K}{(1+r)^{N-n}} - S_n,$$

per ogni $n = 0, \ldots, N-1$.

La circostanza che in assenza di arbitraggi nel modello di mercato multiperiodale binomiale un'opzione call europea, e quindi anche un'opzione put, sia replicabile è un caso particolare della più generale proprietà di *completezza* di un tale mercato.

Definizione 222 *Diciamo che un mercato è* completo *se ogni derivato è replicabile.*

Si può provare che

Proposizione 223 *Nel modello CRR multiperiodale la proprietà di completezza comporta l'assenza di BS-portafogli d'arbitraggio.*

Proof. *Per rendersi conto della validità dell'asserto è sufficente osservare che dalla dimostrazione del Teorema 215 si evince come la proprietà di completezza di mercato dia modo di costruire una probabilità neutrale al rischio. In virtù dell'esistenza di una tale probabilità la Proposizione 199 assicura allora la non esistenza di BS-portafogli d'arbitraggio.* $\square$

Il Teorema 215 e la Proposizione 223 posono essere riassunti nel fondamentale risultato

Teorema 224 *Nel modello CRR multiperiodale l'assenza di BS-portafogli d'arbitraggio è equivalente alla completezza di mercato.*

In definitiva, nel modello CRR multiperiodale l'assenza di BS-portafogli d'arbitraggio, l'esistenza di un'unica probabilità neutrale al rischio e la replicabilità di ogni derivato sono proprietà equivalenti.

3.5.3 American Options

Consideriamo ancora il modello di mercato CRR multiperiodale articolato in N periodi di contrattazione, individuati dalla successione finita di tempi $0 \equiv t_0 < t_1 < \cdots < t_{n-1} < t_N \equiv T$, con $\Delta t \equiv t_n - t_{n-1} = T/N$, per ogni $n = 1, \ldots, N$, in cui sia possibile investire in un bond B di tasso di rendimento non rischioso $r^{(B)}_{n-1,n} \equiv r > 0$, per ogni $n = 1, \ldots, N$, e in uno stock S, che non distribuisca dividendi, di tasso di rendimento rischioso $r^{(S)}_{n-1,n} \equiv (u-1, d-1)$, per opportuni $u, d > 0$, tali che $u > 1 + r > d$, per ogni $n = 1, \ldots, N$. Sia quindi $\left(\Omega, \mathcal{E}, \tilde{\mathbf{P}}\right) \equiv \tilde{\Omega}$ lo spazio di probabilità CRR dotato della probabilità neutrale al rischio $\tilde{\mathbf{P}} : \mathcal{E} \to \mathbb{R}_+$ e della filtrazione $(\mathcal{F}_n)_{n=0}^N \equiv \mathfrak{F}$ generata dal processo di rumore $(\beta_n)_{n=1}^N$ o equivalentemente dal processo dei prezzi $(S_n)_{n=0}^N$ (cfr Sezione 3.5). Ricordiamo che la probabilità neutrale al rischio $\tilde{\mathbf{P}}$ è caratterizzata dalla distribuzione

$$(\tilde{p}, \tilde{q}) \equiv \left(\frac{1+r-d}{u-d}, \frac{u-(1+r)}{u-d}\right),$$

tale che

$$\tilde{\mathbf{P}}\left(r^{(S)}_{n-1,n} = u-1\right) = \tilde{p} \quad \text{e} \quad \tilde{\mathbf{P}}\left(r^{(S)}_{n-1,n} = d-1\right) = \tilde{q}.$$

per ogni $n = 1, \ldots, N$. Continuiamo a supporre che nel mercato sia possibile investire in opzioni europee d'acquisto e vendita, rispettivamente C e P, di sottostante S, maturità T, prezzo d'esercizio K, payoff $C_N \equiv C_T \equiv (S_T - K)^+$ e $P_N \equiv P_T \equiv (K - S_T)^+$ e valori di mercato $(C_n)_{n=0}^{N-1}$ e $(P_n)_{n=0}^{N-1}$, dati dai rispettivi prezzi di non arbitraggio al tempo $t = t_n$, per ogni $n = 0, 1, \ldots, N-1$, o equivalentemente dalle attese condizionate scontate dei rispettivi payoff allo stesso tempo. Formalmente,

$$C_0 = X_1^C B_0 + Y_1^C S_0 = \frac{1}{(1+r)^N} \tilde{\mathbf{E}}\left[C_N\right], \quad C_n = X_n^C B_n + Y_n^C S_n = \frac{1}{(1+r)^{N-n}} \tilde{\mathbf{E}}_n\left[C_N\right],$$

e

$$P_0 = X_1^P B_0 + Y_1^P S_0 = \frac{1}{(1+r)^N} \tilde{\mathbf{E}}\left[P_N\right], \quad P_n = X_n^P B_n + Y_n^P S_n = \frac{1}{(1+r)^{N-n}} \tilde{\mathbf{E}}_n\left[P_N\right],$$

per ogni $n = 0, 1, \ldots, N-1$, essendo $\Pi^C \equiv \left(\left(X_n^C, Y_n^C\right)_{n=0}^N\right)$ e $\Pi^P \equiv \left(\left(X_n^P, Y_n^P\right)_{n=0}^N\right)$ rispettivamente i portafogli replicanti dell'opzione call e put. Supponiamo infine che sia possibile investire anche in opzioni americane d'acquisto e vendita, rispettivamente AC e AP, sempre di sottostante S, con la stessa maturità T e lo stesso prezzo d'esercizio K delle opzioni europee, di rispettivo valore di mercato $(AC_n)_{n=0}^N$ e $(AP_n)_{n=0}^N$. Ricordiamo che, a differenza di un'opzione call [risp. put] europea, un'opzione call [risp. put] americana può essere esercitata una sola volta a un tempo t_n, per un qualsiasi $n = 0, 1, \ldots, N$, sempre allo stesso prezzo d'esercizio K. Quindi, a priori, per un'opzione call [risp. put] americana non è possibile applicare il risultato di replicabilità relativo alla corrispondente opzioni call [risp. put] europea. Infatti, dipendendo oltre che dal valore S_n dello stock anche dal tempo t_n d'esercizio non è a priori possibile interpretare il payoff un'opzione call [risp. put] americana come una semplice variabile aleatoria reale su $\tilde{\Omega}$. Ovviamente ciò rende più complesso il problema della determinazione del suo valore di mercato AC_n [risp. AP_n] al tempo $t = t_n$, per $n = 0, 1, \ldots, N$. Si pone inoltre il problema della determinazione del tempo ottimale d'esercizio t_{n^*}, per un opportuno $n^* \in \{0, 1, \ldots, N\}$.

Definizione 225 *Il payoff di un'opzione americana d'acquisto [risp. vendita] con maturità T e prezzo d'esercizio K esercitata al tempo $t_n \leqslant T$ è dato da*

$$(S_n - K)^+ \equiv \max\{S_n - K, 0\} \quad \text{[risp. } (K - S_n)^+ \equiv \max\{K - S_n, 0\}\text{]}, \quad \forall n = 0, 1, \ldots, N.$$

Inoltre, se esercitata a un tempo $t = t_{n_0} \leqslant T$, per un certo $n_0 \in \{0, 1, \ldots, N\}$, il payoff diventa nullo per ogni $t = t_n$ con $n = n_0 + 1, \ldots, N$.

Osservazione 226 *Il valore di mercato di un'opzione americana d'acquisto [risp. vendita] con maturità T e prezzo d'esercizio K al tempo $t_n \leqslant T$ (non esercitata a un tempo $t < t_n$) soddisfa la disuguaglianza*

$$AC_n \geqslant C_n \quad e \quad AP_n \geqslant P_n \tag{3.151}$$

per ogni $n = 0, 1, \ldots, N$. In particolare,

$$AC_N = C_N = (S_N - K)^+ \quad e \quad AP_N = P_N = (K - S_N)^+ . \tag{3.152}$$

Ciò è immediata conseguenza della circostanza che un'opzione americana comporta, a priori, il diritto aggiuntivo della scelta del tempo d'esercizio rispetto alla corrispondente opzione europea con stessa maturità T e stesso strike price K. Con un ragionamento di non arbitraggio, essendo la validità dell'Equazione (3.152) evidente, supponiamo che per un qualche tempo $t = t_{n_0}$, con $n_0 \in \{0, 1, \ldots, N-1\}$, e all'occorrenza di un qualche $\omega_0 \in \Omega$, si verifichi

$$C_{n_0}(\omega_0) > AC_{n_0}(\omega_0)$$

allora potremmo vendere allo scoperto la call europea al prezzo $C_{n_0}(\omega_0)$ comprando con il ricavato una call americana al prezzo $AC_{n_0}(\omega_0)$ e investendo la differenza $C_{n_0}(\omega_0) - AC_{n_0}(\omega_0) > 0$ nell'acquisto del bond B. Quindi, lasciando il portafoglio così costituito inalterato sino al tempo $t = t_N \equiv T$, stante la (3.152), il payoff della call americana alla maturità consentirebbe di coprire lo scoperto sulla call europea e l'investimento sul bond garantirebbe un payoff positivo. Avremmo infatti

$$C_N(\omega_0) = AC_N(\omega_0) \qquad \text{e} \qquad (C_{n_0}(\omega_0) - AC_{n_0}(\omega_0))(1+r)^{N-n_0} > 0.$$

Un ragionamento del tutto analogo può applicarsi nel caso di una put americana.

Proposizione 227 *Nel modello di mercato CRR multiperiodale in assenza di BS-portafogli d'arbitraggio, il valore di un'opzione call americana con maturità T e prezzo d'esercizio K al tempo $t_n \leqslant T$ (non esercitata a un tempo $t < t_n$) soddisfa la disuguaglianza*

$$AC_n \geqslant \left(S_n - \frac{K}{(1+r)^{N-n}}\right)^+, \tag{3.153}$$

per ogni $n = 0, 1, \ldots, N$. Il valore di un'opzione americana di vendita con maturità T e prezzo d'esercizio K al tempo $t_n \leqslant T$ (non esercitata a un tempo $t < t_n$) soddisfa la disuguaglianza

$$AP_n \geqslant (K - S_n)^+, \tag{3.154}$$

per ogni $n = 0, 1, \ldots, N$.

Proof. *L'Equazione (3.153) è immediata conseguenza delle Equazioni (3.142) e (3.151). Per quanto riguarda l'Equazione (3.154), osserviamo che in conseguenza delle Equazioni (3.146) e (3.151) si ha chiaramente*

$$AP_n \geqslant 0, \tag{3.155}$$

per ogni $n = 0, 1, \ldots, N$. Inoltre se risultasse

$$AP_{n_0}(\omega_0) < K - S_{n_0}(\omega_0)$$

per un qualche $n_0 \in \{0, 1, \ldots, N\}$ e all'occorrenza di un qualche $\omega_0 \in \Omega$, prendendo a prestito la cifra K sarebbe possibile comprare lo stock S al prezzo $S_{n_0}(\omega_0)$, la put AP al prezzo $AP_{n_0}(\omega_0)$ e investire nel bond la differenza $K - (S_{n_0}(\omega_0) - AP_{n_0}(\omega_0))$. Quindi esercitando immediatamente la put potremmo vendere lo stock in nostro possesso al prezzo K e ripianare immediatamente il debito contratto prima della decorrenza degli interessi. Rimarrebbe un saldo complesssivo delle operazioni pari a

$$K - (S_{n_0}(\omega_0) - AP_{n_0}(\omega_0)) > 0$$

investito nel bond, che al tempo $t = t_N$ produrrebbe un capitale

$$(K - (S_{n_0}(\omega_0) - AP_{n_0}(\omega_0)))(1 + r)^{N - n_0}.$$

Avremmo pertanto costruito un arbitraggio. In assenza di BS-portafogli d'arbitraggio deve allora aversi

$$AP_n \geqslant K - S_n, \tag{3.156}$$

per ogni $n = 0, 1, \ldots, N$. Combinando le Equazioni (3.155) e (3.156) segue immediatamente la desiderata (3.154). □

Proposizione 228 *L'esercizio anticipato di una call americana non è mai ottimale.*

Proof. *Consideriamo un call americana e supponiamo che il tempo ottimale d'esercizio sia $t = t_{n_0}$, per un qualche $n_0 \in \{0, 1, \ldots, N-1\}$ e all'occorrenza di un qualche $\omega_0 \in \Omega$. Stante l'Equazione (3.142), abbiamo*

$$C_{n_0}(\omega_0) \geqslant \left(S_{n_0}(\omega_0) - \frac{K}{(1 + r)^{N - n_0}}\right)^+.$$

D'altra parte, se la call americana fosse esercitata avremmo

$$AC_{n_0}(\omega_0) = (S_{n_0}(\omega_0) - K)^+.$$

Quindi risulterebbe

$$C_{n_0}(\omega_0) > AC_{n_0}(\omega_0)$$

il che contraddirebbe la (3.151). □

Proposizione 229 *In assenza di BS portafogli d'arbitraggio, il prezzo di una call americana AC ad ogni tempo uguaglia il prezzo di una call europea C con la stessa maturità T e lo stesso prezzo d'esercizio K. In simboli,*

$$AC_n = C_n, \tag{3.157}$$

per ogni $n = 0, 1, \ldots, N$.

Proof. *Poichè l'esercizio anticipato di una call americana non è mai ottimale essa va esercitata alla maturità* $T \equiv t_N$*, tempo in cui diventa di fatto una call europea e si ha pertanto*

$$AC_N = C_N.$$

Ma allora la non esistenza di BS-portafogli d'arbitraggio e la permanenza in vita della call americana per ogni $n = 0, 1, \ldots, N$ *comporta la possibilità di applicare il Corollario 212 e ottenere la (3.157).*

Presentiamo anche una dimostrazione alternativa è. Dall'Osservazione 226 sappiamo che

$$AC_n \geqslant C_n$$

per ogni $n = 0, 1, \ldots, N-1$ *e che*

$$AC_N = C_N.$$

Supponiamo quindi che si verifichi

$$AC_{n_0}(\omega_0) > C_{n_0}(\omega_0)$$

per un qualche $n_0 \in \{0, 1, \ldots, N-1\}$ *e all'occorrenza di un qualche* $\omega_0 \in \Omega$*, allora potremmo vendere la call americana allo scoperto al prezzo* $AC_{n_0}(\omega_0)$ *comprando col ricavato una call europea di prezzo* $C_{n_0}(\omega_0)$ *e investendo la differenza* $AC_{n_0}(\omega_0) - C_{n_0}(\omega_0)$ *nel bond. Se l'acquirente della call americana non la dovesse esercitare prima della maturità, al tempo* $t = t_N$ *l'esercizio della call europea ci consentirebbe di fronteggiare un eventuale esercizio della call americana a saldo zero e per di più dall'investimento sul bond avremmo ricavato payoff positivo. Infatti*

$$C_N(\omega_0) = AC_N(\omega_0) \quad e \quad (AC_{n_0}(\omega_0) - C_{n_0}(\omega_0))(1+r)^{N-n_0} > 0,$$

Se invece l'acquirente della call la esercitasse a un tempo $t = t_n$ *con* $n < N$*, ci troveremmo a dovergli corrispondere la cifra* $S_n(\omega_0) - K$*. D'altra parte stante l'Equazione (3.142), la call europea in nostro possesso avrebbe un valore di mercato* $C_n(\omega_0) \geqslant S_n(\omega_0) - \frac{K}{(1+r)^{N-n}}$*. Quindi potremmo venderla sul mercato con un saldo tra le due call pari a*

$$-(S_n(\omega_0) - K) + C_n(\omega_0) \geqslant -(S_n(\omega_0) - K) + S_n(\omega_0) - \frac{K}{(1+r)^{N-n}} = K - \frac{K}{(1+r)^{N-n}} \geqslant 0.$$

Inoltre, per la posizione aperta su bond, avremmo a disposizione il payoff positivo

$$(AC_{n_0}(\omega_0) - C_{n_0}(\omega_0))(1+r)^{n-n_0}.$$

In definitiva, avremmo realizzato un arbitraggio. □

Un'ulteriore semplice considerazione finanziaria: essendo l'esercizio anticipato dell'opzione call americana sempre inappropriato, il diritto all'esercizio anticipato deve valere zero. Quindi la call americana deve avere lo stesso valore di una call europea in ogni tempo antecedente la maturità. Notare che l'eventuale esercizio anticipato al tempo $t = t_{n^*}$ dell'opzione call americana, per un qualche $n_0 \in \{0, 1, \ldots, N-1\}$ e all'occorrenza di un qualche $\omega_0 \in \Omega$ produrrebbe un payoff

$$S_{n_0}(\omega_0) - K < S_n(\omega_0) - \frac{K}{(1+r)^{N-n^*}} \leqslant C_{n_0}(\omega_0) \leqslant AC_{n_0}(\omega_0)$$

inferiore al suo valore di mercato. Quindi piuttosto che l'esercizio anticipato converrebbe la sua vendita sul mercato.

In merito alle put americane il discorso è alquanto più complesso e l'esercizio anticipato può risultare conveniente. Per rendersi conto di ciò, ragionando in termini finanziari, ma esternamente al modello CRR, immaginiamo di avere una put americana con maturità T su uno stock il cui prezzo subisca una caduta fino al valore $S_{t_0}(\omega_0) = 0$, a un certo tempo $t_0 < T$ e in una qualche circostanza ω_0. In questo caso, l'esercizio dell'opzione put al tempo $t = t_0$ restituirebbe esattamente il massimo payoff possibile $(K - S_{t_0}(\omega_0))^+ = K^+ = K$. Pertanto, se l'opzione non venisse esercitata immediatamente si perderebbe quanto meno una parte del capitale $K(1+r)^{N-n_0}$ generato alla maturità dall'incasso del prezzo d'esercizio K al tempo $t = t_0$. Ciò indipendentemente dall'evoluzione futura del prezzo dello stock.

Per la determinazione del tempo d'esercizio e del prezzo di una put americana occorre introdurre la nozione di tempo d'arresto. Tuttavia con alcune semplici considerazioni basate sull'assenza di portafogli d'arbitraggio è possibile provare che

Proposizione 230 *In assenza di BS portafogli d'arbitraggio, il prezzo di una put americana AP al tempo $t = t_0 \equiv 0$ soddisfa*

$$K \geqslant AP_0 \geqslant (K - S_0)^+ . \tag{3.158}$$

Proof. *Se fosse*

$$(K - S_0)^+ > AP_0,$$

sarebbe possibile acquistare la put americana al prezzo AP_0 ed esercitarla istantaneamente realizzando un payoff

$$(K - S_0)^+ - AP_0 > 0.$$

Se altresì fosse

$$AP_0 > K,$$

sarebbe possibile vendere allo scoperto la put al prezzo AP_0 e investire il ricavato sul bond ottenendo ad ogni tempo $t = t_n$ un payoff $AP_0(1+r)^n$ sicuramente superore l valore massimo K del possibile payoff derivante dal possibile esercizio della put ad ogni tempo $t = t_n$. Infatti, l'eventuale esercizio della put a un qualsiasi tempo $t = t_n$, con $n \in \{1, \dots, N\}$, e per un qualsiasi esito $\omega \in \Omega$, comporterebbe un esborso di entità K a fronte dell'acquisizione del titolo di prezzo $S_n(\omega)$ e darebbe quindi luogo al payoff complessivo

$$AP_0(1+r)^n - K + S_n(\omega) > 0,$$

mentre se la put non venisse esercitata otterremmo un payoff finale di

$$AP_0(1+r)^N > 0.$$

□

Proposizione 231 *Nel caso di opzioniam ericane la relazione di call-put parity diventa*

$$AC_0 - S_0 + K \geqslant AP_0 \geqslant AC_0 + \frac{K}{(1+r)^N} - S_0. \tag{3.159}$$

Proof. *Dall'Equazione (3.148), abbiamo chiaramente*

$$P_0 = C_0 + \frac{K}{(1+r)^N} - S_0.$$

Allora, considerando le Equazioni (3.151) e (3.157), otteniamo

$$AP_0 \geqslant AC_0 + \frac{K}{(1+r)^N} - S_0.$$

Per provare la prima parte dell'Equazione (3.159), supponiamo che si abbia

$$AP_0 + S_0 > AC_0 + K \tag{3.160}$$

allora vendendo allo scoperto sia la put americana di prezzo d'esercizio K e maturità T che il titolo rischioso disporremmo di una somma tale da poter comprare la call americana di prezzo d'esercizio K e maturità T e investire nel bond la somma K. In caso d'esercizio anticipato della put a un tempo $t = t_{n_0}$, con $n_0 \in \{0, 1, \ldots, N-1\}$, e per un certo $\omega_0 \in \Omega$, dovremmo pagare all'acquirente della put il prezzo K ottenendone in cambio il titolo rischioso di valore $S_{n_0}(\omega_0)$. D'altra parte, ricaveremmo la somma K dall'investimento nel bond che al tempo $t = t_{n_0}$ frutterebbe un pay-off pari a $K(1+r)^{n_0}$ e useremmo il titolo rischioso ottenuto in cambio di K per coprire il nostro scoperto sul titolo rischioso. Ci ritroveremmo pertanto con un pay-off complessivo pari a

$$AC_{n_0}(\omega_0) + K\left((1+r)^{n_0} - 1\right) \geqslant 0.$$

Da notare che se $n_0 > 0$ il pay-off sarebbe strettamente positivo. Se invece la put non venisse esercitata anticipatamente, considerata l'Equazione (3.152), alla maturità $t = t_N \equiv T$ ci ritroveremmo con un pay-off complessivo pari a

$$\begin{aligned}
&AC_N - AP_N - S_N + K(1+r)^N \\
&= (S_N - K)^+ - (K - S_N)^+ - S_N + K(1+r)^N \\
&= \begin{cases} S_N - K - S_N + K(1+r)^N = K\left((1+r)^N - 1\right) > 0, & \textit{se } S_N \geqslant K, \\ -(K - S_N) - S_N + K(1+r)^N = K\left((1+r)^N - 1\right) > 0, & \textit{se } S_N < K \end{cases}
\end{aligned}$$

In definitiva, valendo la (3.160), potremmo costituire un d'arbitraggio. L'assenza di BS-portafogli d'arbitraggio impone allora che si abbia

$$AC_0 + K \geqslant AP_0 + S_0.$$

La (3.159) è quindi completamente provata. □

Per caratterizzare esattamente il valore di un'opzione put americana nell'ambito del modello CRR multiperiodale ad ogni tempo $t = t_n$ è necessario introdurre la nozione di *tempo d'arresto*.

Definizione 232 *Chiamiamo* tempo d'arresto *rispetto alla filtrazione $\mathfrak{F}$, più brevemente $\mathfrak{F}$-*tempo d'arresto*, una variabile aleatoria reale $\nu : \Omega \to \mathbb{R}$ tale che*

$$\nu(\Omega) \subseteq \{0, 1, \ldots, N\} \quad e \quad \{\omega \in \Omega : \nu(\omega) \leqslant n\} \equiv \{\nu \leqslant n\} \in \mathcal{F}_n, \quad \forall n \in \{0, 1, \ldots, N\}. \tag{3.161}$$

Osservazione 233 *Comunque data una variabile aleatoria reale $\nu : \Omega \to \mathbb{R}$ risulta*

$$\{\nu \leqslant n\} \in \mathcal{F}_n, \quad \forall n \in \{0, 1, \ldots, N\} \Leftrightarrow \{\nu = n\} \in \mathcal{F}_n, \quad \forall n \in \{0, 1, \ldots, N\}. \tag{3.162}$$

Il senso di questa definizione è che l'evento in cui il tempo d'arresto prende un valore pari ad n dev'essere osservabile alla luce dell'informazione in $\mathcal{F}_n$, che si rende disponibile al tempo $t = t_n$. In altri termini, non possiamo scoprire che la variabile aleatoria tempo d'arresto avrebbe dovuto prendere un valore pari ad n solo dopo che il tempo $t = t_n$ sia trascorso. Quindi, un tempo d'arresto vuole rappresentare un modo di poter associare ad ogni esisto del fenomeno aleatorio un valore n tra $0, 1, \ldots, N$ osservabile alla luce dell'informazione $\mathcal{F}_n$ disponibile sino al tempo $t = t_n$ senza dover attendere che si renda informazione disponibile in tempi successivi.

Esempio 234 *Consideriamo le variabili aleatorie $\nu_1 : \Omega \to \mathbb{R}$ e $\nu_2 : \Omega \to \mathbb{R}$ definite ponendo*

$$\nu_1(\omega) \stackrel{def}{=} \omega_1 + 1, \quad \forall \omega \in \Omega, \ \omega \equiv (\omega_n)_{n=1}^N,$$

e

$$\nu_2(\omega) \stackrel{def}{=} \omega_2 + 1, \quad \forall \omega \in \Omega, \ \omega \equiv (\omega_n)_{n=1}^N.$$

Si puo facilmente verificare che ν_1 è un tempo d'arresto, ma ν_2 non lo è. Infatti, per fissare le idee, consideriamo $N = 3$, allora

$$\Omega \equiv \{(0,0,0), (0,0,1), (0,1,0), (0,1,1), (1,0,0), (1,0,1), (1,1,0), (1,1,1)\}.$$

Abbiamo quindi

$$\nu_1((0,0,0)) = \nu_1((0,0,1)) = \nu_1((0,1,0)) = \nu_1((0,1,1)) = 1,$$

e

$$\nu_1((1,0,0)) = \nu_1((1,0,1)) = \nu_1((1,1,0)) = \nu_1((1,1,1)) = 2.$$

Pertanto,

$$\{\nu_1 \leqslant 0\} = \varnothing, \quad \{\nu_1 \leqslant 1\} = E_0, \quad \{\nu_1 \leqslant 2\} = \Omega \quad \{\nu_1 \leqslant 3\} = \Omega,$$

essendo $E_0 \equiv \{(0,0,0), (0,0,1), (0,1,0), (0,1,1)\}$. Tenuto conto che $E_0 \in \mathcal{F}_1$, risulta allora

$$\{\nu_1 \leqslant n\} \in \mathcal{F}_n, \quad \forall n \in \{0,1,2,3\}.$$

Abbiamo poi

$$\nu_2((0,0,0)) = \nu_2((0,0,1)) = \nu_2((1,0,0)) = \nu_2((1,0,1)) = 1$$

e

$$\nu_2((0,1,0)) = \nu_2((0,1,1)) = \nu_2((1,1,0)) = \nu_2((1,1,1)) = 2.$$

Pertanto,

$$\{\nu_2 \leqslant 0\} = \varnothing, \quad \{\nu_2 \leqslant 1\} = E_{0,0} \cup E_{1,0}, \quad \{\nu_2 \leqslant 2\} = \Omega \quad \{\nu_3 \leqslant 3\} = \Omega,$$

essendo $E_{0,0} \equiv \{(0,0,0), (0,0,1)\}$ e $E_{1,0} \equiv \{(1,0,0), (1,0,1)\}$. Tenuto conto che $E_{0,0} \cup E_{1,0} \in \mathcal{F}_2 - \mathcal{F}_1$, risulta

$$\{\nu_2 \leqslant 1\} \notin \mathcal{F}_1.$$

Come si voleva mostrare.

Abbiamo

Teorema 235 *Il processo dei prezzi $(AP_n)_{n=0}^N$ di un'opzione put americana soddisfa l'equazione*

$$AP_n = \max_{\nu \in \mathcal{N}(n,N)} \tilde{\mathbf{E}}\left[\frac{1}{(1+r)^{\nu-n}}(K - S_\nu)^+ \mid \mathcal{F}_n\right], \tag{3.163}$$

per ogni $n = 0,1,\ldots,N$, essendo $\mathcal{N}(n,N) \equiv \mathcal{N}(t_n,T)$ l'insieme degli $\mathfrak{F}$-tempi d'arresto che soddisfano la condizione $n \leqslant \nu \leqslant N$, o equivalentemente $t_n \leqslant \nu \leqslant T$, per ogni $n = 0,1,\ldots,N$. Inoltre, il tempo d'arresto $\nu_n^* \in \mathcal{N}(n,N)$ per il quale vale l'Equazione (3.163), noto come tempo d'arresto ottimale *successivo al tempo $t = t_n$, è dato da*

$$\nu_n^*(\omega) = \min\left\{u \geqslant n : AP_u(\omega) = (K - S_u(\omega))^+\right\}.$$

Teorema 236 *Il processo dei prezzi $(AP_n)_{n=0}^N$ di un'opzione put americana può anche ottenersi con l'induzione retroattiva definita da*

$$AP_N = (K - S_N)^+ = P_N \tag{3.164}$$

e

$$AP_n = \max\left\{(K - S_n)^+, \frac{1}{1+r}\tilde{\mathbf{E}}_n[AP_{n+1}]\right\} = \max\left\{(K - S_n)^+, \frac{1}{1+r}\left(AP_{n+1}^+\tilde{p} + AP_{n+1}^-\tilde{q}\right)\right\} \tag{3.165}$$

per ogni $n = N - 1, \dots, 1, 0$. *A sua volta, il tempo d'arresto ottimale nell'insieme* $\mathcal{N}(n, N)$ *può anche ottenersi con l'induzione retroattiva definita da*

$$\nu_N^* (\omega) = N \tag{3.166}$$

per ogni $\omega \in \Omega$ *e*

$$\nu_n^* (\omega) = n 1_{\{AP_n = (K - S_n)^+\}} (\omega) + \nu_{n+1}^* (\omega) 1_{\{AP_n > (K - S_n)^+\}} (\omega) \tag{3.167}$$

per ogni $n = N - 1, \dots, 1, 0$ *ed ogni* $\omega \in \Omega$.

Da notare che stante l'Equazione (3.154), gli eventi $\{AP_n = (K - S_n)^+\}$ e $\{AP_n > (K - S_n)^+\}$ costituiscono una partizione di Ω.

Definizione 237 *Considerata un'opzione put americana AP chiamiamo* valore d'esercizio immediato (current payoff) *[risp.* valore d'esercizio atteso (expected payoff)*] di AP al tempo* $t = t_n$*, per ogni* $n = 0, 1, \dots, N - 1$*, la variabile aleatoria*

$$(K - S_n)^+ \quad \textit{[risp. } \frac{1}{1+r} \tilde{\mathbf{E}}_n [AP_{n+1}] \textit{]}$$

Osservazione 238 *Alla luce della Definizione 237, il Teorema 3.165 può essere riformulato enunciando che il valore di mercato* $AP_n (\omega)$*, al tempo* $t = t_n$ *e all'occorrenza dell'esito* $\omega \in \tilde{\Omega}$*, di un'opzione put americana non ancora esercitata è dato dal massimo tra il valore d'esercizio immediato e il valore d'esercizio atteso. Inoltre, l'opzione put americana va esercitata non appena tale massimo è il valore d'esercizio immediato.*

Osservazione 239 *Si ha*

$$\nu_0^* (\omega) = 0 \Leftrightarrow (K - S_0)^+ \geqslant \frac{1}{1+r} \tilde{\mathbf{E}}_0 [AP_1] \quad e \quad \nu_0^* (\omega) = \nu_1^* (\omega) \Leftrightarrow \frac{1}{1+r} \tilde{\mathbf{E}}_0 [AP_1] > (K - S_0)^+ ,$$

per ogni $\omega \in \Omega$
Proof. *È sufficiente osservare che si ha*

$$\{AP_0 = (K - S_0)^+\} = \begin{cases} \Omega \Leftrightarrow (K - S_0)^+ \geqslant \frac{1}{1+r} \tilde{\mathbf{E}}_0 [AP_1] , \\ \varnothing \Leftrightarrow (K - S_0)^+ < \frac{1}{1+r} \tilde{\mathbf{E}}_0 [AP_1] . \end{cases}$$

□

Esempio 240 *In relazione all'esempio presentato nello script, abbiamo*

$$AP_5 (\omega) = \max \{K - S_5 (\omega), 0\} = \begin{cases} 22.622, & \textit{se } S_5 (\omega) = 77.378, \\ 6.332, & \textit{se } S_5 (\omega) = 93.688, \\ 0, & \textit{altrimenti}, \end{cases}$$

e

$$\nu_5^* (\omega) = 5.$$

Abbiamo poi

$$\begin{aligned} AP_4 (\omega) &= \max \left\{ K - S_4 (\omega), \frac{1}{1+r} \tilde{\mathbf{E}} [AP_5 \mid S_4 = S_4 (\omega)] \right\} \\ &= \max \left\{ K - S_4 (\omega), \frac{1}{1+r} \left(AP_5^+ \tilde{p} + AP_5^- \tilde{q} \right) \right\} \\ &= \begin{cases} 18.549, & \textit{se } S_4 (\omega) = 81.451, \\ 3.015, & \textit{se } S_4 (\omega) = 98.598, \\ 0, & \textit{altrimenti}, \end{cases} \end{aligned}$$

e

$$\nu_4^*(\omega) = \begin{cases} 5, & \textit{if } \sum_{k=1}^4 \omega_k = 1, \\ 4, & \textit{altrimenti}, \end{cases}$$

dove

$$\left\{\omega \in \Omega : \sum_{k=1}^{4} \omega_k = 1\right\} = \left\{(0,0,0,1,\omega_5), (0,0,1,0,\omega_5), (0,1,0,0,\omega_5), (1,0,0,0,\omega_5)\right\},$$

con ω_5 *che prende entrambi i possibili valori* 0 *e* 1. *Similmente,*

$$\begin{aligned} AP_3(\omega) &= \max\left\{K - S_3(\omega), \frac{1}{1+r}\tilde{\mathbf{E}}\left[AP_4 \mid S_3 = S_3(\omega)\right]\right\} \\ &= \max\left\{K - S_3(\omega), \frac{1}{1+r}\left(AP_4^+\tilde{p} + AP_4^-\tilde{q}\right)\right\} \\ &= \begin{cases} 14.263, & \textit{se } S_3(\omega) = 85.737, \\ 1.436, & \textit{se } S_3(\omega) = 103.787, \\ 0, & \textit{altrimenti}, \end{cases} \end{aligned}$$

e

$$\nu_3^*(\omega) = \begin{cases} \nu_4^*(\omega), & \textit{se } \sum_{k=1}^3 \omega_k = 1, \\ 3, & \textit{altrimenti}, \end{cases} = \begin{cases} 5, & \textit{se } \sum_{k=1}^3 \omega_k = 1 \wedge \omega_4 = 0, \\ 4, & \textit{se } \sum_{k=1}^3 \omega_k = 1 \wedge \omega_4 = 1, \\ 3, & \textit{altrimenti}, \end{cases}$$

dove

$$\left\{\omega \in \Omega : \sum_{k=1}^{3} \omega_k = 1\right\} = \left\{(0,0,1,\omega_4,\omega_5), (0,1,0,\omega_4,\omega_5), (1,0,0,\omega_4,\omega_5)\right\}$$

con ω_4 *e* ω_5 *che prendono ciascuno entrambi i possibili valori* 0 *e* 1. *Inoltre,*

$$\left\{\omega \in \Omega : \sum_{k=1}^{3} \omega_k = 1\right\}^c = \left\{\omega \in \Omega : \sum_{k=1}^{3} \omega_k = 0\right\} \cup \left\{\omega \in \Omega : \sum_{k=1}^{3} \omega_k = 2\right\} \cup \left\{\omega \in \Omega : \sum_{k=1}^{3} \omega_k = 3\right\},$$

con

$$\left\{\omega \in \Omega : \sum_{k=1}^{3} \omega_k = 0\right\} = \left\{(0,0,0,\omega_4,\omega_5)\right\},$$

$$\left\{\omega \in \Omega : \sum_{k=1}^{3} \omega_k = 2\right\} = \left\{(0,1,1,\omega_4,\omega_5), (1,0,1,\omega_4,\omega_5), (1,1,0,\omega_4,\omega_5)\right\},$$

e

$$\left\{\omega \in \Omega : \sum_{k=1}^{3} \omega_k = 3\right\} = \left\{(1,1,1,\omega_4,\omega_5)\right\},$$

sempre con ω_4 *e* ω_5 *che prendono ciascuno entrambi i possibili valori* 0 *e* 1. *Abbiamo poi*

$$\begin{aligned} AP_2(\omega) &= \max\left\{K - S_2(\omega), \frac{1}{1+r}\tilde{\mathbf{E}}\left[AP_3 \mid S_2 = S_2(\omega)\right]\right\} \\ &= \max\left\{K - S_2(\omega), \frac{1}{1+r}\left(AP_3^+\tilde{p} + AP_3^-\tilde{q}\right)\right\} \\ &= \begin{cases} 9.75, & \textit{se } S_2(\omega) = 90.25, \\ 0.684, & \textit{se } S_2(\omega) = 109.25, \\ 0, & \textit{altrimenti}, \end{cases} \end{aligned}$$

e

$$\nu_2^*(\omega) = \begin{cases} \nu_3^*(\omega), & \textit{se } \sum_{k=1}^{2}\omega_k = 1, \\ 2, & \textit{altrimenti}, \end{cases} = \begin{cases} 5, & \textit{se } \sum_{k=1}^{2}\omega_k = 1 \wedge \omega_3 = 0 \wedge \omega_4 = 0, \\ 4, & \textit{se } \sum_{k=1}^{3}\omega_k = 1 \wedge \omega_3 = 0 \wedge \omega_4 = 1, \\ 3, & \textit{se } \sum_{k=1}^{3}\omega_k = 1 \wedge \omega_3 = 1, \\ 2, & \textit{altrimenti}, \end{cases}$$

dove

$$\left\{\omega \in \Omega : \sum_{k=1}^{2}\omega_k = 1\right\} = \{(0,1,\omega_3,\omega_4,\omega_5), (1,0,\omega_3,\omega_4,\omega_5)\}$$

con ω_3*,* ω_4 *e* ω_5 *che prendono ciascuno entrambi i possibili valori* 0 *e* 1*. Inoltre,*

$$\left\{\omega \in \Omega : \sum_{k=1}^{2}\omega_k = 1\right\}^c = \left\{\omega \in \Omega : \sum_{k=1}^{2}\omega_k = 0\right\} \cup \left\{\omega \in \Omega : \sum_{k=1}^{2}\omega_k = 2\right\},$$

con

$$\left\{\omega \in \Omega : \sum_{k=1}^{2}\omega_k = 0\right\} = \{(0,0,\omega_3,\omega_4,\omega_5)\}, \quad e \quad \left\{\omega \in \Omega : \sum_{k=1}^{2}\omega_k = 2\right\} = \{(1,1,\omega_3,\omega_4,\omega_5)\},$$

sempre con ω_3*,* ω_4 *e* ω_5 *che prendono ciascuno entrambi i possibili valori* 0 *e* 1*. Infine,*

$$\begin{aligned} AP_1(\omega) &= \max\left\{K - S_1(\omega), \frac{1}{1+r}\tilde{\mathbf{E}}\left[AP_2 \mid S_1 = S_1(\omega)\right]\right\} \\ &= \max\left\{K - S_1(\omega), \frac{1}{1+r}\left(AP_2^+\tilde{p} + AP_2^-\tilde{q}\right)\right\} \\ &= \begin{cases} 5, & \textit{se } S_1(\omega) = 95, \\ 0.326, & \textit{se } S_1(\omega) = 115, \end{cases} \end{aligned}$$

e

$$\nu_2^*(\omega) = \begin{cases} \nu_2^*(\omega), & \textit{se } \sum_{k=1}^{1}\omega_k = 1, \\ 1, & \textit{altrimenti}, \end{cases} = \begin{cases} 5, & \textit{se } \omega_1 = 1 \wedge \omega_2 = 0 \wedge \omega_3 = 0 \wedge \omega_4 = 0, \\ 4, & \textit{se } \omega_1 = 1 \wedge \omega_2 = 0 \wedge \omega_3 = 0 \wedge \omega_4 = 1, \\ 3, & \textit{se } \omega_1 = 1 \wedge \omega_2 = 0 \wedge \omega_3 = 1, \\ 2, & \textit{se } \omega_1 = 1 \wedge \omega_2 = 1, \\ 1, & \textit{altrimenti}, \end{cases}$$

essendo chiaramente

$$\left\{\omega \in \Omega : \sum_{k=1}^{1}\omega_k = 1\right\} = \{\omega \in \Omega : \omega_1 = 1\} \quad e \quad \left\{\omega \in \Omega : \sum_{k=1}^{1}\omega_k = 1\right\}^c = \{\omega \in \Omega : \omega_1 = 0\}.$$

3.5.4 Calibrazione

Nonostante la sua semplicità, il modello di Cox Ross Rubinstein si presta piuttosto bene a simulare mercati finanziari reali, consentendo la valutazione teorica del prezzo di opzioni europee, nonchè di altri derivati più sofisticati, abbastanza prossima ai prezzi reali di mercato. Questa prossimità è peraltro accentuata dalla circostanza che vi è un notevole numero di agenti che fanno uso di implementazioni computazionali di questo modello, proprio per stimare gli ipotetici i prezzi corretti *(fair prices)* di mercato. Ciò porta ad una sorta di autorealizzazione delle aspettative che finisce paradossalmente con lo spingere il prezzo reale verso il prezzo teorico. Ovviamente, per potere concretamente adoperare il modello bisogna preliminarmente *calibrarlo*, ossia specificare i valori opportuni per u, d e per $\tilde{p}$ che consentano di ottenere risultati del modello in sufficiente accordo con i dati reali. A tale scopo, cominciamo con l'osservare che avendo il modello natura moltiplicativa, il prezzo dello stock non è mai negativo e pertanto è possibile rifarsi al logaritmo del prezzo dello stock come variabile fondamentale.

In effetti, l'uso del logaritmo conduce ad ottenere relazioni semplici per la selezione dei parametri. Ricordiamo di avere ipotizzato che tutti i periodi di contrattazione abbiano la stessa durata, ossia $t_n - t_{n-1} = T/N \equiv \Delta t$, per ogni $n = 1, \dots N$, e quindi $t_n = n\Delta t$, per ogni $n = 0, \dots N$. Considerato il processo dei prezzi $(S_n)_{n=0}^N$ dello stock S rilevati ai tempi $t = t_n$ nell'intervallo $[0, T]$, per $n = 0, \dots N$, assumiamo si abbia

$$\ln(S_n) - \ln(S_{n-1}) = \ln\left(\frac{S_n}{S_{n-1}}\right) \sim \mathcal{N}\left(\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t, \sigma^2 \Delta t\right), \quad \forall n = 1, \dots, N$$

per opportuni $\mu \in \mathbb{R}$, $\sigma > 0$. Ciò significa assumere che i rendimenti logaritmici dello stock S siano normalmente equidistribuiti[10]

$$\mathbf{E}\left[\ln\left(\frac{S_n}{S_{n-1}}\right)\right] = \left(\mu - \frac{1}{2}\sigma^2\right)\Delta t \quad \text{e} \quad \mathbf{D}^2\left[\ln\left(\frac{S_n}{S_{n-1}}\right)\right] = \sigma^2 \Delta t,$$

per ogni $n = 1, \dots, N$. Da notare che in conseguenza di questa ipotesi si ha che le variabili aleatorie S_n/S_{n-1} è sono lognormalmente equidistribuite. Più specificamente,

$$\frac{S_n}{S_{n-1}} = \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t + \sigma\sqrt{\Delta t} Z_n\right), \quad Z_n \sim \mathcal{N}(0, 1),$$

per ogni $n = 1, \dots, N$. Si ha allora, [11],

$$\mathbf{E}\left[\frac{S_n}{S_{n-1}}\right] = \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t\right)\exp\left(\frac{1}{2}\sigma^2\Delta t\right) = \exp(\mu\Delta t) \quad \text{e} \quad \mathbf{D}^2\left[\frac{S_n}{S_{n-1}}\right] = \exp(2\mu\Delta t)\left(\exp\left(\sigma^2\Delta t\right) - 1\right),$$

[10] Questa è un'ipotesi che ha contrassegnato le origini della finanza matematica e che alla luce dell'enorme mole di dati accumulatisi in tempi più recenti può essere assunta solo come un'approssimazione alquanto grossolana.

[11] If X is a log-normally distributed random variable, then, setting $\mathbf{E}[\ln(X)] \equiv \mu$ e $\mathbf{D}^2[\ln(X)] \equiv \sigma^2$, we can write

$$X = \exp(\mu + \sigma Z), \quad Z \sim \mathcal{N}(0, 1).$$

As a consequence,

$$\mathbf{E}[X] = \mathbf{E}\left[e^{\mu + \sigma Z}\right] = \mathbf{E}\left[e^{\mu} e^{\sigma Z}\right] = e^{\mu}\mathbf{E}\left[e^{\sigma Z}\right]$$

e

$$\begin{aligned}\mathbf{D}^2[X] &= \mathbf{E}\left[X^2\right] - \mathbf{E}[X]^2 = \mathbf{E}\left[e^{2(\mu+\sigma Z)}\right] - e^{2\mu}\mathbf{E}\left[e^{\sigma Z}\right]^2 \\ &= e^{2\mu}\mathbf{E}\left[e^{2\sigma Z}\right] - e^{2\mu}\mathbf{E}\left[e^{\sigma Z}\right]^2 = e^{2\mu}\left(\mathbf{E}\left[e^{2\sigma Z}\right] - \mathbf{E}\left[e^{\sigma Z}\right]^2\right).\end{aligned}$$

On the other hand,

$$\begin{aligned}\mathbf{E}\left[e^{\sigma Z}\right] &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{\sigma z} e^{-\frac{1}{2}z^2} dz = \frac{e^{\frac{1}{2}\sigma^2}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\sigma^2} e^{\sigma z - \frac{1}{2}z^2} dz \\ &= \frac{e^{\frac{1}{2}\sigma^2}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\sigma^2 + \sigma z - \frac{1}{2}z^2} dz = \frac{e^{\frac{1}{2}\sigma^2}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{(z-\sigma)^2}{2}} dz \\ &= \frac{e^{\frac{1}{2}\sigma^2}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = e^{\frac{1}{2}\sigma^2}.\end{aligned}$$

It clearly follows

$$\mathbf{E}\left[e^{2\sigma Z}\right] = e^{2\sigma^2} \quad \text{and} \quad \mathbf{E}\left[e^{\sigma Z}\right]^2 = e^{\sigma^2}.$$

In the end,

$$\mathbf{E}[X] = e^{\mu} e^{\frac{1}{2}\sigma^2} = e^{\mu + \frac{1}{2}\sigma^2}$$

and

$$\mathbf{D}^2[X] = e^{2\mu}\left(e^{2\sigma^2} - e^{\sigma^2}\right) = e^{2\mu} e^{\sigma^2}\left(e^{\sigma^2} - 1\right) = e^{2\mu + \sigma^2}\left(e^{\sigma^2} - 1\right).$$

per ogni $n = 1, \ldots, N$. Se assumiamo inoltre che le variabili aleatorie $\ln\left(\frac{S_N}{S_{N-1}}\right), \ldots, \ln\left(\frac{S_1}{S_0}\right)$ siano indipendenti[12], posto

$$X_n \equiv \ln\left(\frac{S_n}{S_{n-1}}\right), \quad \forall n = 1, \ldots, N,$$

$$\bar{X}_N \equiv \frac{1}{N}\sum_{n=1}^{N} X_n, \quad \text{e} \quad S_N^2(X) \equiv \frac{1}{N-1}\sum_{n=1}^{N}\left(X_n - \bar{X}_N\right)^2,$$

abbiamo[13]

$$\bar{X}_N \sim \mathcal{N}\left(\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t, \frac{\sigma^2 \Delta t}{N}\right),$$

per cui

$$\mathbf{E}\left[\bar{X}_N\right] = \left(\mu - \frac{1}{2}\sigma^2\right)\Delta t, \quad \mathbf{D}^2\left[\bar{X}_N\right] = \frac{\sigma^2 \Delta t}{N}.$$

Inoltre, abbiamo[14]

$$\frac{(N-1)\, S_N^2(X)}{\sigma^2 \Delta t} \sim \chi^2_{N-1},$$

per cui

$$\mathbf{E}\left[S_N^2(X)\right] = \sigma^2 \Delta t, \quad \mathbf{D}^2\left[S_N^2(X)\right] = \frac{2}{N-1}\sigma^4 \Delta t^2.$$

Per quanto precedentemente osservato, possiamo usare $\bar{X}_N$ come stimatore corretto di $\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t$ con errore quadratico pari a $\frac{\sigma^2 \Delta t}{N}$ e possiamo usare $S_N^2(X)$ come stimatore corretto di $\sigma^2 \Delta t$ con errore quadratico pari a $\frac{2}{N-1}\sigma^4 \Delta t^2$. Notiamo ancora che potendo scrivere

$$\frac{S_1}{S_0} = \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t + \sigma\sqrt{\Delta t} Z_1\right), \ldots, \frac{S_N}{S_{N-1}} = \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t + \sigma\sqrt{\Delta t} Z_N\right),$$

con $Z_0, \ldots, Z_N \sim \mathcal{N}(0,1)$ indipendenti. Risulta

$$\begin{aligned}
\frac{S_N}{S_0} &= \frac{S_1}{S_0}\frac{S_2}{S_1}\cdots\frac{S_{N-1}}{S_{N-2}}\frac{S_N}{S_{N-1}} = \exp\left(N\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t + \sigma\sqrt{\Delta t}\sum_{n=1}^{N} Z_n\right) \\
&= \exp\left(N\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t + \sqrt{N\Delta t}\sigma\left(\frac{1}{\sqrt{N}}\sum_{n=1}^{N} Z_n\right)\right) \\
&= \exp\left(N\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t + \sqrt{N\Delta t}\sigma Z\right) \\
&= \exp\left(T\left(\mu - \frac{1}{2}\sigma^2\right) + \sqrt{T}\sigma Z\right),
\end{aligned}$$

con $Z \equiv \frac{1}{\sqrt{N}}\sum_{n=1}^{N} Z_n \sim \mathcal{N}(0,1)$. Otteniamo allora

$$\mathbf{E}\left[\frac{S_N}{S_0}\right] = \exp\left(T\left(\mu - \frac{1}{2}\sigma^2\right)\right)\exp\left(+\frac{1}{2}T\sigma^2\right) = \exp(\mu T)$$

e

$$\mathbf{D}^2\left[\frac{S_N}{S_0}\right] = \exp\left(2T\left(\mu - \frac{1}{2}\sigma^2\right) + T\sigma^2\right)\left(\exp\left(T\sigma^2\right) - 1\right) = \exp(2T\mu)\left(\exp\left(T\sigma^2\right) - 1\right).$$

[12] Anche questa è un'ipotesi che ha contrassegnato le origini della finanza matematica e non può più essere assunta come pienamente valida.

[13] $X \sim N(\mu, \sigma^2) \Rightarrow \bar{X}_N \sim N(\mu, \sigma^2/N)$

[14] $X \sim N(\mu, \sigma^2) \Rightarrow (N-1)\, S_N^2(X)/\sigma^2 \sim \chi^2_{N-1}$

Inoltre,

$$\ln\left(\frac{S_N}{S_0}\right) = T\left(\mu - \frac{1}{2}\sigma^2\right) + \sqrt{T}\sigma Z.$$

Quindi,

$$\mathbf{E}\left[\ln\left(\frac{S_N}{S_0}\right)\right] = T\left(\mu - \frac{1}{2}\sigma^2\right) \quad \text{e} \quad \mathbf{D}^2\left[\ln\left(\frac{S_N}{S_0}\right)\right] = T\sigma^2.$$

Peraltro, nel modello abbiamo

$$S_N = \beta_N \cdots \beta_1 S_0.$$

Pertanto,

$$\ln\left(\frac{S_N}{S_0}\right) = \ln\left(\beta_N \cdots \beta_1\right) = \sum_{n=1}^{N} \ln\left(\beta_n\right).$$

Risulta allora,

$$\mathbf{E}\left[\ln\left(\frac{S_N}{S_0}\right)\right] = \mathbf{E}\left[\sum_{n=1}^{N} \ln\left(\beta_n\right)\right] = \sum_{n=1}^{N} \mathbf{E}\left[\ln\left(\beta_n\right)\right] = N\mathbf{E}\left[\ln\left(\beta_1\right)\right] = N\left(p\ln\left(u\right) + q\ln\left(d\right)\right).$$

Quindi

$$T\left(\mu - \frac{1}{2}\sigma^2\right) = N\left(p\ln\left(u\right) + q\ln\left(d\right)\right),$$

ovvero

$$\Delta t\left(\mu - \frac{1}{2}\sigma^2\right) = p\ln\left(u\right) + q\ln\left(d\right).$$

Inoltre

$$\mathbf{D}^2\left[\ln\left(\frac{S_T}{S_0}\right)\right] = \mathbf{D}^2\left[\sum_{n=1}^{N} \ln\left(\beta_n\right)\right] = \sum_{n=1}^{N} \mathbf{D}^2\left[\ln\left(\beta_n\right)\right] = N\mathbf{D}^2\left[\ln\left(\beta_1\right)\right],$$

essendo

$$\begin{aligned}
\mathbf{D}^2\left[\ln\left(\beta_1\right)\right] &= \mathbf{E}\left[\ln\left(\beta_1\right)^2\right] - \mathbf{E}\left[\ln\left(\beta_1\right)\right]^2 \\
&= p\ln(u)^2 + q\ln(d)^2 - \left(p\ln(u) + q\ln(d)\right)^2 \\
&= p\ln(u)^2 + q\ln(d)^2 - p^2\ln(u)^2 - q\ln(d)^2 - 2pq\ln(u)\ln(d) \\
&= p\left(1-p\right)\ln(u)^2 + q\left(1-q\right)\ln(d)^2 - 2pq\ln(u)\ln(d) \\
&= pq\ln(u)^2 + qp\ln(d)^2 - 2pq\ln(u)\ln(d) \\
&= pq\left(\ln(u) - \ln(d)\right)^2 \\
&= pq\ln\left(\frac{u}{d}\right)^2.
\end{aligned}$$

Otteniamo allora

$$T\sigma^2 = Npq\ln\left(\frac{u}{d}\right)^2,$$

da cui

$$\Delta t\sigma^2 = pq\ln\left(\frac{u}{d}\right)^2.$$

In definitiva, abbiamo costruito il sistema di equazioni

$$p\ln(u) + q\ln(d) = \left(\mu - \frac{1}{2}\sigma^2\right)\Delta t, \qquad pq\ln\left(\frac{u}{d}\right)^2 = \sigma^2\Delta t, \tag{3.168}$$

nelle effettive incognite p, u e d, essendo $q \equiv 1-p$, i cui membri di destra, nell'ipotesi di equidistribuzione normale e indipendenza dei rendimenti logaritmici dello stock, sono stimati dalla media e varianza campionaria, $\bar{X}_N$ e $S_N^2(X)$, dei rendimenti logaritmici stessi. Posto allora

$$m \equiv \mu - \frac{1}{2}\sigma^2, \tag{3.169}$$

e seguendo la scelta di Cox, Ross, Rubinstein

$$d \equiv 1/u, \tag{3.170}$$

dalle (3.168) otteniamo

$$(2p-1)\ln(u) = m\Delta t, \qquad 4p(1-p)\ln^2(u) = \sigma^2\Delta t. \tag{3.171}$$

Sommando i quadrati dei due membri della prima delle equazioni (3.171) ai due membri della seconda, segue

$$(2p-1)^2\ln^2(u) + 4p(1-p)\ln^2(u) = m^2\Delta t^2 + \sigma^2\Delta t, \tag{3.172}$$

ossia,

$$\ln^2(u) = m^2\Delta t^2 + \sigma^2\Delta t. \tag{3.173}$$

Ciò comporta

$$\ln(u) = \sqrt{m^2\Delta t^2 + \sigma^2\Delta t} = \sigma\sqrt{\Delta t + \left(\frac{m}{\sigma}\right)^2\Delta t^2},$$

da cui

$$u = e^{\sigma\sqrt{\Delta t + \left(\frac{m}{\sigma}\right)^2\Delta t^2}} \approx 1 + \sigma\sqrt{\Delta t}. \tag{3.174}$$

Per la (3.170), ne segue

$$d = e^{-\sigma\sqrt{\Delta t + \left(\frac{m}{\sigma}\right)^2\Delta t^2}} \approx 1 - \sigma\sqrt{\Delta t}. \tag{3.175}$$

Dalla (3.171), otteniamo poi

$$p = \frac{1}{2}\left(\frac{m\Delta t + \ln(u)}{\ln(u)}\right) = \frac{1}{2}\left(1 + \frac{m\Delta t}{\ln(u)}\right) = \frac{1}{2}\left(1 + \frac{m\Delta t}{\sigma\sqrt{\Delta t + \left(\frac{m}{\sigma}\right)^2\Delta t^2}}\right) = \frac{1}{2}\left(1 + \frac{m}{\sigma}\frac{\sqrt{\Delta t}}{\sqrt{1 + \left(\frac{m}{\sigma}\right)^2\Delta t}}\right).$$

Infine, scegliendo l'ampiezza Δt dei sottoperiodi in cui si è diviso il periodo di contrattazione $[0,T]$ molto piccola rispetto a T, ovvero scegliendo il numero N dei sottoperiodi molto grande, i parametri del reticolo binomiale possono essere scelti come segue

$$p = \frac{1}{2}\left(1 + \frac{m}{\sigma}\sqrt{\Delta t}\right), \tag{3.176}$$

$$u = e^{\sigma\sqrt{\Delta t}}, \tag{3.177}$$

$$d = e^{-\sigma\sqrt{\Delta t}}. \tag{3.178}$$

Con questa scelta il modello binomiale viene calibrato in modo da dare luogo a una corrispondenza abbastanza buona tra prezzi i prezzi dei derivati teoricamente stimati e i prezzi reali di mercato. Da notare che in questo modo si è calibrato il modello in funzione della distribuzione di probabilità oggettiva (p,q). Volendo calibrare il modello in funzione della distribuzione di probabilità neutrale al rischio $(\tilde{p},\tilde{q})$ vanno usate le equazioni

$$\tilde{p} = \frac{1+r-d}{u-d}, \tag{3.179}$$

$$u = e^{\sigma\sqrt{\Delta t}}, \tag{3.180}$$

$$d = e^{-\sigma\sqrt{\Delta t}}, \tag{3.181}$$

dove r è il tasso di rendimento non rischioso relativo all'intervallo Δt, che va supposto costante negli intervalli $[t_{n-1}, t_n]$ al variate di $n = 1, \ldots, N$. Il tasso r va stimato mediante i dati del mercato. A tal fine può essere utile considerare la seguente proposizione.

Proposizione 241 *Siano r e ρ il tasso di interesse non rischioso e il tasso d'interesse non rischioso continuamente composto corrisposti rispettivamente negli intervalli di tempo Δt e $[0, T]$ per l'investimento nel bond. La relazione tra r e ρ è data da*

$$r = \exp(\rho \Delta t) - 1, \qquad \rho = \frac{\log(1+r)}{\Delta t} \tag{3.182}$$

Proof. *Infatti, considerato il tasso d'interesse non rischioso r_T relativo al periodo $[0, T]$, si ha*

$$(1+r)^N = 1 + r_T = \exp(\rho T),$$

da cui

$$N \ln(1+r) = \rho T,$$

che comporta

$$\ln(1+r) = \rho \frac{T}{N} = \rho \Delta t.$$

L'Equazione 3.182 segue immediatamente. □

Notare che risulta

$$\frac{B_n}{B_{n-1}} = \exp(\rho \Delta t),$$

per ogni $n = 1, \ldots, N$.

Capitolo 4

Modelli di Prezzo a Tempo Continuo

I modelli di mercato a tempo continuo sono caratterizzati da una evoluzione temporale in cui l'insieme dei tempi $\mathbb{T}$ è un intervallo di $\mathbb{R}_0^+$, in genere $\mathbb{T} \equiv [0, T]$, per un qualche $T > 0$, o $\mathbb{T} \equiv \mathbb{R}_0^+$, secondochè i processi siano impiegati nell'ambito di modelli ad orizzonte temporale finito o infinito. Una prima opportunità di considerare modelli a tempo continuo deriva dall'osservazione che con l'avvento dei mercati telematici l'aggiornamento delle quotazioni dei titoli in un mercato finanziario è molto frequente. L'esperienza mostra che la frequenza di variazione delle quotazioni sul mercato è limitata solo dal tipo di tecnologia disponibile per veicolare le informazioni e nonostante gli intervalli tra un aggiornamento e l'altro non siano arbitrariamente piccoli, certamente non minori del tempo di reazione del sistema che disciplina la pubblicazione e l'esecuzione degli ordinativi immessi dagli operatori, risulta ormai di fatto impossibile distinguere tra una successione di quotazioni di mercato di un titolo e una successione ricavata dal rilevamento sperimentale dei valori di una qualsiasi grandezza variabile con continuità. Una seconda ragione che spiega la rapida diffusione dei modelli a tempo continuo è di carattere più pragmatico. Infatti, per lo studio di tali modelli sono disponibili le tecniche del calcolo stocastico, che consentono di estrarre più risultati di quanti ne consentano le tecniche dei modelli a tempo discreto. Talvolta accade che l'accordo tra i risultati conseguiti e le osservazioni empiriche è tale che di per sè conduce ad accettare la validità del modello proposto. Il decennio 1969-1979 viene considerato una sorta di "decennio d'oro" per le teorie di asset pricing basate sui modelli a tempo continuo. In particolare, all'inizio degli anni '70, Fisher Black, Myron Scholes e Robert Merton hanno dato un fondamentale contributo alla teoria di valutazione delle opzioni, sviluppando l'omonimo modello di Black & Scholes. La formula che deriva da questo modello ha avuto un'enorme influenza sul modo in cui gli operatori valutano le opzioni ed effettuano le coperture, tanto che la sua importanza è stata celebrata nel 1997, purtroppo dopo la scomparsa di F. Black, con l'assegnazione del premio Nobel per l'economia a M. Scholes e R. Merton.

4.1 Modello di Black & Scholes

Il modello di Black & Scholes si riferisce alle opzioni europee sui titoli rischiosi. L'equazione centrale del modello, nota per l'appunto come *formula di di Black & Scholes*, è stata derivata da F. Black e M. Scholes in un lavoro del 1973, sulla base di precedenti ricerche di R. Merton e Paul Samuelson. L'idea alla base del modello è che, in assenza di arbitraggio e sotto ulteriori ipotesi di natura essenzialmente tecnica, si possa prezzare un titolo derivato, in particolare un'opzione, mediante un portafoglio replicante. Le ulteriori ipotesi del modello di Black & Scholes sono:

1. nel mercato vengono scambiati un titolo non rischioso, diciamo un *bond*, un titolo rischioso, *stock*, e opzioni europee d'acquisto, *european call option*, e vendita, *european put option*, con sottostante lo stock;

2. è possibile scambiare frazioni arbitrarie di ogni titolo sul mercato;

3. è consentita la vendita allo scoperto di ogni titolo sul mercato;

4. nel mercato non sussistono costi di transazione, tassazione, né frizioni di altro tipo;

5. sia il bond, che lo stock, che le opzioni possono essere scambiati sul mercato a ogni istante dell'intervallo di tempo $[0, T]$ (dove T è generalmente espresso in termini di anni);

6. il tasso d'interesse non rischioso r (generalmente annuale continuamente composto) pagato del bond nell'intervallo di tempo $[0, T]$ è costante;

7. il prezzo dello stock nell'intervallo di tempo $[0, T]$ segue un moto browniano geometrico caratterizzato da un tasso di rendimento atteso (generalmente annuale) μ e da una volatilità (generalmente annualizzata) σ.

Come immediate conseguenze delle ipotesi del modello, si ottiene che il processo di prezzo del bond $(B_t)_{t=0}^T \equiv B$ è modellato dall'equazione differenziale deterministica

$$dB_t = rB_t dt, \tag{4.1}$$

di soluzione

$$B_t = B_0 e^{rt},$$

al variare di $t \in [0, T]$, e il processo di prezzo del titolo rischioso $(S_t)_{t=0}^T \equiv S$ è modellato dall'equazione differenziale stocastica

$$dS_t = \mu S_t \, dt + \sigma S_t \, dW_t, \tag{4.2}$$

di soluzione

$$S_t = S_0 \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W_t\right),$$

al variare di $t \in [0, T]$. La variazione quadratica del processo S è allora data da

$$d\left[S\right]_t = \sigma S_t^2 \, dt, \tag{4.3}$$

al variare di $t \in [0, T]$.

Proposizione 242 *Il valore di una qualsiasi somma di denaro M_0 al tempo $t = 0$ ha un valore capitalizzato al tempo t di*

$$M_t = M_0 e^{rt} \tag{4.4}$$

per ogni $t \in [0, T]$. Viceversa, il valore di una somma M_T di denaro al tempo $t = T$ ha un valore scontato al tempo t di

$$M_t = M_T e^{-r(T-t)} \tag{4.5}$$

per ogni $t \in [0, T]$.
Proof. *La disponibilità di denaro M_0 al tempo $t = 0$ consente l'acquisto di*

$$x = \frac{M_0}{B_0}$$

unità di titolo non rischioso di valore di mercato B_0. La liquidazione di questo investimento al tempo t produce un'ammontare M_t secondo la formula

$$M_t = xB_t = \frac{M_0}{B_0} B_0 e^{rt} = M_0 e^{rt}.$$

Viceversa, volendo produrre un montante M_T al tempo $t = T$ è necessario liquidare

$$x = \frac{M_T}{B_T}$$

unità di titolo non rischioso di valore di mercato B_T. D'altra parte l'acquisto di tali unità di titolo non rischioso al tempo t richiede l'investimento di una somma di denaro M_t pari a

$$M_t = xB_t = \frac{M_T}{B_T}B_t = \frac{M_T}{e^{r(T-t)}B_t}B_t = M_T e^{-r(T-t)}.$$

□

Definizione 243 *Chiamiamo* portafoglio *del modello di Black & Scholes un processo stocastico $\Pi \equiv (\Pi_t)_{t=0}^T \equiv (X_t, Y_t)_{t=0}^T$ a stati in $\mathbb{R}^2$, le cui componenti $(X_t)_{t=0}^T$ e $(Y_t)_{t=0}^T$ siano processi* predicibili *con termini X_t ed Y_t rispettivamente rappresentanti le quantità di bond e di stock allocate in portafoglio al tempo t, al variare di $t \in [0, T]$.*

Definizione 244 *Chiamiamo* valore *del portafoglio $\Pi \equiv (X_t, Y_t)_{t=0}^T$ il processo stocastico $(V_t(\Pi))_{t=0}^T$ definito ponedo*

$$V_t(\Pi) \stackrel{def}{=} X_t B_t + Y_t S_t, \quad t \in [0, T].$$

Definizione 245 *Diciamo che un portafoglio $\Pi \equiv (X_t, Y_t)_{t=0}^T$ è* autofinanziante *se le componenti $(X_t)_{t=0}^T$ e $(Y_t)_{t=0}^T$ sono tali che*

$$\int_0^T |X_s| \, ds < \infty, \qquad \int_0^T Y_s^2 \, ds < \infty, \tag{4.6}$$

e

$$d\left(X_t B_t + Y_t S_t\right) = X_t \, dB_t + Y_t \, dS_t, \tag{4.7}$$

per ogni $t \in [0, T]$.

Proposizione 246 *In termini differenziali la condizione (??) di autofinanziamento si esprime come*

$$X_t B_t + Y_t S_t = X_0 B_0 + Y_0 S_0 + \int_0^t X_s \, dB_s + \int_0^t Y_s \, dS_s \tag{4.8}$$

al variare di $t \in [0, T]$.
Proof. *Integrando ambo i membri dell'Equazione (4.7), otteniamo*

$$X_t B_t + Y_t S_t - X_0 B_0 + Y_0 S_0 = \int_0^t X_s \, dB_s + \int_0^t Y_s \, dS_s,$$

per ogni $t \in [0, T]$, cioè la (4.8). Viceversa, differenziando ambo i membri dell'Equazione (4.8) segue immediatamente la (4.7). □

Ovviamente, stanti la (4.1) e la (4.2), risulta

$$X_t \, dB_t = rX_t B_t \, dt \quad \text{e} \quad Y_t \, dS_t = \mu Y_t S_t \, dt + \sigma Y_t S_t \, dW_t,$$

per ogni $t \in [0, T]$. In forma integrale,

$$\int_0^t X_s \, dB_s = r \int_0^t X_s B_s \, ds$$

e

$$\int_0^t Y_s \, dS_s = \mu \int_0^t Y_s S_s \, ds + \sigma \int_0^t Y_s S_s \, dW_s,$$

per ogni $t \in [0, T]$.

Osservazione 247 *Se* $\Pi \equiv (X_t, Y_t)_{t=0}^T$ *è un portafoglio autofinanziane, si ha*

$$d\left(X_t B_t + Y_t S_t\right) = \left(r X_t B_t + \mu Y_t S_t\right) dt + \sigma Y_t S_t \, dW_t \tag{4.9}$$

Definizione 248 *Diciamo che un portafoglio autofinanziante* $\Pi \equiv (X_t, Y_t)_{t=0}^T$ *è un* arbitraggio *se a partire da un valore iniziale nullo,* $V_0(\Pi) = 0$*, risulta*

$$V_T \geqslant 0, \qquad e \qquad \mathbf{P}(V_T > 0) > 0. \tag{4.10}$$

Lemma 249 *Siano* U *e* V *due qualsiasi titoli del mercato con processi di prezzo* $(U_t)_{t=0}^T$ *e* $(V_t)_{t=0}^T$ *rispettivamente strutturati in modo tale che valga l'uguaglianza*

$$U_T = V_T. \tag{4.11}$$

Se nel mercato non sono possibili portafogli d'arbitraggio risulta necessariamente

$$U_t = V_t \tag{4.12}$$

per ogni $t \in [0, T]$.
Proof. *Se a un certo tempo* $t_0 \in [0, T)$ *risultasse*

$$U_{t_0} > V_{t_0},$$

allora vendute allo scoperto n_U *unita di titolo* U*, con il ricavato* $n_U U_{t_0}$ *si potrebbero acquistare* $Y_{t_0} \equiv n_U$ *unità di titolo* V *con un esborso pari a* $n_U V_{t_0}$ *ed investire il surpuls* $n_U \left(U_{t_0} - V_{t_0}\right)$ *nell'acquisto di* $X_{t_0} \equiv n_U \left(U_{t_0} - V_{t_0}\right) / B_{t_0}$ *unità di titolo non rischioso con esborso pari a* $n_U \left(U_{t_0} - V_{t_0}\right)$*. Infatti, si ha chiaramente*

$$n_U U_{t_0} = n_U \left(U_{t_0} - V_{t_0}\right) + n_U V_{t_0}.$$

Consideriamo quindi il portafoglio $\Pi \equiv (X_t, Y_t)_{t=0}^T$ *tale che*

$$X_t = \begin{cases} 0 & \text{if } t \in [0, t_0) \\ X_{t_0} & \text{if } t \in [t_0, T] \end{cases} \qquad Y_t = \begin{cases} 0 & \text{if } t \in [0, t_0) \\ Y_{t_0} & \text{if } t \in [t_0, T] \end{cases}.$$

Al tempo $t = T$ *la liquidazione di tale portafoglio in merito alla posizione sul titolo* V *produrrebbe un introito* $Y_T V_T = Y_{t_0} V_T = n_U V_T$ *tale da garantire la possibilità di ripianare lo scoperto della posizione su* U*, nel frattempo maturato a* $n_U U_T$*, e al contempo la liquidazione della posizione sul titolo non rischioso garantirebbe un introito certo pari a*

$$X_T U_T = X_{t_0} U_T = \frac{n_U \left(U_{t_0} - V_{t_0}\right)}{B_{t_0}} B_T = \frac{n_U \left(U_{t_0} - V_{t_0}\right)}{B_{t_0}} e^{(T - t_0)} B_{t_0} = n_U \left(U_{t_0} - V_{t_0}\right) e^{(T - t_0)} > 0.$$

Si sarebbe quindi realizzato un arbitraggio. Un arbitraggio del tutto analogo si potrebbe realizzare se al tempo $t_0 \in [0, T)$ *fosse* $U_{t_0} < V_{t_0}$*. Non rimane che concludere circa la veridicità della tesi del Lemma.*
□

Consideriamo adesso un'opzione europea d'acquisto C sul titolo rischioso S. Ricordiamo che un'opzione europea d'acquisto [risp. di vendita] è un contratto tra due parti contraenti, un titolare ed un garante. Il titolare dietro il pagamento di un premio, all'istante della stipula del contratto, si riserva il diritto, ma non l'obbligo, di acquistare dal garante [risp. vendere al garante] un attivo finanziario rischioso ad una scadenza e a un prezzo pattuiti all'atto della stipula del contratto. Il garante, in cambio del premio, si fa carico dell'obbligo di soddisfare il titolare.

Indicati con:

- $t = 0$ l'istante della stipula del contratto d'opzione;

- C_0 il valore del premio, prezzo dell'opzione al tempo $t = 0$;
- $t = T$ l'istante di scadenza, cosiddetta *maturity*, del contratto d'opzione (espressa in termini di anni);
- K il prezzo d'esercizio, cosiddetto *exercise* o *strike price*, pattuito nel contratto d'opzione;

il titolare di un'opzione d'acquisto [risp. di vendita] sarà interessato ad esercitare il proprio diritto solo se risulterà

$$S_T - K > 0 \qquad [\text{risp. } S_T - K < 0].$$

Pertanto, alla scadenza T, il valore di un'opzione europea d'acquisto [risp. di vendita] è quantificato da

$$C(T, S_T) \equiv (S_T - K)^+ \equiv \max\{S_T - K, 0\} \qquad [\text{risp. } P(T, S_T) \equiv (K - S_T)^+ \equiv \max\{0, S_T - K\}].$$

Si pongono i seguenti problemi:

- la valutazione del giusto prezzo d'esercizio, C_0, per il contratto di opzione *(option pricing)* che dia al titolare dell'opzione la possibilità di realizzare un utile alla scadenza;
- l'individuazione di una stategia finanziaria di copertura *(hedging strategy)*, ossia un opportuno portafoglio $\Pi \equiv (X_t, Y_t)_{t=0}^T$, che dia al garante dell'opzione la possibilità di produrre, a partire dall'incasso del premio, una ricchezza almeno pari al valore dell'opzione.

Da notare che la vendita di un'opzione d'acquisto, cui non faccia seguito un'adeguata strategia di copertura, può causare al garante perdite potenzialmente illimitate alla scadenza.

Definizione 250 *Chiamiamo* strategia di copertura *(*hedging strategy*) dall'esercizio dell'opzione call [resp. put] un portafoglio autofinanziante $\Pi \equiv (X_t, Y_t)_{t=0}^T$ tale che all'istante terminale $t = T$ si abbia*

$$X_T B_T + Y_T S_T = C_T \qquad [\textit{risp. } X_T B_T + Y_T S_T = P_T]. \tag{4.13}$$

Lemma 251 *Sia $(C_t)_{t=0}^T$ [risp. $(P_t)_{t=0}^T$] il processo dei prezzi di una opzione call [resp. put] sul titolo rischioso S e sia $\Pi \equiv (X_t, Y_t)_{t=0}^T$ una strategia di copertura dall'esercizio dell'opzione. In assenza d'arbitraggi, deve necessariamente aversi*

$$X_t B_t + Y_t S_t = C_t \qquad [\textit{risp. } X_t B_t + Y_t S_t = P_t], \tag{4.14}$$

per ogni $t \in [0, T]$.
Proof. *The statement of Lemma 251 is a clear consequence of Lemma 249.* □

Teorema 252 *Sia $(C_t)_{t=0}^T$ [risp. $(P_t)_{t=0}^T$] il processo dei prezzi di una opzione call [resp. put] sul titolo rischioso S. Si supponga che esista una strategia di copertura dall'esercizio dell'opzione call o dell'opzione put. In assenza d'arbitraggi vale l'equazione di* put-call parity

$$P_t = C_t - S_t + K e^{-r(T-t)}, \tag{4.15}$$

per ogni $t \in [0, T]$.
Proof. *Osserviamo preliminarmente che dalla definizione stessa di opzione call e put deve aversi*

$$P_T = C_T - S_T + K. \tag{4.16}$$

Infatte, se $S_T > K$ allora $C_T = S_T - K$, $P_T = 0$ e la (4.16) è verificata. Se $S_T < K$ allora $C_T = 0$, $P_T = K - S_T$ e la (4.16) è ancora verificata. Infine se $S_T = K$ la (4.16) è chiaramente vera. Supponiamo *ora che esista una strategia di copertura dall'esercizio dell'opzione call, sia essa* $\Pi \equiv (X_t, Y_t)_{t=0}^T$. *Da defizione risulta*

$$C_T = X_T S_T + Y_T B_T$$

ma allora per la (4.14) del Lemma 251 deve aversi

$$C_t = X_t S_t + Y_t B_t$$

per ogni $t \in [0, T]$. *D'altra parte stante la (4.16) deve anche aversi*

$$P_T = X_T S_T + Y_T B_T - S_T + K$$

Quindi

$$P_T = (X_T - 1) S_T + \left(Y_T + \frac{K}{B_T} \right) B_T.$$

Considerato che una somma di denaro K *al tempo* T *ha un valore* $Ke^{-r(T-t)}$ *al tempo* t *(see Equation (4.5) nella Proposizione 242), riapplicando la (4.14) del Lemma 251, otteniamo*

$$P_t = (X_t - 1) S_t + \left(Y_t + \frac{Ke^{-r(T-t)}}{B_t} \right) B_t,$$

Ossia

$$P_t = X_t S_t + Y_t B_t - S_t + Ke^{-r(T-t)} = C_t - S_t + Ke^{-r(T-t)}$$

come desiderato. □

In assenza d'arbitraggi, stante la put-call parity (see (4.15) in Theorem 4.15), ci si può limitare a considerare il problema dell'option pricing e dell'individuazione di una hedging strategy in riferimento alla sola opzione call.

Il risultato di Black & Scholes consente di determinare una formula esplicita per il valore di C_t ad ogni istante $t \in [0, T]$. In particolare il premio C_0 da pagare all'atto della sottoscrizione del contratto, nonchè la composizione del portafoglio di copertura Π ad ogni istante $t \in [0, T]$.

Teorema 253 (Black & Scholes European Call Option Price) *Sia* $(C_t)_{t=0}^T$ *il processo dei prezzi di una opzione call sul titolo rischioso* S *e sia* $\Pi \equiv (X_t, Y_t)_{t=0}^T$ *una strategia di copertura dall'esercizio dell'opzione. Risulta*

$$C_t \equiv C(t, S_t) = \begin{cases} S_t \Phi(d_+(T-t, S_t)) - Ke^{-r(T-t)} \Phi(d_-(T-t, S_t)) & 0 \leqslant t < T, S_t \geqslant 0 \\ (S_T - K)^+ & t = T, S_T \geqslant 0 \end{cases}. \tag{4.17}$$

essendo $\Phi(x)$ *la funzione di distribuzione della variabile aleatoria Gaussiana standardizzata,*

$$\Phi(x) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-z^2/2} dz, \tag{4.18}$$

per ogni $x \in \mathbb{R}$, *ed essendo*

$$d_\pm(T-t, S_t) \equiv \frac{1}{\sigma\sqrt{T-t}} \left(\ln\left(\frac{S_t}{K}\right) + (r \pm \frac{\sigma^2}{2})(T-t) \right), \tag{4.19}$$

per ogni $t \in [0, T)$ *ed ogni* $S_t > 0$. *Inoltre, la stategia di copertura è data da*

$$X_t = \frac{C(t, S_t) - Y_t S_t}{B_t}, \qquad Y_t = \partial_x C(t, S_t). \tag{4.20}$$

In particolare,

$$C_0 = S_0\Phi(d_+(T,S_0)) - Ke^{-rT}\Phi(d_-(T,S_0)) \tag{4.21}$$

e

$$X_0 = \frac{C_0 - Y_0S_0}{B_0}, \qquad Y_0 = \partial_x C\,(t,x)\,|_{(t,x)=(0,S_0)}. \tag{4.22}$$

Proof. *Lo schema della prova consiste nel supporre, in assenza d'arbitraggio, l'esistenza di una strategia di copertura dall'esercizio dell'opzione e grazie a questa ipotesi determinare sia un'equazione per il prezzo dell'opzione che un'equazione per la stessa strategia di copertura in funzione del prezzo dell'opzione. Quindi, determinata una soluzione dell'equazione per il prezzo dell'opzione viene anche garantita l'esistenza della strategia di copertura.*

Assunta, in assenza d'arbitraggio, l'esistenza di una strategia di copertura dall'esercizio dell'opzione, cioè un portafoglio autofinanziante $\Pi \equiv (X_t, Y_t)_{t=0}^T$ *tale che*

$$X_tS_t + Y_tB_t = C_t \tag{4.23}$$

per ogni $t \in t \in [0,T]$ *(see Equation 4.14 in Lemma 251), si osserva che la dinamica deterministica del bond da modo di stabilire una relazione funzionale tra il prezzo* C_t *dell'opzione con il solo termine stocastico* S_t *del tipo*

$$C_t = C(t,S_t). \tag{4.24}$$

Ipotizzando allora che una tale relazione C *possa essere rappresentata da una funzione sufficientemente regolare, precisamente di classe* $C^{1,2}([0,T]\times\mathbb{R}_+;\mathbb{R})$*, ci troviamo nelle condizioni per poter applicare la formula di Ito ottenendo*

$$dC_t = \partial_t C(t,x)|_{x=S_t}dt + \partial_x C(t,x)|_{x=S_t}dS_t + \frac{1}{2}\partial_{xx}C(t,x)|_{x=S_t}d\,[S]_t\,. \tag{4.25}$$

Combinando la (4.25) con la (4.2) e (4.3), risulta allora

$$\begin{aligned} dC_t &= \partial_t C(t,x)|_{x=S_t}dt + \partial_x C(t,x)|_{x=S_t}(\mu S_t\,dt + \sigma S_t\,dW_t) + \frac{1}{2}\partial_{xx}C(t,x)|_{x=S_t}\sigma S_t^2\,dt \\ &= \left(\partial_t C(t,S_t) + \mu\partial_x C(t,S_t)S_t + \frac{1}{2}\sigma\partial_{xx}C(t,S_t)S_t^2\right)dt + \sigma\partial_x C(t,S_t)S_t\,dW_t. \end{aligned} \tag{4.26}$$

D'altra parte, considerando la (4.23) in forma differenziale e tenendo conto della condizione d'autofinanziamento (see Equation (4.9) nell'Osservazione 247) possiamo scrivere

$$\begin{aligned} &(rX_tB_t + \mu Y_tS_t)\,dt + \sigma Y_tS_t\,dW_t \\ &= \left(\partial_t C(t,S_t) + \mu\partial_x C(t,S_t)S_t + \frac{1}{2}\sigma\partial_{xx}C(t,S_t)S_t^2\right)dt + \sigma\partial_x C(t,S_t)S_t\,dW_t. \end{aligned} \tag{4.27}$$

La condizione che le strategie di portafoglio abbiano lo stesso profilo di rischio dell'opzione suggerisce di uguagliare i coefficienti del differenziale stocastico dW_t *ottenendo la condizione di copertura*

$$Y_t = \partial_x C\,(t,S_t)\,. \tag{4.28}$$

Quindi, considerata la (4.28), l'Equazione (4.27) diviene

$$rX_tB_t = \partial_t C\,(t,S_t) + \frac{1}{2}\sigma^2\partial_{xx}C\,(t,S_t)\,S_t^2. \tag{4.29}$$

Combinando la (4.23) e (4.28) segue

$$X_t\,B_t = C\,(t,S_t) - \partial_x C\,(t,S_t)\,S_t. \tag{4.30}$$

Sostituendo il risultato di quest'ultima nell'Equazione (4.29) otteniamo

$$\partial_t C(t, S_t) = rC(t, S_t) - \partial_x C(t, S_t) S_t - \frac{1}{2}\sigma^2 \partial_{xx} C(t, S_t) S_t^2. \tag{4.31}$$

L'equazione (??) suggerisce allora di cercare C come la soluzione dell'equazione differenziale parabolica

$$\partial_t C(t, x) = rC(t, x) - rx\partial_x C(t, x) - \frac{1}{2}\sigma^2 x^2 \partial_{xx} C(t, x), \tag{4.32}$$

con dato terminale

$$C(T, x) = \max\{x - K, 0\}, \quad t = T, \tag{4.33}$$

che è la celebre equazione di Black & Scholes *avente per soluzione la funzione data dalla (4.17). Un'equazione di questo tipo è nota come* backward Kolmogorov equation. *Infine, l'equazione (4.20) segue immediatamente combinando la (4.28) e (4.30).* □

Esempio 254 *Consideriamo un'opzione call europea su un sottostante S che quoti $S_0 = \$52$ all'atto dell'acquisto della call. Assumiamo che il prezzo d'esercizio della call sia $K = \$50$ e la sua maturità sia $T = 3$ mesi. Assumiamo inoltre S non paghi dividenti nel periodo di vita della call, che la volatilità del sottostante σ sia pari al 30% annuo e che il tasso d'interesse non rischioso r sia del 2% annuo. Desideriamo calcolare il prezzo d'acquisto della call.*

Dal punto di vista di un practitioner, la stima annua della volatilità del sottostante σ ottenuta dai prezzi di chiusura giornalieri aggiustati è effettuata sulla base di una media di 252 dati (average number of trading days in a year). L'opzione giunge a maturità in tre mesi che corrispondono mediamente a 3×21 (average number of trading days in a month) giorni di mercato. Per omogeneità dei tempi di riferimento, ciò suggerisce di porre $T = 63/252 = 0.25$. Possiamo allora riassumere i dati disponibili nella seguente tabella

$$S_0 = \$52, \quad K = \$50, \quad r = 0.02, \quad \sigma = 0.3, \quad T = 0.25.$$

Applicando la (4.19), otteniamo

$$\begin{aligned} d_+(T, S_0) &= \frac{1}{0.3\sqrt{0.25}}\left(\ln\left(\frac{52}{50}\right) + \left(0.02 + \frac{0.3^2}{2}\right)0.25\right) = 0.3698, \\ d_-(T, S_0) &= \frac{1}{0.3\sqrt{0.25}}\left(\ln\left(\frac{52}{50}\right) + \left(0.02 - \frac{0.3^2}{2}\right)0.25\right) = 0.2198. \end{aligned}$$

Dalla (4.21) segue allora

$$C_0 = \$\left(52\Phi(0.3698) - 50e^{-0.02\times 0.25}\Phi(0.2198)\right) = \$4.2972$$

The NASDAQ composite is composed of over 2,500 securities listed in the NASDAQ exchange. The NASDAQ Composite Index includes almost every security listed in the NASDAQ exchange. The NASDAQ 100 index tracks the 100 largest non-financial stocks listed on the NASDAQ exchange. The NASDAQ 100 is very focused on the technology sector, and is a widely-held tracking index for futures, options and exchange-traded fund trading.

$$dW_t dW_t = dt \qquad \mathbf{E}\left[W_t^2\right] = t$$

$$f(t, x)$$

$$f(t, x) = tx....t^2x....te^x...\log(tx) = \frac{x}{tx} = \frac{1}{t}$$

$$\left(\frac{\partial f}{\partial x}\right)_{x=Y_t}$$

Esempio 255 *In data* $T_0 \equiv 2019/06/28$ *risultava una quotazione di chiusura di* $S_{T_0} \equiv \$7,693.75$ *del future sul Nasdaq 100 denominato E-Mini Sept '19 (NQU19.CME) con maturità* $T_1 = 2019/09/20$*. Sempre in data* $T_0 \equiv 2019/06/28$ *la Treasury Yield Curve presentava un rendimento richiesto annualizzato,* annualized asked yield, *con maturità al* $T_1 = 2019/09/20$ *del* 2.27% *(see Wall Street Journal, https://www.wsj.com/market-data/bonds/treasuries). Note that the annualized asked yield of the treasury yield curve is computed on the basis of the day count factor (DCF)*

$$\boldsymbol{DCF}(T_1 - T_0) = \frac{360 \times (Y_1 - Y_0) + 30 \times (M_1 - M_0) + (D_1 - D_0)}{360},$$

which, in our case, yields

$$\boldsymbol{DCF}(T_1 - T_0) = \frac{360 \times 0 + 30 \times (9 - 6) + (20 - 28)}{360} = \frac{82}{360} = 0.2278.$$

Now, suppose that at t_0 *we want to compute the price of European call and put options with underlying the Nasdaq 100 E-Mini Sep '19, different strikes, and maturity on* $2019/09/19$ *(see https://www.barchart.com/futures/ To apply the Black & Scholes formula, we only need to determine the Nasdaq 100 E-Mini Sep '19 volatility. To this task we download the dividend adjusted last prices of the Nasdaq 100 E-Mini Sep '19* $(s_t)_{t=t_0}^{T_0}$*, from* $t_0 = 2019/01/02$ *to* $T_0 \equiv 2019/06/28$ *(see https://www.barchart.com/futures/quotes/NQU19/price-history/historical) and we compute the logarithm returns* $(r_t)_{t=t_0}^{T_0}$*, where*

$$r_t \overset{def}{=} \ln\left(\frac{s_t}{s_{t-1}}\right), \quad \forall t = t_0 + 1, \ldots, T_0.$$

From the sample scatter and line plot we have:

1. *a slight evidence for nonstationarity;*
2. *no clear evidence for autoregression;*
3. *some evidence for non normality;*
4. *a slight evidence for heteroskedasticity;*
5. *a slight evidence for some structure break point.*

The KPSS test does not allow to reject the null hypothesis of stationarity at the 0.05 *significance level and the Augmented Dickey-Fuller test rejects the null hypothesis of unit root at the* 0.05 *significance level. Hence, we consider the logarithm return time series as it were stationary.*

The visual inspection of the autocorrelogram and partial autocorrelogram provides no evidence of autocorrelation. This is confirmed by the Ljung-Box test which does not allow to reject the null hypothesis of no autocorrelation at the 0.05 *significance level. Hence, we consider the logarithm return time series as it had no autocorrelation.*

The visual inspection of the QQ-plot of the logarithm return time series against the standard Gaussian distribution provides a rather strong evidence that the logarithm returns are not normally distributed. More precisely that the logarithm returns have an heavy tail distribution. This is confirmed by the Lilliefors, Jarque-Bera, Shapiro-Wilk, and D'Agostino-Pearson tests all rejecting the normality assumption at the significance level $\alpha = 0.05$*.*

The Breusch-Pagan, Score and F tests for hereoskedasticity (on the linear model associated to the time series) provide no evidence of heteroskedasticity.

No evidence for structural break points in the logarithm return time series.

In light of the above evidences, besides the rejection of the normality null, the time series might be modelled by Gaussian white noise. Nevertheless, the rejection of the normality null is a serious issue *to prevent this possibility. On the other hand, it is possible to observe that the transaction volumes are very thin, but in the most recent part of the time series. This might suggest to select the most recent part of the time series filtering on the transaction volume in such a way that there are still enough terms to apply the normality tests (at least* 20 *terms). In our case, selecting the last* 20 *terms of the time series we have that the transation volume is always larger than* $\$5,000$*, and the Lilliefors, Jarque-Bera, Shapiro-Wilk, and D'Agostino-Pearson tests do not allow to reject the normality assumption at the significance level* $\alpha = 0.05$*, not even at* $\alpha = 0.1$*. On the basis of this smaller data sample we can estimate a daily standard deviation* $\hat{\sigma}_d = 0.00773$*. According to the standard convention, this corresponds to an estimate of the annualized standard deviation given by*

$$\hat{\sigma}_y = \hat{\sigma}_d \times \sqrt{360} = 0.00773 \times \sqrt{360} = 0.1467 = 14.67\%$$

Finally we can apply the Black & Scholes formula for the price of a Call [resp. Put] option with maturity $T_1^* = 2019/09/19$ *in date* $T_0 \equiv 2019/06/28$ *with strike prices at the money* $K_{A.M.} \equiv \$7,690$*, in the money* $K_{I.M.} \equiv \$5,000$*, and out the money* $K_{O.M.} \equiv \$9,000$*. We have*

$$S_0 \equiv S_{T_0} = \$7,693.75, \quad r_0 = 2.27\%, \quad \hat{\sigma}_y = 14.67\%, \quad T \equiv T_1^* - T_0 = 0.225, \quad K \equiv \begin{cases} \$9,000 \\ \$7,690 \\ \$5,000 \end{cases}.$$

Applicando la (4.19), otteniamo

$$\begin{aligned} d_+(T, S_0) &= \frac{1}{\sigma\sqrt{T}}\left(\ln\left(\frac{S_0}{K_{A.M.}}\right) + (r + \frac{\sigma^2}{2})T\right) \\ &= \frac{1}{0.1467\sqrt{0.225}}\left(\ln\left(\frac{7693.75}{7690.00}\right) + \left(0.0227 + \frac{0.1467^2}{2}\right)0.225\right) \\ &= 0.11620, \\ d_-(T, S_0) &= \frac{1}{\sigma\sqrt{T}}\left(\ln\left(\frac{S_0}{K_{O.M.}}\right) + (r - \frac{\sigma^2}{2})T\right) \\ &= \frac{1}{0.1467\sqrt{0.225}}\left(\ln\left(\frac{7693.75}{7690.00}\right) + \left(0.0227 - \frac{0.1467^2}{2}\right)0.225\right) \\ &= 0.05612 \end{aligned}$$

Dalla (4.21) segue allora

$$\begin{aligned} C_0(T, K_{A.M.}) &= S_0\Phi(d_+(T, S_0)) - Ke^{-rT}\Phi(d_-(T, S_0)) \\ &= \$\left(7693.75\Phi(0.11620) - 9000.00e^{-0.0225\times 0.225}\Phi(0.05612)\right) \\ &= \$235.10 \end{aligned}$$

Similarly,

$$C_0(T, K_{O.M.}) = \$2.98 \quad \textit{and} \quad C_0(T, K_{I.M.}) = 2719.22$$

Note that the last available market price of the American call options on the Nasdaq 100 E-Mini Sep '19 at T_0 *on* $02:20$*am CDT with strikes* $K_{A.M.}$*,* $K_{O.M.}$*, and* $K_{I.M.}$*(not available the European call closing prices) were*

$$C_0^{MP}(T, K_{A.M.}) = \$267.00, \quad C_0^{MP}(T, K_{O.M.}) = \$2.40, \quad C_0^{MP}(T, K_{I.M.}) = \$2690.50,$$

respectively.

Come già osservato, in assenza d'arbitraggi, la put-call parity (see (4.15) in Theorem 4.15) consente di determinare elementarmente una formula esplicita per il valore di una put di sottostante S ad ogni istante $t \in [0, T]$. Abbiamo infatti

Corollary 256 (Black & Scholes European Put Option Price) *Sia $(P_t)_{t=0}^T$ il processo dei prezzi di una opzione put sottostante S e sia $\Pi \equiv (X_t, Y_t)_{t=0}^T$ una strategia di copertura dall'esercizio dell'opzione. Risulta*

$$P_t \equiv P(t, S_t) = \begin{cases} Ke^{-r(T-t)}\Phi(-d_-(T-t, S_t)) - S_t\Phi(-d_+(T-t, S_t)) & 0 \leqslant t < T, S_t \geqslant 0 \\ (K - S_T)^+ & t = T, S_T \geqslant 0 \end{cases} -, \tag{4.34}$$

con la stessa definizione delle funzioni Φ e $d_\pm$ che in Theorem 253. Inoltre, la stategia di copertura è data da

$$X_t = \frac{P(t, S_t) - Y_t S_t}{B_t}, \qquad Y_t = \partial_x P(t, S_t). \tag{4.35}$$

In particolare,

$$P_0 = Ke^{-rT}\Phi(-d_-(T, S_0)) - S_0(\Phi(-d_+(T, S_0))) +$$

e

$$X_0 = \frac{P_0 - Y_0 S_0}{B_0}, \qquad Y_0 = \partial_x P(t, x)\,|_{(t,x)=(0,S_0)}.$$

Proof. *Combinando la (4.17) con la put-call-parity (4.15), otteniamo*

$$P_t = Ke^{-r(T-t)}(1 - \Phi(d_-(T-t, S_t))) - S_t(1 - \Phi(d_+(T-t, S_t))).$$

D'altra parte per le proprietà della distribuzione di probabilità Gaussiana standard

$$1 - \Phi(d_\pm(T-t, S_t)) = \Phi(-d_\pm(T-t, S_t))$$

e la (4.34) segue immediatamente. □

Definizione 257 (Black & Scholes Equation) *Chiamiamo Black & Scholes Equation l'equazione differenziale parabolica*

$$\partial_t F(t, x) = rF(t, x) - rx\partial_x F(t, x) - \frac{1}{2}\sigma^2 x^2 \partial_{xx} F(t, x). \tag{4.36}$$

per una funzione reale F avente per dominio un sottoinsieme di $\mathbb{R}_+ \times \mathbb{R}$.

E' importante osservare che nelle Equazioni (4.17) e (4.34) non compare il parametro μ, il tasso di rendimento atteso del titolo rischioso sottostante l'opzione. Dal punto di vista finanziario ciò significa che le valutazioni dei prezzo delle opzioni espresse dalla (4.17) sono valutazioni neutrali al rischio.

Si può anche osservare che ogni funzione $F(t, S_t)$ che soddisfa la (4.36) rappresenta il prezzo teorico di un derivato negoziabile nel mercato sul sottostante S. Ciò che specifica il tipo di derivato è il dato terminale che per un'opzione call europea è assegnato dalla (4.33). Viceversa, se una funzione $F(t, S_t)$ non soddisfa l'Equazione (4.36) essa non può rappresentare il prezzo di un derivato negoziabile nel mercato di sottostante S. Se così fosse si presenterebbero opportunità di arbitraggio per gli operatori.

Esempio 258 (Black & Scholes Forward Price) *Un contratto forward su un titolo rischioso S che non paga dividendi dal momento dell'acquisto fino alla maturità T con strike price K è un derivato di sottostante S. Pertanto la valutazione neutrale al rischio $F(t, S_t)$ del forward deve soddisfare l'Equazione (4.36) al variare di $t \in [0, T]$. Inoltre, per la natura stessa dei contratti forward, al tempo $t = T$ deve chiaramente aversi*

$$F(T, S_T) = S_T - K.$$

Combinando questo dato terminale con l'idea di valutazione neutrale al rischio, siamo allora condotti a ipotizzare la valutazione

$$F(t, S_t) = S_t - Ke^{-r(T-t)}, \tag{4.37}$$

al variare di $t \in [0,T]$. *D'altra parte, per la funzione*

$$F(t,x) = x - Ke^{-r(T-t)},$$

si ha

$$\partial_t F(t,x) = -rKe^{-r(T-t)}, \quad \partial_x F(t,x) = 1, \quad \partial_{xx} F(t,x) = 0.$$

e sostituendo tali derivate nella (4.36) otteniamo

$$-rKe^{-r(T-t)} = rF(t,x) - rx$$

ossia la funzione $F(t,x)$ *soddisfa l'equazione di Black & Scholes. Quindi la valutazione ipotizzata ha lo stesso valore terminale del forward e soddisfa l'equazione di Black & Scholes. Se assumiamo che nel mercato non vi siano possibilità di arbitraggio non rimane che concludere che la (**??**) è proprio l'equazione che esprime la valutazione neutrale al rischio per un forward sul sottostante* S. *Conseguenza di quanto appena presentato è che le osservazioni al tempo* $t = 0$ *dei prezzi di mercato* S_0 *di un titolo rischioso ed* F_0 *di un forward con sottostante lo stesso titolo, maturità* T *(rispetto al tempo* $t = 0$*) e strike price* K *permettono una stima del tasso di rendimento non rischioso valutato dal mercato nell'intervallo di tempo* $[0,T]$. *Infatti, dalla (**??**) segue facilmente*

$$r = -\frac{1}{T}\ln\left(\frac{S_0 - F_0}{K}\right). \tag{4.38}$$

L'unico parametro della formula di Black & Scholes che non può essere osservato direttamente è la volatilità del prezzo del sottostante. E' qui opportuno citare un approccio per la stima di tale parametro che utilizza la cosiddetta *volatilità implicita (implied volatility)*. Si tratta della volatilità implicita nel prezzo di mercato delle opzioni.

Per dare un'idea di questo approccio, supponiamo che il valore di una call europea, scritta su un titolo che non paga dividendi, sia pari a \$1.875 quando $S_0 = \$21$, $K = \$20$, $r = 0.1$ e $T = 0.25$. La volatilità implicita è quel valore di σ che, inserito nell'equazione (**??**), ci consente di ottenere un valore teorico dell'opzione pari a quello di mercato ($C(0,S_0) = \$1.875$). Purtroppo, non è possibile invertire l'equazione (**??**) in modo da esprimere σ in funzione di S_0, K, r, T e C_0. Tuttavia, si può utilizzare una procedura iterativa per trovare il σ implicito, che in questo esempio ci restituisce una volatilità pari a 0.235 ossia al 23.5 per cento annuo.

Le volatilità implicite nei prezzi delle opzioni possono essere utilizzate per verificare le opinioni del mercato circa la volatilità attesa per un particolare titolo. Spesso gli analisti calcolano le volatilità implicite nei prezzi delle opzioni, scritte su un certo titolo, che sono più attivamente negoziate per poi calcolare i prezzi di altre opzioni, scritte sullo stesso titolo, ma meno attivamente negoziate.

Finora si è assunto che il sottostante su cui è scritta l'opzione non paghi dividendi. In pratica, questa circostanza non si verifica quasi mai. Grazie ad un breve esempio riguardante un'opzione europea di acquisto, cercheremo ora di estendere il modello di Black & Scholes in modo da tener conto dei dividendi. Assumeremo che i dividendi pagati durante la vita dell'opzione e le relative date di pagamento possano essere previsti con esattezza. Se le opzioni sono di breve durata quest'assunzione non è irragionevole. Se le opzioni sono di lunga durata si assume, in genere, che il tasso al quale vengono pagati i dividendi alle relative scadenze sia noto. Assumeremo, inoltre, che la data di pagamento dei dividendi coincida con la data di stacco. In questa data il prezzo dell'azione si riduce in misura pari al valore del dividendo staccato.

Le opzioni europee possono essere analizzate assumendo che il prezzo dell'azione sia la somma di due componenti: una componente priva di rischio, che verrà utilizzata per pagare i dividendi distribuiti durante la vita dell'opzione, e una componente rischiosa. La componente priva di rischio è pari, in ogni momento, alla somma dei dividendi che verranno pagati durante la vita dell'opzione, attualizzati dalle date di stacco in base al tasso di interesse privo di rischio. Alla data di scadenza dell'opzioni i dividendi saranno già stati pagati e la componente priva di rischio non esisterà più. Pertanto, la formula di Black & Scholes è corretta se S_0 è uguale alla componente rischiosa e σ è la volatilità del processo seguito dalla componente rischiosa. La formula di Black & Scholes può essere quindi utilizzata se dal prezzo dell'azione vengono dedotti i dividendi che verranno distribuiti durante la vita dell'opzione, attualizzati in base al tasso di interesse privo di rischio. Vanno inclusi nei calcoli solo i dividendi il cui stacco avviene durante la vita dell'opzione.

Esempio 259 *Si consideri un'opzione di acquisto europea scritta su un titolo che paga un dividendo tra* 2 *mesi e uno tra* 5 *mesi. Ci si attende che entrambi i dividendi siano pari a* \$0.5*. Il prezzo corrente del titolo è di* \$40*, il prezzo d'esercizio è di* \$40*, la volatilità del prezzo dell'azione è del* 30% *annuo, il tasso di interesse privo di rischio è pari al* 9% *annuo e la vita residua dell'opzione è di* 6 *mesi. Il valore attuale dei dividendi è pari a*

$$\$\left(0.5e^{-0.1667\times 0.09}+0.5e^{-0.4167\times 0.09}\right)=\$0.9741.$$

Il prezzo dell'opzione call può essere calcolato in base alla formula di Black & Scholes ponendo $S_0 = \$\,(40-0.9741) = \$39.0259$*,* $K=\$40$*,* $r=0.09$*,* $\sigma=0.3$ *e* $T=0.5$*, da cui*

$$\begin{aligned} d_{+}(T,S_0) &= \frac{1}{0.3\sqrt{0.5}}\left[\ln\left(\frac{39.0259}{40}\right)+\left(0.09+\frac{0.3^2}{2}\right)\right]=0.2017 \\ d_{-}(T,S_0) &= \frac{1}{0.3\sqrt{0.5}}\left[\ln\left(\frac{39.0259}{40}\right)+\left(0.09-\frac{0.3^2}{2}\right)\right]=-0,0104. \end{aligned}$$

Dunque, il prezzo della call risulta pari a

$$C(0,S_0)=\$\left(39.0259\Phi\left(0.2017\right)-40.e^{-0.09\times 0,5}\Phi\left(-0.0104\right)\right)=\$3.67$$

ossia a \$3.67*.*

4.1.1 Greeks

From the Black and Scholes model we can derive several measures of sensitivity of the option price with respect to changes in the parameters on which the oprion price depends. The name *greeks* comes from the customary use of greek letters to denote them. From Equation (4.17), it is rather evident that we can think on the option price not only as a differentiable function of the time to maturity $T-t\equiv\tau$ and the underlying stock price $S_t\equiv x$, but also of the risk free rate r and the stock price volatility σ. On the other hand, since the strike price K can usually take only a finite number of predetermined values, the sensitivity of the the option price with respect to changes in the strike price K is less interesting and less frequently consider. Formally, we can think that

$$C\equiv C\left(x,\tau,r,\sigma,K\right)\qquad [\text{resp. } P\equiv P\left(x,\tau,r,\sigma,K\right)].$$

This allows to consider the first and second order partial derivatives of the option price with respect to the variables entering the price function, to say the greeks.

First Order Greeks

Definizione 260 *We call the* Delta *of the call [resp. put] option the derivative of the call [resp. put] option price with respect to the underlying asset value. Formally,*

$$\Delta_C\overset{def}{=}\partial C_x\left(x,\tau,r,\sigma,K\right)\qquad [\textit{resp. } \Delta_P\overset{def}{=}\partial P_x\left(x,\tau,r,\sigma,K\right)].$$

The Delta measures the sensitivity of the option price to variation in the underlying asset price. Namely, the rate of change of the option price with respect to the change of the underlying asset price when everything else is unchanged.

Osservazione 261 *We have*

$$\Delta_C - \Delta_P = 1.$$

Definizione 262 *We call the* dual Delta *of the call [resp. put] option the derivative of the call [resp. put] option price with respect to the strike price. Formally,*

$$\Delta'_C \stackrel{def}{=} \partial C_K(x,\tau,r,\sigma,K) \qquad \textit{[resp. } \Delta'_P \stackrel{def}{=} \partial P_K(x,\tau,r,\sigma,K)\textit{].}$$

Definizione 263 *We call the* Theta *of the call [resp. put] option the opposite of the derivative of the call [resp. put] option price with respect to the time to maturity. Formally,*

$$\times_C \stackrel{def}{=} -\partial_\tau C(x,\tau,r,\sigma,K) \qquad \textit{[resp. } \times_P \stackrel{def}{=} -\partial_\tau P(x,\tau,r,\sigma,K)\textit{].}$$

The Theta measures the sensitivity of the price to the time to the maturity. Namely, the rate of change of the option price as the time to maturity passes and everything else is unchanged.

Definizione 264 *We call the* Rho *of the call [resp. put] option the opposite of derivative of the call [resp. put] option price with respect to the risk-free interest rate. Formally,*

$$\rho_C \stackrel{def}{=} -\partial_r C(x,\tau,r,\sigma,K) \qquad \textit{[resp. } \rho_P \stackrel{def}{=} -\partial_r P(x,\tau,r,\sigma,K)\textit{],}$$

The Rho measures the sensitivity of the option to variation of market risk free interest rate. Namely, the rate of change of the option price with respect to the change of the risk free interest rate when everything else is unchanged.

Definizione 265 *We call* Vega *of the call [resp. put] option the derivative of the call [resp. put] option price with respect to the underlying asset volatility. That is,*

$$\mathcal{V}_C \stackrel{def}{=} \partial_\sigma C(x,\tau,r,\sigma,K) \qquad \textit{[resp. } \mathcal{V}_P \stackrel{def}{=} \partial_\sigma P(x,\tau,r,\sigma,K)\textit{].}$$

The Vega measures the sensitivity of the option price to variation of the underlying asset price volatility. Namely, the amount of money per underlying share that the option price will gain [resp. lose] as the asset volatility rise [resp. fall] by 1% and everything else is unchanged.

Definizione 266 *We call* Omega *or* lambda *or* elasticity *of the call [resp. put] option the percentage change in the option price per percentage change in the underlying asset price. That is*

$$\Omega_C \equiv \lambda_C \stackrel{def}{=} \frac{x\partial_x C(x,\tau,r,\sigma,K)}{C(x,\tau,r,\sigma,K)} \qquad \textit{[resp. } \Omega_P \equiv \lambda_P \stackrel{def}{=} \frac{x\partial_x P(x,\tau,r,\sigma,K)}{P(x,\tau,r,\sigma,K)}\textit{].}$$

The Omega is a measure of the *financial leverage* in option investing.

Second Order Greeks

Definizione 267 *We call the* Gamma *of the call [resp. put] option the derivative of the option Delta with respect to the underlying asset price. That is the second derivative of the call [resp. put] option price twice with respect to the underlying asset price. Formally,*

$$\Gamma_C \overset{def}{=} \partial_x \Delta_C (x, \tau, r, \sigma, K) = \partial_{x,x} C (x, \tau, r, \sigma, K) \qquad \textit{[resp. } \Gamma_P \overset{def}{=} \partial_x \Delta_P (x, \tau, r, \sigma, K) = \partial_{x,x} P (x, \tau, r, \sigma, K) \textit{].}$$

The Gamma measures the rate of change of the option Delta with respect to the change of the market price of the underlying, when everything else is unchanged.

Definizione 268 *We call the* dual Gamma *of the call [resp. put] option the derivative of the option dual Delta with respect to the underlying asset price. That is the second derivative of the call [resp. put] option price twice with respect to the strike price. Formally,*

$$\Gamma'_C \overset{def}{=} \partial_K \Delta'_C (x, \tau, r, \sigma, K) = \partial_{K,K} C' (x, \tau, r, \sigma, K) \qquad \textit{[resp. } \Gamma_P \overset{def}{=} \partial_K \Delta'_P (x, \tau, r, \sigma, K) = \partial_{x,x} P (x, \tau, r, \sigma, K) \textit{].}$$

Definizione 269 *We call the* Vanna *or* DdeltaDvol *or* DvegaDspot *of the call [resp. put] option the derivative of the option Delta with respect to the underlying asset price volatility. That is the second derivative of the call [resp. put] option price once with respect to the underlying asset price and once to the underlying asset price volatility. Formally,*

$$\text{Vanna}_C \overset{def}{=} \partial_\sigma \Delta_C (x, \tau, r, \sigma, K) = \partial_{x,\sigma} C (x, \tau, r, \sigma, K) \qquad \textit{[resp. } \text{Vanna}_P \overset{def}{=} \partial_x \Delta_P (x, \tau, r, \sigma, K) = \partial_{x,\sigma} P (x, \tau, r, \sigma, K$$

Definizione 270 *We call the* Charm *or* delta decay *of the call [resp. put] option the opposite derivative of the option Delta with respect to the time to maturity. That is the opposite of the second derivative of the call [resp. put] option price once with respect to the underlying asset price and once to the time to maturity. Formally,*

$$\text{Charm}_C \overset{def}{=} -\partial_\tau \Delta_C (x, \tau, r, \sigma, K) = -\partial_{x,\tau} C (x, \tau, r, \sigma, K) \qquad \textit{[resp. } \text{Charm}_P \overset{def}{=} -\partial_\tau \Delta_P (x, \tau, r, \sigma, K) = -\partial_{x,\tau} P (x,$$

Definizione 271 *We call the* Vomma *or* DvegaDvol *or* vega convexity *of the call [resp. put] option the derivative of the option Vega with respect to the underlying asset price volatility. That is the second derivative of the call [resp. put] option price twice with respect to the underlying asset price volatility. Formally,*

$$\text{Vomma}_C \overset{def}{=} \partial_\sigma \mathcal{V}_C (x, \tau, r, \sigma, K) = \partial_{\sigma,\sigma} C (x, \tau, r, \sigma, K) \qquad \textit{[resp. } \text{Vomma}_P \overset{def}{=} \partial_\sigma \mathcal{V}_C (x, \tau, r, \sigma, K) = \partial_{\sigma,\sigma} P (x, \tau, r, \sigma,$$

Definizione 272 *We call the* Veta *or* DvegaDtime *of the call [resp. put] option the derivative of the option Vega with respect to the time to maturity. That is the second derivative of the call [resp. put] option price once with respect to the underlying asset price volatility and to the time to maturity. Formally,*

$$\text{Veta}_C \overset{def}{=} \partial_\tau \mathcal{V}_C (x, \tau, r, \sigma, K) = \partial_{\sigma,\tau} C (x, \tau, r, \sigma, K) \qquad \textit{[resp. } \text{Veta}_P \overset{def}{=} \partial_\tau \mathcal{V}_P (x, \tau, r, \sigma, K) = \partial_{\sigma,\tau} P (x, \tau, r, \sigma, K) \textit{].}$$

Definizione 273 *We call the* Vera *of the call [resp. put] option the derivative of the option Rho with respect to the asset price volatility. That is the second derivative of the call [resp. put] option price once with respect to the risk free rate and once to the underlying asset price volatility. Formally,*

$$\text{Vera}_C \overset{def}{=} \partial_\sigma \rho_C (x, \tau, r, \sigma, K) = \partial_{r,\sigma} C (x, \tau, r, \sigma, K) \qquad \textit{[resp. } \text{Vera}_P \overset{def}{=} \partial_\sigma \rho_P (x, \tau, r, \sigma, K) = \partial_{r,\sigma} P (x, \tau, r, \sigma, K) \textit{].}$$

4.2 Modello di Prezzo a Tempo Continuo

Il poter disporre di un modello che descriva adeguatamente l'andamento dei prezzi dei titoli un mercato finanziario risulta particolarmente interessante non solo a fini previsionali, ma anche perchè le formule di valutazione degli strumenti finanziari derivati richiedono la conoscenza del processo che regola tale andamento. I prezzi dei titoli, unitamente all'ovvia caratteristica di variare nel tempo, presentano una forte dipendenza da accadimenti economici aleatori. Pertanto il modello matematico più naturale per la loro descrizione consiste nel processo stocastico. In effetti, i modelli finanziari più utilizzati si basano sull'ipotesi che la dinamica dei prezzi dei titoli sia regolata da un processo stocastico. In particolare, si assume spesso che tale processo sia a parametro temporale continuo ed a variabili aleatorie di tipo continuo, per sfruttare le caratteristiche di importanti e matematicamente ben noti processi stocastici quali il processo di Wiener, il moto browniano, il moto browniano geometrico. Abbiamo già discusso l'opportunità di ipotizzare una continuità temporale. Riguardo la continuità in termini di valori, bisogna osservare che in realtà i prezzi dei titoli assumono solo valori discreti, ad esempio in Italia i prezzi sono multipli di 1 millesimo di euro, e nei mercati americani sono multipli di 1/8 di dollaro. Tuttavia, anche in questo caso i benefici matematici che derivano da una rappresentazione continua sono di gran lunga superiori alla perdita di realisticità del modello.

Cercheremo ora di illustrare la nozione di arbitraggio nel contesto dei modelli a tempo continuo, in particolare con riferimento al modello di Black & Scholes per la prezzatura delle opzioni.

Sia $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ uno spazio di probabilità, sia $(W_t)_{t\geqslant 0} \equiv W$ un processo di Wiener unidimensionale standard su Ω, sia $\left(\mathfrak{F}_t^W\right)_{t\geqslant 0}$ la filtrazione da esso generata e sia $(X_t)_{t\geqslant 0}$ un processo stocastico reale su Ω.

Definizione 274 *Diciamo che $(X_t)_{t\geqslant 0}$ è un modello di* prezzo *per un titolo del mercato azionario, o un* moto browniano geometrico*, se*

1. *i* tassi di rendimento *del processo sono indipendenti, cioè l'incremento relativo $(X_t - X_s)/X_s$ del titolo, è indipendente dal σ-algebra $\mathfrak{F}_s^W$;*

2. *i tassi di rendimento del processo sono* stazionari*, cioè la distribuzione dell'incremento relativo $(X_t - X_s)/X_s$ uguaglia la distribuzione dell'incremento relativo $(X_{t-s} - X_0)/X_0$ per tutti gli $s, t \geqslant 0$, con $t > s$;*

3. *le traiettorie del processo sono continue.*

Osserviamo che in un mercato finanziario reale, fino al tempo corrente s ci si trova nella situazione di avere osservato una particolare traiettoria del processo stocastico, rappresentante l'andamento del prezzo del titolo, traiettoria che si è realizzata tra tutte quelle che sarebbero state possibili. Ovviamente le traiettorie dei prezzi futuri sono incerte e la loro realizzazione può essere espressa solo in termini di distribuzione di probabilità. A tale fine occorre calibrare il modello matematico candidato alla loro rappresentazione e per fare ciò è necessario imporre delle restrizioni sulle caratteristiche del modello stesso, in modo da rendere possibile l'applicazione dei metodi di inferenza statistica anche in presenza di una sua sola realizzazione. In particolare, assumere che il processo stocastico modello abbia incrementi indipendenti e normalmente distribuiti rende conto della circostanza che la realizzazione di una specifica traiettoria fino al tempo corrente, piuttosto che un'altra, non fornisce una migliore indicazione sulla probabilità di realizzarsi delle traiettorie future e che il prezzo all'istante corrente s è il migliore predittore statistico del prezzo ad un istante futuro t con una deviazione linearmente dipendente dall'intervallo di tempo $t - s$.

Si consideri, ad esempio, un'azione che arriva oggi a quotare \$10; le previsioni sul prezzo futuro non tengono in alcun conto della storia del prezzo del titolo. In altri termini, non ha alcuna importanza se il titolo quotava ieri \$5 piuttosto che \$15. La sola informazione rilevante è la quotazione odierna. La stessa varierà solo al sopraggiungere di nuove informazioni nel mercato che modificano le aspettative future circa il vero valore del titolo stesso, mentre rimarrà di \$10 in assenza di tali informazioni. L'incertezza sul prezzo futuro viene espressa in termini della distribuzione di probabilità gaussiana di media \$10 e di deviazione proporzionale al trascorrere del tempo.

Queste caratteristiche del modello sono coerenti con la forma debole di efficienza dei mercati. Secondo questa teoria, dovuta a Fama, il prezzo corrente di un'azione racchiude in sè tutta l'informazione resasi disponibile nel mercato fino all'istante corrente ed incorporata nella serie storica dei prezzi. E' la stessa competizione all'interno dei mercati che assicura la validità della forma debole di efficienza. La presenza di moltissimi investitori che osservano il mercato azionario proprio per trarne profitto fa sì che i prezzi delle azioni racchiudano in sè tutta l'informazione disponibile fino al presente. Se la forma debole dell'efficienza dei mercati non fosse valida, i cultori dell'analisi tecnica potrebbero realizzare profitti superiori alla media interpretando i grafici che illustrano la serie passata dei prezzi azionari, ma c'è ben poca evidenza che siano in grado di riuscirci. Per di più, a nostro avviso, l'analisi tecnica è anche poco sensata concettualmente. Ad esempio, se si scoprisse che, dopo una particolare evoluzione dei prezzi nel passato, le probabilità di rialzo di una certa azione sono pari al 65 per cento, non appena quella particolare evoluzione si manifestasse nuovamente gli investitori cercherebbero di comprare il titolo e la domanda aumenterebbe immediatamente. Ne seguirebbe un immediato aumento delle quotazioni tale da eliminare l'effetto osservato.

Riprendiamo adesso l'analisi dei modelli matematici oggetti del nostro interesse.

Il seguente risultato è un'applicazione del fondamentale Teorema di Levy sulla caratterizzazione del moto browniano.

Teorema 275 *Se $(X_t)_{t\geqslant 0}$ è un modello di prezzo per un titolo del mercato azionario, o un moto browniano geometrico, allora è possibile determinare due costanti reali μ e σ, con $\sigma \geqslant 0$, tali che per ogni $t \geqslant 0$ il processo $(X_t)_{t\geqslant 0}$ sia l'unica soluzione dell'equazione differenziale stocastica* (SDE)

$$dX_t = \mu X_t \, dt + \sigma X_t \, dW_t, \tag{4.39}$$

di dato iniziale X_0 assegnato e che quindi si può scrivere

$$S_t = S_0 \exp\left(mt + \sigma W_t\right), \tag{4.40}$$

essendo $m = \mu - \frac{1}{2}\sigma^2$.

$$\mathbf{E}\left[S_t/S_0\right] = \mathbf{E}\left[\exp\left(mt + \sigma W_t\right)\right] = \mathbf{E}\left[e^{mt} e^{\sigma W_t}\right] = e^{mt}\mathbf{E}\left[e^{\sigma W_t}\right] = e^{\left(\mu - \frac{1}{2}\sigma^2\right)t} e^{\frac{1}{2}\sigma^2} = e^{\mu t}$$

and

$$\begin{aligned}\mathbf{D}^2\left[S_t/S_0\right] &= \mathbf{D}^2\left[\exp\left(mt + \sigma W_t\right)\right] = \mathbf{D}^2\left[e^{mt} e^{\sigma W_t}\right] = e^{2mt}\mathbf{D}^2\left[e^{\sigma W_t}\right] = e^{2mt}\left(e^{2\sigma^2 t} - e^{\sigma^2 t}\right) \\ &= e^{2\left(\mu - \frac{1}{2}\sigma^2\right)t}\left(e^{2\sigma^2 t} - e^{\sigma^2 t}\right) = e^{2\mu t} e^{-\sigma^2 t}\left(e^{2\sigma^2 t} - e^{\sigma^2 t}\right) = e^{2\mu t}\left(e^{\sigma^2 t} - 1\right).\end{aligned}$$

since $W_t \sim N\left(0, \sqrt{t}\right)$.

Note that

$$\frac{S_t}{S_0} = \exp\left(mt + \sigma W_t\right)$$

This implies

$$\ln\left(\frac{S_t}{S_0}\right) = mt + \sigma W_t$$

Therefore,

$$\mathbf{E}\left[\ln\left(\frac{S_t}{S_0}\right)\right] = mt = \mu - \tfrac{1}{2}\sigma^2 \quad \text{and} \quad \mathbf{D}^2\left[\ln\left(\frac{S_t}{S_0}\right)\right] = \sigma^2 t.$$

We have also

$$\frac{dS_t}{S_t} = \mu\, dt + \sigma\, dW_t$$

Now

$$\begin{aligned} d\ln(S_t) &= \left(\frac{\partial}{\partial x}\ln(x)\right)_{x=S_t} dS_t + \frac{1}{2}\left(\frac{\partial^2}{\partial x^2}\ln(x)\right)_{x=S_t} d[S]_t \\ &= \left(\frac{1}{x}\right)_{x=S_t} dS_t + \left(-\frac{1}{x^2}\right)_{S_t} \sigma^2 S_t^2 dt, \end{aligned}$$

since, by the Ito table

$$d[S]_t = (\mu S_t\, dt + \sigma S_t\, dW_t)(\mu S_t\, dt + \sigma S_t\, dW_t) = \sigma^2 S_t^2 dt.$$

As a consequence,

$$d\ln(S_t) = \frac{dS_t}{S_t} - \frac{1}{2}\sigma^2 dt$$

This implies

$$d\ln(S_t) + \frac{1}{2}\sigma^2 dt = \frac{dS_t}{S_t} = \mu\, dt + \sigma\, dW_t$$

Integrating

$$\int_0^t d\ln(S_s) + \int_0^t \frac{1}{2}\sigma^2 ds = \int_0^t \frac{dS_s}{S_s} = \int_0^t \mu\, ds + \int_0^t \sigma\, dW_s$$

That is

$$\ln(S_t) - \ln(S_0) + \frac{1}{2}\sigma^2 t = \mu t + \sigma W_t.$$

$$\ln\left(\frac{S_t}{S_0}\right) = \left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W_t$$

and exponentiating

$$\frac{S_t}{S_0} = \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W_t\right),$$

that is

$$S_t = S_0 \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W_t\right).$$

Osservazione 276 *Chiaramente, la variabile aleatoria* $\ln(S_t/S_0)$ *ha distribuzione normale con* $\mathbf{E}[\ln(S_t/S_0)] = mt$ *e* $\mathbf{D}^2[\ln(S_t/S_0)] = \sigma^2 t$*.*

Il moto browniano geometrico spesso è chiamato processo log-normale in ragione del fatto che, per ogni t, il $\log(X_t) = \log(X_0) + mt + \sigma W_t$ è normalmente distribuito. Da notare che se $(X_t)_{t\geqslant 0}$ è un modello di prezzo per un titolo del mercato azionario, allora si ha formalmente

$$\mathbf{E}_t[dX_t] = \mu X_t\, dt \quad \text{e} \quad \mathbf{D}_t^2[dX_t] = \sigma^2 X_t^2\, dt\ ,$$

per cui si può pensare a μ come al tasso di rendimento istantaneo atteso per ogni unità di conto investita nel titolo al tempo t, e si può pensare a σ come alla deviazione istantanea standard dal tasso di rendimento istantaneo atteso. In effetti, con un certo abuso di notazione, molti autori scrivono

$$\mathbf{E}_t[dX_t/X_t] = \mu\, dt \quad \text{e} \quad \mathbf{D}_t^2[dX_t/X_t] = \sigma^2\, dt\ .$$

Definizione 277 *Diciamo che un processo di prezzi per un titolo* $(X_t)_{t\geqslant 0}$ *rappresenta una* valuta *se* $\sigma = 0$*, ed in questo caso il coefficiente* μ *è anche noto come* tasso di interesse composto continuo. *Diciamo che il processo* $(X_t)_{t\geqslant 0}$ *rappresenta un* titolo rischioso *se* $\sigma > 0$*, ed in questo caso i coefficienti* μ *e* σ *sono rispettivamente chiamati* tasso di rendimento istantaneo atteso *e* volatilità *del titolo rischioso.*

Osserviamo che se un processo di prezzi per un titolo $(X_t)_{t\geqslant 0}$ rappresenta una valuta, allora $(X_t)_{t\geqslant 0}$ è soluzione dell'equazione differenziale deterministica

$$dX_t = \mu X_t \, dt,$$

per cui risulta

$$X_t = X_0 \exp(\mu t).$$

Inoltre, il tasso di interesse composto continuo, detto anche tasso a breve, rappresenta il tasso secondo il quale si accumulano i capitali depositati in assenza di rischio.

Parte I

Appendix

Capitolo 5

Constrained Optimization

Let $\mathbb{D}$ an open subset of $\mathbb{R}^N$, let $f \in C^1(\mathbb{D};\mathbb{R})$, let $g_k \in C^1(\mathbb{R}^N;\mathbb{R})$ for $k = 1, \ldots, K$, let $\mathbb{G} \equiv \{\mathbf{x} \in \mathbb{R}^N \mid g_k(\mathbf{x}) = 0, \quad k = 1, \ldots, K\}$, and let $\hat{\mathbf{x}} \in \mathbb{D} \cap \mathbb{G}$.

Definizione 278 *We say that f has a* local minimum *[resp.* maximum*] point at $\hat{\mathbf{x}}$* under the constraint $\mathbb{G}$*, if there exists $r > 0$, such that*

$$f(\mathbf{x}) \geqslant f(\hat{\mathbf{x}}) \quad [res.\ f(\mathbf{x}) \leqslant f(\hat{\mathbf{x}})], \quad \forall \mathbf{x} \in \mathbb{D} \cap \mathbb{G} \cap \mathbb{B}(\hat{\mathbf{x}}; r).$$

Teorema 279 *Assume the function f has a critical point at $\hat{\mathbf{x}}$ under the constraint $\mathbb{G}$ and $rank(J_{\hat{\mathbf{x}}}(g_1, \ldots, g_K)) = K \leqslant N - 1$. Then, introducing the Lagrangian function $L : \mathbb{D} \times \mathbb{R}^K \to \mathbb{R}$ given by*

$$L(\mathbf{x}, \lambda) \stackrel{def}{=} f(\mathbf{x}) - \sum_{k=1}^{K} \lambda_k g_k(\mathbf{x}), \quad (\mathbf{x}, \lambda) \in \mathbb{D} \times \mathbb{R}^K$$

there exists a unique $\hat{\lambda} \equiv (\hat{\lambda}_1, \ldots, \hat{\lambda}_K) \in \mathbb{R}^K$ such that L has a critical point at $(\hat{\mathbf{x}}, \hat{\lambda})$.

Let $\mathbb{D}$ an open subset of $\mathbb{R}^2$, let $f \in C^1(\mathbb{D};\mathbb{R})$, let $h \in C^1(\mathbb{R};\mathbb{R})$, let $\mathbb{H} \equiv \{\mathbf{x} \in \mathbb{R}^2 \mid h(\mathbf{x}) \leqslant 0\}$, and let $\hat{\mathbf{x}} \in \mathbb{D} \cap \mathbb{H}$.

Teorema 280 *Assume the function f has a maximum point at $\hat{\mathbf{x}}$ under the constraint $\mathbb{H}$, that is*

$$f(\hat{\mathbf{x}}) = \max_{h(\mathbf{x}) \leqslant 0} \{f(\mathbf{x}),\ \mathbf{x} \in \mathbb{D}\},$$

and $rank(J_{\hat{\mathbf{x}}}(h)) = 1$. Then, introducing the Lagrangian function $L : \mathbb{D} \times \mathbb{R} \to \mathbb{R}$ given by

$$L(\mathbf{x}, \lambda) \stackrel{def}{=} f(\mathbf{x}) - \mu h(\mathbf{x}), \quad (\mathbf{x}, \mu) \in \mathbb{D} \times \mathbb{R}$$

there exists $\hat{\mu} \in \mathbb{R}$ such that the following conditions are fulfilled

optimality $\partial_{x_n} L(\hat{\mathbf{x}}, \hat{\lambda}) = 0$*, for $n = 1, 2$;*

slackness $\hat{\mu} h(\hat{\mathbf{x}}) = 0$*;*

multiplier feasibility $\hat{\lambda} \geqslant 0$*;*

constrain feasibility $g(\hat{\mathbf{x}}) \leqslant 0$*.*

Let $\mathbb{D}$ an open subset of $\mathbb{R}^N$, let $f \in C^1(\mathbb{D};\mathbb{R})$, let $g_j, h_k \in C^1(\mathbb{R}^N;\mathbb{R})$, for $j = 1, \ldots, J$, $k = 1, \ldots, K$, let $\mathbb{G} \equiv \{\mathbf{x} \in \mathbb{R}^N \mid g_j(\mathbf{x}) = 0, \quad j = 1, \ldots, J\}$, $\mathbb{H} \equiv \{\mathbf{x} \in \mathbb{R}^N \mid h_k(\mathbf{x}) \leqslant 0, \quad k = 1, \ldots, K\}$, and let $\hat{\mathbf{x}} \in \mathbb{D} \cap \mathbb{G} \cap \mathbb{H}$.

Teorema 281 *Assume the function f has a maximum point at $\hat{\mathbf{x}}$ under the constraint $\mathbb{G} \cap \mathbb{H}$, that is*

$$f(\hat{\mathbf{x}}) = \max_{\substack{g_1(\mathbf{x})=0,\dots,g_J(\mathbf{x})=0 \\ h_1(\mathbf{x})\leqslant 0,\dots,h_K(\mathbf{x})\leqslant 0}} \{f(\mathbf{x}),\ \mathbf{x} \in \mathbb{D}\},$$

and

$$rank \begin{pmatrix} \partial_{x_1} g_1(\hat{\mathbf{x}}) & \cdots & \partial_{x_N} g_1(\hat{\mathbf{x}}) \\ \vdots & \ddots & \vdots \\ \partial_{x_1} g_J(\hat{\mathbf{x}}) & \cdots & \partial_{x_N} g_J(\hat{\mathbf{x}}) \\ \partial_{x_1} h_1(\hat{\mathbf{x}}) & \cdots & \partial_{x_N} h_1(\hat{\mathbf{x}}) \\ \vdots & \ddots & \vdots \\ \partial_{x_1} h_K(\hat{\mathbf{x}}) & \cdots & \partial_{x_N} h_K(\hat{\mathbf{x}}) \end{pmatrix} = J + K_b,$$

where K_b is the number of binding inequality constraints. Then, introducing the Lagrangian function $L : \mathbb{D} \times \mathbb{R}^J \times \mathbb{R}^K \to \mathbb{R}$, given by

$$L(\mathbf{x}, \lambda, \mu) = f(\mathbf{x}) - \sum_{j=1}^{J} \lambda_j g_j(\mathbf{x}) - \sum_{k=1}^{K} \mu_k h_k(\mathbf{x}), \quad (\mathbf{x}, \lambda, \mu) \in \mathbb{D} \times \mathbb{R}^J \times \mathbb{R}^K,$$

there exist $(\hat{\lambda}, \hat{\mu}) \in \mathbb{R}^J \times \mathbb{R}^K$, $\hat{\lambda} \equiv (\hat{\lambda}_1, \dots, \hat{\lambda}_J)$, $\hat{\mu} \equiv (\hat{\mu}_1, \dots, \hat{\mu}_K)$, such that the following conditions are fulfilled:

optimality $\partial_{x_n} L(\hat{\mathbf{x}}, \hat{\lambda}, \hat{\mu}) = 0$, *for* $n = 1, \dots, N$*;*

I feasibility $g_j(\hat{\mathbf{x}}) = 0$, *for* $j = 1, \dots, J$*;*

II feasibility $h_k(\hat{\mathbf{x}}) \leqslant 0$, *for* $k = 1, \dots, K$*;*

non negativity $\hat{\mu}_k \geqslant 0$, *for* $k = 1, \dots, K$*;*

slackness $\hat{\mu}_k h_k(\hat{\mathbf{x}}) = 0$, *for* $k = 1, \dots, K$.

Osservazione 282 *The optimality condition 281 can be reformulated in a vector form as*

$$\nabla f(\hat{\mathbf{x}}) = \sum_{j=1}^{J} \hat{\lambda}_j \nabla g_j(\hat{\mathbf{x}}) + \sum_{k=1}^{K} \hat{\mu}_k \nabla h_k(\hat{\mathbf{x}}),$$

which highlights the circumstance that the gradient of the objective function f computed at a candidate maximum point has to be a linear combination of the gradients of the constraining functions.

In economic and financial applications, it is customary to present the inequality constraints in a minimization problem in the form

$$h_1(\mathbf{x}) \geqslant 0, \dots, h_K(\mathbf{x}) \geqslant 0,$$

instead of

$$h_1(\mathbf{x}) \leqslant 0, \dots, h_K(\mathbf{x}) \leqslant 0,$$

That is

$$\mathbb{H} \equiv \{\mathbf{x} \in \mathbb{R}^N \mid h_k(\mathbf{x}) \geqslant 0, \quad k = 1, \dots, K\}.$$

In light of this, the analogue of Theorem 281 for a minimization problem takes the following formulation.

Teorema 283 *Assume the function f has a minimum point at $\check{\mathbf{x}}$ under the constraint $\mathbb{G} \cap \mathbb{H}$, that is*

$$f(\check{\mathbf{x}}) = \min_{\substack{g_1(\mathbf{x})=0,\ldots,g_J(\mathbf{x})=0 \\ h_1(\mathbf{x})\geqslant 0,\ldots,h_K(\mathbf{x})\geqslant 0}} \{f(\mathbf{x}),\ \mathbf{x} \in \mathbb{D}\},$$

and

$$\operatorname{rank} \begin{pmatrix} \partial_{x_1} g_1(\check{\mathbf{x}}) & \cdots & \partial_{x_N} g_1(\check{\mathbf{x}}) \\ \vdots & \ddots & \vdots \\ \partial_{x_1} g_J(\check{\mathbf{x}}) & \cdots & \partial_{x_N} g_J(\check{\mathbf{x}}) \\ \partial_{x_1} h_1(\check{\mathbf{x}}) & \cdots & \partial_{x_N} h_1(\check{\mathbf{x}}) \\ \vdots & \ddots & \vdots \\ \partial_{x_1} h_K(\check{\mathbf{x}}) & \cdots & \partial_{x_N} h_K(\check{\mathbf{x}}) \end{pmatrix} = J + K_b,$$

where K_b is the number of binding inequality constraints. Then, introducing the Lagrangian function $L : \mathbb{D} \times \mathbb{R}^J \times \mathbb{R}^K \to \mathbb{R}$, given by

$$L(\mathbf{x}, \lambda, \mu) = f(\mathbf{x}) - \sum_{j=1}^{J} \lambda_j g_j(\mathbf{x}) - \sum_{k=1}^{K} \mu_k h_k(\mathbf{x}), \quad \forall (\mathbf{x}, \lambda, \mu) \in \mathbb{D} \times \mathbb{R}^J \times \mathbb{R}^K,$$

there exist $(\check{\lambda}, \check{\mu}) \in \mathbb{R}^J \times \mathbb{R}^K$, $\check{\lambda} \equiv (\check{\lambda}_1, \ldots, \check{\lambda}_J)$, $\check{\mu} \equiv (\check{\mu}_1, \ldots, \check{\mu}_K)$, such that the following conditions are fulfilled:

optimality $\partial_{x_n} L(\check{\mathbf{x}}, \check{\lambda}, \check{\mu}) = 0$, *for* $n = 1, \ldots, N$;

I feasibility $g_j(\check{\mathbf{x}}) = 0$, *for* $j = 1, \ldots, J$;

II feasibility $h_k(\check{\mathbf{x}}) \geqslant 0$, *for* $k = 1, \ldots, K$;

non negativity $\check{\mu}_k \geqslant 0$, *for* $k = 1, \ldots, K$;

slackness $\check{\mu}_k h_k(\check{\mathbf{x}}) = 0$, *for* $k = 1, \ldots, K$.

Capitolo 6

Stochastic Processes

The goal of the theory of *stochastic processes* is to build and study mathematical models of real-world phenomena evolving in time under the relevant influence of some random perturbation, the so called *stochastic phenomena*. For instance, the time evolution of many variables in Economics and Finance such as gross domestic product of nations, employment rates, commodity prices, bond prices, stock prices, derivative prices, interest rates, exchange rates, is clearly affected by the frequent occurrence of random events. Hence, these variables describe features of some underlying stochastic phenomenon and advocate the application of stochastic processes to study their time evolution. Roughly speaking, a stochastic process is a temporally indexed family of random variables, defined on the same probability space and taking values in some measurable space. We will make this notion more precise in what follows.

6.1 Basic Definitions and Notations

Let $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ be a *complete* probability space[1], let $\mathbb{T}$ be a nonempty subset of the Euclidean real line $\mathbb{R}$, and let $(\mathbb{X}, \mathcal{M}) \equiv \mathbb{X}$ be a measurable space.

Definizione 284 *A map $X : \mathbb{T} \times \Omega \to \mathbb{X}$ is said to be a* stochastic process *on Ω with* time set $\mathbb{T}$ *and* state space $\mathbb{X}$*, if the t-partial map $X_t : \Omega \to \mathbb{X}$ given by*

$$X_t(\omega) \overset{def}{=} X(t, \omega), \quad \forall \omega \in \Omega,$$

is a random variable, for every $t \in \mathbb{T}$.

When the time set $\mathbb{T}$ is a subset of $\mathbb{Z}$ [resp. an interval of $\mathbb{R}$], we speak of *discrete-time* [resp. *continuous-time*] stochastic process. In case of discrete-time stochastic process, we typically assume $\mathbb{T} \equiv \{1, \ldots, T\}$, for some $T \in \mathbb{N}$, or $\mathbb{T} \equiv \mathbb{N}$ or even $\mathbb{T} \equiv \mathbb{Z}$. We also assume $\mathbb{T} \equiv \{0, 1, \ldots, T\}$ or $\mathbb{T} \equiv \hat{n}$ when we want stress that something special occurs at the initial time $t = 0$. In case of continuous-time stochastic process, we typically assume $\mathbb{T} \equiv [0, T]$, for some $T \in \mathbb{R}_{++}$, or $\mathbb{T} \equiv [0, +\infty)$. Either $\mathbb{T}$ is a subset of $\mathbb{Z}$ or an interval of $\mathbb{R}$, it is naturally equipped with the subspace topology. This is the discrete topology or the Euclidean topology according to whether $\mathbb{T} \subseteq \mathbb{Z}$ or $\mathbb{T} \subseteq \mathbb{R}$. As a consequence, $\mathbb{T}$ becomes a measurable space if equipped with the Borel σ-algebra $\mathcal{B}(\mathbb{T})$. Recall that in case $\mathbb{T} \subseteq \mathbb{Z}$ the Borel σ-algebra $\mathcal{B}(\mathbb{T})$ concides with the discrete σ-algebra $\mathcal{P}(\mathbb{T})$. This measurable structure on $\mathbb{T}$

[1] Introduce the family of all *negligible* events of the probability space Ω, that is

$$\mathcal{N} \equiv \{N \subseteq \Omega : N \subseteq E,\ E \in \mathcal{E},\ \mathbf{P}(E) = 0\}.$$

We say that Ω is *complete* if $\mathcal{N} \subseteq \mathcal{E}$. Any probability space can be completed. Therefore, we will speak of probability spaces in the sense of *complete* probability spaces.

naturally leads to the measurable structure $\mathcal{B}(\mathbb{T}) \otimes \mathcal{E}$ on $\mathbb{T} \times \Omega$. The latter plays a role to introduce several properties of stochastic processes.

The measurable space $\mathbb{X}$ is usually a complete separable metrizable space (*Polish space*), or a locally compact space with countable base (*LCCB space*). In these cases, the σ-algebra $\mathcal{M}$ is assumed to be the Borel σ-algebra $\mathcal{B}(\mathbb{X})$, so that $\mathbb{X}$ is referred to as a *Borel measurable topological space.* To produce interesting results, it is often necessary to assign a measure $\mu : \mathcal{M} \to \bar{\mathbb{R}}_+$ on $\mathbb{X}$. When $\mathbb{X} \equiv \mathbb{R}$ equipped with the Lebesgue-Borel measure $\mu_L : \mathcal{B}(\mathbb{R}) \to \bar{\mathbb{R}}_+$, we speak of *real* stochastic process. When $\mathbb{X} \equiv \mathbb{R}^N$, for some $N \in \mathbb{N}$, equipped with the Lebesgue-Borel measure $\mu_{L^N} : \mathcal{B}\left(\mathbb{R}^N\right) \to \bar{\mathbb{R}}_+$, we speak of *$N$-dimensional* or *$N$-variate real* stochastic process.

Notation 285 *It is customary to identify a stochastic process $X : \mathbb{T}\times\Omega \to \mathbb{X}$ with the family $(X_t)_{t\in\mathbb{T}} \equiv X$ of the t-partial maps and speak of the t-partial map as the* random variable *in the process at time (index) t. In what follows, we will also adopt this convention whenever the measurable structure of the cartesian product $\mathbb{T} \times \Omega$ is not involved.*

Let $(X_t)_{t\in\mathbb{T}} \equiv X$ be a real stochastic process on a probability space $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$.

Esempio 286 (Dirac process) *We call X a* Dirac process *if X_t is a Dirac random variable concentrated at some $x_t \in \mathbb{R}$, for every $t \in \mathbb{T}$, that is,*

$$X_t = x_t, \quad \mathbf{P}(X_t = x_t) = 1, \quad \forall t \in \mathbb{T},$$

where $x_t \in \mathbb{R}$ may vary on varying of $t \in \mathbb{T}$.

Esempio 287 (independent Bernoulli process) *Fixed any $p \in (0,1)$, we call X an* independent Bernoulli process *with success probability p if:*

1. *X_t is a standard Bernoulli random variable with success probability parameter p, for every $t \in \mathbb{T}$, that is,*
$$X_t = \begin{cases} 1 & \mathbf{P}(X_t = 1) = p \\ 0 & \mathbf{P}(X_t = 0) = 1 - p \end{cases} \qquad \forall t \in \mathbb{T},$$

2. *the random variables in the process X are (totally) independent.*

Esempio 288 (independent Rademacher process) *We call X an* independent Rademacher process *with success probability p if:*

1. *X_t is a standard Rademacher random variable, for every $t \in \mathbb{T}$, that is,*
$$X_t = \begin{cases} 1 & \mathbf{P}(X_t = 1) = 1/2 \\ -1 & \mathbf{P}(X_t = 0) = 1/2 \end{cases} \qquad \forall t \in \mathbb{T},$$

2. *the random variables in the process X are are independent.*

Esempio 289 (independent binomial process) *Fixed any $n \in \mathbb{N}$ and $p \in (0,1)$, we call X an* independent Binomial process *with number of trials n and success probability p if:*

1. *X_t is a standard Binomial random variable with number of trials [resp. success probability] parameter n [resp. p], for every $t \in \mathbb{T}$, that is,*
$$X_t = k, \quad \mathbf{P}(X_t = k) = \binom{n}{k}, \quad \forall k = 0, 1, \ldots, n, \quad t \in \mathbb{T};$$

2. *the random variables in the process X are independent.*

Esempio 290 (independent Poisson process) *Fixed any $\lambda \in \mathbb{R}_{++}$, we call X an* independent Poisson process *with rate λ if:*

1. *X_t is a standard Poisson random variable with rate parameters λ, for every $t \in \mathbb{T}$, that is,*

$$X_t = n, \quad \mathbf{P}(X_t = n) = e^{-\lambda}\frac{\lambda^n}{n!}, \quad \forall n \in \hat{n}, \quad t \in \mathbb{T};$$

2. *the random variables in the process X are independent.*

Esempio 291 (independent Gaussian process) *Fixed any $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}_{++}$, we call X an* independent Gaussian *process with mean μ and variance σ^2 if:*

1. *X_t is a Gaussian random variable with mean [resp. variance] parameter μ [resp. σ^2], that is, X_t has a density $f_{X_t} : \mathbb{R} \to \mathbb{R}_+$ given by*

$$f_{X_t}(x) \overset{def}{=} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad \forall x \in \mathbb{R};$$

2. *the random variables in the process X are independent.*

A less trivial example of stochastic process it the so called *random walk*.

Esempio 292 (random walk) *Let $(Z_n)_{n\geqslant 1} \equiv Z$ be a sequence of independent and identically distributed random variables on a probability space Ω with states in $\mathbb{R}^N$, for some $N \in \mathbb{N}$, and let X_0 be a N-variate random variable which is independent of the random variables in Z. We call a* random walk *starting from X_0 with* innovation *Z the stochastic process $(X_t)_{t\in\hat{n}} \equiv X$ given by*

$$X_n \overset{def}{=} X_{n-1} + Z_n, \quad \forall n \in \mathbb{N}.$$

Note that Z is itself a stochastic process with time set $\mathbb{N}$ and state space $\mathbb{R}^N$. The random variable X_0 is often a Dirac random variable centered at x_0, for some $x_0 \in \mathbb{R}^N$. For instance, if Z is an independent Bernoulli [resp. Rademacher, Gaussian] process, the random walk X, referred to as Bernoulli *[resp.* Gaussian*]* random walk *has many important applications in modeling.*

Let $(X_t)_{t\in\mathbb{T}} \equiv X$ be a stochastic process on Ω with state space $\mathbb{X}$.

Definizione 293 *Given any sample point $\omega \in \Omega$, we call the ω-*path *or ω-*trajectory *or ω-*realization *of the process X the map $\omega_X : \mathbb{T} \to \mathbb{X}$ given by*

$$\omega_X(t) \overset{def}{=} X_t(\omega), \quad \forall t \in \mathbb{T}.$$

Assume that the time set $\mathbb{T}$ is an interval of $\mathbb{R}_+$ and let $\mathbb{X}$ be a Borel measurable topological space.

Le traiettorie del processo aleatorio sono tante quanti sono i possibili punti dello spazio campionario Ω. Fissato un istante corrente $\hat{t} \in \mathbb{T}$ l'evoluzione passata del fenomeno stocastico, ossia l'evoluzione dall'istante iniziale fino all'istante corrente $\hat{t}$ è manifesta ed è rappresentabile da una specifica traiettoria del processo. Invece, a priori non si conosce cosa accadrà successivamente all'istante corrente $\hat{t}$, per cui l'evoluzione futura del processo è rappresentabile come la famiglia di tutte le possibili traiettorie che si diramano da $X_{\hat{t}}(\Omega)$, ciascuna con una sua probabilità di realizzazione. I grafici osservabili nella vita quotidiana con cui vengono descritti dei fenomeni stocastici, sono in realtà le rappresentazioni della traiettoria del fenomeno che si è realizzata da un passato più o meno remoto fino all'istante corrente. Esaustive rappresentazioni grafiche dell'evoluzione futura di un processo stocastico non sono in generale possibili, con la notevole eccezione dei celebri diagrammi ad albero dei processi a numero finito di stati.

Definizione 294 *We say that the process X is* continuous *if the sample paths of the process are almost surely continuous, that is, there exist a neglegible event $E \in \mathcal{E}$ such that the ω-sample path $\omega_X : \mathbb{T} \to \mathbb{X}$ is continuous for every $\omega \in \Omega - E$.*

Definizione 295 *We say that the process X is* right-continuous *[resp.* left-continuous*] if the sample paths of the process are almost surely right-continuous [resp. left-continuous]*[2]*.*

Definizione 296 *We say that the process X is* rcll *or* cadlag[3] *[resp.* lcrl *or* caglad[4]*], if the sample paths of the process are almost surely right-continuous with left-hand limits [resp. left-continuous with right-hand limits]*[5]*.*

6.2 Stochastic Processes and Time Series

Let $T \in \mathbb{N}$ and, for some $N \in \mathbb{N}$, let $\mathbb{R}^N$ be the Euclidean N-dimensional real space.

Definizione 297 *We call an N-*variate real time series *of lenght T any sequence $(x_t)_{t=1}^T$ of points in $\mathbb{R}^N$.*

Let $(x_t)_{t=1}^T \equiv x$ be an N-variate real time series of lenght T and let $(X_t)_{t \in \mathbb{T}} \equiv X$ be a stochastic process on a probability space Ω with time set $\mathbb{T}$, discrete or continuous, and state space $\mathbb{R}^N$.

Definizione 298 *We say that the stochastic process X is a* model *for x if $\{1, \ldots, T\} \subseteq \mathbb{T}$ and the time series x may be thought as the restriction to $\{1, \ldots, T\}$ of a sample path of the process, that is*

$$x_t = \omega_X(t),$$

for some $\omega \in \Omega$ and every $t \in \{1, \ldots, T\}$.

The *Time Series Analysis* is the collection of methods and techniques which allow to infer the structure of stochastic process which is a "good model" for a time series. There are several reasons to determine a "good model". Without a "good model" the *interpretation*[6] and the *forecasting* [7] of a time series cannot be pursued. Not to mention more sophisticated tasks such as the *control*[8], the *simulation*[9] and the *hypothesis testing*[10]. While inferring a model for a time series is rather easy[11], inferring a "good model" might be a very difficult task. This difficulty is due to the fact the inference has to be based on the analysis of few realizations, typically a single one of the stochastic process which we aim to candidate as a "good model".

---

[2] The notion of almost surely right-continuous [resp. left-continuous] paths has to be intended in the same sense of Definition 294.

[3] The word *cadlag* is a French acronym for *continu à droite avec des limites à gauche.*

[4] The word *caglad* is a French acronym for *continu à gauche avec des limites à droite.*

[5] The notion of almost surely rcll [resp. lcrl] paths has to be intended in the same sense of Definition 294.

[6] That is the specification of the role played by various variables in the evolution of the pattern.

[7] That is the statistical prediction of the future evolution of the pattern.

[8] That is the statistical prediction of the influence that a policy on some variables might exert on the evolution of the future pattern. For instance, the effect that the monetary policy of a central bank might exert on the economy growth.

[9] That is the statistical description of future scenarios related to the evolution of pattern. For instance the prediction of major earthquakes following a pattern of seismic activity.

[10] That is the statistical reliability of some conjectures. For instance, the confirmation or refutation of the global warming conjecture.

[11] Any polynomial $P : \mathbb{T} \to \mathbb{R}^N$ of degree not lower than T such that

$$P(t) = x_t$$

for every $t \in \{1, \ldots, T\}$, is a model for the time series $(x_t)_{t=1}^T$.

6.3 Filtrations

Let $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ be a complete probability space, let $\mathbb{T}$ be a time set (nonempty subset of the Euclidean real line $\mathbb{R}$), and let $(\mathcal{F}_t)_{t\in\mathbb{T}} \equiv \mathfrak{F}$ be a family of sub-σ-algebras of $\mathcal{E}$ indexed on $\mathbb{T}$.

Definizione 299 *We say that $\mathfrak{F}$ a filtration on Ω if we have*

$$\mathcal{F}_s \subseteq \mathcal{F}_t, \quad \forall s, t \in \mathbb{T} \text{ s.t. } s < t.$$

Let $(\mathcal{F}_t)_{t\in\mathbb{T}} \equiv \mathfrak{F}$ a filtration on Ω.

Definizione 300 *We set*

$$\mathcal{F}_\infty \equiv \bigvee_{t\in\mathbb{T}} \mathcal{F}_t \equiv \sigma\left(\bigcup\nolimits_{t\in\mathbb{T}} \mathcal{F}_t\right).$$

Osservazione 301 *We clearly have $\mathcal{F}_\infty \subseteq \mathcal{E}$. In addition, if $\mathbb{T} \equiv [0, T]$, for some $T > 0$, we have $\mathcal{F}_\infty = \mathcal{F}_T$.*

Definizione 302 *If $\mathbb{T}$ has a first element, say 0, we say that $\mathfrak{F}$ is* complete *if $\mathcal{F}_0$ contains all neglegible events of $\mathcal{E}$.*

Let $(X_t)_{t\in\mathbb{T}} \equiv X$ be a stochastic process on Ω with states in a measurable space $(\mathbb{X}, \mathcal{M}) \equiv \mathbb{X}$ and let $(\mathcal{F}_t)_{t\in\mathbb{T}} \equiv \mathfrak{F}$ be a filtration on Ω.

Definizione 303 *We say that the process X is $\mathfrak{F}$-adapted if the random variable X_t of the process is $\mathcal{F}_t$-measurable, for every $t \in \mathbb{T}$, that is*

$$\{X_t \in M\} \in \mathcal{F}_t,$$

for any $M \in \mathcal{M}$ on varying of $t \in \mathbb{T}$.

Definizione 304 *We call the* filtration generated *by the process X the family $\left(\mathcal{F}_t^X\right)_{t\in\mathbb{T}} \equiv \mathfrak{F}^X$ of sub-σ-algebras of $\mathcal{E}$ given by*

$$\mathcal{F}_t^X \overset{def}{=} \sigma\left(X_s;\ s \leqslant t\right), \quad \forall t \in \mathbb{T},$$

where $\sigma(X_s;\ s \leqslant t)$ is the σ-algebra generated by the random variables X_s of the process as s varies in $\mathbb{T}$, up to and including t.

Osservazione 305 *Any process X is $\mathfrak{F}^X$-adapted. Eventually, $\mathfrak{F}^X$ is the smallest filtration with respect to which the process X is adapted.*

The notion of filtration aims to model the information flow progressively provided by a stochastic phenomenon to an observer with persistent memory[12]. Such an information flow is made of all events which can be discriminated or questions which can be answered by the observer at the current time. Hence, each sub-σ-algebra $\mathcal{F}_t$ of a filtration $(\mathcal{F}_t)_{t\in\mathbb{T}}$, represents the information available to the observer up to and including the time t. On the other hand, each t-partial map X_t of a stochastic process represents a quantitative observation, that is a measurement, on the stochastic phenomenon made by the observer at the time t. Therefore, the notion of adaptness expresses the possibility of making the measurement X_t in light of the avilable information $\mathcal{F}_t$. In particular, the notion of filtration generated by a stochastic process aims to model the minimum information flow which has to be progressively available to an observer to make the measurements represented by the stochastic process.

[12]We assume that the observer does not forget the past.

Definizione 306 *For any $t \in \mathbb{T}$ write $\mathbb{T}_{t]} \equiv \{s \in \mathbb{T} : s \leqslant t\}$. We say that the process X is* progressively measurable *with respect to the filtration $\mathfrak{F}$ if the restriction of the map X to $\mathbb{T}_{t]} \times \Omega$ is $\mathcal{B}(\mathbb{T}_{t]}) \otimes \mathcal{F}_t$ measurable, for every $t \in \mathbb{T}$, that is,*

$$\left\{X_{|\mathbb{T}_{t]} \times \Omega} \in M\right\} \in \mathcal{B}(\mathbb{T}_{t]}) \otimes \mathcal{F}_t, \quad \forall M \in \mathcal{M},$$

for every $t \in \mathbb{T}$. Recall that the restriction of the map X to $\mathbb{T}_{t]} \times \Omega$ is given by

$$X_{|\mathbb{T}_{t]} \times \Omega}(s, \omega) \stackrel{def}{=} X(s, \omega), \quad \forall (s, \omega) \in \mathbb{T}_{t]} \times \Omega.$$

Osservazione 307 *Any progressively measurable process is measurable.*

Assume that the set $\mathbb{T}$ is an interval of $\mathbb{R}_+$.

Definizione 308 *We call the $\mathfrak{F}$-predictable σ-algebra and denote it by $\mathcal{P}(\mathfrak{F})$ the σ-algebra generated by all $\mathfrak{F}$-adapted left continuous stochastic processes, that is,*

$$\mathcal{P}(\mathfrak{F}) \stackrel{def}{=} \bigvee_{X \in \mathrm{X}_{l.c.}(\mathfrak{F})} \mathcal{F}_\infty(X) \equiv \sigma\left(\bigcup_{X \in \mathrm{X}_{l.c.}(\mathfrak{F})} \mathcal{F}_\infty(X)\right),$$

where we write $\mathrm{X}_{l.c.}(\mathfrak{F})$ for the set of all $\mathfrak{F}$-adapted left continuos stochastic processes.

Definizione 309 *We say that the process X is* $\mathfrak{F}$-predictable *if X is $(\mathcal{B}(\mathbb{T}) \otimes \mathcal{P}(\mathfrak{F}), \mathcal{M})$-measurable.*

Osservazione 310 *Any $\mathfrak{F}$-adapted left-continuous process, in particular any continuous process, is $\mathfrak{F}$-predictable.*

Assume that the set $\mathbb{T}$ is a subset of $\mathbb{Z}$. In this case the definition of predictable process changes significantly.

Definizione 311 *In case the time set $\mathbb{T}$ is discrete we say that X is* $\mathfrak{F}$-predictable *if the random variable X_t of the process is $\mathcal{F}_{t-1}$-measurable, for every $t \in \mathbb{T}$ such that $t-1 \in \mathbb{T}$, that is,*

$$\{X_t \in M\} \in \mathcal{F}_{t-1},$$

for any $M \in \mathcal{M}$ on varying of $t \in \mathbb{T}$ such that $t-1 \in \mathbb{T}$. In addition, if $\mathbb{T}$ has a first element, say 0, *we require that X_0 is $\mathcal{F}_0$ measurable.*

6.4 Consistent Families of Finite-Dimensional Distributions

Let $(X_t)_{t \in \mathbb{T}} \equiv X$ a stochastic process on a probability space $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ with states in a measurable space $(\mathbb{X}, \mathcal{M}) \equiv \mathbb{X}$.

Definizione 312 *For any $L \in \mathbb{N}$, we call a* time multi-index *of lenght L any sequence $(t_\ell)_{\ell=1}^L$ of L distinct elements of $\mathbb{T}$. A multi-index $(t_\ell)_{\ell=1}^L$ is said to be* increasing *if $t_\ell < t_{\ell+1}$ for every $\ell = 1, \ldots, L-1$.*

Osservazione 313 *Any time multi-index [resp. increasing multi-index] can be identified with a finite permutation [resp. combination] of the elements of $\mathbb{T}$.*

Notation 314 *We write $\mathcal{P}_{fin}(\mathbb{T})$ [resp. $\mathcal{C}_{fin}(\mathbb{T})$] for the set of all time multi-indices [resp. increasing time multi-indices].*

Definizione 315 *For any* $t \in \mathcal{P}_{fin}(\mathbb{T})$, $t \equiv (t_\ell)_{\ell=1}^L$, *we call* t-marginal distribution *of* X *the map* $P_{X;t} : \otimes_{\ell=1}^L \mathcal{M} \to \mathbb{R}_+$, *where* $\otimes_{\ell=1}^L \mathcal{M}$ *is the* tensor product *σ-algebra*[13] *on* $\mathbb{X}^L$, *given by*

$$P_{X;t}(M) \overset{def}{=} \mathbf{P}\{(X_{t_1}, \dots, X_{t_L}) \in M\}, \quad \forall M \in \otimes_{\ell=1}^L \mathcal{M}. \tag{6.1}$$

In particular, if we assume that $\mathbb{T}$ *has a first element, say* 0, *we call the* initial distribution *of* X *the distribution of the random variable* X_0. *Namely, the map* $P_{X_0} : M \to \mathbb{R}_+$ *given by*

$$P_{X_0}(M) \overset{def}{=} \mathbf{P}(X_0 \in M), \quad \forall M \in \mathcal{M}. \tag{6.2}$$

Proposizione 316 *Given any* $t \in \mathcal{P}_{fin}(\mathbb{T})$, $t \equiv (t_\ell)_{\ell=1}^L$, *the t-marginal distribution of X is a probability measure on* $\mathbb{X}^L$, *which is also known as the* joint distribution *of the random variables* $X_{t_1}, \dots, X_{t_L}$.

Lemma 317 *Assume* $\mathbb{X} \equiv \mathbb{R}^N$, *for some* $N \in \mathbb{N}$. *We have*

$$\mathbf{P}(X_{t_1} \leqslant x_1, \dots, X_{t_L} \leqslant x_L) = \mathbf{P}\left(X_{t_{\pi(1)}} \leqslant x_{\pi(1)}, \dots, X_{t_{\pi(L)}} \leqslant x_{\pi(L)}\right) \tag{6.3}$$

for every $t \in \mathcal{P}_{fin}(\mathbb{T})$, $t \equiv (t_\ell)_{\ell=1}^L$, *every permutation* π *on the set* $\{1, \dots, L\}$, *and every* $(x_1, \dots, x_L) \in \times_{\ell=1}^L \mathbb{R}^N$. *In addition,*

$$\lim_{x_L \to +\infty} \cdots \lim_{x_{K+1} \to +\infty} \mathbf{P}\left(X_{t_1} \leqslant x_1, \dots, X_{t_K} \leqslant x_K, X_{t_{K+1}} \leqslant x_{K+1}, \dots, X_{t_L} \leqslant x_L\right) = \mathbf{P}(X_{t_1} \leqslant x_1, \dots, X_{t_K} \leqslant x_K), \tag{6.4}$$

for every $K \in \mathbb{N}$ *such that* $K \leqslant L - 1$, *and every* $(x_1, \dots, x_K) \in \times_{k=1}^K \mathbb{R}^N$.

Proof. See (**?**). □

Let $\{P_t\}_{t \in \mathcal{P}_{fin}(\mathbb{T})}$ be a family of set functions such that $P_t : \otimes_{\ell=1}^L \mathcal{M} \to \mathbb{R}_+$ for every $t \in \mathcal{P}_{fin}(\mathbb{T})$, $t \equiv (t_\ell)_{\ell=1}^L$.

Definizione 318 *We say that* $\{P_t\}_{t \in \mathcal{P}_{fin}(\mathbb{T})}$ *is a* consistent family of finite-dimensional probability distributions *on* $\mathbb{X}$ *if:*

1. *the set function* $P_t : \otimes_{\ell=1}^L \mathcal{M} \to \mathbb{R}_+$ *is a probability distribution, for every* $t \in \mathcal{P}_{fin}(\mathbb{T})$, $t \equiv (t_\ell)_{\ell=1}^L$;
2. *we have*
$$P_{(t_\ell)_{\ell=1}^L}(M_1 \times \cdots \times M_L) = P_{(t_{\pi(\ell)})_{\ell=1}^L}\left(M_{\pi(1)} \times \cdots \times M_{\pi(\ell)}\right) \tag{6.5}$$
for every $t \in \mathcal{P}_{fin}(\mathbb{T})$, $t \equiv (t_\ell)_{\ell=1}^L$, *every permutation* π *of* $\{1, \dots, L\}$, *and all* $M_1, \dots, M_L \in \mathcal{M}$;
3. *we have*
$$P_{(t_\ell)_{\ell=1}^L}\left(\underbrace{M_1 \times \cdots \times M_{k-1} \times \mathbb{X} \times M_{k+1} \times \cdots \times M_L}_{k}\right) = P_{(t_\ell)_{\ell=1, \ell \neq k}^L}(M_1 \times \cdots \times M_{k-1} \times M_{k+1} \times \cdots \times M_L), \tag{6.6}$$
for every $t \in \mathcal{P}_{fin}(\mathbb{T})$, $t \equiv (t_\ell)_{\ell=1}^L$, *all* $M_1, \dots, M_{k-1}, M_{k+1}, \dots, M_L \in \mathcal{M}$, *and any* $k = 1, \dots, L$, *where* $(t_\ell)_{\ell=1, \ell \neq k}^L$ *is the time-multiindex obtained by suppressing the kth term of* $(t_\ell)_{\ell=1}^L$.

[13] Thi is the σ-algebra on $\mathbb{X}^L$ generated by the family of all sets of the form $\times_{\ell=1}^L M_k$, where $M_k \in \mathcal{M}$, for every $\ell = 1, \dots, L$. In symbols

$$\otimes_{\ell=1}^L \mathcal{M} = \sigma\left(\left\{\bigtimes_{\ell=1}^L M_\ell, : M_\ell \in \mathcal{M}\ \forall \ell = 1, \dots, L\right\}\right).$$

Osservazione 319 *Let* $\{P_t^X\}_{t\in\mathcal{P}_{fin}(\mathbb{T})}$ *be the family of the marginal distributions of the stochastic process* X. *Then* $\{P_t^X\}_{t\in\mathcal{P}_{fin}(\mathbb{T})}$ *is a* consistent family of finite-dimensional probability distributions *on* $\mathbb{X}$.

Of the utmost importance for applications is the following theorem

Teorema 320 (Daniell-Kolmogorov) *Given any consistent family* $\{P_t\}_{t\in\mathcal{P}_{fin}(\mathbb{T})}$ *of finite-dimensional distributions on a measurable space* $\mathbb{X}$, *there exist a probability space* Ω *and a stochastic process* $(X_t)_{t\in\mathbb{T}} \equiv X$ *on* Ω *with state space* $\mathbb{X}$ *such that* $\{P_t\}_{t\in\mathcal{P}_{fin}(\mathbb{T})}$ *turns out to be the family of the marginal distributions of* X.

Proof. See (**?**). □

Definizione 321 *Assume* $\mathbb{X} \equiv \mathbb{R}^N$, *for some* $N \in \mathbb{N}$. *For any* $t \equiv (t_\ell)_{\ell=1}^L \in \mathcal{P}_{fin}(\mathbb{T})$, *we call the* t-marginal distribution function *of the stochastic process* X *the function* $F_{X;t} : \times_{\ell=1}^L \mathbb{R}^N \to \mathbb{R}_+$ *given by*

$$F_{X;t}(x_1, \ldots, x_L) \overset{def}{=} \mathbf{P}(X_{t_1} \leqslant x_1, \ldots, X_{t_L} \leqslant x_L), \quad \forall (x_1, \ldots, x_L) \in \bigtimes_{\ell=1}^{L} \mathbb{R}^N. \tag{6.7}$$

In particular, if we assume that $\mathbb{T}$ *has a first element, say* 0, *we call the* initial distribution function *of* X *the distribution of the random variable* X_0. *That is to say the function* $F_{X_0} : \mathbb{R}^N \to \mathbb{R}_+$ *given by*

$$F_{X_0}(x) \overset{def}{=} \mathbf{P}(X_0 \leqslant x), \quad \forall x \in \mathbb{R}^N. \tag{6.8}$$

Osservazione 322 *For any* $t \equiv (t_\ell)_{\ell=1}^L \in \mathcal{P}_{fin}(\mathbb{T})$, *the* t-*marginal distribution function of the time series* X *is just the joint distribution function of the* N-*variate real random variables* $X_{t_1}, \ldots, X_{t_L}$ *in the process* X.

Let $\{F_{X;t}\}_{t\in\mathcal{P}_{fin}(\mathbb{T})}$ be the set of all marginal distribution functions of the stochastic process X.

Teorema 323 *We have*

$$F_{X;(t_\ell)_{\ell=1}^L}(x_1, \ldots, x_L) = F_{X;(t_{\pi(\ell)})_{\ell=1}^L}\left(x_{\pi(1)}, \ldots, x_{\pi(L)}\right), \tag{6.9}$$

for every $(t_\ell)_{\ell=1}^L \in \mathcal{P}_{fin}(\mathbb{T})$, *every permutation* π *on the set* $\{1, \ldots, L\}$, *and every* $(x_1, \ldots, x_L) \in \times_{\ell=1}^L \mathbb{R}^N$. *In addition,*

$$\lim_{x_L \to +\infty} \cdots \lim_{x_{K+1} \to +\infty} F_{X;(t_\ell)_{\ell=1}^L}(x_1, \ldots, x_K, x_{K+1}, \ldots, x_L) = F_{X;(t_\ell)_{\ell=1}^K}(x_1, \ldots, x_K), \tag{6.10}$$

for every $K \in \mathbb{N}$ *such that* $K \leqslant L - 1$, *and every* $(x_1, \ldots, x_K) \in \times_{\ell=1}^K \mathbb{R}^N$.

Definizione 324 *We say that the process* X *is* Gaussian *if the random variables* $X_{t_1}, \ldots, X_{t_L}$ *are jointly Gaussian distributed, for every* $(t_\ell)_{\ell=1}^L \in \mathcal{C}_{fin}(\mathbb{T})$. *That is to say, the random variable* $\sum_{\ell=1}^L c_\ell X_{t_\ell}$ *is Gaussian distributed*[14], *for every* $(c_1, \ldots, c_L) \in \mathbb{R}^L$.

---

[14] There exist $\mu \in \mathbb{R}^N$ and $\Sigma \in \mathbb{R}^{N^2}$ such that

$$\sum_{k=1}^{K} c_k X_{t_k} \sim N(\mu, \Sigma),$$

where by a Gaussian distribution with mean μ and $\Sigma = 0$ we mean the Dirac distribution concentrated in μ.

6.5 Kth-Order Processes

Let $(X_t)_{t\in\mathbb{T}} \equiv X$ be a stochastic process on a probability space Ω with state space $\mathbb{R}^N$.

Definizione 325 *Given any $K \in \mathbb{N}$, We say that X is a Kth-order process if all the random variables in X have finite Kth moment.*

Osservazione 326 *If X is a Kth-order process, for some $K \in \mathbb{N}$, then X is a Jth-order process for every $1 \leqslant J \leqslant K$.*

Let $(X_t)_{t\in\mathbb{T}} \equiv X$ be a 1st-order process on a probability space Ω with state space $\mathbb{R}^N$.

Definizione 327 *We call the* mean function *of X the map $\mu_X : \mathbb{T} \to \mathbb{R}^N$ given by*

$$\mu_X(t) \overset{def}{=} \mathbf{E}[X_t], \quad \forall t \in \mathbb{T}. \tag{6.11}$$

Esempio 328 (Dirac process) *If X is a Dirac process (see Example 286), we have*

$$\mu_X(t) = x_t,$$

for every $t \in \mathbb{T}$.

Esempio 329 (independent Bernoulli process) *If X is an independent Bernoulli process with success probability p (see Example 287), we have*

$$\mu_X(t) = p,$$

for every $t \in \mathbb{T}$.

Esempio 330 (independent binomial process) *If X is an independent binomial process with number of trials n and success probability p (see Example 289), we have*

$$\mu_X(t) = np,$$

for every $t \in \mathbb{T}$.

Esempio 331 (independent Poisson process) *If X is an independent Poisson process with rate parameter λ (see Example 290) we have*

$$\mu_X(t) = \lambda,$$

for every $t \in \mathbb{T}$.

Esempio 332 (independent Gaussian process) *If X is an independent Gaussian process with with mean μ and variance σ^2 (see Example 291) we have*

$$\mu_X(t) = \mu,$$

for every $t \in \mathbb{T}$.

Let $(X_t)_{t\in\mathbb{T}} \equiv X$ be a 2nd-order process on a probability space Ω with states in $\mathbb{R}^N$.

Definizione 333 *We call the* variance-covariance function *of X the map $\Sigma_X : \mathbb{T} \to \mathbb{R}^{N^2}$ given by*

$$\Sigma_X(t) \overset{def}{=} Var(X_t), \quad \forall t \in \mathbb{T}, \tag{6.12}$$

where $Var(X_t)$ is the variance-covariance of the N-variate random variable X_t. Note that in the case $N = 1$, the variance-covariance function of X reduces to the variance function *$\sigma_X^2 : \mathbb{T} \to \mathbb{R}$ given by*

$$\sigma_X^2(t) \overset{def}{=} D^2[X_t], \quad \forall t \in \mathbb{T}. \tag{6.13}$$

Osservazione 334 *We have*

$$\Sigma_X(t) = \begin{pmatrix} \mathbf{D}^2\left[X_t^{(1)}\right] & Cov\left(X_t^{(1)}, X_t^{(2)}\right) & \cdots & Cov\left(X_t^{(1)}, X_t^{(N-1)}\right) & kCov\left(X_t^{(1)}, X_t^{(N)}\right) \\ Cov\left(X_t^{(2)}, X_t^{(1)}\right) & \mathbf{D}^2\left[X_t^{(2)}\right] & \cdots & Cov\left(X_t^{(2)}, X_t^{(N-1)}\right) & Cov\left(X_t^{(2)}, X_t^{(N)}\right) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ Cov\left(X_t^{(N-1)}, X_t^{(1)}\right) & Cov\left(X_t^{(N-1)}, X_t^{(2)}\right) & \cdots & \mathbf{D}^2\left[X_t^{(N-1)}\right] & Cov\left(X_t^{(N-1)}, X_t^{(N)}\right) \\ Cov\left(X_t^{(N)}, X_t^{(1)}\right) & Cov\left(X_t^{(N)}, X_t^{(2)}\right) & \cdots & Cov\left(X_t^{(N)}, X_t^{(N-1)}\right) & \mathbf{D}^2\left[X_t^{(N)}\right] \end{pmatrix}, \tag{6.14}$$

where $X_t^{(1)}, X_t^{(2)}, \ldots, X_t^{(N-1)}, X_t^{(N)}$ *are the entries of the* N*-variate real random variable* X_t.

Esempio 335 (Dirac process) *If* X *is a Dirac process (see Example 286), we have*

$$\sigma_X^2(t) = 0,$$

for every $t \in \mathbb{T}$.

Esempio 336 (independent Bernoulli process) *If* X *is an independent Bernoulli process with success probability* p *(see Example 287), we have*

$$\sigma_X^2(t) = p(1-p),$$

for every $t \in \mathbb{T}$.

Definizione 337 *We call the* correlation function *of* X *the map* $\mathrm{P}_X : \mathbb{T} \to \mathbb{R}^{N^2}$ *given by*

$$\mathrm{P}_X(t) \stackrel{def}{=} \begin{cases} Corr(X_t), & \forall t \in \mathbb{T} \text{ s.t. } \det(\operatorname{diag}\Sigma_X(t)) \neq 0 \\ 0, & \forall t \in \mathbb{T} \text{ s.t. } \det(\operatorname{diag}\Sigma_X(t)) = 0 \end{cases} \tag{6.15}$$

where $Corr(X_t)$ *is the correlation of the* N*-variate random variable* X_t *and* $\operatorname{diag}\Sigma_X(t)$ *is the diagonal matrix having for diagonal entries the corresponding diagonal entries of* $\Sigma_X(t)$. *Note that in the case* $N = 1$, *the correlation function of* X *reduces to the trivial function* $\rho_X : \mathbb{T} \to \mathbb{R}$ *given by*

$$\rho_X(t) = \begin{cases} 1, & \forall t \in \mathbb{T} \text{ s.t. } \sigma_X^2(t) \neq 0 \\ 0, & \forall t \in \mathbb{T} \text{ s.t. } \sigma_X^2(t) = 0 \end{cases}.$$

Osservazione 338 *Assume that* $\det(\operatorname{diag}\Sigma_X(t)) \neq 0$. *We have*

$$\mathrm{P}_X(t) = \begin{pmatrix} 1 & Corr\left(X_t^{(1)}, X_t^{(2)}\right) & \cdots & Corr\left(X_t^{(1)}, X_t^{(N-1)}\right) & Corr\left(X_t^{(1)}, X_t^{(N)}\right) \\ Corr\left(X_t^{(2)}, X_t^{(1)}\right) & 1 & \cdots & Corr\left(X_t^{(2)}, X_t^{(N-1)}\right) & Corr\left(X_t^{(2)}, X_t^{(N)}\right) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ Corr\left(X_t^{(N-1)}, X_t^{(1)}\right) & Corr\left(X_t^{(N-1)}, X_t^{(2)}\right) & \cdots & 1 & Corr\left(X_t^{(N-1)}, X_t^{(N)}\right) \\ Corr\left(X_t^{(N)}, X_t^{(1)}\right) & Corr\left(X_t^{(N)}, X_t^{(2)}\right) & \cdots & Corr\left(X_t^{(N)}, X_t^{(N-1)}\right) & 1 \end{pmatrix}, \tag{6.16}$$

where $X_t^{(1)}, X_t^{(2)}, \ldots, X_t^{(N-1)}, X_t^{(N)}$ *are the entries of the* N*-variate real random variable* X_t.

Esempio 339 (Dirac process) *If* X *is a Dirac process (see Example 286), we have*

$$\rho_X(t) = 0,$$

for every $t \in \mathbb{T}$.

Esempio 340 (independent Bernoulli process) *If X is an independent Bernoulli process with success probability p (see Example 287), we have*

$$\rho_X(t) = 1,$$

for every $t \in \mathbb{T}$.

Definizione 341 *We call the* autocovariance function *of X the map $\Gamma_X : \mathbb{T} \times \mathbb{T} \to \mathbb{R}^{N^2}$ given by*

$$\Gamma_X(s,t) \stackrel{def}{=} Cov(X_s, X_t), \quad \forall s,t \in \mathbb{T}. \tag{6.17}$$

We call the autocorrelation function *of X the map $\rho_X : \mathbb{T} \times \mathbb{T} \to \mathbb{R}^{N^2}$ given by*

$$\mathrm{P}_X(s,t) \stackrel{def}{=} \begin{cases} \operatorname{diag}\Sigma_X(s)^{-\frac{1}{2}}\,\Gamma_X(s,t)\operatorname{diag}\Sigma_X(t)^{-\frac{1}{2}}, & \forall s,t \in \mathbb{T} \text{ s.t. } \det(\operatorname{diag}\Sigma_X(s)\operatorname{diag}\Sigma_X(t)) \neq 0 \\ 0, & \forall s,t \in \mathbb{T} \text{ s.t. } \det(\operatorname{diag}\Sigma_X(s)\operatorname{diag}\Sigma_X(t)) = 0 \end{cases} \tag{6.18}$$

where $\operatorname{diag}\Sigma_X(s)$ *[resp.* $\operatorname{diag}\Sigma_X(t)$*] is the diagonal matrix having for diagonal entries the corresponding diagonal entries of $\Sigma_X(s)$ [resp. $\Sigma_X(t)$], and* $\det : \mathbb{R}^{N^2} \to \mathbb{R}$ *is the determinant function. Note that in the case $N = 1$, setting $\Gamma_X(s,t) \equiv \gamma_X(s,t)$ and $\mathrm{P}_X(s,t) \equiv \rho_X(s,t)$, we have*

$$\rho_X(s,t) \stackrel{def}{=} \begin{cases} \frac{\gamma_X(s,t)}{\sigma_X(s)\sigma_X(t)}, & \forall s,t \in \mathbb{T} \text{ s.t. } \sigma_X(s)\sigma_X(t) \neq 0 \\ 0 & \forall s,t \in \mathbb{T} \text{ s.t. } \sigma_X(s)\sigma_X(t) = 0 \end{cases}$$

where $\sigma_X : \mathbb{T} \to \mathbb{R}$ is the standard deviation function *of X, that is the square root of the variance function of X.*

Osservazione 342 *We have*

$$\Gamma_X(t,t) = \Sigma_X(t) \quad \text{and} \quad \mathrm{P}_X(t,t) = \mathrm{P}_X(t), \tag{6.19}$$

for every $t \in \mathbb{T}$.

Osservazione 343 *We have*

$$\gamma_X(s,t) \equiv \mathbf{E}\left[(X_s - \mathbf{E}[X_s])(X_t - \mathbf{E}[X_t])^{\intercal}\right] = \mathbf{E}\left[X_s X_t^{\intercal}\right] - \mathbf{E}[X_s]\,\mathbf{E}[X_t]^{\intercal},$$

for all $s,t \in \mathbb{T}$.

Proposizione 344 *Write $\left(X_s^{(1)}, \ldots, X_s^{(N)}\right)$ [resp. $\left(X_t^{(1)}, \ldots, X_t^{(N)}\right)$] for the entries of X_s [resp. X_t]. We have*

$$\gamma_X(s,t) = \begin{pmatrix} Cov\left(X_s^{(1)}, X_t^{(1)}\right) & Cov\left(X_s^{(1)}, X_t^{(2)}\right) & \cdots & Cov\left(X_s^{(1)}, X_t^{(N-1)}\right) & Cov\left(X_s^{(1)}, X_t^{(N)}\right) \\ Cov\left(X_s^{(2)}, X_t^{(1)}\right) & Cov\left(X_s^{(2)}, X_t^{(2)}\right) & \cdots & Cov\left(X_s^{(2)}, X_t^{(N-1)}\right) & Cov\left(X_s^{(2)}, X_t^{(N)}\right) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ Cov\left(X_s^{(N-1)}, X_t^{(1)}\right) & Cov\left(X_s^{(N-1)}, X_t^{(2)}\right) & \cdots & Cov\left(X_s^{(N-1)}, X_t^{(N-1)}\right) & Cov\left(X_s^{(N-1)}, X_t^{(N)}\right) \\ Cov\left(X_s^{(N)}, X_t^{(1)}\right) & Cov\left(X_s^{(N)}, X_t^{(2)}\right) & \cdots & Cov\left(X_s^{(N)}, X_t^{(N-1)}\right) & Cov\left(X_s^{(N)}, X_t^{(N)}\right) \end{pmatrix}.$$

Osservazione 345 *In the case $N = 1$, we have*

$$\gamma_X(s,t) = \gamma_X(t,s) \quad \text{and} \quad \rho_X(s,t) = \rho_X(t,s)$$

for all $s,t \in \mathbb{T}$.

Osservazione 346 *In the case $N > 1$, we have, in general,*

$$\gamma_X(s,t) \neq \gamma_X(t,s) \quad \text{and} \quad \rho_X(s,t) \neq \rho_X(t,s)$$

for some $s, t \in \mathbb{T}$.

Definizione 347 *We say that the random variables in the process X are* uncorrelated *if we have*

$$\gamma_X(s,t) = 0 \quad \text{or} \quad \rho_X(s,t) = 0 \tag{6.20}$$

for all $s, t \in \mathbb{T}$ s.t. $s \neq t$.

Esempio 348 (Bernoulli White Noise) *If X is a Bernoulli white noise, we have*

$$\gamma_X(s,t) = \begin{cases} p(1-1) & \text{if } s = t \\ 0 & \text{if } s \neq t \end{cases} \quad \text{and} \quad \rho_X(s,t) = \begin{cases} 1 & \text{if } s = t \\ 0 & \text{if } s \neq t \end{cases}$$

for every $s, t \in \mathbb{T}$.

6.6 Strong-Sense Stationary (SSS) Processes

Let $(X_t)_{t\in\mathbb{T}} \equiv \mathbf{X}$ be a stochastic process on a probability space $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ with state space a measurable space $(\mathbb{X}, \mathcal{M}) \equiv \mathbb{X}$ and let $\mathcal{C}_{fin}(\mathbb{T})$ be the set of all increasing time multi-indices.

Definizione 349 *Given any $\mathbf{t} \in \mathcal{C}_{fin}(\mathbb{T})$, such that $t \equiv (t_\ell)_{\ell=1}^L$ for some $L \in \mathbb{N}$, and any $\tau \in \mathbb{R}$, we call* τ-shift of $\mathbf{t}$ *the time multi-index $\mathbf{t}_\tau$ given by*

$$\mathbf{t}_\tau \overset{def}{=} (t_1 + \tau, \ldots, t_L + \tau). \tag{6.21}$$

Note that in Time Series literature the time multi-index t_τ is often referred to as τ-lag of t.

Definizione 350 *We say that the process $\mathbf{X}$ is* strong-sense stationary *(*SSS*) or* strongly stationary *if we have*

$$\mathbf{P}((X_{t_1}, \ldots, X_{t_L}) \in M) = \mathbf{P}((X_{t_1+\tau}, \ldots, X_{t_L+\tau}) \in M), \tag{6.22}$$

for every $\mathbf{t} \in \mathcal{C}_{fin}(\mathbb{T})$, such that $\mathbf{t} \equiv (t_\ell)_{\ell=1}^L$ for some $L \in \mathbb{N}$, every $\tau \in \mathbb{R}$, such that $\mathbf{t}_\tau \equiv (t_\ell + \tau)_{\ell=1}^L \in \mathcal{C}_{fin}(\mathbb{T})$, and every $M \in \otimes_{\ell=1}^L \mathcal{M}$.

Osservazione 351 *Considering the t-marginal [resp. τ-shifted-t-marginal] distribution of the process X, that is the probability $P_{X;t} : \otimes_{\ell=1}^L \mathcal{M} \to \mathbb{R}_+$ [resp. $P_{X;t_\tau} : \otimes_{\ell=1}^L \mathcal{M} \to \mathbb{R}_+$] (see Definition 315), Equation (6.22) can be rewritten as*

$$P_{X;t} = P_{X;t_\tau}, \quad \forall (t,\tau) \in \mathcal{C}_{fin}(\mathbb{T}) \times \mathbb{R} \text{ s.t. } t_\tau \in \mathcal{C}_{fin}(\mathbb{T}). \tag{6.23}$$

Definizione 352 *Assume that $\mathbb{X} \equiv \mathbb{R}^N$, for some $N \in \mathbb{N}$. Then X is a SSS process, if and only if we have*

$$\mathbf{P}(X_{t_1} \leqslant x_1, \ldots, X_{t_L} \leqslant x_L) = \mathbf{P}(X_{t_1+\tau} \leqslant x_1, \ldots, X_{t_L+\tau} \leqslant x_L), \tag{6.24}$$

for every $t \in \mathcal{C}_{fin}(\mathbb{T})$, $t \equiv (t_\ell)_{\ell=1}^L$, every $\tau \in \mathbb{R}$ such that $(t_\ell + \tau)_{\ell=1}^L \in \mathcal{C}_{fin}(\mathbb{T})$, and every $(x_1, \ldots, x_L) \in \times_{\ell=1}^L \mathbb{R}^N$.

Osservazione 353 *Considering the t-marginal [resp. the τ-shifted-t-marginal] distribution function of the proces X, that is the function $F_{X;t} : \times_{\ell=1}^L \mathbb{R}^N \to \mathbb{R}_+$ [resp. $F_{X;t_\tau} : \times_{\ell=1}^L \mathbb{R}^N \to \mathbb{R}_+$] (see Definition 321), Equation (6.24) can be rewritten as*

$$F_{X;t} = F_{X;t_\tau}, \quad \forall (t,\tau) \in \mathcal{C}_{fin}(\mathbb{T}) \times \mathbb{R} \text{ s.t. } t_\tau \in \mathcal{C}_{fin}(\mathbb{T}). \tag{6.25}$$

Proposizione 354 *If the process X is SSS, then the random variables in X are identically distributed. The converse is not true.*

Proof. As a particular case of Definition 350, we have

$$\mathbf{P}(X_s \in M) = \mathbf{P}(X_t \in M),$$

for all $s, t \in \mathbb{T}$ and for any $M \in \mathcal{M}$. This proves that the random variables in X are identically distributed. To show that the converse is not true, consider the discrete real random variables Y and Z given by the following distribution table

Y/Z	0	1	2
0	0	$1/7$	$1/7$
1	$2/7$	0	$1/7$
2	0	$2/7$	0

we have

$$\mathbf{P}(Y = 0, Z = 1) = 1/7 \quad \text{and} \quad \mathbf{P}(Z = 0, Y = 1) = 2/7.$$

Therefore, the 2-variate real random variables (Y, Z) and (Z, Y) have different distributions. On the other hand, the distributions of Y and Z are given by

$$\mathbf{P}(Y = 0) = 2/7, \quad \mathbf{P}(Y = 1) = 3/7, \quad \mathbf{P}(Y = 2) = 2/7$$

and

$$\mathbf{P}(Z = 0) = 2/7, \quad \mathbf{P}(Z = 1) = 3/7, \quad \mathbf{P}(Z = 2) = 2/7.$$

Hence, the random variables Y and Z have the same distribution. Now, consider the real stochastic process $(X_t)_{t=1}^{3} \equiv \mathbf{X}$ given by

$$X_1 \stackrel{\text{def}}{=} Y, \quad X_2 \stackrel{\text{def}}{=} Z, \quad X_3 \stackrel{\text{def}}{=} Y.$$

The random variables in $\mathbf{X}$ have the same distribution but the distribution of $(X_1, X_2) = (Y, Z)$ is different than the distribution of $(X_{1+1}, X_{2+1}) = (X_2, X_3) = (Z, Y)$. This prevents the process $\mathbf{X}$ from being strong-sense stationary. Note that

$$\mathbf{E}[Y] = 1, \quad \mathbf{E}[Z] = 1,$$

Furthermore, since

$$\begin{aligned}
&\mathbf{P}(YZ = 0) = \mathbf{P}(Y = 0 \vee Z = 0) = \mathbf{P}(Y = 0) + \mathbf{P}(Y = 0) - \mathbf{P}(Y = 0, Z = 0) = 4/7, \\
&\mathbf{P}(YZ = 1) = \mathbf{P}(Y = 1, Z = 1) = 0, \\
&\mathbf{P}(YZ = 2) = \mathbf{P}(Y = 1, Z = 2 \vee Y = 2, Z = 1) = \mathbf{P}(Y = 1, Z = 2) + \mathbf{P}(Y = 2, Z = 1) = 3/7, \\
&\mathbf{P}(YZ = 4) = \mathbf{P}(Y = 2, Z = 2) = 0.
\end{aligned}$$

we have

$$\mathbf{E}[YZ] = 6/7.$$

It follows,

$$\mathbf{E}[YZ] - \mathbf{E}[Y]\,\mathbf{E}[Z] = -1/7.$$

This shows that Y and Z are not uncorrelated. A fortiori, Y and Z are not independent. This clearly implies that the random variables in the process the process $(X_\ell)_{\ell=1}^{3}$ are not independent. □

The crucial reason why the process presented in Proposition 354 fails to be SSS is that the random variables in the process, altough identically distributed, are not indpendent. On the contrary, we have

Proposizione 355 *Assume that the random variables in the process X are independent and identically distributed. Then X is a SSS process.*

Proof. For simplicity we prove the claim in the case $\mathbb{X} \equiv \mathbb{R}^N$, for some $N \in \mathbb{N}$. An analogous proof which uses the definition of product probability can be given when $\mathbb{X}$ is a general measurable space. Now, since the random variables in X are identically distributed, there exists a function $F : \mathbb{R}^N \to \mathbb{R}_+$ such that

$$F_{X_t}(x) = F(x)$$

for every $t \in \mathbb{T}$ and every $x \in \mathbb{R}^N$. In addition, the N-variate real random variables in X are independent. We can then write

$$\begin{aligned} F_{X_{t_1+\tau},\dots,X_{t_L+\tau}}(x) &= F_{X_{t_1+\tau}}(x_1)\cdots F_{X_{t_L+\tau}}(x_L) = F(x_1)\cdots F(x_L) = F_{X_{t_1}}(x_1)\cdots F_{X_{t_L}}(x_L) \\ &= F_{X_{t_1},\dots,X_{t_L}}(x), \end{aligned} \tag{6.26}$$

for every $t \in \mathcal{C}_{fin}(\mathbb{Z})$, $t \equiv, (t_\ell)_{\ell=1}^L$, every $\tau \in \mathbb{R}$ such that $(t_\ell + \tau)_{\ell=1}^L \in \mathcal{C}_{fin}(\mathbb{T})$, and every $x \in \times_{\ell=1}^L \mathbb{R}^N$, $x \equiv (x_1, \dots, x_L)$. Considering Remark 353, the desired claim follows. □

Clear consequence of Proposition 355 are the following

Corollary 356 *The stochastic processes presented in Examples 287-291 are SSS processes.*

Corollary 357 *Assume that X is a Kth-order SSS process, then the moments of X of all orders up to the Kth included are time invariant.*

Stochastic processes with independent and identically distributed random variables constitute a simple and important class of SSS processes. However, differently than identical distribution, independence of the random variables in the process X is not a necessary condition for the strong stationarity of X. This is shown by the following simple example.

Esempio 358 *Let Y be a random variable on a probability space Ω and states in $\mathbb{R}^N$, for some $N \in \mathbb{N}$, with finite second order moment and zero mean. Fixed any time set $\mathbb{T}$, write*

$$X_t \overset{def}{=} Y, \quad \forall t \in \mathbb{T},$$

and consider the process $(X_t)_{t\in\mathbb{T}} \equiv \mathbf{X}$. Then the random variables in X are not independent, not even uncorrelated, and X is a SSS process.

Discussion. Since $\mathbf{E}[Y] = 0$, we have

$$\mu_{\mathbf{X}}(t) = \mathbf{E}[X_t] = \mathbf{E}[Y] = 0,$$

for every $t \in \mathbb{T}$. Furthermore, setting $\mathbf{D}^2[Y] \equiv \Sigma_Y$, we have

$$\Sigma_{\mathbf{X}}(t) = \mathbf{D}^2[X_t] = \mathbf{D}^2[Y] \equiv \Sigma_Y,$$

for every $t \in \mathbb{T}$, and

$$\gamma_{\mathbf{X}}(s,t) = Cov(Y,Y) = \mathbf{E}\left[Y^2\right] - \mathbf{E}[Y]^2 = \Sigma_Y,$$

for all $s, t \in \mathbb{T}$. Hence, the process X has constant mean, variance-covariance, and autocovariance function. In particular, as a consequence of the constant non-zero autocovariance function, the random variables in X are not uncorrelated. Now, we clearly have

$$\mathbf{P}(X_{t_1} \leqslant x_1, \dots, X_{t_L} \leqslant x_L) = \mathbf{P}(Y \leqslant x_1, \dots, Y \leqslant x_L) = \mathbf{P}(X_{t_1+\tau} \leqslant x_1, \dots, X_{t_L} \leqslant x_{n+\tau}),$$

for every $t \in \mathcal{C}_{fin}(\mathbb{T})$, $t \equiv (t_\ell)_{\ell=1}^L$, every $\tau \in \mathbb{R}$ such that $(t_\ell + \tau)_{\ell=1}^L \in \mathcal{C}_{fin}(\mathbb{T})$, and every $(x_1, \ldots, x_L) \in \times_{\ell=1}^L \mathbb{R}^N$. This yields the strong stationarity of X. Note that for each $\omega \in \Omega$ we have

$$X_t(\omega) = Y(\omega),$$

on varying of $t \in \mathbb{T}$. In words, the paths of the process X are horizontal straight lines with intercept $Y(\omega)$. $\square$

Less trivial examples are the following.

Esempio 359 *Fixed any time set $\mathbb{T} \subseteq \mathbb{R}$, let $(Y_t)_{t\in\mathbb{T}} \equiv \mathbf{Y}$ be a real stochastic process on a probability space Ω of independent standard Rademacher random variables and let Z be a standard Gaussian random variable on Ω, which is independent of the random variables in Y. Set*

$$X_t \overset{def}{=} Y_t Z, \quad \forall t \in \mathbb{T},$$

and consider the process $(X_t)_{t\in\mathbb{T}} \equiv \mathbf{X}$. Then the random variables in $\mathbf{X}$ are standard Gaussian distributed, uncorrelated but not independent, and the process $\mathbf{X}$ is strongly stationary.

Discussion. By virtue of the independence between Z and the random variables in Y, we have

$$\mu_{\mathbf{X}}(t) = \mathbf{E}[X_t] = \mathbf{E}[Y_t Z] = \mathbf{E}[Y_t]\,\mathbf{E}[Z] = 0$$

and

$$\Sigma_{\mathbf{X}}(t) = \mathbf{D}^2[X_t] = \mathbf{D}^2[Y_t Z] = \mathbf{E}\left[Y_t^2 Z^2\right] = \mathbf{E}\left[Y_t^2\right]\mathbf{E}\left[Z^2\right] = 1,$$

for every $t \in \mathbb{T}$. Furthermore, since also the random variables in Y are independent, we have

$$\gamma_{\mathbf{X}}(s,t) = Cov(Y_s Z, Y_t Z) = \mathbf{E}\left[Y_s Y_t Z^2\right] = \mathbf{E}[Y_s]\,\mathbf{E}[Y_t]\,\mathbf{E}\left[Z^2\right] = 0,$$

for all $s, t \in \mathbb{T}$ such that $s \neq t$. That is the random variables in X are uncorrelated. The lack of independence of the random variables in X can be argued by considering the random variable

$$X_t^2 = Y_t^2 Z^2 = Z^2,$$

for every $t \in \mathbb{T}$, and observing that

$$Cov\left(X_s^2, X_t^2\right) = Cov\left(Z^2, Z^2\right) = \mathbf{D}^2\left[Z^2\right] = 3,$$

for all $s, t \in \mathbb{T}$. This, preventing the pairwise independence of X_s^2 and X_t^2, prevents the independence of the random variables in X. In the end, applying the total probability theorem and considering again the independence between Z and the random variables in Y, we have

$$
\begin{aligned}
&\mathbf{P}\left(X_{t_1} \leqslant x_1, \ldots, X_{t_L} \leqslant x_L\right) \\
&= \mathbf{P}\left(Y_{t_1} Z \leqslant x_1, \ldots, Y_{t_L} Z \leqslant x_L\right) \\
&= \sum_{(e_1,\ldots,e_L)\in\{-1,1\}^L} \mathbf{P}\left(Y_{t_1} Z \leqslant x_1, \ldots, Y_{t_L} Z \leqslant x_L \mid Y_{t_1} = e_1, \ldots, Y_{t_L} = e_L\right) \mathbf{P}\left(Y_{t_1} = e_1, \ldots, Y_{t_L} = e_L\right) \\
&= \sum_{(e_1,\ldots,e_L)\in\{-1,1\}^L} \mathbf{P}\left(e_1 Z \leqslant x_1, \ldots, e_L Z \leqslant x_L \mid Y_{t_1} = e_1, \ldots, Y_{t_L} = e_L\right) \mathbf{P}\left(Y_{t_1} = e_1, \ldots, Y_{t_L} = e_L\right) \\
&= \sum_{(e_1,\ldots,e_L)\in\{-1,1\}^L} \mathbf{P}\left(e_1 Z \leqslant x_1, \ldots, e_L Z \leqslant x_L\right) \mathbf{P}\left(Y_{t_1} = e_1, \ldots, Y_{t_L} = e_L\right) \\
&= \sum_{(e_1,\ldots,e_L)\in\{-1,1\}^L} \mathbf{P}\left(e_1 Z \leqslant x_1, \ldots, e_L Z \leqslant x_L\right) \mathbf{P}\left(Y_{t_1} = e_1\right) \cdots \mathbf{P}\left(Y_{t_L} = e_L\right) \\
&= \sum_{(e_1,\ldots,e_L)\in\{-1,1\}^L} \mathbf{P}\left(e_1 Z \leqslant x_1, \ldots, e_L Z \leqslant x_L\right) \mathbf{P}\left(Y_{t_1+\tau} = e_1\right) \cdots \mathbf{P}\left(Y_{t_L+\tau} = e_L\right) \\
&= \sum_{(e_1,\ldots,e_L)\in\{-1,1\}^L} \mathbf{P}\left(e_1 Z \leqslant x_1, \ldots, e_L Z \leqslant x_L\right) \mathbf{P}\left(Y_{t_1+\tau} = e_1, \ldots, Y_{t_L+\tau} = e_L\right) \\
&= \sum_{(e_1,\ldots,e_L)\in\{-1,1\}^L} \mathbf{P}\left(e_1 Z \leqslant x_1, \ldots, e_L Z \leqslant x_L \mid Y_{t_1+\tau} = e_1, \ldots, Y_{t_L+\tau} = e_L\right) \mathbf{P}\left(Y_{t_1+\tau} = e_1, \ldots, Y_{t_L+\tau} = e_L\right) \\
&= \sum_{(e_1,\ldots,e_L)\in\{-1,1\}^L} \mathbf{P}\left(Y_{t_1+\tau} Z \leqslant x_1, \ldots, Y_{t_L+\tau} Z \leqslant x_L \mid Y_{t_1+\tau} = e_1, \ldots, Y_{t_L+\tau} = e_L\right) \mathbf{P}\left(Y_{t_1+\tau} = e_1, \ldots, Y_{t_L+\tau} = e_L\right) \\
&= \mathbf{P}\left(Y_{t_1+\tau} Z \leqslant x_1, \ldots, Y_{t_L+\tau} Z \leqslant x_L\right)
\end{aligned}
$$

for every $\mathbf{t} \in \mathcal{C}_{fin}(\mathbb{T})$, such that $\mathbf{t} \equiv (t_\ell)_{\ell=1}^L$ for some $L \in \mathbb{N}$, every $\tau \in \mathbb{R}$, such that $\mathbf{t}_\tau \equiv (t_\ell + \tau)_{\ell=1}^L \in \mathcal{C}_{fin}(\mathbb{T})$, and every $(x_1, \ldots, x_L) \in \times_{\ell=1}^L \mathbb{R}^N$. This proves that X is SSS. □

Esempio 360 *Fixed any time set $\mathbb{T}$, let $(Y_t)_{t\in\mathbb{T}} \equiv \mathbf{Y}$ be a real stochastic process on Ω with independent random variables $Y_t \sim Unif(-1,1)$, for every $t \in \mathbb{T}$, and let Z be a standard Gaussian random variable on Ω, which is independent of the random variables in $\mathbf{Y}$. Set*

$$X_t \overset{def}{=} Y_t + Z, \quad \forall t \in \mathbb{T},$$

and consider the process $(X_t)_{t\in\mathbb{T}} \equiv \mathbf{X}$. Then the random variables in X are not independent, not even uncorrelated, and the process X is SSS.

Discussion. We have

$$\mu_X(t) = \mathbf{E}[X_t] = \mathbf{E}[Y_t + Z] = \mathbf{E}[Y_t] + \mathbf{E}[Z] = 0$$

and, thanks to the independence relationship between Z and the random variables in Y,

$$\Sigma_X(t) = \mathbf{D}^2[X_t] = \mathbf{D}^2[Y_t + Z] = \mathbf{D}^2[Y_t] + \mathbf{D}^2[Z] = \frac{1}{3} + 1 = \frac{4}{3},$$

for every $t \in \mathbb{T}$. Furthermore, since also the random variables in Y are independent,

$$\gamma_X(s,t) = Cov(Y_s + Z, Y_t + Z) = Cov(Y_s, Y_t) + Cov(Y_s, Z) + Cov(Z, Y_t) + Cov(Z, Z) = \mathbf{D}^2[Z] = 1,$$

for all $s, t \in \mathbb{T}$ such that $s \neq t$. That is the random variables in X are not uncorrelated. Now, given any $t \in \mathcal{C}_{fin}(\mathbb{T})$, $t \equiv (t_\ell)_{\ell=1}^L$, consider the random vectors

$$X^{(L+1)} \equiv (X_{t_1}, \ldots, X_{t_L}, Z)^\intercal \quad \text{and} \quad Y^{(L+1)} \equiv (Y_{t_1}, \ldots, Y_{t_L}, Z)^\intercal.$$

Note that the random vector $Y^{(L+1)}$ is absolutely continuous and writing $f_{Y^{(L+1)}} : \mathbb{R}^{L+1} \to \mathbb{R}_+$ for the density of $Y^{(L+1)}$ we have

$$f_{Y^{(L+1)}}(y_1, \ldots, y_L, z) = f_{Y_{t_1}}(y_1) \cdots f_{Y_{t_L}}(y_L) f_Z(z) = f_U(y_1) \cdots f_U(y_L) f_Z(z),$$

where $f_{Y_{t_\ell}} : \mathbb{R} \to \mathbb{R}_+$ is the density of the entry Y_{t_ℓ} of $Y^{(L+1)}$, for $\ell = 1, \ldots, L$, and $f_U : \mathbb{R} \to \mathbb{R}_+$ [resp. $f_Z : \mathbb{R} \to \mathbb{R}_+$] is the density of $Unif(-1,1)$ [resp. $N(0,1)$], given by

$$f_U(y) \overset{\text{def}}{=} 1_{[-1,1]}(y), \quad y \in \mathbb{R} \quad [\text{resp. } {}_Z(z) \overset{\text{def}}{=} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad z \in \mathbb{R}].$$

In addition, setting

$$A \equiv \begin{pmatrix} 1 & \cdots & 0 & 1 \\ \vdots & \ddots & 0 & 1 \\ 0 & \cdots & 1 & 1 \\ 0 & \cdots & 0 & 1 \end{pmatrix}$$

such that $\det(A) = 1$ and

$$A^{-1} \equiv \begin{pmatrix} 1 & \cdots & 0 & -1 \\ \vdots & \ddots & 0 & -1 \\ 0 & \cdots & 1 & -1 \\ 0 & \cdots & 0 & 1 \end{pmatrix},$$

we have

$$X^{(L+1)} = AY^{(L+1)} \quad \text{and} \quad Y^{(L+1)} = A^{-1}X^{(L+1)}$$

and

$$A^{-1}(x_1, \ldots, x_L, z)^\intercal = (x_1 - z, \ldots, x_L - z, z)^\intercal$$

for every $(x_1, \ldots, x_L, z) \in \mathbb{R}^{L+1}$. It follows that also the random vector $X^{(L+1)}$ is absolutely continuous and the density $f_{X^{(L+1)}} : \mathbb{R}^{L+1} \to \mathbb{R}_+$ fulfills

$$\begin{aligned} f_{X^{(L+1)}}(x_1, \ldots, x_L, z) &= f_{Y^{(L+1)}}\left(A^{-1}(x_1, \ldots, x_L, z)^\intercal\right) \\ &= f_{Y^{(L+1)}}(x_1 - z, \ldots, x_L - z, z) \\ &= f_U(x_1 - z) \cdots f_U(x_L - z) f_Z(z). \end{aligned}$$

It follows,

$$f_{X_{t_1},\ldots,X_{t_L}}(x_1, \ldots, x_L) = \int_{\mathbb{R}} f_{X^{(L+1)}}(x_1, \ldots, x_L, z)\, d\mu_L(z) = \int_{-\infty}^{+\infty} f_U(x_1 - z) \cdots f_U(x_L - z) f_Z(z)\, dz.$$

Therefore, since the last term in the equality chain depends on the time multi-index t only through its lenght L, we can conclude that the process X is SSS. □

An important example of stochastic process which is not a SSS process is the random walk.

Esempio 361 (random walk) *The random walk process X presented in Example 292 is not a SSS process.*

Discussion. Given any $s, t \in \hat{n}$ such that $s < t$, we have

$$X_t = X_s + \sum_{n=s+1}^{t} Z_n.$$

Now, since the random variables X_s and all the random variables in the sum $\sum_{n=s+1}^{t} Z_n$ are independent, we have

$$F_{X_t} = F_{Z_s} * F_{Z_{s+1}} * \cdots * F_{Z_t}, \tag{6.27}$$

where $*$ is the convolution operator. In the large generality of cases, Equation (6.27) implies that

$$F_{X_t} \neq F_{X_s}.$$

Considering Proposition 354, the latter prevents that X is a SSS process. $\square$

Strong-sense stationarity is preserved under rather general transformations.

Let $(X_t)_{t\in\mathbb{T}} \equiv \mathbf{X}$ be a stochastic process on a probability space $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ with state space $\mathbb{R}^M$, for some $M \in \mathbb{N}$.

Proposizione 362 *Assume that the process* $\mathbf{X}$ *is SSS. In addition, for a fixed finite sequence* $\left(t_k^{(0)}\right)_{k=1}^{n}$, *for some* $n \in \mathbb{N}$, *assume that the time set* $\mathbb{T}$ *satisfies the folllowing properties*

1. *the sequence* $\left(t_k^{(0)} + t\right)_{k=1}^{n} \in \mathcal{C}_{fin}(\mathbb{T})$ *for every* $t \in \mathbb{T}$.

2. *the sequences* $\left(t_1^{(0)} + t_1, t_2^{(0)} + t_1, \ldots, t_n^{(0)} + t_1, \ldots, t_1^{(0)} + t_L, t_2^{(0)} + t_L, \ldots, t_n^{(0)} + t_L\right)$ *and* $\left(t_1^{(0)} + t_1 + \tau, t_2^{(0)} + t_1\right.$ $\mathcal{C}_{fin}(\mathbb{T})$ *for every* $\mathbf{t} \in \mathcal{C}_{fin}(\mathbb{T})$, *such that* $\mathbf{t} \equiv (t_\ell)_{\ell=1}^{L}$ *for some* $L \in \mathbb{N}$, *and every* $\tau \in \mathbb{R}$ *such that* $\mathbf{t}_\tau \equiv (t_\ell + \tau)_{\ell=1}^{L} \in \mathcal{C}_{fin}(\mathbb{T})$.

   Then, given any Borel map $g : \times_{k=1}^{n} \mathbb{R}^M \to \mathbb{R}^N$, *for some* $N \in \mathbb{N}$, *the process* $(Y_t)_{t\in\mathbb{T}} \equiv \mathbf{Y}$ *on the probability space* Ω *with state space* $\mathbb{R}^N$ *given by*

$$Y_t \stackrel{def}{=} g\left(X_{t_1^{(0)}+t}, X_{t_2^{(0)}+t} \ldots, X_{t_n^{(0)}+t}\right), \quad \forall t \in \mathbb{T}, \tag{6.28}$$

   is SSS. Note that Conditions 1 and 2 on the time set $\mathbb{T}$ *have the only purpose to allow the transformation of the process* $\mathbf{X}$ *by the map* g *and the check of the SSS property of the transformed process. For instance, if* $\mathbb{T} = \mathbb{N}$ *or* $\mathbb{T} = \mathbb{Z}$ *or* $\mathbb{T} = \mathbb{R}_+$ *they are obviously satisfied.*

Proof. Given any $\mathbf{t} \in \mathcal{C}_{fin}(\mathbb{T})$, such that $\mathbf{t} \equiv (t_\ell)_{\ell=1}^{L}$ for some $L \in \mathbb{N}$, any $\tau \in \mathbb{R}$, such that $\mathbf{t}_\tau \equiv (t_\ell + \tau)_{\ell=1}^{L} \in \mathcal{C}_{fin}(\mathbb{T})$, and any $(y_1, \ldots, y_L) \in \times_{\ell=1}^{L} \mathbb{R}^N$, recalling (6.24) in Proposition 352, we can write

$$\begin{aligned}
&\mathbf{P}\left(Y_{t_1} \leqslant y_1, \ldots, Y_{t_L} \leqslant y_L\right) \\
&= \mathbf{P}\left(g\left(X_{t_1^{(0)}+t_1}, X_{t_2^{(0)}+t_1}, \ldots, X_{t_n^{(0)}+t_1}\right) \leqslant y_1, \ldots, g\left(X_{t_1^{(0)}+t_L}, X_{t_2^{(0)}+t_L}, \ldots, X_{t_n^{(0)}+t_L}\right) \leqslant y_L\right) \\
&= \mathbf{P}\left(\left(X_{t_1^{(0)}+t_1}, X_{t_2^{(0)}+t_1}, \ldots, X_{t_n^{(0)}+t_1}\right) \in g^{-1}\left((-\infty, y_1]\right), \ldots, \left(X_{t_1^{(0)}+t_L}, X_{t_2^{(0)}+t_L}, \ldots, X_{t_n^{(0)}+t_L}\right) \in g^{-1}\left((-\infty, y_L]\right)\right) \\
&= \mathbf{P}\left(\left(X_{t_1^{(0)}+t_1}, X_{t_2^{(0)}+t_1}, \ldots, X_{t_n^{(0)}+t_1}, \ldots, X_{t_1^{(0)}+t_L}, X_{t_2^{(0)}+t_L}, \ldots, X_{t_n^{(0)}+t_L}\right) \in \bigtimes_{\ell=1}^{L} g^{-1}\left((-\infty, y_\ell]\right)\right) \\
&= \mathbf{P}\left(\left(X_{t_1^{(0)}+t_1+\tau}, X_{t_2^{(0)}+t_1+\tau}, \ldots, X_{t_n^{(0)}+t_1+\tau}, \ldots, X_{t_1^{(0)}+t_L+\tau}, X_{t_2^{(0)}+t_L+\tau}, \ldots, X_{t_n^{(0)}+t_L+\tau}\right) \in \bigtimes_{\ell=1}^{L} g^{-1}\left((-\infty, y_\ell]\right)\right) \\
&= \mathbf{P}\left(\left(X_{t_1^{(0)}+t_1+\tau}, X_{t_2^{(0)}+t_1+\tau}, \ldots, X_{t_n^{(0)}+t_1+\tau}\right) \in g^{-1}\left((-\infty, y_1]\right), \ldots, \left(X_{t_1^{(0)}+t_L+\tau}, X_{t_2^{(0)}+t_L+\tau}, \ldots, X_{t_n^{(0)}+t_L+\tau}\right) \in g^{-1}\right. \\
&= \mathbf{P}\left(g\left(X_{t_1^{(0)}+t_1+\tau}, X_{t_2^{(0)}+t_1+\tau}, \ldots, X_{t_n^{(0)}+t_1+\tau}\right) \leqslant y_1, \ldots, g\left(X_{t_1^{(0)}+t_L+\tau}, X_{t_2^{(0)}+t_L+\tau}, \ldots, X_{t_n^{(0)}+t_L+\tau}\right) \leqslant y_L\right) \\
&= \mathbf{P}\left(Y_{t_1+\tau} \leqslant y_1, \ldots, Y_{t_L+\tau} \leqslant y_L\right).
\end{aligned}$$

This proves that $\mathbf{Y}$ is SSS. $\square$

Esempio 363 *Let* $\mathbf{X}$ *be a SSS real stochastic process on a probability space* Ω*. Then, for any fixed* $K \in \mathbb{N}$*, the process* $(Y_t)_{t\in\mathbb{T}}$ *on* Ω *given by*

$$Y_t \overset{def}{=} X_t^K, \quad \forall t \in \mathbb{T},$$

is SSS.

Esempio 364 *Let* $\mathbf{X}$ *be a SSS strictly positive stochastic process on a probability space* Ω*. Then, for any fixed* $p \in \mathbb{R}$*, the process* $(Y_t)_{t\in\mathbb{T}}$ *on* Ω *given by*

$$Y_t \overset{def}{=} X_t^p, \quad \forall t \in \mathbb{T},$$

is SSS.

Esempio 365 *Let* X *be a SSS real stochastic process on a probability space* Ω*. Then the process* $(Y_t)_{t\in\mathbb{T}}$ *given by*

$$Y_t \overset{def}{=} \exp(X_t), \quad \forall t \in \mathbb{T},$$

is strongly stationary.

Esempio 366 *Let* X *be a SSS strictly positive stochastic process on a probability space* Ω*. Then, the process* $(Y_t)_{t\in\mathbb{T}}$ *given by*

$$Y_t \overset{def}{=} \log(X_t), \quad \forall t \in \mathbb{T},$$

is strongly stationary.

Esempio 367 *Let* X *be a SSS real stochastic process on a probability space* Ω*. For simplicity, assume that* $\mathbb{T} \equiv \mathbb{N}$*. Then, for any fixed* $n \in \mathbb{N}$*, the process* $(Y_t)_{t\in\mathbb{N}}$ *given by*

$$Y_t \overset{def}{=} X_t \cdot \cdots \cdot X_{t+n}, \quad \forall t \in \mathbb{N},$$

is strongly stationary.

Let $(X_t)_{t\in\mathbb{T}} \equiv X$ be a stochastic process on a probability space $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ with state space $\mathbb{R}^N$, for some $N \in \mathbb{N}$.

Proposizione 368 *If the process* X *is SSS, then its entries* $X^{(1)}, \ldots, X^{(N)}$ *are also SSS. The converse is not true.*

Proof. We have

$$X_t^{(n)} \overset{\text{def}}{=} \pi_n(X_t), \quad \forall t \in \mathbb{T},$$

where $\pi_n : \mathbb{R}^N \to \mathbb{R}$ is the nth canonical projection of $\mathbb{R}^N$ on $\mathbb{R}$, for any $n = 1, \ldots, N$. Now, $\pi_n : \mathbb{R}^N \to \mathbb{R}$ is a continuous map. This, on account of Proposition **??**, shows that under the SSS assumption for X the entries $X^{(1)}, \ldots, X^{(N)}$ are also SSS.

To show that the converse is not true, consider the discrete real random variables Y and Z introduced in Proposition 354. We know that Y and Z have the same distribution but the random vectors $(Y, Z)^\intercal$ and $(Z, Y)^\intercal$ have different distributions. Now, consider the real stochastic processes $\left(X_t^{(1)}\right)_{t=1}^2$ and $\left(X_t^{(2)}\right)_{t=1}^2$ given by

$$X_1^{(1)} \overset{\text{def}}{=} Y, \quad X_2^{(1)} \overset{\text{def}}{=} Z \quad \text{and} \quad X_1^{(2)} \overset{\text{def}}{=} Z, \quad X_2^{(2)} \overset{\text{def}}{=} Y.$$

Since Y and Z have the same distribution, the processes $\left(X_t^{(1)}\right)_{t=1}^2$ and $\left(X_t^{(2)}\right)_{t=1}^2$ are clearly SSS. On the other hand, since $(Y,Z)^\intercal$ and $(Z,Y)^\intercal$ have different distributions, the process $(X_t)_{t=1}^2$ with states in $\mathbb{R}^2$ given by

$$X_t \overset{\text{def}}{=} \left(X_t^{(1)}, X_t^{(2)}\right)^\intercal, \quad \forall t = 1, 2,$$

is not SSS. $\square$

Proposizione 369 *If the entries entries $X^{(1)}, \dots, X^{(N)}$ of the process X are strongly stationary and independent, then the process X is strongly stationary.*

Proof. Assume that the real processes $X^{(1)}, \dots, X^{(N)}$ are SSS and independent. Then we can write

$$\begin{aligned}
&\mathbf{P}\left(X_{t_1} \leqslant x_1, \dots, X_{t_L} \leqslant x_L\right) \\
&= \mathbf{P}\left(\left(X_{t_1}^{(1)} \leqslant x_1^{(1)}, \dots, X_{t_1}^{(N)} \leqslant x_1^{(N)}\right), \dots, \left(X_{t_L}^{(1)} \leqslant x_L^{(1)}, \dots, X_{t_L}^{(N)} \leqslant x_L^{(N)}\right)\right) \\
&= \mathbf{P}\left(\bigcap_{\ell=1}^{L} \left\{X_{t_\ell}^{(1)} \leqslant x_\ell^{(1)}, \dots, X_{t_\ell}^{(N)} \leqslant x_\ell^{(N)}\right\}\right) \\
&= \mathbf{P}\left(\bigcap_{\ell=1}^{L} \bigcap_{n=1}^{N} \left\{X_{t_\ell}^{(n)} \leqslant x_\ell^{(n)}\right\}\right) = \mathbf{P}\left(\bigcap_{n=1}^{N} \bigcap_{\ell=1}^{L} \left\{X_{t_\ell}^{(n)} \leqslant x_\ell^{(n)}\right\}\right) \\
&= \mathbf{P}\left(\bigcap_{n=1}^{N} \left\{X_{t_1}^{(n)} \leqslant x_1^{(n)}, \dots, X_{t_L}^{(n)} \leqslant x_L^{(n)}\right\}\right) \\
&= \prod_{n=1}^{N} \mathbf{P}\left(X_{t_1}^{(n)} \leqslant x_1^{(n)}, \dots, X_{t_L}^{(n)} \leqslant x_L^{(n)}\right) = \prod_{n=1}^{N} \mathbf{P}\left(X_{t_1+\tau}^{(n)} \leqslant x_1^{(n)}, \dots, X_{t_L+\tau}^{(n)} \leqslant x_L^{(n)}\right) \\
&= \mathbf{P}\left(\bigcap_{n=1}^{N} \left\{X_{t_1+\tau}^{(n)} \leqslant x_1^{(n)}, \dots, X_{t_L+\tau}^{(n)} \leqslant x_L^{(n)}\right\}\right) \\
&= \mathbf{P}\left(\bigcap_{n=1}^{N} \bigcap_{\ell=1}^{L} \left\{X_{t_\ell+\tau}^{(n)} \leqslant x_\ell^{(n)}\right\}\right) = \mathbf{P}\left(\bigcap_{\ell=1}^{L} \bigcap_{n=1}^{N} \left\{X_{t_\ell+\tau}^{(n)} \leqslant x_\ell^{(n)}\right\}\right) \\
&= \mathbf{P}\left(\bigcap_{\ell=1}^{n} \left\{X_{t_\ell+\tau}^{(1)} \leqslant x_\ell^{(1)}, \dots, X_{t_\ell+\tau}^{(m)} \leqslant x_\ell^{(m)}\right\}\right) \\
&= \mathbf{P}\left(\left(X_{t_1+\tau}^{(1)} \leqslant x_1^{(1)}, \dots, X_{t_1+\tau}^{(N)} \leqslant x_1^{(N)}\right), \dots, \left(X_{t_L+\tau}^{(1)} \leqslant x_L^{(1)}, \dots, X_{t_L+\tau}^{(N)} \leqslant x_L^{(N)}\right)\right) \\
&= \mathbf{P}\left(X_{t_1+\tau} \leqslant x_1, \dots, X_{t_L+\tau} \leqslant x_L\right),
\end{aligned}$$

for every $t \in \mathcal{C}_{fin}(\mathbb{T})$, $t \equiv (t_\ell)_{\ell=1}^L$, every $\tau \in \mathbb{R}$ such that $(t_\ell + \tau)_{\ell=1}^L \in \mathcal{C}_{fin}(\mathbb{T})$, and every $(x_1, \dots, x_L) \in \times_{\ell=1}^L \mathbb{R}^N$. This proves that X is SSS. $\square$

6.6.1 Processes with Strict-Sense Stationary Increments

Let $(X_t)_{t\in\mathbb{T}} \equiv X$ be a stochastic process on a probability space $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ with state space $\mathbb{R}^N$, for some $N \in \mathbb{N}$.

Definizione 370 *We call the random variable $X_t - X_s$ the* increment *of the process X corresponding at the time increment $t - s$, for all $s, t \in \mathbb{T}$ suche that $s < t$.*

Definizione 371 *We say that the process X has* strict-sense stationary increments *or X is a* SSSI *process, if we have*

$$\mathbf{P}\left(X_{t_2}-X_{t_1}\leqslant x_1,\ldots,X_{t_{L+1}}-X_{t_L}\leqslant x_L\right)=\mathbf{P}\left(X_{t_2+\tau}-X_{t_1+\tau}\leqslant x_1,\ldots,X_{t_{L+1}+\tau}-X_{t_L+\tau}\leqslant x_L\right), \tag{6.29}$$

for every $t\in\mathcal{C}_{fin}(\mathbb{T})$, $t\equiv(t_\ell)_{\ell=1}^{L+1}$, every $\tau\in\mathbb{R}$ such that $(t_\ell+\tau)_{\ell=1}^{L+1}\in\mathcal{C}_{fin}(\mathbb{T})$, and every $(x_1,\ldots,x_L)\in\bigtimes_{\ell=1}^{L}\mathbb{R}^N$.

Proposizione 372 *Assume that X is a SSS process. Then X is a* SSSI *process.*

□For simplicity in the proof, assume further that X is a real process. Hence, consider the characteristc functions of the random vectors $\left(X_{t_2}-X_{t_1},\ldots,X_{t_{L+1}}-X_{t_L}\right)$ and $\left(X_{t_2+\tau}-X_{t_1+\tau},\ldots,X_{t_{L+1}+\tau}-X_{t_L+\tau}\right)$, that is the functions $\varphi_{\left(X_{t_2}-X_{t_1},\ldots,X_{t_{L+1}}-X_{t_L}\right)}:\mathbb{R}^L\to\mathbb{R}$ and $\varphi_{\left(X_{t_2}+\tau-X_{t_1+\tau},\ldots,X_{t_{L+1}}+\tau-X_{t_L+\tau}\right)}:\mathbb{R}^L\to\mathbb{R}$ given by

$$\varphi_{\left(X_{t_2}-X_{t_1},\ldots,X_{t_{L+1}}-X_{t_L}\right)}(u)\overset{\text{def}}{=}\int_\Omega e^{i\sum_{\ell=1}^{L}u_\ell\left(X_{t_{\ell+1}}-X_{t_\ell}\right)}\,d\mathbf{P},\quad\forall u\in\mathbb{R}^L \tag{6.30}$$

and

$$\varphi_{\left(X_{t_2}+\tau-X_{t_1+\tau},\ldots,X_{t_{L+1}}+\tau-X_{t_L+\tau}\right)}(u)\overset{\text{def}}{=}\int_\Omega e^{i\sum_{\ell=1}^{L}u_\ell\left(X_{t_{\ell+1}+\tau}-X_{t_\ell+\tau}\right)}\,d\mathbf{P},\quad\forall u\in\mathbb{R}^L \tag{6.31}$$

We have

$$\begin{aligned}\sum_{\ell=1}^{L}u_\ell\left(X_{t_{\ell+1}}-X_{t_\ell}\right)&=-u_1X_{t_1}+(u_1-u_2)X_{t_2}+\cdots+(u_{L-1}-u_L)X_{t_L}+u_LX_{t_{L+1}}\\&=g\left(u,X_{t_1},X_{t_2},\ldots,X_{t_L},X_{t_{L+1}}\right)\end{aligned} \tag{6.32}$$

and

$$\begin{aligned}\sum_{\ell=1}^{L}u_\ell\left(X_{t_{\ell+1}+\tau}-X_{t_\ell+\tau}\right)&=-u_1X_{t_1+\tau}+(u_1-u_2)X_{t_2+\tau}+\cdots+(u_{L-1}-u_L)X_{t_L+\tau}+u_LX_{t_{L+1}+\tau}\\&=g\left(u_1,\ldots,u_L,X_{t_1+\tau},X_{t_2+\tau},\ldots,X_{t_L+\tau},X_{t_{L+1}+\tau}\right),\end{aligned} \tag{6.33}$$

where $g:\mathbb{R}^L\times\mathbb{R}^{L+1}\to\mathbb{R}$ is the Borel function given by

$$\begin{aligned}&g(u_1,\ldots,u_L,x_1,\ldots,x_L,x_{L+1})\overset{\text{def}}{=}-u_1x_1+(u_1-u_2)x_2+\cdots+(u_{L-1}-u_L)x_L+u_Lx_{L+1},\\&\quad\forall(u_1,\ldots,u_L,x_1,\ldots,x_L,x_{L+1})\in\mathbb{R}^L\times\mathbb{R}^{L+1}\end{aligned}$$

On the other hand, since X is a SSS process, it follows that the random vectors $\left(X_{t_1},X_{t_2},\ldots,X_{t_L},X_{t_{L+1}}\right)$ and $\left(X_{t_1+\tau},X_{t_2+\tau},\ldots,X_{t_L+\tau},X_{t_{L+1}+\tau}\right)$ have the same distribution. This implies that also the random variables on the left hand side of Equations (6.32) and (6.33), which are the same Borel function of identically distributed random vectors, have the same distribution for every $u\in\mathbb{R}^L$. Thus, the left hand side of Equations (6.30) and (6.31) are equal for every $u\in\mathbb{R}^L$, every $t\in\mathcal{C}_{fin}(\mathbb{T})$, $t\equiv(t_\ell)_{\ell=1}^{L+1}$, and every $\tau\in\mathbb{R}$ such that $(t_\ell+\tau)_{\ell=1}^{L+1}\in\mathcal{C}_{fin}(\mathbb{T})$. In the end, having the same characteristic function, the random vectors $\left(X_{t_2}-X_{t_1},\ldots,X_{t_{L+1}}-X_{t_L}\right)$ and $\left(X_{t_2+\tau}-X_{t_1+\tau},\ldots,X_{t_{L+1}+\tau}-X_{t_L+\tau}\right)$ have the same distribution. □

There exist SSSI processes which are not SSS processes. The most basic in this class of processes is the random walk, which is not a SSS process (see Example 361).

Esempio 373 (random walk) *The random walk process X presented in Example (292) is a SSSI process.*

Discussion. Write

$$S_u^v \equiv \sum_{s=u+1}^{v} Z_s$$

for all $u, v \in \mathbb{T}$ such that $u \leqslant v$. Note that, since the random variables in the process Z are independent and identically distributed, we have

$$F_{S_u^v} = *_{s=u}^{v} F_{Z_s} = *_{s=1}^{v-u} F_{Z_s},$$

for all $u, v \in \mathbb{T}$ such that $u \leqslant v$, where $*$ is the convolution product and $F_{S_u^v} : \mathbb{R}^N \to \mathbb{R}$ [resp. $F_{Z_s} : \mathbb{R}^N \to \mathbb{R}$] is the distribution function of the random variable S_u^v [resp. Z_s]. In addition, the random variables $S_{u_1}^{v_1}, \ldots, S_{u_1}^{v_L}$ are independent for all $u_1, \ldots, u_L, v_1, \ldots, v_L \in \mathbb{T}$ such that $u_1 \leqslant v_1 < u_2 \leqslant v_2 < \ldots u_L \leqslant v_L$. Now, we have

$$\left(X_{t_2} - X_{t_1}, \ldots, X_{t_{L+1}} - X_{t_L}\right) = \left(\sum_{s=t_1+1}^{t_2} Z_s, \ldots, \sum_{s=t_L+1}^{t_{L+1}} Z_s,\right) = \left(S_{t_1+1}^{t_2}, \ldots, S_{t_L+1}^{t_{L+1}}\right) \tag{6.34}$$

and

$$\left(X_{t_2+\tau} - X_{t_1+\tau}, \ldots, X_{t_{L+1}+\tau} - X_{t_L+\tau}\right) = \left(\sum_{s=t_1+\tau+1}^{t_2+\tau} Z_s, \ldots, \sum_{s=t_L+\tau+1}^{t_{L+1}+\tau} Z_s\right) = \left(S_{t_1+\tau+1}^{t_2+\tau}, \ldots, S_{t_L+\tau+1}^{t_L+\tau+1}\right), \tag{6.35}$$

for every $t \in \mathcal{C}_{fin}(\mathbb{T})$, $t \equiv (t_\ell)_{\ell=1}^{L+1}$, and every $\tau \in \mathbb{R}$ such that $(t_\ell + \tau)_{\ell=1}^{L+1} \in \mathcal{C}_{fin}(\mathbb{T})$. On the other hand, for what observed above, the distribution functions of the random vectors on the right hand side of Equations (6.34) and (6.35) $F_{S_{t_1+1}^{t_2}, \ldots, S_{t_L+1}^{t_L+1}} : \times_{\ell=1}^{L} \mathbb{R}^N \to \mathbb{R}$ and $F_{S_{t_1+\tau+1}^{t_2+\tau}, \ldots, S_{t_L+\tau+1}^{t_L+\tau+1}} : \times_{\ell=1}^{L} \mathbb{R}^N \to \mathbb{R}$ are given by

$$F_{S_{t_1+1}^{t_2}, \ldots, S_{t_L+1}^{t_L+1}} = F_{S_{t_1+1s}^{t_2}} \cdot \cdots \cdot F_{S_{t_L+1}^{t_L+1}} = *_{s=1}^{t_2-t_1} F_{Z_s} \cdot \cdots \cdot *_{s=1}^{t_{L+1}-t_L} F_{Z_s}$$

and

$$F_{S_{t_1+\tau+1}^{t_2+\tau}, \ldots, S_{t_L+\tau+1}^{t_L+\tau+1}} = F_{S_{t_1+\tau+1}^{t_2+\tau}} \cdot \cdots \cdot F_{S_{t_L+\tau+1}^{t_L+\tau+1}} = *_{s=1}^{t_2-t_1} F_{Z_s} \cdot \cdots \cdot *_{s=1}^{t_{L+1}-t_L} F_{Z_s}$$

respectively. This shows that the random vectors on the left hand side of Equations (6.34) and (6.35) have the same distribution, that is X is a SSSI process. □

6.6.2 Processes with Independent Increments

Let $(X_t)_{t\in\mathbb{T}} \equiv X$ be a stochastic process on a probability space $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ with state space $\mathbb{R}^N$, for some $N \in \mathbb{N}$.

Definizione 374 *We say that the process X has* independent increments *process, or X is a* II *process, if the increments $X_{t_2} - X_{t_1}, \ldots, X_{t_{L+1}} - X_{t_L}$ are independent for every $t \in \mathcal{C}_{fin}(\mathbb{T})$, $t \equiv (t_\ell)_{\ell=1}^{L+1}$. In case $\mathbb{T}$ has a first element, say 0, we also require that the random variable X_0 be independent of any set of increments.*

Esempio 375 (random walk) *The random walk process X presented in Example (292) is a II process.*

Discussion. We already observed that the increments of the random walk are independent while discussing Example 373. □

Inportant for the characterization of Markov processes if the following result.

Teorema 376 *Assume that the process X has* independent increments. *Then the increment $X_t - X_s$ is independent of the σ-algebra $\mathcal{F}_s^X$ of the filtration $\left(\mathcal{F}_t^X\right)_{t\in\mathbb{T}} \equiv \mathfrak{F}^X$ generated by the process for all $s, t \in \mathbb{T}$, such that $s < t$. Conversely if the increment $X_t - X_s$ is independent of $\mathcal{F}_s^X$ for all $s, t \in \mathbb{T}$, such that $s < t$, then the increments $X_{t_2} - X_{t_1}, \ldots, X_{t_{L+1}} - X_{t_L}$ are independent for every $t \in \mathcal{C}_{fin}(\mathbb{T})$, $t \equiv (t_\ell)_{\ell=1}^{L+1}$.*

Definizione 377 *We call a SSSI process with independent increments a* strict sense stationary independent increment *(SSSII) process.*

Proposizione 378 *The process X is a SSSII if and only if X is a II process and the increment $X_t - X_s$ is stationary for all $s, t \in \mathbb{T}$, such that $s < t$.*

6.7 Weak-Sense Stationary (WSS) Processes

Let $(X_t)_{t\in\mathbb{T}} \equiv X$ be a stochastic process on a probability space Ω with state space $\mathbb{R}^N$.

Definizione 379 *Assume that X is a 2nd-order process. Then we say that X is* weak-sense stationary *(*WSS*) or* weakly stationary *or also* covariance stationary *if we have:*

1. $\mu_X(s) = \mu_X(t)$, *for all $s, t \in \mathbb{T}$;*
2. $\gamma_X(s,t) = \gamma_X(s+\tau, t+\tau)$ *for all $s, t \in \mathbb{T}$ and every $\tau \in \mathbb{R}$ s.t. $s+\tau$ and $t+\tau \in \mathbb{T}$.*

Osservazione 380 *If X is weakly stationary, given any $t_0 \in \mathbb{T}$, we have*

$$\mu_X(t) = \mu_X(t_0), \quad \Sigma_X(t) = \Sigma_X(t_0), \quad \mathrm{P}_X(t) = \mathrm{P}_X(t_0),$$

for every $t \in \mathbb{T}$.

Definizione 381 *If the process X is weakly stationary, we write μ_X [resp. Σ_X, P_X] for the costant value of the mean [resp. variance-covariance, correlation] function of X and we refer to μ_X [resp. Σ_X, P_X] as the* mean *[resp.* variance-covariance, correlation*] of X.*

Proposizione 382 *If X is weakly stationary, given any $t_0 \in \mathbb{T}$, we have*

$$\gamma_X(s,t) = \gamma_X(t_0, t_0 + (t-s)) \quad \textit{and} \quad \rho_X(s,t) = \rho_X(t_0, t_0 + (t-s)),$$

for all $s, t \in \mathbb{T}$ such that $t_0 + (t-s) \in \mathbb{T}$.

Proof. Under the assumption of weakly covariance, for all $s, t \in \mathbb{T}$ s.t. $t_0 + (t-s) \in \mathbb{T}$, setting $\tau \equiv s - t_0$, we can write

$$\gamma_X(s,t) = \gamma_X(s+\tau, t+\tau) = \gamma_X(s + (t_0 - s), t + (t_0 - s)) = \gamma_X(t_0, t_0 + (t-s))$$

In addition, we have

$$\Sigma_X(t) = \Sigma_X(s) = \Sigma_X(t_0 + (t-s)) = \Sigma_X(t_0),$$

for all $s, t \in \mathbb{T}$. Therefore, in the non-trival case $\Sigma_X(t) \neq 0$ for every $t \in \mathbb{T}$, we have

$$\rho_X(s,t) = (\operatorname{diag}\Sigma_X(s))^{-\frac{1}{2}} \gamma_X(s,t) (\operatorname{diag}\Sigma_X(t))^{-\frac{1}{2}} = (\operatorname{diag}\Sigma_X(t_0))^{-\frac{1}{2}} \gamma_X(t_0, t_0 + (t-s)) \left(\operatorname{diag}\Sigma_X\left(_{t_0+(t-s)}\right)\right)^{-\frac{1}{2}} =$$

This completes the proof. $\square$

Definizione 383 *If X is weakly stationary, fixed any $t_0 \in \mathbb{T}$ write*

$$\mathbb{T}_0 \equiv \{\tau \in \mathbb{R} : t_0 + \tau \in \mathbb{T}\}.$$

We call the map $\gamma_{X,t_0} : \mathbb{T}_0 \to \mathbb{R}^{m^2}$ given by

$$\gamma_{X,t_0}(\tau) \stackrel{def}{=} \gamma_X(t_0, t_0 + \tau), \quad \forall \tau \in \mathbb{T}_0, \tag{6.36}$$

the reduced autocovariance function *of X referred to t_0 and, for any $\tau \in \mathbb{T}_0$, we call the matrix $\gamma_{X,t_0}(\tau)$ the* autocovariance *of X* at lag τ. *We call the map $\rho_{X,t_0} : \mathbb{T}_0 \to \mathbb{R}^{m^2}$ given by*

$$\rho_{X,t_0}(\tau) \stackrel{def}{=} \rho_X(t_0, t_0 + \tau), \quad \forall \tau \in \mathbb{T}_0, \tag{6.37}$$

a reduced autocorrelation function *of X and, for any $\tau \in \mathbb{T}_0$, we call the matrix $\rho_{X,t_0}(\tau)$ the* autocorrelation *or the* serial correlation *of X* at lag τ.

The autocovariance and autocorrealtion functions express the temporal "linear dependence" of the random variables in a weakly stationary process.

Proposizione 384 *Assume that X is weakly stationary and, for some $t_0 \in \mathbb{T}$, let $\gamma_{X,t_0} : \mathbb{T}_0 \to \mathbb{R}$ [resp. $\rho_{X,t_0} : \mathbb{T}_0 \to \mathbb{R}$] be a reduced autocovariance [resp. autocorrelation] function of X. Then we have*

$$\rho_{X,t_0}(\tau) = (\operatorname{diag} \Sigma_X(t_0))^{-\frac{1}{2}} \gamma_{X,t_0}(\tau) (\operatorname{diag} \Sigma_X(t_0))^{-\frac{1}{2}} \tag{6.38}$$

and

$$(\operatorname{diag} \Sigma_X(t_0))^{\frac{1}{2}} \rho_{X,t_0}(\tau) (\operatorname{diag} \Sigma_X(t_0))^{\frac{1}{2}} = \gamma_{X,t_0}(\tau) \tag{6.39}$$

for every $\tau \in \mathbb{T}_0$.

Proof. Under the assumption of weakly covariance, we can write

$$\begin{aligned}\rho_{X,t_0}(\tau) &= \rho_X(t_0, t_0 + \tau) = (\operatorname{diag} \Sigma_X(t_0))^{-\frac{1}{2}} \gamma_X(t_0, t_0 + \tau) (\operatorname{diag} \Sigma_X(t_0 + \tau))^{-\frac{1}{2}} \\ &= (\operatorname{diag} \Sigma_X(t_0))^{-\frac{1}{2}} \gamma_X(t_0, t_0 + \tau) (\operatorname{diag} \Sigma_X(t_0))^{-\frac{1}{2}} = (\operatorname{diag} \Sigma_X(t_0))^{-\frac{1}{2}} \gamma_{X,t_0}(\tau) (\operatorname{diag} \Sigma_X(t_0))^{-\frac{1}{2}}.\end{aligned}$$

This proves (6.38). Equation (6.39) immediately follows. $\square$

Osservazione 385 *For any $t_0 \in \mathbb{T}$ we have $0 \in \mathbb{T}_0$. Furthermore,*

$$\gamma_{X,t_0}(0) = \Sigma_X \quad \textit{and} \quad \rho_{X,t_0}(0) = \mathrm{P}_X,$$

where Σ_X [resp. P_X] is the variance-autocovariance [resp. correlation] of X.

Osservazione 386 *In the case $N = 1$, assume that X is weakly stationary and let $\gamma_{X,t_0} : \mathbb{T}_0 \to \mathbb{R}$ [resp. $\rho_{X,t_0} : \mathbb{T}_0 \to \mathbb{R}$] be a reduced autocovariace [resp. autocorrelation] function of X, for some $t_0 \in \mathbb{T}$. Then we have*

$$\rho_{X,t_0}(\tau) = \frac{\gamma_{X,t_0}(\tau)}{\gamma_{X,t_0}(0)} = \frac{\gamma_{X,t_0}(\tau)}{\sigma_X^2}, \tag{6.40}$$

for every $\tau \in \mathbb{T}_0$. In particular,

$$\rho_{X,t_0}(0) = 1. \tag{6.41}$$

Proposizione 387 *In the case $N = 1$, assume that X is weakly stationary and let $\gamma_{X,t_0} : \mathbb{T}_0 \to \mathbb{R}$ [resp. $\rho_{X,t_0} : \mathbb{T}_0 \to \mathbb{R}$] be a reduced autocovariace [resp. autocorrelation] function of X, for some $t_0 \in \mathbb{T}$. Then we have*

$$\gamma_{X,t_0}(-\tau) = \gamma_{X,t_0}(\tau) \quad \textit{[resp. } \rho_{X,t_0}(-\tau) = \rho_{X,t_0}(\tau)\textit{]}, \tag{6.42}$$

for every $\tau \in \mathbb{T}_0$ such that $-\tau \in \mathbb{T}_0$. Furthermore,

$$|\gamma_{X,t_0}(\tau)| \leqslant \sigma_X^2 \quad \textit{[resp. } |\rho_{X,t_0}(\tau)| \leqslant 1\textit{]}, \tag{6.43}$$

for every $\tau \in \mathbb{T}_0$.

Proof. We have

$$\gamma_{X,t_0}(\tau) \equiv \gamma_X(t_0, \tau) = Cov\left(X_{t_0}, X_{t_0+\tau}\right). \tag{6.44}$$

On the other hand, when $m = 1$

$$Cov\left(X_{t_0}, X_{t_0+\tau}\right) = Cov\left(X_{t_0+\tau}, X_{t_0}\right) \equiv \gamma_X(t_0 + \tau, t_0). \tag{6.45}$$

Now, the weakly stationarity of X implies

$$\gamma_X(t_0 + \tau, t_0) = \gamma_X((t_0 + \tau) - \tau, t_0 - \tau) = \gamma_X(t_0, t_0 - \tau) \equiv \gamma_{X,t_0}(-\tau). \tag{6.46}$$

Combining (6.44)-(6.46) the desired (**??**) immediately follows. □

Osservazione 388 *If X is a 2nd-order strongly stationary process, then X is weakly stationary.*

Proposizione 389 *Assume that the random variables in X are uncorrelated and identically distributed. Then X is a weakly stationary process.*

Esempio 390 *Let $m = 1$, let $(Y_t)_{t\in\mathbb{N}} \equiv Y$ be a real stochastic process on Ω of independent and identically distributed standard Gaussian random variables, $Y_t \sim N(0,1)$, for every $t \in \mathbb{N}$, and let Z be a standard Rademacher random variable on Ω which is independent of Y_t, for every $t \in \mathbb{N}$. Set*

$$X_t \stackrel{def}{=} \begin{cases} Y_{\frac{t+1}{2}} & \textit{if } t \in \mathbb{D} \\ Y_{\frac{t}{2}} Z & \textit{if } t \in \mathbb{E} \end{cases}$$

where $\mathbb{D}$ [resp. $\mathbb{E}$] is the subset of all odd [resp. even] numbers of $\mathbb{N}$. Then $(X_t)_{t\in\mathbb{N}}$ is a weakly stationary process which is not strongly stationary.

Discussion. We have

$$\mu_X(t) = \mathbf{E}[X_t] = \begin{cases} \mathbf{E}\left[Y_{\frac{t+1}{2}}\right] = 0 & \text{if } t \in \mathbb{D} \\ \mathbf{E}\left[Y_{\frac{t}{2}} Z\right] = \mathbf{E}\left[Y_{\frac{t}{2}}\right]\mathbf{E}[Z] = 0 & \text{if } t \in \mathbb{E} \end{cases}$$

and, since Z is independent of the random variables in Y,

$$\Sigma_X(t) = \mathbf{D}^2[X_t] = \begin{cases} \mathbf{D}^2\left[Y_{\frac{t+1}{2}}\right] = 1 & \text{if } t \in \mathbb{D} \\ \mathbf{D}^2\left[Y_{\frac{t}{2}} Z\right] = \mathbf{E}\left[Y_{\frac{t}{2}}^2 Z^2\right] = \mathbf{E}\left[Y_{\frac{t}{2}}^2\right]\mathbf{E}\left[Z^2\right] = 1 & \text{if } t \in \mathbb{E} \end{cases}.$$

Furthermore, assuming $s < t$, since the random variables in Y are also independent, we have

$$\begin{aligned} \gamma_X(s,t) &= Cov(X_s, X_t) = \mathbf{E}[X_s X_t] - \mathbf{E}[X_s]\mathbf{E}[X_t] \\ &= \mathbf{E}[X_s X_t] = \begin{cases} \mathbf{E}\left[Y_{\frac{s+1}{2}} Y_{\frac{t+1}{2}}\right] = \mathbf{E}\left[Y_{\frac{s+1}{2}}\right]\mathbf{E}\left[Y_{\frac{t+1}{2}}\right] = 0 & \text{if } s,t \in \mathbb{D} \\ \mathbf{E}\left[Y_{\frac{s}{2}} Z Y_{\frac{t+1}{2}}\right] = \mathbf{E}\left[Y_{\frac{s}{2}}\right]\mathbf{E}\left[Y_{\frac{t+1}{2}}\right]\mathbf{E}[Z] = 0 & \text{if } s \in \mathbb{E}, t \in \mathbb{D} \\ \mathbf{E}\left[Y_{\frac{s+1}{2}} Y_{\frac{t}{2}} Z\right] = \mathbf{E}\left[Y_{\frac{s+1}{2}} Y_{\frac{t}{2}}\right]\mathbf{E}[Z] = 0 & \text{if } s \in \mathbb{D}, t \in \mathbb{E} \\ \mathbf{E}\left[Y_{\frac{s}{2}} Z Y_{\frac{t}{2}} Z\right] = \mathbf{E}\left[Y_{\frac{s}{2}}\right]\mathbf{E}\left[Y_{\frac{t}{2}}\right]\mathbf{E}\left[Z^2\right] = 0 & \text{if } s,t \in \mathbb{E} \end{cases}. \end{aligned}$$

To show that $(X_t)_{t\in\mathbb{N}}$ is not strongly stationary, consider

$$\mathbf{P}(X_1 \leqslant a, X_2 \leqslant b) \quad \text{and} \quad \mathbf{P}(X_2 \leqslant a, X_3 \leqslant b),$$

on varying of $a, b \in \mathbb{R}$. We have

$$\begin{aligned}
&\mathbf{P}(X_1 \leqslant a, X_2 \leqslant b) \\
&= \mathbf{P}(Y_1 \leqslant a, Y_1 Z \leqslant b) \\
&= \mathbf{P}(Y_1 \leqslant a, Y_1 Z \leqslant b \mid Z = 1)\, \mathbf{P}(Z = 1) + \mathbf{P}(Y_1 \leqslant a, Y_1 Z \leqslant b \mid Z = -1)\, \mathbf{P}(Z = -1) \\
&= \mathbf{P}(Y_1 \leqslant a, Y_1 \leqslant b)\, \mathbf{P}(Z = 1) + \mathbf{P}(Y_1 \leqslant a, -Y_1 \leqslant b)\, \mathbf{P}(Z = 1) \\
&= \frac{1}{2}\left(\mathbf{P}(Y_1 \leqslant a \wedge b) + \mathbf{P}(-b \leqslant Y_1 \leqslant a)\right)
\end{aligned}$$

and

$$\begin{aligned}
&\mathbf{P}(X_2 \leqslant a, X_3 \leqslant b) \\
&= \mathbf{P}(Y_1 Z \leqslant a, Y_2 \leqslant b) \\
&= \mathbf{P}(Y_1 Z \leqslant a, Y_2 \leqslant b \mid Z = 1)\, \mathbf{P}(Z = 1) + \mathbf{P}(Y_1 Z \leqslant a, Y_2 \leqslant b \mid Z = -1)\, \mathbf{P}(Z = -1) \\
&= \mathbf{P}(Y_1 \leqslant a, Y_2 \leqslant b)\, \mathbf{P}(Z = 1) + \mathbf{P}(-Y_1 \leqslant a, Y_2 \leqslant b)\, \mathbf{P}(Z = -1) \\
&= \frac{1}{2}\left(\mathbf{P}(Y_1 \leqslant a, Y_2 \leqslant b) + \mathbf{P}(-Y_1 \leqslant a, Y_2 \leqslant b)\right) \\
&= \frac{1}{2}\left(\mathbf{P}(Y_1 \leqslant a)\, \mathbf{P}(Y_2 \leqslant b) + \mathbf{P}(Y_1 \leqslant a)\, \mathbf{P}(Y_2 \leqslant b)\right) \\
&= \mathbf{P}(Y_1 \leqslant a)\, \mathbf{P}(Y_2 \leqslant b).
\end{aligned}$$

Therefore, if the process were strongly stationary, we should have

$$\frac{1}{2}\left(\mathbf{P}(Y_1 \leqslant a \wedge b) + \mathbf{P}(-b \leqslant Y_1 \leqslant a)\right) = \mathbf{P}(Y_1 \leqslant a)\, \mathbf{P}(Y_2 \leqslant b)$$

for all $a, b \in \mathbb{R}$. On the other hand, choosing $a = b = 1$, we have

$$\begin{aligned}
\frac{1}{2}\left(\mathbf{P}(Y_1 \leqslant a \wedge b) + \mathbf{P}(-b \leqslant Y_1 \leqslant a)\right) &= \frac{1}{2}\left(\mathbf{P}(Y_1 \leqslant 1) + \mathbf{P}(-1 \leqslant Y_1 \leqslant 1)\right) \\
&= \frac{1}{2}\left(\Phi(1) + (\Phi(1) - \Phi(-1))\right) \\
&= \frac{1}{2}(0.8413 + 0.6827) = 0.7620
\end{aligned}$$

and

$$\mathbf{P}(Y_1 \leqslant a)\, \mathbf{P}(Y_2 \leqslant b) = \mathbf{P}(Y_1 \leqslant 1)\, \mathbf{P}(Y_2 \leqslant 1) = \Phi(1)^2 = 0.7079.$$

These prevent the process to be strongly stationary. □

Proposizione 391 *Assume that the random variables in X are uncorrelated and that for a suitable $(\mu, \Sigma) \in \mathbb{R}^m \times \mathbb{R}^{m^2}$ we have $\mathbf{E}[X_t] = \mu$ and $\Sigma_X(t) = \Sigma$, for every $t \in \mathbb{T}$. Then X is a weakly stationary process.*

Esempio 392 *Let $m = 1$ and let $(X_t)_{t\in\mathbb{N}} \equiv X$ be a real stochastic process on Ω of independent random variables with generalized normal distribution having location parameter $\mu \equiv 0$, scale parameter $\alpha_t \equiv \sqrt{\Gamma(1/\beta_t)/\Gamma(3/\beta_t)}$, and shape parameter $\beta_t \equiv (3t-2)/t$, on varying of $t \in \mathbb{N}$. Then X is a weakly stationary process which is not strongly stationary.*

Discussion. It is sufficent to observe that the density function $f_{X_t}:\mathbb{R}\to\mathbb{R}_+$ of a random variable X_t in X having location parameter $\mu=0$ is given by

$$f_{X_t}(x)\overset{\text{def}}{=}\frac{1}{2}\operatorname{sgn}(x)\frac{\gamma\left(\frac{1}{\beta_t},\left(\frac{|x|}{\alpha_t}\right)^{\beta_t}\right)}{2\Gamma(1/\beta_t)},\quad\forall x\in\mathbb{R},$$

where γ denotes the lower incomplete gamma function, and that the mean μ_{X_t} and variance $\sigma^2_{X_t}$ of X_t are given by

$$\mu_{X_t}=0,\quad\sigma^2_{X_t}=\alpha_t^2\frac{\Gamma(3/\beta_t)}{\Gamma(1/\beta_t)}=1,$$

for every $t\in\mathbb{N}$. Therefore, Proposition 391 can be invoked to state that X is weakly stationary, while the different distributions of the random variables X_t on varying of $t\in\mathbb{N}$, prevent X from being strongly stationary. □

As a consequence of Proposition 391, it is clearly seen that different processes may have the same sutocorrelation function.

Teorema 393 (Wold Representation Theorem) *Assume that the process X has mean zero, that is $\mu_X(t)=0$ for every $t\in\mathbb{Z}$. Then we can write*

$$X_t=\sum_{n=0}^{\infty}\psi_n W_{t-n}+D_t,\quad\forall t\in\mathbb{Z},$$

where $(W_t)_{t\in\mathbb{Z}}\equiv W$ is a 2nd-order real stocastic process with mean zero, $\mu_W(t)=0$ for every $t\in\mathbb{Z}$, constant variance, $\sigma^2_W(t)=\sigma^2_W>0$ for every $t\in\mathbb{Z}$, and uncorrelated random variables, $\gamma_W(s,t)=0$ for all $s,t\in\mathbb{Z}$[15]*, the sequence $(\psi_n)_{n\geqslant 0}$ is such that $\psi_0=1$ and is square summable, $\sum_{n=0}^{\infty}\psi_n^2<\infty$, and $(D_t)_{t\in\mathbb{Z}}\equiv D$ is a 2nd-order real stocastic process uncorrelated with W, that is $\mathbf{E}[D_tW_s]=0$ for all $s,t\in\mathbb{Z}$, and fulfilling*

$$\mathbf{P}[D_{t+\tau}\mid X_{t-1},X_{t-s},\dots]=D_{t+\tau},\quad\forall t\in\mathbb{Z},\tau\in\hat{n}.$$

Proof. . □

6.7.1 Processes with Wide-Sense Stationary Increments

6.7.2 Partial Autocorrelation Function

Let $(X_t)_{t\in\mathbb{Z}}\equiv X$ be a wide-sense stationary real stochastic process on a probability space Ω with time set $\mathbb{T}\equiv\mathbb{Z}$.

Definizione 394 *We call the* partial autocorrelation function *of X the map $\phi_X:\mathbb{Z}\to\mathbb{R}$ given by*

$$\phi_X(\tau)\overset{def}{=}\begin{cases}1 & \text{if } \tau=0\\ Corr(X_t,X_{t+\tau}) & \text{if } |\tau|=1\\ Corr(X_t-\mathbf{P}[X_t\mid X_{t+1},\dots,X_{t+\tau-1}],X_{t+\tau}-\mathbf{P}[X_{t+\tau}\mid X_{t+1},\dots,X_{t+\tau-1}]) & \text{if } \tau\geqslant 2\\ Corr(X_t-\mathbf{P}[X_t\mid X_{t-1},\dots,X_{t+\tau+1}],X_{t+\tau}-\mathbf{P}[X_{t+\tau}\mid X_{t-1},\dots,X_{t+\tau+1}]) & \text{if } \tau\leqslant -2\end{cases}\tag{6.47}$$

where $\mathbf{P}[X_t\mid X_{t+1},\dots,X_{t+\tau-1}]$ and $\mathbf{P}[X_t\mid X_{t-1},\dots,X_{t+\tau+1}]$ [resp. $\mathbf{P}[X_{t+\tau}\mid X_{t+1},\dots,X_{t+\tau-1}]$ and $\mathbf{P}[X_{t+\tau}\mid X_{t-1},\dots,X_{t+\tau+1}]$] are the orthogonal projection of X_t [resp. $X_{t+\tau}$] on $X_{t+1},\dots,X_{t+\tau-1}$ and $X_{t-1},\dots,X_{t+\tau+1}$, respectively, and

$$Corr(U,V)\equiv\frac{Cov(U,V)}{\mathbf{D}^2[U]\,\mathbf{D}^2[V]}\equiv\frac{\mathbf{E}[(U-\mathbf{E}[U])(V-\mathbf{E}[V])]}{\mathbf{D}^2[U]\,\mathbf{D}^2[V]}.$$

[15] Briefly, the process W is a *weak white noise* (see Definition (**??**)).

Osservazione 395 *Assume $\tau \geqslant 2$. Write*

$$\Sigma_{X_{t+1},\ldots,X_{t+\tau-1}} \equiv \mathbf{E}\left[(X_{t+1}-\mathbf{E}[X_{t+1}],\ldots,X_{t+\tau-1}-\mathbf{E}[X_{t+\tau-1}])^\intercal (X_{t+1}-\mathbf{E}[X_{t+1}],\ldots,X_{t+\tau-1}-\mathbf{E}[X_{t+\tau-1}])\right]$$

and

$$\Gamma_{X_t,(X_{t+1},\ldots,X_{t+\tau-1})} = \left(\mathbf{E}[(X_t-\mathbf{E}[X_t])(X_{t+1}-\mathbf{E}[X_{t+1}])],\ldots,\mathbf{E}[(X_t-\mathbf{E}[X_t])(X_{t+\tau-1}-\mathbf{E}[X_{t+\tau-1}])]\right)$$

[resp. $\Gamma_{X_{t+\tau},(X_{t+1},\ldots,X_{t+\tau-1})} = \left(\mathbf{E}[(X_{t+\tau}-\mathbf{E}[X_{t+\tau}])(X_{t+1}-\mathbf{E}[X_{t+1}])],\ldots,\mathbf{E}[(X_t-\mathbf{E}[X_t])(X_{t+\tau-1}-\mathbf{E}[X_{t+\tau-1}$

and assume that the matrix $\Sigma_{X_{t+1},\ldots,X_{t+\tau-1}}$ is non singular. Then we have

$$\mathbf{P}[X_t \mid X_{t+1},\ldots,X_{t+\tau-1}] = \Gamma_{X_t,(X_{t+1},\ldots,X_{t+\tau-1})}\Sigma^{-1}_{X_{t+1},\ldots,X_{t+\tau-1}}(X_{t+1},\ldots,X_{t+\tau-1})^\intercal$$

[resp. $\mathbf{P}[X_{t+\tau} \mid X_{t+1},\ldots,X_{t+\tau-1}] = \Gamma_{X_{t+\tau},(X_{t+1},\ldots,X_{t+\tau-1})}\Sigma^{-1}_{X_{t+1},\ldots,X_{t+\tau-1}}(X_{t+1},\ldots,X_{t+\tau-1})^\intercal$ *].*

Osservazione 396 *Assume $\tau \leqslant -2$. Write*

$$\Sigma_{X_{t-1},\ldots,X_{t+\tau+1}} \equiv \mathbf{E}\left[(X_{t-1}-\mathbf{E}[X_{t-1}],\ldots,X_{t+\tau+1}-\mathbf{E}[X_{t+\tau+1}])^\intercal (X_{t-1}-\mathbf{E}[X_{t-1}],\ldots,X_{t+\tau+1}-\mathbf{E}[X_{t+\tau+1}])\right]$$

and

$$\Gamma_{X_t,(X_{t-1},\ldots,X_{t+\tau+1})} = \left(\mathbf{E}[(X_t-\mathbf{E}[X_t])(X_{t-1}-\mathbf{E}[X_{t-1}])],\ldots,\mathbf{E}[(X_t-\mathbf{E}[X_t])(X_{t+\tau+1}-\mathbf{E}[X_{t+\tau+1}])]\right)$$

[resp. $\Gamma_{X_{t+\tau},(X_{t-1},\ldots,X_{t+\tau+1})} = \left(\mathbf{E}[(X_{t+\tau}-\mathbf{E}[X_{t+\tau}])(X_{t-1}-\mathbf{E}[X_{t-1}])],\ldots,\mathbf{E}[(X_t-\mathbf{E}[X_t])(X_{t+\tau+1}-\mathbf{E}[X_{t+\tau+1}$

and assume that the matrix $\Sigma_{X_{t-1},\ldots,X_{t+\tau+1}}$ is non singular. Then we have

$$\mathbf{P}[X_t \mid X_{t-1},\ldots,X_{t+\tau+1}] = \Gamma_{X_t,(X_{t-1},\ldots,X_{t+\tau+1})}\Sigma^{-1}_{X_{t-1},\ldots,X_{t+\tau+1}}(X_{t-1},\ldots,X_{t+\tau+1})^\intercal$$

[resp. $\mathbf{P}[X_{t+\tau} \mid X_{t-1},\ldots,X_{t+\tau+1}] = \Gamma_{X_{t+\tau},(X_{t-1},\ldots,X_{t+\tau+1})}\Sigma^{-1}_{X_{t-1},\ldots,X_{t+\tau+1}}(X_{t-1},\ldots,X_{t+\tau+1})^\intercal$ *].*

Teorema 397 (Durbin-Levison Recursion) *Fix any $\tau \in \mathbb{N}$ such that $\tau \geqslant 2$ and write*

$$\rho_X(s) \equiv Corr(X_t, X_{t+s}), \quad \forall s = 1,\ldots,\tau,$$

Hence, consider the vector

$$\rho_X^{(\tau)} \equiv (\rho_X(1),\ldots,\rho_X(\tau))^\intercal$$

and the matrix

$$\mathrm{P}_X(\tau) \equiv \begin{pmatrix} 1 & \rho_X(1) & \cdots & \rho_X(\tau-2) & \rho_X(\tau-1) \\ \rho_X(1) & 1 & \cdots & \rho_X(\tau-3) & \rho_X(\tau-2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_X(\tau-2) & \rho_X(\tau-3) & \cdots & 1 & \rho_X(1) \\ \rho_X(\tau-1) & \rho_X(\tau-2) & \cdots & \rho_X(1) & 1 \end{pmatrix}$$

Then the solution

$$\phi^{(\tau)} \equiv (\phi_{\tau,1},\ldots,\phi_{\tau,\tau})^\intercal$$

of the equation

$$\mathrm{P}_X(\tau)\,\phi^{(\tau)} = \rho_X^{(\tau)}$$

fulfills

$$\phi_{\tau,\tau} = \phi_X(\tau).$$

In particular, if the matrix $\mathrm{P}_X(\tau)$ is non-singular, we have

$$\phi_X(\tau) = \frac{\det(\mathrm{P}^*_X(\tau))}{\det(\mathrm{P}_X(\tau))},$$

where $\mathrm{P}^*_X(\tau)$ is a matrix with the same columns of $\mathrm{P}_X(\tau)$ but the τth column which is replaced by the vector $\rho_X^{(\tau)}$.

Osservazione 398 *Using the abbreviation*

$$\rho_X(\tau) \equiv \rho_\tau.$$

we have

$$\phi_X(1) = \rho_1, \quad \phi_X(2) = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2-\rho_1^2}{1-\rho_1^2}, \quad \phi_X(3) = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} = \frac{\rho_3\left(1-\rho_1^2\right)+\rho_1\left(\rho_1^2+\rho_2^2-2\rho_2\right)}{(1-\rho_2)\left(1+\rho_2-2\rho_1^2\right)},$$

$$\phi_X(4) = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \rho_1 \\ \rho_1 & 1 & \rho_1 & \rho_2 \\ \rho_2 & \rho_1 & 1 & \rho_3 \\ \rho_3 & \rho_2 & \rho_1 & \rho_4 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & 1 & \rho_1 & \rho_2 \\ \rho_2 & \rho_1 & 1 & \rho_1 \\ \rho_3 & \rho_2 & \rho_1 & 1 \end{vmatrix}} = \dots$$

6.8 Weakly Stationary Stochastic Processes

Let $(X)_{t\in\mathbb{T}} \equiv X$ be a 2nd-order real stochastic process on a probability space Ω.

Definizione 399 *We call* autocovariace function *of X the map $\gamma_X : \mathbb{T}\times\mathbb{T}\to\mathbb{R}$ given by*

$$\gamma_X(s,t) \overset{def}{=} Cov(X_s, X_t), \quad \forall s,t\in\mathbb{T}. \tag{6.48}$$

Definizione 400 *We call* autocorrelation *or* Pearson correlation function *of X the map $\rho_X : \mathbb{T}\times\mathbb{T}\to\mathbb{R}$ given by*

$$\rho_X(s,t) \overset{def}{=} \frac{\gamma_X(s,t)}{\mathbf{D}[X_s]\,\mathbf{D}[X_t]}, \quad \forall s,t\in\mathbb{T}. \tag{6.49}$$

Definizione 401 *We say that the process X is* weakly stationary *or* autocovariance stationary *if we have*

1. $\mathbf{E}[X_t] = \mathbf{E}[X_s]$ *for all $s,t\in\mathbb{T}$;*
2. $\gamma_X(s,t) = \gamma_X(s+h,t+h)$ *for all $s,t\in\mathbb{T}$ and every $h\in\mathbb{R}$ such that $s+h, t+h\in\mathbb{T}$.*

Proposizione 402 *If X is a weakly stationary process, then given any $t_0\in\mathbb{T}$, we have*

$$\mathbf{D}^2[X_t] = \mathbf{D}^2[X_{t_0}], \tag{6.50}$$

for every $t\in\mathbb{T}$.

Proof. Given any $t_0\in\mathbb{T}$, for every $t\in\mathbb{T}$ consider $h\equiv t-t_0$. Then, under the assumption of weakly covariance, we can write

$$\mathbf{D}^2[X_{t_0}] \equiv \gamma_X(t_0,t_0) = \gamma_X(t_0+(t-t_0), t_0+(t-t_0)) = \gamma_X(t,t) \equiv \mathbf{D}^2[X_t],$$

as claimed. □

Osservazione 403 *If X is a weakly stationary process, then given any $t_0 \in \mathbb{T}$, we have*

$$\mathbf{D}[X_t] = \mathbf{D}[X_{t_0}],$$

for every $t \in \mathbb{T}$.

Notation 404 *If X is a weakly stationary process, then we write μ_X [resp. σ_X] for the constant values of $\mathbf{E}[X_t]$ [resp. $\mathbf{D}[X_t]$] on varying of $t \in \mathbb{T}$.*

Proposizione 405 *If X is a weakly stationary process, then given any $t_0 \in \mathbb{T}$, we have*

$$\gamma_X(s,t) = \gamma_X(t_0, t_0 + (t-s)), \tag{6.51}$$

and

$$\rho_X(s,t) = \frac{\mathbf{E}\left[(X_{t_0} - \mu_X)\left(X_{t_0+(t-s)} - \mu_X\right)\right]}{\sigma_X^2}. \tag{6.52}$$

for all $s, t \in \mathbb{T}$ such that $t_0 + (t-s) \in \mathbb{T}$.

Proof. Given any $t_0 \in \mathbb{T}$, for all $s, t \in \mathbb{T}$ consider $h \equiv t_0 - s$. Then, under the assumption of weakly covariance, we have

$$\gamma_X(s,t) = \gamma_X(s + (t_0 - s), t + (t_0 - s)) = \gamma_X(t_0, t_0 + (t-s)),$$

which proves (6.51). In addition, we can write

$$\mathbf{D}[X_t] = \mathbf{D}[X_s] \equiv \sigma_X \quad \text{and} \quad \mathbf{E}\left[X_{t_0+(t-s)}\right] \equiv \mu_X.$$

Therefore,

$$\rho_X(s,t) = \frac{\gamma_X(s,t)}{\mathbf{D}[X_s]\,\mathbf{D}[X_t]} = \frac{\gamma_X(t_0, t_0 + (t-s))}{\sigma_X^2} \tag{6.53}$$

and

$$\gamma_X(t_0, t_0 + (t-s)) = cov\left(X_{t_0}, X_{t_0+(t-s)}\right) = \mathbf{E}\left[(X_{t_0} - \mathbf{E}[X_{t_0}])\left(X_{t_0+(t-s)} - \mathbf{E}\left[X_{t_0+(t-s)}\right]\right)\right] = \mathbf{E}\left[(X_{t_0} - \mu_X)\left(X_{t_0+}\right.\right. \tag{6.54}$$

Combining (6.53) and (6.54) the desired (**??**) immediately follows. □

Definizione 406 *If X is a weakly stationary process, given any $t_0 \in \mathbb{T}$, write*

$$\mathbb{T}_0 \equiv \{\tau \in \mathbb{R} : t_0 + \tau \in \mathbb{T}\}$$

we call the map $\gamma_{X,t_0} : \mathbb{T}_0 \to \mathbb{R}$ given by

$$\gamma_{X,t_0}(\tau) \stackrel{def}{=} \gamma_X(t_0, t_0 + \tau), \quad \forall \tau \in \mathbb{T}, \tag{6.55}$$

the reduced autocovariace function *of X referred to t_0 and we call the map $\rho_{X,t_0} : \mathbb{T}_0 \to \mathbb{R}$ given by*

$$\rho_{X,t_0}(\tau) \stackrel{def}{=} \rho_X(t_0, t_0 + \tau), \quad \forall \tau \in \mathbb{T}, \tag{6.56}$$

the reduced autocorrelation function *of X referred to t_0.*

Osservazione 407 *If X is a 2nd-order stationary process, then X is weakly stationary.*

Osservazione 408 *Let $(X_t)_{t \geqslant 1} \equiv X$ be a sequence of N-dimensional uncorrelated random vectors on Ω having finite 2nd-order moment such that for a suitable $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+$ we have $\mathbf{E}[X_t] = \mu$ and $\mathbf{D}[X_t] = \sigma$ for every $t \geqslant 1$. Then X is a weakly stationary process.*

Osservazione 409 *Let $(X_t)_{t \geqslant 1} \equiv X$ be a sequence of uncorrelated and identically distributed N-dimensional random vectors on Ω having finite 2nd-order moment. Then X is a weakly stationary process.*

6.9 Gaussian Processes

Let $(X)_{t\in\mathbb{T}} \equiv X$ be a real stochastic process on a probability space $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$, and let $\mathcal{P}_{fin}(\mathbb{T})$ be the set of all multi-indices on the time set $\mathbb{T}$.

Definizione 410 *We say that the process X is Gaussian if the random variables $X_{t_1}, \ldots, X_{t_L}$ are jointly normally distributed, for every $(t_\ell)_{\ell=1}^L \in \mathcal{P}_{fin}(\mathbb{T})$, that is to say, the random variable $\sum_{\ell=1}^L c_\ell X_{t_\ell}$ is normally distributed*[16]*, for every $(t_\ell)_{\ell=1}^L \in \mathcal{P}_{fin}(\mathbb{T})$ and every $(c_1, \ldots, c_L) \in \mathbb{R}^L$.*

Proposizione 411 *The process X is Gaussian if and only if, for every $t \in \mathcal{P}_{fin}(\mathbb{T})$, $t \equiv (t_\ell)_{\ell=1}^L$, there exist independent standard Gaussian random variables $Z_1, \ldots, Z_{K_t}$, for some $K_t \in \mathbb{N}$, a vector $\mu_t \in \mathbb{R}^L$, and a matrix $A_t \in \mathbb{R}^{L\times K}$, generally all depending on $\not{t}$, such that*

$$(X_{t_1}, \ldots, X_{t_L})^\intercal = \mu_t^\intercal + A_t \ (Z_1, \ldots Z_{K_t})^\intercal . \tag{6.57}$$

Proof. See (**?**). □

Proposizione 412 *The process X is Gaussian if and only if, for every $t \in \mathcal{P}_{fin}(\mathbb{T})$, $t \equiv (t_\ell)_{\ell=1}^L$, there exist a vector $\mu_t \in \mathbb{R}^L$, $\mu_t \equiv \left(\mu_1^{(t)}, \ldots, \mu_L^{(t)}\right)^\intercal$ and a non-negative definite symmetric matrix $\Sigma_t \in \mathbb{R}^{L^2}$, $\Sigma_t \equiv \left(\sigma_{k,\ell}^{(t)}\right)_{k,\ell=1}^L$, generally depending on $\not{t}$, such that the characteristic function $\varphi_{X_{t_1},\ldots,X_{t_L}} : \mathbb{R}^L \to \mathbb{C}$ of the random vector $(X_{t_1}, \ldots, X_{t_L})^\intercal$ is given by*

$$\varphi_{X_{t_1},\ldots,X_{t_L}}(u_1, \ldots, u_L) = \exp\left(-\frac{1}{2}\textstyle\sum_{k,\ell=1}^L \sigma_{k,\ell}^{(t)} u_k u_\ell + i \sum_{\ell=1}^L \mu_\ell^{(t)} u_\ell\right), \tag{6.58}$$

for every $(u_1, \ldots, u_L) \in \mathbb{R}^L$.

Proof. See (**?**). □

Proposizione 413 *If the process X is Gaussian, with reference to the notation of Propositions 411 and 412 we have*

1. *$\mu_\ell^{(t)} \equiv \mathbf{E}[X_{t_\ell}]$, for every $\ell = 1, \ldots, L$;*
2. *$\sigma_{k,\ell}^{(t)} \equiv \mathbf{E}[(X_{t_k} - \mu_k)(X_{t_\ell} - \mu_\ell)]$, for all $k, \ell = 1, \ldots, L$;*
3. *$\Sigma_t = A_t A_t^\intercal$.*

for every $t \in \mathcal{P}_{fin}(\mathbb{T})$, $t \equiv (t_\ell)_{\ell=1}^L$.

Proof. See (**?**). □

Proof. . □

Definizione 414 *We say that a Gaussian process X is non-degenerate if the autocovariance matrix $\Sigma_t \equiv \left(\sigma_{k,\ell}^{(t)}\right)_{k,\ell=1}^L$ of the random variables $X_{t_1}, \ldots, X_{t_L}$ is positive definite, for every $t \in \mathcal{P}_{fin}(\mathbb{T})$, $t \equiv (t_\ell)_{\ell=1}^L$.*

[16] There exist $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}_+$ such that

$$\sum_{k=1}^n c_k X_{t_k} \sim N(\mu, \sigma^2),$$

where by a normal distribution with mean μ and $\sigma^2 = 0$ we mean the Dirac distribution concentrated in μ.

Proposizione 415 *If X is a non-degenerate Gaussian process, then the random variables $X_{t_1},\dots,X_{t_L}$ are jointly abolutely continuous with joint density $f_{X_{t_1},\dots,X_{t_L}}:\mathbb{R}^L\to\mathbb{R}_+$ given by*

$$f_{X_{t_1},\dots,X_{t_L}}(x_1,\dots,x_L)=\frac{1}{\sqrt{(2\pi)^L\det(\Sigma_t)}}\exp\left(-\frac{1}{2}(x-\mu_t)^\intercal\Sigma_t^{-1}(x-\mu_t)\right), \tag{6.59}$$

for every $x\in\mathbb{R}^L$, $x\equiv(x_1,\dots,x_L)$, where, for every $t\in\mathcal{P}_{fin}(\mathbb{T})$, $t\equiv(t_\ell)_{\ell=1}^L$, the vector $\mu_t\in\mathbb{R}^L$, $\mu_t\equiv\left(\mu_1^{(t)},\dots,\mu_L^{(t)}\right)^\intercal$ [resp. the matrix $\Sigma_t\in\mathbb{R}^{L^2}$, $\Sigma_t\equiv\left(\sigma_{k,\ell}^{(t)}\right)_{k,\ell=1}^L$], generally depending on t, is given by (1) [resp. (2)] in Proposition 413.

Proof. See (?). □

A useful result to generate Gaussian processes from Gaussian processes is the following

Proposizione 416 *Let $(Y)_{t\in\mathbb{T}}\equiv Y$ a real process on a probability space Ω. Assume that there exists a Gaussian process $(X)_{t\in\mathbb{S}}\equiv X$ on Ω such that for every $t\in\mathcal{P}_{fin}(\mathbb{T})$, $t\equiv(t_\ell)_{\ell=1}^L$, there exist a vector $b_t\in\mathbb{R}^L$, a full rank matrix $A_t\in\mathbb{R}^{L\times K}$, and a multi-index $s_t\in\mathcal{P}_{fin}(\mathbb{S})$, $s_t\equiv\left(s_k^{(t)}\right)_{k=1}^K$, generally depending on t, such that*

$$(Y_{t_1},\dots,Y_{t_L})^\intercal=b_t+A_t\left(X_{s_1^{(t)}},\dots,X_{s_k^{(t)}}\right)^\intercal.$$

Then Y is a Gaussian process.

Proof. See (?). □

As a first important consequence of Proposition 416, we have

Proposizione 417 *If a process X is Gaussian then X has Gaussian distributed increments.*

Proof. For any $t\in\mathcal{C}_{fin}(\mathbb{T})$, $t\equiv(t_\ell)_{\ell=1}^{L+1}$, we have

$$\left(X_{t_2}-X_{t_1},X_{t_3}-X_{t_2},\dots,X_{t_L}-X_{t_{L-1}},X_{t_{L+1}}-X_{t_L}\right)=A_t\left(X_{t_1},X_{t_2},X_{t_3}\dots,X_{t_L-1},X_{t_{L+1}},X_{t_{L+1}}\right)^\intercal$$

where the matrix $A_t\in\mathbb{R}^{L\times(L+1)}$ is given by

$$A_t=\begin{pmatrix}-1 & 1 & 0 & ... & 0 & 0 & 0\\ 0 & -1 & 1 & ... & 0 & 0 & 0\\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots\\ 0 & 0 & 0 & \cdots & -1 & 1 & 0\\ 0 & 0 & 0 & \cdots & 0 & -1 & 1\end{pmatrix}.$$

Now, we clearly have $\operatorname{rank}(A_t)=L$. Therefore, thanks to Proposition 416, from the Gaussianity of X the Gaussianity of any vector of increments follows. □

Proposizione 418 *If X is a wide sense stationary Gaussian process, then X is strict sense stationary.*

Proof. It is sufficient to observe that when X is a Gaussian process, then its marginal distributions are completely determined by means and variance-covariance matrices (see e.g. G. Lindgreen, H. Rootzén, M. Sandsten, Stationary Stochastic Processes for Scientists and Engineers, CRC press, § 1.3.2, p. 14, see also Flandoli AUTCap2 p. ix). □

6.10 Markov Processes

The definition of Markov process aims to formalize the idea that the probability distribution of the states of an observable of a stochastic phenomenon, which are future with respect to a current time, depends only on the current state and is not influenced by the past states. Accordingly, a stochastic process is said to be a *Markov process* if the probability distribution of the future states of the process given all information accumulated from the past up to the current time is the same as the probability distribution given the current state. Only the knowledge of the current state of the process is useful to forecast the future paths of a Markov process, the past history plays no role. Otherwise saying, a Markov process has no memory of the past.

Stock prices are often modelled as Markov processes, on the ground of Eugen Fama's Efficient Market Hypothesis (e.g. see (**?**)): the price sensitive information which becames progressively available in the market is quickly incorporated in the price of the stocks, so that the past pattern of the prices has no forecasting value[17]. However, under suitable assumptions of stationarity the history of the stock prices is still useful to determine some characteristics of the modeling Markov processes, for instance the average return, the volatility, and the other so-called "Greeks".

Let $(X_t)_{t\in\mathbb{T}} \equiv X$ be a stochastic process on a probability space $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ with states in a measurable space $(\mathbb{X}, \mathcal{M}) \equiv \mathbb{X}$ and let $(\mathcal{F}_t)_{t\in\mathbb{T}} \equiv \mathfrak{F}$ be a filtration on Ω.

Definizione 419 *We say that X is an $\mathfrak{F}$-*Markov process *or that the process X is $\mathfrak{F}$-*Markovian *if X is $\mathfrak{F}$-adapted and for all $s, t \in \mathbb{T}$ such that $s < t$ we have*

$$\mathbf{P}(X_t \in M \mid \mathcal{F}_s) \overset{\mathbf{P}\text{-}a.s.}{=} \mathbf{P}(X_t \in M \mid X_s), \quad \forall M \in \mathcal{M},$$

where $\mathbf{P}(X_t \in M \mid \mathcal{F}_s)$ [resp. $\mathbf{P}(X_t \in M \mid X_s)$] stands for the conditional probability of the event $\{X_t \in M\}$ given the σ-algebra $\mathcal{F}_s$ [resp. the σ-algebra $\sigma(X_s)$ generated by the random variable X_s]. The reference to the filtration $\mathfrak{F}$ is usually omitted when $\mathfrak{F}$ is intended to be the filtration $\mathfrak{F}^X$ generated by the process.

Proposizione 420 *The process X is $\mathfrak{F}$-Markovian if and only if X is $\mathfrak{F}$-adapted and for every $(t_k)_{k=1}^L \in \mathcal{C}_{fin}(\mathbb{T})$ we have*

$$\mathbf{P}\left(X_{t_L} \in M \mid X_{t_1}, \dots, X_{t_{L-1}}\right) \overset{\mathbf{P}\text{-}a.s.}{=} \mathbf{P}\left(X_{t_n} \in M \mid X_{t_{L-1}}\right), \quad \forall M \in \mathcal{M},$$

where $\mathbf{P}\left(X_{t_L} \in M \mid X_{t_1}, \dots, X_{t_{L-1}}\right)$ [resp. $\mathbf{P}\left(X_{t_L} \in M \mid X_{t_{L-1}}\right)$] stands for the conditional probability of the event $\{X_{t_L} \in M\}$ given the σ-algebra $\sigma\left(X_{t_1}, \dots, X_{t_{L-1}}\right)$ generated by the random variables $X_{t_1}, \dots, X_{t_{L-1}}$ [resp. the σ-algebra $\sigma\left(X_{t_{L-1}}\right)$ generated by the random variable $X_{t_{L-1}}$].

Esempio 421 *Let $\mathbb{T} \equiv \mathbb{N}$, and let $X \equiv (X_n)_{n\geqslant 0}$ be a sequence of independent random variables on Ω with states in $\mathbb{X}$. Then X is a Markov process.*

Proof. The independence of the random variables of the sequence $(X_n)_{n\geqslant 0}$ implies that, for every $M \in \mathcal{M}$, the event $\{X_n \in M\}$ is independent of the σ-field $\sigma(X_1, \dots, X_m)$ for every $m \geqslant 0$ such that $m < n$. Therefore, we have trivially

$$\mathbf{P}(X_n \in M \mid X_1, \dots, X_m) \overset{\mathbf{P}\text{-a.s.}}{=} \mathbf{P}(X_n \in M) \overset{\mathbf{P}\text{-a.s.}}{=} \mathbf{P}(X_n \in M \mid X_m),$$

as desired. □

The following functional characterization of Markov processes deserves to be reported.

Let $L^\infty(\mathbb{X})$ be the Banach space of all bounded Borel-measurable complex-valued functions on $\mathbb{X}$ equipped with the usual supremum norm $\|\cdot\|_\infty : L^\infty(\mathbb{X}) \to \mathbb{R}_+$ given by

$$\|f\|_\infty \overset{\text{def}}{=} \sup_{x\in\mathbb{X}} \{|f(x)|\}, \quad \forall f \in L^\infty(\mathbb{X}).$$

[17] In contrast, the advocates of technical analysis aim to forecast the future pattern of stock prices relying on the past one.

Teorema 422 *The process X is $\mathfrak{F}$-Markovian if and only if X is $\mathfrak{F}$-adapted and for all $s,t \in \mathbb{T}$ such that $s < t$ we have*

$$\mathbf{E}\left[f\left(X_t\right) \mid \mathcal{F}_s\right] \overset{\mathbf{P}\text{-}a.s.}{=} \mathbf{E}\left[f\left(X_t\right) \mid X_s\right], \quad \forall f \in L^{\infty}\left(\mathbb{X}\right), \quad \forall s,t \in \mathbb{T},\ s < t, \tag{6.60}$$

where $\mathbf{E}\left[f\left(X_t\right) \mid \mathcal{F}_s\right]$ [resp. $\mathbf{E}\left[f\left(X_t\right) \mid X_s\right]$] stands for the conditional expectation of the random variable $f \circ X_t$ given the σ-algebra $\mathcal{F}_s$ [resp. $\sigma\left(X_s\right)$].

Proof. see Chung 1.1 $\square$

6.11 Martingales and Semimartingales

The definition of martingale aims to formalize the idea that the probability distribution of the future measurements of some observable of a stochastic phenomenon is given by the current probability distribution of the observable. In light of this, a stochastic process is called a *martingale* if the probability distribution of the future states of the process conditional on all information accumulated from the past up to the current time, is given by the probability distribution of the current states. Otherwise saying, the current expectation of the future states of the process conditional to all information accumulated from the past up to the current time, is the current state.

Note that idea of martingale is slightly but significantly different from the idea of Markov process. In fact, while the future probability distribution of the states of a Markov process conditional on the past information is its future probability distribution conditional on the present information, the future probability distribution of the states of a martingale conditional on the past information is the present probability distribution.

Martingales have been widely exploited as models for "fair betting", namely sequences of bets for which the knowledge of the past outcomes doesn't help to improve the future performances of the gambler. In addition, the idea of martingale plays a central role in many models of financial markets, which show how suitably detrended stock prices turn to be martingales for risk-neutral investors.

NEW

Let $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ be a complete probability space, let $(X_t)_{t\in\mathbb{T}} \equiv X$ be real stochastic process on Ω, and let $(\mathcal{F}_t)_{t\in\mathbb{T}} \equiv \mathfrak{F}$ be a filtration on Ω fulfilling the usual conditions.

Definizione 423 *We say that X is an $\mathfrak{F}$-*martingale*, if X is an $\mathfrak{F}$-adapted first order process and for all $s,t \in \mathbb{T}$ such that $s < t$ we have*

$$\mathbf{E}\left[X_t \mid \mathcal{F}_s\right] \overset{\mathbf{P}\text{-}a.s.}{=} X_s. \tag{6.61}$$

We will omit the reference to the filtration $\mathfrak{F}$ when $\mathfrak{F}$ is intended to be the filtration $\mathfrak{F}^X$ generated by the process.

It is easy to check that

Proposizione 424 *The process X is an $\mathfrak{F}$-martingale, if and only if X is an $\mathfrak{F}$-adapted first order process and for all $s,t \in \mathbb{T}$ such that $s < t$ we have*

$$\mathbf{E}\left[X_t 1_F\right] = \mathbf{E}\left[X_s 1_F\right], \quad \forall F \in \mathcal{F}_s. \tag{6.62}$$

Osservazione 425 *If the process X is an $\mathfrak{F}$-martingale, we have*

$$\mathbf{E}\left[X_t\right] = \mathbf{E}\left[X_s\right], \quad \forall s,t \in \mathbb{T} \ s.t.\ s < t.$$

Definizione 426 *We say that X is an $\mathfrak{F}$-*submartingale *[resp. $\mathfrak{F}$-*supermartingale*], if X is an $\mathfrak{F}$-adapted first order process and for all $s,t \in \mathbb{T}$ such that $s < t$ we have*

$$\mathbf{E}[X_t \mid \mathcal{F}_s] \overset{\mathbf{P}\text{-}a.s.}{\geqslant} X_s \qquad \text{[resp. } \mathbf{E}[X_t \mid \mathcal{F}_s] \overset{\mathbf{P}\text{-}a.s.}{\leqslant} X_s\text{].} \tag{6.63}$$

Also in this case we will omit the reference to the filtration $\mathfrak{F}$ when $\mathfrak{F}$ is intended to be the filtration $\mathfrak{F}^X$ generated by the process.

In contrast to martingale, *submartingales* [resp. a *supermartingale*] are are processes to model the idea that information on a stochastic phenomenon, accumulated by some observer from the past up to the current time increases [resp. decreases] the observer's current expectations on her future measurements of some observable of the phenomenon, with respect to her current measurements. Therefore, a submartingale [resp. supermartingale] is a natural model for "favorable betting", [resp. "unfavorable betting"] which are sequences of bets such that the knowledge of the past outcomes of the bet increases [resp. decreases] the future performances of the gambler.

Proposizione 427 *Assume that X is an $\mathfrak{F}$-adapted first order non-decreasing [resp. non-increasing] process. Then X is an $\mathfrak{F}$-*submartingale *[resp. an $\mathfrak{F}$-*supermartingale].

Proof. We only need to prove that **??** of Definition 426 holds true. In fact, since the process X is non-decreasing [resp. non-increasing], thanks to the monotonicity and stability properties of the conditional expectation operator, we have

$$\mathbf{E}[X_t \mid \mathcal{F}_s] \geqslant \mathbf{E}[X_s \mid \mathcal{F}_s] = X_s \quad \text{[resp. } \mathbf{E}[X_t \mid \mathcal{F}_s] \leqslant \mathbf{E}[X_s \mid \mathcal{F}_s] = X_s\text{], a.s. on } \Omega \text{, whenever } 0 \leqslant s \leqslant t,$$

as desired. □

Esempio 428 *Let X be a real random variable with finite second order moment on a probability space $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ and let $(Z_n)_{n\geqslant 1}$ be a sequence of independent centered random variables with finite second order moment which are independent of X. Consider the sequence of random variables $(X_n)_{n\geqslant 1}$ given by*

$$X_1 \overset{def}{=} X, \quad X_2 \overset{def}{=} X_1 + Z_1 X, \quad X_{n+1} \overset{def}{=} X_n + X \sum_{k=1}^{n} Z_k, \quad \forall n \geqslant 2$$

we have

$$\mathbf{E}[X_{n+1} \mid \mathcal{F}_n] = \mathbf{E}[X_{n+1} \mid X_1, \ldots, X_n] = \mathbf{E}\left[X_n + X \sum_{k=1}^{n} Z_k \mid X_1, \ldots, X_n\right]$$

6.12 Brownian Motion

Let $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ be a complete probability space, let $\mathbb{T}$ be an interval of $\mathbb{R}_+$ with first element 0, let $(\mathcal{F}_t)_{t\in\mathbb{T}} \equiv \mathfrak{F}$ be a filtration on Ω fulfilling the usual conditions, and let $(X_t)_{t\in\mathbb{T}} \equiv X$ be a stochastic process on Ω with states in $\mathbb{R}^N$

Definizione 429 *We say that X is a K-*dimensional $\mathfrak{F}$-Brownian motion *if there exist $\mu \in \mathbb{R}^K$ and a nonsingular K-order matrix A such that we have:*

1. *the process X is $\mathfrak{F}$-adapted;*
2. *the increment $X_t - X_s$ is independent of the σ-algebra $\mathcal{F}_s$, for all $s,t \in \mathbb{T}$ such that $s < t$;*

3. *the increment $X_t - X_s$ is normally distributed with mean $(t-s)\mu$ and autocovariance matrix $(t-s)\Sigma$*[18]*, where $\Sigma \equiv AA^{\intercal}$, for all $s,t \in \mathbb{T}$ such that $s < t$;*

4. *the initial distribution of the process is the Dirac measure in $\mathbb{R}^K$ concentrated at some $x_0 \in \mathbb{R}^K$*[19]*;*

5. *the process X is a.s. continuous.*

We will omit the reference to the filtration $\mathfrak{F}$ when $\mathfrak{F}$ is intended to be the filtration $\mathfrak{F}^X$ generated by the process.

Definizione 430 *A K-dimensional Brownian motion with drift vector $\mu \equiv 0$, diffusion matrix $\Sigma \equiv I_K$, where I_K is the K-order identity matrix, and initial distribution concentrated at* 0 *is called a* standard K-dimensional Brownian motion *or a K-dimensional* Wiener process.

Teorema 431 *[Lévy] Assume that the process X satisfies the following conditions:*

1. *the increment $X_t - X_s$ is independent of the σ-algebra $\mathcal{F}_s^X$ in the filtration $\mathfrak{F}^X$ generated by the process, for all $s,t \in \mathbb{T}$ such that $s < t$;*

2. *X has stationary increments;*

3. *X is almost surely continuous;*

4. *the initial distribution of the process is the Dirac measure in $\mathbb{R}^K$ concentrated at some $x_0 \in \mathbb{R}^K$.*

Then there exist $\mu \in \mathbb{R}^K$ and a nonsingular K-order matrix A, such that the increment $X_t - X_s$ is normally distributed with mean $(t-s)\mu$ and covariance matrix $(t-s)\Sigma$, where $\Sigma \equiv AA^{\intercal}$, for all $s,t \in \mathbb{T}$ such that $s < t$.

6.12.1 Brownian Motion as a Markov Process

Let $(B_t)_{t\geqslant 0} \equiv B$ be a K-dimensional $\mathfrak{F}$-Brownian motion on Ω.

Teorema 432 *The process B is $\mathfrak{F}$-Markovian.*

Proof. For all $s,t \in \mathbb{T}$ such that $s < t$ the increment $B_t - B_s$ is independent of the σ-algebra $\mathcal{F}_s$ and the random variable B_s is $\mathcal{F}_s$-measurable. Therefore, thanks to Karatzas & Shevre (**?**, 5.9 p.74), for every $M \in \mathcal{B}\left(\mathbb{R}^K\right)$ we can write

$$\begin{aligned}\mathbf{P}\left(B_t \in M \mid \mathcal{F}_s\right) &= \mathbf{P}\left(\left(B_t - B_s\right) + B_s \in M \mid \mathcal{F}_s\right) \\ &= \mathbf{P}\left(\left(B_t - B_s\right) + B_s \in M \mid B_s\right) \\ &= \mathbf{P}\left(B_t \in M \mid B_s\right),\end{aligned}$$

as desired. □

[18] The distribution function $F_{X_t - X_s} : \mathbb{R}^K \to \mathbb{R}$ of the random variable $X_t - X_s$ is given by

$$F_{X_t-X_s}(x) \stackrel{\text{def}}{=} \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_K} f\left(t-s; x_0, y\right) dy_1 \dots dy_K, \quad \forall x \in \mathbb{R}^K,$$

where

$$f\left(t; x_0, y\right) \stackrel{\text{def}}{=} \frac{1}{(2\pi t)^{K/2} (\det \Sigma)^{1/2}} \exp\left(\frac{-\left(y - x_0 - t\mu\right)\Sigma^{-1}\left(y - x_0 - t\mu\right)^{\intercal}}{2t}\right), \quad \forall t \geqslant 0, y \in \mathbb{R}^K.$$

[19] We recall that the Dirac measure on $\mathbb{R}^K$ concentrated at $x_0 \in \mathbb{R}^K$ is the map $D_{x_0} : \mathcal{B}\left(\mathbb{R}^d\right) \to \mathbb{R}_+$ given by

$$D_{x_0}(\mathsf{B}) = \begin{cases} 1, & \text{if } x_0 \in B \\ 0, & \text{if } x_0 \notin B \end{cases} \quad \forall B \in \mathcal{B}\left(\mathbb{R}^K\right).$$

6.12.2 Brownian Motion as a Martingale

Let $(B_t)_{t\geqslant 0} \equiv B$ be a K-dimensional $\mathfrak{F}$-Brownian motion on Ω.

Teorema 433 *Assume that B has drift vector 0. Then B is an $\mathfrak{F}$-martingale.*

Proof. For all $s, t \in \mathbb{T}$ such that $s < t$ the increment $B_t - B_s$ is independent of the σ-algebra $\mathcal{F}_s$ and has mean 0. Furthermore, the random variable B_s is $\mathcal{F}_s$-measurable. We then have

$$\begin{aligned} \mathbf{E}\left[B_t | \mathcal{F}_s\right] &= \mathbf{E}\left[(B_t - B_s) + B_s | \mathcal{F}_s\right] \\ &= \mathbf{E}\left[B_t - B_s | \mathcal{F}_s\right] + \mathbf{E}\left[B_s | \mathcal{F}_s\right] \\ &= \mathbf{E}\left[B_t - B_s\right] + B_s \\ &= B_s, \end{aligned}$$

which proves our claim. □

6.13 Itō Stochastic Integral

6.13.1 Basic About Stochastic Integration

Let $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ be a complete probability space, let $\mathbb{T}$ be an interval of $\mathbb{R}_+$ with first element 0, let $(\mathcal{F}_t)_{t\in\mathbb{T}} \equiv \mathfrak{F}$ be a filtration on Ω fulfilling the usual conditions, and let $(X_t)_{t\in\mathbb{T}} \equiv X$ be a K-dimensional real stochastic process on Ω.

Definizione 434 *We say that a function $f : \mathbb{T} \times \Omega \to \mathbb{R}$ is Ito-integrable, if:*

1. *f is measurable with respect to the σ-algebras $\mathcal{B}(\mathbb{T}) \otimes \mathfrak{F}_\infty$ and $\mathcal{B}(\mathbb{R})$;*
2. *the t-partial map $f_t : \Omega \to \mathbb{R}$ given by*

$$f_t(\omega) \overset{def}{=} f(t, \omega), \quad \forall \omega \in \Omega,$$

is an $(\mathfrak{F}_t, \mathcal{B}(\mathbb{R}))$-random variable, for every $t \in \mathbb{T}$;

3. *we have*

$$\mathbf{E}\left[\int_0^T f(t, \omega)^2 \; d\mu_L(t)\right] < \infty.$$

for every $T \in \mathbb{T}$.

We denote by $I(\mathbb{T} \times \Omega; \mathbb{R})$ the set of all Ito-integrable functions.

Note that

$$\mathbf{E}\left[\int_0^T f(t, \omega)^2 \; d\mu_L(t)\right] \equiv \int_\Omega \left(\int_0^T f(t, \omega)^2 \; d\mu_L(t)\right) \; d\mathbf{P}(\omega) = \int_{[0,T]\times\Omega} f(t, \omega)^2 \; d(\mu_L \otimes \mathbf{P})(t, \omega).$$

We define the Ito integral for a suitable set of *elementary functions* in $I(\mathbb{T} \times \Omega; \mathbb{R})$. Then, since every function $f \in I(\mathbb{T} \times \Omega; \mathbb{R})$ can be considers as a limit of a sequence of such elementary functions, we define the Ito integral for the function f as the limit of integrals of the elementary functions approximating f.

Definizione 435 *A function $\phi \in I(\mathbb{T} \times \Omega; \mathbb{R})$ is said to be* elementary *if for some $T \in \mathbb{T}$ there exists $(t_n)_{n=0}^{N} \in \mathcal{C}_{fin}(\mathbb{T})$ with $t_0 \equiv 0$ and $t_n \equiv T$ and a corresponding sequence $(F_n)_{n=0}^{N-1}$ of essentially bounded $(\mathcal{F}_{t_k}, \mathcal{B}(\mathbb{R}))$-random variables such that*

$$\phi(t,\omega) = \sum_{n=0}^{N-1} F_n(\omega) 1_{[t_n,t_{n+1}[}(t),$$

where we write $1_{[t_n,t_{n+1}[}(t)$ for the indicator function of the interval $[t_n, t_{n+1}[$, for every $n = 0, 1, \ldots, N-1$. We denote by $I_e(\mathbb{T} \times \Omega; \mathbb{R})$ the subset of all elementary functions of $I(\mathbb{T} \times \Omega; \mathbb{R})$.

Definizione 436 *We call* Ito integral *of the elementary function $\phi \in I(\mathbb{T} \times \Omega; \mathbb{R})$ the real random variable*

$$\int_0^T \phi(t,\omega)\ dB_t \equiv \sum_{n=0}^{N-1} F_n(\omega)\left(B_{t_{n+1}}(\omega) - B_{t_n}(\omega)\right).$$

Proposizione 437 *For any $\phi \in I_e(\mathbb{T} \times \Omega; \mathbb{R})$ the process $(S_t)_{t\in\mathbb{T}}$ given by*

$$S_t \overset{def}{=} \begin{array}{ll} \sum_{\substack{m=0,\ldots,n \\ t_n<t<t_{n+1}}} F_n(\omega)\left(B_t(\omega) - B_{t_n}(\omega)\right) & \text{if } t < T \\ \int_0^T \phi(t,\omega)\ dB_t & \text{if } t \geqslant T \end{array},$$

is a second order $\mathfrak{F}^B$-martingale, where $\mathfrak{F}^B$ is the filtration generated by the Brownian motion $(B_t)_{t\in\mathbb{T}}$.

Teorema 438 *For every $f \in I(\mathbb{T} \times \Omega; \mathbb{R})$ and for every $T \in \mathbb{T}$ there exists a sequence $(\phi_n)_{n\geqslant 0}$ in $I_e(\mathbb{T} \times \Omega; \mathbb{R})$ such that*

$$\lim_{n\to\infty} \mathbf{E}\left[\int_0^T \left(\phi_n(t,\omega) - f(t,\omega)\right)^2\ d\mu_L(t)\right] = 0. \tag{6.64}$$

Moreover, for any pair of sequences $(\phi'_n)_{n\geqslant 1}, (\phi''_n)_{n\geqslant 1}$ in $I_e(\mathbb{T} \times \Omega; \mathbb{R})$ fulfilling (6.64) we have

$$\lim_{n\to\infty} \left\| \int_0^T \phi'_n(t,\omega)\ dB_t - \int_0^T \phi''_n(t,\omega)\ dB_t \right\|_2 = 0,$$

where $\|\cdot\|_2$ is the standard norm in $L^2(\Omega; \mathbb{R})$.

Proof. . □

Definizione 439 *We call the* Ito integral *of the function $f \in I(\mathbb{T} \times \Omega; \mathbb{R})$ in $[0,T]$ the real random variable*

$$\int_0^T f(t,\omega)\ dB_t = \|\cdot\|_2 - \lim_{n\to\infty} \int_0^T \phi_n(t,\omega)\ dB_t.$$

where $(\phi_n)_{n\geqslant 0}$ is a any sequence in $I_e(\mathbb{T} \times \Omega; \mathbb{R})$ fulfilling (6.64) of Theorem 438.

The following theorem is one of the main achievements of the Itō stochastic integration theory.

Teorema 440 *[Itō isometry] For every function $f \in I(\mathbb{T} \times \Omega; \mathbb{R})$ we have*

$$\mathbf{E}\left[\left(\int_0^T f(t,\omega)\ dB_t\right)^2\right] = \mathbf{E}\left[\int_0^T f^2(t,\omega)\ d\mu_L(t)\right],$$

for every $T \in \mathbb{T}$.

Proposizione 441 (Properties of the Itō integral) *For all $f, g \in I(\mathbb{T} \times \Omega; \mathbb{R})$ we have*

1. *the integral* $\int_0^T f(t,\omega)\, dB_t$ *is an* $(\mathfrak{F}_T, \mathcal{B}(\mathbb{R}))$*-random variable, for every* $T \in \mathbb{T}$*;*

2. *we have*

$$\mathbf{E}\left[\int_0^T f(t,\omega)\, dB_t\right] = 0,$$

for every $T \in \mathbb{T}$*;*

3. *we have*

$$\int_0^T (\alpha f(t,\omega) + \beta g(t,\omega))\, dB_t = \alpha \int_0^T f(t,\omega)\, dB_t + \beta \int_0^T g(t,\omega)\, dB_t,$$

for every $\alpha, \beta \in \mathbb{R}$ *and for every* $T \in \mathbb{T}$*.*

Teorema 442 *For every function* $f \in I(\mathbb{T} \times \Omega; \mathbb{R})$ *the stochastic process* $(S_t)_{t\in\mathbb{T}}$ *given by*

$$S_t \overset{def}{=} \int_0^t f(s,\omega)\, dB_s, \quad \forall t \in \mathbb{T},$$

is a second order $\mathfrak{F}^B$*-martingale, where* $\mathfrak{F}^B$ *is the filtration generated by the Brownian motion* $(B_t)_{t\in\mathbb{T}}$*.*

6.13.2 The Itō Formula

Let $(B_t)_{t\in\mathbb{T}} \equiv B$ be a one-dimensional Brownian motion, let $\left(\mathcal{F}_t^B\right)_{t\in\mathbb{T}} \equiv \mathfrak{F}^B$ the σ-algebra generated by B and let $(X_t)_{t\in\mathbb{T}} \equiv X$ be a real stochastic process on Ω.

Definizione 443 *We say that* $(X_t)_{t\geqslant 0}$ *is an* Ito process *if we can write*

$$X_t = X_0 + \int_0^t H_s\, ds + \int_0^t K_s\, dB_s, \quad \forall t \geqslant 0 \tag{6.65}$$

where

- X_0 is a $\left(\mathcal{F}_0^B, B\right)$-random variable;
- $(H_t)_{t\geqslant 0}$ and $(K_t)_{t\geqslant 0}$ are $\mathfrak{F}^B$-adapted real stochastic processes on Ω such that

$$\int_0^t |H_s|\, ds < \infty \qquad \text{and} \qquad \int_0^t K_s^2\, ds < \infty, \tag{6.66}$$

for every $t \geqslant 0$.

It is customary to write Equation (6.65) in the shorter differential form

$$dX_t = H_t\, dt + K_t\, dB_t. \tag{6.67}$$

This formal writing suggests an interesting idea. Loosely speaking, we can think on the the increase dX_t of the process X as it were made by two parts. An endogenous part $H_t\, dt$ related to the inner structure of the process and an exogenous part $K_t\, dB_t$ due to the interaction between the process and the noise disturbance. Eventually, we have

Proposizione 444 *Assume that the processes* $(H_t)_{t\geqslant 0}$ *and* $(K_t)_{t\geqslant 0}$ *fulfill*

$$\mathbf{E}\left[\int_0^t H_s^2 ds\right] < \infty \qquad \textit{and} \qquad \mathbf{E}\left[\int_0^t K_s^2 ds\right] < \infty.$$

Then, we have

$$D^+\mathbf{E}_t[X_u]\big|_{u=t} = H_t \qquad \textit{and} \qquad D^+\mathbf{D}_t^2[X_u]\big|_{u=t} = K_t^2,$$

per ogni $t \geqslant 0$*, where* $\mathbf{E}_t[X_u] \equiv \mathbf{E}\left[X_u \mid \mathfrak{F}_t^W\right]$*,* $\mathbf{D}_t^2[X_u] \equiv \mathbf{E}_t\left[X_u^2\right] - \mathbf{E}_t[X_u]^2 \equiv \mathbf{D}^2\left[X_u \mid \mathfrak{F}_t^W\right]$ *and* D^+ *denotes the right derivative operator.*

In light of the above proposition, we can loosely say that an Itō process is characterized by a *conditional expected drift* [resp. *variance*] *rate* H_t [resp. K_t^2] at time t and by a slight abuse of notation we can write

$$\mathbf{E}_t\left[dX_t\right] = H_t\,dt \qquad \text{and} \qquad \mathbf{D}_t^2\left[dX_t\right] = K_t^2 dt.$$

Let $(X_t)_{t\in\mathbb{T}} \equiv X$ be an Itō process on Ω with conditional expected drift [resp. variance] rate process $(H_t)_{t\in\mathbb{T}}$ [resp. $\left(K_t^2\right)_{t\in\mathbb{T}}$] (see Definition 443) and let $f \in C^{1,2}\left(\mathbb{R}_+ \times \mathbb{R}; \mathbb{R}\right)$

Teorema 445 *[Itō formula] Let $(Y_t)_{t\geqslant 0}$ be the real sochastic process given by $Y_t \overset{def}{=} f(t, X_t)$, for every $t \geqslant 0$. We then have*

$$Y_t = f\left(0, X_0\right) + \int_0^t \partial_s f\left(s, X_s\right)\, ds + \int_0^t \partial_x f\left(s, X_s\right)\, dX_s + \tag{6.68}$$
$$+ \frac{1}{2}\int_0^t \partial_{x,x} f\left(s, X_s\right) K_s^2\, ds,$$

where

$$\int_0^t \partial_x f\left(s, X_s\right)\, dX_s = \int_0^t \partial_x f\left(s, X_s\right) H_s\, ds + \int_0^t \partial_x f\left(s, X_s\right) K_s\, dB_s.$$

In differential notation

$$dY_t = \partial_t f\left(t, X_t\right)\, dt + \partial_x f\left(t, X_t\right)\left(H_t\, dt + K_t\, dN_t\right) + \frac{1}{2}\partial_{x,x} f\left(t, X_t\right) K_t^2\, dt.$$

We hav also

Teorema 446 *Let $(X_t)_{t\geqslant 0}$ and $(Y_t)_{t\geqslant 0}$ Itō process on Ω such that*

$$X_t = X_0 + \int_0^t H_s\, ds + \int_0^t K_s\, dB_s \qquad \text{and} \qquad Y_t = Y_0 + \int_0^t G_s\, ds + \int_0^t L_s\, dB_s.$$

We then have

$$X_t Y_t = X_0 Y_0 + \int_0^t X_s\, dY_s + \int_0^t Y_s\, dX_s + \langle X, Y\rangle_t\,,$$

where

$$\langle X, Y\rangle_t = \int_0^t K_s L_s\, ds.$$

is the so-called cross variation *of the processes $(X_t)_{t\geqslant 0}$ and $(Y_t)_{t\geqslant 0}$.*

6.14 Stochastic Differential Equations

6.14.1 Introductory material

Consider the differential equation for the real stochastic process $(X_t)_{t\geqslant 0}$

$$\frac{dX_t}{dt} = \mu\left(t, X_t\right) + \sigma\left(t, X_t\right) Z_t, \tag{6.69}$$

where $\mu\left(t, x\right)$ and $\sigma\left(t, x\right)$ are measurable real functions defined on $\mathbb{R}_+ \times \mathbb{R}$ with some additional regularity property and $(Z_t)_{t\geqslant 0}$ is an one-dimensional *white noise*. Itô's interpretation of Equation (6.69) is that $(X_t)_{t\geqslant 0}$ satisfies the stochastic integral equation

$$X_t = X_0 + \int_0^t \mu\left(t, X_t\right)\, ds + \int_0^t \sigma\left(t, X_t\right)\, dW_s, \tag{6.70}$$

where $(W_t)_{t\geqslant 0}$ is an one-dimensional Wiener process (see Definition 430). Equation (6.70) can be rewritten in the differential form

$$dX_t = \mu(t, X_t)\ dt + \sigma(t, X_t)\ dW_t. \tag{6.71}$$

Formally we have then

$$Z_t = \frac{dW_t}{dt}.$$

However, we have to tackle some important issues.

- Existence and uniqueness theorems for such stochastic differential equation.
- Qualitative properties of the solutions.
- Explicit solutions.

6.14.2 Ito's Theory for Stochastic Differential Equations

Let $b : \mathbb{R}_+ \times \mathbb{R}^K \to \mathbb{R}^K$, $b(t,x) \equiv (b_k(t,x))_{k=1,\dots,K}$, be a Borel-measurable K-dimensional vector field on $\mathbb{R}_+ \times \mathbb{R}^K$ and let $\sigma : \mathbb{R}_+ \times \mathbb{R}^K \to \mathbb{R}^{K\times L}$, $\sigma(t,x) \equiv (\sigma_{k,\ell}(t,x))_{k=1,\dots,K,\ \ell=1,\dots,L}$, be a Borel-measurable $K \times L$ matrix field on $\mathbb{R}_+ \times \mathbb{R}^K$.

Definizione 447 *The problem of finding a probability space $\Omega \equiv (\Omega, \mathcal{E}, \mathbf{P})$, a filtration $(\mathcal{F}_t)_{t\geqslant 0} \equiv \mathfrak{F}$ on Ω, an $\mathfrak{F}$-adapted standard L-dimensional Brownian motion $B \equiv (B_t)_{t\geqslant 0}$ on Ω and an $\mathfrak{F}$-adapted K-dimensional process $(X_t)_{t\geqslant 0}$ on Ω such that for every $k = 1,\dots,K$ and every $t \in [0,+\infty[$ we have*

$$X_t^{(k)} = X_0^{(k)} + \sum_{\ell=1}^{L} \int_0^t \sigma_{k,\ell}(t, X_s)\ dB_s^{(\ell)} + \int_0^t b_k(t, X_s)\ dt, \tag{6.72}$$

where $X_0 \equiv \left(X_0^{(1)}, \dots, X_0^{(K)}\right)$ is a given K-dimensional random vector is called weak stochastic differential equation *with* diffusion coefficient $\sigma\sigma^\top$*, where $\sigma^\top$ is the transpose of σ,* drift coefficient b*, and* initial state X_0*. This problem is commonly written in the differential form*

$$dX_t^{(k)} = \sum_{\ell=1}^{L} \sigma_{k,\ell}(t, X_s)\ dB_s^{(\ell)} + b_k(t, X_s)\ dt. \tag{6.73}$$

The 4-tuple $(\Omega, \mathfrak{F}, B, X)$ is called a weak solution *of Equation (6.73).*

Definizione 448 *Equation (6.72) is called a* strong stochastic differential equation *when the probability space Ω, the filtration $\mathfrak{F}$, and the Brownian motion B are given in advance. In this case, the process X by alone is called a* strong solution *to (6.72).*

Teorema 449 *Assume that $b : \mathbb{R}_+ \times \mathbb{R}^K$ and $\sigma : \mathbb{R}_+ \times \mathbb{R}^K \to \mathbb{R}^{K\times L}$ fulfill the following assumptions:*

1. local boundedness *in the time variable and* sublinear growth *in the space variable, that is to say for any $T > 0$ there exists $A > 0$ such that*

$$|b(t,x)|_K \leqslant A(1 + |x|_K) \qquad \text{and} \qquad |\sigma(t,x)|_{K\times L} \leqslant A(1 + |x|_K),$$

for every $(t,x) \in \mathbb{R}_+ \times \mathbb{R}^K$ such that $t \leqslant T$, where $|\cdot|_K$ [resp. $|\cdot|_{K\times L}$] is the standard norm on $\mathbb{R}^K$ [resp. $\mathbb{R}^{K\times L}$];

2. local lipschitzianity *in the space variable, that is to say for any $T, R > 0$ there exists $B > 0$ such that we have*

$$|b(t,x) - b(t,y)|_K \leqslant B|x-y|_K \qquad \text{and} \qquad |\sigma(t,x) - \sigma(t,y)|_{K\times L} \leqslant B|x-y|_K,$$

for all $(t,x),(t,y) \in \mathbb{R}_+ \times \mathbb{R}^K$ such that $t \leqslant T$ and $|x|_K \vee |y|_K \leqslant R$.

Then for any K-dimensional random vector with finite second order moment X_0 there exists a unique second order progressively measurable stochastic process $(X_t)_{t\leqslant T}$ with almost sure square integrable paths solution of Equation (6.72).

Proof. See Baldi (**?**). □

6.14.3 Ornstein-Uhlenbeck Equations

Given a complete probability space $\Omega \equiv (\Omega, \mathcal{E}, \mathbf{P})$, a filtration $(\mathcal{F}_t)_{t\geqslant 0} \equiv \mathfrak{F}$ on Ω, and an $\mathfrak{F}$-adapted L-dimensional Wiener process $(W_t)_{t\geqslant 0} \equiv W$ on Ω, let us consider the stochastic differential equation

$$dX_t = (A(t)X_t + b(t))\,dt + \sigma(t)\,dW_t, \tag{6.74}$$

where $A(t)$ is a real K-order matrix process, $b(t)$ is a real K-dimensional vector process, $\sigma(t)$ is a real $K \times L$ matrix process, and $(X_t)_{t\geqslant 0} \equiv X$ is a real K-dimensional random vector on Ω which is independent of W.

Proposizione 450 *Assume that the matrix processes $A(t)$ and $\sigma(t)$, and the vector process $b(t)$ are deterministic, measurable and locally bounded. Then there exists a unique $\mathfrak{F}$-adapted K-dimensional stochastic process $(X_t)_{t\geqslant 0}$ on Ω solution of the Ornstein-Uhlenbeck equation (6.74).*

To determine the explicit form of such a solution we first consider the deterministic Cauchy problem corresponding to (6.74), That is

$$\begin{array}{l} \frac{d}{dt}X(t) = A(t)X(t) + b(t) \\ X(0) = X_0 \end{array}. \tag{6.75}$$

We have

Proposizione 451 *The Cauchy problem (6.75) has a unique solution $X(t)$ given by*

$$X(t) = \Xi(t)\left(X + \int_0^t \Xi^{-1}(s)\,b(s)\,ds\right), \tag{6.76}$$

where $\Xi(t)$ is the invertible K-order square matrix which in turn solves the Cauchy problem

$$\begin{cases} \frac{d}{dt}\Xi(t) = A(t)\Xi(t) \\ \Xi(0) = I_K \end{cases} \tag{6.77}$$

I_K being the K-order identity matrix.

Proof. Assume that $X_1(t)$ and $X_2(t)$ are both solution of the Cauchy Problem (6.75). Then the difference $\Delta(t) \equiv X_2(t) - X_1(t)$ solves the Cauchy Problem

$$\begin{array}{l} \frac{d}{dt}\Delta(t) = A(t)\Delta(t) \\ \Delta(0) = 0 \end{array}. \tag{6.78}$$

On the other hand, under the assumptions considered for $A(t)$, the Cauchy Problem (6.78) is well known to have unique solution

$$\Delta(t) = \Delta(0) = 0,$$

for every $t \geqslant 0$. This implies

$$X_1(t) = X_2(t),$$

for every $t \geqslant 0$. Now, the Cauchy problem (6.77), has a unique (absolutely continuous) K-order square matrix solution $\Xi(t)$ defined for every $t \geqslant 0$. The column vectors of $\Xi(t)$ are a set of fundamental solutions of the homogeneous equation

$$\dot{X}(t) = A(t) X(t) \tag{6.79}$$

associated to the Cauchy problem (6.75). Moreover, for every $t \geqslant 0$ the matrix $\Xi(t)$ is nonsingular.

In fact, suppose there exists $t_0 > 0$ such that $\Xi(t_0)$ is singular. Then there should exist a nonzero K-dimensional vector Y such that

$$\Xi(t_0) Y = 0.$$

On the other hand, the K-dimensional vector $\Xi(t) Y$ is a solution to Equation (6.79) which vanishes at t_0. Hence $\Xi(t) Y$ vanishes for every $t \geqslant 0$. In particular we would have

$$\Xi(0) Y = Y = 0,$$

which contradicts the choice of Y.

The above argument allows us to conclude that $X(t)$ given by (6.76) is well defined. In addition, a straightforward computation gives

$$\begin{aligned}
\dot{X}(t) &= \dot{\Xi}(t) \left(X + \int_0^t \Xi^{-1}(s) b(s)\, ds \right) + \Xi(t) \Xi^{-1}(t) b(t) \\
&= A(t) \Xi(t) \left(X + \int_0^t \Xi^{-1}(s) b(s)\, ds \right) + b(t) \\
&= A(t) X(t) + b(t),
\end{aligned}$$

which completes the proof. □

Osservazione 452 *If A is independent of t, then the solution of the Cauchy problem (6.77) is given by*

$$\Xi(t) = \exp(At).$$

Teorema 453 *The solution $(X_t)_{t \geqslant 0}$ of the stochastic differential equation (6.74) is given by*

$$X_t = \Xi(t) \left(X + \int_0^t \Xi^{-1}(s) b(s)\, ds + \int_0^t \Xi^{-1}(s) \sigma(s)\, dW_s \right), \tag{6.80}$$

for every $t \geqslant 0$.

Proof. Writing

$$\Xi(t) = \Xi(0) + \int_0^t \dot{\Xi}(s)\, ds$$

and applying the Itō integration by parts rule to the right hand side of (**??**), we have

$$
\begin{aligned}
X_t &= \Xi(0)\, X + \int_0^t \Xi(s) \left(\Xi^{-1}(s)\, b(s)\ ds + \Xi^{-1}(s)\, \sigma(s)\ dW_s\right) \\
&\quad + \int_0^t \dot{\Xi}(s) \left(X + \int_0^s \Xi^{-1}(r)\, b(r)\ dr + \int_0^s \Xi^{-1}(r)\, \sigma(r)\ dW_r\right) ns \\
&= X + \int_0^t b(s)\ ds + \int_0^t \sigma(s)\ dW_s \\
&\quad + \int_0^t A(s)\, \Xi(s) \left(X + \int_0^s \Xi^{-1}(r)\, b(r)\ dr + \int_0^s \Xi^{-1}(r)\, \sigma(r)\ dW_r\right) ds \\
&= X + \int_0^t b(s)\ ds + \int_0^t \sigma(s)\ dW_s + \int_0^t A(s)\, X_s\, ds.
\end{aligned}
$$

Therefore, the process $(X_t)_{t\geqslant 0}$ solves the integral counterpart of Equation (6.74). $\square$

Corollary 454 *Suppose that* $\mathbf{E}\left[\left\|X^2\right\|\right] < \infty$. *We then have*

$$
\mathbf{E}[X_t] = \Xi(t) \left(\mathbf{E}[X] + \int_0^t \Xi^{-1}(s)\, b(s)\ ds\right) \tag{6.81}
$$

and

$$
Cov(X_s, X_t) = \Xi(s) \left(\mathbf{D}^2[X] + \int_0^{s\wedge t} \Xi^{-1}(r)\, \sigma(r) \left(\Xi^{-1}(r)\, \sigma(r)\right)^{\intercal}\ dr\right) \Xi^{\intercal}(t)\,. \tag{6.82}
$$

for all $s, t \geqslant 0$. *In particular* $\mathbf{E}[X_t] \equiv \mu(t)$ *and* $\mathbf{D}^2[X_t] \equiv \Sigma(t)$ *solve the linear equations*

$$
\dot{\mu}(t) = A(t)\, \mu(t) + b(t)
$$

and

$$
\dot{\Sigma}(t) = A(t)\, \Sigma(t) + \Sigma(t)\, A^{\intercal}(t) + \sigma(t)\, \sigma^{\intercal}(t)\,.
$$

Proof. Thanks to Theorem 453, we have

$$
\begin{aligned}
\mathbf{E}[X_t] &= \mathbf{E}\left[\Xi(t) \left(X + \int_0^t \Xi^{-1}(s)\, b(s)\ ds + \int_0^t \Xi^{-1}(s)\, \sigma(s)\ dW_s\right)\right] \\
&= \Xi(t) \left(\mathbf{E}[X] + \mathbf{E}\left[\int_0^t \Xi^{-1}(s)\, b(s)\ ds\right] + \mathbf{E}\left[\int_0^t \Xi^{-1}(s)\, \sigma(s)\ dW_s\right]\right) \\
&= \Xi(t) \left(\mathbf{E}[X] + \int_0^t \Xi^{-1}(s)\, b(s)\ ds\right).
\end{aligned}
$$

This implies

$$
X_t - \mathbf{E}[X_t] = \Xi(t) \left(X - \mathbf{E}[X] + \int_0^t \Xi^{-1}(s)\, \sigma(s)\ dW_s\right),
$$

and

$$
\begin{aligned}
Cov\left(X_s, X_t\right) &= \mathbf{E}\left[\left(\Xi\left(s\right)\left(X - \mathbf{E}\left[X\right] + \int_0^s \Xi^{-1}\left(r\right)\sigma\left(r\right)\, dW_r\right)\right)\right. \\
&\quad \left. \cdot \left(\Xi\left(t\right)\left(X - \mathbf{E}\left[X\right] + \int_0^t \Xi^{-1}\left(s\right)\sigma\left(s\right)\, dW_s\right)\right)^{\intercal}\right] \\
&= \Xi\left(s\right)\mathbf{E}\left[\left(X - \mathbf{E}\left[X\right] + \int_0^s \Xi^{-1}\left(r\right)\sigma\left(r\right)\, dW_r\right)\right. \\
&\quad \left. \cdot \left(\left(X - \mathbf{E}\left[X\right]\right)^T + \int_0^t \left(\Xi^{-1}\left(r\right)\sigma\left(r\right)\right)^{\intercal}\, dW_r\right)\right]\Xi^{\intercal}\left(t\right) \\
&= \Xi\left(s\right)\left(\mathbf{E}\left[\left(X - \mathbf{E}\left[X\right]\right)\left(X - \mathbf{E}\left[X\right]\right)^{\intercal}\right] + \mathbf{E}\left[\left(X - \mathbf{E}\left[X\right]\right)\int_0^t \left(\Xi^{-1}\left(r\right)\sigma\left(r\right)\right)^{\intercal}\, dW_r\right]\right. \\
&\mathbf{E}\left[\int_0^s \Xi^{-1}\left(r\right)\sigma\left(r\right)\, dW_r\,\left(X - \mathbf{E}\left[X\right]\right)^{\intercal}\right] \\
&\quad \left. + \mathbf{E}\left[\int_0^s \Xi^{-1}\left(r\right)\sigma\left(r\right)\, dW_r \int_0^t \left(\Xi^{-1}\left(r\right)\sigma\left(r\right)\right)^{\intercal}\, dW_r\right]\right)\Xi^{\intercal}\left(t\right). \qquad (6.83)
\end{aligned}
$$

Now,

$$
\mathbf{D}^2\left[X\right] = \mathbf{E}\left[\left(X - \mathbf{E}\left[X\right]\right)\left(X - \mathbf{E}\left[X\right]\right)^{\intercal}\right], \qquad (6.84)
$$

and, since the Wiener process $\left(W_t\right)_{t\geqslant 0}$ is independent of X,

$$
\mathbf{E}\left[\left(X - \mathbf{E}\left[X\right]\right)\int_0^t \left(\Xi^{-1}\left(r\right)\sigma\left(r\right)\right)^{\intercal}\, dW_r\right] = \mathbf{E}\left[\left(X - \mathbf{E}\left[X\right]\right)\right]\mathbf{E}\left[\int_0^t \left(\Xi^{-1}\left(r\right)\sigma\left(r\right)\right)^{\intercal}\, dW_r\right] = 0 \qquad (6.85)
$$

and

$$
\mathbf{E}\left[\int_0^s \Xi^{-1}\left(r\right)\sigma\left(r\right)\, dW_r\,\left(X - \mathbf{E}\left[X\right]\right)^{\intercal}\right] = \mathbf{E}\left[\int_0^s \Xi^{-1}\left(r\right)\sigma\left(r\right)\, dW_r\right]\mathbf{E}\left[\left(X - \mathbf{E}\left[X\right]\right)^{\intercal}\right] = 0. \qquad (6.86)
$$

Moreover, we can write

$$
\begin{aligned}
&\int_0^s \Xi^{-1}\left(r\right)\sigma\left(r\right)\, dW_r \int_0^t \left(\Xi^{-1}\left(r\right)\sigma\left(r\right)\right)^{\intercal}\, dW_r \\
&= \left(\int_0^{s\wedge t} \Xi^{-1}\left(r\right)\sigma\left(r\right)\, dW_r + \int_{s\wedge t}^s \Xi^{-1}\left(r\right)\sigma\left(r\right)\, dW_r\right) \\
&\cdot \left(\int_0^{s\wedge t} \left(\Xi^{-1}\left(r\right)\sigma\left(r\right)\right)^{\intercal}\, dW_r + \int_{s\wedge t}^t \left(\Xi^{-1}\left(r\right)\sigma\left(r\right)\right)^{\intercal}\, dW_r\right)
\end{aligned}
$$

and the properties of the Itō integral imply

$$
\begin{aligned}
&\mathbf{E}\left[\int_0^s \Xi^{-1}\left(r\right)\sigma\left(r\right)\, dW_r \int_0^t \left(\Xi^{-1}\left(r\right)\sigma\left(r\right)\right)^{\intercal}\, dW_r\right] \\
&= \mathbf{E}\left[\int_0^{s\wedge t} \Xi^{-1}\left(r\right)\sigma\left(r\right)\, dW_r \int_0^{s\wedge t} \left(\Xi^{-1}\left(r\right)\sigma\left(r\right)\right)^{\intercal}\, dW_r\right] \\
&= \mathbf{E}\left[\int_0^{s\wedge t} \Xi^{-1}\left(r\right)\sigma\left(r\right)\left(\Xi^{-1}\left(r\right)\sigma\left(r\right)\right)^{\intercal}\, dr\right] \\
&= \int_0^{s\wedge t} \Xi^{-1}\left(r\right)\sigma\left(r\right)\left(\Xi^{-1}\left(r\right)\sigma\left(r\right)\right)^{\intercal}\, dr. \qquad (6.87)
\end{aligned}
$$

Combining (6.83) (6.84) (6.85) (6.86) and (6.87) Equation (**??**) immediately follows. □

Osservazione 455 *In the case $A(t) \equiv A$ independent of t, the solution of Equation (6.74) is given by*

$$X_t = \exp(tA)\left(x + \int_0^t \exp(-sA)\, b(s)\; ds + \int_0^t \exp(-sA)\, \sigma(s)\; dW_s\right),$$

where

$$\exp(tA) \equiv \sum_{n=0}^{\infty} \frac{t^n}{n!} A^n.$$

Moreover, for all $s, t \geqslant 0$, we have

$$X_t = \exp((t-s)A)\left(X_s + \exp(sA)\left(\int_s^t \exp(-uA)\, b(u)\; du + \int_s^t \exp(-uA)\, \sigma(u)\; dW_u\right)\right).$$

As a consequence,

$$\mathbf{E}[X_t] = \exp(tA)\left(\mathbf{E}[x] + \int_0^t \exp(-sA)\, b(s)\; ds\right)$$

and

$$Cov(X_s, X_t) = \exp(sA)\left(\mathbf{D}^2[x] + \int_0^{s\wedge t} \exp(-rA)\, \sigma(r)\, \sigma(r)^{\intercal} \exp(-rA^{\intercal})\; dr\right) \exp(tA^{\intercal}).$$

In particular,

$$\mathbf{D}^2[X_t] = \exp(tA)\left(\mathbf{D}^2[x] + \int_0^t \exp(-rA)\, \sigma(r)\, \sigma(r)^{\intercal} \exp(-rA^{\intercal})\; dr\right) \exp(tA^{\intercal}).$$

Proof. The claim immediately follows combining Theorem 453 and Corollary **??** with Remark 452. Now, We can write

$$\begin{aligned}
X_t &= \exp((t-s)A)\exp(sA)\left(x + \int_0^s \exp(-uA)\, b(u)\; ds + \int_s^t \exp(-uA)\, b(u)\; ds\right. \\
&\quad \left. + \int_0^s \exp(-uA)\, \sigma(u)\; dW_u + \int_s^t \exp(-uA)\, \sigma(u)\; dW_u\right) \\
&= \exp((t-s)A)\left(\exp(sA)\left(x + \int_0^s \exp(-uA)\, b(u)\; ds + \int_0^t \exp(-uA)\, b(u)\; ds\right)\right. \\
&\quad \left. + \exp(sA)\left(\int_s^t \exp(-uA)\, b(u)\; ds + \int_s^t \exp(-uA)\, \sigma(u)\; dW_u\right)\right) \\
&= \exp((t-s)A)\left(X_s + \exp(sA)\left(\int_s^t \exp(-uA)\, b(u)\; ds + \int_s^t \exp(-uA)\, \sigma(u)\; dW_u\right)\right)
\end{aligned}$$

... □

Osservazione 456 *In the case $K = L = 1$, $b(t) = 0$, $A(t) \equiv -\alpha$, for $\alpha > 0$, and $\sigma(t) = \sigma > 0$, Equation (6.74) becomes*

$$\begin{cases} dX_t = -\alpha X_t\, dt + \sigma\, dW_t \\ X_{t=0} = X_0 \end{cases} \tag{6.88}$$

which is the classical Ornstein-Uhlenbeck equation. The solution of Equation (6.88) is the Ornstein-Uhlenbeck process $(X_t)_{t\geqslant 0}$ given by

$$X_t = X_0 e^{-\alpha t} + \sigma e^{-\alpha t} \int_0^t e^{\alpha s}\, dW_s.$$

Moreover if $\mathbf{E}\left[\left\|X_0^2\right\|\right] < \infty$ *we have*

$$\begin{aligned}
\mathbf{E}\left[X_t\right] &= \mathbf{E}\left[X_0\right] e^{-\alpha t}, \\
\mathbf{D}^2\left[X_t\right] &= e^{-2\alpha t}\left(\mathbf{D}^2\left[X_0\right] + \frac{\sigma^2}{2\alpha}\left(e^{2\alpha t} - 1\right)\right) \\
Cov\left(X_s, X_t\right) &= e^{-\alpha(s+t)}\left(\mathbf{D}^2\left[X_0\right] + \frac{\sigma^2}{2\alpha}\left(e^{2\alpha(s\wedge t)} - 1\right)\right).
\end{aligned}$$

Proof. We have

$$\begin{aligned}
X_t &= Xe^{-\alpha t} + \sigma e^{-\alpha t}\int_0^t e^{\alpha r}\, dW_r \\
&= e^{-\alpha(t-s)}\left(Xe^{-\alpha s} + \sigma e^{-\alpha s}\int_0^s e^{\alpha r}\, dW_r\right) + \sigma e^{-\alpha t}\int_s^t e^{\alpha r}\, dW_r \\
&= e^{-\alpha(t-s)}X_s + \sigma e^{-\alpha t}\int_s^t e^{\alpha r}\, dW_r,
\end{aligned}$$

for all $0 \leqslant s < t$. Therefore,

$$\begin{aligned}
\mathbf{E}\left[X_t\right] &= e^{-\alpha(t-s)}\mathbf{E}\left[X_s\right] + \sigma e^{-\alpha t}\mathbf{E}\left[\int_s^t e^{\alpha r}\, dW_r\right] \\
&= e^{-\alpha(t-s)}\mathbf{E}\left[X_s\right] \\
&= e^{-\alpha t}\mathbf{E}\left[X_0\right],
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{E}\left[X_t^2\right] &= \mathbf{E}\left[\left(e^{-\alpha(t-s)}X_s + \sigma e^{-\alpha t}\int_s^t e^{\alpha r}\, dW_r\right)^2\right] \\
&= \mathbf{E}\left[e^{-2\alpha(t-s)}X_s^2 + 2\sigma e^{-\alpha t}X_s\int_s^t e^{\alpha r}\, dW_r + \sigma^2 e^{-2\alpha t}\left(\int_s^t e^{\alpha r}\, dW_r\right)^2\right] \\
&= e^{-2\alpha(t-s)}\mathbf{E}\left[X_s^2\right] + 2\sigma e^{-\alpha t}\mathbf{E}\left[X_s\int_s^t e^{\alpha r}\, dW_r\right] + \sigma^2 e^{-2\alpha t}\mathbf{E}\left[\left(\int_s^t e^{\alpha r}\, dW_r\right)^2\right] \\
&= e^{-2\alpha(t-s)}\mathbf{E}\left[X_s^2\right] + \sigma^2 e^{-2\alpha t}\int_s^t e^{2\alpha r}\, dr \\
&= e^{-2\alpha(t-s)}\mathbf{E}\left[X_s^2\right] + \frac{\sigma^2}{2\alpha}e^{-2\alpha t}\left(e^{2\alpha t} - e^{2\alpha s}\right) \\
&= e^{-2\alpha(t-s)}\left(\mathbf{E}\left[X_s^2\right] + \frac{\sigma^2}{2\alpha}\left(e^{2\alpha(t-s)} - 1\right)\right),
\end{aligned}$$

since

$$\begin{aligned}
\mathbf{E}\left[X_s\int_s^t e^{\alpha r}\, dW_r\right] &= \mathbf{E}\left[\mathbf{E}\left[X_s\int_s^t e^{\alpha r}\, dW_r\right]|\mathfrak{F}_s\right] \\
&= \mathbf{E}\left[X_s\mathbf{E}\left[\int_s^t e^{\alpha r}\, dW_r\right]|\mathfrak{F}_s\right] = 0
\end{aligned}$$

and, by the Itō isometry,

$$\mathbf{E}\left[\left(\int_s^t e^{\alpha r}\, dW_r^2\right)^2\right] = \mathbf{E}\left[\int_s^t e^{2\alpha r}\, dr\right] = \int_s^t e^{2\alpha r}\, dr.$$

We then have

$$
\begin{aligned}
\mathbf{n}^2[X_t] &= \mathbf{E}\left[X_t^2\right] - \mathbf{E}[X_t]^2 \\
&= e^{2\alpha(t-s)}\left(\mathbf{E}\left[X_s^2\right] + \frac{\sigma^2}{2\alpha}\left(e^{-2\alpha(t-s)} - 1\right)\right) - e^{2\alpha(t-s)}\mathbf{E}[X_s]^2 \\
&= e^{2\alpha(t-s)}\left(\mathbf{E}\left[X_s^2\right] - \mathbf{E}[X_s]^2 + \frac{\sigma^2}{2\alpha}\left(e^{-2\alpha(t-s)} - 1\right)\right) \\
&= e^{2\alpha(t-s)}\left(\mathbf{D}^2[X_s] + \frac{\sigma^2}{2\alpha}\left(e^{-2\alpha(t-s)} - 1\right)\right) \\
&= e^{2\alpha t}\left(\mathbf{D}^2[X_0] + \frac{\sigma^2}{2\alpha}\left(e^{-2\alpha t} - 1\right)\right),
\end{aligned}
$$

and

$$
\begin{aligned}
Cov(X_s, X_t) &= \mathbf{E}[(X_s - \mathbf{E}[X_s])(X_t - \mathbf{E}[X_t])] \\
&= \mathbf{E}\left[(X_s - \mathbf{E}[X_s])\left(e^{\alpha(t-s)}X_s + \sigma e^{-\alpha t}\int_s^t e^{\alpha r}\,dW_r - e^{\alpha(t-s)}\mathbf{E}[X_s]\right)\right] \\
&= \mathbf{E}\left[e^{\alpha(t-s)}X_s^2 + \sigma e^{-\alpha t}X_s\int_s^t e^{\alpha r}\,dW_r - 2e^{\alpha(t-s)}\mathbf{E}[X_s]X_s + e^{\alpha(t-s)}\mathbf{E}[X_s]^2\right] \\
&= e^{\alpha(t-s)}\mathbf{E}\left[X_s^2\right] + \sigma e^{-\alpha t}\mathbf{E}\left[X_s\int_s^t e^{\alpha r}\,dW_r\right] - e^{\alpha(t-s)}\mathbf{E}[X_s]^2 \\
&= e^{\alpha(t-s)}\mathbf{D}^2[X_s] \\
&= e^{\alpha(t+s)}\left(\mathbf{D}^2[X_0] + \frac{\sigma^2}{2\alpha}\left(e^{-2\alpha s} - 1\right)\right),
\end{aligned}
$$

which completes our proof. □

Osservazione 457 *If $(X_t)_{t\geqslant 0}$ is the Ornstein-Uhlenbeck process solution of Equation (6.88), and the initial random variable X is normally distributed with null mean and variance $\sigma^2/2\alpha$, then $(X_t)_{t\geqslant 0}$ is a weakly stationary zero-mean Gaussian process with covariance function*

$$
\gamma(s,t) = \frac{\sigma^2}{2\alpha}e^{-\alpha|t-s|}.
$$

The classical Ornstein-Uhlenbeck equation corresponds to the Langevin equation for the Brownian motion of a particle with friction.

Osservazione 458 *In the case $K = L = 1$, $b(t) = \mu$, $A(t) = -1$, and $\sigma(t) = \sigma > 0$, Equation (6.74) becomes*

$$
\begin{cases} dX_t = (m - X_t)\,dt + \sigma\,dW_t \\ X_{t=0} = X_0 \end{cases} \tag{6.89}
$$

which is the classical mean reverting Ornstein-Uhlenbeck equation. The solution of Equation (6.89) is the mean reverting Ornstein-Uhlenbeck process $(X_t)_{t\geqslant 0}$ given by

$$
X_t = \mu + (X_0 - \mu)e^{-t} + \sigma e^{-t}\int_0^t e^s\,dW_s.
$$

Moreover if $\mathbf{E}\left[\left\|X^2\right\|\right] < \infty$ we have

$$
\mathbf{E}[X_t] = \mu + (\mathbf{E}[X_0] - \mu)e^{-t}, \tag{6.90}
$$

$$
cov(X_s, X_t) = e^{-(s+t)}\left(\mathbf{D}^2[X_0] + \frac{\sigma^2}{2}\left(e^{2(s\wedge t)} - 1\right)\right). \tag{6.91}
$$

Osservazione 459 *We have*

$$cov\left(X_t - X_{t-\Delta t}, X_{t+\Delta t} - X_t\right) = -\left(e^{-\Delta t} - 1\right)^2 \left(\frac{\sigma^2}{2} + e^{-(2t-\Delta t)} \left(\frac{\sigma^2}{2} - \mathbf{D}^2\left[X_0\right]\right)\right)$$

for all $t, \Delta t \geqslant 0$.

Proof. The bilinearity property of the convariance functional allows us to write

$$\begin{aligned}
&cov\left(X_t - X_{t-\Delta t}, X_{t+\Delta t} - X_t\right) \\
&= cov\left(X_t, X_{t+\Delta t}\right) - cov\left(X_t, X_t\right) - cov\left(X_{t-\Delta t}, X_{t+\Delta t}\right) + cov\left(X_{t-\Delta t}, X_t\right),
\end{aligned}$$

where, thanks to (6.91),

$$\begin{aligned}
cov\left(X_t, X_{t+\Delta t}\right) &= e^{-(2t+\Delta t)} \left(\mathbf{D}^2\left[X\right] + \frac{\sigma^2}{2}\left(e^{2t} - 1\right)\right) \\
cov\left(X_t, X_t\right) &= e^{-2t} \left(\mathbf{D}^2\left[X\right] + \frac{\sigma^2}{2}\left(e^{2t} - 1\right)\right) \\
cov\left(X_{t-\Delta t}, X_{t+\Delta t}\right) &= e^{-2t} \left(\mathbf{D}^2\left[X\right] + \frac{\sigma^2}{2}\left(e^{2(t-\Delta t)} - 1\right)\right) \\
cov\left(X_{t-\Delta t}, X_t\right) &= e^{-(2t-\Delta t)} \left(\mathbf{D}^2\left[X\right] + \frac{\sigma^2}{2}\left(e^{2(t-\Delta t)} - 1\right)\right).
\end{aligned}$$

We then have

$$\begin{aligned}
&cov\left(X_t - X_{t-\Delta t}, X_{t+\Delta t} - X_t\right) \\
&= \left(e^{-(2t+\Delta t)} - e^{-2t} - e^{-2t} + e^{-(2t-\Delta t)}\right) \mathbf{D}^2\left[X\right] \\
&+ \frac{\sigma^2}{2}\left(e^{-\Delta t} - e^{-(2t+\Delta t)} - 1 + e^{-2t} - e^{-2\Delta t} + e^{-2t} + e^{-\Delta t} - e^{-(2t-\Delta t)}\right) \\
&= e^{-(2t-\Delta t)}\left(e^{-2\Delta t} - 2e^{-\Delta t} + 1\right) \mathbf{D}^2\left[X\right] - \frac{\sigma^2}{2}\left(e^{-(2t-\Delta t)}\left(e^{-2\Delta t} - 2e^{-\Delta t} + 1\right) + \left(e^{-\Delta t} - 1\right)^2\right), \\
&= e^{-(2t-\Delta t)}\left(e^{-\Delta t} - 1\right)^2 \mathbf{D}^2\left[X\right] - \frac{\sigma^2}{2}\left(e^{-(2t-\Delta t)}\left(e^{-\Delta t} - 1\right)^2 + \left(e^{-\Delta t} - 1\right)^2\right) \\
&= -\left(e^{-\Delta t} - 1\right)^2 \left(\frac{\sigma^2}{2} + e^{-(2t-\Delta t)}\left(\frac{\sigma^2}{2} - \mathbf{D}^2\left[X\right]\right)\right)
\end{aligned}$$

as desired. □

Capitolo 7

Stochastic Models for Time Series

7.1 Ergodic Processes

Let $(X_t)_{t\in\mathbb{T}} \equiv \mathbf{X}$ be a stochastic process of order 2 on a probability space $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ with state space $\mathbb{R}^N$, for some $N \in \mathbb{N}$. Fixed any $t_0 \in \mathbb{T}$, consider the N-variate real random variable X_{t_0} and let $X_{t_0}^{(1)}, \dots, X_{t_0}^{(n)}$ be a simple random sample of size $n \in \mathbb{N}$ drawn from X_{t_0}. From the basic Statistics, we know that the simple sample mean of size $n \in \mathbb{N}$ drawn from X_{t_0}, which is given by

$$\bar{X}_{t_0,n} \overset{\text{def}}{=} \frac{1}{n}\sum_{k=1}^{n} X_{t_0}^{(k)},$$

is an *unbiased estimator* of $\mathbf{E}[X_{t_0}] \equiv \mu_{\mathbf{X}}(t_0)$, that is

$$\mathbf{E}\left[\bar{X}_{t_0,n}\right] = \mu_{\mathbf{X}}(t_0).$$

As a consequence. the mean square error of $\bar{X}_{t_0,n}$ is given by

$$MSE\left(\bar{X}_{t_0,n}\right) = \operatorname{trace}\left(Var\left(X_{t_0}\right)\right),$$

where $\operatorname{trace} : \mathbb{R}^{N^2} \to \mathbb{R}$ is the *trace operator* given by

$$\operatorname{trace}\left(\left(a_{k,\ell}\right)_{k,\ell=1}^{N}\right) \overset{\text{def}}{=} \sum_{k=1}^{N} a_{k,k}, \quad \forall \left(a_{k,\ell}\right)_{k,\ell=1}^{N} \in \mathbb{R}^{N^2}.$$

Moreover, by virtue of the Law of Large Numbers for random variables with finite moment of order 2, we know that $\bar{X}_{t_0,n}$ is *mean square error consistent* that is

$$\bar{X}_{t_0,n} \overset{\mathbf{L}^2}{\to} \mu_{\mathbf{X}}(t_0).$$

A fortiori,

$$\bar{X}_{t_0,n} \overset{\mathbf{P}}{\to} \mu_{\mathbf{X}}(t_0).$$

Therefore, the sample mean $\bar{X}_{t_0,n}$ is a "good" estimator for $\mu_{\mathbf{X}}(t_0)$. Recall that the estimate of $\mu_{\mathbf{X}}(t_0)$ by means of the estimator $\bar{X}_{t_0,n}$, on the occurrence of an outcome $\omega \in \Omega$, can be written as

$$\bar{X}_{t_0,n}(\omega) \equiv \frac{1}{n}\sum_{k=1}^{n} X_{t_0}^{(k)}(\omega) \equiv \frac{1}{n}\sum_{k=1}^{n} x_{t_0}^{(k)},$$

where $x_{t_0}^{(1)} \equiv X_{t_0}^{(1)}(\omega), \dots, x_{t_0}^{(n)} \equiv X_{t_0}^{(n)}(\omega)$ are the realizations of the independent and identically distributed random variables $X_{t_0}^{(1)}, \dots, X_{t_0}^{(n)}$. Now, if we assume that the process $\mathbf{X}$ is at least weakly stationary, since

$$\mu_{\mathbf{X}}(t) = \mu_{\mathbf{X}}(t_0) \equiv \mu_{\mathbf{X}},$$

for every $t \in \mathbb{T}$, we could use $\bar{X}_{t_0,n}$ as a "good" estimator for the mean $\mu_{\mathbf{X}}$ of the process. In this context, the estimator $\bar{X}_{t_0,n}$ is referred to as the *ensemble average estimator* of size n of $\mu_{\mathbf{X}}$. However, unless the random variables in the process $\mathbf{X}$ are independent and identically distributed, we cannot observe the realization of n independent copies of X_{t_0}, but only the realizations $X_{t_1}(\omega), \dots, X_{t_T}(\omega)$ of the random variables in $\mathbf{X}$ at different times, $t_1, \dots, t_T \in \mathbb{T}$, for some $T \in \mathbb{N}$, and we know that even in the case of strong stationary processess these variables do not need to be independent, neither uncorrelated. Hence, the question arises to what extent we can use the information provided by $X_{t_1}(\omega), \dots, X_{t_T}(\omega)$ to estimate some traits of the process $\mathbf{X}$. This leads to introduce the idea of *ergodicity*.

For simplicity, assume that $(X_t)_{t\in\mathbb{T}} \equiv \mathbf{X}$ is a WSS process. Furthermore, assume that $\mathbb{T} \equiv \mathbb{N}$ or $\mathbb{T} \equiv \mathbb{Z}$. Write $\mu_{\mathbf{X}}$ [resp. $\Sigma_{\mathbf{X}}$] for the constant value of the mean [resp. variance-covariance] function of $\mathbf{X}$ and write $\Gamma_{\mathbf{X},1}$ or $\Gamma_{\mathbf{X},0}$ [resp. $\mathrm{P}_{\mathbf{X},1}$ or $\mathrm{P}_{\mathbf{X},0}$] for the reduced autocovariance [resp. autocorrelation] function of $\mathbf{X}$ referred to $t_0 = 1$ or $t_0 = 0$, according to whether $\mathbb{T} \equiv \mathbb{N}$ or $\mathbb{T} \equiv \mathbb{Z}$.

Definizione 460 *Fixed any $T \in \mathbb{N}$, we call the* time average estimator *of size T of $\mathbf{X}$ the statistic*

$$\bar{X}_T \overset{def}{=} \begin{cases} \frac{1}{T}\sum\limits_{t=1}^{T} X_t, & \text{if } \mathbb{T} \equiv \mathbb{N}, \\ \frac{1}{2T+1}\sum\limits_{t=-T}^{T} X_t, & \text{if } \mathbb{T} \equiv \mathbb{Z}. \end{cases} \tag{7.1}$$

Definizione 461 *Fixed any $T \in \mathbb{N}$, we call the* time variance-covariance estimator *of size T of $\mathbf{X}$ the statistic*

$$S^2_{\mathbf{X},T} \overset{def}{=} \begin{cases} \frac{1}{T}\sum\limits_{t=1}^{T} \left(X_t - \bar{X}_T\right)\left(X_t - \bar{X}_T\right)^{\intercal}, & \text{if } \mathbb{T} \equiv \mathbb{N}, \\ \frac{1}{2T+1}\sum\limits_{t=-T}^{T} \left(X_t - \bar{X}_T\right)\left(X_t - \bar{X}_T\right)^{\intercal}, & \text{if } \mathbb{T} \equiv \mathbb{Z}. \end{cases} \tag{7.2}$$

Definizione 462 *Fixed any $T \in \mathbb{N}$, we call* time autocovariance estimator *of size T of $\mathbf{X}$ at lag (shift) τ the statistic*

$$C_{\mathbf{X},T}(\tau) \overset{def}{=} \begin{cases} \frac{1}{T}\sum\limits_{t=1}^{T-\tau} \left(X_t - \bar{X}_T\right)\left(X_{t+\tau} - \bar{X}_T\right)^{\intercal}, & \text{if } \mathbb{T} \equiv \mathbb{N}, \\ \frac{1}{2T+1}\sum\limits_{t=-T}^{T-\tau} \left(X_t - \bar{X}_T\right)\left(X_{t+\tau} - \bar{X}_T\right)^{\intercal}, & \text{if } \mathbb{T} \equiv \mathbb{Z}, \end{cases} \quad \forall \tau = 0, 1, \dots, T-1. \tag{7.3}$$

Osservazione 463 *We have*

$$C_{\mathbf{X},T}(0) = S^2_{\mathbf{X},T}.$$

Definizione 464 *Fixed any $T \in \mathbb{N}$, we call* time autocovariance essential estimator *of size T of $\mathbf{X}$ at lag (shift) τ the statistic*

$$G_{\mathbf{X},T}(\tau) \overset{def}{=} \begin{cases} \frac{1}{T}\sum\limits_{t=1}^{T-\tau} X_t X_{t+\tau}, & \text{if } \mathbb{T} \equiv \mathbb{N}, \\ \frac{1}{2T+1}\sum\limits_{t=-T}^{T-\tau} X_t X_{t+\tau}, & \text{if } \mathbb{T} \equiv \mathbb{Z}, \end{cases} \quad \forall \tau = 0, 1, \dots, T-1. \tag{7.4}$$

Note that in Equations (7.3) and (7.4) the factor $1/T$ [resp. $1/(2T+1)$] is sometimes replaced by the factor $1/(T-\tau)$ [resp. $1/(2T+1-\tau)$].

Definizione 465 *Fixed any $T \in \mathbb{N}$, assume that* $\det(\operatorname{diag}(C_{\mathbf{X},T}(0))) \neq 0$*, where* $\det : \mathbb{R}^{N^2} \to \mathbb{R}$ *is the determinant function on $\mathbb{R}^{N^2}$ and* $\operatorname{diag}(C_{\mathbf{X},T}(0))$ *is the diagonal matrix having for diagonal entries the corresponding diagonal entries of $C_{\mathbf{X},T}(0)$. We call* time autocorrelation estimator *of size T of $\mathbf{X}$ at the shift τ the statistic*

$$R_{\mathbf{X},T}(\tau) \overset{def}{=} \operatorname{diag}(C_{\mathbf{X},T}(0))^{-\frac{1}{2}} C_{\mathbf{X},T}(\tau) \operatorname{diag}(C_{\mathbf{X},T}(0))^{-\frac{1}{2}}, \quad \forall \tau = 0, 1, \dots, T-1. \tag{7.5}$$

Proposizione 466 *For any $T \in \mathbb{N}$, the* time average estimator $\bar{X}_T$ *of size T of $\mathbf{X}$ is an unbiased estimator of $\mu_{\mathbf{X}}$ and its mean square error is given by*

$$\operatorname{trace}\left(Var\left(\bar{X}_T\right)\right) = \begin{cases} \frac{1}{T}\left(\sum\limits_{\tau=-(T-1)}^{T-1}\left(1-\frac{|\tau|}{T}\right)\operatorname{trace}\left(\Gamma_{\mathbf{X},1}\left(\tau\right)\right)\right), & \text{if } \mathbb{T}\equiv\mathbb{N}, \\ \frac{1}{2T+1}\left(\sum\limits_{\tau=-2T}^{2T}\left(1-\frac{|\tau|}{2T+1}\right)\operatorname{trace}\left(\Gamma_{\mathbf{X},0}\left(\tau\right)\right)\right), & \text{if } \mathbb{T}\equiv\mathbb{Z}, \end{cases} \tag{7.6}$$

where trace : $\mathbb{R}^{N^2} \to \mathbb{R}$ *is the trace function on $\mathbb{R}^{N^2}$ and $\Gamma_{X,1}(\tau)$ [resp. $\Gamma_{X,0}(\tau)$] is the value at the shift τ of the reduced autocovariance of* $\mathbf{X}$ *referred to $t_0 = 1$ [resp. $t_0 = 0$]. (see Definition* **??***).*

Proof. The unbiasedness of the estimator $\bar{X}_T$ being rather evident, we restrict ourselves to compute $Var\left(\bar{X}_T\right)$. To this goal, assume first $\mathbb{T} \equiv \mathbb{N}$. Then, on account of the weak stationarity of $\mathbf{X}$, a straightforward computation yields

$$\begin{aligned} Var\left(\bar{X}_T\right) &= Var\left[\frac{1}{T}\sum_{t=1}^{T}X_t\right] = \frac{1}{T^2}Var\left[\sum_{t=1}^{T}X_t\right] \\ &= \frac{1}{T^2}\left(\mathbf{E}\left[\left(\sum_{t=1}^{T}X_t\right)\left(\sum_{t=1}^{T}X_t\right)^{\intercal}\right] - \mathbf{E}\left[\sum_{t=1}^{T}X_t\right]\mathbf{E}\left[\sum_{t=1}^{T}X_t\right]^{\intercal}\right) \\ &= \frac{1}{T^2}\mathbf{E}\left[\left(\sum_{t=1}^{T}X_t\right)\left(\sum_{t=1}^{T}X_t^{\intercal}\right)\right] - \left(\sum_{t=1}^{T}\mathbf{E}\left[X_t\right]\right)\left(\sum_{t=1}^{T}\mathbf{E}\left[X_t\right]\right)^{\intercal} \\ &= \frac{1}{T^2}\left(\mathbf{E}\left[\sum_{s,t=1}^{T}X_sX_t^{\intercal}\right] - \left(\sum_{t=1}^{T}\mathbf{E}\left[X_t\right]\right)\left(\sum_{t=1}^{T}\mathbf{E}\left[X_t\right]^{\intercal}\right)\right) \\ &= \frac{1}{T^2}\left(\sum_{s,t=1}^{T}\mathbf{E}\left[X_sX_t^{\intercal}\right] - \sum_{s,t=1}^{T}\mathbf{E}\left[X_s\right]\mathbf{E}\left[X_t\right]^{\intercal}\right) \\ &= \frac{1}{T^2}\sum_{s,t=1}^{T}\left(\mathbf{E}\left[X_sX_t^{\intercal}\right] - \mathbf{E}\left[X_s\right]\mathbf{E}\left[X_t^{\intercal}\right]\right) \\ &= \frac{1}{T^2}\sum_{s,t=1}^{T}Cov\left(X_s,X_t\right) \\ &= \frac{1}{T^2}\sum_{s,t=1}^{T}\Gamma_{\mathbf{X},1}\left(t-s\right), \end{aligned} \tag{7.7}$$

where

$$
\begin{aligned}
&\sum_{s,t=1}^{T} \Gamma_{\mathbf{X},1}(t-s) \\
&= \sum_{t=1}^{T} \Gamma_{\mathbf{X},1}(t-1) + \sum_{t=1}^{T} \Gamma_{\mathbf{X},1}(t-2) + \cdots + \sum_{t=1}^{T} \Gamma_{\mathbf{X},1}(t-(T-1)) + \sum_{t=1}^{T} \Gamma_{\mathbf{X},1}(t-T) \\
&(\Gamma_{\mathbf{X},1}(0) + \Gamma_{\mathbf{X},1}(1) + \cdots + \Gamma_{\mathbf{X},1}(T-2) + \Gamma_{\mathbf{X},1}(T-1)) \\
&+ (\Gamma_{\mathbf{X},1}(-1) + \Gamma_{\mathbf{X},1}(0) + \cdots + \Gamma_{\mathbf{X},1}(T-3) + \Gamma_{\mathbf{X},1}(T-2)) \\
&+ \cdots + (\Gamma_{\mathbf{X},1}(2-T) + \Gamma_{\mathbf{X},1}(1-T) + \cdots + \Gamma_{\mathbf{X},1}(0) + \Gamma_{\mathbf{X},1}(1)) \\
&+ (\Gamma_{\mathbf{X},1}(1-T) + \Gamma_{\mathbf{X},1}(2-T) + \cdots + \Gamma_{\mathbf{X},1}(-1) + \Gamma_{\mathbf{X},1}(0)) \\
&T\Gamma_{\mathbf{X},1}(0) + (T-1)(\Gamma_{\mathbf{X},1}(1) + \Gamma_{\mathbf{X},1}(-1)) + \cdots + 2(\Gamma_{\mathbf{X},1}(T-2) + \Gamma_{\mathbf{X},1}(2-T)) + \Gamma_{\mathbf{X},1}(T-1) + \Gamma_{\mathbf{X},1}(1-T) \\
&= \sum_{\tau=-(T-1)}^{T-1} (T-|\tau|)\, \Gamma_{\mathbf{X},1}(\tau). \qquad (7.8)
\end{aligned}
$$

Combining (7.7) and (7.8), it follows,

$$
Var\left(\bar{X}_T\right) = \frac{1}{T^2} \sum_{\tau=-(T-1)}^{T-1} (T-|\tau|)\, \Gamma_{\mathbf{X},1}(\tau) = \frac{1}{T} \left(\sum_{\tau=-(T-1)}^{T-1} \left(1 - \frac{|\tau|}{T}\right) \Gamma_{\mathbf{X},1}(\tau) \right).
$$

Now, assume $\mathbb{T} \equiv \mathbb{Z}$. In this case, with a similar computation as above, we can write

$$
Var\left(\bar{X}_T\right) = \frac{1}{(2T+1)^2} \sum_{s,t=-T}^{T} \Gamma_{\mathbf{X},0}(t-s), \qquad (7.9)
$$

where

$$
\begin{aligned}
&\sum_{s,t=-T}^{T} \Gamma_{\mathbf{X},0}(t-s) \\
&= \sum_{t=-T}^{T} \Gamma_{\mathbf{X},0}(t+T) + \sum_{t=-T}^{T} \Gamma_{\mathbf{X},0}(t+T-1) + \cdots \\
&+ \sum_{t=-T}^{T} \Gamma_{\mathbf{X},0}(t+1) + \sum_{t=-T}^{T} \Gamma_{\mathbf{X},0}(t) + \sum_{t=-T}^{T} \Gamma_{\mathbf{X},0}(t-1) + \cdots \\
&+ \sum_{t=-T}^{T} \Gamma_{\mathbf{X},0}(t-T+1) + \sum_{t=-T}^{T} \Gamma_{\mathbf{X},0}(t-T) \\
&= \Gamma_{\mathbf{X},0}(0) + \Gamma_{\mathbf{X},0}(1) + \cdots + \Gamma_{\mathbf{X},0}(T-1) + \Gamma_{\mathbf{X},0}(T) + \Gamma_{\mathbf{X},0}(T+1) + \cdots + \Gamma_{\mathbf{X},0}(2T-1) + \Gamma_{\mathbf{X},0}(2T) \\
&+ \Gamma_{\mathbf{X},0}(-1) + \Gamma_{\mathbf{X},0}(0) + \cdots + \Gamma_{\mathbf{X},0}(T-2) + \Gamma_{\mathbf{X},0}(T-1) + \Gamma_{\mathbf{X},0}(T) + \cdots + \Gamma_{\mathbf{X},0}(2T-2) + \Gamma_{\mathbf{X},0}(2T-1) \\
&+ \cdots \\
&+ \Gamma_{\mathbf{X},0}(1-T) + \Gamma_{\mathbf{X},0}(2-T) + \cdots + \Gamma_{\mathbf{X},0}(0) + \Gamma_{\mathbf{X},0}(1) + \Gamma_{\mathbf{X},0}(2) + \cdots + \Gamma_{\mathbf{X},0}(T) + \Gamma_{\mathbf{X},0}(T+1) \\
&+ \Gamma_{\mathbf{X},0}(-T) + \Gamma_{\mathbf{X},0}(1-T) + \cdots + \Gamma_{\mathbf{X},0}(-1) + \Gamma_{\mathbf{X},0}(0) + \Gamma_{\mathbf{X},0}(1) + \cdots + \Gamma_{\mathbf{X},0}(T-1) + \Gamma_{\mathbf{X},0}(T) \\
&+ \Gamma_{\mathbf{X},0}(-(T+1)) + \Gamma_{\mathbf{X},0}(-T) + \cdots + \Gamma_{\mathbf{X},0}(-2) + \Gamma_{\mathbf{X},0}(-1) + \Gamma_{\mathbf{X},0}(0) + \cdots + \Gamma_{\mathbf{X},0}(T-2) + \Gamma_{\mathbf{X},0}(T-1) \\
&+ \cdots \\
&+ \Gamma_{\mathbf{X},0}(-2T+1) + \Gamma_{\mathbf{X},0}(-2T+2) + \cdots + \Gamma_{\mathbf{X},0}(-T) + \Gamma_{\mathbf{X},0}(-T+1) + \Gamma_{\mathbf{X},0}(-T+2) + \cdots + \Gamma_{\mathbf{X},0}(0) + \Gamma_{\mathbf{X},0}(1) \\
&+ \Gamma_{\mathbf{X},0}(-2T) + \Gamma_{\mathbf{X},0}(-2T+1) + \cdots + \Gamma_{\mathbf{X},0}(-(T-1)) + \Gamma_{\mathbf{X},0}(-T) + \Gamma_{\mathbf{X},0}(-T+1) + \cdots + \Gamma_{\mathbf{X},0}(-1) + \Gamma_{\mathbf{X},0}(0) \\
&= (2T+1)\Gamma_{\mathbf{X},0}(0) + 2T\left(\Gamma_{\mathbf{X},0}(1) + \Gamma_{\mathbf{X},0}(-1)\right) + (2T-1)\left(\Gamma_{\mathbf{X},0}(2) + \Gamma_{\mathbf{X},0}(-2)\right) \\
&+ \cdots + 3\left(\Gamma_{\mathbf{X},0}(2T-2) + \Gamma_{\mathbf{X},0}(-(2T-2))\right) + 2\left(\Gamma_{\mathbf{X},0}(2T-1) + \Gamma_{\mathbf{X},0}(-(2T-1))\right) + \left(\Gamma_{\mathbf{X},0}(2T) + \Gamma_{\mathbf{X},0}(-2T)\right) \\
&= \sum_{\tau=-2T}^{2T} (2T+1-|\tau|)\,\Gamma_{\mathbf{X},0}(\tau) \qquad (7.10)
\end{aligned}
$$

Combining (7.9) and (7.10), we obtain

$$
Var\left(\bar{X}_T\right) = \frac{1}{(2T+1)^2} \sum_{\tau=-2T}^{2T} (2T+1-|\tau|)\,\Gamma_{\mathbf{X},0}(\tau) = \frac{1}{2T+1} \sum_{\tau=-2T}^{2T} \left(1 - \frac{|\tau|}{2T+1}\right) \Gamma_{\mathbf{X},0}(\tau).
$$

The linearity of the trace operator allows to complete the proof. $\square$

Proposizione 467 *In the case $N = 1$, for any $T \in \mathbb{N}$, we have*

$$
\mathbf{D}^2\left[\bar{X}_T\right] = \begin{cases} \frac{\sigma_X^2}{T}\left(1 + 2\sum_{\tau=1}^{T-1}\left(1 - \frac{\tau}{T}\right)\rho_{\mathbf{X},1}(\tau)\right), & \text{if } \mathbb{T} \equiv \mathbb{N} \\ \frac{\sigma_X^2}{2T+1}\left(1 + 2\sum_{\tau=1}^{2T}\left(1 - \frac{\tau}{2T+1}\right)\rho_{\mathbf{X},0}(\tau)\right), & \text{if } \mathbb{T} \equiv \mathbb{Z} \end{cases}. \qquad (7.11)
$$

where $\rho_{X,1}(\tau)$ [resp. $\rho_{X,1}(\tau)$] is the value at the shift τ of the reduced autocorrelation of $\mathbf{X}$ referred to $t_0 = 1$ [resp. $t_0 = 0$] (see Definition 383).

Proof. When $N = 1$, thanks to Equation (6.40) in Remark 386, we can write

$$
\mathbf{D}^2\left[\bar{X}_T\right] = \begin{cases} \frac{\sigma_X^2}{T}\left(\sum_{\tau=-(T-1)}^{T-1}\left(1 - \frac{|\tau|}{T}\right)\rho_{\mathbf{X},1}(\tau)\right), & \text{if } \mathbb{T} \equiv \mathbb{N}, \\ \frac{\sigma_X^2}{2T+1}\left(\sum_{\tau=-2T}^{2T}\left(1 - \frac{|\tau|}{2T+1}\right)\rho_{\mathbf{X},0}(\tau)\right), & \text{if } \mathbb{T} \equiv \mathbb{Z}. \end{cases} \qquad (7.12)
$$

Now, by virtue of (**??**) in Remark **??** and (6.42) in Proposition 387, we have

$$\begin{aligned}\sum_{\tau=-(T-1)}^{T-1}\left(1-\frac{|\tau|}{T}\right)\rho_{\mathbf{X},0}(\tau) &= 1+\sum_{\tau=-(T-1)}^{1}\left(1-\frac{|\tau|}{T}\right)\rho_{\mathbf{X},0}(\tau)+\sum_{\tau=1}^{T-1}\left(1-\frac{|\tau|}{T}\right)\rho_{\mathbf{X},0}(\tau)\\ &= 1+\sum_{\tau=1}^{T-1}\left(1-\frac{\tau}{T}\right)\rho_{\mathbf{X},0}(-\tau)+\sum_{\tau=1}^{T-1}\left(1-\frac{\tau}{T}\right)\rho_{\mathbf{X},0}(\tau)\\ &= 1+2\sum_{\tau=1}^{T-1}\left(1-\frac{\tau}{T}\right)\rho_{\mathbf{X},0}(\tau) \qquad (7.13)\end{aligned}$$

Similarly,

$$\sum_{\tau=-2T}^{2T}\left(1-\frac{|\tau|}{2T+1}\right)\rho_{X,0}(\tau)=1+2\sum_{\tau=1}^{2T}\left(1-\frac{\tau}{2T+1}\right)\rho_{X,0}(\tau). \qquad (7.14)$$

Combining (7.12)-(7.14), the desidered result follows. □

Definizione 468 *We say that the process* $\mathbf{X}$ *is* probability *[resp.* mean-square*]* ergodic in the mean *if*

$$\bar{X}_T \xrightarrow{\boldsymbol{P}} \mu_{\mathbf{X}} \quad [resp.\ \bar{X}_T \xrightarrow{\boldsymbol{L}^2} \mu_{\mathbf{X}}], \qquad (7.15)$$

as $T\to\infty$.

Proposizione 469 *The process* $\mathbf{X}$ *is mean-square ergodic in the mean if and only if*

$$\lim_{T\to+\infty}\operatorname{trace}\left(Var\left(\bar{X}_T\right)\right)=0. \qquad (7.16)$$

A SSS process does not need to be mean-square ergodic in the mean, not even in probability.

Esempio 470 *With reference to Example* **??***, set* $\mathbb{T}=\mathbb{N}$*. Then the process* $\mathbf{X}$ *is not probability ergodic in the mean.*

Discussion. We know that

$$\mu_{\mathbf{X}}(t)=0 \quad \text{and} \quad \Sigma_{\mathbf{X}}(t)=\Sigma_Y$$

for every $t\in\mathbb{T}$. Furthermore,

$$\Gamma_{\mathbf{X}}(s,t)=\Sigma_Y,$$

for all $s,t\in\mathbb{T}$. As a consequence, the reduced covariance function $\Gamma_{\mathbf{X},0}:\mathbb{Z}\to\mathbb{R}^{N^2}$ is given by

$$\Gamma_{\mathbf{X},0}(\tau)=\Gamma_{\mathbf{X}}(0,\tau)=\Sigma_Y,$$

for every $\tau\in\mathbb{N}$. Probability ergodicity in the mean requires that

$$\bar{X}_T\xrightarrow{\mathbf{P}}0$$

as $T\to\infty$. On the other hand,

$$\bar{X}_T\equiv\frac{1}{T}\sum_{t=1}^{T}X_t=\frac{1}{T}\sum_{t=1}^{T}Y=Y.$$

This prevents the ergodicity. □

Esempio 471 *With reference to Example 359, set* $\mathbb{T}=\mathbb{N}$*. Then the process* X *is probability ergodic in the mean.*

Discussion. We know that

$$\mu_{\mathbf{X}}(t) = 0 \quad \text{and} \quad \Sigma_{\mathbf{X}}(t) = 1.$$

Furthermore

$$\gamma_{\mathbf{X}}(s, t) = 0,$$

for all $s, t \in \mathbb{T}$ such that $s \neq t$. As a consequence, the reduced covariance function $\gamma_{\mathbf{X},1} : \mathbb{N} \to \mathbb{R}$ is given by

$$\gamma_{\mathbf{X},1}(\tau) = \gamma_{\mathbf{X}}(1, 1 + \tau) = 0,$$

for every $\tau \in \hat{n}$. We have also

$$\bar{X}_T = \frac{1}{T}\sum_{t=1}^{T} X_t = \frac{1}{T}\sum_{t=1}^{T} Y_t Z = \left(\frac{1}{T}\sum_{t=1}^{T} Y_t\right) Z.$$

Hence,

$$\mathbf{E}\left[\bar{X}_T\right] = \mathbf{E}\left[\left(\frac{1}{T}\sum_{t=1}^{T} Y_t\right) Z\right] = \left(\frac{1}{T}\sum_{t=1}^{T} \mathbf{E}\left[Y_t\right]\right) \mathbf{E}\left[Z\right] = 0 = \mu_{\mathbf{X}}(t),$$

for every $t \in \mathbb{T}$. Now, we have

$$\bar{Y}_T \xrightarrow{\mathbf{P}} 0.$$

This implies that also

$$\bar{X}_T \xrightarrow{\mathbf{P}} 0.$$

□

Esempio 472 *With reference to Example 360, set* $\mathbb{T} = \mathbb{N}$. *Then the process* $\mathbf{X}$ *is not probability ergodic in the mean.*

Discussion. We know that

$$\mu_{\mathbf{X}}(t) = 0 \quad \text{and} \quad \Sigma_{\mathbf{X}}(t) = \frac{4}{3}.$$

Furthermore

$$\gamma_{\mathbf{X}}(s, t) = 1,$$

for all $s, t \in \mathbb{T}$ such that $s \neq t$. As a consequence, the reduced covariance function $\gamma_{\mathbf{X},1} : \mathbb{N} \to \mathbb{R}$ is given by

$$\gamma_{\mathbf{X},1}(t) = \gamma_{\mathbf{X}}(1, 1 + \tau) = 1,$$

for every $\tau \in \hat{n}$. Probability ergodicity in the mean requires that

$$\bar{X}_T \xrightarrow{\mathbf{P}} 0$$

as $T \to \infty$. We have,

$$\bar{X}_T = \frac{1}{T}\sum_{t=1}^{T} X_t = \frac{1}{T}\sum_{t=1}^{T} (Y_t + Z) = \frac{1}{T}\sum_{t=1}^{T} Y_t + Z.$$

Hence,

$$\mathbf{E}\left[\bar{X}_T\right] = \mathbf{E}\left[\frac{1}{T}\sum_{t=1}^{T} Y_t + Z\right] = \frac{1}{T}\sum_{t=1}^{T} \mathbf{E}\left[Y_t\right] + \mathbf{E}\left[Z\right] = 0 = \mu_{\mathbf{X}}(t),$$

for every $t \in \mathbb{T}$. On the other hand, we have

$$\bar{X}_T - Z = \bar{Y}_T \xrightarrow{\mathbf{P}} 0.$$

This implies

$$\bar{X}_T \xrightarrow{\mathbf{P}} Z,$$

which prevents the ergodicity. □

Teorema 473 (Law of large numbers) *Assume that the random variables in the process* $\mathbf{X}$ *are uncorrelated. Then the process* $\mathbf{X}$ *is mean-square ergodic in the mean. A fortiori is probability ergodic in the mean.*
Proof. *We have*

$$\begin{aligned}\left(\bar{X}_T - \mathbf{E}\left[\bar{X}_T\right]\right)\left(\bar{X}_T - \mathbf{E}\left[\bar{X}_T\right]\right)^{\intercal} &= \left(\frac{1}{T}\sum_{t=1}^{T} X_t - \mu_{\mathbf{X}}\right)\left(\frac{1}{T}\sum_{t=1}^{T} X_t - \mu_{\mathbf{X}}\right)^{\intercal} \\ &= \frac{1}{T^2}\left(\sum_{t=1}^{T}\left(X_t - \mu_{\mathbf{X}}\right)\right)\left(\sum_{t=1}^{T}\left(X_t - \mu_{\mathbf{X}}\right)\right)^{\intercal} \\ &= \frac{1}{T^2}\sum_{s,t=1}^{T}\left(X_s - \mu_{\mathbf{X}}\right)\left(X_t - \mu_{\mathbf{X}}\right)^{\intercal}.\end{aligned}$$

Therefore,

$$\begin{aligned}Var\left(\bar{X}_T\right) &= \mathbf{E}\left[\left(\bar{X}_T - \mathbf{E}\left[\bar{X}_T\right]\right)\left(\bar{X}_T - \mathbf{E}\left[\bar{X}_T\right]\right)^{\intercal}\right] = \frac{1}{T^2}\mathbf{E}\left[\sum_{s,t=1}^{T}\left(X_s - \mu_{\mathbf{X}}\right)\left(X_t - \mu_{\mathbf{X}}\right)^{\intercal}\right] \\ &= \frac{1}{T^2}\sum_{s,t=1}^{T}\mathbf{E}\left[\left(X_s - \mu_{\mathbf{X}}\right)\left(X_t - \mu_{\mathbf{X}}\right)^{\intercal}\right] \\ &= \frac{1}{T^2}\sum_{t=1}^{T}\mathbf{E}\left[\left(X_t - \mu_{\mathbf{X}}\right)\left(X_t - \mu_{\mathbf{X}}\right)^{\intercal}\right] + \frac{1}{T^2}\sum_{\substack{s,t=1\\ s\neq t}}^{T}\mathbf{E}\left[\left(X_s - \mu_{\mathbf{X}}\right)\left(X_t - \mu_{\mathbf{X}}\right)^{\intercal}\right] \\ &= \frac{1}{T^2}\sum_{t=1}^{T}\Sigma_{\mathbf{X}} + \frac{1}{T^2}\sum_{\substack{s,t=1\\ s\neq t}}^{T} Cov\left(X_s, X_t\right).\end{aligned}$$

Now, since the random variables in $\mathbf{X}$ *are uncorrelated, we obtain*

$$Var\left(\bar{X}_T\right) = \frac{1}{T^2}\sum_{t=1}^{T}\Sigma_{\mathbf{X}} = \frac{\Sigma_{\mathbf{X}}}{T}.$$

The latter implies

$$\lim_{T\to+\infty}\operatorname{trace}\left(Var\left(\bar{X}_T\right)\right) = \lim_{T\to+\infty}\frac{\operatorname{trace}\left(\Sigma_{\mathbf{X}}\right)}{T} = 0.$$

Hence, the desired result follows from Proposition 469. □

To obtain the claim of Theorem 473, for simplicity, we have referred to a WSS process $\mathbf{X}$ with uncorrelated random variables. However, the same claims holds true in more general settings. Eventually, it is possible to prove

Teorema 474 (Law of large numbers - first generalization) *Let $(X_t)_{t\in\mathbb{N}} \equiv \mathbf{X}$ a real stochastic process of order 2 on a probability space $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$. Write $\mu_t \equiv \mathbf{E}[X_t]$ and $\sigma_t^2 \equiv \mathbf{D}^2[X_t]$, for every $t \in \mathbb{N}$. Assume there exist $\mu \in \mathbb{R}$ and $\sigma^2 \in \mathbb{R}_+$ such that*

$$\lim_{T\to+\infty} \frac{1}{T}\sum_{t=1}^{T} \mu_t \equiv \mu$$

and

$$\sigma_t^2 \leqslant \sigma^2,$$

for every $t \in \mathbb{N}$. Moreover, assume that the random variables in $\mathbf{X}$ are uncorrelated. Then we have

$$\bar{X}_T \xrightarrow{\boldsymbol{L}^2} \mu.$$

A fortiori,

$$\bar{X}_T \xrightarrow{\boldsymbol{P}} \mu.$$

Teorema 475 (Law of the large numbers - second generalization) *Let $(X_t)_{t\in\mathbb{N}} \equiv \mathbf{X}$ a real stochastic process of order 2 on a probability space $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$. Assume there exists $\mu_{\mathbf{X}} \in \mathbb{R}$ such that*

$$\mathbf{E}[X_t] = \mu_{\mathbf{X}},$$

for every $t \in \mathbb{N}$ and

$$\lim_{T\to+\infty} \frac{1}{T^2}\sum_{s,t=1}^{T} Cov(X_s, X_t) = 0.$$

Then the process $\mathbf{X}$ is mean-square ergodic in the mean. A fortiori is probability ergodic in the mean.

Assume again that $(X_t)_{t\in\mathbb{T}} \equiv \mathbf{X}$ is a WSS process. Furthermore, assume that $\mathbb{T} \equiv \mathbb{N}$, we have

Teorema 476 (Slutsky) *The process $\mathbf{X}$ is mean-square ergodic in the mean if and only if*

$$\lim_{T\to+\infty} \frac{1}{T}\sum_{\tau=0}^{T-1} \Gamma_{\mathbf{X},1}(\tau) = 0.$$

Osservazione 477 *We have*

$$\frac{1}{T}\sum_{\tau=0}^{T-1} \Gamma_{\mathbf{X},1}(\tau) = Cov\left(\bar{X}_T, X_T\right).$$

Proof. A straightforward computation yields

$$Cov\left(\bar{X}_T, X_T\right) = Cov\left(\frac{1}{T}\sum_{t=1}^{T} X_t, X_T\right) = \frac{1}{T}\sum_{t=1}^{T} Cov(X_t, X_T) = \frac{1}{T}\sum_{\tau=0}^{T-1} Cov(X_{T-\tau}, X_T) = \frac{1}{T}\sum_{\tau=0}^{T-1} \Gamma_{\mathbf{X},1}(\tau),$$

as desidered. □

Lemma 478 *Let (a_n) a sequence in $\mathbb{R}^L$, for some $L \in \mathbb{N}$, such that there exists*

$$\lim_{n\to\infty} a_n = a. \tag{7.17}$$

Then, we have

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} a_k = a. \tag{7.18}$$

Proof. Under Assumption (478), for any $\varepsilon > 0$ there exists m_ε such that for every $n > m_\varepsilon$ we have

$$\|a_n - a\|_2 < \frac{\varepsilon}{2}.$$

Therefore, we can write

$$\begin{aligned}\left\|\frac{1}{n}\sum_{k=1}^{n} a_k - a\right\|_2 &= \left\|\frac{1}{n}\sum_{k=1}^{n} (a_k - a)\right\|_2 \leqslant \frac{1}{n}\sum_{k=1}^{n} \|a_k - a\|_2 \\ &= \frac{1}{n}\left(\sum_{k=1}^{m_\varepsilon} \|a_k - a\|_2 + \sum_{k=m_\varepsilon+1}^{n} \|a_k - a\|_2\right) \\ &< \frac{1}{n}\left(\sum_{k=1}^{m_\varepsilon} \|a_k - a\|_2 + \sum_{k=m_\varepsilon+1}^{n} \frac{\varepsilon}{2}\right) \\ &= \frac{1}{n}\sum_{k=1}^{m_\varepsilon} \|a_k - a\|_2 + \frac{n - m_\varepsilon}{n}\frac{\varepsilon}{2} \\ &< \frac{1}{n}\sum_{k=1}^{m_\varepsilon} \|a_k - a\|_2 + \frac{\varepsilon}{2}.\end{aligned}$$

In the end, since the sum $\sum_{k=1}^{m_\varepsilon} \|a_k - a\|_2$ does not depend on n, we can choose $n_\varepsilon > m_\varepsilon$ such that

$$\frac{1}{n}\sum_{k=1}^{m_\varepsilon} \|a_k - a\|_2 < \frac{\varepsilon}{2}.$$

We then obtain

$$\left\|\frac{1}{n}\sum_{k=1}^{n} a_k - a\right\|_2 < \varepsilon,$$

for every $n > n_\varepsilon$, which proves Equation (**??**). $\square$

Teorema 479 *Assume that*

$$\lim_{\tau\to+\infty} \Gamma_{\mathbf{X},1}(\tau) = 0.$$

Then the process X is mean-square ergodic in the mean.

Teorema 480 *In the case $N = 1$, assume that*

$$\lim_{\tau\to+\infty} \rho_{X,1}(\tau) = 0.$$

Then the process X is mean-square ergodic in the mean.

Assume again that $\mathbb{T} \equiv \mathbb{N}$ or $\mathbb{T} \equiv \mathbb{Z}$.

Definizione 481 *We say that the process* $\mathbf{X}$ *is* probability *[resp.* mean-square*]* ergodic in the variance-covariance *if*

$$S^2_{\mathbf{X},T} \xrightarrow{\boldsymbol{P}} \Sigma_{\mathbf{X}} \quad \textit{[resp. } S^2_{\mathbf{X},T} \xrightarrow{\boldsymbol{L}^2} \Sigma_{\mathbf{X}}\textit{]},$$

as $T \to +\infty$.

Definizione 482 *We say that the process* $\mathbf{X}$ *is* probability *[resp.* mean-square*]* ergodic in the autocovariance *if*

$$C_{\mathbf{X},T}(\tau) \xrightarrow{\boldsymbol{P}} \Gamma_{\mathbf{X},1}(\tau) \quad or \quad C_{\mathbf{X},T}(\tau) \xrightarrow{\boldsymbol{P}} \Gamma_{\mathbf{X},0}(\tau) \quad [resp.\ C_{\mathbf{X},T}(\tau) \xrightarrow{\boldsymbol{L}^2} \Gamma_{\mathbf{X},1}(\tau) \quad or \quad C_{\mathbf{X},T}(\tau) \xrightarrow{\boldsymbol{L}^2} \Gamma_{\mathbf{X},0}(\tau)\,],$$

according to whether $\mathbb{T} \equiv \mathbb{N}$ *or* $\mathbb{T} \equiv \mathbb{Z}$ *for every* $\tau \in \mathbb{T}_0$*, as* $T \to +\infty$*.*

Osservazione 483 *If the process* $\mathbf{X}$ *is ergodic in the autocovariance, then* $\mathbf{X}$ *is also ergodic in the variance-covariance.*

Definizione 484 *We say that the process* $\mathbf{X}$ *is* probability *[resp.* mean-square*]* ergodic in the wide sense*, if it is ergodic both in the mean and the autocovariance.*

Lemma 485 *For simplicity set* $\mathbb{T} \equiv \mathbb{N}$*. Assume that* $\mathbf{X}$ *is a WSS real process of order* 4*. Assume also that*

$$G_{\mathbf{X},T}(\tau) \xrightarrow{\boldsymbol{P}} \mathbf{E}[X_t X_{t+\tau}] \quad [resp.\ G_{\mathbf{X},T}(\tau) \xrightarrow{\boldsymbol{L}^2} \mathbf{E}[X_t X_{t+\tau}]\,],$$

for some $\tau \in \mathbb{T}_0$*. Then*

$$C_{\mathbf{X},T}(\tau) \xrightarrow{\boldsymbol{P}} \Gamma_{\mathbf{X},1}(\tau) \quad [resp.\ C_{\mathbf{X},T}(\tau) \xrightarrow{\boldsymbol{L}^2} \gamma_{\mathbf{X},1}(\tau)\,].$$

????

Proof. *The process*

$$X_s X_t - Cov(X_s, X_t)$$

$$\mathbf{E}[X_s X_t - Cov(X_s, X_t)] = \mathbf{E}[X_s X_t] - Cov(X_s, X_t)$$

$$Y_t \stackrel{def}{=} X_t - \bar{X}_T$$

is a stationary process of order 4 *such that*

$$\mathbf{E}[Y_t] = \mathbf{E}\left[X_t - \bar{X}_T\right] = \mathbf{E}[X_t] - \mathbf{E}\left[\bar{X}_T\right] = 0$$

and

$$Cov(Y_s, Y_t) = \mathbf{E}[Y_s Y_t] = \mathbf{E}[Y_s Y_t]$$

Under the assumtion on the process $\mathbf{X}$ *we have*

$$\begin{aligned} C_{\mathbf{X},T}(\tau) &= \frac{1}{T}\sum_{t=1}^{T-\tau}\left(X_t - \bar{X}_T\right)\left(X_{t+\tau} - \bar{X}_T\right) \\ &= \frac{1}{T}\sum_{t=1}^{T-\tau} X_t X_{t+\tau} - \frac{\bar{X}_T}{T}\sum_{t=1}^{T-\tau} X_t - \frac{\bar{X}_T}{T}\sum_{t=1}^{T-\tau} X_{t+\tau} + \frac{\bar{X}_T}{T}\sum_{t=1}^{T-\tau}\bar{X}_T \\ &= G_{\mathbf{X},T}(\tau) - \frac{\bar{X}_T}{T}\left(\sum_{t=1}^{T-\tau} X_t + \sum_{t=1}^{T-\tau} X_{t+\tau} - \sum_{t=1}^{T-\tau}\bar{X}_T\right) \end{aligned}$$

Now,

$$\frac{1}{T}\sum_{t=1}^{T-\tau} X_t = \frac{1}{T}\sum_{t=1}^{T} X_t - \frac{1}{T}\sum_{t=T-\tau+1}^{T} X_t = \bar{X}_T - \frac{1}{T}\sum_{t=1}^{\tau} X_{T-\tau+t}$$

and

$$\frac{1}{T}\sum_{t=1}^{T-\tau} X_{t+\tau} = \frac{1}{T}\sum_{t=1}^{T} X_t - \frac{1}{T}\sum_{t=1}^{\tau} X_t = \bar{X}_T - \frac{1}{T}\sum_{t=1}^{\tau} X_t.$$

Therefore,

$$\begin{aligned}
C_{\mathbf{X},T}(\tau) &= \frac{1}{T}\sum_{t=1}^{T-\tau} X_t X_{t+\tau} - \bar{X}_T^2 + \frac{\bar{X}_T}{T}\sum_{t=1}^{\tau} X_{T-\tau+t} - \bar{X}_T^2 + \frac{\bar{X}_T}{T}\sum_{t=1}^{\tau} X_t + \frac{T-\tau}{T}\bar{X}_T^2 \\
&= \frac{1}{T}\sum_{t=1}^{T-\tau} X_t X_{t+\tau} + \frac{\bar{X}_T}{T}\left(\sum_{t=1}^{\tau} X_{T-\tau+t} + \sum_{t=1}^{\tau} X_t\right) - \frac{T+\tau}{T}\bar{X}_T^2 \\
&= \frac{1}{T}\sum_{t=1}^{T-\tau} X_t X_{t+\tau} + \frac{\bar{X}_T}{T}\left(\sum_{t=1}^{\tau} X_{T-\tau+t} + \sum_{t=1}^{\tau} X_t - \frac{T+\tau}{T}\sum_{t=1}^{T} X_t\right)
\end{aligned}$$

On the other hand,

$$\Gamma_{\mathbf{X},1}(\tau) = \mathbf{E}\left[(X_t - \mathbf{E}[X_t])(X_{t+\tau} - \mathbf{E}[X_{t+\tau}])\right] = \mathbf{E}[X_t X_{t+\tau}] - \mathbf{E}[X_t]\,\mathbf{E}[X_{t+\tau}] = \mathbf{E}[X_t X_{t+\tau}] - \mu_{\mathbf{X}}^2$$

Now,

$$\frac{1}{T}\sum_{t=1}^{T-\tau} X_t X_{t+\tau} \xrightarrow{\boldsymbol{P}} \mathbf{E}[X_t X_{t+\tau}] \quad \textit{[resp. } \frac{1}{T}\sum_{t=1}^{T-\tau} X_t X_{t+\tau} \xrightarrow{\boldsymbol{L}^2} \mathbf{E}[X_t X_{t+\tau}]\textit{]}$$

$$\frac{1}{T}\sum_{t=1}^{T-\tau} X_t X_{t+\tau} \xrightarrow{\boldsymbol{P}} \mathbf{E}[X_t X_{t+\tau}] \quad \textit{[resp. } \frac{1}{T}\sum_{t=1}^{T-\tau} X_t X_{t+\tau} \xrightarrow{\boldsymbol{L}^2} \mathbf{E}[X_t X_{t+\tau}]\textit{]},$$

$$, \quad \frac{1}{T}\sum_{t=1}^{\tau} X_{T-\tau+t} \xrightarrow{\boldsymbol{L}^2} 0$$

We have

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T-\tau}\left(X_t - \bar{X}_T\right)\left(X_{t+\tau} - \bar{X}_T\right) &= \frac{1}{T}\sum_{t=1}^{T-\tau}\left(X_t X_{t+\tau} - X_t\bar{X}_T - \bar{X}_T X_{t+\tau} + \bar{X}_T\bar{X}_T\right) \\
&= \frac{1}{T}\sum_{t=1}^{T-\tau} X_t X_{t+\tau} - \frac{1}{T}\sum_{t=1}^{T-\tau} X_t\bar{X}_T - \frac{1}{T}\sum_{t=1}^{T-\tau}\bar{X}_T X_{t+\tau} + \frac{1}{T}\sum_{t=1}^{T-\tau}\bar{X}_T\bar{X}_T \\
&= \frac{1}{T}\sum_{t=1}^{T-\tau} X_t X_{t+\tau} - \frac{\bar{X}_T}{T}\sum_{t=1}^{T-\tau} X_t - \frac{\bar{X}_T}{T}\sum_{t=1}^{T-\tau} X_{t+\tau} + \frac{T-\tau}{T}\bar{X}_T^2
\end{aligned}$$

$$\frac{1}{T}\sum_{t=1}^{T-\tau} X_t \xrightarrow{\boldsymbol{L}^2} \bar{X}_T, \quad \frac{1}{T}\sum_{t=1}^{T-\tau} X_{t+\tau} \xrightarrow{\boldsymbol{L}^2} \bar{X}_T,$$

$$\frac{1}{T}\sum_{t=1}^{T-\tau} X_t = \frac{1}{T}\sum_{t=1}^{T} X_t - \frac{1}{T}\sum_{t=T-\tau+1}^{T} X_t = \bar{X}_T - \frac{1}{T}\sum_{t=1}^{\tau} X_{T-\tau+t}, \quad \frac{1}{T}\sum_{t=1}^{\tau} X_{T-\tau+t} \xrightarrow{\boldsymbol{L}^2} 0$$

$$\frac{1}{T}\sum_{t=1}^{T-\tau} X_{t+\tau} = \frac{1}{T}\sum_{t=1}^{T} X_t - \frac{1}{T}\sum_{t=1}^{\tau} X_t = \bar{X}_T - \frac{1}{T}\sum_{t=1}^{\tau} X_t, \quad \frac{1}{T}\sum_{t=1}^{\tau} X_t \xrightarrow{\boldsymbol{L}^2} 0$$

□

Teorema 486 *For simplicity set $\mathbb{T} \equiv \mathbb{N}$. Assume that $\mathbf{X}$ is a WSS real process of order 4 such that*

$$\mathbf{E}[X_s X_{s+t} X_{s+u} X_{s+t+u}] = \mathbf{E}[X_1 X_{1+t} X_{1+u} X_{1+t+u}], \tag{7.19}$$

for every $s \in \mathbb{N}$ and for all $t, u \in \hat{n}$. Moreover, assume that

$$\lim_{u\to+\infty} Cov\left(X_1 X_{1+t}, X_{1+u} X_{1+t+u}\right) = 0, \tag{7.20}$$

or equivalently

$$\lim_{u\to+\infty} \mathbf{E}\left[X_1 X_{1+t} X_{1+u} X_{1+t+u}\right] = \mathbf{E}\left[X_1 X_{1+t}\right]^2, \tag{7.21}$$

for every $t \in \hat{n}$. Then X is mean square ergodic in the wide sense.

Proof. Since $\mathbf{X}$ is a WSS real process, we have

$$Cov\left(X_1, X_{1+t}\right) = Cov\left(X_{1+u}, X_{1+t+u}\right),$$

that is

$$\mathbf{E}\left[X_1 X_{1+t}\right] - \mathbf{E}\left[X_1\right]\mathbf{E}\left[X_{1+t}\right] = \mathbf{E}\left[X_{1+u} X_{1+t+u}\right] - \mathbf{E}\left[X_{1+u}\right]\mathbf{E}\left[X_{1+t+u}\right].$$

for every $u \in \hat{n}$. On the other hand,

$$\mathbf{E}\left[X_1\right]\mathbf{E}\left[X_{1+t}\right] = \mu_{\mathbf{X}}^2 = \mathbf{E}\left[X_{1+u}\right]\mathbf{E}\left[X_{1+t+u}\right].$$

It then follows

$$\mathbf{E}\left[X_1 X_{1+t}\right] = \mathbf{E}\left[X_{1+u} X_{1+t+u}\right]$$

for every $u \in \hat{n}$. The equivalence between Equations (7.20) and (7.21) immediately follows. Now, fixed any $t \in \hat{n}$, consider the process $\left(Y_s^{(t)}\right)_{s\in\mathbb{N}} \equiv \mathbf{Y}^{(t)}$, given by

$$Y_s^{(t)} \stackrel{\text{def}}{=} X_s X_{s+t}, \quad \forall s \in \mathbb{N}.$$

Since $\mathbf{X}$ is a WSS real process of order 4 we have

$$\mathbf{E}\left[Y_s^{(t)}\right] = \mathbf{E}\left[X_s X_{s+t}\right] = Cov\left(X_s, X_{s+t}\right) + \mathbf{E}\left[X_s\right]\mathbf{E}\left[X_{s+t}\right] = \gamma_{\mathbf{X},1}(t) + \mu_{\mathbf{X}}^2$$

and

$$\mathbf{E}\left[Y_{s+u}^{(t)}\right] = \mathbf{E}\left[X_{s+u} X_{s+u+t}\right] = Cov\left(X_{s+u}, X_{s+u+t}\right) + \mathbf{E}\left[X_{s+u}\right]\mathbf{E}\left[X_{s+u+t}\right] = \gamma_{\mathbf{X},1}(t) + \mu_{\mathbf{X}}^2,$$

for every $s \in \mathbb{N}$ and for every $u \in \hat{n}$. In particular,

$$\mathbf{E}\left[Y_1^{(t)}\right] = \mathbf{E}\left[Y_{1+u}^{(t)}\right] = \gamma_{\mathbf{X},1}(t) + \mu_{\mathbf{X}}^2,$$

for any $t \in \hat{n}$. Moreover, thanks to Equation (7.19), we have

$$\mathbf{E}\left[Y_s^{(t)} Y_{s+u}^{(t)}\right] = \mathbf{E}\left[X_s X_{s+t} X_{s+u} X_{s+t+u}\right] = \mathbf{E}\left[X_1 X_{1+t} X_{1+u} X_{1+t+u}\right] = \mathbf{E}\left[Y_1^{(t)} Y_{1+u}^{(t)}\right].$$

What shown above implies

$$\begin{aligned}
\gamma_{\mathbf{Y}^{(t)}}(s, s+u) &= \mathbf{E}\left[Y_s^{(t)} Y_{s+u}^{(t)}\right] - \mathbf{E}\left[Y_s^{(t)}\right]\mathbf{E}\left[Y_{s+u}^{(t)}\right] \\
&= \mathbf{E}\left[Y_1^{(t)} Y_{1+u}^{(t)}\right] - \left(\gamma_{\mathbf{X},1}(t) + \mu_{\mathbf{X}}^2\right)^2 \\
&= \mathbf{E}\left[Y_1^{(t)} Y_{1+u}^{(t)}\right] - \mathbf{E}\left[Y_1^{(t)}\right]\mathbf{E}\left[Y_{1+u}^{(t)}\right] \\
&= \gamma_{\mathbf{Y}^{(t)}}(1, 1+u) \\
&\equiv \gamma_{\mathbf{Y}^{(t)},1}(u).
\end{aligned}$$

Hence, also $\mathbf{Y}^{(t)}$ is a WSS process, for any $t \in \hat{n}$. In the end, by virtue of Equation (7.20) we have

$$\lim_{u\to+\infty} \gamma_{\mathbf{Y}^{(t)},1}(u) = 0.$$

We can then apply Theorem 479 to obtain the mean square ergodicity in the mean of the process $\mathbf{Y}^{(t)}$, for any $t \in \hat{n}$. It then follows

$$\bar{Y}_T^{(t)} \xrightarrow{\mathbf{L}^2} \mathbf{E}\left[Y_1^{(t)}\right] = \gamma_{\mathbf{X},1}(t) + \mu_{\mathbf{X}}^2,$$

that is

$$\frac{1}{T}\sum_{s=1}^{T} X_s X_{s+t} \xrightarrow{\mathbf{L}^2} \gamma_{\mathbf{X},1}(t) + \mu_{\mathbf{X}}^2 = \mathbf{E}[X_s X_{s+t}],$$

for any $t \in \hat{n}$. On account of Lemma 485, this implies the desired result. □

From a direct ispection of a time series $\mathbf{x}$, there is no way to check the ergodicity of the stochastic process $\mathbf{X}$ which might have generated $\mathbf{x}$. However, the ergodicity of the process $\mathbf{X}$ chosen as a model for at least one component of $\mathbf{x}$ is necessary to make inferences from $\mathbf{x}$. As it is shown by the presented results, ergodicity is related to the asymptotic independence of the random variables in the process. This is a rather natural property of many natural noise processes. In fact, it is rather natural to think that as time runs out the influence of the past states of the noise, which affects a stochastic phenomenon, on its current states vanishes.

7.2 Strong White Noise (SWN)

Let $\mathbb{T} \subseteq \mathbb{R}$ and let $(W_t)_{t\in\mathbb{T}} \equiv W$ be a 2nd-order stochastic process on a probability space Ω with state space $\mathbb{R}^N$, for some $N \in \mathbb{N}$.

Definizione 487 (Strong White Noise) *We say that W is an N-variate* strong white noise *(*SWN*) or an N-variate* independent and identically distributed *noise (*IIDN*) if the random variables in W are independent and identically distributed with mean* 0*. In case $N = 1$, we usually speak of* strong white noise *with no reference to N.*

Notation 488 (Strong White Noise) *To denote that W is an N-variate strong white noise we write $W \sim SWN^N(\Sigma_W)$ or $W \sim IIDN^N(\Sigma_W)$, where Σ_W is the common variance-covariance matrix of the random variables in W. In case $N = 1$, we write $W \sim SWN(\sigma_W^2)$ or $W \sim IID^N(\sigma_W^2)$, where σ_W^2 is the common variance of the random variables in W.*

Let $(W_t)_{t\in\mathbb{T}} \equiv W \sim SWN^N(\Sigma_W)$.

Osservazione 489 *The mean [resp. variance-covariance] function $\mu_W : \mathbb{T} \to \mathbb{R}^N$ [resp. $\Sigma_W : \mathbb{T} \to \mathbb{R}^{N^2}$] satisfies*

$$\mu_W(t) = 0, \quad \text{[resp. } \Sigma_W(t) = \Sigma_W\text{]}, \tag{7.22}$$

for every $t \in \mathbb{T}$. The autocovariance function $\gamma_W : \mathbb{T}^2 \to \mathbb{R}^N$ and the autocorrelation function $\rho_W : \mathbb{T} \times \mathbb{T} \to \mathbb{R}^{N^2}$ statisfy

$$\gamma_W(s,t) = \rho_W(s,t) = 0,$$

for all $s, t \in \mathbb{T}$ such that $s \neq t$.

Osservazione 490 *Fixed any $t_0 \in \mathbb{T}$, write*

$$\mathbb{T}_0 \equiv \{\tau \in \mathbb{R} : t_0 + \tau \in \mathbb{T}\}.$$

The reduced autocovariance function $\gamma_{W,t_0} : \mathbb{T}_0 \to \mathbb{R}^{N^2}$ and the reduced autocorrelation function $\rho_{W,t_0} : \mathbb{T}_0 \to \mathbb{R}^{N^2}$ of W referred to t_0 satisfy

$$\gamma_{W,t_0}(\tau) = \begin{cases} \Sigma_W, & \text{if } \tau = 0, \\ 0, & \text{if } \tau \neq 0, \end{cases} \quad \text{and} \quad \rho_{W,t_0}(\tau) = \begin{cases} I_N, & \text{if } \tau = 0, \\ 0, & \text{if } \tau \neq 0, \end{cases}$$

where I_N is the identity matrix in $\mathbb{R}^{N^2}$.

Osservazione 491 *Assume that $\mathbb{T} \equiv \mathbb{Z}$. The partial autocorrelation function $\phi_W : \mathbb{Z} \to \mathbb{R}^{N^2}$ satisfies*

$$\phi_W(\tau) = 0,$$

for every $\tau \geqslant 1$.

Proposizione 492 *The process W is strict-sense stationary.*

Proposizione 493 *The process W is mean-square ergodic in the wide sense.*

Definizione 494 (Gaussian White Noise) *We say that W is a* Gaussian white noise *if the random variables in W are Gaussian distributed.*

Notation 495 (Gaussian White Noise) *To denote that W is a Gaussian white noise we write $W \sim GWN^N(\Sigma_W)$, where Σ_W is the common variance-covariance matrix of the N-variate random variables in W. In case $N = 1$, we set $\Sigma_W \equiv \sigma_W^2$ and bypass the reference to N.*

Proposizione 496 (Gaussian White Noise) *Let W be a Gaussian process. Assume that the random variables in W are uncorrelated with mean zero and same variance-covariance. Then W is a Gaussian white noise.*

□It is enough to observe that jointly Gaussian distributed and uncorrelated N-variate random variables are independent and that the distribution of a Gaussian random variable is completely determined by the mean and the variance-covariance matrix. □

Let $(W_t)_{t\in\mathbb{T}} \equiv W$ be a strong real white noise with time set $\mathbb{T} \equiv \hat{n}$ and variance σ_W^2. Let $T > 0$, and, for any $\tau = 1, \ldots, T-1$, let $\Gamma_{W,T}(\tau)$ [resp. $\mathrm{P}_{W,T}(\tau)$] be the time sample autocovariance [resp. autocorrelation] of length T of W at lag τ.

Teorema 497 *Given any $S \in \mathbb{N}$, write I_S for the identity matrix of order S. We have*

1. *as $T \to \infty$, the vector $(\Gamma_{W,T}(1), \ldots, \Gamma_{W,T}(S))^\intercal$ converges in distribution to $N\left(0, \frac{\sigma_W^2}{T} I_S\right)$;*
2. *as $T \to \infty$, the vector $(\mathrm{P}_{W,T}(1), \ldots, \mathrm{P}_{W,T}(S))^\intercal$ converges in distribution to $N\left(0, \frac{1}{T} I_S\right)$.*

Assume that T is large and fix any $S \in \mathbb{N}$ such that $S \ll T$. As a consequence of Theorem (497), the vector $(\mathrm{P}_{W,T}(1), \ldots, \mathrm{P}_{W,T}(S))^\intercal$ is approximately Gaussian distributed with mean 0 and variance $1/T$. It follows that, given any $\alpha \in (0,1)$, an approximate confidence interval at the confidence level of $100(1-\alpha)\,\%$ for the realization

$$\hat{\mathrm{P}}_{W,T}(\tau) \equiv \frac{\sum_{t=1}^{T-\tau} (w_t - \bar{w}_T)(w_{t+\tau} - \bar{w}_T)}{\sum_{t=1}^{T} (w_t - \bar{w}_T)^2}$$

of $\mathrm{P}_{W,T}(\tau)$, corresponding to a realization $(w_t)_{t=1}^T \equiv w$ of W, is given by

$$\left(-\frac{z_{\alpha/2}}{\sqrt{T}}, \frac{z_{\alpha/2}}{\sqrt{T}}\right),$$

for every $\tau = 1, \ldots, S$, where $z_{\alpha/2}$ is the upper tail critical value of level $\alpha/2$ of $N(0,1)$. In addition, there is evidence against the null hypothesis $H_0 : \mathbf{E}[\mathrm{P}_{W,T}(\tau)] = 0$ at the approximate significance level of $100\alpha\%$ when

$$\left|\hat{\mathrm{P}}_{W,T}(\tau)\right| > \frac{z_{\alpha/2}}{\sqrt{T}} \Leftrightarrow \mathbf{P}\left(|Z| \geqslant \sqrt{T}\left|\hat{\mathrm{P}}_{W,T}(\tau)\right|\right) < \alpha/2.$$

In particular, we have the approximate confidence interval

$$\begin{array}{ll}\left(-\frac{1.645}{\sqrt{T}}, \frac{1.645}{\sqrt{T}}\right) & \text{at } 90\% \text{ c.l.} \\ \left(-\frac{1.96}{\sqrt{T}}, \frac{1.96}{\sqrt{T}}\right) & \text{at } 95\% \text{ c.l.} \\ \left(-\frac{2.575}{\sqrt{T}}, \frac{2.575}{\sqrt{T}}\right) & \text{at } 99\% \text{ c.l.}\end{array}$$

and there is evidence against the null hypothesis $H_0 : \mathbf{E}\left[\mathrm{P}_{W,T}(\tau)\right] = 0$ when

$$\begin{array}{lll}\left|\hat{\mathrm{P}}_{W,T}(\tau)\right| > \frac{1.645}{\sqrt{T}} \Leftrightarrow \mathbf{P}\left(|Z| \geqslant \sqrt{T}\left|\hat{\mathrm{P}}_{W,T}(\tau)\right|\right) < 0.01 & \text{at nearly } 10\% \text{ s.l.} \\ \left|\hat{\mathrm{P}}_{W,T}(\tau)\right| > \frac{1.96}{\sqrt{T}} \Leftrightarrow \mathbf{P}\left(|Z| \geqslant \sqrt{T}\left|\hat{\mathrm{P}}_{W,T}(\tau)\right|\right) < 0.05 & \text{at nearly } 5\% \text{ s.l.} \\ \left|\hat{\mathrm{P}}_{W,T}(\tau)\right| > \frac{2.575}{\sqrt{T}} \Leftrightarrow \mathbf{P}\left(|Z| \geqslant \sqrt{T}\left|\hat{\mathrm{P}}_{W,T}(\tau)\right|\right) < 0.01 & \text{at nearly } 1\% \text{ s.l.}\end{array}.$$

Corollary 498 *Given any $\alpha \in (0,1)$, if $100(1-\alpha)\%$ of the realizations of the time sample autocorrelations of length T of W, on varying of the time shift $\tau = 1, \ldots, S$, fall outside the interval*

$$\left(\frac{z_{\alpha/2}}{\sqrt{T}}, \frac{z_{\alpha/2}}{\sqrt{T}}\right),$$

then there is evidence against the null hypothesis that the time series is modeled by a strong white noise at the approximate significance level of $100\alpha\%$.

Let $(X_t)_{t\in\mathbb{N}} \equiv X$ be a real stochastic process and let $S, T \in \mathbb{N}$ such that $S < T$.

Definizione 499 *We call Ljung-Box statistic of lenght S on W the statistic*

$$Q_{W.T}(S) \stackrel{def}{=} T(T+2)\sum_{\tau=1}^{S} \frac{\mathrm{P}_{X,T}(\tau)}{T-\tau}.$$

Teorema 500 *Assume that T is "large"*[1] *and $S \ll T$. Then under the null hypothesis that W is a 2nd order white noise with variance σ_W^2, the Ljung-Box statistic $Q_{W.T}(S)$ has a Chi-square distribution with S degrees of freedom. In symbols, $Q_{W.T}(S) \sim \chi_S^2$. As a consequence, there is evidence against the null hypothesis $H_0 : W \sim SWN\left(\sigma_W^2\right)$ at the approximate significance level of $100\alpha\%$, for any $\alpha \in (0,1)$, when*

$$q_{w.T}(S) > \chi^2_{u,\alpha,S} \Leftrightarrow \mathbf{P}\left(\chi_S^2 \geqslant q_{w.T}(S)\right) < \alpha/2,$$

where $q_{w.T}(S)$ is the realization of $Q_{W.T}(S)$ corresponding to a realization $(w_t)_{t=1}^T \equiv w$ of W and $\chi^2_{u,\alpha,S}$ is the upper α-critical value of the chi-square distribution with S degrees of freedom.

Note that the Ljung-Box is Portmanteau test: the null hypothesis is well specified but the alternative hypothesis is more loosely specified. Actually, when there is evidence against the null hypothesis of strong white noise the alternative suggests that data are not likely to be generated via an independent sequence of random variable due to the apparent presence of serial correlation.Parameter Estimation

Let $T \in \mathbb{N}$, and let $(W_t)_{t=1}^T \equiv W$ be a strong white noise on a probability space Ω with state space $\mathbb{R}$ and variance σ_W^2. Let $\bar{W}_T$ [resp. $S_T^2(W)$] be the time sample mean $\bar{W}_T$ [resp. time sample variance] of length T of W (see Definitions **??** and **??**).

Osservazione 501 *The statistic $\bar{W}_T$ is an unbiased estimator of $\mu_W = 0$ with mean quadratic error given by $Var\left(\bar{W}_T\right)$, (see Equation (7.6) in Proposition 466).*

[1] As a rule of thumb of the pre-computer age, T is intended to be *large* if $T \geqslant 30$, according to some authors, or $T \geqslant 40$, according to others.

Proposizione 502 *Assume that T is "large". Then the approximate confidence interval for $\mu_W = 0$, at the confidence level of $100(1-\alpha)\%$, for any $\alpha \in (0,1)$, is given by*

$$\left(\bar{w}_T - z_{\alpha/2}\frac{s_T(W)}{\sqrt{T}},\ \bar{w}_T + z_{\alpha/2}\frac{s_T(W)}{\sqrt{T}}\right),$$

where $z_{\alpha/2}$ is the upper tail critical value of level $\alpha/2$ of the standard Gaussian random variable and $s_T(W)$ is the realization of the time sample standard deviation $S_T(W) \equiv \sqrt{S_T^2(W)}$. In addition, there is evidence against the null hypothesis $H_0 : \mu_W = 0$ at the approximate significance level of $100\alpha\%$, for any $\alpha \in (0,1)$, when

$$\left|\frac{\bar{w}_T\sqrt{T}}{s_T(W)}\right| > z_{\alpha/2} \Leftrightarrow \mathbf{P}\left(|Z| \geqslant \left|\frac{\bar{w}_T\sqrt{T}}{s_T(W)}\right|\right) < \alpha,$$

where Z is a standard Gaussian random variable. In particular, we have approximatively

$$\begin{array}{ll}\left(\bar{w}_T - 1.645\frac{s_T(W)}{\sqrt{T}},\ \bar{w}_T + 1.645\frac{s_T(W)}{\sqrt{T}}\right) & \text{at } 90\% \text{ c.l.}\\ \left(\bar{w}_T - 1.96\frac{s_T(W)}{\sqrt{T}},\ \bar{w}_T + 1.96\frac{s_T(W)}{\sqrt{T}}\right) & \text{at } 95\% \text{ c.l.}\\ \left(\bar{w}_T - 2.575\frac{s_T(W)}{\sqrt{T}},\ \bar{w}_T + 2.575\frac{s_T(W)}{\sqrt{T}}\right) & \text{at } 99\% \text{ c.l.}\end{array}$$

and there is evidence against the null hypothesis $H_0 : \mu_W = 0$ when

$$\begin{array}{lll}\left|\frac{\bar{w}_T\sqrt{T}}{s_T(W)}\right| > 1.645 \Leftrightarrow \mathbf{P}\left(|Z| \geqslant \left|\frac{\bar{w}_T\sqrt{T}}{s_T(W)}\right|\right) < 0.1 & \text{at nearly } 10\% \text{ s.l.}\\ \left|\frac{\bar{w}_T\sqrt{T}}{s_T(W)}\right| > 1.96 \Leftrightarrow \mathbf{P}\left(|Z| \geqslant \left|\frac{\bar{w}_T\sqrt{T}}{s_T(W)}\right|\right) < 0.05 & \text{at nearly } 5\% \text{ s.l.}\\ \left|\frac{\bar{w}_T\sqrt{T}}{s_T(W)}\right| > 2.575 \Leftrightarrow \mathbf{P}\left(|Z| \geqslant \left|\frac{\bar{w}_T\sqrt{T}}{s_T(W)}\right|\right) < 0.01 & \text{at nearly } 1\% \text{ s.l.}\end{array}.$$

Proposizione 503 *Assume that W is Gaussian. Then the confidence interval for $\mu_W = 0$, at the confidence level of $100(1-\alpha)\%$, for any $\alpha \in (0,1)$, is given by*

$$\left(\bar{w}_T - t_{T-1,\alpha/2}\frac{s_T(W)}{\sqrt{T-1}},\ \bar{w}_T + t_{T-1,\alpha/2}\frac{s_T(W)}{\sqrt{T-1}}\right),$$

where $t_{T-1,\alpha/2}$ is the upper tail critical value of level $\alpha/2$ of the Student random variable with $T-1$ degree of freedom and $s_T(W)$ is the realization of the time sample standard deviation $S_T(W) \equiv \sqrt{S_T^2(W)}$. In addition, there is evidence against the null hypothesis $H_0 : \mu_W = 0$ at the significance level of $100\alpha\%$, for any $\alpha \in (0,1)$, when

$$\left|\frac{\bar{w}_T}{s_T(W)/\sqrt{T-1}}\right| > t_{T-1,\alpha/2} \Leftrightarrow \mathbf{P}\left(|X_{T-1}| \geqslant \left|\frac{\bar{w}_T}{s_T(W)/\sqrt{T-1}}\right|\right) < \alpha/2,$$

where X_{T-1} is a Student random variable with $T-1$ degree of freedom.

Proposizione 504 *Assume that W is a fourth order process and T is "large". Then the approximate confidence interval for σ_W, at the confidence level of $100(1-\alpha)\%$, for any $\alpha \in (0,1)$, is given by*

$$\left(\frac{s_T^2(W)}{1 - z_{\alpha/2}\sqrt{(kurt_T(W)-1)/T}},\frac{s_T^2(W)}{1 + z_{\alpha/2}\sqrt{(kurt_T(W)-1)/T}}\right),$$

where $z_{\alpha/2}$ is the upper tail critical value of level $\alpha/2$ of the standard Gaussian random variable and $kurt_T(W)$ is the realization of the time sample kurtosis $Kurt_T(W)$ of W. In addition, there is evidence *against the null hypothesis* $H_0 : \sigma_W = \sigma$ *at the approximate significance level of* $100\alpha\%$, *for any* $\alpha \in (0,1)$, *when*

$$\left|\frac{s_T^2(W) - \sigma^2}{\sigma^2\sqrt{(kurt_T(W) - 1)/n}}\right| > z_{\alpha/2} \Leftrightarrow \mathbf{P}\left(|Z| \geqslant \left|\frac{s_T^2(W) - \sigma^2}{\sigma^2\sqrt{(kurt_T(W) - 1)/n}}\right|\right) < \alpha,$$

where Z *is a standard Gaussian random variable.*

Proposizione 505 *Assume that* W *is Gaussian. Then the confidence interval for* σ_W^2, *at the confidence level of* $100(1-\alpha)\%$, *for any* $\alpha \in (0,1)$, *is given by*

$$\left(\frac{Ts_T^2(W)}{\chi^2_{u,\alpha/2,T-1}}, \frac{Ts_T^2(W)}{\chi^2_{\ell,\alpha/2,T-1}}\right)$$

where $\chi^2_{u,\alpha/2,T-1}$ *[resp.* $\chi^2_{\ell,\alpha/2,T-1}$*] is the upper [resp. lower]* $\alpha/2$*-critical value of the chi-square distribution with* T *degrees of freedom and* $s_T^2(W)$ *is the realization of the time sample variance* $S_T^2(W)$. *In addition, there is evidence against the null hypothesis* $H_0 : \sigma_W = \sigma$ *at the significance level of* $100\alpha\%$, *for any* $\alpha \in (0,1)$, *when*

$$\frac{Ts_T^2(W)}{\sigma^2} < \chi^2_{\ell,\alpha/2,T-1} \text{ or } \frac{Ts_T^2(W)}{\sigma^2} > \chi^2_{u,\alpha/2,T-1} \Leftrightarrow \min\left\{\mathbf{P}\left(\chi^2_{T-1} < \frac{Ts_T^2(W)}{\sigma^2}\right), \mathbf{P}\left(\chi^2_{T-1} > \frac{Ts_T^2(W)}{\sigma^2}\right)\right\} < \alpha/2,$$

where χ^2_{T-1} *is a Chi-square random variable with* $T-1$ *degree of freedom.*

Note that a Chi-square test is rather sensitive to deviations from the Gaussian distribution. Unlike the Student distribution, the Chi-square distribution is not robust to deviation from normality of the population distribution. If the white noise distribution is not Gaussian or close enough to Gaussian, possibly the null hypothesis will be mistakenly rejected.

Assume that W is a real Gaussian white noise, $W \sim GWN(\sigma_W^2)$. Then also the forecast $\hat{W}_{T+k,T}$ of the state W_{T+k} is Gaussian for every $k \in \mathbb{N}$. Therefore, for any $\alpha \in (0,1)$, a $100(1-\alpha)\%$ prediction interval for the true value of the state W_{T+k} is given by

$$\left(\hat{W}_{T+k,T} - z_{\alpha/2}\sigma_W(T),\ \hat{W}_{T+k,T} + z_{\alpha/2}\sigma_W(T)\right), \tag{7.23}$$

where $z_{\alpha/2} \equiv z^+_{\alpha/2}$ is the upper tail critical value of level $\alpha/2$ of the standard Gaussian random variable. The realization of the predicton interval is then

$$\left(\hat{w}_{T+k,T} - z_{\alpha/2}\hat{\sigma}_W(T),\ \hat{w}_{T+k,T} + z_{\alpha/2}\hat{\sigma}_W(T)\right), \tag{7.24}$$

where $\hat{w}_{T+k,T}$ is the realization of the estimator $\hat{W}_{T+k,T}$ of the state W_{T+k} and $\hat{\sigma}_W(T)$ is the estimated value of the parameter σ_W.

7.3 Prediction of Future States and Prediction Intervals

Let $(W_t)_{t\in\mathbb{T}} \equiv \mathbf{W}$ be a strong white noise on a probability space $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ with state space $\mathbb{R}^N$, for some $N \in \mathbb{N}$. For sake of simplicity, assume that $\mathbb{T} \equiv \mathbb{N}$. Write $\mu_\mathbf{W}$ [resp. $\Sigma_\mathbf{W}$] for the constant value of the mean [resp. variance-covariance] function of $\mathbf{W}$. For any $S, T \in \mathbb{N}$, write W_T [resp. W_{T+S}] for the state of the process $\mathbf{W}$ at the current time T [resp. future time $T+S$] and write $\hat{W}_{T+S|T}$ for the minimum square error predictor of the future state W_{T+S} of the process, given the information $\mathcal{F}_T \equiv \sigma(W_1, \ldots, W_T)$ generated by the process $\mathbf{W}$ itself up to the time T included.

Proposizione 506 (Future State Predictor of a SWN) *The time average estimator $\bar{W}_T$ is a point estimator for W_{T+S}, for all $S, T \in \mathbb{N}$.*

□Writing $\mathbf{E}[\cdot \mid \mathcal{F}_T]$ for the conditional expectation operator given the information $\mathcal{F}_T$, we know that

$$\hat{W}_{T+S|T} \overset{\text{def}}{=} \underset{Y \in L^2(\Omega_{\mathcal{F}_T};\mathbb{R}^N)}{\arg\min} \left\{ \mathbf{E}\left[(W_{T+S} - Y)^2 \right] \right\} = \mathbf{E}[W_{T+S} \mid \mathcal{F}_T], \tag{7.25}$$

for all $S, T \in \mathbb{N}$. Now, since the N-variate random variables in $\mathbf{W}$ are independent and the mean function of $\mathbf{W}$ is constant, we have

$$\mathbf{E}[W_{T+S} \mid \mathcal{F}_T] = \mathbf{E}[W_{T+S}] = \mu_{\mathbf{W}}, \tag{7.26}$$

for all $S, T \in \mathbb{N}$. Combining (7.25) and (7.26), we obtain

$$\hat{W}_{T+S|T} = \mu_{\mathbf{W}}, \tag{7.27}$$

for all $S, T \in \mathbb{N}$. On the other hand, $\bar{W}_T$ is a point estimator for $\mu_{\mathbf{W}}$ and the desired claim follows. □

Proposizione 507 (Predition Intervals for Future States of a GSWN) *In case $N = 1$, assume that $\mathbf{W}$ is Gaussian, that is $\mathbf{W} \sim GWN(\sigma^2_{\mathbf{W}})$, for some $\sigma_{\mathbf{W}} > 0$. Then, a prediction interval for the state W_{T+S}, for any $S \in \mathbb{N}$, at the confidence level of $100(1-\alpha)\%$, for any $\alpha \in (0,1)$, is given by*

$$\left(\bar{W}_T - t_{T-1,\alpha/2} S_{\mathbf{W},T} \sqrt{1 + \frac{1}{T}},\ \bar{W}_T + t_{T-1,\alpha/2} S_{\mathbf{W},T} \sqrt{1 + \frac{1}{T}} \right), \tag{7.28}$$

where $t_{T-1,\alpha/2}$ is the upper tail critical value of level $\alpha/2$ of the standard Student's t_{T-1} distribution with $T-1$ degree of freedom and $S_{\mathbf{W},T} \equiv \sqrt{S^2_{\mathbf{W},T}}$ is the time standard deviation of $\mathbf{W}$. Hence, a realization of the prediction interval (7.28) is given by

$$\left(\bar{w}_T - t_{T-1,\alpha/2} s_{\mathbf{W},T} \sqrt{1 + \frac{1}{T}},\ \bar{w}_T + t_{T-1,\alpha/2} s_{\mathbf{W},T} \sqrt{1 + \frac{1}{T}} \right), \tag{7.29}$$

where $\bar{w}_T$ [resp. $s_{\mathbf{W},T}$] is the realization of the time average estimator $\bar{W}_T$ [resp. time standard deviation $S_{\mathbf{W},T}$] of $\mathbf{W}$.

□By virtue of Proposition (506) and the Gaussianity assumption on $\mathbf{W}$, the statistic

$$\frac{W_{T+S} - \hat{W}_{T+S|T}}{S_{\mathbf{W},T}\sqrt{1 + \frac{1}{T}}} = \frac{W_{T+S} - \bar{W}_T}{S_{\mathbf{W},T}\sqrt{1 + \frac{1}{T}}} \equiv X$$

has the standard Student's t_{T-1} distribution with $T-1$ degrees of freedom. In fact, since $\mathbf{W}$ is a Gaussian strong white noise, the random variable $(W_{T+S} - \bar{W}_T)/\sigma_{\mathbf{W}}\sqrt{1 + \frac{1}{T}} \equiv Z$ is Gaussian distributed. Moreover, we have

$$\mathbf{E}[Z] = \frac{\mathbf{E}[W_{T+S}] - \mathbf{E}[\bar{W}_T]}{\sigma_{\mathbf{W}}\sqrt{1 + \frac{1}{T}}} = 0$$

and

$$\mathbf{D}^2[Z] = \frac{\mathbf{D}^2[W_{T+S}] + \mathbf{D}^2[\bar{W}_T]}{\sigma^2_{\mathbf{W}}\left(1 + \frac{1}{T}\right)} = \frac{\sigma^2_{\mathbf{W}} + \frac{1}{T}\sigma^2_{\mathbf{W}}}{\sigma^2_{\mathbf{W}}\left(1 + \frac{1}{T}\right)} = 1.$$

That is $Z \sim N(0,1)$. On the other hand, the random variable $(T-1)\, S^2_{\mathbf{W},T}/\sigma^2_{\mathbf{W}} \equiv Y$ has the chi-square distribution χ^2_{T-1} with $T-1$ degrees of freedom. It follows that the statistic

$$\frac{Z}{\sqrt{Y/(T-1)}} \equiv \frac{\frac{W_{T+S}-\bar{W}_T}{\sigma_{\mathbf{W}}\sqrt{1+\frac{1}{T}}}}{\sqrt{\frac{(T-1)S^2_{\mathbf{W},T}}{\sigma^2_{\mathbf{W}}}/(T-1)}} = X$$

has the standard Student's t_{T-1} distribution with $T-1$ degrees of freedom. As a consequence, we can write

$$\mathbf{P}\left(\left|\frac{W_{T+S}-\hat{W}_{T+S|T}}{S_{\mathbf{W},T}\sqrt{1+\frac{1}{T}}}\right| < t_{T-1,\alpha/2}\right) < 1-\alpha,$$

where $t_{T-1,\alpha/2}$ is the upper tail critical value of level $\alpha/2$ of the standard Student's t_{T-1} distribution with $T-1$ degree of freedom. Now, we have

$$\begin{aligned}
\left|\frac{W_{T+S}-\hat{W}_{T+S|T}}{S_{\mathbf{W},T}\sqrt{1+\frac{1}{T}}}\right| < t_{T-1,\alpha/2} &\Leftrightarrow -t_{T-1,\alpha/2} < \frac{W_{T+S}-\hat{W}_{T+S|T}}{S_{\mathbf{W},T}\sqrt{1+\frac{1}{T}}} < t_{T-1,\alpha/2} \\
&\Leftrightarrow \hat{W}_{T+S|T} - t_{T-1,\alpha/2}S_{\mathbf{W},T}\sqrt{1+\frac{1}{T}} < W_{T+S} < \hat{W}_{T+S|T} + t_{T-1,\alpha/2}S_{\mathbf{W},T}\sqrt{1+\frac{1}{T}} \\
&\Leftrightarrow \bar{W}_T - t_{T-1,\alpha/2}S_{\mathbf{W},T}\sqrt{1+\frac{1}{T}} < W_{T+S} < \bar{W}_T + t_{T-1,\alpha/2}S_{\mathbf{W},T}\sqrt{1+\frac{1}{T}}
\end{aligned}$$

Therefore,

$$\mathbf{P}\left(\bar{W}_T - t_{T-1,\alpha/2}S_{\mathbf{W},T}\sqrt{1+\frac{1}{T}} < W_{T+S} < \bar{W}_T + t_{T-1,\alpha/2}S_{\mathbf{W},T}\sqrt{1+\frac{1}{T}}\right) < 1-\alpha,$$

which shows that (7.28) is a prediction interval for the state W_{T+S}, for any $S \in \mathbb{N}$, at the confidence level of $100\,(1-\alpha)\,\%$, for any $\alpha \in (0,1)$. □

7.4 Weak White Noise

Let $\mathbb{T} \subseteq \mathbb{Z}$, and let $(W_t)_{t\in\mathbb{T}} \equiv \mathbf{W}$ be a stochastic process of order 2 on a probability space $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ with state space $\mathbb{R}^N$, for some $N \in \mathbb{N}$.

Definizione 508 (Weak White Noise) *We say that* $\mathbf{W}$ *is a* weak white noise *if the random variables in* $\mathbf{W}$ *are uncorrelated, have mean zero, and the same variance-covariance matrix. That is the mean [resp variance-covariance] function* $\mu_{\mathbf{W}} : \mathbb{T} \to \mathbb{R}^N$ *[resp.* $\Sigma_{\mathbf{W}} : \mathbb{T} \to \mathbb{R}^{N^2}$*] satisfies*

$$\mu_{\mathbf{W}}(t) = 0 \qquad [\textit{resp. } \Sigma_{\mathbf{W}}(t) = \Sigma_{\mathbf{W}}],$$

for every t*, where* $\Sigma_{\mathbf{W}}$ *is a suitable positive semidefinite symmetric matrix, and*

$$\Gamma_{\mathbf{W}}(s,t) = 0,$$

for all $s, t \in \mathbb{T}$ *such that* $s \neq t$*.*

Osservazione 509 *If* $\mathbf{W}$ *is a strong white noise, then* $\mathbf{W}$ *is a weak white noise.*

Note that when the finiteness of the moment of order 2 of the random variables in $\mathbf{W}$ is not assumed Remark 509 is not longer true.

Esempio 510 *Let $(p_t)_{t\in\mathbb{N}}$ be a sequence of points in $(0,1)$ and let W_t be the Bernoulli random variable with success parameter p_t given by*

$$W_t \overset{def}{=} \begin{cases} \sqrt{\frac{1-p_t}{p_t}}, & \mathbf{P}\left(W_t = \sqrt{\frac{1-p_t}{p_t}}\right) = p_t, \\ -\sqrt{\frac{p_t}{1-p_t}}, & \mathbf{P}\left(W_t = -\sqrt{\frac{p_t}{1-p_t}}\right) = 1-p_t, \end{cases} \qquad \forall t \in \mathbb{N}.$$

We have

$$\mathbf{E}[W_t] = \sqrt{\frac{1-p_t}{p_t}}p_t - \sqrt{\frac{p_t}{1-p_t}}(1-p_t) = 0$$

and

$$\mathbf{D}^2[W_t] = \frac{1-p_t}{p_t}p_t + \frac{p_t}{1-p_t}(1-p_t) = 1.$$

Therefore, the random variables in the stochastic process $(W_t)_{t\in\mathbb{N}} \equiv \mathbf{W}$ have the same expectation and variance; they have different distribution, though. Now, assume that the random variables in $\mathbf{W}$ are independent. Then $\mathbf{W}$ is a weak white noise which is not a strong white noise.

Esempio 511 *Let $(X_t)_{t\in\mathbb{N}} \equiv \mathbf{X}$ be a standard Rademacher white noise, in symbols $\mathbf{X} \sim RWN(1/2)$, and let Z be a standard Gaussian random variable which is independent of $\mathbf{X}$. Consider the process $(W_t)_{t\in\mathbb{N}} \equiv \mathbf{W}$ given by*

$$W_t \overset{def}{=} ZX_t, \quad \forall t \in \mathbb{N}.$$

By virtue of the independence of Z and $\mathbf{X}$, we have

$$\mathbf{E}[W_t] = \mathbf{E}[ZX_t] = \mathbf{E}[Z]\,\mathbf{E}[X_t] = 0 \quad \textit{and} \quad \mathbf{D}^2[W_t] = \mathbf{E}\left[W_t^2\right] = \mathbf{E}\left[Z^2X_t^2\right] = \mathbf{E}\left[Z^2\right]\mathbf{E}\left[X_t^2\right] = 1,$$

for every $t \in \mathbb{N}$. Hence,

$$\mu_{\mathbf{W}}(t) = 0 \quad \textit{and} \quad \sigma^2_{\mathbf{W}}(t) = 1,$$

for every $t \in \mathbb{N}$. In addition,

$$Cov(W_s, W_t) = \mathbf{E}[W_sW_t] = \mathbf{E}[ZX_sZX_t] = \mathbf{E}\left[Z^2X_sX_t\right] = \mathbf{E}\left[Z^2\right]\mathbf{E}[X_s]\,\mathbf{E}[X_t] = 0,$$

for all $s,t \in \mathbb{N}$, such that $s \neq t$. Hence, $\mathbf{W}$ is a weak white noise such that

$$\gamma_{\mathbf{W},1}(\tau) = \begin{cases} 1 \equiv \sigma^2_{\mathbf{W}}, & \textit{if } \tau = 0, \\ 0, & \textit{if } \tau \in \mathbb{N}. \end{cases}$$

On the other hand,

$$\begin{aligned} Cov\left(W_s^2, W_t^2\right) &= \mathbf{E}\left[W_s^2W_t^2\right] - \mathbf{E}\left[W_s^2\right]\mathbf{E}\left[W_t^2\right] = \mathbf{E}\left[Z^4X_s^2X_t^2\right] - 1 \\ &= \mathbf{E}\left[Z^4\right]\mathbf{E}\left[X_s^2\right]\mathbf{E}\left[X_t^2\right] - 1 = \mathbf{E}\left[Z^4\right]\mathbf{E}\left[X_s^2\right]\mathbf{E}\left[X_t^2\right] - 1 \\ &= 3 - 1 = 2, \end{aligned}$$

for all $s,t \in \mathbb{N}$, such that $s \neq t$. Therefore, $\mathbf{W}$ is not a strong white noise. Now, we have

$$\begin{aligned} G_{\mathbf{W},T}(0) = S^2_{\mathbf{W},T} &= \frac{1}{T}\sum_{t=1}^{T}\left(W_t - \bar{W}_T\right)\left(W_t - \bar{W}_T\right) \\ &= \frac{1}{T}\sum_{t=1}^{T}\left(ZX_t - Z\bar{X}_T\right)\left(ZX_t - Z\bar{X}_T\right) \\ &= \frac{Z^2}{T}\sum_{t=1}^{T}\left(X_t - \bar{X}_T\right)\left(X_t - \bar{X}_T\right) \\ &= Z^2S^2_{\mathbf{X},T}. \end{aligned}$$

Since $S^2_{\mathbf{X},T} \xrightarrow{\boldsymbol{P}} \sigma^2_{\mathbf{X}} \equiv 1$*, as* $T \to \infty$*, by virtue of the properties of the convergence in probability, it follows that* $G_{\mathbf{W},T}(0) \xrightarrow{\boldsymbol{P}} Z^2$*, as* $T \to \infty$*. This prevents that* $G_{\mathbf{W},T}(0) \xrightarrow{\boldsymbol{P}} \gamma_{\mathbf{W},1}(0) = 1$*. Thus the weak white noise* $\mathbf{W}$ *is not ergodic in probability and, a fortiori, in mean square, in the wide sense.*

$$\begin{aligned} G^2_{\mathbf{W},T}(\tau) &= \left(\frac{1}{T}\sum_{t=1}^{T-\tau}\left(W_t - \bar{W}_T\right)\left(W_{t+\tau} - \bar{W}_T\right)\right)^2 = \left(\frac{1}{T}\sum_{t=1}^{T-\tau}\left(ZX_t - Z\bar{X}_T\right)\left(ZX_{t+\tau} - Z\bar{X}_T\right)\right)^2 \\ &= \left(\frac{Y^2}{T}\sum_{t=1}^{T-\tau}\left(X_t - \bar{X}_T\right)\left(X_{t+\tau} - \bar{X}_T\right)\right)^2 = Y^4\left(\frac{1}{T}\sum_{t=1}^{T-\tau}\left(X_t - \bar{X}_T\right)\left(X_{t+\tau} - \bar{X}_T\right)\right)^2, \end{aligned}$$

for every $\tau \in \mathbb{N}$*. As a consequence,*

$$\begin{aligned} \mathbf{E}\left[\left(G_{\mathbf{W},T}(\tau) - \gamma_{\mathbf{W},1}(\tau)\right)^2\right] = \mathbf{E}\left[\Gamma^2_{\mathbf{W},T}(\tau)\right] &= \mathbf{E}\left[Y^4\left(\frac{1}{T}\sum_{t=1}^{T-\tau}\left(X_t - \bar{X}_T\right)\left(X_{t+\tau} - \bar{X}_T\right)\right)^2\right] \\ &= \mathbf{E}\left[Y^4\right]\mathbf{E}\left[\left(\frac{1}{T}\sum_{t=1}^{T-\tau}\left(X_t - \bar{X}_T\right)\left(X_{t+\tau} - \bar{X}_T\right)\right)^2\right] \\ &= \mathbf{E}\left[Y^4\right]\mathbf{E}\left[\Gamma^2_{\mathbf{X},T}(\tau)\right] \\ &= \mathbf{E}\left[Y^4\right]\mathbf{E}\left[\left(\Gamma_{\mathbf{X},T}(\tau) - \gamma_{\mathbf{X},1}(\tau)\right)^2\right]. \end{aligned}$$

It follows

$$\lim_{T\to\infty}\mathbf{E}\left[\left(\Gamma_{\mathbf{W},T}(\tau) - \gamma_{\mathbf{W},1}(\tau)\right)^2\right] = \mathbf{E}\left[Y^4\right]\lim_{T\to\infty}\mathbf{E}\left[\left(\Gamma_{\mathbf{X},T}(\tau) - \gamma_{\mathbf{X},1}(\tau)\right)^2\right] = 0.$$

This implies the mean square ergodicity in the wide sense of $\mathbf{W}$*.*

Let $\mathbf{W} \sim WWN^N(\Sigma_{\mathbf{W}})$. For simplicity assume that $\mathbb{T} \equiv \mathbb{N}$.

Osservazione 512 *Remarks* **??**, **??**, *and* **??** *hold true.*

Proposizione 513 (WWN as a WSS process) *The process* $\mathbf{W}$ *is weak-sense stationary.*

Proposizione 514 (WWN as an ergodic process in the mean) *The process* $\mathbf{W}$ *is mean square ergodic in the mean.*
Proof. *Immediate consequence of Definition 508 and Proposition 473* □

Osservazione 515 *Remark* **??** *holds true.*

Osservazione 516 *Remark* **??** *holds true.*

Proposizione 517 *If* W *is a weak white noise, then* W *is mean-square ergodic. Moreover, if* W *is a fourth order process, then* W *is mean square ergodic in the wide sense.*
Proof.

$$Cov\left(W_1 W_{1+s}, W_{1+t} W_{1+s+t}\right) = \mathbf{E}\left[W_1 W_{1+s} W_{1+t} W_{1+s+t}\right] - \mathbf{E}\left[W_1 W_{1+s}\right]\mathbf{E}\left[W_{1+t} W_{1+s+t}\right]$$

Now, for every $s > 0$*, we have*

$$\mathbf{E}\left[W_1 W_{1+s}\right] = \mathbf{E}\left[W_1\right]\mathbf{E}\left[W_{1+s}\right] = 0 \quad and \quad \mathbf{E}\left[W_{1+t} W_{1+s+t}\right] = \mathbf{E}\left[W_{1+t}\right]\mathbf{E}\left[W_{1+s+t}\right] = 0$$

Hence,

$$\lim_{t\to+\infty} Cov\left(W_1W_{1+s}, W_{1+t}W_{1+s+t}\right) = \lim_{t\to+\infty} \mathbf{E}\left[W_1W_{1+s}W_{1+t}W_{1+s+t}\right]$$

$$\lim_{t\to+\infty} \mathbf{E}\left[W_1W_2W_{1+t}W_{2+t}\right] \neq 0$$

Consider the random wariables W_1, W_2, W_3, W_4 *such that*

$$\mathbf{E}\left[W_1\right] = \mathbf{E}\left[W_2\right] = \mathbf{E}\left[W_3\right] = \mathbf{E}\left[W_4\right] = 0,$$

$$Cov\left(W_s, W_t\right) = 0$$

for every $s, t = 1, \ldots, 4$ *such that* $s \neq t$ *and*

$$\mathbf{E}\left[W_1W_2W_3W_4\right] \neq 0$$

Consider a sequence of independent Rademacher random variables $(Z_t)_{t\in\mathbb{N}}$ *which are also independent of* W_1, W_2, W_3, W_4*. Set* $Z_0 \equiv 1$ *and define*

$$\begin{aligned}
&Y_1 \overset{def}{=} W_1, \quad Y_2 \overset{def}{=} W_2, \quad Y_3 \overset{def}{=} W_3, \quad Y_4 \overset{def}{=} W_4, \\
&Y_5 \overset{def}{=} Z_1W_1, \quad Y_6 \overset{def}{=} Z_1W_2, \quad Y_7 \overset{def}{=} Z_1W_3, \quad Y_8 \overset{def}{=} Z_1W_4 \\
&Y_9 \overset{def}{=} Z_2W_1, \quad Y_{10} \overset{def}{=} Z_2W_2, \quad Y_{11} \overset{def}{=} Z_2W_3, \quad Y_{12} \overset{def}{=} Z_2W_4 \\
&Y_{13} \overset{def}{=} Z_3W_1, \quad Y_{14} \overset{def}{=} Z_3W_2, \quad Y_{15} \overset{def}{=} Z_3W_3, \quad Y_{16} \overset{def}{=} Z_3W_4, \\
&\ldots
\end{aligned}$$

we have

$$\mathbf{E}\left[Y_t\right] = \mathbf{E}\left[Z_qW_{r+1}\right] = \mathbf{E}\left[Z_q\right]\mathbf{E}\left[W_{r+1}\right] = 0$$

where $q \in \hat{n}$ *and* $r \in \{0, 1, 2, 3\}$ *are such that*

$$t - 1 = 4q + r$$

$$Y_s = Z_{q_s}W_{r_s+1}, \quad Y_t = Z_{q_t}W_{r_t+1}$$

$$Y_sY_t = Z_{q_s}Z_{q_t}W_{r_s+1}W_{r_t+1}$$

se $s \neq t$ *then* $q_s \neq q_t$ *or* $q_s = q_t$ *and* $r_s \neq r_t$*. It follows*

$$\mathbf{E}\left[Y_sY_t\right] = \mathbf{E}\left[Z_{q_s}Z_{q_t}W_{r_s+1}W_{r_t+1}\right] = \begin{cases} \mathbf{E}\left[Z_{q_s}\right]\mathbf{E}\left[Z_{q_t}\right]\mathbf{E}\left[W_{r_s+1}W_{r_t+1}\right] = 0 & \text{if } q_s \neq q_t \\ \mathbf{E}\left[Z_{q_t}^2\right]\mathbf{E}\left[W_{r_s+1}\right]\mathbf{E}\left[W_{r_s+1}\right] = 0 & \text{if } q_s = q_t \end{cases}$$

The random variables in the process $(Y_t)_{t\in\mathbb{N}}$ *are scorrelated*

$$\begin{aligned}
\mathbf{E}\left[Y_1Y_1Y_2Y_2\right] &= \mathbf{E}\left[W_1W_1W_2W_2\right] = \mathbf{E}\left[W_1^2W_2^2\right] \\
\mathbf{E}\left[Y_1Y_1Y_3Y_3\right] &= \mathbf{E}\left[W_1W_1W_3W_3\right] = \mathbf{E}\left[W_1^2W_3^2\right]
\end{aligned}$$

$$\begin{aligned}
\mathbf{E}\left[Y_1Y_2Y_3Y_4\right] &= \mathbf{E}\left[W_1W_2W_3W_4\right] \\
\mathbf{E}\left[Y_1Y_2Y_7Y_8\right] &= \mathbf{E}\left[W_1W_2Z_1W_3Z_1W_4\right] = \mathbf{E}\left[Z_1^2\right]\mathbf{E}\left[W_1W_2W_3W_4\right] = \mathbf{E}\left[W_1W_2W_3W_4\right] \\
\mathbf{E}\left[Y_1Y_2Y_{11}Y_{12}\right] &= \mathbf{E}\left[W_1W_2Z_2W_3Z_2W_4\right] = \mathbf{E}\left[Z_2^2\right]\mathbf{E}\left[W_1W_2W_3W_4\right] = \mathbf{E}\left[W_1W_2W_3W_4\right]
\end{aligned}$$

□

Esempio 518 *Let* $(X_t)_{t\in\mathbb{N}} \equiv \mathbf{X}$ *a real standard white noise, in symbols* $\mathbf{X} \sim GWN\left(\sigma_X^2\right)$, *for* $\sigma_X^2 \equiv 1$, *and let* Y *be a standard Gaussian random variable which is independent of* $\mathbf{X}$.*Consider the process* $(W_t)_{t\in\mathbb{N}} \equiv \mathbf{W}$ *given by*

$$W_t \overset{def}{=} YX_t, \quad \forall t \in \mathbb{N}.$$

We have

$$\mathbf{E}[W_t] = \mathbf{E}[YX_t] = \mathbf{E}[Y]\,\mathbf{E}[X_t] = 0 \quad \textit{and} \quad \mathbf{D}^2[W_t] = \mathbf{E}\left[W_t^2\right] = \mathbf{E}\left[Y^2X_t^2\right] = \mathbf{E}\left[Y^2\right]\mathbf{E}\left[X_t^2\right] = 1,$$

for every $t \in \mathbb{N}$. *In addition,*

$$Cov\,(W_s, W_t) = \mathbf{E}[W_sW_t] = \mathbf{E}[YX_sYX_t] = \mathbf{E}\left[Y^2X_sX_t\right] = \mathbf{E}\left[Y^2\right]\mathbf{E}[X_sX_t] = \mathbf{E}[X_s]\,\mathbf{E}[X_t] = 0,$$

for all $s, t \in \mathbb{N}$, *such that* $s \neq t$. *Hence,* $\mathbf{W}$ *is a weak white noise. This implies that*

$$\gamma_{\mathbf{W},1}(\tau) = 0,$$

for every $\tau \in \mathbb{N}$. *On the other hand,*

$$\begin{aligned} Cov\left(W_s^2, W_t^2\right) &= \mathbf{E}\left[W_s^2W_t^2\right] - \mathbf{E}\left[W_s^2\right]\mathbf{E}\left[W_t^2\right] = \mathbf{E}\left[Y^4X_s^2X_t^2\right] - 1 \\ &= \mathbf{E}\left[Y^4\right]\mathbf{E}\left[X_s^2X_t^2\right] - 1 = \mathbf{E}\left[Y^4\right]\mathbf{E}\left[X_s^2\right]\mathbf{E}\left[X_t^2\right] - 1 \\ &= 3 - 1 = 2, \end{aligned}$$

for all $s, t \in \mathbb{N}$, *such that* $s \neq t$. *Therefore,* $\mathbf{W}$ *is not a strong white noise. Now, we have*

$$\begin{aligned} \Gamma_{\mathbf{W},T}^2(\tau) &= \left(\frac{1}{T}\sum_{t=1}^{T-\tau}\left(W_t - \bar{W}_T\right)\left(W_{t+\tau} - \bar{W}_T\right)\right)^2 = \left(\frac{1}{T}\sum_{t=1}^{T-\tau}\left(YX_t - Y\bar{X}_T\right)\left(YX_{t+\tau} - Y\bar{X}_T\right)\right)^2 \\ &= \left(\frac{Y^2}{T}\sum_{t=1}^{T-\tau}\left(X_t - \bar{X}_T\right)\left(X_{t+\tau} - \bar{X}_T\right)\right)^2 = Y^4\left(\frac{1}{T}\sum_{t=1}^{T-\tau}\left(X_t - \bar{X}_T\right)\left(X_{t+\tau} - \bar{X}_T\right)\right)^2, \end{aligned}$$

for every $\tau \in \mathbb{N}$. *As a consequence,*

$$\begin{aligned} \mathbf{E}\left[\left(\Gamma_{\mathbf{W},T}(\tau) - \gamma_{\mathbf{W},1}(\tau)\right)^2\right] = \mathbf{E}\left[\Gamma_{\mathbf{W},T}^2(\tau)\right] &= \mathbf{E}\left[Y^4\left(\frac{1}{T}\sum_{t=1}^{T-\tau}\left(X_t - \bar{X}_T\right)\left(X_{t+\tau} - \bar{X}_T\right)\right)^2\right] \\ &= \mathbf{E}\left[Y^4\right]\mathbf{E}\left[\left(\frac{1}{T}\sum_{t=1}^{T-\tau}\left(X_t - \bar{X}_T\right)\left(X_{t+\tau} - \bar{X}_T\right)\right)^2\right] \\ &= \mathbf{E}\left[Y^4\right]\mathbf{E}\left[\Gamma_{\mathbf{X},T}^2(\tau)\right] \\ &= \mathbf{E}\left[Y^4\right]\mathbf{E}\left[\left(\Gamma_{\mathbf{X},T}(\tau) - \gamma_{\mathbf{X},1}(\tau)\right)^2\right]. \end{aligned}$$

It follows

$$\lim_{T\to\infty}\mathbf{E}\left[\left(\Gamma_{\mathbf{W},T}(\tau) - \gamma_{\mathbf{W},1}(\tau)\right)^2\right] = \mathbf{E}\left[Y^4\right]\lim_{T\to\infty}\mathbf{E}\left[\left(\Gamma_{\mathbf{X},T}(\tau) - \gamma_{\mathbf{X},1}(\tau)\right)^2\right] = 0.$$

This implies the mean square ergodicity in the wide sense of $\mathbf{W}$.

Osservazione 519 (Gaussian White Noise) *If* W *is a Gaussian process, then* W *is a Gaussian white noise.*

Esempio 520 *The stochastic processes* $(X_t)_{t\in\mathbb{N}}$ *presented in Examples 390 and 392 are weak white noises which are not strong white noises.*

Esempio 521 *Let $N = 1$ and let $(X_t)_{t\in\mathbb{N}} \equiv X$ be a real stochastic process on Ω of independent and identically distributed random variables having mean $\mu_{X_t} = 0$ and variance $\sigma^2_{X_t} = 1$, for every $t \in \mathbb{N}$. Fixed any $p \in \mathbb{N}$ consider the process $(W_t)_{t\in\mathbb{N}} \equiv W$ given by*

$$W_t \stackrel{def}{=} X_t \cdot \cdots \cdot X_{t+p}, \quad \forall t \in \mathbb{N}.$$

Then the process X is a weak white noises which is strict-sense stationary but, in general, is not a strong white noise.

Discussion. Since the random variables in X are independent, we have

$$\mathbf{E}[W_t] = \mathbf{E}[X_t \cdot \cdots \cdot X_{t+p}] = \mathbf{E}[X_t] \cdot \cdots \cdot \mathbf{E}[X_{t+p}] = 0$$

and

$$\mathbf{D}^2[W_t] = \mathbf{E}\left[W_t^2\right] = \mathbf{E}\left[X_t^2 \cdot \cdots \cdot X_{t+p}^2\right] = \mathbf{E}\left[X_t^2\right] \cdot \cdots \cdot \mathbf{E}\left[X_{t+p}^2\right] = 1,$$

for every $t \in \mathbb{N}$. Moreover,

$$\begin{aligned} Cov(W_t, W_s) &= \mathbf{E}[W_t W_s] = \mathbf{E}[X_s \cdot \cdots \cdot X_{s+p} X_t \cdot \cdots \cdot X_{t+p}] \\ &= \mathbf{E}[X_s]\,\mathbf{E}[X_{s+1} \cdot \cdots \cdot X_{s+p} X_t \cdot \cdots \cdot X_{t-1+p}]\,\mathbf{E}[X_{t+p}] \\ &= 0 \end{aligned}$$

for every $s, t \in \mathbb{N}$, such that $s < t$. This shows that W is a weak white noise. The strongly stationary of W is an immediate consequence of Proposition **??** (see also Example **??**). On the other hand, in general, the random variables in W may be not independent see Example (**??**). This prevents W from being a strong white noise. □##

Definizione 522 *A sequence $(X_t)_{t\geqslant 1} \equiv X$ of uncorrelated N-dimensional random vectors on Ω having null mean and the same finite second order moment is called an N-dimensional or N-variate* weak white noise.

Notation 523 *To denote that X is an N-dimensional weak white noise we will write $X \sim WWN^N\left(\sigma^2\right)$ where $\sigma \in \mathbb{R}_+$ is such that $\sigma^2 = \mathbf{D}^2[X_t] = \mathbf{E}^2[X_t]$ for any $t \in \mathbb{N}$. The reference to the dimension N will be omitted in the case $N = 1$.*

Osservazione 524 *Let $(X_t)_{t\geqslant 1} \equiv X$ be an N-dimensional weak white noise, $X \sim WWN^N\left(\sigma^2\right)$, Assume that X has Gaussian distribution. Then $X \sim GWN^N\left(\sigma^2\right)$.*

Notation 525 *To denote that $(X)_{t\in\mathbb{N}} \equiv X$ is an N-dimensional white noise we will write $X \sim WN^N\left(\sigma^2\right)$ where $\sigma^2 \equiv \mathbf{D}^2[X_t]$ for any $t \in \mathbb{N}$. The reference to the dimension N will be omitted in the case $N = 1$.*

Osservazione 526 *Let $(X)_{t\in\mathbb{N}} \equiv X$ be a sequence of uncorrelated N-dimensional random vectors on Ω having null mean and the same finite variance σ^2. Assume that X is jointly Gaussian distributed. Then X is an N-dimensional* Gaussian white noise, *$X \sim GWN^N\left(\sigma^2\right)$.*

For any $\tau = 1, \ldots, T-1$ and let $\rho_{W.T}(\tau)$ be the time sample autocorrelation of length T at lag τ of W, that is

$$\rho_{W.T}(\tau) = \frac{\sum_{t=1}^{T-\tau}\left(X_t - \bar{X}_T\right)\left(X_{t+\tau} - \bar{X}_T\right)}{\sum_{t=1}^{T}\left(X_t - \bar{X}_T\right)^2}.$$

Teorema 527 *Assume that T is "large" and fix any $S \in \mathbb{N}$ such that $S \ll T$. Then the vector $(\rho_W(1), \ldots, \rho_W(S))^\intercal$ is approximately Gaussian distributed with mean 0 and variance $1/n$. In symbols*

$$(\rho_{W,T}(1), \ldots, \rho_{W,T}(S))^\intercal \sim N(0, I_S/n)$$

where I_S is the identity matrix of order S. As a consequence, the confidence interval for $\mathbf{E}[\rho_{W,T}(\tau)] = 0$ at the confidence level of $100(1-\alpha)\%$, for any $\alpha \in (0,1)$, is given by

$$\left(\rho_{w,T}(\tau) - \frac{z_{\alpha/2}}{\sqrt{T}},\ \rho_{w,T}(\tau) + \frac{z_{\alpha/2}}{\sqrt{T}}\right),$$

for every $\tau = 1, \ldots, S$, where $\rho_{w,T}(\tau)$ is the realization of $\rho_{W,T}(\tau)$ and $z_{\alpha/2}$ is the upper tail critical value of level $\alpha/2$ of the standard Gaussian random variable. In addition, there is evidence against the null hypothesis $H_0 : \mathbf{E}[\rho_W(\tau)] = 0$ at the approximate significance level of $100\alpha\%$, for any $\alpha \in (0,1)$, when

$$|\rho_{w.T}(\tau)| > \frac{z_{\alpha/2}}{\sqrt{T}} \Leftrightarrow \mathbf{P}\left(|Z| \geqslant \sqrt{T}\,|\rho_{w,T}(\tau)|\right) < \alpha/2,$$

In particular, we have approximatively

$$\begin{array}{ll}\left(\rho_{w,T}(\tau) - \frac{1.645}{\sqrt{T}},\ \rho_{w,T}(\tau) + \frac{1.645}{\sqrt{T}}\right) & \text{at } 90\% \text{ c.l.}\\ \left(\rho_{w,T}(\tau) - \frac{1.96}{\sqrt{T}},\ \rho_{w,T}(\tau) + \frac{1.96}{\sqrt{T}}\right) & \text{at } 95\% \text{ c.l.}\\ \left(\rho_{w,T}(\tau) - \frac{2.575}{\sqrt{T}},\ \rho_{w,T}(\tau) + \frac{2.575}{\sqrt{T}}\right) & \text{at } 99\% \text{ c.l.}\end{array}$$

and there is evidence against the null hypothesis $H_0 : \mathbf{E}[\rho_W(\tau)] = 0$ when

$$\begin{array}{ll}|\rho_{w.T}(\tau)| > \frac{1.645}{\sqrt{T}} \Leftrightarrow \mathbf{P}\left(|Z| \geqslant \sqrt{T}\,|\rho_{w,T}(\tau)|\right) < 0.01 & \text{at nearly } 10\% \text{ s.l.}\\ |\rho_{w.T}(\tau)| > \frac{1.96}{\sqrt{T}} \Leftrightarrow \mathbf{P}\left(|Z| \geqslant \sqrt{T}\,|\rho_{w,T}(\tau)|\right) < 0.05 & \text{at nearly } 5\% \text{ s.l.}\\ |\rho_{w.T}(\tau)| > \frac{2.575}{\sqrt{T}} \Leftrightarrow \mathbf{P}\left(|Z| \geqslant \sqrt{T}\,|\rho_{w,T}(\tau)|\right) < 0.01 & \text{at nearly } 1\% \text{ s.l.}\end{array}.$$

Corollary 528 *When $100(1-\alpha)\%$ of the realizations of the time sample autocorrelations of length T at lag τ of W, on varying of the lag $\tau = 1, \ldots, S$, fall outside the interval*

$$\left(\frac{z_{\alpha/2}}{\sqrt{T}},\ \frac{z_{\alpha/2}}{\sqrt{T}}\right),$$

there is evidence against the null hypothesis that the time series is modeled by a strong white noise at the approximate significance level of $100\alpha\%$.

For any $S \in \mathbb{N}$ such that $S < T$ let $Q_{W.T}(S)$ be the Ljung-Box statistic of lenght S on W given by

$$Q_{W.T}(S) = T(T+2)\sum_{\tau=1}^{S} \frac{\rho_{W.T}(\tau)}{T-\tau}.$$

Teorema 529 *Assume that T is "large" and $S \ll T$. Then under the null hypothesis that W is a strong white noise (actually much weaker see Tsay Analysis of Financial Time Series p. 27), the Ljung-Box statistic $Q_{W.T}(S)$ has a Chi-square distribution with S degrees of freedom. In symbols, $Q_{W.T}(S) \sim \chi^2_S$. As a consequence, there is evidence against the null hypothesis $H_0 : W \sim SWN$ at the approximate significance level of $100\alpha\%$, for any $\alpha \in (0,1)$, when*

$$q_{W.T}(S) > \chi^2_{u,\alpha,S} \Leftrightarrow \mathbf{P}\left(\chi^2_S \geqslant q_{W.T}(S)\right) < \alpha/2,$$

where $\chi^2_{u,\alpha,S}$ is the upper α-critical value of the chi-square distribution with S degrees of freedom and χ^2_S is a chi-square random variable with S degrees of freedom.

Note that the Ljung-Box is Portmanteau test: the null hypothesis is well specified but the alternative hypothesis is more loosely specified. Actually, when there is evidence against the null hypothesis of strong white noise the alternative suggests that data are not likely to be generated via an independent sequence of random variable due to the apparent presence of serial correlation.

7.5 Random Walks

In this section we generalize the random walk process, presented in Example 292, in a direction which is at the basis of the formulation of the main statistical tests to investigate the stationarity of a time series.

Let $(X_t)_{t\in\hat{n}} \equiv X$ be a stochastic process on a probability space Ω with states in $\mathbb{R}^N$, for some $N \in \mathbb{N}$, let $\alpha, \beta \in \mathbb{R}^N$, and let $(W_t)_{t\in\mathbb{N}} \equiv W$ be a weak white noise on Ω with states in $\mathbb{R}^N$ and variance-covariance matrix Σ_W, in symbols $W \sim WWN^N(\Sigma_W)$. We assume that the random variable X_0 has finite second moment and is independent of the random variables in W. We set $\mu_{X_0} \equiv \mathbf{E}[X_0]$ and $\Sigma_{X_0} \equiv Var(X_0)$.

Definizione 530 *We call X an N-*variate random walk *with* drift *and* linear trend *if the random variables in X are given by*

$$X_t \stackrel{def}{=} \alpha + \beta t + X_{t-1} + W_t, \quad \forall t \in \mathbb{N}, \tag{7.30}$$

The random vector [resp. the distribution of the random vector] X_0 is referred to as the initial state *[resp.* initial distribution*] of the random walk X, in case $X_0 \equiv x_0 \in \mathbb{R}^N$, we also call x_0 the* starting point *of the random walk; the vector α, [resp. β] is referred to as the* drift *[resp.* trend*]* coefficient *or* parameter *of X*[2]*; the weak white noise W is referred to as the* state innovation*. In case $N = 1$, we usually speak of real random walk with no reference to N. Also the explicit reference to the innovation process W is usually omitted when not necessary.*

Osservazione 531 *In some circumstances, it is more appropriate to consider a stocastic process $(X_t)_{t\in\mathbb{T}} \equiv X$ and a weak white noise $(W_t)_{t\in\mathbb{T}} \equiv W$ with time set $\mathbb{T} \equiv \mathbb{Z}$. In this case we say that X is an N-*variate random walk *with innovation W if the random variables in X are given by Equation (7.30) on varying of $t \in \mathbb{Z}$. Clearly, in this case there is no initial state of the process.????*

Notation 532 *To denote that X is a N-variate random walk we write $X \sim RW^N$. In case $N = 1$, we simply write $X \sim RW$. We also write σ_W^2 [resp. $\sigma_{X_0}^2$] instead of Σ_W [resp. Σ_{X_0}].*

Let $(X_t)_{t\in\hat{n}} \equiv X$ be an N-variate random walk with innovation $(W_t)_{t\in\mathbb{N}} \equiv W$ satisfying Equation (7.30), for some $\alpha, \beta \in \mathbb{R}^N$.

Proposizione 533 *We have*

$$X_t = X_0 + \alpha t + \frac{1}{2}\beta t(t+1) + \sum_{s=1}^{t} W_s, \tag{7.31}$$

for every $t \in \mathbb{N}$. More generally,

$$X_t = X_s + \alpha(t-s) + \frac{1}{2}\beta(t-s)(t+s+1) + \sum_{r=s+1}^{t} W_r, \tag{7.32}$$

for all $s, t \in \hat{n}$ such that $0 \leqslant s < t$.

Proof. According to (7.30), Equation (7.31) is clearly true for $t = 1$. Assume inductively that (7.31) is true for some $t > 1$ and consider the case $t + 1$. We can then write

$$\begin{aligned} X_{t+1} &= \alpha + \beta(t+1) + X_t + W_{t+1} \\ &= \alpha + \beta(t+1) + X_0 + \alpha t + \frac{1}{2}\beta t(t+1) + \sum_{s=1}^{t} W_s + W_{t+1} \\ &= X_0 + \alpha(t+1) + \frac{1}{2}\beta(t+1)(t+2) + \sum_{s=1}^{t+1} W_s, \end{aligned}$$

[2] When we want to stress that $\alpha \neq 0$ and $\beta \neq 0$, [resp. $\alpha \neq 0$ and $\beta = 0$, resp. $\alpha = 0$ and $\beta \neq 0$] we call X a *random walk with drift and trend* [resp. with *drift*, resp. with *trend*].

which is the desired Equation (7.31) in the case $t+1$. By virtue of the Induction Principle, we can conclude that (7.31) is true for every $t \geqslant 1$. Now, considering any s such that $0 \leqslant s \leqslant t$, we can write

$$\begin{aligned}
X_t &= X_0 + \alpha t + \frac{1}{2}\beta t\,(t+1) + \sum_{s=1}^{t} W_s \\
&= X_0 + \alpha s + \alpha\,(t-s) + \frac{1}{2}\beta s\,(s+1) + \frac{1}{2}\beta\,(t-s)\,(t+s+1) + \sum_{r=1}^{s} W_r + \sum_{r=s+1}^{t} W_r \\
&= X_s + \alpha\,(t-s) + \frac{1}{2}\beta\,(t-s)\,(t+s+1) + \sum_{r=s+1}^{t} W_r.
\end{aligned}$$

This proves (7.32). $\square$

Let $\left(\mathcal{F}_t^{X_0,W}\right)_{t\in\hat{n}} \equiv \mathfrak{F}^{X_0,W}$ the filtration generated by the initial state X_0 of the random walk X and the innovation process W, that is

$$\mathcal{F}_0^{X_0,W} \overset{\text{def}}{=} \sigma\,(X_0) \quad \text{and} \quad \mathcal{F}_t^{X_0,W} \overset{\text{def}}{=} \sigma\,(X_0, W_1, \ldots, W_t)\,, \quad \forall t \in \mathbb{N},$$

where $\sigma\,(X, Y, Z, \ldots)$ denotes the σ-algebra generated by the random variables $X, Y, Z, \ldots$

Corollary 534 *The random walk X is adapted to $\mathfrak{F}^{X_0,W}$.*

Proof. The claim is an immediate consequence of Equation (7.31). $\square$

Note that when $X_0 \equiv x_0 \in \mathbb{R}^N$ or we consider a random walk X with time set $\mathbb{T} = \mathbb{Z}$, the process X turns out to adapted to the filtration $\left(\mathcal{F}_t^W\right)_{t\in\mathbb{T}} \equiv \mathfrak{F}^W$ generated by the innovation process W, that is

$$\mathcal{F}_t^W \equiv \bigvee_{s\in\mathbb{T}, s\leqslant t} \sigma\,(W_s)\,.$$

Corollary 535 *If the innovation W is a Kth-order process, then the random walk X is also a Kth-order process for every $K \geqslant 2$.*

Proof. Recalling that the random variables with finite Kth moment constitute a Banach space, the claim is an immediate consequence of Equation (7.31). $\square$

Corollary 536 *The random variables $X_1, \ldots, X_t$ in the random walk X are uncorrelated with the random variables $W_{t+1}, W_{t+2}, \ldots$ of the innovation process W for every $t \in \mathbb{N}$.*

Corollary 537 *Under the assumption that the innovation W is a strong white noise, the random variables $X_1, \ldots, X_t$ in the random walk X are independent of the random variables $W_{t+1}, W_{t+2}, \ldots$ of the innovation process W for every $t \in \mathbb{N}$.*

Corollary 538 *Under the assumption that the innovation W is a strong white noise, the random walk X is a Markov process, that is*

$$\mathbf{P}\left(X_t \in B \mid \mathcal{F}_s^{X_0,W}\right) = \mathbf{P}\left(X_t \in B \mid \sigma\,(X_s)\right), \tag{7.33}$$

for every $B \in \mathcal{B}\left(\mathbb{R}^N\right)$ and all $s, t \in \hat{n}$ such that $0 \leqslant s < t$.

Proof. Considering Equation (7.32), we can write

$$\mathbf{P}\left(X_t \in B \mid \mathcal{F}_s^{X_0,W}\right) = \mathbf{P}\left(X_s + \alpha\,(t-s) + \frac{1}{2}\beta\,(t-s)\,(t+s+1) + \sum_{r=s+1}^{t} W_r \in B \mid \mathcal{F}_s^{X_0,W}\right), \tag{7.34}$$

for every $B \in \mathcal{B}\left(\mathbb{R}^N\right)$, for every $t \geqslant s$. Now, the random variable $X_s + \alpha(t-s) + \frac{1}{2}\beta(t-s)(t+s+1)$ is clearly $\sigma(X_s)$-measurable, where $\sigma(X_s) \subseteq \mathcal{F}_s^{X_0,W}$, and, under the assumption that W is a strong white noise, the random variable $\sum_{r=s+1}^{t} W_r$ is independent of $\mathcal{F}_s^{X_0,W}$. Therefore (see Karatzas-Shevre (**?**, Problem 5.9 p.74) (1996)),

$$\mathbf{P}\left(X_s + \alpha(t-s) + \frac{1}{2}\beta(t-s)(t+s+1) + \sum_{r=s+1}^{t} W_r \in B \mid \mathcal{F}_s^{X_0,W}\right)$$
$$= \mathbf{P}\left(X_s + \alpha(t-s) + \frac{1}{2}\beta(t-s)(t+s+1) + \sum_{r=s+1}^{t} W_r \in B \mid \sigma(X_s)\right), \tag{7.35}$$

for every $B \in \mathcal{B}\left(\mathbb{R}^N\right)$. Combining Equations (7.34) and (7.35), we show the Markov property. □

Corollary 539 *Under the assumption that the innovation W is a strong white noise and $\alpha = \beta = 0$, the random walk X is a martingale, that is*

$$\mathbf{E}\left[X_t \mid \mathcal{F}_s^{X_0,W}\right] = X_s, \tag{7.36}$$

for all $s, t \in \hat{n}$ such that $0 \leqslant s < t$.

Proof. Considering Equation (7.32), if $\alpha = \beta = 0$, we can write

$$\mathbf{E}\left[X_t \mid \mathcal{F}_s^{X_0,W}\right] = \mathbf{E}\left[X_s + \sum_{r=s+1}^{t} W_r \mid \mathcal{F}_s^{X_0,W}\right] \tag{7.37}$$

On the other hand, since $\sigma(X_s) \subseteq \mathcal{F}_s^{X_0,W}$ and, under the assumption that W is a strong white noise, the random variable $\sum_{r=s+1}^{t} W_r$ is independent of $\mathcal{F}_s^{X_0,W}$, thanks to the properties of the conditional expectation operator, we have

$$\begin{aligned}\mathbf{E}\left[X_s + \sum_{r=s+1}^{t} W_r \mid \mathcal{F}_s^{X_0,W}\right] &= \mathbf{E}\left[X_s \mid \mathcal{F}_s^{X_0,W}\right] + \sum_{r=s+1}^{t} \mathbf{E}\left[W_r \mid \mathcal{F}_s^{X_0,W}\right] \\ &= X_s + \sum_{r=s+1}^{t} \mathbf{E}\left[W_r\right] \\ &= X_s. \end{aligned} \tag{7.38}$$

Combining Equations (**??**) and (7.38), we obtain the desired (7.36). □

Proposizione 540 *The mean function $\mu_X : \hat{n} \to \mathbb{R}^N$ of X is given by*

$$\mu_X(t) = \mu_{X_0} + \alpha t + \frac{1}{2}\beta t(t+1), \tag{7.39}$$

for every $t \in \hat{n}$.

Proof. Considering Equation (7.31), since

$$\mathbf{E}\left[W_s\right] = 0,$$

for all $s = 1, \ldots, t$, thanks to the properties of the expectation operator, we have

$$\begin{aligned}\mu_X(t) = \mathbf{E}\left[X_t\right] &= \mathbf{E}\left[X_0 + \alpha t + \frac{1}{2}\beta t(t+1) + \sum_{s=1}^{t} W_s\right] \\ &= \mathbf{E}\left[X_0\right] + \alpha t + \frac{1}{2}\beta t(t+1) + \sum_{s=1}^{t} \mathbf{E}\left[W_s\right] \\ &= \mu_{X_0} + \alpha t + \frac{1}{2}\beta t(t+1),\end{aligned}$$

for every $t \in \hat{n}$. □

Proposizione 541 *The variance-covariance function* $\Sigma_X : \hat{n} \to \mathbb{R}^{N^2}$ *of* X *is given by*

$$\Sigma_X(t) = \Sigma_{X_0} + t\Sigma_W, \tag{7.40}$$

for every $t \in \hat{n}$.

Proof. The result being evident if $t = 0$, consider $t \in \mathbb{N}$. Combining Equations (7.31) and (7.39), we obtain

$$X_t - \mathbf{E}[X_t] = X_0 - \mathbf{E}[X_0] + \sum_{s=1}^{t} W_s.$$

Therefore, recalling that X_0 is independent of the random variables in W, we can write

$$\begin{aligned}
\Sigma_X(t) &= \mathbf{E}\left[(X_t - \mathbf{E}[X_t])(X_t - \mathbf{E}[X_t])^\intercal\right] \\
&= \mathbf{E}\left[(X_0 - \mathbf{E}[X_0])(X_0 - \mathbf{E}[X_0])^\intercal\right] + \mathbf{E}\left[(X_0 - \mathbf{E}[X_0])\left(\sum_{s=1}^{t} W_s\right)^\intercal\right] \\
&+ \mathbf{E}\left[\left(\sum_{s=1}^{t} W_s\right)(X_0 - \mathbf{E}[X_0])^\intercal\right] + \mathbf{E}\left[\left(\sum_{s=1}^{t} W_s\right)\left(\sum_{s=1}^{t} W_s\right)^\intercal\right] \\
&= Var(X_0) + \mathbf{E}[X_0 - \mathbf{E}[X_0]]\,\mathbf{E}\left[\sum_{s=1}^{t} W_s^\intercal\right] \\
&+ \mathbf{E}\left[\sum_{s=1}^{t} W_s^\intercal\right]\mathbf{E}[X_0 - \mathbf{E}[X_0]]^\intercal + \mathbf{E}\left[\left(\sum_{s=1}^{t} W_s\right)\left(\sum_{s=1}^{t} W_s^\intercal\right)\right] \\
&= \Sigma_{X_0} + \mathbf{E}\left[\left(\sum_{s=1}^{t} W_s\right)\left(\sum_{s=1}^{t} W_s^\intercal\right)\right],
\end{aligned} \tag{7.41}$$

where,

$$\begin{aligned}
\mathbf{E}\left[\left(\sum_{s=1}^{t} W_r\right)\left(\sum_{s=1}^{t} W_r^\intercal\right)\right] &= \mathbf{E}\left[\left(\sum_{r,s=1}^{t} W_r W_s^\intercal\right)\right] = \sum_{r,s=1}^{t} \mathbf{E}[W_r W_s^\intercal] \\
&= \sum_{\substack{r,s=1 \\ r=s}}^{t} \mathbf{E}[(W_r W_s^\intercal)] + \sum_{\substack{r,s=1 \\ r\neq s}}^{t} \mathbf{E}[(W_r W_s^\intercal)].
\end{aligned} \tag{7.42}$$

On the other hand,

$$\sum_{\substack{r,s=1 \\ r=s}}^{t} \mathbf{E}[(W_r W_s^\intercal)] = \sum_{r=1}^{t} \mathbf{E}[(W_r W_r^\intercal)] = \sum_{r=1}^{t} Var(W_r) = \sum_{r=1}^{t} \Sigma_W = t\Sigma_W \tag{7.43}$$

and

$$\mathbf{E}[(W_r W_s^\intercal)] = 0, \tag{7.44}$$

for all $r, s = 1, \ldots, t$ such that $r \neq s$. In the end, combining (7.41)-(7.44), the desired Equation (7.40) follows. □

Proposizione 542 *The autocovariance function* $\gamma_X : \hat{n} \times \hat{n} \to \mathbb{R}^{N^2}$ *of* X *is given by*

$$\gamma_X(s,t) = \Sigma_{X_0} + s\Sigma_W, \tag{7.45}$$

for all $s, t \in \hat{n}$ *such that* $s \leqslant t$.

Proof. Considering Equation (7.32), we can write

$$\begin{aligned}\gamma_X(s,t) = Cov(X_s, X_t) &= Cov\left(X_s, X_s + \alpha(t-s) + \frac{1}{2}\beta(t-s)(t+s+1) + \sum_{r=s+1}^{t} W_r\right) \\ &= Var(X_s) + Cov\left(X_s, \sum_{r=s+1}^{t} W_r\right). \end{aligned} \tag{7.46}$$

for all $s, t \in \hat{n}$ such that $s \leqslant t$. Now, applying Equation (7.40), we have

$$Var(X_s) = \Sigma_X(t) = \Sigma_{X_0} + s\Sigma_W. \tag{7.47}$$

Furthermore,

$$Cov\left(X_s, \sum_{r=s+1}^{t} W_t\right) = 0. \tag{7.48}$$

In fact, W is a weak white noise and

$$X_s = X_0 + \alpha s + \frac{1}{2}\beta s(s+1) + \sum_{r=1}^{s} W_r.$$

Hence, the random variables X_s and $\sum_{r=s+1}^{t} W_r$ are uncorrelated. Combining (7.46)-(7.48) we then obtain Equation (7.45). $\square$

Proposizione 543 *The autocorrelation function $\rho_X : \hat{n} \times \hat{n} \to \mathbb{R}^{N^2}$ of X is given by*

$$\rho_X(s,t) = \begin{cases} \operatorname{diag}(\Sigma_{X_0})^{-1/2} \Sigma_{X_0} \operatorname{diag}(\Sigma_{X_0})^{-1/2} & \text{if } 0 = s = t \\ \sqrt{\frac{s}{t}} \operatorname{diag}\left(\frac{1}{s}\Sigma_{X_0} + \Sigma_W\right)^{-1/2} \left(\frac{1}{s}\Sigma_{X_0} + \Sigma_W\right) \operatorname{diag}\left(\frac{1}{t}\Sigma_{X_0} + \Sigma_W\right)^{-1/2} & \text{if } 0 < s \leqslant t \end{cases}. \tag{7.49}$$

In particular, if $X_0 \equiv x_0 \in \mathbb{R}^N$ we have

$$\rho_X(s,t) = \begin{cases} 0 & \text{if } 0 = s = t \\ \sqrt{\frac{s}{t}} \operatorname{diag}(\Sigma_W)^{-1/2} (\Sigma_W) \operatorname{diag}(\Sigma_W)^{-1/2} & \text{if } 0 < s \leqslant t \end{cases}. \tag{7.50}$$

Proof. The result being evident if $s = t = 0$, consider any $s.t \in \mathbb{N}$ such that $s \leqslant t$. On account of (**??**) and (7.45), we can write

$$\begin{aligned}\rho_X(s,t) &= \operatorname{diag} \Sigma_X(s)^{-\frac{1}{2}} \gamma\alpha_X(s,t) \operatorname{diag} \Sigma_X(t)^{-\frac{1}{2}} \\ &= \operatorname{diag}(\Sigma_{X_0} + s\Sigma_W)^{-\frac{1}{2}} (\Sigma_{X_0} + s\Sigma_W) \operatorname{diag}(\Sigma_{X_0} + t\Sigma_W)^{-\frac{1}{2}} \\ &= s^{-\frac{1}{2}} s t^{-\frac{1}{2}} \operatorname{diag}\left(\frac{1}{s}\Sigma_{X_0} + \Sigma_W\right)^{-\frac{1}{2}} \left(\frac{1}{s}\Sigma_{X_0} + \Sigma_W\right) \operatorname{diag}\left(\frac{1}{t}\Sigma_{X_0} + \Sigma_W\right)^{-\frac{1}{2}} \\ &= \sqrt{\frac{s}{t}} \operatorname{diag}\left(\frac{1}{s}\Sigma_{X_0} + \Sigma_W\right)^{-\frac{1}{2}} \left(\frac{1}{s}\Sigma_{X_0} + \Sigma_W\right) \operatorname{diag}\left(\frac{1}{t}\Sigma_{X_0} + \Sigma_W\right)^{-\frac{1}{2}},\end{aligned}$$

for all $s, t \in \mathbb{N}$ such that $s \leqslant t$. $\square$

Osservazione 544 *In the case $N = 1$, we have*

$$\rho_X(s,t) = \begin{cases} 1 & \text{if } 0 = s = t \\ \sqrt{\frac{s}{t} \frac{\left(\frac{1}{s}\sigma_{X_0}^2 + \sigma_W^2\right)}{\left(\frac{1}{t}\sigma_{X_0}^2 + \sigma_W^2\right)}} & \text{if } 0 < s \leqslant t \end{cases} \tag{7.51}$$

In particular, if $X_0 \equiv x_0 \in \mathbb{R}$ we have

$$\rho_X(s,t) = \begin{cases} 0 & \text{if } 0 = s = t \\ \sqrt{\frac{s}{t}} & \text{if } 0 < s \leqslant t \end{cases}. \tag{7.52}$$

Let $(X_t)_{t\in\hat{n}} \equiv X$ be an N-variate random walk with innovation $(W_t)_{t\in\mathbb{N}} \equiv W$ satisfying Equation (7.30), for some $\alpha, \beta \in \mathbb{R}^N$.

Proposizione 545 *Assume that X_0 is Gaussian, possibly degenerate. In addition, assume that the innovation process W is Gaussian, in symbols $W \sim GWN(\Sigma_W^2)$. Then we have*

$$X_t \sim N\left(\mu_X(t), \Sigma_X^2(t)\right), \tag{7.53}$$

for every $t \in \mathbb{N}$, where $\mu_X(t)$ and $\Sigma_X^2(t)$ are given by (7.39) and (7.40), respectively. In addition, the process X is Gaussian. As a consequence, X has also Gaussian increments.

Proof. Recall that the assumption $W \sim GWN(\Sigma_W^2)$ implies that the random variables in W are independent. Hence, thanks to the independence and Gaussianity of the random variables on the right hand side of Equation (7.31), Equation (7.53) immediately follows. Moreover, in case X_0 is degenerate, referring Equation (7.31) to the standardized random variables in W and considering Proposition 411, the Gaussianity of the random walk X immediately follows. In case X_0 is not degenerate, but is independent of the random variables in W, referring Equation (7.31) to the standardization of X_0 and the random variables in W and still considering Proposition 411, we obtain again the Gaussianity of the $AR(1)$ process X. In the end, the Gaussianity of the increments of X comes from Proposition 417. □

Definizione 546 *In light of Proposition 545, we call* Gaussian *a random walk with Gaussian innovation W.*

Proposizione 547 *The random walk is not weakly stationary.*

Proof. Immediate consequence of Equations (7.39)-(7.45). □

Despite, not weakly stationary, a simple transformation can make a random walk an almost stationary process.

Proposizione 548 *The process $(\Delta X_t)_{t\in\mathbb{N}} \equiv \Delta X$ given by*

$$\Delta X_t \overset{def}{=} X_t - X_{t-1}, \quad \forall t \in \mathbb{N} \tag{7.54}$$

is a white noise with a linear trend. In particular, if X has no trend, ΔX is weakly stationary and ergodic.

Proof. Replacing s with $t-1$ into Equation (7.32), we can write

$$\begin{aligned} X_t &= X_{t-1} + \alpha\left(t-(t-1)\right) + \frac{1}{2}\beta\left(t-(t-1)\right)\left(t+(t-1)+1\right) + \sum_{r=(t-1)+1}^{t} W_r \\ &= X_{t-1} + \alpha + \beta t + W_t, \end{aligned}$$

for every $t \in \mathbb{N}$. As a consequence,

$$\Delta X_t = \alpha + \beta t + W_t,$$

for every $t \in \mathbb{N}$. the claim follows from Definition **??**. □

7.6 Autoregressive Processes - AR Processes

7.6.1 First Order Autoregressive Process - AR(1) Process

Let $(X_t)_{t\in\hat{n}} \equiv X$ be a stochastic process on a probability space Ω with states in $\mathbb{R}^N$, for some $N \in \mathbb{N}$, and let $(W_t)_{t\in\mathbb{N}} \equiv W$ be a weak white noise on Ω with states in $\mathbb{R}^N$ and variance-covariance matrix Σ_W, in symbols $W \sim WWN(\Sigma_W)$. We assume that the random variable X_0 has finite 2nd moment and is independent of the random variables in W. We set $\mu_{X_0} \equiv \mathbf{E}[X_0]$ and $\sigma^2_{X_0} \equiv \mathbf{D}^2[X_0]$.

Definizione 549 *We say that X is a* 1st-order N-variate autoregressive process *with* innovation W, *if the random variables in X are given by*

$$X_t \overset{def}{=} \alpha + \beta t + \phi X_{t-1} + W_t, \quad \forall t \in \mathbb{N}, \tag{7.55}$$

for some vectors $\alpha, \beta \in \mathbb{R}^N$ and a matrix $\phi \in \mathbb{R}^{N^2}$. The random vector [resp. the distribution of the random vector] X_0 is referred to as the initial state *[resp.* initial distribution*] of the process X; in case $X_0 \equiv x_0 \in \mathbb{R}^N$, we simply call x_0 the* starting point *of X; the vector α [resp. β] is referred to as the* drift, *[resp.* trend*]* coefficient *or* parameter *of X; the matrix ϕ is referred to as the* regression coefficient *or* parameter *of X. In case $N = 1$, we usually speak of* 1st-order real autoregressive process *with no reference to N. Also the reference to the innovation process W is usually omitted when not necessary.*

Osservazione 550 *In some circumstances it is more appropriate to consider a stochastic process $(X_t)_{t\in\mathbb{T}} \equiv X$ and a weak white noise $(W_t)_{t\in\mathbb{T}} \equiv W$ with time set $\mathbb{T} \equiv \mathbb{Z}$. In this case, we say that X is a* 1st-order N-variate autoregressive process *with* innovation W *if the random variables in X fulfill Equation (7.55), for every $t \in \mathbb{Z}$. Clearly, in this case, there is no mention to an initial state of the process.*

Notation 551 *To denote that X is a 1st-order N-variate autoregressive process we write $X \sim AR(1)^N$. In case $N = 1$, we simply write $X \sim AR(1)$.*

Osservazione 552 *When $\phi \equiv I_N$, where I_N is the identity matrix in $\mathbb{R}^{N^2}$, Equation (7.55) becomes the random walk equation (7.30).*

In what follows, we restrain our attention to $AR(1)$ processes X, satisfying Equation (7.55), for which the innovation W is a real weak white noise with variance σ^2_W, in symbols $W \sim WWN\left(\sigma^2_W\right)$, for some $\sigma_W > 0$, and the coefficients α, β, and ϕ are all real numbers. Recall that in this case both the autocovariance and autocorrelation functions of X are symmetric, that is

$$\gamma_X(s,t) = \gamma_X(t,s) \quad \text{and} \quad \rho_X(s,t) = \rho_X(t,s)$$

for all $s, t \in \mathbb{N}$. Furthermore, we will consider the technical assumption $\phi \neq 1$ to distinguish a 1st-order autoregressive process from a random walk.

The following result states that an $AR(1)$ processes is essentially characterized by an $AR(1)$ noise with the same regression coefficient. This is rather useful to achieve correct OLS parameter estimations.

Proposizione 553 *Assume that $(X_t)_{t\in\hat{n}} \equiv X$ is an $AR(1)$ process with innovation $(W_t)_{t\in\mathbb{N}} \equiv W$ satisfying Equation (7.55), for some $\alpha, \beta, \phi \in \mathbb{R}$. Then there exist $\tilde{\alpha}, \tilde{\beta} \in \mathbb{R}$ such that we can write*

$$X_t = \tilde{\alpha} + \tilde{\beta} t + Y_t, \tag{7.56}$$

for every $t \in \mathbb{N}$, where $(Y_t)_{t\in\mathbb{N}} \equiv Y$ is an $AR(1)$ process with innovation W solution of

$$Y_t = \phi Y_{t-1} + W_t, \tag{7.57}$$

for every $t \in \mathbb{N}$. Conversely, assume that X is a process satisfying Equation (7.57), for some $\tilde{\alpha}, \tilde{\beta} \in \mathbb{R}$, where Y is an $AR(1)$ process with innovation W solution of (7.56), for some $\phi \in \mathbb{R}$. Then X is an $AR(1)$ process with innovation W satisfying Equation (7.55), for suitable $\alpha, \beta \in \mathbb{R}$.

Proof. Set

$$Y_t \overset{\text{def}}{=} -\frac{\alpha(1-\phi)-\beta\phi}{(1-\phi)^2} - \frac{\beta}{1-\phi}t + X_t, \quad \forall t \in \mathbb{N}. \tag{7.58}$$

Since X_t satisfies Equation (7.56), from (7.58), we obtain

$$\begin{aligned}
Y_t &= -\frac{\alpha(1-\phi)-\beta\phi}{(1-\phi)^2} - \frac{\beta}{1-\phi}t + \alpha + \beta t + \phi X_{t-1} + W_t \\
&= \frac{\alpha(1-\phi)^2-\alpha(1-\phi)+\beta\phi}{(1-\phi)^2} + \frac{\beta(1-\phi)-\beta}{1-\phi}t + \phi X_{t-1} + W_t \\
&= -\phi\frac{\alpha(1-\phi)-\beta}{(1-\phi)^2} - \phi\frac{\beta}{1-\phi}t + \phi X_{t-1} + W_t \\
&= -\phi\frac{\alpha(1-\phi)-\beta}{(1-\phi)^2} - \phi\frac{\beta}{1-\phi} - \phi\frac{\beta}{1-\phi}(t-1) + \phi X_{t-1} + W_t \\
&= -\phi\frac{\alpha(1-\phi)-\beta\phi}{(1-\phi)^2} - \phi\frac{\beta}{1-\phi}(t-1) + \phi X_{t-1} + W_t \\
&= \phi Y_{t-1} + W_t
\end{aligned}$$

for every $t \in \mathbb{N}$. Therefore, Y_t solves Equation (7.57). Moreover, writing

$$\tilde{\alpha} \equiv \frac{\alpha(1-\phi)-\beta\phi}{(1-\phi)^2} \qquad \text{and} \qquad \tilde{\beta} \equiv \frac{\beta}{1-\phi} \tag{7.59}$$

also Equation (7.56) is trivially satisfied. Conversely, assume that X is a process satisfying Equation (7.56), for some $\tilde{\alpha}, \tilde{\beta} \in \mathbb{R}$, then we have

$$X_{t-1} = \tilde{\alpha} + \tilde{\beta}(t-1) + Y_{t-1}, \tag{7.60}$$

for every $n \in \mathbb{N}$. Hence, combining (7.56) with (7.57) and (7.60), it follows

$$X_t - \tilde{\alpha} - \tilde{\beta}t = \phi X_{t-1} - \phi\tilde{\alpha} - \phi\tilde{\beta}(t-1) + W_t,$$

tha is

$$X_t = \tilde{\alpha}(1-\phi) + \tilde{\beta}\phi + \tilde{\beta}(1-\phi)t + \phi X_{t-1} + W_t.$$

for every $n \in \mathbb{N}$. In the end, X satisfies Equation (7.55) for

$$\alpha = \tilde{\alpha}(1-\phi) + \tilde{\beta}\phi \qquad \text{and} \qquad \beta = \tilde{\beta}(1-\phi). \tag{7.61}$$

In the end, note that Equations (7.59) and (7.61) are equivalent. □

Let $(X_t)_{t\in\hat{n}} \equiv X$ be an $AR(1)$ process with innovation $(W_t)_{t\in\mathbb{N}} \equiv W$ satisfying Equation (7.55) for some $\alpha, \beta, \phi \in \mathbb{R}$.

Proposizione 554 *We have*

$$X_t = \phi^t X_0 + \ell(t;\alpha,\beta) + \textstyle\sum_{s=1}^{t}\phi^{t-s}W_s, \tag{7.62}$$

for every $t \in \mathbb{N}$, where

$$\ell(t;\alpha,\beta) \equiv \alpha\frac{1-\phi^t}{1-\phi} + \beta\frac{t-(t+1)\phi+\phi^{t+1}}{(1-\phi)^2}. \tag{7.63}$$

More generally,

$$X_t = \phi^{t-s}X_s + \ell(s,t;\alpha,\beta) + \textstyle\sum_{r=s+1}^{t}\phi^{t-r}W_r, \tag{7.64}$$

for all $s, t \in \mathbb{N}$ such that $0 \leqslant s < t$, where

$$\ell(s,t;\alpha,\beta) \equiv \alpha\frac{1-\phi^{t-s}}{1-\phi} + \beta\frac{t-(t+1)\phi-s\phi^{t-s}+(s+1)\phi^{t-s+1}}{(1-\phi)^2}. \tag{7.65}$$

Proof. According to (7.55), Equation (7.62) is clearly true for $t = 1$. Assume inductively that (7.62) holds true for some $t > 1$ and consider the case $t + 1$. We can then write

$$\begin{aligned}
X_{t+1} &= \alpha + \beta(t+1) + \phi X_t + W_{t+1} \\
&= \alpha + \beta(t+1) + \phi\left(\phi^t X_0 + \alpha\frac{1-\phi^t}{1-\phi} + \beta\frac{t-(t+1)\phi+\phi^{t+1}}{(1-\phi)^2} + \textstyle\sum_{s=1}^{t}\phi^{t-s}W_s\right) + W_{t+1} \\
&= \phi^{t+1}X_0 + \alpha\left(\frac{1-\phi^t}{1-\phi}\phi + 1\right) + \beta\left(\frac{t-(t+1)\phi+\phi^{t+1}}{(1-\phi)^2}\phi + t + 1\right) + \textstyle\sum_{s=1}^{t}\phi^{t+1-s}W_s + W_{t+1} \\
&= \phi^{t+1}X_0 + \alpha\frac{1-\phi^{t+1}}{1-\phi} + \beta\frac{t+1-(t+2)\phi+\phi^{t+2}}{(1-\phi)^2} + \textstyle\sum_{s=1}^{t+1}\phi^{t+1-s}W_s,
\end{aligned}$$

which is the desired Equation (7.62) in the case $t + 1$. By virtue of the Induction Principle, we can conclude that (7.62) holds true for every $t \geqslant 1$. Now, considering any s such that $0 \leqslant s \leqslant t$, we can write

$$\begin{aligned}
X_t &= \phi^t X_0 + \alpha\frac{1-\phi^t}{1-\phi} + \beta\frac{t-(t+1)\phi+\phi^{t+1}}{(1-\phi)^2} + \textstyle\sum_{s=1}^{t}\phi^{t-s}W_s \\
&= \phi^{t-s}\phi^s X_0 + \alpha\frac{1-\phi^s}{1-\phi}\phi^{t-s} + \frac{1-\phi^{t-s}}{1-\phi}\alpha \\
&+ \beta\frac{s-(s+1)\phi+\phi^{s+1}}{(1-\phi)^2}\phi^{t-s} + \beta\frac{t-(t+1)\phi-s\phi^{t-s}+(s+1)\phi^{t-s+1}}{(1-\phi)^2} \\
&+ \phi^{t-s}\textstyle\sum_{r=1}^{s}\phi^{s-r}W_r + \sum_{r=s+1}^{t}\phi^{t-r}W_r \\
&= \phi^{t-s}\left(\phi^s X_0 + \alpha\frac{1-\phi^s}{1-\phi} + \beta\frac{s-(s+1)\phi+\phi^{s+1}}{(1-\phi)^2} + \textstyle\sum_{r=1}^{s}\phi^{s-r}W_r\right) \\
&+ \alpha\frac{1-\phi^{t-s}}{1-\phi} + \beta\frac{t-(t+1)\phi-s\phi^{t-s}+(s+1)\phi^{t-s+1}}{(1-\phi)^2} + \textstyle\sum_{r=s+1}^{t}\phi^{t-r}W_r,
\end{aligned}$$

which is the desired (7.64). □

Let $\left(\mathcal{F}_t^{X_0,W}\right)_{t\in\hat{n}} \equiv \mathfrak{F}^{X_0,W}$ the filtration generated by the initial state X_0 of the $AR(1)$ process X and the innovation process W, that is

$$\mathcal{F}_0^{X_0,W} \overset{\text{def}}{=} \sigma(X_0) \quad \text{and} \quad \mathcal{F}_t^{X_0,W} \overset{\text{def}}{=} \sigma(X_0, W_1, \ldots, W_t), \quad \forall t \in \mathbb{N},$$

where $\sigma(X, Y, Z, \ldots)$ denotes the σ-algebra generated by the random variables $X, Y, Z, \ldots$

Corollary 555 *The $AR(1)$ process X is adapted to $\mathfrak{F}^{X_0,W}$.*

Proof. The claim is an immediate consequence of Equations (7.62) and (7.63). □

Note that when $X_0 \equiv x_0 \in \mathbb{R}^N$ or we consider an $AR(1)$ process X with time set $\mathbb{T} = \mathbb{Z}$, the process X turns out to adapted to the filtration $\left(\mathcal{F}_t^W\right)_{t\in\mathbb{T}} \equiv \mathfrak{F}^W$ generated by the innovation process W, that is

$$\mathcal{F}_t^W \equiv \bigvee_{s\in\mathbb{T}, s\leqslant t} \sigma(W_s).$$

Corollary 556 *If the innovation W is a Kth-order process, then the $AR(1)$ process X is also a Kth-order process for every $K \geqslant 2$.*

Proof. Recalling that the random variables with finite Kth moment constitute a Banach space, the claim is an immediate consequence of Equations (7.62) and (7.63). □

Corollary 557 *The random variables $X_1, \ldots, X_t$ in the $AR(1)$ process X are uncorrelated with the random variables $W_{t+1}, W_{t+2}, \ldots$ of the innovation process W for every $t \in \mathbb{N}$.*

Corollary 558 *Under the assumption that W is a strong white noise, the random variables $X_1, \ldots, X_t$ in the $AR(1)$ process X are independent of the random variables $W_{t+1}, W_{t+2}, \ldots$ of the innovation process W for every $t \in \mathbb{N}$.*

Corollary 559 *Under the assumption that W is a strong white noise, the $AR(1)$ process X is a Markov process.*

Proof. Considering Equation (7.64), we can write

$$\mathbf{P}\left(X_t \in B \mid \mathcal{F}_s^W\right) = \mathbf{P}\left(\phi^{t-s}X_s + \ell(s,t;\alpha,\beta) + \textstyle\sum_{r=s+1}^{t}\phi^{t-r}W_r \in B \mid \mathcal{F}_s^W\right), \tag{7.66}$$

for every $B \in \mathcal{B}\left(\mathbb{R}^N\right)$, for every $t \geqslant s$. Now, under the assumption that W is a strong white noise, the random variable $\sum_{r=s+1}^{t}\phi^{t-r}W_r$ is independent of $\mathcal{F}_s^{X_0,W}$ and $X_s + \ell(t,s)$ is clearly $\sigma(X_s)$-measurable. Therefore

$$\begin{aligned} &\mathbf{P}\left(\phi^{t-s}X_s + \ell(s,t;\alpha,\beta) + \textstyle\sum_{r=s+1}^{t}\phi^{t-r}W_r \in B \mid \mathcal{F}_s^W\right) \\ &= \mathbf{P}\left(\phi^{t-s}X_s + \ell(s,t;\alpha,\beta) + \textstyle\sum_{r=s+1}^{t}\phi^{t-r}W_r \in B \mid \sigma(X_s)\right), \end{aligned} \tag{7.67}$$

for every $B \in \mathcal{B}\left(\mathbb{R}^N\right)$. Combining Equations (7.66) and (7.67), we show the Markov property. □

Proposizione 560 *The mean function $\mu_X : \hat{n} \to \mathbb{R}$ of X is given by*

$$\mu_X(t) = \begin{cases} \mu_{X_0}, & \text{if } t = 0, \\ \frac{\alpha(1-\phi)-\phi\beta}{(1-\phi)^2} + \frac{\beta}{1-\phi}t + \left(\mu_{X_0} - \frac{\alpha(1-\phi)-\phi\beta}{(1-\phi)^2}\right)\phi^t, & \text{if } t > 0, \end{cases} \tag{7.68}$$

for every $t \in \hat{n}$.

Proof. Equation (7.68) is evident if $t = 0$. Hence, consider the case $t > 0$. On account of Equation (7.62), since W is a white noise, we have

$$\begin{aligned} \mu_X(t) &= \mathbf{E}[X_t] = \mathbf{E}\left[\phi^t X_0 + \ell(t;\alpha,\beta) + \textstyle\sum_{s=1}^{t}\phi^{t-s}W_s\right] \\ &= \phi^t\mathbf{E}[X_0] + \ell(t;\alpha,\beta) + \textstyle\sum_{s=1}^{t}\phi^{t-s}\mathbf{E}[W_s] \\ &= \phi^t\mu_{X_0} + \alpha\frac{1-\phi^t}{1-\phi} + \beta\frac{t-(t+1)\phi+\phi^{t+1}}{(1-\phi)^2} \\ &= \frac{\alpha}{1-\phi} - \frac{\beta\phi}{(1-\phi)^2} + \frac{\beta}{1-\phi}t + \left(\mu_{X_0} - \frac{\alpha}{1-\phi} + \frac{\beta\phi}{(1-\phi)^2}\right)\phi^t \\ &= \frac{\alpha(1-\phi)-\beta\phi}{(1-\phi)^2} + \frac{\beta}{1-\phi}t + \left(\mu_{X_0} - \frac{\alpha(1-\phi)-\beta\phi}{(1-\phi)^2}\right)\phi^t, \end{aligned}$$

as desired. □

Proposizione 561 *The variance function $\sigma_X^2 : \hat{n} \to \mathbb{R}_+$ of X is given by*

$$\sigma_X^2(t) = \begin{cases} \left(\sigma_{X_0}^2 - \frac{\sigma_W^2}{1-\phi^2}\right)\phi^{2t} + \frac{\sigma_W^2}{1-\phi^2}, & \text{if } \phi \neq -1, \\ \sigma_{X_0}^2 + t\sigma_W^2, & \text{if } \phi = -1, \end{cases} \tag{7.69}$$

for every $t \in \hat{n}$.

Proof. Equation (7.69) is evident if $t=0$. When $t>0$, considering the properties of the variance operator and that the random variables in W are uncorrelated, a straightforward computation yields

$$\begin{aligned}\mathbf{D}^2\left[X_t\right] &= \mathbf{D}^2\left[\phi^t X_0+\ell\left(t;\alpha,\beta\right)+\textstyle\sum_{s=1}^t\phi^{t-s}W_s\right]\\ &= \phi^{2t}\mathbf{D}^2\left[X_0\right]+\textstyle\sum_{s=1}^t\phi^{2(t-s)}\mathbf{D}^2\left[W_s\right]\\ &= \sigma_{X_0}^2\phi^{2t}+\sigma_W^2\textstyle\sum_{s=1}^t\left(\phi^2\right)^{t-s}.\end{aligned} \tag{7.70}$$

Now, we have

$$\textstyle\sum_{s=1}^t\left(\phi^2\right)^{t-s}=\begin{cases}\frac{1-\phi^{2t}}{1-\phi^2}, & \text{if } \phi\neq-1\\ t, & \text{if } \phi=-1\end{cases}.$$

Replacing the latter in (7.70), the desired (7.69) follows. □

Proposizione 562 *The autocovariance function $\gamma_X:\mathbb{N}\times\mathbb{N}\to\mathbb{R}$ of X is given by*

$$\gamma_X\left(s,t\right)=\begin{cases}\left(\left(\sigma_{X_0}^2-\frac{\sigma_W^2}{1-\phi^2}\right)\phi^{2s}+\frac{\sigma_W^2}{1-\phi^2}\right)\phi^{t-s}, & \textit{if } \phi\neq-1,\\ \left(-1\right)^{t-s}\left(\sigma_{X_0}^2+s\sigma_W^2\right), & \textit{if } \phi=-1,\end{cases} \tag{7.71}$$

and the autocorrelation function $\rho_X:\mathbb{N}\times\mathbb{N}\to\mathbb{R}$ of X is given by

$$\rho_X\left(s,t\right)=\begin{cases}\dfrac{\left(\phi^{2s}\left(\sigma_{X_0}^2-\frac{\sigma_W^2}{1-\phi^2}\right)+\frac{\sigma_W^2}{1-\phi^2}\right)^{1/2}}{\left(\phi^{2t}\left(\sigma_{X_0}^2-\frac{\sigma_W^2}{1-\phi^2}\right)+\frac{\sigma_W^2}{1-\phi^2}\right)^{1/2}}\phi^{t-s}, & \textit{if } \phi\neq-1,\\ \left(-1\right)^{t-s}\sqrt{\frac{s}{t}\frac{\left(\frac{1}{s}\sigma_{X_0}^2+\sigma_W^2\right)}{\left(\frac{1}{t}\sigma_{X_0}^2+\sigma_W^2\right)}}, & \textit{if } \phi=-1,\end{cases} \tag{7.72}$$

for all $s,t\in\mathbb{N}$ such that $s<t$.

Proof. Considering the properties of the covariance function, Equation (7.64) implies

$$\begin{aligned}\gamma_X\left(s,t\right)=Cov\left(X_s,X_t\right) &= Cov\left(X_s,\phi^{t-s}X_s+\ell\left(s,t;\alpha,\beta\right)+\textstyle\sum_{r=s+1}^t\phi^{t-r}W_r\right)\\ &= \phi^{t-s}Cov\left(X_s,X_s\right)+\textstyle\sum_{r=s+1}^t\phi^{t-r}Cov\left(X_s,W_r\right),\end{aligned} \tag{7.73}$$

for all $s,t\in\mathbb{N}$ such that $s<t$. On the other hand, by Equation (7.62),

$$\begin{aligned}Cov\left(X_s,W_r\right) &= Cov\left(\phi^s X_0+\ell\left(s,t;\alpha,\beta\right)+\textstyle\sum_{q=1}^s\phi^{t-q}W_q,W_r\right)\\ &= \phi^s Cov\left(X_0,W_r\right)+\textstyle\sum_{q=1}^s\phi^{t-q}Cov\left(W_q,W_r\right)\\ &= 0,\end{aligned} \tag{7.74}$$

because X_0 is uncorrelated with the random variables in W, we have $q<r$, and W is a white noise. Combining (7.73) and (7.74), it follows

$$Cov\left(X_s,X_t\right)=\phi^{t-s}\mathbf{D}^2\left[X_s\right]. \tag{7.75}$$

The latter, on account of (7.69), yields the desired (7.71). Still as a consequence of (7.75), we have

$$\rho_X\left(s,t\right)=Corr\left(X_s,X_t\right)=\frac{Cov\left(X_t,X_s\right)}{\mathbf{D}\left[X_t\right]\mathbf{D}\left[X_s\right]}=\frac{\phi^{t-s}\mathbf{D}^2\left[X_s\right]}{\mathbf{D}\left[X_t\right]\mathbf{D}\left[X_s\right]}=\phi^{t-s}\frac{\mathbf{D}\left[X_s\right]}{\mathbf{D}\left[X_t\right]}. \tag{7.76}$$

On the other hand, on account of (7.69), we obtain

$$\phi^{t-s}\frac{\mathbf{D}[X_s]}{\mathbf{D}[X_t]} = \begin{cases} \dfrac{\left(\phi^{2s}\left(\sigma_{X_0}^2-\frac{\sigma_W^2}{1-\phi^2}\right)+\frac{\sigma_W^2}{1-\phi^2}\right)^{1/2}}{\left(\phi^{2t}\left(\sigma_{X_0}^2-\frac{\sigma_W^2}{1-\phi^2}\right)+\frac{\sigma_W^2}{1-\phi^2}\right)^{1/2}}\phi^{t-s}, & \text{if } \phi \neq -1, \\ (-1)^{t-s}\dfrac{\left(\sigma_{X_0}^2+s\sigma_W^2\right)^{1/2}}{\left(\sigma_{X_0}^2+t\sigma_W^2\right)^{1/2}}, & \text{if } \phi = -1. \end{cases} \tag{7.77}$$

Combining (7.76) and (7.77), it clearly follows Equation (7.72). □

Proposizione 563 (Yule-Walker Equations for $AR(1)$) *We have*

$$\gamma_X(t,t) = \phi\gamma_X(t,t-1) + \sigma_W^2, \qquad \textit{and} \qquad \gamma_X(t,t-1) = \phi\gamma_X(t-1,t-1), \tag{7.78}$$

for every $t \in \mathbb{N}$.

Proof. Considering Equation 7.55 in Definition 549, thanks to the properties of the covariance function, we can write

$$\begin{aligned}\gamma_X(t,t) &= Cov(X_t, X_t) = Cov(X_t, \alpha + \beta t + \phi X_{t-1} + W_t) \\ &= \phi Cov(X_t, X_{t-1}) + Cov(X_t, W_t).\end{aligned} \tag{7.79}$$

On the other hand, thanks to Equation (7.62), we have

$$\begin{aligned}Cov(X_t, W_t) &= Cov\left(\phi^t X_0 + \ell(s,t;\alpha,\beta) + \textstyle\sum_{s=1}^{t}\phi^{t-s}W_s, W_t\right) \\ &= \phi^t Cov(X_0, W_t) + \textstyle\sum_{s=1}^{t-1}\phi^{t-s}Cov(W_s, W_t) + Cov(W_t, W_t),\end{aligned} \tag{7.80}$$

for every $t \in \mathbb{N}$. Now, X_0 is independent of the random variables in W and W is a weak white noise, Therefore,

$$Cov(X_0, W_t) = 0 \qquad \text{and} \qquad Cov(W_s, W_t) = 0, \tag{7.81}$$

for all $s, t \in \mathbb{N}$ such that $s < t$. Combining Equations (7.79)-(7.81) we obtain the first equation in (7.78). Similarly, still considering Equation 7.55, we can write

$$\begin{aligned}Cov(X_{t-1}, X_t) &= Cov(X_{t-1}, \alpha + \beta t + \phi X_{t-1} + W_t) \\ &= \phi Cov(X_{t-1}, X_{t-1}) + Cov(X_{t-1}, W_t)\end{aligned} \tag{7.82}$$

and applying again Equation (7.62), a computation analogous to (7.80) yields

$$\begin{aligned}Cov(X_{t-1}, W_t) &= Cov\left(\phi^{t-1}X_0 + \ell(s,t;\alpha,\beta) + \textstyle\sum_{s=1}^{t-1}\phi^{t-1-s}W_s, W_t\right) \\ &= \phi^{t-1}Cov(X_0, W_t) + \textstyle\sum_{s=1}^{t-1}\phi^{t-1-s}Cov(W_s, W_t).\end{aligned} \tag{7.83}$$

For the same reasons as in (7.81) we clearly have

$$Cov(X_{t-1}, W_t) = 0, \tag{7.84}$$

for every $t \in \mathbb{N}$. Hence, combining Equations (7.82)-(7.81) the second equation in (7.78) follows. □

Let $(X_t)_{t\in\hat{n}} \equiv X$ be an $AR(1)$ process on a probability space Ω with innovation $(W_t)_{t\in\mathbb{N}} \equiv W$.

Proposizione 564 *Assume that X_0 is Gaussian, possibly degenerate. In addition, assume that the innovation process W is Gaussian, in symbols $W \sim GWN(\sigma_W^2)$. Then we have*

$$X_t \sim N\left(\mu_X(t), \sigma_X^2(t)\right), \tag{7.85}$$

for every $t \in \mathbb{N}$, where $\mu_X(t)$ and $\sigma_X^2(t)$ are given by (7.68) and (7.69), respectively. In addition, the process X is Gaussian. As a consequence, X has also Gaussian increments.

Proof. Thanks to the independence and Gaussianity of the random variables on the right hand side of Equation (7.62) in Proposition 554, Equation (7.85) immediately follows. Moreover, in case X_0 is degenerate, referring Equation (7.62) to the standardized random variables in W and considering Proposition 411, the Gaussianity of the $AR(1)$ process X immediately follows. In case X_0 is not degenerate, but is independent of the random variables in W, referring Equation (7.62) to the standardization of X_0 and the random variables in W and still considering Proposition 411, we obtain again the Gaussianity of the $AR(1)$ process X. In the end, the Gaussianity of the increments of X comes from Proposition 417. □

Definizione 565 *In light of Proposition 564 we call* Gaussian *an $AR(1)$ process with Gaussian innovation W.*

So far we have obtained our results on $AR(1)$ processes considering $\phi \neq 1$. However, by virtue of Equation (7.62), it is not difficult to recognize that when $|\phi| > 1$ the impact of past shocks, represented by the random variables in W, on the current evolution of the random variables in the process X increases over time. This circumstance is not of great interest for financial or economic modelling. In fact, in most of the evolutions of real variables from Finance or Economics the impact of the past shocks diminishes over time and eventually becomes negligible. In addition, when $\phi = -1$ we are not aware of possible applications of the corresponding $AR(1)$ in modeling. To this goal we consider the following

Assumption 566 *Assume that we have*

$$|\phi| < 1. \tag{7.86}$$

Definizione 567 *The condition $|\phi| < 1$ is usually referred to as* casuality assumption.

Assumption 568 *Assume that we have*

$$\beta = 0 \tag{7.87}$$

Proposizione 569 *Under Assumptions (7.86) and (7.87), we have*

$$\lim_{t \to +\infty} \mu_X(t) = \frac{\alpha}{1-\phi}, \qquad \lim_{t \to +\infty} \sigma_X^2(t) = \frac{\sigma_W^2}{1-\phi^2}. \tag{7.88}$$

In addition,

$$\gamma_X(s,t) \simeq \frac{\phi^{t-s}}{1-\phi^2}\sigma_W^2 \quad \text{and} \quad \rho_X(s,t) \simeq \phi^{t-s}. \tag{7.89}$$

for all $s, t \in \mathbb{N}$ such that $1 \ll s < t$.

The structure of Equations (7.68)-(7.72) suggests a remarkable characterization of the initial state X_0 of the $AR(1)$ process X which implies important properties.

Definizione 570 *We say that the initial state X_0 is a* steady state *of the $AR(1)$ process X if we have*

$$\mu_{X_0} = \frac{\alpha}{1-\phi} \quad \text{and} \quad \sigma_{X_0}^2 = \frac{\sigma_W^2}{1-\phi^2}. \tag{7.90}$$

Osservazione 571 *Under Assumptions (7.86) and (7.87), assume in addition that X_0 is a steady state of the $AR(1)$ process X. Then we have*

$$\mu_X(t) = \frac{\alpha}{1-\phi} \quad \text{and} \quad \sigma_X^2(t) = \frac{\sigma_W^2}{1-\phi^2} \tag{7.91}$$

for every $t \in \hat{n}$. Moreover,,

$$\gamma_X(s,t) = \frac{\phi^{t-s}\sigma_W^2}{1-\phi^2} \quad \text{and} \quad \rho_X(s,t) = \phi^{t-s}, \tag{7.92}$$

for all $s, t \in \hat{n}$ such that $s < t$.

Proposizione 572 *Under Assumptions (7.86) and (7.87), assume in addition that X_0 is a steady state of the $AR(1)$ process X. Then the $AR(1)$ process X is weak sense stationary. In particular, we can consider the reduced autocovariance and autocorrelation function of the process X referred to 0, which are given by*

$$\gamma_{X,0}(t) = \frac{\phi^t \sigma_W^2}{1-\phi^2} \quad \text{and} \quad \rho_{X,0}(t) = \phi^t,$$

for every $t \in \hat{n}$.

Proof. Considering Definition 379, the weak stationarity is immediate consequence of Equations (7.92) and (7.92). □

Proposizione 573 *Under Assumptions (7.86) and (7.87), assume in addition that X_0 is a steady state of the $AR(1)$ process X. Then the $AR(1)$ process X is mean square ergodic in the mean. Moreover,if X_0 has finite 4th-order moment and the innovation W is a 4th-order strong white noise, X is mean-square ergodic in the wide sense.*

Proof. Under Assumption **??**, we have

$$\lim_{t\to\infty} \gamma_{X,0}(t) = \lim_{t\to\infty} \frac{\phi^t \sigma_W^2}{1-\phi^2} = 0.$$

Hence, the mean-square ergodicity in the mean follows from Theorem **??**. Now, under the assumption that X_0 has finite 4th moment and the innovation W is a 4th order strong white noise, with reference to the case $k = 0$, we have

$$\begin{aligned}
&Cov\left(X_0 X_k, X_t X_{t+k}\right) \\
&= Cov\left(X_0^2, X_t^2\right) = Cov\left(X_0^2, (\alpha + \phi X_{t-1} + W_t)^2\right) \\
&= Cov\left(X_0^2, \alpha^2 + \phi^2 X_{t-1}^2 + W_t^2 + 2\alpha\phi X_{t-1} + 2\alpha W_t + 2\phi X_{t-1} W_t\right) \\
&= \phi^2 Cov\left(X_0^2, X_{t-1}^2\right) + Cov\left(X_0^2, W_t^2\right) + 2\alpha\phi Cov\left(X_0^2, X_{t-1}\right) + 2\alpha Cov\left(X_0^2, W_t\right) + 2\phi Cov\left(X_0^2, X_{t-1} W_t\right),
\end{aligned} \tag{7.93}$$

where

$$Cov\left(X_0^2, W_t^2\right) = \mathbf{E}\left[X_0^2 W_t^2\right] - \mathbf{E}\left[X_0^2\right]\mathbf{E}\left[W_t^2\right] = \mathbf{E}\left[W_t^2\right]\mathbf{E}\left[W_t^2\right] - \mathbf{E}\left[X_0^2\right]\mathbf{E}\left[W_t^2\right] = 0, \tag{7.94}$$

$$Cov\left(X_0^2, W_t\right) = \mathbf{E}\left[X_0^2 W_t\right] - \mathbf{E}\left[X_0^2\right]\mathbf{E}\left[W_t\right] = \mathbf{E}\left[X_0^2\right]\mathbf{E}\left[W_t\right] = 0, \tag{7.95}$$

and, by virtue of Corollary 558,

$$\begin{aligned}
Cov\left(X_0^2, X_{t-1} W_t\right) &= \mathbf{E}\left[X_0^2 X_{t-1} W_t\right] - \mathbf{E}\left[X_0^2\right]\mathbf{E}\left[X_{t-1} W_t\right] \\
&= \mathbf{E}\left[X_0^2 X_{t-1}\right]\mathbf{E}\left[W_t\right] - \mathbf{E}\left[X_0^2\right]\mathbf{E}\left[X_{t-1}\right]\mathbf{E}\left[W_t\right] = 0.
\end{aligned} \tag{7.96}$$

In addition, for $t > 1$,

$$\begin{aligned}
Cov\left(X_0^2, X_{t-1}\right) &= Cov\left(X_0^2, \phi^{t-1} X_0 + \alpha\frac{1-\phi^{t-1}}{1-\phi} + \textstyle\sum_{s=1}^{t-1} \phi^{t-1-s} W_s\right) \\
&= Cov\left(X_0^2, \phi^{t-1} X_0\right) + \textstyle\sum_{s=1}^{t-1} \phi^{t-1-s} Cov\left(X_0^2, W_s\right) \\
&= \phi^{t-1}\mathbf{E}\left[X_0^3\right] - \phi^{t-1}\mathbf{E}\left[X_0^2\right]\mathbf{E}\left[X_0\right] \\
&= \phi^{t-1}\left(\mathbf{E}\left[X_0^3\right] - \mathbf{E}\left[X_0^2\right]\mathbf{E}\left[X_0\right]\right)
\end{aligned} \tag{7.97}$$

and, since

$$
\begin{aligned}
X_{t-1}^2 &= \left(\phi^{t-1}X_0 + \alpha\frac{1-\phi^{t-1}}{1-\phi} + \textstyle\sum_{s=1}^{t-1}\phi^{t-1-s}W_s\right)^2 \\
&= \phi^{2(t-1)}X_0^2 + \alpha^2\left(\frac{1-\phi^{t-1}}{1-\phi}\right)^2 + \textstyle\sum_{s=1}^{t-1}\phi^{2(t-1-s)}W_s^2 + 2\alpha\dfrac{1-\phi^{t-1}}{1-\phi}\phi^{t-1}X_0 \\
&+ 2\phi^{t-1}\textstyle\sum_{s=1}^{t-1}\phi^{t-1-s}X_0W_s + 2\alpha\dfrac{1-\phi^{t-1}}{1-\phi}\textstyle\sum_{s=1}^{t-1}\phi^{t-1-s}W_s + 2\textstyle\sum_{\substack{s,r=1\\ r<s}}^{t-1}\phi^{(t-1-r)}\phi^{(t-1-s)}W_rW_s,
\end{aligned}
$$

we have

$$
\begin{aligned}
Cov\left(X_0^2, X_{t-1}^2\right) &= \phi^{2(t-1)}Cov\left(X_0^2, X_0^2\right) + \textstyle\sum_{s=1}^{t-1}\phi^{2(t-1-s)}Cov\left(X_0^2, W_s^2\right) \\
&+ 2\alpha\frac{1-\phi^{t-1}}{1-\phi}\phi^{t-1}Cov\left(X_0^2, X_0\right) + 2\phi^{t-1}\textstyle\sum_{s=1}^{t-1}\phi^{t-1-s}Cov\left(X_0^2, X_0W_s\right) \\
&+ 2\alpha\frac{1-\phi^{t-1}}{1-\phi}\textstyle\sum_{s=1}^{t-1}\phi^{t-1-s}Cov\left(X_0^2, W_s\right) + 2\textstyle\sum_{\substack{s,r=1\\ r<s}}^{t-1}\phi^{(t-1-r)}\phi^{(t-1-s)}Cov\left(X_0^2, W_rW_s\right).
\end{aligned}
\tag{7.98}
$$

Considering that for $r, s = 1, \ldots, t-1$, $r < s$, the random variable X_0 is independent of W_r and W_s, which are independent of each other, we have

$$Cov\left(X_0^2, W_s^2\right) = \mathbf{E}\left[X_0^2W_s^2\right] - \mathbf{E}\left[X_0^2\right]\mathbf{E}\left[W_s^2\right] = \mathbf{E}\left[X_0^2\right]\mathbf{E}\left[W_s^2\right] - \mathbf{E}\left[X_0^2\right]\mathbf{E}\left[W_s^2\right] = 0, \tag{7.99}$$

$$Cov\left(X_0^2, X_0W_s\right) = \mathbf{E}\left[X_0^3W_s\right] - \mathbf{E}\left[X_0^2\right]\mathbf{E}\left[X_0W_s\right] = \mathbf{E}\left[X_0^3\right]\mathbf{E}\left[W_s\right] - \mathbf{E}\left[X_0^2\right]\mathbf{E}\left[X_0\right]\mathbf{E}\left[W_s\right] = 0, \tag{7.100}$$

$$Cov\left(X_0^2, W_s\right) = \mathbf{E}\left[X_0^2W_s\right] - \mathbf{E}\left[X_0^2\right]\mathbf{E}\left[W_s\right] = \mathbf{E}\left[X_0^2\right]\mathbf{E}\left[W_s\right] = 0, \tag{7.101}$$

and

$$
\begin{aligned}
Cov\left(X_0^2, W_rW_s\right) &= \mathbf{E}\left[X_0^2W_rW_s\right] - \mathbf{E}\left[X_0^2\right]\mathbf{E}\left[W_rW_s\right] \\
&= \mathbf{E}\left[X_0^2\right]\mathbf{E}\left[W_rW_s\right] - \mathbf{E}\left[X_0^2\right]\mathbf{E}\left[W_r\right]\mathbf{E}\left[W_s\right] = \mathbf{E}\left[X_0^2\right]\mathbf{E}\left[W_r\right]\mathbf{E}\left[W_s\right] = 0.
\end{aligned}
\tag{7.102}
$$

Combining (7.98)-(7.102), we obtain

$$Cov\left(X_0^2, X_{t-1}^2\right) = \phi^{2(t-1)}\mathbf{D}^2\left[X_0^2\right] + 2\alpha\frac{1-\phi^{t-1}}{1-\phi}\phi^{t-1}\left(\mathbf{E}\left[X_0^3\right] - \mathbf{E}\left[X_0^2\right]\mathbf{E}\left[X_0\right]\right). \tag{7.103}$$

Combining (7.93)-(7.97) and (7.103), it then follows,

$$Cov\left(X_0X_k, X_tX_{t+k}\right) = \phi^{t-1}\left(\mathbf{E}\left[X_0^3\right] - \mathbf{E}\left[X_0^2\right]\mathbf{E}\left[X_0\right]\right) + \phi^{2(t-1)}\mathbf{D}^2\left[X_0^2\right] + 2\alpha\frac{1-\phi^{t-1}}{1-\phi}\phi^{t-1}\left(\mathbf{E}\left[X_0^3\right] - \mathbf{E}\left[X_0^2\right]\mathbf{E}\left[X_0\right]\right.$$

Therefore,

$$\lim_{t\to+\infty} Cov\left(X_0X_k, X_tX_{t+k}\right) = 0.$$

Now, for $k > 0$, we have

$$X_0X_k = X_0\left(\alpha + \phi X_{k-1} + W_k\right) = \alpha X_0 + \phi X_{t+k-1} + X_0W_k$$

and

$$
\begin{aligned}
X_tX_{t+k} &= \left(\alpha + \phi X_{t-1} + W_t\right)\left(\alpha + \phi X_{t+k-1} + W_{t+k}\right) \\
&= \alpha^2 + \alpha\left(W_t + W_{t+k}\right) + \alpha\phi\left(X_{t-1} + W_{t+k-1}\right) + \phi\left(X_{t-1}W_{t+k} + X_{t+k-1}W_t\right) \\
&+ \phi^2X_{t-1}X_{t+k-1} + W_tW_{t+k}.
\end{aligned}
$$

As a consequence,

$$
\begin{aligned}
&Cov\left(X_0X_k, X_tX_{k+t}\right)\\
&= \alpha^2\left(Cov\left(X_0, W_t\right) + Cov\left(X_0, W_{t+k}\right)\right) + \alpha^2\phi\left(Cov\left(X_0, X_{t-1}\right) + Cov\left(X_0, W_{t+k-1}\right)\right)\\
&+ \alpha\phi\left(Cov\left(X_0, X_{t-1}W_{t+k}\right) + Cov\left(X_0, X_{t+k-1}W_t\right)\right) + \alpha\phi^2 Cov\left(X_0, X_{t-1}X_{t+k-1}\right) + \alpha Cov\left(X_0, W_tW_{t+k}\right)\\
&+ \alpha\phi\left(Cov\left(X_0X_{k-1}, W_t\right) + Cov\left(X_0X_{k-1}, W_{t+k}\right)\right) + \alpha\phi^2\left(Cov\left(X_0X_{k-1}, X_{t-1}\right) + Cov\left(X_0X_{k-1}, W_{t+k-1}\right)\right)\\
&+ \phi^2\left(Cov\left(X_0X_{k-1}, X_{t-1}W_{t+k}\right) + Cov\left(X_0X_{k-1}, X_{t+k-1}W_t\right)\right) + \phi^3 Cov\left(X_0X_{k-1}, X_{t-1}X_{t+k-1}\right)\\
&+ \phi Cov\left(X_0X_{k-1}, W_tW_{t+k}\right)\\
&+ \alpha\left(Cov\left(X_0W_k, W_t\right) + Cov\left(X_0W_k, W_{t+k}\right)\right) + \alpha\phi\left(Cov\left(X_0W_k, X_{t-1}\right) + Cov\left(X_0W_k, W_{t+k-1}\right)\right)\\
&+ \phi\left(Cov\left(X_0W_k, X_{t-1}W_{t+k}\right) + Cov\left(X_0W_k, X_{t+k-1}W_t\right)\right) + \phi^2 Cov\left(X_0W_k, X_{t-1}X_{t+k-1}\right)\\
&+ Cov\left(X_0W_k, W_tW_{t+k}\right)
\end{aligned}
$$

for every $t \in \mathbb{N}$ and $k \in \mathbb{N}$. Now, we clearly have

$$Cov\left(X_0, W_t\right) = Cov\left(X_0, W_{t+k}\right) = Cov\left(X_0, W_{t+k-1}\right) = 0$$

and

$$Cov\left(X_0, W_tW_{t+k}\right) = \mathbf{E}\left[X_0W_tW_{t+k}\right] - \mathbf{E}\left[X_0\right]\mathbf{E}\left[W_tW_{t+k}\right] = \mathbf{E}\left[X_0\right]\mathbf{E}\left[W_t\right]\mathbf{E}\left[W_{t+k}\right] - \mathbf{E}\left[X_0\right]\mathbf{E}\left[W_t\right]\mathbf{E}\left[W_{t+k}\right] = 0$$

Furthermore, by virtue of Corollary 558, since $\max\{k-1, t-1\} < t+k$,

$$
\begin{aligned}
Cov\left(X_0, X_{t-1}W_{t+k}\right) &= \mathbf{E}\left[X_0X_{t-1}W_{t+k}\right] - \mathbf{E}\left[X_0\right]\mathbf{E}\left[X_{t-1}W_{t+k}\right]\\
&= \mathbf{E}\left[X_0X_{t-1}\right]\mathbf{E}\left[W_{t+k}\right] - \mathbf{E}\left[X_0\right]\mathbf{E}\left[X_{t-1}\right]\mathbf{E}\left[W_{t+k}\right] = 0,
\end{aligned}
$$

$$
\begin{aligned}
Cov\left(X_0X_{k-1}, W_t\right) &= \mathbf{E}\left[X_0X_{k-1}W_t\right] - \mathbf{E}\left[X_0X_{t-1}\right]\mathbf{E}\left[W_t\right] = \mathbf{E}\left[X_0X_{k-1}\right]\mathbf{E}\left[W_t\right] = 0,\\
Cov\left(X_0X_{k-1}, W_{t+k}\right) &= \mathbf{E}\left[X_0X_{k-1}W_{t+k}\right] - \mathbf{E}\left[X_0X_{t-1}\right]\mathbf{E}\left[W_{t+k}\right] = \mathbf{E}\left[X_0X_{k-1}\right]\mathbf{E}\left[W_{t+k}\right] = 0,
\end{aligned}
$$

and

$$
\begin{aligned}
Cov\left(X_0X_{k-1}, X_{t-1}W_{t+k}\right) &= \mathbf{E}\left[X_0X_{k-1}X_{t-1}W_{t+k}\right] - \mathbf{E}\left[X_0X_{k-1}\right]\mathbf{E}\left[X_{t-1}W_{t+k}\right]\\
&= \mathbf{E}\left[X_0X_{k-1}X_{t-1}\right]\mathbf{E}\left[W_{t+k}\right] - \mathbf{E}\left[X_0X_{k-1}\right]\mathbf{E}\left[X_{t-1}\right]\mathbf{E}\left[W_{t+k}\right] = 0.
\end{aligned}
$$

We have also

$$Cov\left(X_0X_{k-1}, W_{t+k-1}\right) = \mathbf{E}\left[X_0X_{k-1}W_{t+k-1}\right] - \mathbf{E}\left[X_0X_{k-1}\right]\mathbf{E}\left[W_{t+k-1}\right] = \mathbf{E}\left[X_0X_{k-1}\right]\mathbf{E}\left[W_{t+k-1}\right] = 0$$

Similarly, for each $k \in \mathbb{N}$, and for t large enough

$$
\begin{aligned}
Cov\left(X_0W_k, W_t\right) &= \mathbf{E}\left[X_0W_kW_t\right] - \mathbf{E}\left[X_0W_k\right]\mathbf{E}\left[W_t\right] = \mathbf{E}\left[X_0W_k\right]\mathbf{E}\left[W_t\right] = 0,\\
Cov\left(X_0W_k, W_{t+k-1}\right) &= \mathbf{E}\left[X_0W_kW_{t+k-1}\right] - \mathbf{E}\left[X_0W_k\right]\mathbf{E}\left[W_{t+k-1}\right] = \mathbf{E}\left[X_0W_k\right]\mathbf{E}\left[W_{t+k-1}\right] = 0,\\
Cov\left(X_0W_k, W_{t+k}\right) &= \mathbf{E}\left[X_0W_kW_{t+k}\right] - \mathbf{E}\left[X_0W_k\right]\mathbf{E}\left[W_{t+k}\right] = \mathbf{E}\left[X_0W_k\right]\mathbf{E}\left[W_{t+k}\right] = 0,\\
Cov\left(X_0X_{k-1}, W_t\right) &= \mathbf{E}\left[X_0X_{k-1}W_t\right] - \mathbf{E}\left[X_0X_{k-1}\right]\mathbf{E}\left[W_t\right] = \mathbf{E}\left[X_0X_{k-1}\right]\mathbf{E}\left[W_t\right] = 0,
\end{aligned}
$$

$$
\begin{aligned}
Cov\left(X_0X_{k-1}, X_{t-1}W_{t+k}\right) &= \mathbf{E}\left[X_0X_{k-1}X_{t-1}W_{t+k}\right] - \mathbf{E}\left[X_0X_{k-1}\right]\mathbf{E}\left[X_{t-1}W_{t+k}\right]\\
&= \mathbf{E}\left[X_0X_{k-1}X_{t-1}\right]\mathbf{E}\left[W_{t+k}\right] - \mathbf{E}\left[X_0X_{k-1}\right]\mathbf{E}\left[X_{t-1}\right]\mathbf{E}\left[W_{t+k}\right] = 0,
\end{aligned}
$$

$$
\begin{aligned}
Cov\left(X_0X_{k-1}, W_tW_{t+k}\right) &= \mathbf{E}\left[X_0X_{k-1}W_tW_{t+k}\right] - \mathbf{E}\left[X_0X_{k-1}\right]\mathbf{E}\left[W_tW_{t+k}\right]\\
&= \mathbf{E}\left[X_0X_{k-1}W_t\right]\mathbf{E}\left[W_{t+k}\right] - \mathbf{E}\left[X_0X_{k-1}\right]\mathbf{E}\left[W_t\right]\mathbf{E}\left[W_{t+k}\right] = 0,
\end{aligned}
$$

$$\begin{aligned}Cov\left(X_0W_k, X_{t-1}W_{t+k}\right) &= \mathbf{E}\left[X_0W_kX_{t-1}W_{t+k}\right] - \mathbf{E}\left[X_0W_k\right]\mathbf{E}\left[X_{t-1}W_{t+k}\right] \\ &= \mathbf{E}\left[X_0W_kX_{t-1}\right]\mathbf{E}\left[W_{t+k}\right] - \mathbf{E}\left[X_0W_k\right]\mathbf{E}\left[X_{t-1}\right]\mathbf{E}\left[W_{t+k}\right] = 0,\end{aligned}$$

$$\begin{aligned}Cov\left(X_0W_k, W_tW_{t+k}\right) &= \mathbf{E}\left[X_0W_kW_tW_{t+k}\right] - \mathbf{E}\left[X_0W_k\right]\mathbf{E}\left[W_tW_{t+k}\right] \\ &=\end{aligned}$$

and

$$\begin{aligned}&Cov\left(X_0X_k, X_tX_{k+t}\right) \\ &= \alpha^2\phi Cov\left(X_0, X_{t-1}\right) \\ &+ \alpha\phi^2 Cov\left(X_0, X_{t-1}X_{t+k-1}\right) \\ &+ \alpha\phi + \alpha\phi^2 Cov\left(X_0X_{k-1}, X_{t-1}\right) \\ &+ \phi^2\left(+Cov\left(X_0X_{k-1}, X_{t+k-1}W_t\right)\right) + \phi^3 Cov\left(X_0X_{k-1}, X_{t-1}X_{t+k-1}\right) \\ &+ \alpha\phi\left(Cov\left(X_0W_k, X_{t-1}\right) +\right) \\ &+ \phi\left(+Cov\left(X_0W_k, X_{t+k-1}W_t\right)\right) + \phi^2 Cov\left(X_0W_k, X_{t-1}X_{t+k-1}\right) +\end{aligned}$$

$$Cov\left(X_0, X_{t-1}\right) = \gamma_{X,0}\left(t-1\right) = \frac{\phi^{t-1}\sigma_W^2}{1-\phi^2}.$$

By virtue of Theorem **??** the desired result follows. □

Osservazione 574 *Under Assumptions (7.86) and (7.87), assume in addition that X_0 is a Gaussian steady state of the $AR(1)$ process X and $W \sim GWN(\sigma_W^2)$, under the further Assumption* **??**, *we have*

$$X_t \sim N\left(\frac{\alpha}{1-\phi}, \frac{\sigma_W^2}{1-\phi^2}\right)$$

for every $t \in \hat{n}$. In addition, the process X is Gaussian.

Note that, by virtue of Proposition 569, in the long run, Assumptions (7.86) and (7.87), alone let us to bypass the need to deal with an initial steady state of the process. Therefore, for statistical purposes, it is customary to think on $AR(1)$ processes X satisfying Assumptions (7.86) and (7.87) as a weak-sense stationary processes. Eventually, an $AR(1)$ process is thought as already been running since enough time in the past to approach at the present the steady state X_0 in which Equations (**??**) and (**??**) are fulfilled.

Parameter Estimation

Let $(x_t)_{t=1}^T \equiv x$ be a univariate real time series, for some $T \geqslant 2$, and let $(X_t)_{t\in\hat{n}} \equiv X$ be an $AR(1)$ processes with innovation $(W_t)_{t\in\mathbb{N}} \equiv W \sim SWN\left(\sigma_W^2\right)$, satisfying Equation (7.55), for some drift [resp. trend, resp. regression] coefficient $\alpha \in \mathbb{R}$ [resp. $\beta \in \mathbb{R}$, resp. $\phi \in \mathbb{R}$], thought of as parameter. The goal is to determine the values of the parameters α, β, and ϕ which allow the "best fit" of the $AR(1)$ process X to the time series x.

Method of Moments The easiest way to estimate the parameters of the $AR(1)$ processes X is undoubtedly the Method of Moments (MM). The MM estimators are consistent and asymptotically normal but not the most efficient. To obtain estimates of the parameters by MM we need to assume the ergodicity of the process X. Therefore, in what follows we assume that the assumption in Proposition 573 are satisfied. Under these assumptions we have

$$\mu_X(t) = \frac{\alpha}{1-\phi}, \quad \sigma_X^2(t) = \frac{\sigma_W^2}{1-\phi^2}, \quad \rho_{X,0}(t) = \phi^t,$$

Hence, applying the method of moments, we can write the equations

$$\frac{\hat{\alpha}_T}{1-\hat{\phi}_T} = \bar{X}_T = \frac{1}{T}\sum_{t=1}^{T} X_t, \tag{7.104}$$

$$\frac{\hat{\sigma}_{W,T}^2}{1-\hat{\phi}_T^2} = S_T^2(X) = \frac{1}{T}\sum_{t=1}^{T}\left(X_t - \bar{X}_T\right)^2, \tag{7.105}$$

and

$$\hat{\phi}_T = \mathrm{P}_{X,T}(1) = \frac{\Gamma_{X,T}(1)}{\Gamma_{X,T}(0)} = \frac{\Gamma_{X,T}(1)}{S_T^2(X)} = \frac{\sum_{t=1}^{T-1}\left(X_t - \bar{X}_T\right)\left(X_{t+1} - \bar{X}_T\right)^{\intercal}}{\sum_{t=1}^{T}\left(X_t - \bar{X}_T\right)^2}, \tag{7.106}$$

which yieds MM estimators $\hat{\alpha}_T$, $\hat{\phi}_T$, and $\hat{\sigma}_{W,T}^2$ for the true values of the parameters of the process α, ϕ, and σ_W^2. The estimates $\hat{\alpha}_T(\omega)$, $\hat{\phi}_T(\omega)$, and $\hat{\sigma}_{W,T}^2(\omega)$ corresponding to the time series x are easily computed by setting

$$\hat{\phi}_T(\omega) = \frac{\sum_{t=1}^{T-1}(x_t - \bar{x}_T)(x_{t+1} - \bar{x}_T)^{\intercal}}{\sum_{t=1}^{T}(x_t - \bar{x}_T)^2}, \quad \hat{\alpha}_T(\omega) = \left(1 - \hat{\phi}_T(\omega)\right)\bar{x}_T, \quad \hat{\sigma}_{W,T}^2 = \left(1 - \hat{\phi}_T^2(\omega)\right)\frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x}_T)^2,$$

where

$$\bar{x}_T = \frac{1}{T}\sum_{t=1}^{T} x_t.$$

OLS Parameter Estimation To simplify notations when dealing with OLS parameter estimation, we think of the initial time of the time series x as $t=0$ instead of $t=1$, but we keep on thinking of the final time as $t=T$, that is we set $x \equiv (x_t)_{t=0}^{T}$. Then, under the assumption that X is actually a model for x, the stochastic Equation (7.55) leads to the set of algebraic equations

$$x_t = \alpha + \beta t + \phi x_{t-1} + w_t, \quad t = 1, \ldots, T, \tag{7.107}$$

where $(w_t)_{t=1}^{T} \equiv w$ is the unobservable realization of the innovation process W corresponding to the time series x. Similarly to the case of a white noise with a linear trend, we introduce the following quadratic function

Definizione 575 *We call the* residual sum of squares *(RSS), for the set of agebraic equations (7.107), the function* $RSS : \mathbb{R}^3 \to \mathbb{R}_+$ *given by*

$$RSS(\alpha, \beta, \phi) \overset{def}{=} \frac{1}{2}\textstyle\sum_{t=1}^{T}\left(x_t - (\alpha + \beta t + \phi x_{t-1})\right)^2. \tag{7.108}$$

Definizione 576 *We call the* ordinary least squares *(OLS) estimate of the parameters* α, β, *and* ϕ *the real numbers* $\hat{\alpha}$, $\hat{\beta}$, *and* $\hat{\phi}$ *such that*

$$\left(\hat{\alpha}, \hat{\beta}, \hat{\phi}\right) = \underset{(\alpha,\beta,\phi)\in\mathbb{R}^3}{\arg\min}\, RSS(\alpha, \beta, \phi). \tag{7.109}$$

We call the ordinary least squares *(OLS) estimate of the parameter* σ_W^2 *the real number* $\hat{\sigma}_W^2$ *given by*

$$\hat{\sigma}_W^2 = \frac{1}{2}\sum_{t=1}^{T}\left(x_t - \left(\hat{\alpha} + \hat{\beta}t + \hat{\phi}x_{t-1}\right)\right)^2,$$

where the vector $\left(\hat{\alpha}, \hat{\beta}, \hat{\phi}\right)$ *satisfies Equation (7.109).*

Proposizione 577 **??***Introducing the* design matrix X*, the* data *vector* x *and the* parameter *vector* θ *given by*

$$X \equiv \begin{pmatrix} 1 & 1 & x_0 \\ 1 & 2 & x_1 \\ \vdots & \vdots & \vdots \\ 1 & T-1 & x_{T-2} \\ 1 & T & x_{T-1} \end{pmatrix}, \qquad x \equiv \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{T-1} \\ x_T \end{pmatrix}, \qquad \theta \equiv \begin{pmatrix} \alpha \\ \beta \\ \phi \end{pmatrix}, \tag{7.110}$$

respectively, unless a special condition for the generation of the time series x*, the matrix* $X^{\intercal}X$ *is invertible and the OLS estimate vector* $\hat{\theta} \equiv \left(\hat{\alpha}, \hat{\beta}, \hat{\phi}\right)$ *of the parameter vector* θ *is given by*

$$\hat{\theta} = \left(X^{\intercal}X\right)^{-1} X^{\intercal}x. \tag{7.111}$$

Proof. By means of the design matrix X, the data vector x and the parameter vector θ given by (7.110), we can write

$$RSS\left(\alpha, \beta, \phi\right) = \frac{1}{2}\left\|X\theta - x\right\|^2.$$

Hence, the optimization problem described by Equation (7.109) in Definition 576, becomes the determination of $\hat{\theta} \in \mathbb{R}^3$ such that

$$\hat{\theta} = \underset{\theta\in\mathbb{R}^3}{\arg\min}\frac{1}{2}\left\|X\theta - x\right\|^2. \tag{7.112}$$

It is well known that under the assumption that the columns of the design matrix X are linearly independent, then then matrix $X^{\intercal}X$ is invertible and Equation (7.112) has a unique solution $\hat{\theta}$ given by Equation (7.111). On the other hand, the linear independence of the columms of the design matrix X is guaranteed unless we have

$$x_t = \lambda + \mu t, \quad \forall t = 0, 1, \ldots, T-1,$$

for some $\lambda, \mu \in \mathbb{R}$, which would imply that the time series x is generated by a linear deterministic process till the time $t = T-1$. In the end, note that, since

$$\sum_{t=1}^{T} t^2 = \frac{1}{6}T\left(T+1\right)\left(2T+1\right) \quad \text{and} \quad \sum_{t=1}^{T} t = \frac{1}{2}T\left(T+1\right),$$

we have

$$X^{\intercal}X = \begin{pmatrix} 1 & 1 & \vdots & 1 & 1 \\ 1 & 2 & \vdots & T-1 & T \\ x_0 & x_1 & \vdots & x_{T-2} & x_{T-1} \end{pmatrix}\begin{pmatrix} 1 & 1 & x_0 \\ 1 & 2 & x_1 \\ \vdots & \vdots & \vdots \\ 1 & T-1 & x_{T-2} \\ 1 & T & x_{T-1} \end{pmatrix} = \begin{pmatrix} T & \frac{1}{2}T\left(T+1\right) & \sum_{t=1}^{T} x_{t-1} \\ \frac{1}{2}T\left(T+1\right) & \frac{1}{6}T\left(T+1\right)\left(2T+1\right) & \sum_{t=1}^{T} t x_{t-1} \\ \sum_{t=1}^{T} x_{t-1} & \sum_{t=1}^{T} t x_{t-1} & \sum_{t=1}^{T} x_{t-1}^2 \end{pmatrix}.$$

Thus, the invertibility of $X^{\intercal}X$ is equivalent to $\det\left(X^{\intercal}X\right) \neq 0$. This observation leads to another proof of Proposition **??**. based on the first order conditions on the function RSS. In fact, setting for brevity $RSS \equiv g$, we can consider the first order conditions

$$\partial_{\alpha} g\left(\alpha, \beta, \phi\right) = \partial_{\beta} g\left(\alpha, \beta, \phi\right) = \partial_{\phi} g\left(\alpha, \beta, \phi\right) = 0. \tag{7.113}$$

a straghtforward computation yields

$$g(\alpha,\beta,\phi)=\frac{1}{2}\Bigg(\textstyle\sum_{t=1}^{T}x_t^2+T\alpha^2+\frac{1}{6}T(T+1)(2T+1)\beta^2+\left(\sum_{t=1}^{T}x_{t-1}^2\right)\phi^2$$
$$-2\left(\textstyle\sum_{t=1}^{T}x_t\right)\alpha-2\left(\sum_{t=1}^{T}tx_t\right)\beta-2\left(\sum_{t=1}^{T}x_tx_{t-1}\right)\phi$$
$$+T(T+1)\alpha\beta+2\left(\textstyle\sum_{t=1}^{T}x_{t-1}\right)a\phi+2\left(\sum_{t=1}^{T}tx_{t-1}\right)\beta\phi\Bigg).$$

As a consequence,

$$\begin{aligned}
\partial_\alpha g(\alpha,\beta,\phi)&=T\alpha+\tfrac{1}{2}T(T+1)\beta+\left(\textstyle\sum_{t=1}^{T}x_{t-1}\right)\phi-\sum_{t=1}^{T}x_t,\\
\partial_\beta g(\alpha,\beta,\phi)&=\tfrac{1}{2}T(T+1)\alpha+\tfrac{1}{6}T(T+1)(2T+1)\beta+\left(\textstyle\sum_{t=1}^{T}tx_{t-1}\right)\phi-\sum_{t=1}^{T}tx_t,\\
\partial_\phi g(\alpha,\beta,\phi)&=\left(\textstyle\sum_{t=1}^{T}x_{t-1}\right)\alpha+\left(\sum_{t=1}^{T}tx_{t-1}\right)\beta+\left(\sum_{t=1}^{T}x_{t-1}^2\right)\phi-\sum_{t=1}^{T}x_tx_{t-1}.
\end{aligned}\tag{7.114}$$

Thanks to (7.113), we obtain that the determinant Δ of the linear system (7.112) is given by

$$\Delta\equiv\begin{pmatrix} T & \frac{1}{2}T(T+1) & \sum_{t=1}^{T}x_{t-1}\\ \frac{1}{2}T(T+1) & \frac{1}{6}T(T+1)(2T+1) & \sum_{t=1}^{T}tx_{t-1}\\ \sum_{t=1}^{T}x_{t-1} & \sum_{t=1}^{T}tx_{t-1} & \sum_{t=1}^{T}x_{t-1}^2\end{pmatrix}=\det\left(X^{\intercal}X\right)\neq 0.$$

Therefore, the solution of Equation (7.109) can be written as

$$\hat{\alpha}=\frac{\Delta^{(1)}}{\Delta},\quad \hat{\beta}=\frac{\Delta^{(2)}}{\Delta},\quad \hat{\phi}=\frac{\Delta^{(3)}}{\Delta},$$

where $\Delta^{(k)}$ is obtained by Δ replacing the kth column with the columns vector $\left(\sum_{t=1}^{T}x_t,\sum_{t=1}^{T}tx_t,\sum_{t=1}^{T}x_tx_{t-1}\right)^{\intercal}$.
□

Maximum Likelihood (ML) Estimates To apply the maximum likelihood method for the parameter estimation, it is customary to consider some assumptions to make the problem tractable. The main assumption regards the distribution of the innovation process W which is assumed to be Gaussian, that is $W\sim SWN\left(\sigma_W^2\right)$. We recall that in this case the $AR(1)$ process X itself is Gaussian (see 564). We also assume that the process X starts from a steady state X_0 and also (**??**) holds true. Therefore the density of any sample $X_1,\dots,X_T$ drawn by the process X is the function $f_{X_1,\dots,X_T}:\mathbb{R}^T\times\left(\mathbb{R}^2\times\mathbb{R}_+\right)\to\mathbb{R}_+$ given by

$$f_{X_1,\dots,X_T}\left(x_1,\dots,x_T;\alpha,\phi,\sigma_W^2\right)=\frac{1}{(2\pi)^{T/2}\det\left(\Sigma_{X_1,\dots,X_T}\right)^{1/2}}\exp\left(-\frac{1}{2}\left(x-\mu_{X_1,\dots,X_T}\right)^{\intercal}\Sigma_{X_1,\dots,X_T}^{-1}\left(x-\mu_{X_1,\dots,X_T}\right)\right),\tag{7.115}$$

for every $x\equiv(x_1,\dots,x_T)\in\mathbb{R}^T$ and each choice of the parameter vector $\left(\alpha,\phi,\sigma_W^2\right)\in\mathbb{R}^2\times\mathbb{R}_+$, where $\mu_{X_1,\dots,X_T}$ [resp. $\Sigma_{X_1,\dots,X_T}$] is the mean vector [resp.variance-covariance matrix] of the sample $X_1,\dots,X_T$ given by

$$\mu_{X_1,\dots,X_T}=\mu(1,\dots,1)\tag{7.116}$$

where

$$\mu\equiv\frac{\alpha}{1-\phi}$$

and

$$\Sigma_{X_1,\dots,X_T}=\left(Cov\left(X_s,X_t\right)\right)_{s,t=1}^{T}=\left(\frac{\sigma_W^2\phi^{|t-s|}}{1-\phi^2}\right)_{s,t=1}^{T}=\frac{\sigma_W^2}{1-\phi^2}\Phi_T,\tag{7.117}$$

where

$$\Phi_T \equiv \begin{pmatrix} 1 & \phi & \cdots & \phi^{T-2} & \phi^{T-1} \\ \phi & 1 & \cdots & \phi^{T-3} & \phi^{T-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \phi^{T-2} & \phi^{T-3} & \cdots & 1 & \phi \\ \phi^{T-1} & \phi^{T-2} & \cdots & \phi & 1 \end{pmatrix}.$$

The likelihood function of the sample $X_1, \ldots, X_T$ is then the function $\mathcal{L}_{X_1,\ldots,X_T} : \left(\mathbb{R}^2 \times \mathbb{R}_+\right) \times \mathbb{R}^T \to \mathbb{R}_+$ given by

$$\mathcal{L}_{X_1,\ldots,X_T}\left(\alpha, \phi, \sigma_W^2; x_1, \ldots, x_T\right)) = f_{X_1,\ldots,X_T}\left(x_1, \ldots, x_T; \alpha, \phi, \sigma_W^2\right),$$

for every $\left(\alpha, \phi, \sigma_W^2\right) \in \mathbb{R}^3 \times \mathbb{R}_+$ and each realization $(x_1, \ldots, x_T) \in \mathbb{R}^T$ of the sample vector $(X_1, \ldots, X_T)$. The likelihood maximization provides estimators $\hat{\alpha}_T, \hat{\phi}_T, \hat{\sigma}_{W,T}^2$ and estimates $\hat{\alpha}_T(\omega), \hat{\phi}_T(\omega), \hat{\sigma}_{W,T}^2(\omega)$ for the (unknown) true value of the parameters α, ϕ, σ_W^2. Such maximum likelihood (ML) estimates are the solution of the maximization problem

$$\left(\hat{\alpha}, \hat{\phi}, \hat{\sigma}_W^2\right) = \underset{(\alpha,\phi,\sigma_W^2)}{\arg\max} \mathcal{L}_{X_1,\ldots,X_T}\left(\alpha, \phi, \sigma_W^2; x_1, \ldots, x_T\right)). \tag{7.118}$$

However, due to the structure of the likelihood function it is more convenient to consider the equivalent problem of maximizing the logarithm of the likelihood function, that is

$$\left(\hat{\alpha}, \hat{\phi}, \hat{\sigma}_W^2\right) = \underset{(\alpha,\phi,\sigma_W^2)}{\arg\max} \log \mathcal{L}_{X_1,\ldots,X_T}\left(\alpha, \phi, \sigma_W^2; x_1, \ldots, x_T\right)), \tag{7.119}$$

where $\log \mathcal{L}_{X_1,\ldots,X_T} : \mathbb{R}^3 \times \mathbb{R}_+ \to \mathbb{R}$ is given by

$$\log \mathcal{L}_{X_1,\ldots,X_T}\left(\alpha, \phi, \sigma_W^2; x_1, \ldots, x_T\right)) = \log\left(\frac{1}{(2\pi)^{T/2} \det\left(\Sigma_{X_1,\ldots,X_T}\right)^{1/2}} \exp\left(-\frac{1}{2}\left(x - \mu_{X_1,\ldots,X_T}\right)^\intercal \Sigma_{X_1,\ldots,X_T}^{-1}\left(x - \mu_{X_1,\ldots}\right.\right.\right.$$

Indeed, since the logarithm function is strictly increasing, Problem (7.119) has the same solution as Problem (7.118), but it turns out to be computationally easier. Now, considering the properties of the logarithm function, we can write

$$\begin{aligned} &\log \mathcal{L}_{X_1,\ldots,X_T}\left(\alpha, \phi, \sigma_W^2; x_1, \ldots, x_T\right)) \\ &= -\frac{1}{2}\left(T \log(2\pi) + \log\det\left(\Sigma_{X_1,\ldots,X_T}\right) + \left(x - \mu_{X_1,\ldots,X_T}\right)^\intercal \Sigma_{X_1,\ldots,X_T}^{-1}\left(x - \mu_{X_1,\ldots,X_T}\right)\right). \end{aligned}$$

On the other hand, thanks to the Cholesky decomposition, we can write

$$\Phi_T = C_T C_T^\intercal,$$

where,

$$C_T = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \phi & \sqrt{1-\phi^2} & 0 & \cdots & 0 & 0 & 0 \\ \phi^2 & \phi\sqrt{1-\phi^2} & \sqrt{1-\phi^2} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \phi^{T-3} & \phi^{T-4}\sqrt{1-\phi^2} & \phi^{T-5}\sqrt{1-\phi^2} & \cdots & \sqrt{1-\phi^2} & 0 & 0 \\ \phi^{T-2} & \phi^{T-3}\sqrt{1-\phi^2} & \phi^{T-4}\sqrt{1-\phi^2} & \cdots & \phi^2\sqrt{1-\phi^2} & \sqrt{1-\phi^2} & 0 \\ \phi^{T-1} & \phi^{T-2}\sqrt{1-\phi^2} & \phi^{T-3}\sqrt{1-\phi^2} & \cdots & \phi\sqrt{1-\phi^2} & \phi\sqrt{1-\phi^2} & \sqrt{1-\phi^2} \end{pmatrix}.$$

Furthermore, setting

$$D_T^2 \equiv \operatorname{diag}(C_T)^2 = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1-\phi^2 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 1-\phi^2 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1-\phi^2 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1-\phi^2 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1-\phi^2 \end{pmatrix}$$

and

$$L_T = C_T D_T^{-1} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \phi & 1 & 0 & \cdots & 0 & 0 & 0 \\ \phi^2 & \phi & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \phi^{T-3} & \phi^{T-4} & \phi^{T-5} & \cdots & 1 & 0 & 0 \\ \phi^{T-2} & \phi^{T-3} & \phi^{T-4} & \cdots & \phi & 1 & 0 \\ \phi^{T-1} & \phi^{T-2} & \phi^{T-3} & \cdots & \phi^2 & \phi & 1 \end{pmatrix}$$

where

$$D_T^{-1} \equiv \left(D_T^2\right)^{-1/2} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{1-\phi^2}} & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{1-\phi^2}} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{\sqrt{1-\phi^2}} & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & \frac{1}{\sqrt{1-\phi^2}} & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & \frac{1}{\sqrt{1-\phi^2}} \end{pmatrix}$$

we obtain the LDLT decomposition

$$\Phi_T = L_T D^2 L_T^\intercal.$$

Considering (7.173), the latter implies

$$\begin{aligned} \det\left(\Sigma_{X_1,\dots,X_T}\right) &= \det\left(\frac{\sigma_W^2}{1-\phi^2} L_T D^2 L_T^\intercal\right) = \left(\frac{\sigma_W^2}{1-\phi^2}\right)^T \det\left(L_T\right)\det\left(D^2\right)\det\left(L_T^\intercal\right) = \left(\frac{\sigma_W^2}{1-\phi^2}\right)^T \det\left(D^2\right) \\ &= \left(\frac{\sigma_W^2}{1-\phi^2}\right)^T \left(1-\phi^2\right)^{T-1} = \frac{\left(\sigma_W^2\right)^T}{1-\phi^2} \end{aligned}$$

and

$$\Sigma_{X_1,\dots,X_T}^{-1} = \frac{1-\phi^2}{\sigma_W^2}\Phi_T^{-1} = \frac{1-\phi^2}{\sigma_W^2}\left(L_T^{-1}\right)^\intercal \left(D^2\right)^{-1} L_T^{-1},$$

where

$$L_T^{-1} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ -\phi & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -\phi & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 0 & 0 & \cdots & -\phi & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -\phi & 1 \end{pmatrix}$$

and

$$\left(D^2\right)^{-1} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \frac{1}{1-\phi^2} & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{1-\phi^2} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{1-\phi^2} & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & \frac{1}{1-\phi^2} & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & \frac{1}{1-\phi^2} \end{pmatrix}.$$

As a consequence

$$\log\det\left(\Sigma_{X_1,\ldots,X_T}\right) = \log\left(\frac{\left(\sigma_W^2\right)^T}{1-\phi^2}\right) = T\log\left(\sigma_W^2\right) - \log\left(1-\phi^2\right).$$

In addition,

$$\left(x-\mu_{X_1,\ldots,X_T}\right)^\intercal \Sigma_{X_1,\ldots,X_T}^{-1}\left(x-\mu_{X_1,\ldots,X_T}\right) = \frac{1-\phi^2}{\sigma_W^2}\left(x_1-\mu,\ldots,x_T-\mu\right)\left(L_T^{-1}\right)^\intercal\left(D^2\right)^{-1}L_T^{-1}\begin{pmatrix} x_1-\mu \\ \vdots \\ x_T-\mu \end{pmatrix},$$

where

$$\left(x_1-\mu,\ldots,x_T-\mu\right)\left(L_T^{-1}\right)^\intercal \equiv \left(y_1,\ldots y_T\right), \quad y_1 = x_1-\mu,\ y_t = x_t-\mu-\phi\left(x_{t-1}-\mu\right), \quad \forall t = 2,\ldots,T.$$

Combining Equations () ... () we then obtain

$$\begin{aligned}\left(x-\mu_{X_1,\ldots,X_T}\right)^\intercal \Sigma_{X_1,\ldots,X_T}^{-1}\left(x-\mu_{X_1,\ldots,X_T}\right) &= \frac{1-\phi^2}{\sigma_W^2}\left(\left(x_1-\mu\right)^2 + \frac{1}{1-\phi^2}\sum_{t=2}^{T}\left(x_t-\mu-\phi\left(x_{t-1}-\mu\right)\right)^2\right) \\ &= \frac{1}{\sigma_W^2}\left(\left(1-\phi^2\right)\left(x_1-\frac{\alpha}{1-\phi}\right)^2 + \sum_{t=2}^{T}\left(x_t-\left(\alpha+\phi x_{t-1}\right)\right)^2\right)\end{aligned}$$

In the end, we have

$$\begin{aligned}&\log\mathcal{L}_{X_1,\ldots,X_T}\left(\alpha,\phi,\sigma_W^2;x_1,\ldots,x_T)\right) \\ &-\frac{1}{2}\left(T\log\left(2\pi\right) - \log\left(1-\phi^2\right) + T\log\left(\sigma_W^2\right) + \frac{1}{\sigma_W^2}\left(\left(1-\phi^2\right)\left(x_1-\frac{\alpha}{1-\phi}\right)^2 + \sum_{t=2}^{T}\left(x_t-\left(\alpha+\phi x_{t-1}\right)\right)^2\right)\right)\end{aligned}$$

Maximum Conditional Likelihood (MCL) Estimates Another approach is to write the joint density $f_{X_1,\ldots,X_T}: \mathbb{R}^T \times \mathbb{R}^3 \to \mathbb{R}_+$ considering the properties of the conditional densities under the Gaussianity assumption on the innovation process $(W_t)_{t\in\hat{n}} \equiv W$. In fact, it is well known that we have

$$f_{X_1,\ldots,X_T}\left(x_1,\ldots,x_T;\alpha,\phi,\sigma_W^2\right) = \left(\prod_{t=s+1}^{T} f_{X_t|X_1,\ldots,X_{t-1}}\left(x_t \mid x_1,\ldots,x_{t-1};\alpha,\phi,\sigma_W^2\right)\right) f_{X_1,\ldots,X_s}\left(x_1,\ldots,x_s;\alpha,\phi,\sigma_W^2\right)$$

for any $1 \leqslant s < t \leqslant T$. In particular,

$$\begin{aligned}&f_{X_1,\ldots,X_T}\left(x_1,\ldots,x_T;\alpha,\phi,\sigma_W^2\right) \\ &= f_{X_T|X_1,\ldots,X_{T-1}}\left(x_T \mid x_1,\ldots,x_{T-1};\alpha,\phi,\sigma_W^2\right) f_{X_{T-1}|X_1,\ldots,X_{T-2}}\left(x_{T-1} \mid x_1,\ldots,x_{T-2};\alpha,\phi,\sigma_W^2\right) \\ &\cdots f_{X_2|X_1}\left(x_2 \mid x_1;\alpha,\phi,\sigma_W^2\right) f_{X_1}\left(x_1;\alpha,\phi,\sigma_W^2\right)\end{aligned}$$

Now, since X_0 is assumed to be a Gaussian steady state and W is Gaussian, the random variable

$$X_1 = \alpha + \phi X_0 + W_1$$

ia also Gaussian with mean $\frac{\alpha}{1-\phi}$ and variance $\frac{\sigma_W^2}{1-\phi^2}$; in symbols $X_1 \sim N\left(\frac{\alpha}{1-\phi}, \frac{\sigma_W^2}{1-\phi^2}\right)$. We then have

$$f_{X_1}\left(x_1; \alpha, \phi, \sigma_W^2\right) = \sqrt{\frac{1-\phi^2}{2\pi\sigma_W^2}} \exp\left(-\frac{1}{2}\left(\frac{1-\phi^2}{\sigma_W^2}\right)\left(x_1 - \frac{\alpha}{1-\phi}\right)^2\right).$$

On the other hand, since

$$X_2 = \alpha + \phi X_1 + W_1,$$

the observation of the realization x_1 of X_1 provides information about the conditional distribution of X_2, which is given by

$$f_{X_2|X_1}\left(x_2; \alpha, \phi, \sigma_W^2\right) = \frac{1}{\sqrt{2\pi\sigma_W^2}} \exp\left(-\frac{1}{2}\frac{\left(x_2 - (\alpha + \phi x_1)\right)^2}{\sigma_W^2}\right)$$

Similarly, the observation of the realization x_2 of X_2 provides information about the conditional distribution of X_3, which is given by

$$f_{X_3|X_1,X_2}\left(x_3; \alpha, \phi, \sigma_W^2\right) = \frac{1}{\sqrt{2\pi\sigma_W^2}} \exp\left(-\frac{1}{2}\frac{\left(x_3 - (\alpha + \phi x_2)\right)^2}{\sigma_W^2}\right).$$

Iterating this argument we can write

$$f_{X_t|X_1,X_2,\dots,X_{t-1}}\left(x_t; \alpha, \phi, \sigma_W^2\right) = \frac{1}{\sqrt{2\pi\sigma_W^2}} \exp\left(-\frac{1}{2}\frac{\left(x_t - (\alpha + \phi x_{t-1})\right)^2}{\sigma_W^2}\right).$$

Therefore, we obtain

$$\begin{aligned}
&f_{X_1,\dots,X_T}\left(x_1,\dots,x_T; \alpha, \phi, \sigma_W^2\right) \\
&= f_{X_T|X_1,\dots,X_{T-1}}\left(x_T \mid x_1,\dots,x_{T-1}; \alpha, \phi, \sigma_W^2\right) f_{X_{T-1}|X_1,\dots,X_{T-2}}\left(x_{T-1} \mid x_1,\dots,x_{T-2}; \alpha, \phi, \sigma_W^2\right) \\
&\cdots f_{X_2|X_1}\left(x_2 \mid x_1; \alpha, \phi, \sigma_W^2\right) f_{X_1}\left(x_1; \alpha, \phi, \sigma_W^2\right) \\
&= \frac{1}{\sqrt{2\pi\sigma_W^2}} \exp\left(-\frac{1}{2}\frac{\left(x_T - (\alpha + \phi x_{T-1})\right)^2}{\sigma_W^2}\right) \frac{1}{\sqrt{2\pi\sigma_W^2}} \exp\left(-\frac{1}{2}\frac{\left(x_{T-1} - (\alpha + \phi x_{T-2})\right)^2}{\sigma_W^2}\right) \\
&\cdots \frac{1}{\sqrt{2\pi\sigma_W^2}} \exp\left(-\frac{1}{2}\frac{\left(x_2 - (\alpha + \phi x_1)\right)^2}{\sigma_W^2}\right) \sqrt{\frac{1-\phi^2}{2\pi\sigma_W^2}} \exp\left(-\frac{1}{2}\left(\frac{1-\phi^2}{\sigma_W^2}\right)\left(x_1 - \frac{\alpha}{1-\phi}\right)^2\right) \\
&= \frac{\left(1-\phi^2\right)^{1/2}}{\left(2\pi\sigma_W^2\right)^{T/2}} \exp\left(-\frac{1}{2}\left(\frac{1-\phi^2}{\sigma_W^2}\right)\left(x_1 - \frac{\alpha}{1-\phi}\right)^2\right) \prod_{t=1}^{T-1} \exp\left(-\frac{1}{2}\frac{\left(x_t - (\alpha + \phi x_1)\right)^2}{\sigma_W^2}\right).
\end{aligned}$$

As a consequence, we can consider the log-likelihhod function $\log \mathcal{L}_{X_1,\dots,X_T} : \mathbb{R}^2 \times \mathbb{R}_+ \to \mathbb{R}$ given by

$$\begin{aligned}
&\log \mathcal{L}_{X_1,\dots,X_T}\left(\alpha, \phi, \sigma_W^2; x_1,\dots,x_T\right)) \\
&= \log\left(\frac{\left(1-\phi^2\right)^{1/2}}{\left(2\pi\sigma_W^2\right)^{T/2}} \exp\left(-\frac{1}{2}\left(\frac{1-\phi^2}{\sigma_W^2}\right)\left(x_1 - \frac{\alpha}{1-\phi}\right)^2\right) \prod_{t=1}^{T-1} \exp\left(-\frac{1}{2}\frac{\left(x_t - (\alpha + \phi x_1)\right)^2}{\sigma_W^2}\right)\right) \\
&= -\frac{1}{2}\left(\log(2\pi) - \log\left(1-\phi^2\right) + T\log\left(\sigma_W^2\right) + \frac{1}{\sigma_W^2}\left(\left(1-\phi^2\right)\left(x_1 - \frac{\alpha}{1-\phi}\right)^2 + \sum_{t=1}^{T-1}\left(x_t - (\alpha + \phi x_1)\right)^2\right)\right).
\end{aligned}$$

Note that the log-likelihood computed by the conditional distribution approach concides with the log-likelihood computed straightforward approach.

Forecasting

Let $(x_t)_{t=1}^{T} \equiv x$ be a univariate real time series, for some $T \geqslant 2$, and let $(X_t)_{t\in\hat{n}} \equiv X$ be an $AR(1)$ processes with innovation $(W_t)_{t\in\mathbb{N}} \equiv W \sim SWN\left(\sigma_W^2\right)$, satisfying Equation (7.55), for some drift [resp. trend, resp. regression] coefficient $\alpha \in \mathbb{R}$ [resp. $\beta \in \mathbb{R}$, resp. $\phi \in \mathbb{R}$], thought of as parameter. Assume to have determined the values $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\phi}$ of the parameters α, β, and ϕ which allow the "best fit" of the $AR(1)$ process X to the time series x. Let X_{T+k} be the state of the process X at k-steps ahead with respect to the state X_T and let $\left(\mathcal{F}_t^{X_0,W}\right)_{t\in\mathbb{N}} \equiv \mathfrak{F}^{X_0,W}$ be the filtration generated by the initial state X_0 of the $AR(1)$ process X and the innovation process W. Write $X_{T+k,T}$ for the forecast of the state X_{T+k} in light of the information represented by $\mathcal{F}_T^W$. We define $X_{T+k,T}$ as the random variable which minimizes the distance from X_{T+k} in term of the mean square norm, in symbols,

$$X_{T+k,T} = \underset{Y\in L^2\left(\Omega_{\mathcal{F}_T^{X_0,W}},\mathbb{R}\right)}{\arg\min} \mathbf{E}\left[Y - X_{T+k}\right]^2, \tag{7.120}$$

where $L^2\left(\Omega_{\mathcal{F}_T^{X_0,W}},\mathbb{R}\right)$ is the (Hibert) space of the random variables which are measurable with respect to the σ-algebra $\mathcal{F}_T^{X_0,W}$ and have finite 2nd moment. As a consequence of Equation (**??**), we have

$$X_{T+k,T} = \mathbf{E}\left[X_{T+k} \mid \mathcal{F}_T^{X_0,W}\right]. \tag{7.121}$$

It is well known that

$$X_{T+k,T} = h\left(X_0, W_1, \ldots, W_T\right), \tag{7.122}$$

for the unique function $h : \mathbb{R}^{T+1} \to \mathbb{R}$, which solves the minimization problem

$$\underset{g:\mathbb{R}^{T+1}\to\mathbb{R} \text{ t. } g(X_0,W_1,\ldots,W)\in L^2\left(\Omega_{\mathcal{F}_T^{X_0,W}},\mathbb{R}\right)}{\arg\min} \mathbf{E}\left[g\left(X_0, W_1, \ldots, W_T\right) - X_{T+k}\right]^2.$$

We recall some important results which depend only on the properties of the conditional expectation operator on the Hilbert space of the random variables with finite 2nd moment and are independent of the $AR(1)$ structure of the process X.

Recall 578 *We have*

$$\mathbf{E}\left[X_{T+k,T}\right] = \mathbf{E}\left[X_{T+k}\right] \tag{7.123}$$

and

$$\mathbf{E}\left[\left(X_{T+k} - X_{T+k,T}\right) X_{T+k,T}\right] = 0, \tag{7.124}$$

for every $k \in \mathbb{N}$.

Corollary 579 *We have*

$$\mathbf{E}\left[X_{T+k,T}^2\right] = \mathbf{E}\left[X_{t+k}X_{T+k,T}\right] \tag{7.125}$$

and

$$\mathbf{D}^2\left[X_{T+k,T}\right] = Cov\left(X_{T+k}, X_{T+k,T}\right), \tag{7.126}$$

for every $k \in \mathbb{N}$.

Proof. Thanks to Equation (7.124) in Recall 578, we have

$$0 = \mathbf{E}\left[\left(X_{t+k} - X_{T+k,T}\right) X_{T+k,T}\right] = \mathbf{E}\left[X_{t+k}X_{T+k,T} - X_{T+k,T}^2\right] = \mathbf{E}\left[X_{t+k}X_{T+k,T}\right] - \mathbf{E}\left[X_{T+k,T}^2\right],$$

for every $k \in \mathbb{N}$. This is Equation (7.125). Hence, considering Equation (7.123) in Proposition 578, we can write

$$\begin{aligned}
Cov\left(X_{T+k}, X_{T+k,T}\right) &= \mathbf{E}\left[X_{T+k}X_{T+k,T}\right] - \mathbf{E}\left[X_{T+k}\right]\mathbf{E}\left[X_{T+k,T}\right] \\
&= \mathbf{E}\left[X_{T+k,T}^2\right] - \mathbf{E}\left[X_{T+k}\right]\mathbf{E}\left[X_{T+k,T}\right] \\
&= \mathbf{E}\left[X_{T+k,T}^2\right] - \mathbf{E}\left[X_{T+k,T}\right]\mathbf{E}\left[X_{T+k,T}\right] \\
&= \mathbf{D}^2\left[X_{T+k,T}\right],
\end{aligned}$$

for every $k \in \mathbb{N}$, which is Equation (7.126). □

Definizione 580 *We call* forecast error *of the forecast $X_{T+k,T}$ of the state X_{T+k} the random variable*

$$E_{T+k,T} \overset{def}{=} X_{T+k} - X_{T+k,T}, \quad \forall k \in \mathbb{N}. \tag{7.127}$$

We call mean squared error *of the forecast $X_{T+k,T}$ of the state X_{T+k}, the positive number*

$$\mathbf{MSE}\left[E_{T+k,T}\right] \overset{def}{=} \mathbf{E}\left[E_{T+k,T}^2\right], \quad \forall k \in \mathbb{N}. \tag{7.128}$$

Osservazione 581 *We have*

$$\mathbf{E}\left[E_{T+k,T}\right] = 0. \tag{7.129}$$

Hence,

$$\mathbf{MSE}\left[E_{T+k,T}\right] = \mathbf{D}^2\left[E_{T+k,T}\right]. \tag{7.130}$$

Proposizione 582 *We have*

$$\mathbf{D}^2\left[E_{T+k,T}\right] = \mathbf{D}^2\left[X_{T+k}\right] - \mathbf{D}^2\left[X_{T+k,T}\right] \tag{7.131}$$

for every $k \in \mathbb{N}$.

Proof. Considering Equations (7.128) and (7.130), a straightforward computation gives

$$\begin{aligned}
\mathbf{D}^2\left[E_{T+k,T}\right] &= \mathbf{E}\left[E_{T+k,T}^2\right] \\
&= \mathbf{E}\left[\left(X_{T+k} - X_{T+k,T}\right)^2\right] \\
&= \mathbf{E}\left[X_{T+k}^2 + X_{T+k,T}^2 - 2X_{T+k}X_{T+k,T}\right] \\
&= \mathbf{E}\left[X_{T+k}^2\right] - \mathbf{E}\left[X_{T+k}X_{T+k,T} - X_{T+k,T}^2\right] - \mathbf{E}\left[X_{T+k}X_{T+k,T}\right] \\
&= \mathbf{E}\left[X_{T+k}^2\right] - \mathbf{E}\left[X_{T+k}\right]^2 + \mathbf{E}\left[X_{T+k}\right]^2 - \mathbf{E}\left[\left(X_{T+k} - X_{T+k,T}\right)X_{T+k,T}\right] - \mathbf{E}\left[X_{T+k}X_{T+k,T}\right] \\
&= \mathbf{D}^2\left[X_{T+k}\right] + \mathbf{E}\left[X_{T+k}\right]\mathbf{E}\left[X_{T+k}\right] - \mathbf{E}\left[X_{T+k}X_{T+k,T}\right] \\
&= \mathbf{D}^2\left[X_{T+k}\right] + \mathbf{E}\left[X_{T+k}\right]\mathbf{E}\left[X_{T+k,T}\right] - \mathbf{E}\left[X_{T+k}X_{T+k,T}\right] \\
&= \mathbf{D}^2\left[X_{T+k}\right] - Cov\left(X_{T+k}, X_{T+k,T}\right) \\
&= \mathbf{D}^2\left[X_{T+k}\right] - \mathbf{D}^2\left[X_{T+k,T}\right],
\end{aligned}$$

for every $k \in \mathbb{N}$, as desired. □

Now, we turn again our attention to $AR(1)$ processes.

Proposizione 583 *We have*

$$X_{T+k,T} = \phi^k X_T + \alpha\frac{1-\phi^k}{1-\phi} + \beta\frac{T+k-(T+k+1)\phi - T\phi^k + (T+1)\phi^{k+1}}{(1-\phi)^2} + \textstyle\sum_{r=T+1}^{T+k}\phi^{T+k-r}W_r. \tag{7.132}$$

As a consequence,

$$\mathbf{D}^2\left[X_{T+k,T}\right] = Cov\left(X_{T+k}, X_{T+k,T}\right) = \phi^{2k}\mathbf{D}^2\left[X_T\right] = \frac{\phi^{2k}}{1-\phi^2}\sigma_W^2 \tag{7.133}$$

Proof. Thanks to Equation (7.64) in Proposition 554, we can write

$$X_{T+k}=\phi^{k}X_{T}+\alpha\frac{1-\phi^{k}}{1-\phi}+\beta\frac{T+k-(T+k+1)\,\phi-T\phi^{k}+(T+1)\,\phi^{k+1}}{(1-\phi)^{2}}+\textstyle\sum_{r=T+1}^{T+k}\phi^{T+k-r}W_{r}.$$

Therefore, by virtue of the properties of the conditional expectation and considering that $W\sim SWN\left(\sigma_{W}^{2}\right)$, we obtain

$$\begin{aligned}
&X_{T+k,T}\\
&=\mathbf{E}\left[\phi^{k}X_{T}+\alpha\frac{1-\phi^{k}}{1-\phi}+\beta\frac{T+k-(T+k+1)\,\phi-T\phi^{k}+(T+1)\,\phi^{k+1}}{(1-\phi)^{2}}+\textstyle\sum_{r=T+1}^{T+k}\phi^{T+k-r}W_{r}\mid\mathcal{F}_{T}^{X_{0},W}\right]\\
&=\phi^{k}\mathbf{E}\left[X_{T}\mid\mathcal{F}_{T}^{X_{0},W}\right]+\alpha\frac{1-\phi^{k}}{1-\phi}+\beta\frac{T+k-(T+k+1)\,\phi-T\phi^{k}+(T+1)\,\phi^{k+1}}{(1-\phi)^{2}}+\textstyle\sum_{r=T+1}^{T+k}\phi^{T+k-r}\mathbf{E}\left[W_{r}\mid\mathcal{F}_{T}^{X_{0},W}\right]\\
&=\phi^{k}X_{T}+\alpha\frac{1-\phi^{k}}{1-\phi}+\beta\frac{T+k-(T+k+1)\,\phi-T\phi^{k}+(T+1)\,\phi^{k+1}}{(1-\phi)^{2}}+\textstyle\sum_{r=T+1}^{T+k}\phi^{T+k-r}\mathbf{E}\left[W_{r}\right]\\
&=\phi^{k}X_{T}+\alpha\frac{1-\phi^{k}}{1-\phi}+\beta\frac{T+k-(T+k+1)\,\phi-T\phi^{k}+(T+1)\,\phi^{k+1}}{(1-\phi)^{2}}.
\end{aligned}$$

It follows

$$X_{T+k,T}-\mathbf{E}\left[X_{T+k,T}\right]=\phi^{k}\left(X_{T}-\mathbf{E}\left[X_{T}\right]\right).$$

Hence,

$$\mathbf{D}^{2}\left[X_{T+k,T}\right]=\mathbf{E}\left[\left(X_{T+k,T}-\mathbf{E}\left[X_{T+k,T}\right]\right)^{2}\right]=\mathbf{E}\left[\phi^{2k}\left(X_{T}-\mathbf{E}\left[X_{T}\right]\right)^{2}\right]=\phi^{2k}\mathbf{D}^{2}\left[X_{T}\right].$$

for every $k\in\mathbb{N}$. Considering Equations (7.126) in Corollary 579 and (**??**) in Remark **??**, the desired (**??**) follows. □

Proposizione 584 *We have*

$$\mathbf{D}^{2}\left[E_{T+k,T}\right]=\frac{1-\phi^{2k}}{1-\phi^{2}}\sigma_{W}^{2},\tag{7.134}$$

for every $k\in\mathbb{N}$.

Proof. Replacing both Equations (7.91) and Equation (7.133) in Proposition 583 into Equation (7.131), the desired result immediately follows. □

Proposizione 585 *Assume that X_0 is Gaussian and W is a Gaussian white noise, $W\sim GWN\left(\sigma_{W}^{2}\right)$. Then also the forecast $X_{T+k,T}$ of the state X_{T+k} is Gaussian for every $k\in\mathbb{N}$. Therefore, for any $\alpha\in(0,1)$, a $100\,(1-\alpha)\,\%$ prediction interval for the true value of the state X_{T+k} is given by*

$$\left(X_{T+k,T}-z_{\alpha/2}\sqrt{\frac{\phi^{2k}}{1-\phi^{2}}\sigma_{W}^{2}},\ X_{T+k,T}+\sqrt{\frac{\phi^{2k}}{1-\phi^{2}}\sigma_{W}^{2}}\right),\tag{7.135}$$

where $z_{\alpha/2}\equiv z_{\alpha/2}^{+}$ is the upper tail critical value of level $\alpha/2$ of the standard Gaussian random variable. The realization of the predicton interval is then

$$\left(x_{T+k,T}-z_{\alpha/2}\sqrt{\frac{\hat{\phi}\,(\omega)^{2k}}{1-\hat{\phi}\,(\omega)^{2}}\hat{\sigma}_{W}^{2}\,(\omega)},\ x_{T+k,T}+\sqrt{\frac{\hat{\phi}\,(\omega)^{2k}}{1-\hat{\phi}\,(\omega)^{2}}\hat{\sigma}_{W}^{2}\,(\omega)}\right),\tag{7.136}$$

where $x_{T+k,T}$ is the realization of the estimator $X_{T+k,T}$ of the state X_{T+k} and $\hat{\phi}\,(\omega)$ [resp. $\hat{\sigma}_{W}\,(\omega)$] is the estimated value of the parameter ϕ [resp. σ_{W}^{2}].

7.6.2 Higher Order Autoregressive Process - AR(p)

Let $(X_t)_{t\in\hat{n}} \equiv X$ be a stochastic process on a probability space $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ with states in $\mathbb{R}^N$, for some $N \in \mathbb{N}$, and let $(W_t)_{t\in\mathbb{N}} \equiv W$ be a weak white noise on Ω with states in $\mathbb{R}^N$ and variance-covariance matrix Σ_W, in symbols $W \sim WWN^N(\Sigma_W)$. We assume that the random variable X_0 has finite second moment and is independent of the random variables in W.

Definizione 586 *We say that X is a* pth-order N-variate autoregressive process *or an* N-variate autoregressive process of order p*, for some $p \in \mathbb{N}$, with* innovation W *if the random variables in X fulfill the system of equations*

$$\begin{aligned}
&X_1 = \alpha + \beta + \phi_1 X_0 + W_1,\\
&X_2 = \alpha + 2\beta + \phi_1 X_1 + \phi_2 X_0 + W_2,\\
&X_3 = \alpha + 3\beta + \phi_1 X_2 + \phi_2 X_1 + \phi_3 X_0 + W_3,\\
&\cdots\\
&X_{p-1} = \alpha + (p-1)\,\beta + \phi_1 X_{p-2} + \phi_2 X_{p-3} + \cdots + \phi_{p-2} X_1 + \phi_{p-1} X_0 + W_{p-1}\\
&X_p = \alpha + \beta p + \phi_1 X_{p-1} + \phi_2 X_{p-2} + \cdots + \phi_p X_0 + W_p\\
&X_t = \alpha + \beta t + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + W_t, \quad \forall t \geqslant p,
\end{aligned} \tag{7.137}$$

for some vectors $\alpha, \beta \in \mathbb{R}^N$ and matrices $\phi_1, \ldots, \phi_p \in \mathbb{R}^{N^2}$. The random vector X_0 [resp. the distribution of the random vector X_0] is referred to as the initial state *[resp. the* initial distribution*] of the process X. When $X_0 \equiv x_0 \in \mathbb{R}^N$, we call x_0 the* starting point *of the autoregressive process. The vector α [resp. β] is referred to as the* drift, *[resp.* trend*]* coefficient *or* parameter *of X*[3]*; the matrices $\phi_1, \phi_2, \ldots, \phi_p \in \mathbb{R}^{N^2}$ are referred to as the* regression coefficients *or* parameters *of X. The reference to the innovation process W is usually omitted when not necessary. In the case $N = 1$, the reference to the dimension N will be omitted and we simply speak of* pth-order real autoregressive process.

Osservazione 587 *In some circumstances it is more appropriate to consider a stochastic process $(X_t)_{t\in\mathbb{T}} \equiv X$ and a weak white noise $(W_t)_{t\in\mathbb{T}} \equiv W$ with time set $\mathbb{T} \equiv \mathbb{Z}$ and states in $\mathbb{R}^N$. In this case, we say that X is a* pth-order N-variate autoregressive process *with* innovation W *if the random variables in X fulfill the equation*

$$X_t = \alpha + \beta t + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + W_t, \quad \forall t \in \mathbb{Z}, \tag{7.138}$$

for some vectors $\alpha, \beta \in \mathbb{R}^N$ and matrices $\phi_1, \ldots, \phi_2 \in \mathbb{R}^{N^2}$.. In this case, there is no reference to an initial distribution of the process X.

Notation 588 *To denote that X is a pth-order autoregressive process with states in $\mathbb{R}^N$ we write $X \sim AR(p)^N$. In the case $N = 1$, we simply write $X \sim AR(p)$.*

In what follows, we restrain our attention to $AR(p)$ processes X, for which α, β, and $\phi_1, \ldots, \phi_p$ are all real numbers and W is a real weak white noise with variance σ_W^2, for some $\sigma_W > 0$.

We have

$$Cov\,(X_t, W_t) = \sigma_W^2$$

Proposizione 589 *Under assumption $\alpha = \beta = 0$, an $AR(p)$ processes X is weakly stationary if and only the roots of the polynomial*

$$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p$$

are not on the unit circle in the complex plane $\mathbb{C}$.

[3] When we want to stress that $\alpha \neq 0$ and $\beta \neq 0$, [resp. $\alpha \neq 0$ and $\beta = 0$, resp. $\alpha = 0$ and $\beta \neq 0$] we call X a *first order autoregressive process with drift and trend* [resp. with *drift*, resp. with *trend*].

Proposizione 590 *Under assumption $\alpha = \beta = 0$, an $AR(p)$ processes X is casual (ergodic?) if and only the roots of the polynomial*

$$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p$$

are outside the unit disk in the complex plane $\mathbb{C}$.

7.6.3 Parameter Estimation

Under the assumption that the $AR(p)$ process X is weakly stationary, we have

$$\begin{aligned} Cov(X_t, X_t) &= Cov(X_t, \alpha + \beta t + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + W_t) \\ &= \phi_1 Cov(X_t, X_{t-1}) + \phi_2 Cov(X_t, X_{t-2}) + \cdots + \phi_p Cov(X_t, X_{t-p}) + Cov(X_t, W_t) \end{aligned}$$

$$\begin{aligned} Cov(X_{t-1}, X_t) &= Cov(X_{t-1}, \alpha + \beta t + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + W_t) \\ &= \phi_1 Cov(X_{t-1}, X_{t-1}) + \phi_2 Cov(X_{t-1}, X_{t-2}) + \cdots + \phi_p Cov(X_{t-1}, X_{t-p}) + Cov(X_{t-1}, W_t) \end{aligned}$$

$$\cdots$$

$$\begin{aligned} Cov(X_{t-p}, X_t) &= Cov(X_{t-p}, \alpha + \beta t + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + W_t) \\ &= \phi_1 Cov(X_{t-p}, X_{t-1}) + \phi_2 Cov(X_{t-p}, X_{t-2}) + \cdots + \phi_p Cov(X_{t-p}, X_{t-p}) + Cov(X_{t-p}, W_t) \end{aligned}$$

Setting

$$Cov(X_t, X_t) \equiv \alpha(0), \quad Cov(X_{t-1}, X_t) \equiv \alpha(1), \ldots, Cov(X_{t-1}, X_t) \equiv \alpha(p)$$

$$\alpha \equiv (\alpha(1), \ldots, \alpha(p))^\intercal, \quad \phi \equiv (\phi_1, \ldots, \phi_p)^\intercal$$

$$\Gamma_p \equiv \begin{pmatrix} \alpha(0) & \alpha(1) & \cdots & \alpha(p-2) & \alpha(p-1) \\ \alpha(1) & \alpha(2) & \cdots & \alpha(p-3) & \alpha(p-2) \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \alpha(p-2) & \alpha(p-3) & \cdots & \alpha(2) & \alpha(1) \\ \alpha(p-1) & \alpha(p-2) & \cdots & \alpha(1) & \alpha(0) \end{pmatrix} \equiv (\alpha(k-\ell))_{k,\ell=1}^p$$

$$\sigma_W^2 = \alpha(0) - \phi^\intercal \alpha$$

$$\alpha = \Gamma_p \phi$$

$$\Gamma_{X,T}(\tau) \equiv \frac{1}{T} \sum_{t=1}^{T-\tau} \left(X_t - \bar{X}_T\right)\left(X_{t+\tau} - \bar{X}_T\right), \quad \forall \tau = 0, 1, \ldots, p-1.$$

We obtain

$$\begin{pmatrix} \Gamma_{X,T}(0) \\ \Gamma_{X,T}(1) \\ \vdots \\ \Gamma_{X,T}(p-2) \\ \Gamma_{X,T}(p-1) \end{pmatrix} = \begin{pmatrix} \Gamma_{X,T}(0) & \Gamma_{X,T}(1) & \cdots & \Gamma_{X,T}(p-2) & \Gamma_{X,T}(p-1) \\ \Gamma_{X,T}(1) & \Gamma_{X,T}(2) & \cdots & \Gamma_{X,T}(p-3) & \Gamma_{X,T}(p-2) \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \Gamma_{X,T}(p-2) & \Gamma_{X,T}(p-3) & \cdots & \Gamma_{X,T}(2) & \Gamma_{X,T}(1) \\ \Gamma_{X,T}(p-1) & \Gamma_{X,T}(p-2) & \cdots & \Gamma_{X,T}(1) & \Gamma_{X,T}(0) \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_{p-1} \\ \phi_p \end{pmatrix}$$

and

$$\hat{\sigma}_W^2 = \Gamma_{X,T}(0) - \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \end{pmatrix} \begin{pmatrix} \Gamma_{X,T}(0) \\ \Gamma_{X,T}(1) \\ \vdots \\ \Gamma_{X,T}(p-2) \\ \Gamma_{X,T}(p-1) \end{pmatrix}$$

It follows

$$\begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_{p-1} \\ \phi_p \end{pmatrix} = \begin{pmatrix} \Gamma_{X,T}(0) & \Gamma_{X,T}(1) & \cdots & \Gamma_{X,T}(p-2) & \Gamma_{X,T}(p-1) \\ \Gamma_{X,T}(1) & \Gamma_{X,T}(2) & \cdots & \Gamma_{X,T}(p-3) & \Gamma_{X,T}(p-2) \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \Gamma_{X,T}(p-2) & \Gamma_{X,T}(p-3) & \cdots & \Gamma_{X,T}(2) & \Gamma_{X,T}(1) \\ \Gamma_{X,T}(p-1) & \Gamma_{X,T}(p-2) & \cdots & \Gamma_{X,T}(1) & \Gamma_{X,T}(0) \end{pmatrix}^{-1} \begin{pmatrix} \Gamma_{X,T}(0) \\ \Gamma_{X,T}(1) \\ \vdots \\ \Gamma_{X,T}(p-2) \\ \Gamma_{X,T}(p-1) \end{pmatrix}$$

and

$$\hat{\sigma}_W^2 = \Gamma_{X,T}(0) - \begin{pmatrix} \Gamma_{X,T}(0) & \Gamma_{X,T}(1) & \cdots & \Gamma_{X,T}(p-2) & \Gamma_{X,T}(p-1) \end{pmatrix} \begin{pmatrix} \Gamma_{X,T}(0) & \Gamma_{X,T}(1) & \cdots & \Gamma_{X,T} \\ \Gamma_{X,T}(1) & \Gamma_{X,T}(2) & \cdots & \Gamma_{X,T} \\ \vdots & \vdots & \cdots & \\ \Gamma_{X,T}(p-2) & \Gamma_{X,T}(p-3) & \cdots & \Gamma_{X,} \\ \Gamma_{X,T}(p-1) & \Gamma_{X,T}(p-2) & \cdots & \Gamma_{X,} \end{pmatrix}$$

7.7 Parameter Estimation

7.7.1 Predictions

One step ahead

$$X_{T+1} = \alpha + \phi X_T + W_{T+1}$$

$$\hat{X}_{T+1} = \mathbf{E}[X_{T+1} \mid X_T] = \mathbf{E}[\alpha + \phi X_T + W_{T+1} \mid X_T] = \alpha + \phi \mathbf{E}[X_T \mid X_T] + \mathbf{E}[W_{T+1} \mid X_T] = \alpha + \phi X_T$$

Error

$$\mathbf{E}\left[\left(X_{T+1} - \hat{X}_{T+1}\right)^2\right] = \mathbf{E}\left[W_{T+1}^2\right] = \mathbf{D}^2[W_{T+1}]$$

Assume $(W_t)_{t\in\mathbb{Z}}$ is Gaussian distributed. Then, at the confidence level $(1-\alpha)\,\%$, the prediction interval for X_{T+1} is given by

$$\left(\hat{X}_{T+1} - t_{\alpha/2,T-1} S_T(W)\sqrt{1+\frac{1}{T}}, \hat{X}_{T+1} + t_{\alpha/2,T-1} S_T(W)\sqrt{1+\frac{1}{T}}\right)$$

a realization of the prediction interval is given by

$$\left(\hat{\alpha} + \hat{\phi} x_T - t_{\alpha/2,T-1} s_T(W)\sqrt{1+\frac{1}{T}}, \hat{\alpha} + \hat{\phi} x_T - t_{\alpha/2,T-1} s_T(W)\sqrt{1+\frac{1}{T}}\right)$$

In contrast with

$$\left(\hat{X}_{T+1} - z_{\alpha/2}\frac{\sigma_W}{\sqrt{T}}, \hat{X}_{T+1} + z_{\alpha/2}\frac{\sigma_W}{\sqrt{T}}\right)$$

Two step ahead

$$\begin{aligned} X_{T+2} &= \alpha + \phi X_{T+1} + W_{T+1} \\ &= \alpha + \phi(\alpha + \phi X_T + W_{T+1}) + W_{T+2} \\ &= \alpha(1+\phi) + \phi^2 X_T + \phi W_{T+1} + W_{T+2} \end{aligned}$$

Hence,

$$\begin{aligned} \hat{X}_{T+2} &= \mathbf{E}[X_{T+2} \mid X_T] = \mathbf{E}\left[\alpha(1+\phi) + \phi^2 X_T + \phi W_{T+1} + W_{T+2} \mid X_T\right] \\ &= \alpha(1+\phi) + \phi^2 X_T \end{aligned}$$

Error

$$\mathbf{E}\left[\left(X_{T+2}-\hat{X}_{T+2}\right)^2\right] = \mathbf{E}\left[\left(\alpha\left(1+\phi\right)+\phi^2 X_T+\phi W_{T+1}+W_{T+2}-\left(\alpha\left(1+\phi\right)+\phi^2 X_T\right)\right)^2\right]$$
$$= \mathbf{E}\left[\left(\phi W_{T+1}+W_{T+2}\right)^2\right]$$
$$= \mathbf{D}^2\left[\left(\phi W_{T+1}+W_{T+2}\right)\right]$$
$$= \phi^2\mathbf{D}^2\left[W_{T+1}\right]+\mathbf{D}^2\left[W_{T+2}\right]$$
$$= \left(1+\phi^2\right)\sigma_W^2$$

S steps ahead

$$\hat{X}_{T+S} = \alpha\left(1+\phi+\phi^2+\cdots+\phi^{S-1}\right)+\phi^S X_T = \alpha\sum_{r=0}^{S}\phi^r+\phi^S X_T.$$

Error

$$\mathbf{E}\left[\left(X_{T+S}-\hat{X}_{T+S}\right)^2\right] = \sigma_W^2\left(1+\phi^2+\phi^4+\cdots+\phi^{2(S-1)}\right) = \sigma_W^2\sum_{r=0}^{S-1}\phi^{2r}.$$

In the limit

$$\lim_{S\to\infty}\hat{X}_{T+S} = \alpha\sum_{r=0}^{\infty}\phi^r = \frac{\alpha}{1-\phi} = \mathbf{E}\left[X_t\right]$$

and

$$\lim_{S\to\infty}\mathbf{E}\left[\left(X_{T+S}-\hat{X}_{T+S}\right)^2\right] = \sigma_W^2\sum_{r=0}^{\infty}\phi^{2r} = \frac{\sigma_W^2}{1-\phi^2} = \mathbf{D}^2\left[X_t\right]$$

The best forecast (conditional expectation) converges to the unconditional expectation. The mean square error converges to the variance of the process.

7.8 First Order Moving Average Process - MA(1)

Let $(X_t)_{t\in\mathbb{N}} \equiv X$ be a stochastic process on a probability space Ω with states in $\mathbb{R}^N$, for some $N\in\mathbb{N}$, and let $(W_t)_{t\in\hat{n}} \equiv W$ be a weak white noise on Ω with states in $\mathbb{R}^N$ and variance-covariance matrix Σ_W, in symbols $W \sim WWN\left(\Sigma_W\right)$.

Definizione 591 *We say that X is a* 1st-order N-variate *with* state innovation W *if the random variables in X satisfy the equation*

$$X_t = \mu + W_t - \theta W_{t-1}, \quad \forall t\in\mathbb{N}, \tag{7.139}$$

for a vector $\mu\in\mathbb{R}^N$ and a matrix $\theta\in\mathbb{R}^{N^2}$. The vector μ [resp. the matrix θ] is referred to as the mean *[resp.* memory weight*]* coefficient *or* parameter *of the process X*[4]*. In case $N=1$, we usually speak of* real 1st-order moving average process. *The explicit reference to the innovation process W is often omitted.*

Osservazione 592 *In some circumstances it is more appropriate to consider a stochastic process $(X_t)_{t\in\mathbb{T}} \equiv X$ and a weak white noise $(W_t)_{t\in\mathbb{T}} \equiv W$ with time set $\mathbb{T}\equiv\mathbb{Z}$. In this case, we say that X is a* 1st-order N-variate moving average process *with* innovation W *if the random variables in X satisfy Equation (7.139), for every $t\in\mathbb{Z}$.*

[4] When we want to stress that $\mu = 0$ we call X a *demeaned 1st-order moving average process.*

Notation 593 *To denote that X is a 1st-order moving average process with states in $\mathbb{R}^N$ we write $X \sim MA^N(1)$. In case $N=1$, we simply write $X \sim MA(1)$.*

In what follows, we restrain our attention to $MA(1)$ processes X with states in $\mathbb{R}$, satisfying Equation (7.139), for which the innovation W is a real weak noise with variance with variance σ_W^2, for some $\sigma_W > 0$, in symbols $W \sim WWN(\sigma_W^2)$, and the coefficients μ and θ are both real numbers. Recall that in this case both the autocovariance and the autocorrelation functions of X are symmetric, that is

$$\gamma_X(s,t) = \gamma_X(t,s) \quad \text{and} \quad \rho_X(s,t) = \rho_X(t,s),$$

for all $s,t \in \mathbb{N}$. Furthermore, we will consider the technical assumption $\theta \neq 0$ to distinguish a 1st-order moving average process from a random walk.

The following Remark, though rather trivial, is useful to stress the role of $MA(1)$ processes as noises. It should be confronted with the analogous less trivial claim (see Proposition 553) concerning $AR(1)$ processes.

Osservazione 594 *Assume that $(X_t)_{t\in\mathbb{N}} \equiv X$ is an $MA(1)$ process with innovation $(W_t)_{t\in\hat{n}} \equiv W$ satisfying Equation (7.139), for some $\mu,\theta \in \mathbb{R}$. Then we can write*

$$X_t = \mu + Y_t, \tag{7.140}$$

for every $t \in \mathbb{N}$, where $(Y_t)_{t\in\hat{n}} \equiv Y$ is a demeaned $MA(1)$ process with innovation W solution of

$$Y_t = W_t - \theta W_{t-1} \tag{7.141}$$

for every $t \in \mathbb{N}$.

Let $(X_t)_{t\in\mathbb{N}} \equiv X$ be an $MA(1)$ process with innovation $(W_t)_{t\in\hat{n}} \equiv W$ satisfying Equation (7.139), for some $\mu,\theta \in \mathbb{R}$.

Proposizione 595 *We have*

$$X_t = \mu \sum_{s=0}^{t-1} \theta^s - \sum_{s=1}^{t-1} \theta^{t-s} X_s + W_t + (-1)^{t-1}\theta^t W_0, \tag{7.142}$$

for every $t \in \mathbb{N}$, $t \geqslant 2$.

Proof. Fom Equation (7.139), we have

$$W_t = -\mu + X_t + \theta W_{t-1} \tag{7.143}$$

for every $t \in \mathbb{N}$. Therefore, still considering (7.139), we can write

$$X_2 = \mu + W_2 - \theta(-\mu + X_1 + \theta W_0) = \mu(1+\theta) - \theta X_1 + W_2 - \theta^2 W_0,$$

which is Equation (7.142) for $t=2$. Hence, assume inductively that (7.142) holds true for some $t > 2$ and consider the case $t+1$. First, from (7.142), we obtain

$$W_{t-1} = X_{t-1} - \mu \sum_{s=0}^{t-2} \theta^s + \sum_{s=1}^{t-2} \theta^{t-1-s} X_s - (-1)^{t-2}\theta^{t-1} W_0. \tag{7.144}$$

Second, considering (7.143) and (7.144), it follows

$$
\begin{aligned}
X_{t+1} &= \mu + W_{t+1} - \theta W_t \\
&= \mu + W_{t+1} - \theta\left(-\mu + X_t + \theta W_{t-1}\right) \\
&= \mu\left(1+\theta\right) + W_{t+1} - \theta X_t - \theta^2 W_{t-1} \\
&= \mu\left(1+\theta\right) + W_{t+1} - \theta X_t - \theta^2\left(X_{t-1} - \mu\sum_{s=0}^{t-2}\theta^s + \sum_{s=1}^{t-2}\theta^{t-1-s}X_s - \left(-1\right)^{t-2}\theta^{t-1}W_0\right) \\
&= \mu\left(1+\theta\right) + \mu\sum_{s=0}^{t-2}\theta^{s+2} - \theta X_t - \theta^2 X_{t-1} - \sum_{s=1}^{t-2}\theta^{t+1-s}X_s + W_{t+1} + \left(-1\right)^{t-2}\theta^{t+1}W_0 \\
&= \mu\sum_{s=0}^{t}\theta^s - \sum_{s=1}^{t}\theta^{t+1-s}X_s + \left(-1\right)^t\theta^{t+1}W_0 + W_{t+1},
\end{aligned}
$$

which is the desired Equation (7.142) in case $t+1$. This proves that Equation (7.142) holds true for every $t \in \mathbb{N}$, $t \geqslant 2$. □

Corollary 596 *Let $\left(\mathcal{F}_t^W\right)_{t\in\mathbb{N}} \equiv \mathfrak{F}^W$ the filtration generated by the innovation process W, in symbols*

$$\mathcal{F}_t^W \equiv \bigvee_{s\in\hat{n},s\leqslant t}\sigma\left(W_s\right),$$

for every $t \in \hat{n}$. Then the $MA\left(1\right)$ process X is adapted to $\mathfrak{F}^W$.

Corollary 597 *If the white noise W is a Kth-order process, then the $AR\left(1\right)$ process X is also a Kth-order process for every $K \geqslant 2$.*

Proof. Recalling that the random variables with finite Kth moment constitute a Banach space, the claim is an immediate consequence of Equation (7.139). □

Corollary 598 *The random variables $X_1,\ldots,X_t$ in the $MA\left(1\right)$ process X are uncorrelated with the random variables $W_{t+1},W_{t+2},\ldots$ of the innovation process W for every $t \in \mathbb{N}$.*

Corollary 599 *Under the assumption that W is a strong white noise, the random variables $X_1,\ldots,X_t$ in the $MA\left(1\right)$ process X are independent of the random variables $W_{t+1},W_{t+2},\ldots$ of the innovation process W for every $t \in \mathbb{N}$.*

Corollary 600 *Under the assumption that W is a strong white noise, the $MA\left(1\right)$ process X is not a Markov process.*

Proof. The claim is a consequence of Equation (7.142). □

Proposizione 601 *The mean function $\mu_X : \mathbb{N} \to \mathbb{R}$ of X is given by*

$$\mu_X\left(t\right) = \mu, \tag{7.145}$$

for every $t \in \mathbb{N}$.

Proof. Since $W \sim WWN\left(\sigma_W^2\right)$, considering Equation (7.139), thanks to the properties of the expectation operator, we have

$$\mathbf{E}\left[X_t\right] = \mathbf{E}\left[\mu + W_t - \theta W_{t-1}\right] = \mu + \mathbf{E}\left[W_t\right] - \theta\mathbf{E}\left[W_{t-1}\right] = \mu,$$

for every $t \in \mathbb{N}$. □

Proposizione 602 *The variance function $\sigma_X^2 : \mathbb{N} \to \mathbb{R}$ of X is given by*

$$\sigma_X^2(t) = \left(1+\theta^2\right)\sigma_W^2, \tag{7.146}$$

for every $t \in \mathbb{N}$.

Proof. Since $W \sim WWN\left(\sigma_W^2\right)$, considering Equation (7.139), thanks to the properties of the variance operator, we have

$$\mathbf{D}^2[X_t] = \mathbf{D}^2[\mu + W_t - \theta W_{t-1}] = \mathbf{D}^2[W_t] + \theta^2\mathbf{D}^2[W_{t-1}] = \left(1+\theta^2\right)\sigma_W^2,$$

for every $t \in \mathbb{N}$. □

Proposizione 603 *The autocovariance function $\gamma_X : \mathbb{N}\times\mathbb{N} \to \mathbb{R}$ of X is given by*

$$\gamma_X(s,t) = \begin{cases} \left(1+\theta^2\right)\sigma_W^2 & \text{if } t-s=0 \\ -\theta\sigma_W^2 & \text{if } |t-s|=1 \\ 0 & \text{if } |t-s|>1 \end{cases}. \tag{7.147}$$

Proof. Since $W \sim WWN\left(\sigma_W^2\right)$, considering Equation (7.139), thanks to the properties of the covariance functional, we have

$$\begin{aligned}\gamma_X(s,t) &= Cov(X_s, X_t) = Cov(\mu + W_s - \theta W_{s-1}, \mu + W_t - \theta W_{t-1}) \\ &= Cov(W_s, W_t) - \theta Cov(W_s, W_{t-1}) - \theta Cov(W_{s-1}, W_t) + \theta^2 Cov(W_{s-1}, W_{t-1}).\end{aligned}$$

Now

$$Cov(W_s, W_t) = \begin{cases} \sigma_W^2, & \text{if } t=s, \\ 0, & \text{if } t\neq s. \end{cases}$$

It follows

$$\gamma_X(s,t) = \begin{cases} \left(1+\theta^2\right)\sigma_W^2 & \text{if } t=s \\ -\theta\sigma_W^2 & \text{if } t=s+1 \text{ or } t=s-1 \\ 0 & \text{if } t>s+1 \text{ or } t<s-1 \end{cases}.$$

The latter clearly implies (7.147). □

Proposizione 604 *The autocorrelation function $\rho_X : \mathbb{N}\times\mathbb{N} \to \mathbb{R}$ of X is given by*

$$\rho_X(s,t) = \begin{cases} 1 & \text{if } t-s=0 \\ -\frac{\theta}{1+\theta^2} & \text{if } |t-s|=1 \\ 0 & \text{if } |t-s|>1 \end{cases}. \tag{7.148}$$

Proof. It clearly follows combining (**??**) of Definition **??** with (7.146) and (7.147) in Propositions 602 and 603, respectively**Proof.**

Osservazione 605 *For all $s,t \in \mathbb{Z}$ such that $|t-s|>1$ the random variables X_s and X_t in X are uncorrelated. Moreover, if W is a strong white noise process X_s and X_t are independent???*

Proposizione 606 (Yule-Walker Equations for $MA(1)$) *We have*

$$\gamma_X(t,t-1) = -\frac{\theta}{1+\theta^2}\gamma_X(t,t), \tag{7.149}$$

for every $t \in \mathbb{N}$.

Proof. Equation (7.149) is an immediate consequence of Equation (7.147).**Proof.**

Proposizione 607 *The $MA(1)$ process X is weak sense stationary. In particular we can consider the reduced autocovariance and autocorrelation functions of the process X referred to 0, which are given by*

$$\gamma_{X,0}(t) = \begin{cases} \left(1+\theta^2\right)\sigma_W^2 & \textit{if } t = 0 \\ -\theta\sigma_W^2 & \textit{if } t = 1 \\ 0 & \textit{if } t > 1 \end{cases} \quad \textit{and} \quad \rho_{X,0}(t) = \begin{cases} 1 & \textit{if } t = 0 \\ -\frac{\theta}{1+\theta^2} & \textit{if } t = 1 \\ 0 & \textit{if } t > 1 \end{cases}. \tag{7.150}$$

Proof. Considering Definition 379, the weak stationarity is immediate consequence of Equations (7.145) and (7.147).

Proposizione 608 *The $MA(1)$ process X is mean square ergodic in the mean. Moreover, if the innovation W is a 4th order process then X is mean square ergodic in the wide sense.*

Proof. The mean square ergodicity in the mean follows from Theorem **??**. With regard to the wide sense ergodicity, observe that we can write

$$X_1X_k = X_1\left(\mu + W_k - \theta W_{k-1}\right) = \mu X_1 + X_1W_k - \theta X_1W_{k-1}$$

and

$$\begin{aligned} X_tX_{t+k} &= \left(\mu + W_t - \theta W_{t-1}\right)\left(\mu + W_{t+k} - \theta W_{t+k-1}\right) \\ &= \mu^2 + \mu\left(W_t + W_{t+k}\right) - \mu\theta\left(W_{t-1} + W_{t+k-1}\right) - \theta\left(W_{t-1}W_{t+k} + W_tW_{t+k-1}\right) + W_tW_{t+k} + \theta^2 W_{t-1}W_{t+k-1}. \end{aligned}$$

As a consequence, considering Corollary 599, we have

$$\begin{aligned} &Cov\left(X_1X_k, X_tX_{k+t}\right) \\ &= \mu^2\left(Cov\left(X_1, W_t\right) + Cov\left(X_1, W_{t+k}\right)\right) + \mu^2\theta\left(Cov\left(X_1, W_{t-1}\right) + Cov\left(X_1, W_{t+k-1}\right)\right) \\ &+ \theta\left(Cov\left(X_1, W_{t-1}W_{t+k}\right) + Cov\left(X_1, W_tW_{t+k-1}\right)\right) + Cov\left(X_1, W_tW_{t+k}\right) + \theta^2 Cov\left(X_1, W_{t-1}W_{t+k-1}\right) \\ &= 0 \end{aligned}$$

for $k \in \mathbb{N}$ and every $t \in \mathbb{N}$ such that $t > 1$. By virtue of Theorem **??**, the desired result follows. □

Proposizione 609 *Assume that the innovation process W is Gaussian, in symbols $W \sim GWN(\sigma_W^2)$. Then we have*

$$X_t \sim N\left(\mu_X(t), \sigma_X^2(t)\right), \tag{7.151}$$

*for every $t \in \mathbb{N}$, where $\mu_X(t)$ and $\sigma_X^2(t)$ are given by (7.145) and (**??**), respectively. In addition, the process X is Gaussian. As a consequence, X has also Gaussian increments.*

Proof. Thanks to the independence and Gaussianity of the random variables on the right hand side of Equation (7.139) in Definition 591, Equation (7.151) immediately follows. Moreover, referring Equation (7.139) to the standardized random variables in W and considering Proposition 411, we obtain also the Gaussianity of the process $(X_t)_{t\in\mathbb{N}}$. In the end, the Gaussianity of the increments of X comes from Proposition 417. □

Definizione 610 *In light of Proposition 609 we call* Gaussian *an $MA(1)$ process with Gaussian innovation W.*

So far we have obtained our results on $AR(1)$ processes considering only $\theta \neq 0$. However, by virtue of Equation (7.142), it is not difficult to recognize that when $|\theta| > 1$ the memory of the past states of the process X on the current state increases over time. Also this circumstance is not of great interest for financial or economic modelling. In fact, in most of the evolutions of real variables from Finance or Economics the the memory of the past states over time and eventually becomes negligible. To this goal we consider the following

Assumption 611 *Assume that we have*

$$|\theta| < 1. \tag{7.152}$$

Definizione 612 *The condition $|\theta| < 1$ is usually referred to as* invertibility assumption.

7.8.1 Parameter Estimation

Let $(x_t)_{t=1}^{T} \equiv x$ be a univariate real time series, for some $T \geqslant 2$, and let $(X_t)_{t\in\mathbb{N}} \equiv X$ be an $MA(1)$ processes with innovation $(W_t)_{t\in\hat{n}} \equiv W \sim WWN\left(\sigma_W^2\right)$, satisfying Equation (7.139), for some mean [resp. memory] coefficient $\mu \in \mathbb{R}$ [resp. $\theta \in \mathbb{R}$], thought of as parameter. The goal is to determine the values of the parameters μ and θ which allow the "best fit" of the $MA(1)$ process X to x.

Method of Moments (MM) Estimates

From Equations (7.145), (7.146), and (**??**), we know that

$$\mu_X(t) = \mu, \qquad \sigma_X^2(t) = \left(1+\theta^2\right)\sigma_W^2,$$

for every $t \in \mathbb{N}$, and

$$\rho_{X,0}(1) = -\frac{\theta}{1+\theta^2}.$$

The application of the method of moments, requires to set

$$\hat{\mu} = \bar{X}_T \equiv \frac{1}{T}\sum_{t=1}^{T} X_t, \tag{7.153}$$

$$\left(1+\hat{\theta}^2\right)\hat{\sigma}_W^2 = S_T^2(X) \equiv \frac{1}{T}\sum_{t=1}^{T}\left(X_t - \bar{X}_T\right)^2, \tag{7.154}$$

$$-\frac{\hat{\theta}}{1+\hat{\theta}^2} = \mathrm{P}_{X,T}(1) \equiv \frac{\Gamma_{X,T}(1)}{\Gamma_{X,T}(0)} = \frac{\Gamma_{X,T}(1)}{S_T^2(X)} = \frac{\sum_{t=1}^{T-1}\left(X_t - \bar{X}_T\right)\left(X_{t+1} - \bar{X}_T\right)^{\intercal}}{\sum_{t=1}^{T}\left(X_t - \bar{X}_T\right)^2}. \tag{7.155}$$

and thereby solve Equations (7.153)-(7.155) to obtain estimators $\hat{\mu}$, $\hat{\theta}$, and $\hat{\sigma}_W^2$ for the parameters μ, θ, and σ_W^2. Note that Equation (7.155) is the Yule Walker Equation.

The determination of $\hat{\mu}$ is straightforward from Equation (7.153). From Equation (7.155) we obtain

$$\hat{\theta} = \frac{1 \pm \sqrt{1 - 4\mathrm{P}_{X,T}^2(1)}}{2\mathrm{P}_{X,T}(1)}$$

Thus, if

$$\left|\mathrm{P}_{X,T}(1)(\omega)\right| > \frac{1}{2},$$

where $\mathrm{P}_{X,T}^2(1)(\omega)$ is the realization of the statistic $\mathrm{P}_{X,T}^2(1)$, none of the roots of Equation (7.155) can provide an estimate of the parameter θ. On the other hand, if

$$\left|\mathrm{P}_{X,T}(1)(\omega)\right| < \frac{1}{2} \tag{7.156}$$

only

$$\hat{\theta}(\omega) = \frac{1 - \sqrt{1 - 4\mathrm{P}_{X,T}^2(1)(\omega)}}{2\mathrm{P}_{X,T}^2(1)(\omega)}$$

satisfies the invertibility condition

$$\left|\hat{\theta}(\omega)\right| < 1. \tag{7.157}$$

In fact, it is easy to check that

$$|x| < \frac{1}{2} \Rightarrow \left|\frac{1+\sqrt{1-4x^2}}{2x}\right| > 1 \quad \text{and} \quad \left|\frac{1-\sqrt{1-4x^2}}{2x}\right| < 1.$$

Thins means that we have to choose

$$\hat{\theta} = \frac{1-\sqrt{1-4\mathrm{P}^2_{X,T}(1)}}{2\mathrm{P}_{X,T}(1)} \tag{7.158}$$

as an estimator for θ, but that this estimator provides an actual estimate $\hat{\theta}(\omega)$ only if Equation (7.156) holds true. However, once we have an estimate $\hat{\theta}(\omega)$ of the parameter θ satisfying the invertibility condition (7.157), we can easily obtain an estimate for σ^2_W by means of the equation

$$\hat{\sigma}^2_W = \frac{S^2_T(X)}{1+\hat{\theta}^2}, \tag{7.159}$$

which is derived by Equations (7.154).

Summarizing the method of moments estimators for the parameters of a $AR(1)$ processes satisfying the invertibility condition are given by Equations (7.153), (7.158), and (7.159), the latter of which provide actual estimates only if Equation (7.156) holds true.

Approximate Non Linear Least Square (ANLLS) Estimates

To simplify notations, when dealing with $ANLLS$ parameter estimation, we think of the initial time of the time series x as $t=-S$ instead of $t=1$, but we keep on thinking of the final time as $t=T$, that is we set $x \equiv (x_t)_{t=-S}^{T}$. In addition, it is convenient to think of the innovation process W as with time set $\mathbb{T} \equiv \mathbb{Z}$. Then, replicating the procedure which leads to Equation (7.142), we can write

$$X_t = \mu(1+\theta) - \theta X_{t-1} + W_t - \theta^2 W_{t-2} \tag{7.160}$$

for every $t \in \mathbb{N}$, which is in the form of an $AR(1)$ process with innovation $\left(\tilde{W}_t\right)_{t\in\mathbb{N}} \equiv \tilde{W}$ given by

$$\tilde{W}_t \stackrel{\text{def}}{=} W_t - \theta^2 W_{t-2} \quad \forall t \in \mathbb{N}. \tag{7.161}$$

The main problem here is that the innovation term is not a weak white noise, not even if W is a Gaussian white noise, since

$$Cov\left(\tilde{W}_s, \tilde{W}_t\right) = Cov\left(W_s - \theta^2 W_{s-2}, W_t - \theta^2 W_{t-2}\right) = \begin{cases} \left(1+\theta^4\right)\sigma^2_W, & \text{if } t=s, \\ -\theta^2\sigma^2_W, & \text{if } |t-s|=2. \end{cases}$$

On the other hand, iterating the above mentioned procedure, we obtain

$$X_t = \mu\left(1+\theta+\theta^2+\cdots+\theta^{S+1}\right) - \theta X_{t-1} - \theta^2 X_{t-2} - \cdots - \theta^{S+1} X_{t-(S+1)} + W_t - \theta^{S+2} W_{t-(S+2)}, \tag{7.162}$$

for every $t \in \mathbb{N}$, which is in the form of an $AR(s+1)$ process with innovation $\left(\tilde{W}_t\right)_{t\in\mathbb{N}} \equiv \tilde{W}$ given by

$$\tilde{W}_t \stackrel{\text{def}}{=} W_t - \theta^{S+2} W_{t-(S+2)} \quad \forall t \in \mathbb{N}. \tag{7.163}$$

Also in this the innovation term is not a weak white noise, since

$$Cov\left(\tilde{W}_s, \tilde{W}_t\right) = Cov\left(W_s - \theta^{S+2} W_{s-(S+2)}, W_t - \theta^{S+2} W_{t-(S+2)}\right) = \begin{cases} \left(1+\theta^{2(S+2)}\right)\sigma^2_W, & \text{if } t=s, \\ -\theta^{S+2}\sigma^2_W, & \text{if } |t-s|=S+2. \end{cases} \tag{7.164}$$

However, under the invertibility assumption $|\theta| < 1$, the larger is S the weaker the random variables in $\tilde{W}$ are correlated, the closer the variance of the process $\tilde{W}$ is to the variance of W. Now, to obtain $ANLLS$ estimates for the parameters μ, θ, and σ_W^2, first we have to determine the values of μ and θ solution of the minimization problem

$$\left(\hat{\mu}, \hat{\theta}\right) = \underset{(\mu,\theta)\in\mathbb{R}^2}{\arg\min} \sum_{t=1}^{T} \left(x_t - \mu\left(1+\theta+\theta^2+\cdots+\theta^{S+1}\right) + \theta x_{t-1} + \theta^2 x_{t-2} + \cdots + \theta^{S+1} x_{t-(S+1)}\right)^2. \tag{7.165}$$

Second, since from Equation (7.162) it follows

$$W_t - \theta^{S+2} W_{t-(S+2)} = X_t - \mu\left(1+\theta+\theta^2+\cdots+\theta^{S+1}\right) + \theta X_{t-1} + \theta^2 X_{t-2} + \cdots + \theta^{S+1} X_{t-(S+1)},$$

considering Equation (7.164), we can estimate the parameter σ_W^2 by means of

$$\hat{\sigma}_W^2 = \frac{\sum_{t=1}^{T} \left(x_t - \mu\left(1+\hat{\theta}+\hat{\theta}^2+\cdots+\hat{\theta}^{S+1}\right) + \hat{\theta} x_{t-1} + \hat{\theta}^2 x_{t-2} + \cdots + \hat{\theta}^{S+1} x_{t-(S+1)}\right)^2}{1+\hat{\theta}^{2(S+2)}}. \tag{7.166}$$

Conditional Non Linear Least Square (CNLLS) Estimates

We consider again the time series x in the form $(x_t)_t^T$, we assume that the $MA(1)$ process X with innovation $W \sim WWN\left(\sigma_W^2\right)$, which satisfies Equation (7.139), is a model for the time series x, and we assume $W_0 = 0$. Then, with reference to Equation (7.142), we can write

$$W_t = -\mu \sum_{s=0}^{t-1} \theta^s + \sum_{s=1}^{t} \theta^{t-s} X_t, \tag{7.167}$$

for every $t \in \mathbb{N}$. Hence, we can determine estimators $\hat{\mu}$ and $\hat{\theta}$ for the parameters μ and θ as the solution of the optimization problem

$$\left(\hat{\mu}(\omega), \hat{\theta}(\omega)\right) \underset{(\mu,\theta)\in\mathbb{R}\times(-1,1)}{\arg\min} \sum_{t=1}^{T} W_t^2.$$

Once we have the estimators $\hat{\mu}$ and $\hat{\theta}$, we can obtain the corresponding estimators $\hat{\sigma}_W^2$ for the variance of the innovation W in the form

$$\hat{\sigma}_W^2 = \frac{1}{T}\left(-\hat{\mu}\sum_{s=1}^{T} \hat{\theta}^{s-1} + \sum_{s=1}^{T} \hat{\theta}^{t-s} X_s\right)^2. \tag{7.168}$$

In terms of the entries of the time series x we can write Equation (7.167) in the form

$$w_t = -\mu \sum_{s=0}^{t-1} \theta^s + \sum_{s=1}^{t} \theta^{t-s} x_t$$

for every $t = 1, \ldots, T$. Therefore, we can determine estimates $\hat{\mu}(\omega)$ and $\hat{\theta}(\omega)$ for the parameters μ and θ as the solution of the optimization problem

$$\left(\hat{\mu}(\omega), \hat{\theta}(\omega)\right) \underset{(\mu,\theta)\in\mathbb{R}\times(-1,1)}{\arg\min} \sum_{t=1}^{T} w_t^2$$

and obtain the estimate $\hat{\sigma}_W^2(\omega)$ for the variance of the innovation W in the form

$$\hat{\sigma}_W^2(\omega) = \frac{1}{T}\left(-\hat{\mu}(\omega)\sum_{s=1}^{T} \hat{\theta}(\omega)^{s-1} + \sum_{s=1}^{T} \hat{\theta}(\omega)^{t-s} x_s\right)^2.$$

Maximum Likelihood (ML) Estimates

As in case of AR processes, to apply the maximum likelihood method for the parameter estimation of a $MA(1)$ process X with innovation $W \sim WWN\left(\sigma_W^2\right)$, which satisfies Equation (7.139) and is a model for the time series x, we need to consider some assumptions to make the problem tractable. However, besides the invertibility assumption $|\theta| < 1$, in this case we only assume that the distribution of the innovation process W is Gaussian, that is $W \sim SWN\left(\sigma_W^2\right)$. Therefore from now on we proceed under the following

Assumption 613 *We assume*

$$|\theta| < 1 \qquad \textit{and} \qquad W \sim SWN\left(\sigma_W^2\right). \tag{7.169}$$

We recall that when the innovation W is Gaussian, the $MA(1)$ process X itself is Gaussian (see Proposition 609). Therefore the density of any sample $X_1, \dots, X_T$ drawn by the process X is the function $f_{X_1,\dots,X_T} : \mathbb{R}^T \times \left(\mathbb{R}^2 \times \mathbb{R}_+\right) \to \mathbb{R}_+$ given by

$$f_{X_1,\dots,X_T}\left(x_1,\dots,x_T;\mu,\theta,\sigma_W^2\right) = \frac{1}{(2\pi)^{T/2}\det\left(\Sigma_{X_1,\dots,X_T}\right)^{1/2}}\exp\left(-\frac{1}{2}\left(x-\mu_{X_1,\dots,X_T}\right)^{\intercal}\Sigma_{X_1,\dots,X_T}^{-1}\left(x-\mu_{X_1,\dots,X_T}\right)\right),$$

for every $x \equiv (x_1,\dots,x_T) \in \mathbb{R}^T$, and each choice of the parameter vector $\left(\mu,\theta,\sigma_W^2\right) \in \mathbb{R}^2 \times \mathbb{R}_+$, where $\mu_{X_1,\dots,X_T}$ [resp. $\Sigma_{X_1,\dots,X_T}$] is the mean vector [resp.variance-covariance matrix] of the sample $X_1,\dots,X_T$ given by

$$\mu_{X_1,\dots,X_T} = \mu(1,\dots,1)$$

and

$$\Sigma_{X_1,\dots,X_T} = \left(Cov\left(X_s,X_t\right)\right)_{s,t=1}^T = \sigma_W^2\Theta_T, \tag{7.170}$$

where

$$\Theta_T \equiv \begin{pmatrix} 1+\theta^2 & \theta & 0 & \cdots & 0 & 0 & 0 \\ \theta & 1+\theta^2 & \theta & \cdots & 0 & 0 & 0 \\ 0 & \theta & 1+\theta^2 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1+\theta^2 & \theta & 0 \\ 0 & 0 & 0 & \cdots & \theta & 1+\theta^2 & \theta \\ 0 & 0 & 0 & \cdots & 0 & \theta & 1+\theta^2 \end{pmatrix}.$$

The likelihood function of the sample $X_1,\dots,X_T$ is then the function $\mathcal{L}_{X_1,\dots,X_T} : \left(\mathbb{R}^2 \times \mathbb{R}_+\right) \times \mathbb{R}^T \to \mathbb{R}_+$ given by

$$\mathcal{L}_{X_1,\dots,X_T}\left(\mu,\theta,\sigma_W^2;x_1,\dots,x_T\right)) = f_{X_1,\dots,X_T}\left(x_1,\dots,x_T;\mu,\theta,\sigma_W^2\right),$$

for every $\left(\mu,\theta,\sigma_W^2\right) \in \mathbb{R}^2 \times \mathbb{R}_+$, and each realization $(x_1,\dots,x_T) \in \mathbb{R}^T$ of the sample vector $(X_1,\dots,X_T)$. The likelihood maximization provides estimators $\hat{\mu}, \hat{\phi}, \hat{\sigma}_W^2$ and estimates $\hat{\mu}(\omega), \hat{\phi}(\omega), \hat{\sigma}_W^2(\omega)$ for the parameters μ, θ, and σ_W^2 and such maximum likelihood (ML) estimates are the solution of the maximization problem

$$\left(\hat{\mu}(\omega), \hat{\phi}(\omega), \hat{\sigma}_W^2(\omega)\right) = \underset{\left(\mu,\theta,\sigma_W^2\right)}{\arg\max}\mathcal{L}_{X_1,\dots,X_T}\left(\mu,\theta,\sigma_W^2;x_1,\dots,x_T\right)). \tag{7.171}$$

Also in this case, due to the structure of the likelihood function it is more convenient to consider the equivalent problem of maximizing the logarithm of the likelihood function, that is

$$\left(\hat{\mu}(\omega), \hat{\phi}(\omega), \hat{\sigma}_W^2(\omega)\right) = \underset{\left(\mu,\theta,\sigma_W^2\right)}{\arg\max}\log\mathcal{L}_{X_1,\dots,X_T}\left(\mu,\theta,\sigma_W^2;x_1,\dots,x_T\right)), \tag{7.172}$$

where $\log \mathcal{L}_{X_1,\ldots,X_T} : \mathbb{R}^2 \times \mathbb{R}_+ \to \mathbb{R}$ is given by

$$\mathcal{L}_{X_1,\ldots,X_T}\left(\mu, \theta, \sigma_W^2; x_1, \ldots, x_T)\right) = \log\left(\frac{1}{(2\pi)^{T/2} \det\left(\Sigma_{X_1,\ldots,X_T}\right)^{1/2}} \exp\left(-\frac{1}{2}\left(x - \mu_{X_1,\ldots,X_T}\right)^{\intercal} \Sigma_{X_1,\ldots,X_T}^{-1}\left(x - \mu_{X_1,\ldots,X_T}\right.\right.\right.$$

Indeed, since the logarithm function is strictly increasing, Problem (7.172) has the same solution as (7.119), but it turns out to be computationally easier.

Now, considering the properties of the logarithm function, we have

$$\mathcal{L}_{X_1,\ldots,X_T}\left(\mu, \theta, \sigma_W^2; x_1, \ldots, x_T)\right) = -\frac{1}{2}\left(T \log(2\pi) + \log\det\left(\Sigma_{X_1,\ldots,X_T}\right) + \left(x - \mu_{X_1,\ldots,X_T}\right)^{\intercal} \Sigma_{X_1,\ldots,X_T}^{-1}\left(x - \mu_{X_1,\ldots,X_T}\right)\right) \tag{7.173}$$

On the other hand, by the Cholesky decomposition, we can write

$$\Theta_T = C_T C_T^{\intercal},$$

where

$$C_T = \begin{pmatrix} \sqrt{\frac{1-\theta^4}{1-\theta^2}} & 0 & 0 & \cdots & 0 & 0 & 0 \\ \theta\sqrt{\frac{1-\theta^2}{1-\theta^4}} & \sqrt{\frac{1-\theta^6}{1-\theta^4}} & 0 & \cdots & 0 & 0 & 0 \\ 0 & \theta\sqrt{\frac{1-\theta^4}{1-\theta^6}} & \sqrt{\frac{1-\theta^8}{1-\theta^6}} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \sqrt{\frac{1-\theta^{2(T-1)}}{1-\theta^{2(T-2)}}} & 0 & 0 \\ 0 & 0 & 0 & \cdots & \theta\sqrt{\frac{1-\theta^{2(T-2)}}{1-\theta^{2(T-1)}}} & \sqrt{\frac{1-\theta^{2T}}{1-\theta^{2(T-1)}}} & 0 \\ 0 & 0 & 0 & \cdots & 0 & \theta\sqrt{\frac{1-\theta^{2(T-1)}}{1-\theta^{2T}}} & \sqrt{\frac{1-\theta^{2(T+1)}}{1-\theta^{2T}}} \end{pmatrix}.$$

Furthermore, setting

$$D_T^2 \equiv \operatorname{diag}\left(C_T\right)^2 = \begin{pmatrix} \frac{1-\theta^4}{1-\theta^2} & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \frac{1-\theta^6}{1-\theta^4} & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \frac{1-\theta^8}{1-\theta^6} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1-\theta^{2(T-1)}}{1-\theta^{2(T-2)}} & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & \frac{1-\theta^{2T}}{1-\theta^{2(T-1)}} & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & \frac{1-\theta^{2(T+1)}}{1-\theta^{2T}} \end{pmatrix}$$

and

$$L \equiv C_T D_T^{-1} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \theta\frac{1-\theta^2}{1-\theta^4} & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \theta\frac{1-\theta^4}{1-\theta^6} & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 0 & 0 & \cdots & \theta\frac{1-\theta^{2(T-2)}}{1-\theta^{2(T-1)}} & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & \theta\frac{1-\theta^{2(T-1)}}{1-\theta^{2T}} & 1 \end{pmatrix}$$

we obtain the $LDLT$ decomposition

$$\Theta_T = L D^2 L^{\intercal}. \tag{7.174}$$

Considering (7.170), the latter implies

$$\det\left(\Sigma_{X_1,\ldots,X_T}\right) = \det\left(\sigma_W^2 LDL^\intercal\right) = \sigma_W^{2T}\det(L)\det(D)\det\left(L^\intercal\right) = \sigma_W^{2T}\det(D) = \sigma_W^{2T}\frac{1-\theta^{2(T+1)}}{1-\theta^2}, \tag{7.175}$$

and

$$\Sigma_{X_1,\ldots,X_T}^{-1} = \sigma_W^{-2}\left(L^{-1}\right)^\intercal\left(D^2\right)^{-1}L^{-1}, \tag{7.176}$$

where

$$L^{-1} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ -\frac{\theta\left(1-\theta^2\right)}{1-\theta^4} & 1 & 0 & \cdots & 0 & 0 & 0 \\ \frac{\theta^2\left(1-\theta^2\right)}{1-\theta^6} & -\frac{\theta\left(1-\theta^4\right)}{1-\theta^6} & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ (-1)^{T-3}\frac{\theta^{T-3}\left(1-\theta^2\right)}{1-\theta^{2(T-2)}} & (-1)^{T-4}\frac{\theta^{T-4}\left(1-\theta^4\right)}{1-\theta^{2(T-2)}} & (-1)^{T-5}\frac{\theta^{T-5}\left(1-\theta^6\right)}{1-\theta^{2(T-1)}} & \cdots & 1 & 0 & 0 \\ (-1)^{T-2}\frac{\theta^{T-2}\left(1-\theta^2\right)}{1-\theta^{2(T-1)}} & (-1)^{T-3}\frac{\theta^{T-3}\left(1-\theta^4\right)}{1-\theta^{2(T-1)}} & (-1)^{T-4}\frac{\theta^{T-4}\left(1-\theta^6\right)}{1-\theta^{2(T-1)}} & \cdots & -\frac{\theta\left(1-\theta^{2(T-2)}\right)}{1-\theta^{2(T-1)}} & 1 & 0 \\ (-1)^{T-1}\frac{\theta^{T-1}\left(1-\theta^2\right)}{1-\theta^{2T}} & (-1)^{T-2}\frac{\theta^{T-2}\left(1-\theta^4\right)}{1-\theta^{2T}} & (-1)^{T-3}\frac{\theta^{T-3}\left(1-\theta^6\right)}{1-\theta^{2T}} & \cdots & \frac{\theta^2\left(1-\theta^{2(T-2)}\right)}{1-\theta^{2T}} & -\frac{\theta\left(1-\theta^{2(T-1)}\right)}{1-\theta^{2T}} & 1 \end{pmatrix}$$

and

$$\left(D^2\right)^{-1} = \begin{pmatrix} \frac{1-\theta^2}{1-\theta^4} & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \frac{1-\theta^4}{1-\theta^6} & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \frac{1-\theta^6}{1-\theta^8} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1-\theta^{2(T-2)}}{1-\theta^{2(T-1)}} & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & \frac{1-\theta^{2(T-1)}}{1-\theta^{2T}} & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & \frac{1-\theta^{2T}}{1-\theta^{2(T+1)}} \end{pmatrix}.$$

As a consequence,

$$\log\det\left(\Sigma_{X_1,\ldots,X_T}\right) = \log\left(\sigma_W^{2T}\frac{1-\theta^{2(T+1)}}{1-\theta^2}\right) = T\log\left(\sigma_W^2\right) + \log\left(1-\theta^{2(T+1)}\right) - \log\left(1-\theta^2\right). \tag{7.177}$$

In addition,

$$\left(x-\mu_{X_1,\ldots,X_T}\right)^\intercal\Sigma_{X_1,\ldots,X_T}^{-1}\left(x-\mu_{X_1,\ldots,X_T}\right) = \sigma_W^{-2}\left(x_1-\mu,\ldots,x_T-\mu\right)\left(L^{-1}\right)^\intercal\left(D^2\right)^{-1}L^{-1}\begin{pmatrix} x_1-\mu \\ \vdots \\ x_T-\mu \end{pmatrix}, \tag{7.178}$$

where

$$\left(x_1-\mu,\ldots,x_T-\mu\right)\left(L^{-1}\right)^\intercal \equiv \left(y_1,\ldots,y_T\right), \quad y_t \equiv x_t-\mu-\theta\frac{1-\theta^{2(t-1)}}{1-\theta^{2t}}y_{t-1}, \quad \forall t=1,\ldots,T. \tag{7.179}$$

Combining Equations (7.178) and (7.179) we then obtain

$$\left(x-\mu_{X_1,\ldots,X_T}\right)^\intercal\Sigma_{X_1,\ldots,X_T}^{-1}\left(x-\mu_{X_1,\ldots,X_T}\right) = \frac{1}{\sigma_W^2}\sum_{t=1}^T\frac{1-\theta^{2t}}{1-\theta^{2(t+1)}}y_t^2. \tag{7.180}$$

In the end, from (7.173), (7.177), and (7.180), it follows

$$\begin{aligned}&\mathcal{L}_{X_1,\ldots,X_T}\left(\mu,\theta,\sigma_W^2;x_1,\ldots,x_T\right)) \\ &= -\frac{1}{2}\left(T\log\left(2\pi\right)+T\log\left(\sigma_W^2\right)+\log\left(1-\theta^{2(T+1)}\right)-\log\left(1-\theta^2\right)+\frac{1}{\sigma_W^2}\sum_{t=1}^{T}\frac{1-\theta^{2t}}{1-\theta^{2(t+1)}}y_t^2\right).\end{aligned} \tag{7.181}$$

The maximization of the log-likelihood in (7.181) has to be performed numerically.

Maximum Conditional Likelihood (MCL) Estimates

Another approach is to write the joint density $f_{X_1,\ldots,X_T}:\mathbb{R}^T\times\mathbb{R}^3\to\mathbb{R}_+$ considering the properties of the conditional densities under the Gaussianity assumption on the innovation process $(W_t)_{t\in\hat{n}}\equiv W$. In fact, it is well known that we have

$$f_{X_1,\ldots,X_T}\left(x_1,\ldots,x_T;\mu,\theta,\sigma_W^2\right)=\left(\prod_{t=s+1}^{T}f_{X_t|X_1,\ldots,X_{t-1}}\left(x_t\mid x_1,\ldots,x_{t-1};\mu,\theta,\sigma_W^2\right)\right)f_{X_1,\ldots,X_s}\left(x_1,\ldots,x_s;\mu,\theta,\sigma_W^2\right)$$

for any $1\leqslant s<t\leqslant T$. In particular,

$$\begin{aligned}&f_{X_1,\ldots,X_T}\left(x_1,\ldots,x_T;\mu,\theta,\sigma_W^2\right)\\ &=f_{X_T|X_1,\ldots,X_{T-1}}\left(x_T\mid x_1,\ldots,x_{T-1};\mu,\theta,\sigma_W^2\right)f_{X_{T-1}|X_1,\ldots,X_{T-2}}\left(x_{T-1}\mid x_1,\ldots,x_{T-2};\mu,\theta,\sigma_W^2\right)\\ &\cdots f_{X_2|X_1}\left(x_2\mid x_1;\mu,\theta,\sigma_W^2\right)f_{X_1}\left(x_1;\mu,\theta,\sigma_W^2\right)\end{aligned}$$

Now, since W is Gaussian, the random variable

$$X_1=\mu+W_1-\theta W_0 \tag{7.182}$$

is Gaussian with mean μ and variance $\left(1+\theta^2\right)\sigma_W^2$, in symbols $X_1\sim N\left(\mu,\left(1+\theta^2\right)\sigma_W^2\right)$, we then have

$$f_{X_1}\left(x_1;\mu,\theta,\sigma_W^2\right)=\frac{1}{\sqrt{2\pi\left(1+\theta^2\right)\sigma_W^2}}\exp\left(-\frac{1}{2}\frac{(x_1-\mu)^2}{\left(1+\theta^2\right)\sigma_W^2}\right).$$

On the other hand, since

$$X_2=\mu+W_2-\theta W_1,$$

cause of Equation (7.182), the observation of the realization x_1 of X_1 provides no information about the realization of X_2 unless we assume that $W_0=0$, as we alread did in the context of $CNLLS$ Estimates. Under this assumption we have

$$f_{X_1|W_0=0}\left(x_1;\mu,\theta,\sigma_W^2\right)=\frac{1}{\sqrt{2\pi\sigma_W^2}}\exp\left(-\frac{1}{2}\frac{(x_1-\mu)^2}{\sigma_W^2}\right)$$

and the observation of the realization x_1 of X_1 allows to observe the realization w_1 of W_1, which provides information about the realization of X_2. In fact, the assumption $W_0=0$ yields

$$w_1=x_1-\mu$$

and

$$X_2=\mu+W_2-\theta\left(x_1-\mu\right)=\mu\left(1+\theta\right)-\theta x_1+W_2.$$

The latter implies

$$f_{X_2|X_1,W_0=0}\left(x_2;\mu,\theta,\sigma_W^2\right) = \frac{1}{\sqrt{2\pi\sigma_W^2}}\exp\left(-\frac{1}{2}\frac{\left(x_2-\mu\left(1+\theta\right)+\theta x_1\right)^2}{\sigma_W^2}\right)$$

Similarly, still the assumption $W_0 = 0$ yields

$$w_2 = x_2 - \mu + \theta\left(x_1-\mu\right) = -\mu\left(1+\theta\right) + \theta x_1 + x_2$$

and

$$X_3 = \mu + W_3 - \theta w_2 = \mu\left(1+\theta+\theta^2\right) - \theta^2 x_1 - \theta x_2 + W_3.$$

From which

$$f_{X_3|X_1,X_2,W_0=0}\left(x_3;\mu,\theta,\sigma_W^2\right) = \frac{1}{\sqrt{2\pi\sigma_W^2}}\exp\left(-\frac{1}{2}\frac{\left(x_3-\mu\left(1+\theta+\theta^2\right)+\theta^2 x_1+\theta x_2\right)^2}{\sigma_W^2}\right).$$

Iterating this argument, we can write

$$f_{X_t|X_1,X_2,\ldots,X_{t-1},W_0=0}\left(x_t;\mu,\theta,\sigma_W^2\right) = \frac{1}{\sqrt{2\pi\sigma_W^2}}\exp\left(-\frac{1}{2}\frac{\left(x_t-\mu\sum_{s=0}^{t-1}\theta^s+\sum_{s=1}^{t-1}\theta^{t-s}x_s\right)^2}{\sigma_W^2}\right).$$

Hence, we obtain

$$\begin{aligned}
&f_{X_1,\ldots,X_T|W_0=0}\left(x_1,\ldots,x_T;\mu,\theta,\sigma_W^2\right)\\
&= f_{X_T|X_1,\ldots,X_{T-1},W_0=0}\left(x_T \mid x_1,\ldots,x_{T-1};\mu,\theta,\sigma_W^2\right) f_{X_{T-1}|X_1,\ldots,X_{T-2},W_0=0}\left(x_{T-1} \mid x_1,\ldots,x_{T-2};\mu,\theta,\sigma_W^2\right)\\
&\cdots f_{X_2|X_1,W_0=0}\left(x_2 \mid x_1;\mu,\theta,\sigma_W^2\right) f_{X_1,W_0=0}\left(x_1;\mu,\theta,\sigma_W^2\right)\\
&= \frac{1}{\sqrt{2\pi\sigma_W^2}}\exp\left(-\frac{1}{2}\frac{\left(x_T-\mu\sum_{t=0}^{T-1}\theta^t+\sum_{t=1}^{T-1}\theta^{T-t}x_s\right)^2}{\sigma_W^2}\right)\frac{1}{\sqrt{2\pi\sigma_W^2}}\exp\left(-\frac{1}{2}\frac{\left(x_{T-1}-\mu\sum_{t=0}^{T-2}\theta^t+\sum_{t=1}^{T-2}\theta^{T-t}x_s\right)^2}{\sigma_W^2}\right)\\
&\cdots\frac{1}{\sqrt{2\pi\sigma_W^2}}\exp\left(-\frac{1}{2}\frac{\left(x_2-\mu\left(1+\theta\right)+\theta x_1\right)^2}{\sigma_W^2}\right)\frac{1}{\sqrt{2\pi\sigma_W^2}}\exp\left(-\frac{1}{2}\frac{\left(x_1-\mu\right)^2}{\sigma_W^2}\right)\\
&= \left(2\pi\sigma_W^2\right)^{-T/2}\prod_{t=1}^{T}\exp\left(-\frac{1}{2}\frac{\left(x_t-\mu\sum_{s=0}^{t-1}\theta^s+\sum_{s=1}^{t-1}\theta^{t-s}x_t\right)^2}{\sigma_W^2}\right)
\end{aligned}$$

Theefore, we can consider the conditional log-likelihhod function $\log \mathcal{L}_{X_1,\dots,X_T|W_0=0} : \mathbb{R}^2 \times \mathbb{R}_+ \to \mathbb{R}$ given by

$$\begin{aligned}
&\log \mathcal{L}_{X_1,\dots,X_T|W_0=0}\left(\mu,\theta,\sigma_W^2;x_1,\dots,x_T\right)) \\
&= \log\left(\left(2\pi\sigma_W^2\right)^{-T/2}\prod_{t=1}^{T}\exp\left(\frac{1}{2}\frac{\left(x_t-\mu\sum_{s=0}^{t-1}\theta^s+\sum_{s=1}^{t-1}\theta^{t-s}x_t\right)^2}{\sigma_W^2}\right)\right) \\
&= -\frac{T}{2}\left(\log\left(2\pi\right)+\log\left(\sigma_W^2\right)\right)-\frac{1}{2\sigma_W^2}\sum_{t=1}^{T}\left(x_t-\mu\sum_{s=0}^{t-1}\theta^s+\sum_{s=1}^{t-1}\theta^{t-s}x_t\right)^2 \\
&= -\frac{T}{2}\left(\log\left(2\pi\right)+\log\left(\sigma_W^2\right)\right)-\frac{1}{2\sigma_W^2}\sum_{t=1}^{T}\left(x_t-\mu\frac{1-\theta^t}{1-\theta}+\sum_{r=1}^{t-1}\theta^{t-s}x_t\right)^2.
\end{aligned} \tag{7.183}$$

As in the case of the log-likelihood, also the maximization of the conditional log-likelihood in (7.183) has to be performed numerically.

7.8.2 Forecasting

Let $(x_t)_{t=1}^T \equiv x$ be a univariate real time series, for some $T \geqslant 2$, and let $(X_t)_{t\in\mathbb{N}} \equiv X$ be a $MA(1)$ processes with innovation $(W_t)_{t\in\hat{n}} \equiv W \sim SWN\left(\sigma_W^2\right)$, satisfying Equation (7.139), for some mean [resp. memory] coefficient $\mu \in \mathbb{R}$ [resp. $\theta \in \mathbb{R}$], thought of as parameter. Assume to have determined the values $\hat{\mu}$ and $\hat{\theta}$ of the parameters μ and θ which allow the "best fit" of the $MA(1)$ process X to the time series x. Let X_{T+S} be the state of the process X at S-steps ahead with respect to the state X_T and let $\left(\mathcal{F}_t^W\right)_{t\in\hat{n}} \equiv \mathfrak{F}^W$ be the filtration generated by the innovation process W. Write $\hat{X}_{T+S|T}$ for the forecast of the state X_{T+S} in light of the information represented by $\mathcal{F}_T^W$. We define $\hat{X}_{T+S|T}$ as the random variable which minimizes the distance from X_{T+S} in term of the mean square norm, in symbols,

$$\hat{X}_{T+S|T} = \underset{Y\in L^2\left(\Omega_{\mathcal{F}_T^W},\mathbb{R}\right)}{\arg\min}\ \mathbf{E}\left[Y-X_{T+S}\right]^2, \tag{7.184}$$

where $L^2\left(\Omega_{\mathcal{F}_T^W},\mathbb{R}\right)$ is the (Hibert) space of the random variables which are measurable with respect to the σ-algebra $\mathcal{F}_T^W$ and have finite 2nd moment. As in case of $AR(1)$ processes, Equation (7.184) implies

$$\hat{X}_{T+S|T} = \mathbf{E}\left[X_{T+S} \mid \mathcal{F}_T^W\right]. \tag{7.185}$$

In addition,

$$\hat{X}_{T+S|T} = h\left(W_0,W_1,\dots,W_T\right), \tag{7.186}$$

for the unique function $h : \mathbb{R}^T \to \mathbb{R}$, which solves the minimization problem

$$\underset{g:\mathbb{R}^T\to\mathbb{R}\text{ s.t. }g(W_1,\dots,W)\in L^2\left(\Omega_{\mathcal{F}_T^{\mathbf{W}}},\mathbb{R}\right)}{\arg\min}\ \mathbf{E}\left[g\left(W_0,W_1,\dots,W_T\right)-X_{T+S}\right]^2.$$

For reader's convenience, we recall again the important results mentioned in the context of the $AR(1)$ processes, which depend only on the properties of the conditional expectation operator on the Hilbert space of the random variables with finite moment of order 2 and are independent on the $MA(1)$ structure of the process X.

Recall 614 *We have*

$$\mathbf{E}\left[\hat{X}_{T+S|T}\right] = \mathbf{E}\left[X_{T+S}\right] \tag{7.187}$$

and

$$\mathbf{E}\left[\left(X_{T+S} - \hat{X}_{T+S|T}\right)\hat{X}_{T+S|T}\right] = 0, \tag{7.188}$$

for all $S, T \in \mathbb{N}$.

Corollary 615 *We have*

$$\mathbf{E}\left[\hat{X}^2_{T+S|T}\right] = \mathbf{E}\left[X_{T+S}\hat{X}_{T+S|T}\right] \tag{7.189}$$

and

$$\mathbf{D}^2\left[\hat{X}_{T+S|T}\right] = Cov\left(X_{T+S}, \hat{X}_{T+S|T}\right), \tag{7.190}$$

for all $S, T \in \mathbb{N}$.

Proof. Since the result does not depend on the $MA(1)$ structure, the proof is the same as the proof of Proposition 579. □

For reader's convenience, we recall that

Definizione 616 *We call* forecast error *of the forecast* $\hat{X}_{T+S|T}$ *of the state* X_{T+S} *the random variable*

$$E_{T+S|T} \overset{def}{=} X_{T+S} - \hat{X}_{T+S|T}, \quad \forall S, T \in \mathbb{N}. \tag{7.191}$$

We call mean squared error *of the forecast* $\hat{X}_{T+S|T}$ *of the state* X_{T+S}*, the positive number*

$$\mathbf{MSE}\left[E_{T+S|T}\right] \overset{def}{=} \mathbf{E}\left[E^2_{T+S|T}\right], \quad \forall S \in \mathbb{N}. \tag{7.192}$$

Osservazione 617 *We have*

$$\mathbf{E}\left[E_{T+S|T}\right] = 0, \tag{7.193}$$

for all $S, T \in \mathbb{N}$. *Hence,*

$$\mathbf{MSE}\left[E_{T+S|T}\right] = \mathbf{D}^2\left[E_{T+S|T}\right], \tag{7.194}$$

for all $S, T \in \mathbb{N}$.

Proposizione 618 *We have*

$$\mathbf{D}^2\left[E_{T+S|T}\right] = \mathbf{D}^2\left[X_{T+S}\right] - \mathbf{D}^2\left[\hat{X}_{T+S|T}\right], \tag{7.195}$$

for all $S, T \in \mathbb{N}$.

Proof. Since the result does not depend on the $MA(1)$ structure, the proof is the same as the proof of Proposition 582 . □

Now, we turn again our attention to $MA(1)$ processes.

Proposizione 619 *We have*

$$\hat{X}_{T+S|T} = \begin{cases} \mu - \theta W_T, & \text{if } S = 1, \\ \mu, & \text{if } S > 1. \end{cases} \tag{7.196}$$

As a consequence,

$$\mathbf{D}^2\left[\hat{X}_{T+S|T}\right] = \begin{cases} \theta^2\sigma^2_W, & \text{if } S = 1, \\ 0, & \text{if } S > 1. \end{cases} \tag{7.197}$$

Proof. Thanks to Equation (7.139), we can write

$$X_{T+S} = \mu + W_{T+S} - \theta W_{T+S-1}$$

Therefore, by virtue of the properties of the conditional expectation and considering that $W \sim SWN\left(\sigma_W^2\right)$, we obtain

$$\begin{aligned}\hat{X}_{T+S|T} &= \mathbf{E}\left[\mu + W_{T+S} - \theta W_{T+S-1} \mid \mathcal{F}_T^W\right] \\ &= \mu + \mathbf{E}\left[W_{T+S} \mid \mathcal{F}_T^W\right] - \theta \mathbf{E}\left[W_{T+S-1} \mid \mathcal{F}_T^W\right] \\ &= \mu + \mathbf{E}\left[W_{T+S}\right] - \theta \mathbf{E}\left[W_{T+S-1} \mid \mathcal{F}_T^W\right],\end{aligned} \tag{7.198}$$

where,

$$\mathbf{E}\left[W_{T+S-1} \mid \mathcal{F}_T^W\right] = \begin{cases} W_T & \text{if } S = 1 \\ \mathbf{E}\left[W_{T+S-1}\right] & \text{if } S > 1 \end{cases}. \tag{7.199}$$

Combining (7.198) and (7.199) the desired Equation (7.196) follows. Now, considering Equation (7.190), we have

$$\mathbf{D}^2\left[\hat{X}_{T+S|T}\right] = Cov\left(X_{T+S}, \hat{X}_{T+S|T}\right) = \begin{cases} Cov\left(\mu + W_{T+1} - \theta W_T, \mu - \theta W_T\right), & \text{if } S = 1, \\ Cov\left(\mu + W_{T+S+1} - \theta W_{T+1}, \mu\right), & \text{if } S > 1 \end{cases} \tag{7.200}$$

where,

$$Cov\left(\mu + W_{T+S} - \theta W_T, \mu - \theta W_T\right) = \theta^2 \sigma_W^2 \tag{7.201}$$

and

$$Cov\left(\mu + W_{T+S+1} - \theta W_{T+1}, \mu\right) = 0. \tag{7.202}$$

From (7.200)-(7.200), we obtain Equation (7.197). □

Corollary 620 *We have*

$$\mathbf{D}^2\left[E_{T+S|T}\right] = \begin{cases} \sigma_{\mathbf{W}}^2, & \text{if } S = 1, \\ \left(1+\theta^2\right)\sigma_{\mathbf{W}}^2, & \text{if } S > 1. \end{cases} \tag{7.203}$$

Proof. Replacing both Equations (7.146) and (7.197) into Equation (7.195), the desired result immediately follows. □

Proposizione 621 *Assume that the state innovation* $\mathbf{W}$ *is a Gaussian white noise,* $\mathbf{W} \sim GWN\left(\sigma_{\mathbf{W}}^2\right)$ *with* $\sigma_{\mathbf{W}}^2 > 0$. *Then also the forecast* $\hat{X}_{T+S|T}$ *of the state* X_{T+S} *is Gaussian for every* $S \in \mathbb{N}$. *Therefore, for any* $\alpha \in (0,1)$, *a prediction interval of confidence level* $100(1-\alpha)\%$ *for the state* X_{T+S} *is given by*

$$\begin{array}{ll}\left(\hat{X}_{T+S|T} - z_{\alpha/2}\sigma_{\mathbf{W}},\ \hat{X}_{T+S|T} + z_{\alpha/2}\sigma_{\mathbf{W}}\right), & \text{if } S = 1, \\ \left(\hat{X}_{T+S|T} - z_{\alpha/2}\sigma_{\mathbf{W}}\sqrt{1+\theta^2},\ \hat{X}_{T+S|T} + z_{\alpha/2}\sigma_{\mathbf{W}}\sqrt{1+\theta^2}\right), & \text{if } S > 1.\end{array} \tag{7.204}$$

where $z_{\alpha/2} \equiv z_{\alpha/2}^{+}$ *is the upper tail critical value of level* $\alpha/2$ *of the standard Gaussian random variable and* $\hat{X}_{T+S|T}$ *is given by (7.196). The realization of the predicton interval is then*

$$\begin{array}{ll}\left(\hat{x}_{T+S|T} - z_{\alpha/2}\sigma_{\mathbf{W},T}(\omega),\ \hat{x}_{T+S|T} + z_{\alpha/2}\sigma_{\mathbf{W},T}(\omega)\right), & \text{if } S = 1, \\ \left(\hat{x}_{T+S|T} - z_{\alpha/2}\sigma_{\mathbf{W},T}(\omega)\sqrt{1+\hat{\theta}_T^2(\omega)},\ \hat{x}_{T+S|T} + z_{\alpha/2}\sigma_{\mathbf{W},T}(\omega)\sqrt{1+\hat{\theta}_T^2(\omega)}\right), & \text{if } S > 1.\end{array} \tag{7.205}$$

where $\hat{x}_{T+S|T}$ *is the realization of the predictor* $\hat{X}_{T+S|T}$ *of the state* X_{T+S} *and* $\hat{\theta}_T(\omega)$ *[resp.* $\hat{\sigma}_{\mathbf{W},T}(\omega)$*] is the estimated value of the parameter* θ *[resp.* $\sigma_{\mathbf{W}}$*].*

7.9 Higher Order Moving Average Process - MA(q)

Let $(W_t)_{t\in\mathbb{Z}} \equiv W$ be a weak white noise on a probability space Ω with states in $\mathbb{R}^N$ and variance-covariance matrix Σ_W. Let $(X_t)_{t\in\mathbb{Z}} \equiv X$ be a stochastic process on Ω with states in $\mathbb{R}^N$.

Definizione 622 *We say that X is a* qth-order moving average process, *for some $q \in \mathbb{N}$, if the random variables in X fulfill the equation*

$$X_t = \mu + W_t - \theta_1 W_{t-1} - \theta_2 W_{t-2} + \cdots - \theta_q W_{t-q}, \quad \forall t \in \mathbb{Z},$$

where $\mu \in \mathbb{R}^N$ [resp. $\theta_1, \ldots, \theta_q \in \mathbb{R}^{N^2}$] is a parameter vector [resp. are parameter matrices] referred to as the mean *[resp.* memory weights*] of the process X.*

Notation 623 *To denote that X is a qth-order moving average process with states in $\mathbb{R}^N$ we write $X \sim MA(q)^N$. In the case $N = 1$, the reference to the dimension N will be omitted.*

In what follows, we restrain our attention to qth-order moving average processes X with states in $\mathbb{R}$, in symbols $X \sim MA(q)$, for which μ and $\theta_1, \ldots, \theta_q$ are all real parameters and is a real weak white noise with variance σ_W^2 for some $\sigma_W > 0$. Recall that in this case both the autocovariance and the autocorrelation functions of X are symmetric, that is

$$\gamma_X(s,t) = \gamma_X(t,s) \quad \text{and} \quad \rho_X(s,t) = \rho_X(t,s)$$

for all $s, t \in \mathbb{Z}$.

Proposizione 624 *The mean function $\mu_X : \mathbb{Z} \to \mathbb{R}$ of X is given by*

$$\mu_X(t) = \mu, \tag{7.206}$$

for every $t \in \mathbb{N}$.

Proof. . $\square$

Proposizione 625 *The variance function $\sigma_X^2 : \mathbb{Z} \to \mathbb{R}$ of X is given by*

$$\sigma_X^2(t) = \left(1 + \sum_{\ell=1}^{q} \theta_\ell^2\right) \sigma_W^2, \tag{7.207}$$

for every $t \in \mathbb{N}$.

Proof. . $\square$

Proposizione 626 *The autocovariance function $\gamma_X : \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}$ of X is given by*

$$\gamma_X(s,t) = \begin{cases} \left(\sum_{\ell=0}^{q-|t-s|} \theta_\ell \theta_{\ell+|t-s|}\right) \sigma_W^2 & \text{if } 0 \leqslant |t-s| \leqslant q \\ 0 & \text{if } |t-s| > q \end{cases} \tag{7.208}$$

where $\theta_0 \equiv 1$.

Proof. Having set $\theta_0 \equiv 1$, when $|t-s| = 0$ Equation (7.208) is just a reformulation of Equation (7.207). Now, assuming $s < t$, we can write

$$\begin{aligned}
\gamma_X(s,t) &= \mathbf{E}\left[(X_s - \mathbf{E}[X_s])(X_t - \mathbf{E}[X_t])\right] \\
&= \mathbf{E}\left[(\theta_0 W_s + \theta_1 W_{s-1} + \cdots + \theta_{q-1} W_{s-q+1} + \theta_q W_{s-q})(\theta_0 W_t + \theta_1 W_{t-1} + \cdots + \theta_{q-1} W_{t-q+1} + \theta_q W_{t-q})\right] \\
&= \theta_0^2 \mathbf{E}[W_s W_t] + \theta_0\theta_1 \mathbf{E}[W_s W_{t-1}] + \cdots + \theta_0\theta_{q-1}\mathbf{E}[W_s W_{t-q+1}] + \theta_0\theta_q \mathbf{E}[W_s W_{t-q}] \\
&+ \theta_1\theta_0 \mathbf{E}[W_{s-1} W_t] + \theta_1^2 \mathbf{E}[W_{s-1} W_{t-1}] + \cdots + \theta_1\theta_{q-1}\mathbf{E}[W_{s-1} W_{t-q+1}] + \theta_1\theta_q \mathbf{E}[W_{s-1} W_{t-q}] \\
&+ \cdots + \theta_{q-1}\theta_0 \mathbf{E}[W_{s-q+1} W_t] + \theta_{q-1}\theta_1 \mathbf{E}[W_{s-q+1} W_{t-1}] + \cdots + \theta_{q-1}^2\mathbf{E}[W_{s-q+1} W_{t-q+1}] + \theta_q^2 \mathbf{E}[W_{s-q} W_{t-q}] \\
&+ \theta_q\theta_0 \mathbf{E}[W_{s-q} W_t] + \theta_q\theta_1 \mathbf{E}[W_{s-q} W_{t-1}] + \cdots + \theta_q\theta_{q-1}\mathbf{E}[W_{s-q} W_{t-q+1}] + \theta_q^2 \mathbf{E}[W_{s-q} W_{t-q}]
\end{aligned} \tag{7.209}$$

Writing $s = t - r$, Equation (7.209) becomes.

$$\begin{aligned}
\gamma_X(s,t) &= \theta_0^2 \mathbf{E}[W_{t-r} W_t] + \theta_0\theta_1 \mathbf{E}[W_{t-r} W_{t-1}] + \cdots + \theta_0\theta_{q-1}\mathbf{E}[W_{t-r} W_{t-q+1}] + \theta_0\theta_q \mathbf{E}[W_{t-r} W_{t-q}] \\
&+ \theta_1\theta_0 \mathbf{E}[W_{t-r-1} W_t] + \theta_1^2 \mathbf{E}[W_{t-r-1} W_{t-1}] + \cdots + \theta_1\theta_{q-1}\mathbf{E}[W_{t-r-1} W_{t-q+1}] + \theta_1\theta_q \mathbf{E}[W_{t-r-1} W_{t-q}] \\
&+ \cdots + \theta_{q-1}\theta_0 \mathbf{E}[W_{t-r-q+1} W_t] + \theta_{q-1}\theta_1 \mathbf{E}[W_{t-r-q+1} W_{t-1}] + \cdots + \theta_{q-1}^2\mathbf{E}[W_{t-r-q+1} W_{t-q+1}] + \theta_{q-1}\theta_q \mathbf{E}[W \\
&+ \theta_q\theta_0 \mathbf{E}[W_{t-r-q} W_t] + \theta_q\theta_1 \mathbf{E}[W_{t-r-q} W_{t-1}] + \cdots + \theta_q\theta_{q-1}\mathbf{E}[W_{t-r-q} W_{t-q+1}] + \theta_q^2 \mathbf{E}[W_{t-r-q} W_{t-q}].
\end{aligned}$$

Therefore,

$$\gamma_X(s,t) = \begin{cases} \theta_0\theta_1 + \theta_1\theta_2 + \cdots + \theta_{q-1}\theta_q & \text{if } r = 1 \\ \theta_0\theta_2 + \theta_1\theta_3 + \cdots + \theta_{q-2}\theta_q & \text{if } r = 2 \\ \cdots & \cdots \\ \theta_0\theta_{q-1} + \theta_1\theta_q & \text{if } r = q-1 \\ \theta_0\theta_q & \text{if } r = q \\ 0 & \text{if } r > q \end{cases} \tag{7.210}$$

For $r \leqslant q$, Equation (7.210) can be more concisely rewritten as

$$\gamma_X(s,t) = \sum_{\ell=0}^{q-r} \theta_\ell \theta_{\ell+r}, \quad r = 1, \ldots, q.$$

From the latter, on account of the symmetry property of the the autocovariance function, the desired (7.208) folows. □

Proposizione 627 *The autocorrelation function $\rho_X : \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}$ of X is given by*

$$\rho_X(s,t) = \begin{cases} \dfrac{\sum_{\ell=0}^{q-|t-s|} \theta_\ell \theta_{\ell+|t-s|}}{\sum_{\ell=0}^{q} \theta_\ell^2} & \text{if } 0 \leqslant |t-s| \leqslant q \\ 0 & \text{if } |t-s| > q \end{cases}$$

where $\theta_0 \equiv 1$.

Proof. . □

Osservazione 628 *For all $s, t \in \mathbb{Z}$ such that $|t-s| > q$ the random variables X_s and X_t in X are uncorrelated. Moreover, if W is a strong white process X_s and X_t are independent.*

Osservazione 629 *Choosing $t_0 \equiv 0$ the reduced autocorrelation function $\rho_{X,0} : \mathbb{Z} \to \mathbb{R}$ of X is given by*

$$\rho_{X,0}(h) = \begin{cases} \dfrac{\sum_{\ell=0}^{q-|t-s|} \theta_\ell \theta_{\ell+|t-s|}}{\sum_{\ell=0}^{q} \theta_\ell^2} & \text{if } 0 \leqslant |h| \leqslant q \\ 0 & \text{if } |h| > q \end{cases}$$

where $\theta_0 \equiv 1$.

7.10 Infinite Order Moving Average Process - MA(∞)

Definizione 630 *We say that X is an* infinite-order moving average process *if the random variables in X fulfill the equation,*

$$X_t = \mu + \textstyle\sum_{n=0}^{\infty} \theta_n W_{t-n}, \quad \forall t \geqslant 1,$$

where $\mu \in \mathbb{R}^m$ [resp. $\theta_1, \ldots, \theta_n, \cdots \in \mathbb{R}^{N^2}$] is a parameter vector [resp. are parameter matrices such that $\sum_{n=0}^{\infty} \|\theta_n\|^2 < \infty$] referred to as the mean *[resp.* memory weights*] of the process X.*

Notation 631 *To denote that X is an infinite-order moving average process with states in $\mathbb{R}^m$ we write $X \sim MA(\infty)^N$. In the case $N = 1$, the reference to the dimension N will be omitted.*

Osservazione 632 *It is common to consider the stronger condition of* absolute summability *of the matrices $\theta_n \equiv \left(\theta_{i,j}^{(n)}\right)_{i,j=1}^{N}$, for every $n \in \mathbb{N}$, that is*

$$\textstyle\sum_{n=0}^{\infty} \left|\theta_{i,j}^{(n)}\right| < \infty, \quad \forall i, j = 1, \ldots, N,$$

rather than the (weaker) condition $\sum_{n=0}^{\infty} \|\theta_n\|^2 < \infty$.

7.10.1 The Wold Representation Theorem

Let $(X_t)_{t\in\mathbb{Z}} \equiv X$ be a weakly stationary real stochastic process on a probability space Ω with time set $\mathbb{T} \equiv \mathbb{Z}$.

Teorema 633 (Wold Representation Theorem) *Assume that the process X has mean zero, that is $\mu_X(t) = 0$ for every $t \in \mathbb{Z}$. Then we can write*

$$X_t = \sum_{n=0}^{\infty} \theta_n W_{t-n} + D_t, \quad \forall t \in \mathbb{Z}, \tag{7.211}$$

where $(W_t)_{t\in\mathbb{Z}} \equiv W$ is a weak white noise, $(\theta_n)_{n\geqslant 0}$ is a real sequence such that $\theta_0 = 1$ and $\sum\limits_{n=0}^{\infty} \theta_n^2 < \infty$, and $(D_t)_{t\in\mathbb{Z}} \equiv D$ is a second order real stocastic process such that

$$\mathbf{E}[D_t W_{t-n}] = 0 \quad \text{and} \quad \mathbf{P}[D_{t+n} \mid X_{t-1}, X_{t-2}, \ldots] = D_{t+n} \tag{7.212}$$

for every $t \in \mathbb{Z}$ and $n \in \hat{n}$.

Proof. . □

Osservazione 634 *The processes W and D in the Wold Representation Theorem are uncorrelated. That is*

$$\mathbf{E}[D_t W_s] = 0,$$

foe all $s, t \in \mathbb{Z}$.

Definizione 635 *The process D in the Wold Representation Theorem is referred to as the* linearly deterministic component *of X. When $D_t = 0$ for every $t \in \mathbb{Z}$, then the process X is said to be* purely non-deterministic*.*

It should be noted that from Equation (7.211) it follows

$$X_t - \left(\sum_{n=1}^{\infty} \theta_n W_{t-n} + \mathbf{P}[D_t \mid X_{t-1}, X_{t-2}, \ldots]\right) = W_t. \tag{7.213}$$

In words, the error in forecasting any weakly stationary real stochastic process (with mean zero) on the basis of a linear combination of the past errors and a linear combination of the past values of the process is a random walk. This is a rather profound result.

7.10.2 Parameter Estimation

Conditional Non Linear Least Square (CNLLS) Estimates

A similar appoorach works for higher-order $MA(q)$ models, except we assume $W_0 = W_{-1} = \cdots = W_q = 0$ and the numerical search is for the vector $\theta_1, \ldots, \theta_q$.

7.10.3 Partial Autocorrelation Function

Let $(X_t)_{t\in\mathbb{Z}} \equiv X$ be a weakly stationary real stochastic process on a probability space Ω with time set $\mathbb{T} \equiv \mathbb{Z}$.

7.11 ARMA(p,q) Processes

In ARMA modeling the data must be transformed to stationary form prior to analysis. If the data are trending, then some form of trend removal is required.

7.12 ARIMA(p,d,q) Processes

Let $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ be a complete probability space, let $(W_n)_{n\in\mathbb{Z}}$ be a weak white noise on Ω with variance σ^2, let $p, d, q \in \hat{n}$, and let $x_0, x_{-1}, \ldots, x_{1-p} \in \mathbb{R}$.

Esempio 636 (Autoregressive Moving Average Processes) *Let $(Y_t)_{t\in\mathbb{N}} \equiv Y$ be a real* autoregressive moving average process *of orders p and q, briefly $ARMA(p,q)$, for some $p, q \in \mathbb{N}$, with no linear trend and Gaussian innovation. To fix ideas, assume that $q > p$. We know that Y satisfies the following equations*

$$\begin{aligned}
&Y_1 = \alpha + \phi_1 Y_0 + U_1 - \theta_1 U_0, \\
&Y_2 = \alpha + \phi_1 Y_1 + \phi_2 Y_0 + U_2 - \theta_1 U_1 - \theta_2 U_0, \\
&\cdots \\
&Y_{p-1} = \alpha + \phi_1 Y_{p-2} + \phi_2 Y_{p-3} + \cdots + \phi_{p-2} Y_1 + \phi_{p-1} Y_0 + U_{p-1} \\
&\quad -\theta_1 U_{p-2} - \theta_2 U_{p-3} - \cdots - \theta_{p-2} U_1 - \theta_{p-1} U_0, \\
&Y_p = \alpha + \phi_1 Y_{p-1} + \phi_2 Y_{p-2} + \cdots + \phi_{p-1} Y_1 + \phi_p Y_0 + U_p \\
&\quad -\theta_1 U_{p-1} - \theta_2 U_{p-2} - \cdots - \theta_{p-1} U_1 - \theta_p U_0, \\
&Y_{p+1} = \alpha + \phi_1 Y_p + \phi_2 Y_{p-1} + \cdots + \phi_{p-1} Y_2 + \phi_p Y_1 + U_{p+1} \\
&-\theta_1 U_p - \theta_2 U_{p-1} - \cdots - \theta_p U_1 - \theta_{p+1} U_0, \\
&\cdots \\
&Y_{q-1} = \alpha + \phi_1 Y_{q-2} + \phi_2 Y_{q-3} + \cdots + \phi_{p-1} Y_{q-p} + \phi_p Y_{q-p-1} + U_{q-1} \\
&\quad -\theta_1 U_{q-2} - \theta_2 U_{q-3} - \cdots - \theta_{q-2} U_1 - \theta_{q-1} U_0, \\
&Y_t = \alpha + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_{p-1} Y_{t-p+1} + \phi_p Y_{t-p} + U_t \\
&\quad -\theta_1 U_{t-1} - \theta_2 U_{t-2} - \cdots - \theta_{q-1} U_{t-q+1} - \theta_q U_{t-q}, \quad \forall t \geqslant q,
\end{aligned} \tag{7.214}$$

where the real parameter α is the drift *of Y, the real parameters $\phi_1, \ldots, \phi_p$ [resp. $\theta_1, \ldots, \theta_q$] are the* regression coefficients *[resp.* memory weights*], and $(U_t)_{t\in\hat{n}} \equiv U$ is a* Gaussian real white noise.

Definizione 637 *We call a sequence $(X_n)_{n\geqslant 1-p} \equiv X$ of real random variables on Ω an* autoregressive integrated moving average *process of orders p, d, q with* drift*, acronym* ARIMA(p, d, q)*, if X fulfills the equation*

$$X_{1-p} = x_{1-p}, \ldots, X_0 = x_0, \qquad \phi^{(p)}(L)(1-L)^d X_n = \alpha + \theta^{(q)}(L) W_n, \quad \forall n \geqslant 1$$

where $\phi^{(p)}(L)$ [resp $\theta^{(q)}(L)$] is the standard p-order AR [q-order MA] polynomial in the lag operator L and the parameter $\alpha \in \mathbb{R}$ is the drift coefficient.

Esempio 638 *An* $ARIMA\left(0,1,0\right)$ *with zero drift process fulfills the equation*

$$X_0 = x_0, \qquad X_n - X_{n-1} = W_n, \quad \forall n \geqslant 1.$$

This can be clearly rewritten as

$$X_0 = x_0, \qquad X_n = X_{n-1} + W_n, \quad \forall n \geqslant 1,$$

which is the equation of a simple random walk, the most basic $I\left(1\right)$ *process.*

Esempio 639 *An* $ARIMA\left(1,0,0\right)$ *with zero drift fulfills the equation*

$$X_0 = x_0, \qquad X_n - \phi_1 X_{n-1} = W_n, \quad \forall n \geqslant 1.$$

This can be clearly rewritten as

$$X_0 = x_0, \qquad X_n = \phi_1 X_{n-1} + W_n, \quad \forall n \geqslant 1,$$

which is the equation of a simple $AR(1)$ *process.*

Esempio 640 *An* $ARIMA\left(p,0,0\right)$ *with zero drift fulfills the equation*

$$X_{1-p} = x_{1-p}, \ldots, X_0 = x_0, \qquad X_n - \phi_1 X_{n-1} - \cdots - \phi_p X_{n-p} = W_n, \quad \forall n \geqslant 1.$$

This can be clearly rewritten as

$$X_{1-p} = x_{1-p}, \ldots, X_0 = x_0, \qquad X_n = \phi_1 X_{n-1} + \cdots + \phi_p X_{n-p} + W_n, \quad \forall n \geqslant 1,$$

which is the equation of a simple $AR(p)$ *process.*

Esempio 641 *An* $ARIMA\left(0,0,1\right)$ *fulfills the equation*

$$X_0 = x_0, \qquad X_n = W_n + \theta_1 W_{n-1}, \quad \forall n \geqslant 1,$$

which is the equation of a simple $MA(1)$ *process.*

Esempio 642 *An* $ARIMA\left(p,1,1\right)$ *with drift fulfills the equation*

$$X_{1-p} = x_{1-p}, \ldots, X_0 = x_0, \qquad \phi^{(p)}\left(L\right)\left(1-L\right)X_n = \beta + \theta^{(1)}\left(L\right)W_n, \quad \forall n \geqslant 1,$$

that is

$$X_{1-p} = x_{1-p}, \ldots, X_0 = x_0, \qquad (X_n - X_{n-1}) - \phi_1\left(X_{n-1} - X_{n-2}\right) - \cdots - \phi_p\left(X_{n-p} - X_{n-p-1}\right) = \beta + W_n + \theta_1 W_{n-1},$$

PIPPAZZo

$$X_0 = x_0, \qquad X_n = W_n + \theta_1 W_{n-1}, \quad \forall n \geqslant 1,$$

$$X_n = \mu + \beta n + \phi X_{n-1} + W_n$$

$$X_{n-1} = \mu + \beta\left(n-1\right) + \phi X_{n-2} + W_{n-1}$$

$$X_n - X_{n-1} = \beta + \phi\left(X_{n-1} - X_{n-2}\right) + W_n - W_{n-1}$$

we end up with $ARIMA\left(1,1,1\right)$

7.13 Linear Processes

7.14 Autoregressive Conditional Heteroskedasticity Processes - ARCH Processes

A frequently observed phenomenon related to financial, and more generally economical, time series is *volatility clustering*: a volatile (high variance) period in the states of a financial time series tends to be followed by another volatile period; otherwise saying volatile periods are often clustered. Intuitively, the market becomes more volatile when important news breaks in and it takes some time for the market to fully digest it. Volatility clustering in the states of a time series implies time-varying conditional variance of the states. In 1982, Robert Engle developed the autoregressive conditional heteroskedasticity ($ARCH$) models to deal with this type of time-varying volatility. For this contribution, he won the 2003 Nobel Prize in Economics[5]. In $ARCH$ processes the variance of the current state is a function of the variance of the previous error states. In particular, the variance of the current state is often related to the squares of the former states. This allows $ARCH$ processes to have the property of *time-varying conditional variance* while retaining the property of *zero conditional mean*. Therefore, although remaining unpredictable in the states, which is a desirable property for several financial time series, $ARCH$ processes can model the volatility clustering.

Another important phenomenon in financial time series is *fat tail*: the distribution of the states of a financial time series shows a higher kurtosis than the kurtosis of the Gaussian distribution. $ARCH$ processes can also model this phenomenon, even under the assumption that the associated innovation process is Gaussian distributed.

7.14.1 Autoregressive Conditional Heteroskedasticity Process of the 1st Order - $ARCH(1)$ Process.

Let $(X_t)_{t\in\hat{n}} \equiv X$ be a real stochastic process on a probability space Ω and let $(W_t)_{t\in\mathbb{N}} \equiv W$ be a real strong white noise with variance σ_W^2, for some $\sigma_W > 0$, in symbols $W \sim SWN\left(\sigma_W^2\right)$. We assume that the random variable X_0 has finite 2nd moment and it is independent of the random variables in W. We also assume $\mathbf{E}[X_0] = 0$ and set $\sigma_{X_0}^2 \equiv \mathbf{D}^2[X_0]$.

Definizione 643 *We say that X is a $1st$-*order autoregressive conditional heteroskedasticity *process with innovation W, if the random variables in X are given by*

$$X_t \stackrel{def}{=} \sigma_t W_t, \quad \forall t \in \mathbb{N}, \tag{7.215}$$

where $(\sigma_t)_{t\in\mathbb{N}}$ is the positive process on Ω given by

$$\sigma_t^2 \stackrel{def}{=} \alpha_0 + \alpha_1 X_{t-1}^2, \quad \forall t \in \mathbb{N}, \tag{7.216}$$

for parameters $\alpha_0, \alpha_1 \in \mathbb{R}_+$ such that $\alpha_0 > 0$ and $\alpha_1 > 0$.

7.14.2 Autoregressive Conditional Heteroskedasticity Process of the qth Order - $ARCH(1)$ Process.

Let $(X_t)_{t\in\hat{n}} \equiv X$ be a real stochastic process on a probability space Ω and let $(W_t)_{t\in\mathbb{N}} \equiv W$ be a real strong white noise with variance σ_W^2, for some $\sigma_W > 0$, in symbols $W \sim SWN\left(\sigma_W^2\right)$. We assume that the random variable X_0 has finite 2nd moment and it is independent of the random variables in W. We also assume $\mathbf{E}[X_0] = 0$ and set $\sigma_{X_0}^2 \equiv \mathbf{D}^2[X_0]$.

[5] Clive Granger shared the prize for co-integration (see `http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2003/press.html`)

Definizione 644 *We say that X is a qth-*order univariate autoregressive conditional heteroskedasticity *process or an* univariate autoregressive conditional heteroskedasticity *process of* order q, *for some $q \in \mathbb{N}$, with innovation W, if*

$$X_t \overset{def}{=} \sigma_t W_t, \quad \forall t \in \mathbb{N}, \tag{7.217}$$

where $(\sigma_t)_{t\in\mathbb{N}}$ is the positive process on Ω given by

$$\sigma_t^2 \overset{def}{=} \alpha_0 + \sum_{s=1,\dots,q \ s.t.\ t-s\geqslant 0} \alpha_s X_{t-s}^2, \quad \forall t \in \mathbb{N}, \tag{7.218}$$

for parameters $\alpha_0, \alpha_1, \dots, \alpha_q \in \mathbb{R}_+$ such that $\alpha_0 > 0$ and $\alpha_q > 0$.

Notation 645 *To denote that X is a* univariate autoregressive conditional heteroskedasticity *process of order q we write $X \sim ARCH(q)$.*

Osservazione 646 *As an immediate consequence of Equations (7.217) and (7.218), we have*

$$\begin{aligned}
&X_1 \overset{def}{=} \sigma_1 W_1, \quad \sigma_1^2 \overset{def}{=} \alpha_0 + \alpha_1 X_0^2 \\
&X_2 \overset{def}{=} \sigma_2 W_2, \quad \sigma_2^2 \overset{def}{=} \alpha_0 + \alpha_1 X_1^2 + \alpha_2 X_0^2 \\
&\vdots \\
&X_t \overset{def}{=} \sigma_t W_t, \quad \sigma_t^2 \overset{def}{=} \alpha_0 + \alpha_1 X_{t-1}^2 + \alpha_2 X_{t-2}^2 + \cdots + \alpha_t X_0^2, \quad \forall t < q \\
&\vdots \\
&X_q \overset{def}{=} \sigma_q W_q, \quad \sigma_q^2 \overset{def}{=} \alpha_0 + \alpha_1 X_{q-1}^2 + \alpha_2 X_{q-2}^2 + \cdots + \alpha_q X_0^2 \\
&\vdots \\
&X_t \overset{def}{=} \sigma_t W_t, \quad \sigma_t^2 \overset{def}{=} \alpha_0 + \alpha_1 X_{t-1}^2 + \alpha_2 X_{t-2}^2 + \cdots + \alpha_q X_{t-q}^2, \quad \forall t > q
\end{aligned}$$

Let $(X_t)_{t\in\mathbb{Z}} \equiv X$ be an $ARCH(q)$ process with innovation $(W_t)_{t\in\mathbb{N}} \equiv W$.

Osservazione 647 *In particular, we have.*

$$\begin{aligned}
&\sigma_1^2 \overset{def}{=} \alpha_0 + \alpha_1 X_0^2, \quad \forall q \in \mathbb{N}. \\
&\sigma_2^2 \overset{def}{=} \begin{cases} \alpha_0 + \alpha_1 X_1^2, & \text{if } q = 1, \\ \alpha_0 + \alpha_1 X_1^2 + \alpha_2 X_0^2, & \text{if } q \geqslant 2. \end{cases} \\
&\sigma_3^2 \overset{def}{=} \begin{cases} \alpha_0 + \alpha_1 X_2^2, & \text{if } q = 1, \\ \alpha_0 + \alpha_1 X_2^2 + \alpha_2 X_1^2, & \text{if } q = 2, \\ \alpha_0 + \alpha_1 X_2^2 + \alpha_2 X_1^2 + \alpha_3 X_0^2, & \text{if } q \geqslant 3. \end{cases} \\
&\sigma_4^2 \overset{def}{=} \begin{cases} \alpha_0 + \alpha_1 X_3^2, & \text{if } q = 1, \\ \alpha_0 + \alpha_1 X_3^2 + \alpha_2 X_2^2, & \text{if } q = 2, \\ \alpha_0 + \alpha_1 X_3^2 + \alpha_2 X_2^2 + \alpha_3 X_1^2, & \text{if } q = 3, \\ \alpha_0 + \alpha_1 X_3^2 + \alpha_2 X_2^2 + \alpha_3 X_1^2 + \alpha_4 X_0^2, & \text{if } q \geqslant 4. \end{cases} \\
&\vdots
\end{aligned}$$

Lemma 648 *Fixed any $q \in \mathbb{N}$, we have*

$$\sigma_1^2 = P_1^{(q)}(X_0), \tag{7.219}$$

where $P_1^{(q)}:\mathbb{R}\to\mathbb{R}$ *is the polynomial given by*

$$P_1^{(q)}(x) \overset{def}{=} \alpha_0+\alpha_1 x^2, \quad \forall x\in\mathbb{R}. \tag{7.220}$$

Furthermore, we can write

$$\sigma_t^2 = P_t^{(q)}(X_0, W_1, \ldots, W_{t-1}), \tag{7.221}$$

for every $t\geqslant 2$*, where* $P_t^{(q)}:\mathbb{R}^t\to\mathbb{R}$ *is the polynomial given recursively by*

$$\begin{aligned}
&P_2^{(q)}(x,w_1) \overset{def}{=} \alpha_0+\alpha_1 P_1^{(q)}(x)\,w_1^2+\alpha_2 x^2, \quad \forall (x,w_1)\in\mathbb{R}^2, \quad q\geqslant 2,\\
&P_3^{(q)}(x,w_1,w_2) \overset{def}{=} \alpha_0+\alpha_1 P_2^{(q)}(x,w_1)\,w_2^2+\alpha_2 P_1^{(q)}(x)\,w_1^2+\alpha_3 x^2, \quad \forall (x,w_1,w_2)\in\mathbb{R}^3, \quad q\geqslant 3,\\
&\ldots\\
&P_{q-1}^{(q)}(x,w_1,w_2,\ldots,w_{q-2}) \overset{def}{=} \alpha_0+\alpha_1 P_{q-2}^{(q)}(x,w_1,\ldots,w_{q-3})\,w_{q-2}^2+\alpha_2 P_{q-3}^{(q)}(x,w_1,\ldots,w_{q-4})\,w_{q-3}^2+\ldots\\
&\qquad +\tilde{\alpha}_{q-2}P_1^{(q)}(x)\,w_1^2+\tilde{\alpha}_{q-1}x^2, \quad \forall (x,w_1,w_2,\ldots,w_{q-2})\in\mathbb{R}^q, \quad q\geqslant 4,\\
&P_q^{(q)}(x,w_1,w_2,\ldots,w_{q-1}) \overset{def}{=} \alpha_0+\alpha_1 P_{q-1}^{(q)}(x,w_1,\ldots,w_{q-2})\,w_{q-1}^2+\alpha_2 P_{q-2}^{(q)}(x,w_1,\ldots,w_{q-3})\,w_{q-2}^2+\ldots\\
&\qquad +\tilde{\alpha}_{q-1}P_1^{(q)}(x)\,w_1^2+\tilde{\alpha}_q x^2, \quad \forall (x,w_1,w_2,\ldots,w_{q-1})\in\mathbb{R}^q, \quad q\geqslant 5,\\
&P_t^{(q)}(x,w_1,w_2,\ldots,w_{t-1}) \overset{def}{=} \alpha_0+\alpha_1 P_{t-1}^{(q)}(x,w_1,\ldots,w_{t-2})\,w_{t-1}^2+\tilde{\alpha}_2 P_{t-2}^{(q)}(x,w_1,\ldots,w_{t-3})\,w_{t-2}^2+\ldots\\
&\qquad +\tilde{\alpha}_{q-1}P_{t-(q-1)}^{(q)}(x,w_1,\ldots,w_{t-q})\,w_{t-(q-1)}^2+\tilde{\alpha}_q P_{t-q}^{(q)}\left(x,w_1,\ldots,w_{t-(q+1)}\right)w_{t-q}^2,\\
&\qquad \forall (x,w_1,w_2,\ldots,w_{t-1})\in\mathbb{R}^t, \quad t>q,
\end{aligned} \tag{7.222}$$

having set

$$\tilde{\alpha}_j \equiv \begin{cases} 0, & \text{if } j\leqslant 0 \text{ or } j>q,\\ \alpha_j, & \text{if } j>0 \text{ and } j\leqslant q,\end{cases} \tag{7.223}$$

$$P_s^{(q)}(x,w_1,\ldots,w_k) \equiv \begin{cases} P_s^{(q)}(x), & \text{if } k=0,\\ 0, & \text{if } k<0.\end{cases} \quad \forall s\in\mathbb{N}. \tag{7.224}$$

and agreeing that duplicate terms in each sum on the right hand side of Equation (7.222) have to be ignored.

Proof. Consider the case $q=1$. Then, according to Definition 644, we clearly have

$$\sigma_1^2 = P_1^{(1)}(X_0),$$

where $P_1^{(1)}:\mathbb{R}\to\mathbb{R}$ is the polynomial given by

$$P_1^{(1)}(x) \overset{\text{def}}{=} \alpha_0+\alpha_1 x^2, \quad \forall x\in\mathbb{R}. \tag{7.225}$$

Furthermore,

$$\sigma_2^2 = \alpha_0+\alpha_1 X_1^2 = \alpha_0+\alpha_1\sigma_1^2 W_1^2 = \alpha_0+\alpha_1 P_1^{(1)}(X_0)\,W_1^2.$$

Hence,

$$\sigma_2^2 = P_2^{(1)}(X_0, W_1)$$

where $P_2^{(1)}:\mathbb{R}^2\to\mathbb{R}$ is the polynomial given by

$$P_2^{(1)}(x,w_1) \overset{\text{def}}{=} \alpha_0+\alpha_1 P_1(x)\,w_1^2 \quad \forall (x,w_1)\in\mathbb{R}^2. \tag{7.226}$$

Similarly,

$$\sigma_3^2 = \alpha_0+\alpha_1 X_2^2 = \alpha_0+\alpha_1\sigma_2^2 W_2^2 = \alpha_0+\alpha_1 P_2^{(1)}(X_0,W_1)\,W_2^2.$$

Hence,
$$\sigma_3^2 = P_3^{(1)}\left(X_0, W_1, W_2\right),$$
where $P_3^{(1)} : \mathbb{R}^3 \to \mathbb{R}$ is the polynomial given by
$$P_3^{(1)}\left(x, w_1, w_3\right) \stackrel{\text{def}}{=} \alpha_0 + \alpha_1 P_2^{(1)}\left(x, w_1\right) w_2^2. \tag{7.227}$$
Applying the Induction Principle, it is then possible to prove that for every $t \in \mathbb{N}$, such that $t \geqslant 3$, we have
$$\sigma_t^2 = P_t^{(1)}\left(X_0, W_1, \ldots, W_{t-1}\right)$$
where $P_t^{(1)} : \mathbb{R}^t \to \mathbb{R}$ is the polynomial given by
$$P_t^{(1)}\left(x, w_1, \ldots, w_{t-1}\right) \stackrel{\text{def}}{=} \alpha_0 + \alpha_1 P_{t-1}^{(1)}\left(x, w_1, \ldots, w_{t-2}\right) w_{t-1}^2 \quad \forall \left(x, w_1, \ldots, w_{t-2}\right) \in \mathbb{R}^{t-1}, \tag{7.228}$$
and we keep (7.224). Now, setting $q = 1$, Equations (7.225)-(7.227) can be obtained by Equation (7.219) and the first formula in Equation (7.222). In addition, under (7.223) and (7.224), Equation (7.228) can be obtained from the last formula in Equation (7.222). In fact, with reference to the last formula in Equation (7.222), for any $t > q = 1$ we have
$$\tilde{\alpha}_2 = \cdots = \tilde{\alpha}_{q-1} = 0$$
and
$$\tilde{\alpha}_q P_{t-q}^{(q)}\left(x, w_1, \ldots, w_{t-(q+1)}\right) w_{t-q}^2 \equiv \alpha_1 P_{t-1}^{(q)}\left(x, w_1, \ldots, w_{t-2}\right) w_{t-1}^2,$$
which duplicates the second term in the sum and has to be ignored. Hence, the last formula in Equation (7.222) just replicates Equation (7.228).

Consider the case $q = 2$. Then, according to Definition 644, we clearly have
$$\sigma_1^2 = P_1^{(2)}\left(X_0\right), \tag{7.229}$$
where $P_1^{(2)} : \mathbb{R} \to \mathbb{R}$ is the polynomial still given by
$$P_1^{(2)}(x) \stackrel{\text{def}}{=} \alpha_0 + \alpha_1 x^2, \quad \forall x \in \mathbb{R},$$
but
$$\sigma_2^2 = \alpha_0 + \alpha_1 X_1^2 + \alpha_2 X_0^2 = \alpha_0 + \alpha_1 \sigma_1^2 W_1^2 + \alpha_2 X_0^2 = \alpha_0 + \alpha_1 P_1^{(2)}\left(X_0\right) W_1^2 + \alpha_2 X_0^2.$$
Hence,
$$\sigma_2^2 = P_2^{(2)}\left(X_0, W_1\right),$$
where $P_1^{(2)} : \mathbb{R}^2 \to \mathbb{R}$ is the polynomial given by
$$P_2^{(2)}\left(x, w_1\right) \stackrel{\text{def}}{=} \alpha_0 + \alpha_1 P_1^{(2)}(x) w_1^2 + \alpha_2 x^2, \quad \forall \left(x, w_1\right) \in \mathbb{R}^2. \tag{7.230}$$
Similarly,
$$\sigma_3^2 = \alpha_0 + \alpha_1 X_2^2 + \alpha_2 X_1^2 = \alpha_0 + \alpha_1 \sigma_2^2 W_2^2 + \sigma_1^2 W_1^2 = \alpha_0 + \alpha_1 P_2^{(2)}\left(X_0, W_1\right) W_2^2 + P_1^{(2)}\left(X_0\right) W_1^2.$$
Hence,
$$\sigma_3^2 = P_3^{(2)}\left(X_0, W_1, W_2\right),$$
where $P_3^{(2)} : \mathbb{R}^3 \to \mathbb{R}$ is the polynomial given by
$$P_3^{(2)}\left(x, w_1, w_2\right) \stackrel{\text{def}}{=} \alpha_0 + \alpha_1 P_2^{(2)}\left(x, w_1\right) w_2^2 + \alpha_2 P_1^{(2)}(x) w_1^2, \quad \forall \left(x, w_1, w_2\right) \in \mathbb{R}^3. \tag{7.231}$$

One more,

$$\sigma_4^2 = \alpha_0+\alpha_1 X_3^2+\alpha_2 X_2^2 = \alpha_0+\alpha_1\sigma_3^2 W_3^2+\alpha_2\sigma_2^2 W_2^2 = \alpha_0+\alpha_1 P_3^{(2)}\left(X_0,W_1,W_2\right)W_3^2+\alpha_2 P_2^{(2)}\left(X_0,W_1\right)W_2^2.$$

Hence,

$$\sigma_4^2 = P_4^{(2)}\left(X_0,W_1,W_2,W_3\right),$$

where $P_4 : \mathbb{R}^4 \to \mathbb{R}$ is the polynomial given by

$$P_4^{(2)}\left(x,w_1,w_2,w_3\right) \stackrel{\text{def}}{=} \alpha_0 + \alpha_1 P_3^{(2)}\left(x,w_1,w_2\right)w_3^2 + \alpha_2 P_2^{(2)}\left(x,w_1\right)w_2^2, \quad \forall\left(x,w_1,w_2,w_3\right) \in \mathbb{R}^3. \tag{7.232}$$

Applying the Induction Principle, it is then possible to prove that for every $t \in \mathbb{N}$, such that $t \geqslant 4$, we have

$$\sigma_t^2 = P_t^{(2)}\left(X_0,W_1,\dots,W_{t-1}\right)$$

where $P_t^{(2)} : \mathbb{R}^t \to \mathbb{R}$ is the polynomial given by

$$\begin{aligned}P_t^{(2)}\left(x,w_1,\dots,w_{t-1}\right) \stackrel{\text{def}}{=}& \alpha_0 + \alpha_1 P_{t-1}^{(2)}\left(x,w_1,\dots,w_{t-2}\right)w_{t-1}^2 \\ &+\alpha_2 P_{t-2}^{(2)}\left(x,w_1,\dots,w_{t-3}\right)w_{t-2}^2, \quad \forall\left(x,w_1,\dots,w_{t-2},w_{t-1}\right) \in \mathbb{R}^t,\end{aligned} \tag{7.233}$$

and we keep (7.224). Now, setting $q = 2$ Equations (7.229)-(7.230) can be obtained by Equation (7.219) and the second formula in Equation (7.222). In addition, under (7.223) and (7.224), Equation (7.228) can be obtained from the last formula in Equation (7.222). In fact, with reference to the last formula in Equation (7.222), for any $t > q = 2$ we have

$$\tilde{\alpha}_2 = \alpha_2, \quad \tilde{\alpha}_3 \cdots = \tilde{\alpha}_{q-1} = 0$$

and

$$\tilde{\alpha}_q P_{t-q}^{(q)}\left(x,w_1,\dots,w_{t-(q+1)}\right)w_{t-q}^2 \equiv \alpha_2 P_{t-2}^{(q)}\left(x,w_1,\dots,w_{t-3}\right)w_{t-2}^2$$

which duplicates the second term in the sum and has to be ignored. Hence, the last formula in Equation (7.222) yields Equation (7.233).

Consider the case $q = 3$. Then, according to Definition 644, we clearly have

$$\sigma_1^2 = P_1^{(3)}\left(X_0\right) \quad \text{and} \quad \sigma_2^2 = P_2^{(3)}\left(X_0,W_1\right),$$

where $P_1^{(3)} : \mathbb{R} \to \mathbb{R}$ and $P_2^{(3)} : \mathbb{R}^2 \to \mathbb{R}$ are the polynomials still given by

$$P_1^{(3)}(x) \stackrel{\text{def}}{=} \alpha_0 + \alpha_1 x^2, \quad \forall x \in \mathbb{R}, \tag{7.234}$$

and

$$P_2^{(3)}(x,w) \stackrel{\text{def}}{=} \alpha_0 + \alpha_1 P_1^{(3)}(x)\,w^2 + \alpha_2 x^2, \quad \forall(x,w) \in \mathbb{R}^2, \tag{7.235}$$

but

$$\begin{aligned}\sigma_3^2 &= \alpha_0 + \alpha_1 X_2^2 + \alpha_2 X_1^2 + \alpha_3 X_0^2 \\ &= \alpha_0 + \alpha_1\sigma_2^2 W_2^2 + \alpha_2\sigma_1^2 W_1^2 + \alpha_3 X_0^2 \\ &= \alpha_0 + \alpha_1 P_2^{(3)}\left(X_0,W_1\right)W_2^2 + \alpha_2 P_1^{(3)}\left(X_0\right)W_1^2 + \alpha_3 X_0^2\end{aligned}$$

Hence,

$$\sigma_3^2 = P_3^{(3)}\left(X_0,W_1,W_2\right),$$

where $P_3^{(3)} : \mathbb{R}^3 \to \mathbb{R}$ is the polynomial given by

$$P_3^{(3)}\left(x,w_1,w_2\right) \stackrel{\text{def}}{=} \alpha_0 + \alpha_1 P_2^{(3)}\left(x,w_1\right)w_2^2 + \alpha_2 P_1^{(3)}(x)\,w_1^2 + \alpha_3 x^2, \quad \forall\left(x,w_1,w_2\right) \in \mathbb{R}^3. \tag{7.236}$$

Furthermore,

$$\begin{aligned}\sigma_4^2 &= \alpha_0 + \alpha_1 X_3^2 + \alpha_2 X_2^2 + \alpha_3 X_1^2 \\ &= \alpha_0 + \alpha_1 \sigma_3^2 W_3^2 + \alpha_2 \sigma_2^2 W_2^2 + \alpha_3 \sigma_1^2 W_1^2 \\ &= \alpha_0 + \alpha_1 P_3^{(3)}\left(X_0, W_1, W_2\right) W_3^2 + \alpha_2 P_2^{(3)}\left(X_0, W_1\right) W_2^2 + \alpha_3^2 P_1^{(3)}\left(X_0\right) W_1^2\end{aligned}$$

Hence,

$$\sigma_4^2 = P_4^{(3)}\left(X_0, W_1, W_2, W_3\right),$$

where $P_4^{(3)} : \mathbb{R}^4 \to \mathbb{R}$ is the polynomial given by

$$\begin{aligned}P_4^{(3)}\left(x, w_1, w_2, w_3\right) \stackrel{\text{def}}{=} &\alpha_0 + \alpha_1 P_3^{(3)}\left(x, w_1, w_2\right) w_3^2 + \alpha_2 P_2^{(3)}\left(x, w_1\right) w_2^2 \\ &+ \alpha_3 P_1^{(3)}\left(x\right), \quad \forall \left(x, w_1, w_2, w_3\right) \in \mathbb{R}^3.\end{aligned} \tag{7.237}$$

One more,

$$\begin{aligned}\sigma_5^2 &= \alpha_0 + \alpha_1 X_4^2 + \alpha_2 X_3^2 + \alpha_3 X_2^2 \\ &= \alpha_0 + \alpha_1 \sigma_4^2 W_4^2 + \alpha_2 \sigma_3^2 W_3^2 + \alpha_3 \sigma_2^2 W_2^2 \\ &= \alpha_0 + \alpha_1 P_4^{(3)}\left(X_0, W_1, W_2, W_3\right) W_4^2 + \alpha_2 P_3^{(3)}\left(X_0, W_1, W_2\right) W_3^2 + \alpha_3^2 P_2^{(3)}\left(X_0, W_1\right) W_2^2.\end{aligned}$$

Hence,

$$\sigma_5^2 = P_5^{(3)}\left(X_0, W_1, W_2, W_3, W_4\right),$$

where $P_5^{(3)} : \mathbb{R}^5 \to \mathbb{R}$ is the polynomial given by

$$\begin{aligned}P_5^{(3)}\left(x, w_1, w_2, w_3, w_4\right) \stackrel{\text{def}}{=} &\alpha_0 + \alpha_1 P_4^{(3)}\left(x, w_1, w_2, w_3\right) w_4^2 + \alpha_2 P_3^{(3)}\left(x, w_1, w_2\right) w_3^2 \\ &+ \alpha_3 P_2^{(3)}\left(x, w_1\right) w_2^2, \quad \forall \left(x, w_1, w_2, w_3, w_4\right) \in \mathbb{R}^4.\end{aligned} \tag{7.238}$$

Again, applying the Induction Principle, it is then possible to prove that for every $t \in \mathbb{N}$, such that $t \geqslant 5$, we have

$$\sigma_t^2 = P_t^{(3)}\left(X_0, W_1, \ldots, W_{t-1}\right)$$

where $P_t^{(3)} : \mathbb{R}^t \to \mathbb{R}$ is the polynomial given by

$$\begin{aligned}P_t^{(3)}\left(x, w_1, \ldots, w_{t-1}\right) \stackrel{\text{def}}{=} &\alpha_0 + \alpha_1 P_{t-1}^{(3)}\left(x, w_1, \ldots, w_{t-2}\right) w_{t-1}^2 + \alpha_1 P_{t-2}^{(3)}\left(x, w_1, \ldots, w_{t-3}\right) w_{t-2}^2 \\ &+ \alpha_3 P_{t-3}^{(3)}\left(x, w_1, \ldots, w_{t-4}\right) w_{t-3}^2, \quad \forall \left(x, w_1, \ldots, w_{t-2}, w_{t-1}\right) \in \mathbb{R}^t,\end{aligned} \tag{7.239}$$

and we keep (7.224). Now, setting $q = 3$ Equations (7.234)-(7.236) can be obtained by Equation (7.219) and the second and third formulas in Equation (7.222). In addition, under (7.223) and (7.224), Equation (7.239) can be obtained from the last formula in Equation (7.222). In fact, with reference to the last formula in Equation (7.222), for any $t > q = 3$ we have

$$\tilde{\alpha}_2 = \alpha_2, \quad \tilde{\alpha}_3 = \alpha_3, \quad \tilde{\alpha}_4 \cdots = \tilde{\alpha}_{q-1} = 0$$

and

$$\tilde{\alpha}_q P_{t-q}^{(q)}\left(x, w_1, \ldots, w_{t-(q+1)}\right) w_{t-q}^2 \equiv \alpha_3 P_{t-3}^{(q)}\left(x, w_1, \ldots, w_{t-4}\right) w_{t-3}^2$$

which duplicates the third term in the sum and has to be ignored. Hence, the last formula in Equation (7.222) yields Equation (7.239).

Consider the case $q > 3$... . $\square$

Lemma 649 *We have*

$$\sigma_t^2 = \alpha_0 \left(1 + \sum_{s_1=1}^{q} \alpha_s W_{t-s}^2 + \sum_{s_1,s_2=1}^{q} \alpha_{s_1}\alpha_{s_2} W_{t-s_1-s_2}^2 W_{t-s_1}^2 + \cdots \right.$$

$$\left. + \sum_{\substack{s_1,s_2,\dots,s_{\check{n}}=1 \\ \check{n}=\max\{n\in\mathbb{N}:\ t-nq\geqslant q\}}}^{q} \alpha_{s_1}\alpha_{s_2}\cdots\alpha_{s_{\check{n}}} W_{t-s_1-s_2-\cdots-s_{\check{n}}}^2 \cdots W_{t-s_1-s_2}^2 W_{t-s_1}^2 \right)$$

$$+ \sum_{\substack{s_1,s_2,\dots,s_{\check{n}}=1 \\ \check{n}=\max\{n\in\mathbb{N}:\ t-nq\geqslant q\}}}^{q} \sum_{s_{\check{n}+1}=1}^{q-1} \alpha_{s_1}\alpha_{s_2}\cdots\alpha_{s_{\check{n}}} \sigma_{t-s_1-s_2-\cdots-s_{\check{n}}-s_{\check{n}+1}}^2 W_{t-s_1-s_2-\cdots-s_{\check{n}}}^2 \cdots W_{t-s_1-s_2}^2 W_{t-s_1}^2$$

for any $t \in \mathbb{N}$.

Proof. Considering Equation (7.217), we can write

$$\begin{aligned}
\sigma_t^2 &= \alpha_0 + \sum_{s_1=1}^{q} \alpha_{s_1}\sigma_{t-s_1}^2 W_{t-s_1}^2 \\
&= \alpha_0 + \sum_{s_1=1}^{q} \alpha_{s_1} \left(\alpha_0 + \sum_{s_2=1}^{q} \alpha_{s_2} X_{t-s_1-s_2}^2 \right) W_{t-s_1}^2 \\
&= \alpha_0 \left(1 + \sum_{s_1=1}^{q} \alpha_{s_1} W_{t-s_1}^2 \right) + \sum_{s_1,s_2=1}^{q} \alpha_{s_1}\alpha_{s_2} X_{t-s_1-s_2}^2 W_{t-s_1}^2 \\
&= \alpha_0 \left(1 + \sum_{s_1=1}^{q} \alpha_{s_1} W_{t-s_1}^2 \right) + \sum_{s_1,s_2=1}^{q} \alpha_{s_1}\alpha_{s_2} \sigma_{t-s_1-s_2}^2 W_{t-s_1-s_2}^2 W_{t-s_1}^2 \\
&= \alpha_0 \left(1 + \sum_{s_1=1}^{q} \alpha_{s_1} W_{t-s_1}^2 \right) + \sum_{s_1,s_2=1}^{q} \alpha_{s_1}\alpha_{s_2} \left(\alpha_0 + \sum_{s_3=1}^{q} \alpha_{s_3} X_{t-s_1-s_2-s_3}^2 \right) W_{t-s_1-s_2}^2 W_{t-s_1}^2 \\
&= \alpha_0 \left(1 + \sum_{s_1=1}^{q} \alpha_{s_1} W_{t-s_1}^2 \right) + \alpha_0 \sum_{s_1,s_2=1}^{q} \alpha_{s_1}\alpha_{s_2} W_{t-s_1-s_2}^2 W_{t-s_1}^2 \\
&+ \sum_{s_1,s_2,s_3=1}^{q} \alpha_{s_1}\alpha_{s_2}\alpha_{s_3} X_{t-s_1-s_2-s_3}^2 W_{t-s_1-s_2}^2 W_{t-s_1}^2 \\
&= \alpha_0 \left(1 + \sum_{s_1=1}^{q} \alpha_{s_1} W_{t-s_1}^2 + \sum_{s_1,s_2=1}^{q} \alpha_{s_1}\alpha_{s_2} W_{t-s_1-s_2}^2 W_{t-s_1}^2 \right) \\
&+ \sum_{s_1,s_2,s_3=1}^{q} \alpha_{s_1}\alpha_{s_2}\alpha_{s_3} \sigma_{t-s_1-s_2-s_3}^2 W_{t-s_1-s_2-s_3}^2 W_{t-s_1-s_2}^2 W_{t-s_1}^2.
\end{aligned}$$

Iterating the above computation n times the desired result clearly follows. □

Let $\left(\mathcal{F}_t^{X_0,W}\right)_{t\in\hat{n}} \equiv \mathfrak{F}^{X_0,W}$ the filtration generated by the inital state X_0 of the $ARCH\,(q)$ process X and the innovation process W, that is

$$\mathcal{F}_0^{X_0,W} \stackrel{\text{def}}{=} \sigma\,(X_0) \quad \text{and} \quad \mathcal{F}_t^{X_0,W} \stackrel{\text{def}}{=} \sigma\,(X_0, W_1, \dots, W_t)\,, \quad \forall t \in \mathbb{N},$$

where $\sigma\,(X, Y, Z, \dots)$ denotes the σ-algebra generated by the random variables $X, Y, Z, \dots$

Osservazione 650 *We clearly have*

$$\sigma\left(W_t\right) \subseteq \mathcal{F}_t, \tag{7.240}$$

and the random variable W_t is independent of $\mathcal{F}_{t-1}$, for every $t \in \mathbb{N}$. In addition,

$$\sigma\left(\sigma_t\right) \subseteq \mathcal{F}_{t-1} \quad \textit{and} \quad \sigma\left(X_t\right) \subseteq \mathcal{F}_t, \tag{7.241}$$

for every $t \in \mathbb{N}$.

Proposizione 651 *If X is an ARCH(q) process, then the random variable W_t is independent of σ_s, for every $t \in \mathbb{N}$ and $s = 1, \ldots, t$.*

Proof. The desired result immediately follows from Equations (7.219) and (7.221) in Lemma 648 and the independence assumptions on the strong white noise W in Definition 644. □

Proposizione 652 *If X is an ARCH(q) process, then both σ and X are second-order the processes.*

Proof. By assumption X_0 has finite second-order moment. Hence, the random variable

$$\sigma_1^2 = \alpha_0 + \alpha_1 X_0^2$$

has finite first-order moment, that is σ_1 has finite second-order moment. As a consequence of Proposition 651, W_1^2, which has finite first order moment, is independent of σ_1^2. It follows

$$\mathbf{E}\left[\sigma_1^2 W_1^2\right] = \mathbf{E}\left[\sigma_1^2\right] \mathbf{E}\left[W_1^2\right] < \infty.$$

This implies that the random variable

$$X_1 = \sigma_1 W_1$$

has finite second-order moment. From Remark **??** we have

$$\sigma_2^2 \overset{\text{def}}{=} \begin{cases} \alpha_0 + \alpha_1 X_1^2, & \text{if } q = 1, \\ \alpha_0 + \alpha_1 X_1^2 + \alpha_2 X_0^2, & \text{if } q \geqslant 2. \end{cases}$$

Hence, in any case, the random variable σ_2^2 has finite first-order moment, that is σ_2 has finite second-order moment. Still as a consequence of Proposition 651, the random variable W_2^2, with finite first order moment, is independent of σ_2^2. It follows

$$\mathbf{E}\left[\sigma_2^2 W_2^2\right] = \mathbf{E}\left[\sigma_2^2\right] \mathbf{E}\left[W_2^2\right] < \infty.$$

This implies that the random variable

$$X_2 = \sigma_2 W_2$$

has finite second-order moment. Iterating this argument, we can prove that σ_t and X_t have finite second-order moment for every $t = 1, \ldots, q-1$. Thereby, fixed any $t \geqslant q$ and assuming that the random variables σ_s and X_s have finite second-order moment for every $s = 1, \ldots, t$, since

$$\sigma_{t+1}^2 = \alpha_0 + \sum_{s=1}^{q} \alpha_s X_{t+1-s}^2,$$

the random variable σ_{t+1}^2 has finite first-order moment, that is σ_{t+1} has finite second-order moment. In addition, since the random variable W_{t+1}^2, which has finite first order moment, is independent of σ_{t+1}^2 we have

$$\mathbf{E}\left[\sigma_{t+1}^2 W_{t+1}^2\right] = \mathbf{E}\left[\sigma_{t+1}^2\right] \mathbf{E}\left[W_{t+1}^2\right] < \infty.$$

This implies that the random variable

$$X_{t+1} = \sigma_{t+1} W_{t+1}$$

has finite second-order moment. Therefore, the random variables σ_s and X_s have finite second-order moment for every $s = 1, \ldots, t+1$. By virtue of the Induction Principle, it then follows that σ_t and X_t have finite second-order moment for every $t \in \mathbb{N}$, which is the desired result. □

Proposizione 653 *We have*

$$\mu_X(t) \equiv \mathbf{E}[X_t] = 0, \tag{7.242}$$

for every $t \in \hat{n}$.

Proof. By assumption,

$$\mathbf{E}[X_0] = 0.$$

Now, thanks to Proposition 651, the noise term W_t is independent of σ_t, for every $t \in \mathbb{N}$. Therefore,

$$\mathbf{E}[X_t] = \mathbf{E}[\sigma_t W_t] = \mathbf{E}[\sigma_t]\,\mathbf{E}[W_t] = 0.$$

for every $t \in \mathbb{N}$. $\square$

Let $\mathbf{E}[\cdot \mid \mathcal{F}]$ [resp. $\mathbf{D}^2[\cdot \mid \mathcal{F}]$] be the conditional expectation [resp. variance] operator, given a σ-algebra $\mathcal{F}$. Recall that for any real random variable Z with finite second order moment the random variable $\mathbf{E}[Z \mid \mathcal{F}]$ is the best estimate of Z, in the mean square sense, given the information represented by $\mathcal{F}$. Moreover,

Recall 654 *For any real random variable* Z *on* Ω *with finite second order moment we have*

$$\mathbf{D}^2[Z \mid \mathcal{F}] = \mathbf{E}\left[Z^2 \mid \mathcal{F}\right] - \mathbf{E}[Z \mid \mathcal{F}]^2. \tag{7.243}$$

With reference to the ARCH(q) process X and the filtration $\mathfrak{F}$ generated by the initial state X_0 of X and the strong white noise prosess W we can state

Proposizione 655 *We have*

$$\mathbf{E}[X_{t-s} \mid \mathcal{F}_t] = X_{t-s}, \tag{7.244}$$

for every $t \in \mathbb{N}$ *and every* $s = 0, 1, \ldots, t$*, and*

$$\mathbf{E}[X_t \mid \mathcal{F}_{t-s}] = 0, \tag{7.245}$$

for every $t \in \mathbb{N}$ *and every* $s = 1, \ldots, t$.

Proof. Thanks to (7.241) in Remark 650 and the tower property of the conditional expectation operator, we have

$$\mathbf{E}[X_t \mid \mathcal{F}_t] = \mathbf{E}[\mathbf{E}[X_t \mid \sigma(X_t)] \mid \mathcal{F}_t] = \mathbf{E}[X_t \mid \sigma(X_t)] = X_t, \tag{7.246}$$

for every $t \in \mathbb{N}$, which is Equation (7.244) for every $t \in \mathbb{N}$ and $s = 0$. To complete the proof of (7.244), it sufficient to observe that since

$$\mathcal{F}_{t-s} \subseteq \mathcal{F}_t \tag{7.247}$$

for every $s = 1, \ldots, t$, thanks to (7.246) and the tower property of the conditional expectation operator, we have

$$\mathbf{E}[X_{t-s} \mid \mathcal{F}_t] = \mathbf{E}[\mathbf{E}[X_{t-s} \mid \mathcal{F}_{t-s}] \mid \mathcal{F}_t] = \mathbf{E}[X_{t-s} \mid \mathcal{F}_{t-s}] = X_{t-s}.$$

Now, again on account of Remark 650 and the transparency property of the conditional expectation, we obtain

$$\mathbf{E}[X_t \mid \mathcal{F}_{t-1}] = \mathbf{E}[\sigma_t W_t \mid \mathcal{F}_{t-1}] = \sigma_t \mathbf{E}[W_t \mid \mathcal{F}_{t-1}] = \sigma_t \mathbf{E}[W_t] = 0, \tag{7.248}$$

for every $t \in \mathbb{N}$. Hence, we have obtained the desired (7.244), for every $t \in \mathbb{N}$ and $s = 1$. On the other hand, thanks to (7.247) and the tower property of the conditional expectation operator, we can write

$$\mathbf{E}[X_t \mid \mathcal{F}_{t-s}] = \mathbf{E}[\mathbf{E}[X_t \mid \mathcal{F}_{t-1}] \mid \mathcal{F}_{t-s}] = 0,$$

for every $s = 2, \ldots, t$. This completes the proof of (7.245). $\square$

To deal with the unconditional variance of the ARCH(q) processes X, we need an important assumption on the parameters $\alpha_0, \alpha_1, \ldots, \alpha_q \in \mathbb{R}_+$.

Assumption 656 *We assume*

$$1-\sigma_W^2\sum_{k=1}^{q}\alpha_k>0, \tag{7.249}$$

where σ_W^2 *is the value of the constant variance function of the strong white noise* W*.*

Proposizione 657 *Under Assumption 656, assume further that the process* X *is asymptotically variance stationary, that is*

$$\lim_{t\to\infty}\mathbf{D}^2\left[X_t\right]=\lim_{t\to\infty}\mathbf{E}\left[X_t^2\right]\equiv\sigma_X^2<\infty \tag{7.250}$$

(see (7.242) in Proposition 653). Then,

$$\sigma_X^2=\frac{\alpha_0\sigma_W^2}{1-\sigma_W^2\sum\limits_{k=1}^{q}\alpha_k}. \tag{7.251}$$

Proof. Thanks to Proposition 651 and the linearity of the expectation operator, we have

$$\begin{aligned}\mathbf{E}\left[X_t^2\right]&=\mathbf{E}\left[\sigma_t^2W_t^2\right]=\mathbf{E}\left[\sigma_t^2\right]\mathbf{E}\left[W_t^2\right]\\&=\mathbf{E}\left[\alpha_0+\sum_{s=1}^{q}\alpha_sX_{t-s}^2\right]\mathbf{E}\left[W_t^2\right]\\&=\left(\alpha_0+\sum_{s=1}^{q}\alpha_s\mathbf{E}\left[X_{t-s}^2\right]\right)\sigma_W^2,\end{aligned} \tag{7.252}$$

for every $t\geqslant q$. From (7.252), passing to the limit as $t\to\infty$, under the hypothesis that the process X is asymptotically variance stationary, it follows

$$\sigma_X^2=\left(\alpha_0+\sum_{s=1}^{q}\alpha_s\sigma_X^2\right)\sigma_W^2. \tag{7.253}$$

Solving Equation (7.253) for σ_X^2, thanks to (7.249) in Assumption 656, we then obtain (7.251). □

Proposizione 658 *Under Assumption 656, the process* X *is asymptotically variance stationary. Thus, Equation (7.251) in Proposition 657 holds true.*

Proof. Assume first $q=1$. In this case, Equation (7.249) in Assumption 656 reduces to

$$1-\alpha_1\sigma_W^2>0. \tag{7.254}$$

We have

$$\mathbf{E}\left[X_1^2\right]=\mathbf{E}\left[\sigma_1^2W_1^2\right]=\mathbf{E}\left[\sigma_1^2\right]\mathbf{E}\left[W_1^2\right]=\left(\alpha_0+\alpha_1\mathbf{E}\left[X_0^2\right]\right)\sigma_W^2.$$

This implies

$$\begin{aligned}\mathbf{E}\left[X_2^2\right]&=\mathbf{E}\left[\sigma_2^2W_2^2\right]=\mathbf{E}\left[\sigma_2^2\right]\mathbf{E}\left[W_2^2\right]=\left(\alpha_0+\alpha_1\mathbf{E}\left[X_1^2\right]\right)\sigma_W^2\\&=\left(\alpha_0+\alpha_1\left(\alpha_0+\alpha_1\mathbf{E}\left[X_0^2\right]\right)\sigma_W^2\right)\sigma_W^2\\&=\left(\alpha_0\left(1+\alpha_1\sigma_W^2\right)+\alpha_1^2\sigma_W^2\mathbf{E}\left[X_0^2\right]\right)\sigma_W^2.\end{aligned}$$

In turn,

$$\begin{aligned}\mathbf{E}\left[X_3^2\right]&=\mathbf{E}\left[\sigma_3^2W_3^2\right]=\mathbf{E}\left[\sigma_3^2\right]\mathbf{E}\left[W_3^2\right]=\left(\alpha_0+\alpha_1\mathbf{E}\left[X_2^2\right]\right)\sigma_W^2\\&=\left(\alpha_0+\alpha_1\left(\alpha_0\left(1+\alpha_1\sigma_W^2\right)+\alpha_1^2\sigma_W^2\mathbf{E}\left[X_0^2\right]\right)\sigma_W^2\right)\sigma_W^2\\&=\left(\alpha_0\left(1+\alpha_1\sigma_W^2+\alpha_1^2\sigma_W^4\right)+\alpha_1^3\sigma_W^4\mathbf{E}\left[X_0^2\right]\right)\sigma_W^2.\end{aligned}$$

The above equations lead to guess that

$$\mathbf{E}\left[X_t^2\right] = \left(\alpha_0 \sum_{s=0}^{t-1} \alpha_1^s \sigma_W^{2s} + \alpha_1^t \sigma_W^{2(t-1)} \mathbf{E}\left[X_0^2\right]\right) \sigma_W^2, \tag{7.255}$$

for every $t \in \mathbb{N}$. This can be proven by applying the Induction Principle. In fact, assuming that Equation (7.255) holds true for some $t \in \mathbb{N}$, we can write

$$\begin{aligned}
\mathbf{E}\left[X_{t+1}^2\right] &= \mathbf{E}\left[\sigma_{t+1}^2 W_{t+1}^2\right] = \mathbf{E}\left[\sigma_{t+1}^2\right] \mathbf{E}\left[W_{t+1}^2\right] \\
&= \mathbf{E}\left[\alpha_0 + \alpha_1 X_t^2\right] \mathbf{E}\left[W_{t+1}^2\right] = \left(\alpha_0 + \alpha_1 \mathbf{E}\left[X_t^2\right]\right) \sigma_W^2 \\
&= \left(\alpha_0 + \alpha_1 \left(\alpha_0 \sum_{s=0}^{t-1} \alpha_1^s \sigma_W^{2s} + \alpha_1^t \sigma_W^{2(t-1)} \mathbf{E}\left[X_0^2\right]\right) \sigma_W^2\right) \sigma_W^2 \\
&= \left(\alpha_0 + \alpha_0 \sum_{s=0}^{t-1} \alpha_1^{s+1} \sigma_W^{2(s+1)} + \alpha_1^{t+1} \sigma_W^{2t} \mathbf{E}\left[X_0^2\right]\right) \sigma_W^2 \\
&= \left(\alpha_0 \left(1 + \sum_{s=0}^{t-1} \alpha_1^{s+1} \sigma_W^{2(s+1)}\right) + \alpha_1^{t+1} \sigma_W^{2t} \mathbf{E}\left[X_0^2\right]\right) \sigma_W^2 \\
&= \left(\alpha_0 \sum_{s=0}^{t} \alpha_1^s \sigma_W^{2s} + \alpha_1^{t+1} \sigma_W^{2t} \mathbf{E}\left[X_0^2\right]\right) \sigma_W^2,
\end{aligned}$$

which replicates (7.255) when replacing $t+1$ to t. Thus, Equation (7.255) holds true. Now, since

$$\sum_{s=0}^{t-1} \alpha_1^s \sigma_W^{2s} = \frac{1 - \alpha_1^t \sigma_W^{2t}}{1 - \alpha_1 \sigma_W^2},$$

from (7.255), we can write

$$\mathbf{E}\left[X_t^2\right] = \frac{\alpha_0 \sigma_W^2}{1 - \alpha_1 \sigma_W^2} \left(1 - \alpha_1^t \sigma_W^{2t}\right) + \alpha_1^t \sigma_W^{2(t-1)} \mathbf{E}\left[X_0^2\right]. \tag{7.256}$$

On the other hand, (7.254) clearly implies that

$$\lim_{t \to \infty} \alpha_1^t \sigma_W^{2t} = 0.$$

Therefore, from (7.256), it follows

$$\lim_{t \to \infty} \mathbf{E}\left[X_t^2\right] = \lim_{t \to \infty} \left(\frac{\alpha_0 \sigma_W^2}{1 - \alpha_1 \sigma_W^2} \left(1 - \alpha_1^t \sigma_W^{2t}\right) + \alpha_1^t \sigma_W^{2(t-1)} \mathbf{E}\left[X_0^2\right]\right) = \frac{\alpha_0 \sigma_W^2}{1 - \alpha_1 \sigma_W^2},$$

which is Equation (7.251) in case $q = 1$.

Now, assume $q = 2$. In this case, Equation (7.249) in Assumption 656 reduces to

$$1 - \sigma_W^2 \left(\alpha_1 + \alpha_2\right) > 0. \tag{7.257}$$

As above, we have

$$\mathbf{E}\left[X_1^2\right] = \mathbf{E}\left[\sigma_1^2 W_1^2\right] = \mathbf{E}\left[\sigma_1^2\right] \mathbf{E}\left[W_1^2\right] = \left(\alpha_0 + \alpha_1 \mathbf{E}\left[X_0^2\right]\right) \sigma_W^2.$$

This implies

$$\begin{aligned}
\mathbf{E}\left[X_2^2\right] &= \mathbf{E}\left[\sigma_2^2 W_2^2\right] = \mathbf{E}\left[\sigma_2^2\right] \mathbf{E}\left[W_2^2\right] = \left(\alpha_0 + \alpha_1 \mathbf{E}\left[X_1^2\right] + \alpha_2 \mathbf{E}\left[X_0^2\right]\right) \sigma_W^2 \\
&= \left(\alpha_0 + \alpha_1 \left(\alpha_0 + \alpha_1 \mathbf{E}\left[X_0^2\right]\right) \sigma_W^2 + \alpha_2 \mathbf{E}\left[X_0^2\right]\right) \sigma_W^2 \\
&= \left(\alpha_0 \left(1 + \alpha_1 \sigma_W^2\right) + \left(\alpha_1^2 \sigma_W^2 + \alpha_2\right) \mathbf{E}\left[X_0^2\right]\right) \sigma_W^2.
\end{aligned}$$

In turn,

$$
\begin{aligned}
\mathbf{E}\left[X_3^2\right] &= \mathbf{E}\left[\sigma_3^2 W_3^2\right] = \mathbf{E}\left[\sigma_3^2\right]\mathbf{E}\left[W_3^2\right] = \left(\alpha_0 + \alpha_1 \mathbf{E}\left[X_2^2\right] + \alpha_2 \mathbf{E}\left[X_1^2\right]\right)\sigma_W^2 \\
&= \left(\alpha_0 + \alpha_1\left(\alpha_0\left(1+\alpha_1\sigma_W^2\right) + \left(\alpha_1^2\sigma_W^2 + \alpha_2\right)\mathbf{E}\left[X_0^2\right]\right)\sigma_W^2 + \alpha_2\left(\alpha_0 + \alpha_1\mathbf{E}\left[X_0^2\right]\right)\sigma_W^2\right)\sigma_W^2 \\
&= \left(\alpha_0 + \alpha_0\alpha_1\left(1+\alpha_1\sigma_W^2\right)\sigma_W^2 + \alpha_0\alpha_2\sigma_W^2 + \alpha_1\left(\alpha_1^2\sigma_W^2+\alpha_2\right)\mathbf{E}\left[X_0^2\right]\sigma_W^2 + \alpha_1\alpha_2\mathbf{E}\left[X_0^2\right]\sigma_W^2\right)\sigma_W^2 \\
&= \left(\alpha_0\left(1+\alpha_1\sigma_W^2+\alpha_1^2\sigma_W^4+\alpha_2\sigma_W^2\right) + \left(\alpha_1^3\sigma_W^4 + 2\alpha_1\alpha_2\sigma_W^2\right)\mathbf{E}\left[X_0^2\right]\right)\sigma_W^2
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbf{E}\left[X_4^2\right] &= \mathbf{E}\left[\sigma_4^2 W_4^2\right] = \mathbf{E}\left[\sigma_4^2\right]\mathbf{E}\left[W_4^2\right] = \left(\alpha_0 + \alpha_1\mathbf{E}\left[X_3^2\right] + \alpha_2\mathbf{E}\left[X_2^2\right]\right)\sigma_W^2 \\
&= \left(\alpha_0 + \alpha_1\left(\alpha_0\left(1+\alpha_1\sigma_W^2+\alpha_1^2\sigma_W^4+\alpha_2\sigma_W^2\right) + \alpha_1\left(\alpha_1^2\sigma_W^2+2\alpha_2\right)\mathbf{E}\left[X_0^2\right]\sigma_W^2\right)\sigma_W^2\right)\sigma_W^2 \\
&+ \alpha_2\left(\left(\alpha_0\left(1+\alpha_1\sigma_W^2\right) + \left(\alpha_1^2\sigma_W^2+\alpha_2\right)\mathbf{E}\left[X_0^2\right]\right)\sigma_W^2\right)\sigma_W^2 \\
&= \left(\alpha_0 + \alpha_0\alpha_1\left(1+\alpha_1\sigma_W^2+\alpha_1^2\sigma_W^4\right)\sigma_W^2 + \alpha_0\alpha_1\alpha_2\sigma_W^4 + \alpha_0\alpha_2\left(1+\alpha_1\sigma_W^2\right)\sigma_W^2\right. \\
&+ \left.\left(\alpha_1^2\left(\alpha_1^2\sigma_W^2+2\alpha_2\right)\sigma_W^4 + \alpha_2\left(\alpha_1^2\sigma_W^2+\alpha_2\right)\sigma_W^2\right)\mathbf{E}\left[X_0^2\right]\right)\sigma_W^2 \text{ OK!} \\
&= \left(\alpha_0\left(1+\alpha_1\sigma_W^2+\alpha_1^2\sigma_W^4+\alpha_1^3\sigma_W^6+\alpha_2\sigma_W^2+2\alpha_1\alpha_2\sigma_W^4\right)\right. \\
&+ \left.\left(\alpha_1^4\sigma_W^6+\alpha_2^2\sigma_W^2+3\alpha_1^2\alpha_2\sigma_W^4\right)\mathbf{E}\left[X_0^2\right]\right)\sigma_W^2 \text{ OK!} \\
&= \left(\alpha_0\left(1+\left(\alpha_1+\alpha_2\right)\sigma_W^2+\left(\alpha_1+\alpha_2\right)^2\sigma_W^4+\alpha_1^3\sigma_W^6-\alpha_2^2\sigma_W^4\right)\right. \\
&+ \left.\left(\alpha_1^4\sigma_W^6+\alpha_2^2\sigma_W^2+3\alpha_1^2\alpha_2\sigma_W^4\right)\mathbf{E}\left[X_0^2\right]\right) \\
&= \left(\alpha_0\left(\sum_{s=0}^{t-2}\left(\alpha_1+\alpha_2\right)^s\sigma_W^{2s} + \alpha_1^{t-1}\sigma_W^{2(t-1)} - \alpha_2^{t-2}\sigma_W^{2(t-1)}\right) + \left(\alpha_1^t\sigma_W^{2(t-1)} + \alpha_2^{t-2}\sigma_W^{2(t-3)} + (t-1)\,\alpha_1^{t-2}\alpha_2\sigma_W^{2(t-2)}\right.\right.
\end{aligned}
$$

$$\sum_{s=0}^{t-2}\left(\alpha_1+\alpha_2\right)^s\sigma_W^{2s} = \frac{1-\left(\alpha_1+\alpha_2\right)^{t-1}\sigma_W^{2(t-1)}}{1-\left(\alpha_1+\alpha_2\right)\sigma_W^2}$$

$$1-\left(\alpha_1+\alpha_2\right)\sigma_W^2 > 0$$

$$\left(\alpha_1+\alpha_2\right)\sigma_W^2 < 1 \Rightarrow \alpha_1\sigma_W^2, \alpha_2\sigma_W^2 < 1$$

$$
\begin{aligned}
&t = 1 \to 0 \\
&t = 2 \to \alpha_2 \\
&t = 3 \to 2\alpha_1\alpha_2\sigma_W^2 \\
&t = 4 \to \alpha_2^2\sigma_W^2 + 3\alpha_1^2\alpha_2\sigma_W^4 \\
&\sum_{s=0}^{t-1}
\end{aligned}
$$

$= 0$

$$\mathbf{E}\left[X_t^2\right] = \left(\alpha_0\sum_{s=0}^{t-1}\alpha_1^s\sigma_W^{2s} + \alpha_1^t\sigma_W^{2(t-1)}\mathbf{E}\left[X_0^2\right]\right)\sigma_W^2$$

$$
\begin{aligned}
X_t^2 &= \mathbf{E}\left[\sigma_t^2 W_t^2\right] = \mathbf{E}\left[\sigma_t^2\right]\mathbf{E}\left[W_t^2\right] = \mathbf{E}\left[\alpha_0 + \sum_{s=1}^{q}\alpha_s X_{t-s}^2\right]\sigma_W^2 = \left(\alpha_0 + \sum_{s=1}^{q}\alpha_s\mathbf{E}\left[X_{t-s}^2\right]\right)\sigma_W^2 \\
&= \left(\alpha_0 + \sum_{s=1}^{q}\alpha_s\mathbf{E}\left[\sigma_{t-s}^2 W_{t-s}^2\right]\right)\sigma_W^2 = \left(\alpha_0 + \sum_{s=1}^{q}\alpha_s\mathbf{E}\left[\sigma_{t-s}^2\right]\mathbf{E}\left[W_{t-s}^2\right]\right)\sigma_W^2 \\
&= \left(\alpha_0 + \sigma_W^2\sum_{s=1}^{q}\alpha_s\mathbf{E}\left[\sigma_{t-s}^2\right]\right)\sigma_W^2
\end{aligned}
$$

$\square$

Proposizione 659 *Under Assumption 656, assume further that $q = 1$ and that we have*

$$\mathbf{E}\left[X_{t_0}^2\right] = \frac{\alpha_0 \sigma_W^2}{1 - \alpha_1 \sigma_W^2}, \tag{7.258}$$

for some $t_0 \in \hat{n}$. Then

$$\mathbf{E}\left[X_t^2\right] = \frac{\alpha_0 \sigma_W^2}{1 - \alpha_1 \sigma_W^2}, \tag{7.259}$$

for every $t \in \mathbb{N}$ such that $t \geqslant t_0$.

Proof. Following the argument in the proof of Proposition 658, it is possible to show that

$$\mathbf{E}\left[X_t^2\right] = \frac{\alpha_0 \sigma_W^2}{1 - \alpha_1 \sigma_W^2}\left(1 - \alpha_1^{(t-t_0)+1} \sigma_W^{2((t-t_0)+1)}\right) + \alpha_1^{(t-t_0)+1} \sigma_W^{2((t-t_0)+1)} \mathbf{E}\left[X_{t_0}^2\right], \tag{7.260}$$

for every $t > t_0$. Combining Equations (7.260) and (7.258), we obtain

$$\mathbf{E}\left[X_t^2\right] = \frac{\alpha_0 \sigma_W^2}{1 - \alpha_1 \sigma_W^2}\left(1 - \alpha_1^{(t-t_0)+1} \sigma_W^{2((t-t_0)+1)}\right) + \alpha_1^{(t-t_0)+1} \sigma_W^{2((t-t_0)+1)} \frac{\alpha_0 \sigma_W^2}{1 - \alpha_1 \sigma_W^2} = \frac{\alpha_0 \sigma_W^2}{1 - \alpha_1 \sigma_W^2},$$

as desired. □

Definizione 660 *Under Assumption 656, assume further that $q = 1$ and that Equation (7.258) holds true for some $t_0 \in \hat{n}$. Then we call X_{t_0} a variance-stationary state of the ARCH(1) process X and we say that X is variance stationary.*

Proposizione 661 *We have*

$$Cov\left(X_t, X_s\right) = 0, \tag{7.261}$$

for all $s, t \in \mathbb{N}$ such that $s < t$.

Proof. On account of (7.242) in Proposition (653), we can write

$$Cov\left(X_t, X_s\right) = \mathbf{E}\left[X_t X_s\right] - \mathbf{E}\left[X_t\right] \mathbf{E}\left[X_s\right] = \mathbf{E}\left[X_t X_s\right]. \tag{7.262}$$

On the other hand, by virtue of the iterated expectation and transparency properties of the conditional expectation operator, Equations (7.244) and (7.245) of Proposition 655 yield

$$\mathbf{E}\left[X_t X_s\right] = \mathbf{E}\left[\mathbf{E}\left[X_t X_s \mid \mathcal{F}_s\right]\right] = \mathbf{E}\left[X_s \mathbf{E}\left[X_t \mid \mathcal{F}_s\right]\right] = 0, \tag{7.263}$$

for all $s, t \in \mathbb{N}$ such that $s < t$. Combining (7.262) and (7.263), Equation (7.261) immediately follows. □

Under Assumption 656, assume further that $q = 1$ and X_0 is a variance-stationary state of the ARCH(1) process X. Then Propositions 653, 659, and 661 characterize the process X as a weak white noise. This is the reason why the check for autoregressive conditional heteroskedasticity is often applied to the residuals of an ARIMA process.

Proposizione 662 *Under Assumption 656, assume further that X_0 is a variance-stationary state of the ARCH(q) process X. We then have*

$$\mathbf{E}\left[X_{t-s}^2 \mid \mathcal{F}_t\right] = X_{t-s}^2, \tag{7.264}$$

for every $t \in \mathbb{N}$ and every $s = 0, 1, \ldots, t$, and

$$\mathbf{E}\left[X_t^2 \mid \mathcal{F}_{t-1}\right] = \begin{cases} \left(\alpha_0 + \sum_{s=1}^{t} \alpha_s X_{t-s}^2\right) \sigma_W^2. & \text{if } 1 < t \leqslant q-1, \\ \left(\alpha_0 + \sum_{s=1}^{q} \alpha_s X_{t-s}^2\right) \sigma_W^2, & \text{if } t \geqslant q. \end{cases} \tag{7.265}$$

Proof. Recalling that X is a second order process (see Proposition 652), Equation (7.264) can be proved in an analogous way as Equation (7.244).

Now, on account of (7.264) and thanks to the linearity and transparency properties of the conditional expectation operator,

$$\mathbf{E}\left[X_t^2 \mid \mathcal{F}_{t-1}\right] = \mathbf{E}\left[\sigma_t^2 W_t^2 \mid \mathcal{F}_{t-1}\right]$$
$$= \begin{cases} \mathbf{E}\left[\left(\alpha_0 + \sum_{s=1}^{t} \alpha_s X_{t-s}^2\right) W_t^2 \mid \mathcal{F}_{t-1}\right] = \mathbf{E}\left[\left(\alpha_0 + \sum_{s=1}^{t} \alpha_s \mathbf{E}\left[X_{t-s}^2 \mid \mathcal{F}_{t-1}\right]\right) W_t^2 \mid \mathcal{F}_{t-1}\right] & \\ = \left(\alpha_0 + \sum_{s=1}^{t} \alpha_s \mathbf{E}\left[X_{t-s}^2 \mid \mathcal{F}_{t-1}\right]\right) \mathbf{E}\left[W_t^2 \mid \mathcal{F}_{t-1}\right] = \left(\alpha_0 + \sum_{s=1}^{t} \alpha_s X_{t-s}^2\right) \mathbf{E}\left[W_t^2 \mid \mathcal{F}_{t-1}\right], & \text{if } 1 < t \leqslant q-1, \\ \mathbf{E}\left[\left(\alpha_0 + \sum_{s=1}^{q} \alpha_s X_{t-s}^2\right) W_t^2 \mid \mathcal{F}_{t-1}\right] = \mathbf{E}\left[\left(\alpha_0 + \sum_{s=1}^{q} \alpha_s \mathbf{E}\left[X_{t-s}^2 \mid \mathcal{F}_{t-1}\right]\right) W_t^2 \mid \mathcal{F}_{t-1}\right] & \\ = \left(\alpha_0 + \sum_{s=1}^{q} \alpha_s \mathbf{E}\left[X_{t-s}^2 \mid \mathcal{F}_{t-1}\right]\right) \mathbf{E}\left[W_t^2 \mid \mathcal{F}_{t-1}\right] = \left(\alpha_0 + \sum_{s=1}^{q} \alpha_s X_{t-s}^2\right) \mathbf{E}\left[W_t^2 \mid \mathcal{F}_{t-1}\right], & \text{if } t \geqslant q, \end{cases} \tag{7.266}$$

Now, since W_t is independent of $\mathcal{F}_{t-1}$,

$$\mathbf{E}\left[W_t^2 \mid \mathcal{F}_{t-1}\right] = \mathbf{E}\left[W_t^2\right] = \sigma_W^2. \tag{7.267}$$

for every $t > 1$. Combining (7.266) and (7.267), the desired (**??**) follows. □

Corollary 663 *We have*

$$\mathbf{D}^2\left[X_t \mid \mathcal{F}_{t-1}\right] = \begin{cases} \left(\alpha_0 + \sum_{s=1}^{t} \alpha_s X_{t-s}^2\right) \sigma_W^2. & \text{if } 1 \leqslant t \leqslant q-1, \\ \left(\alpha_0 + \sum_{s=1}^{q} \alpha_s X_{t-s}^2\right) \sigma_W^2, & \text{if } t \geqslant q. \end{cases} \tag{7.268}$$

Proof. Corollary follows immediately in light of (7.243) in Recall 654, (7.245) inProposition 655, and (**??**) in Proposition **??**. □

Recall 664 *Given a real random variable X on a probability space Ω with finite fourth order moment we call the* skewness *[resp.* kurtosis*] of X, and we denote it by the symbol $Skew_X$ [resp. $Kurt_X$], the standardized third [resp. fourth] order moment of X. That is*

$$Skew_X \stackrel{def}{=} \mathbf{E}\left[\left(\frac{X - \mathbf{E}[X]}{\mathbf{D}[X]}\right)^3\right] \qquad \textit{[resp. } Kurt_X \stackrel{def}{=} \mathbf{E}\left[\left(\frac{X - \mathbf{E}[X]}{\mathbf{D}[X]}\right)^4\right] \textit{].}$$

To proceed further, dealing also with the "fat tails" phenomenon, we need an additional regularity assumption on the strong white noise W.

Assumption 665 *Assume that the strong white noise W is a fourth order process, which implies*

$$\mathbf{E}\left[W_t^4\right] \equiv \mathbf{E}\left[(W_t - \mathbf{E}[W_t])^4\right] \equiv \mu_W^4 < \infty \tag{7.269}$$

and

$$Kurt_W \equiv \mathbf{E}\left[\left(\frac{W_t - \mathbf{E}[W_t]}{\mathbf{D}[W_t]}\right)^4\right] = \frac{\mu_W^4}{\sigma_W^4}, \tag{7.270}$$

for every $t \in \mathbb{N}$.

Proposizione 666 *Under Assumption 665, assume further that X_0 has finite fourth-order moment, that is $\mathbf{E}\left[X_0^4\right] < \infty$. Then, both the positive process σ and the ARCH(1) process X have finite fourth-order moment.*

Proof. By hypothesis X_0 has finite fourth-order moment. Hence,

$$\sigma_1^2 = \alpha_0 + \alpha_1 X_0^2$$

has finite second-order moment, that is σ_1 has finite fourth-order moment. On account of Proposition 651, W_1 is independent of σ_1. Therefore, W_1^4 is independent of σ_1^4. It follows

$$\mathbf{E}\left[\sigma_1^4 W_1^4\right] = \mathbf{E}\left[\sigma_1^4\right]\mathbf{E}\left[W_1^4\right] < \infty.$$

This implies that

$$X_1 = \sigma_1 W_1$$

has finite fourth-order moment. As a consequence,

$$\sigma_2^2 = \alpha_0 + \alpha_1 X_1^2 + \alpha_2 X_0^2$$

has finite second-order moment, that is σ_2 has finite fourth-order moment. We know that W_2 is independent of σ_2. Hence, also W_2^4 is independent of σ_2^4. It follows

$$\mathbf{E}\left[\sigma_2^4 W_2^4\right] = \mathbf{E}\left[\sigma_2^4\right]\mathbf{E}\left[W_2^4\right] < \infty.$$

This implies that

$$X_2 = \sigma_2 W_2$$

has finite fourth-order moment. Iterating this argument, we can prove that σ_t and X_t have finite fourth-order moment for every $t = 1, \dots, q-1$. Hence, assuming that σ_t and X_t have finite fourth-order moment for some $t \geqslant q$, essentially the same argument as the one above allows to prove that σ_{t+1} and X_{t+1} have finite fourth-order moment. By virtue of the Induction Principle, it then follows that σ_t and X_t have finite fourth-order moment for every $t \in \mathbb{N}$, which is the desired result. □

Proposizione 667 *Under Assumption 665, assume further that $q = 1$ and that X_0 is a stationary state with finite moment of the fourth-order, that is $\mathbf{E}\left[X_0^4\right] < \infty$. Then we have*

$$\mathbf{E}\left[X_t^4\right] = \left(\alpha_0^2 \frac{1+\alpha_1\sigma_W^2}{1-\alpha_1\sigma_W^2}\sum_{s=0}^{t-1}\alpha_1^{2s}\sigma_W^{4s}Kurt_W^s + \alpha_1^{2t}\sigma_W^{4(t-1)}Kurt_W^{t-1}\mathbf{E}\left[X_0^4\right]\right)\sigma_W^4 Kurt_W, \tag{7.271}$$

for every $t \in \mathbb{N}$. As as an immediate consequence,

$$\mathbf{E}\left[\sigma_t^4\right] = \alpha_0^2 \frac{1+\alpha_1\sigma_W^2}{1-\alpha_1\sigma_W^2}\sum_{s=0}^{t-1}\alpha_1^{2s}\sigma_W^{4s}Kurt_W^s + \alpha_1^{2t}\sigma_W^{4(t-1)}Kurt_W^{t-1}\mathbf{E}\left[X_0^4\right],$$

for every $t \in \mathbb{N}$.

Proof. By virtue of Propositon 666, both the positive process σ and the ARCH(1) process X have finite fourth-order moment. Thanks to Proposition 651, Equations (7.269) and (7.270) in Assumption 665, and Equation (7.259) in Proposition 659, we can write

$$\begin{aligned}
\mathbf{E}\left[X_1^4\right] &= \mathbf{E}\left[\sigma_1^4 W_1^4\right] = \mathbf{E}\left[\left(\alpha_0 + \alpha_1 X_0^2\right)^2 W_1^4\right] \\
&= \mathbf{E}\left[\left(\alpha_0 + \alpha_1 X_0^2\right)^2\right]\mathbf{E}\left[W_1^4\right] \\
&= \left(\alpha_0^2 + 2\alpha_1\alpha_0\mathbf{E}\left[X_0^2\right] + \alpha_1^2\mathbf{E}\left[X_0^4\right]\right)\sigma_W^4 Kurt_W \\
&= \left(\alpha_0^2 + 2\frac{\alpha_0^2\alpha_1\sigma_W^2}{1-\alpha_1\sigma_W^2} + \alpha_1^2\mathbf{E}\left[X_0^4\right]\right)\sigma_W^4 Kurt_W \\
&= \left(\alpha_0^2\frac{1+\alpha_1\sigma_W^2}{1-\alpha_1\sigma_W^2} + \alpha_1^2\mathbf{E}\left[X_0^4\right]\right)\sigma_W^4 Kurt_W.
\end{aligned}$$

As a consequence,

$$
\begin{aligned}
\mathbf{E}\left[X_2^4\right] &= \mathbf{E}\left[\sigma_2^4 W_2^4\right] = \mathbf{E}\left[\left(\alpha_0 + \alpha_1 X_1^2\right)^2 W_2^4\right] \\
&= \mathbf{E}\left[\left(\alpha_0 + \alpha_1 X_1^2\right)^2\right] \mathbf{E}\left[W_2^4\right] \\
&= \left(\alpha_0^2 + 2\alpha_0\alpha_1 \mathbf{E}\left[X_1^2\right] + \alpha_1^2 \mathbf{E}\left[X_1^4\right]\right) \sigma_W^4 Kurt_W \\
&= \left(\alpha_0^2 + 2\alpha_0\alpha_1 \frac{\alpha_0 \sigma_W^2}{1-\alpha_1\sigma_W^2} + \alpha_1^2 \left(\alpha_0^2 \frac{1+\alpha_1\sigma_W^2}{1-\alpha_1\sigma_W^2} + \alpha_1^2 \mathbf{E}\left[X_0^4\right]\right) \sigma_W^4 Kurt_W\right) \sigma_W^4 Kurt_W \\
&= \left(\alpha_0^2 \frac{1+\alpha_1\sigma_W^2}{1-\alpha_1\sigma_W^2} + \alpha_1^2 \left(\alpha_0^2 \frac{1+\alpha_1\sigma_W^2}{1-\alpha_1\sigma_W^2} + \alpha_1^2 \mathbf{E}\left[X_0^4\right]\right) \sigma_W^4 Kurt_W\right) \sigma_W^4 Kurt_W \\
&= \left(\alpha_0^2 \frac{1+\alpha_1\sigma_W^2}{1-\alpha_1\sigma_W^2} \left(1 + \alpha_1^2 \sigma_W^4 Kurt_W\right) + \alpha_1^4 \sigma_W^4 Kurt_W \mathbf{E}\left[X_0^4\right]\right) \sigma_W^4 Kurt_W
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbf{E}\left[X_3^4\right] \\
&= \mathbf{E}\left[\sigma_3^4 W_3^4\right] = \mathbf{E}\left[\left(\alpha_0 + \alpha_1 X_2^2\right)^2 W_3^4\right] \\
&= \mathbf{E}\left[\left(\alpha_0 + \alpha_1 X_2^2\right)^2\right] \mathbf{E}\left[W_3^4\right] \\
&= \left(\alpha_0^2 + 2\alpha_0\alpha_1 \mathbf{E}\left[X_2^2\right] + \alpha_1^2 \mathbf{E}\left[X_2^4\right]\right) \sigma_W^4 Kurt_W \\
&= \left(\alpha_0^2 + 2\alpha_0\alpha_1 \frac{\alpha_0 \sigma_W^2}{1-\alpha_1\sigma_W^2} + \alpha_1^2 \left(\alpha_0^2 \frac{1+\alpha_1\sigma_W^2}{1-\alpha_1\sigma_W^2} \left(1 + \alpha_1^2 \sigma_W^4 Kurt_W\right) + \alpha_1^4 \sigma_W^4 Kurt_W \mathbf{E}\left[X_0^4\right]\right) \sigma_W^4 Kurt_W\right) \sigma_W^4 Kurt_W \\
&= \left(\alpha_0^2 \frac{1+\alpha_1\sigma_W^2}{1-\alpha_1\sigma_W^2} \left(1 + \alpha_1^2 \sigma_W^4 Kurt_W + \alpha_1^4 \sigma_W^8 Kurt_W^2\right) + \alpha_1^6 \sigma_W^8 Kurt_W^2 \mathbf{E}\left[X_0^4\right]\right) \sigma_W^4 Kurt_W.
\end{aligned}
$$

In light of what shown above, Equation (**??**) is true for $t = 1, 2, 3$. Hence, assume that Equation (**??**) holds true for some $t \in \mathbb{N}$. We can then write

$$
\begin{aligned}
&\mathbf{E}\left[X_{t+1}^4\right] \\
&= \mathbf{E}\left[\sigma_t^4 W_t^4\right] = \mathbf{E}\left[\left(\alpha_0 + \alpha_1 X_t^2\right)^2 W_t^4\right] \\
&= \mathbf{E}\left[\left(\alpha_0 + \alpha_1 X_t^2\right)^2\right] \mathbf{E}\left[W_t^4\right] \\
&= \left(\alpha_0^2 + 2\alpha_0\alpha_1 \mathbf{E}\left[X_t^2\right] + \alpha_1^2 \mathbf{E}\left[X_t^4\right]\right) \sigma_W^4 Kurt_W \\
&= \left(\alpha_0^2 + 2\alpha_0\alpha_1 \frac{\alpha_0 \sigma_W^2}{1-\alpha_1\sigma_W^2} + \alpha_1^2 \left(\alpha_0^2 \frac{1+\alpha_1\sigma_W^2}{1-\alpha_1\sigma_W^2} \sum_{s=0}^{t-1} \alpha_1^{2s} \sigma_W^{4s} Kurt_W^s + \alpha_1^{2t} \sigma_W^{4(t-1)} Kurt_W^{t-1} \mathbf{E}\left[X_0^4\right]\right) \sigma_W^4 Kurt_W\right) \sigma_W^4 Ku \\
&= \left(\alpha_0^2 \frac{1+\alpha_1\sigma_W^2}{1-\alpha_1\sigma_W^2} + \alpha_0^2 \frac{1+\alpha_1\sigma_W^2}{1-\alpha_1\sigma_W^2} \sum_{s=0}^{t-1} \alpha_1^{2(s+1)} \sigma_W^{4(s+1)} Kurt_W^{s+1} + \alpha_1^{2(t+1)} \sigma_W^{4t} Kurt_W^t \mathbf{E}\left[X_0^4\right]\right) \sigma_W^4 Kurt_W \\
&= \left(\alpha_0^2 \frac{1+\alpha_1\sigma_W^2}{1-\alpha_1\sigma_W^2} + \alpha_0^2 \frac{1+\alpha_1\sigma_W^2}{1-\alpha_1\sigma_W^2} \sum_{s=1}^{t} \alpha_1^{2s} \sigma_W^{4s} Kurt_W^s + \alpha_1^{2(t+1)} \sigma_W^{4t} Kurt_W^t \mathbf{E}\left[X_0^4\right]\right) \sigma_W^4 Kurt_W \\
&= \left(\alpha_0^2 \frac{1+\alpha_1\sigma_W^2}{1-\alpha_1\sigma_W^2} \sum_{s=0}^{t} \alpha_1^{2s} \sigma_W^{4s} Kurt_W^s + \alpha_1^{2(t+1)} \sigma_W^{4t} Kurt_W^t \mathbf{E}\left[X_0^4\right]\right) \sigma_W^4 Kurt_W,
\end{aligned}
$$

which replicates (**??**) when replacing $t+1$ to t. By virtue of the Induction Principle we can conlude that (**??**) holds true for every $t \in \mathbb{N}$. $\square$

We need also another important assumption on the parameters of the ARCH(q) process X.

Assumption 668 *If $q = 1$, we require that*

$$1 - \alpha_1^2 \sigma_W^4 Kurt_W > 0. \tag{7.272}$$

Proposizione 669 *Under the assumptions of Lemma* **??** *and Assumption 668, we have that the $ARCH(1)$ process X is asymptotically fourth order stationary, and we have*

$$\mu_X^4 \equiv \lim_{t\to\infty} \mathbf{E}\left[(X_t - \mathbf{E}[X_t])^4\right] = \lim_{t\to\infty} \mathbf{E}\left[X_t^4\right] = \alpha_0^2 \frac{1+\alpha_1\sigma_W^2}{1-\alpha_1\sigma_W^2} \frac{\sigma_W^4 Kurt_W}{1-\alpha_1^2\sigma_W^4 Kurt_W}.$$

Proof. With reference to Equation (**??**) in Lemma **??**, we can write

$$\sum_{s=0}^{t-1} \alpha_1^{2s}\sigma_W^{4s} Kurt_W^s = \frac{1-\alpha_1^{2t}\sigma_W^{4t} Kurt_W^t}{1-\alpha_1^2\sigma_W^4 Kurt_W}. \tag{7.273}$$

On account of (7.273), Equation (**??**) in Lemma **??** yields

$$\begin{aligned}
\mathbf{E}\left[X_t^4\right] &= \left(\alpha_0^2 \frac{1+\alpha_1\sigma_W^2}{1-\alpha_1\sigma_W^2} \frac{1-\alpha_1^{2t}\sigma_W^{4t} Kurt_W^t}{1-\alpha_1^2\sigma_W^4 Kurt_W} + \alpha_1^{2t}\sigma_W^{4(t-1)} Kurt_W^{t-1}\mathbf{E}\left[X_0^4\right]\right)\sigma_W^4 Kurt_W \\
&= \alpha_0^2 \frac{1+\alpha_1\sigma_W^2}{1-\alpha_1\sigma_W^2} \frac{\sigma_W^4 Kurt_W}{1-\alpha_1^2\sigma_W^4 Kurt_W}\left(1-\alpha_1^{2t}\sigma_W^{4t} Kurt_W^t\right) + \alpha_1^{2t}\sigma_W^{4t} Kurt_W^t \mathbf{E}\left[X_0^4\right].
\end{aligned}$$

Therefore, by virtue of Assumption 668, we obtain

$$\lim_{t\to\infty} \mathbf{E}\left[X_t^4\right] = \alpha_0^2 \frac{1+\alpha_1\sigma_W^2}{1-\alpha_1\sigma_W^2} \frac{\sigma_W^4 Kurt_W}{1-\alpha_1^2\sigma_W^4 Kurt_W},$$

as desired. □

Proposizione 670 *Under the assumptions of Lemma* **??** *and Assumption 668, assume further that we have*

$$\mathbf{E}\left[X_{t_0}^4\right] = \alpha_0^2 \frac{1+\alpha_1\sigma_W^2}{1-\alpha_1\sigma_W^2} \frac{\sigma_W^4 Kurt_W}{1-\alpha_1^2\sigma_W^4 Kurt_W}, \tag{7.274}$$

for some $t_0 \in \mathbb{N}$. Then,

$$\mathbf{E}\left[X_t^4\right] = \alpha_0^2 \frac{1+\alpha_1\sigma_W^2}{1-\alpha_1\sigma_W^2} \frac{\sigma_W^4 Kurt_W}{1-\alpha_1^2\sigma_W^4 Kurt_W}, \tag{7.275}$$

for every $t \geqslant t_0$. As an immediate consequence,

$$\mathbf{E}\left[\sigma_t^4\right] = \alpha_0^2 \frac{1+\alpha_1\sigma_W^2}{1-\alpha_1\sigma_W^2} \frac{1}{1-\alpha_1^2\sigma_W^4 Kurt_W}, \tag{7.276}$$

for every $t \geqslant t_0$.

Proof. Following the argument in the proof of Lemma **??** and Propositions 670, it is possible to show that

$$\mathbf{E}\left[X_t^4\right] = \alpha_0^2 \frac{1+\alpha_1\sigma_W^2}{1-\alpha_1\sigma_W^2} \frac{\sigma_W^4 Kurt_W}{1-\alpha_1^2\sigma_W^4 Kurt_W}\left(1-\alpha_1^{2(t-t_0)}\sigma_W^{4(t-t_0)} Kurt_W^{t-t_0}\right) + \alpha_1^{2(t-t_0)}\sigma_W^{4(t-t_0)} Kurt_W^{t-t_0}\mathbf{E}\left[X_{t_0}^4\right]. \tag{7.277}$$

Therefore, replacing (7.274) in (7.277), we clearly obtain the desired (7.275). □

Corollary 671 *Under the assumptions of Lemma* **??** *and Assumption 668, assume further that Equation (7.274) in Proposition 670 holds true for some some* $t_0 \in \hat{n}$. *Then we have*

$$Kurt_{X_t} \equiv \mathbf{E}\left[\left(\frac{X_t - \mathbf{E}[X_t]}{\mathbf{D}[X_t]}\right)^4\right] = \left(1 - \alpha_1^2\sigma_W^4\right)\frac{\sigma_W^2 Kurt_W}{1 - \alpha_1^2\sigma_W^4 Kurt_W}, \tag{7.278}$$

for every $t \geqslant t_0$.

Proof. Clearly

$$Kurt_{X_t} = \frac{\mathbf{E}\left[X_t^4\right]}{\mathbf{E}\left[X_t^2\right]^2}. \tag{7.279}$$

On account of (7.259) in Proposition 659 and 7.275 in Proposition 670, we then obtain

$$\begin{aligned} Kurt_{X_t} &= \alpha_0^2 \frac{1 + \alpha_1\sigma_W^2}{1 - \alpha_1\sigma_W^2}\frac{\sigma_W^4 Kurt_W}{1 - \alpha_1^2\sigma_W^4 Kurt_W}\left(\frac{1 - \alpha_1\sigma_W^2}{\alpha_0\sigma_W^2}\right)^2 \\ &= \left(1 + \alpha_1\sigma_W^2\right)\left(1 - \alpha_1\sigma_W^2\right)\frac{\sigma_W^2 Kurt_W}{1 - \alpha_1^2\sigma_W^4 Kurt_W}, \end{aligned}$$

which is the desired (7.278). □

Definizione 672 *Under the assumptions of Lemma* **??**, *assume that Equation (7.274) holds true for some* $t_0 \in \hat{n}$. *Then we call* X_{t_0} *a kurtosis-stationary state of the* $ARCH(1)$ *process* X *and we say that* X *is kurtosis stationary.*

Teorema 673 *Under the assumptions of Lemma* **??**, *assume further that* W *is a Gaussian strong white noise and, for sake of simplicity, that* X_0 *is a kurtosis-stationary state of the* $ARCH(1)$ *process* X. *Then we have*

$$Kurt_X = 3\frac{1 - \alpha_1^2}{1 - 3\alpha_1^2} > 3. \tag{7.280}$$

Proof. The result immediately follows from Equation (7.278) in Corollary 671 and the Gaussianity assumption on W. □

Proposizione 674 *Under Assumption 665, assume further that* X_0 *has finite fourth-order moment, that is* $\mathbf{E}\left[X_0^4\right] < \infty$. *Then the process* $\left(X_t^2\right)_{t\in\hat{n}} \equiv X^2$ *is an* $AR(q)$ *process with initial states* $X_0^2, X_1^2, \ldots, X_{q-1}^2$. *More specifically, we have*

$$\begin{aligned} X_1^2 &= \tilde{\alpha}_0 + \tilde{\alpha}_1 X_0^2 + V_1, \\ X_2^2 &= \tilde{\alpha}_0 + \tilde{\alpha}_1 X_1^2 + \tilde{\alpha}_2 X_0^2 + V_2, \\ &\cdots \\ X_q^2 &= \tilde{\alpha}_0 + \sum_{s=1}^{q}\tilde{\alpha}_s X_{t-s}^2 + V_q, \\ X_t^2 &= \tilde{\alpha}_0 + \sum_{s=1}^{q}\tilde{\alpha}_s X_{t-s}^2 + V_t, \quad \forall t > q, \end{aligned} \tag{7.281}$$

where $\tilde{\alpha}_s \equiv \sigma_W^2\alpha_s$, *for every* $s = 0, 1, \ldots, q$ *and the process* $(V_t)_{t\in\mathbb{N}} \equiv V$ *given by*

$$V_t \overset{def}{=} \sigma_t^2\left(W_t^2 - \sigma_W^2\right), \tag{7.282}$$

for every $t \in \mathbb{N}$, *is a weak white noise. As a consequence, the optimal forecast for the square of an* $ARCH(q)$ *process* X *is an* $AR(q)$ *process.*

Proof. Combining Equations (7.217) and (7.218) in Definition 644, we can write

$$\begin{aligned}
&X_1^2 = \sigma_1^2 W_1^2 = \sigma_1^2\sigma_W^2 + \sigma_1^2\left(W_1^2 - \sigma_W^2\right) = \sigma_W^2\left(\alpha_0 + \alpha_1 X_0^2\right) + \sigma_1^2\left(W_1^2 - \sigma_W^2\right), \\
&X_2^2 = \sigma_2^2 W_2^2 = \sigma_2^2\sigma_W^2 + \sigma_2^2\left(W_2^2 - \sigma_W^2\right) = \sigma_W^2\left(\alpha_0 + \alpha_1 X_1^2 + \alpha_2 X_0^2\right) + \sigma_2^2\left(W_2^2 - \sigma_W^2\right), \\
&\dots \\
&X_q^2 = \sigma_q^2 W_q^2 = \sigma_q^2\sigma_W^2 + \sigma_q^2\left(W_q^2 - \sigma_W^2\right) = \sigma_W^2\left(\alpha_0 + \sum_{s=1}^{q}\alpha_s X_{t-s}^2\right) + \sigma_q^2\left(W_q^2 - \sigma_W^2\right) \\
&X_t^2 = \sigma_t^2 W_t^2 = \sigma_t^2\sigma_W^2 + \sigma_t^2\left(W_q^2 - \sigma_W^2\right) = \sigma_W^2\left(\alpha_0 + \sum_{s=1}^{t}\alpha_s X_{t-s}^2\right) + \sigma_t^2\left(W_q^2 - \sigma_W^2\right), \quad \forall t > q.
\end{aligned} \tag{7.283}$$

This yields the desired result, provided we can prove that the process $(V_t)_{t\in\mathbb{N}}$ given by (7.282) is a a weak white noise. Now, under Assumption 665, thanks to Proposition 666, we have that the process $(V_t)_{t\in\mathbb{N}}$ has finite second-order moment. On account of Proposition (651), which guarantees the independence of σ_t^2 and W_t^2 for every $t \in \mathbb{N}$, we have also

$$\mathbf{E}[V_t] = \mathbf{E}\left[\sigma_t^2\left(W_t^2 - \sigma_W^2\right)\right] = \mathbf{E}\left[\sigma_t^2\right]\mathbf{E}\left[\left(W_t^2 - \sigma_W^2\right)\right] = \mathbf{E}\left[\sigma_t^2\right]\left(\mathbf{E}\left[W_t^2\right] - \sigma_W^2\right) = 0.$$

Thanks to the same Proposition (651),

$$\mathbf{D}^2[V_t] = \mathbf{E}\left[\sigma_t^4\left(W_t^2 - \sigma_W^2\right)^2\right] = \mathbf{E}\left[\sigma_t^4\right]\mathbf{E}\left[\left(W_t^2 - \sigma_W^2\right)^2\right] \tag{7.284}$$

On the other hand,

$$\mathbf{E}\left[\left(W_t^2 - \sigma_W^2\right)^2\right] = \mathbf{E}\left[W_t^4 - 2W_t^2\sigma_W^2 + \sigma_W^4\right] = \mathbf{E}\left[W_t^4\right] - 2\sigma_W^2\mathbf{E}\left[W_t^2\right] + \sigma_W^4 = \sigma_W^4\left(Kurt_W - 1\right) \tag{7.285}$$

and, in case $q = 1$,

$$\mathbf{E}\left[\sigma_t^4\right] = \alpha_0^2\frac{1 + \alpha_1\sigma_W^2}{1 - \alpha_1\sigma_W^2}\frac{1}{1 - \alpha_1^2\sigma_W^4 Kurt_W}, \tag{7.286}$$

Combining (7.284)-(7.286), we obtain

$$\mathbf{D}^2[V_t] = \alpha_0^2\sigma_W^4\frac{1 + \alpha_1\sigma_W^2}{1 - \alpha_1\sigma_W^2}\frac{Kurt_W - 1}{1 - \alpha_1^2\sigma_W^4 Kurt_W},$$

for every $t \in \mathbb{N}$. In addition, on account of Equations (7.240) and (7.241) in Remark 650, we have

$$\mathbf{E}[V_s \mid \mathcal{F}_{t-1}] = \mathbf{E}\left[\sigma_s^2\left(W_s^2 - \sigma_W^2\right) \mid \mathcal{F}_{t-1}\right] = \sigma_s^2\left(W_s^2 - \sigma_W^2\right) \tag{7.287}$$

and

$$\mathbf{E}[V_t \mid \mathcal{F}_{t-1}] = \mathbf{E}\left[\sigma_t^2\left(W_t^2 - \sigma_W^2\right) \mid \mathcal{F}_{t-1}\right] = \sigma_t^2\mathbf{E}\left[W_t^2 - \sigma_W^2 \mid \mathcal{F}_{t-1}\right] = \sigma_t^2\left(\mathbf{E}\left[W_t^2\right] - \sigma_W^2\right) = 0, \tag{7.288}$$

for every $s, t \in \mathbb{N}$ such that $s < t$. Therefore, by virtue of the tower property of the conditional expectation operator,

$$\begin{aligned}
Cov(V_s, V_t) &= \mathbf{E}[V_s V_t] - \mathbf{E}[V_s]\mathbf{E}[V_t] = \mathbf{E}[V_s V_t] \\
&= \mathbf{E}[\mathbf{E}[V_s V_t \mid \mathcal{F}_{t-1}]] = \mathbf{E}[V_s\mathbf{E}[V_t \mid \mathcal{F}_{t-1}]] \\
&= 0,
\end{aligned}$$

for every $s, t \in \mathbb{N}$ such that $s < t$. This completes the proof, in case $q = 1$, that the process V is a weak white noise. In the end, to realize that the optimal forecast for the square of an ARCH(q) process X

is an AR(q) process it is sufficient to observe that from Equation (7.281), on account of (7.288), we obtain

$$
\begin{aligned}
&\mathbf{E}\left[X_1^2 \mid \mathcal{F}_0\right] = \mathbf{E}\left[\tilde{\alpha}_0 + \tilde{\alpha}_1 X_0^2 + V_1 \mid \mathcal{F}_0\right] = \tilde{\alpha}_0 + \tilde{\alpha}_1 X_0^2 + \mathbf{E}\left[V_1 \mid \mathcal{F}_0\right] = \tilde{\alpha}_0 + \tilde{\alpha}_1 X_0^2,\\
&\mathbf{E}\left[X_2^2 \mid \mathcal{F}_1\right] = \mathbf{E}\left[\tilde{\alpha}_0 + \tilde{\alpha}_1 X_1^2 + \tilde{\alpha}_2 X_0^2 + V_2 \mid \mathcal{F}_1\right] = \tilde{\alpha}_0 + \tilde{\alpha}_1 X_1^2 + \tilde{\alpha}_2 X_0^2 + \mathbf{E}\left[V_2 \mid \mathcal{F}_1\right] = \tilde{\alpha}_0 + \tilde{\alpha}_1 X_1^2 + \tilde{\alpha}_2 X_0^2,\\
&\dots\\
&\mathbf{E}\left[X_q^2 \mid \mathcal{F}_{q-1}\right] = \mathbf{E}\left[\tilde{\alpha}_0 + \sum_{s=1}^{q} \tilde{\alpha}_s X_{t-s}^2 + V_q \mid \mathcal{F}_{q-1}\right] = \tilde{\alpha}_0 + \sum_{s=1}^{q} \tilde{\alpha}_s X_{t-s}^2 + \mathbf{E}\left[V_q \mid \mathcal{F}_{q-1}\right] = \tilde{\alpha}_0 + \sum_{s=1}^{q} \tilde{\alpha}_s X_{t-s}^2,\\
&\mathbf{E}\left[X_t^2 \mid \mathcal{F}_{t-1}\right] = \mathbf{E}\left[\tilde{\alpha}_0 + \sum_{s=1}^{q} \tilde{\alpha}_s X_{t-s}^2 + V_t \mid \mathcal{F}_{t-1}\right] = \tilde{\alpha}_0 + \sum_{s=1}^{q} \tilde{\alpha}_s X_{t-s}^2 + \mathbf{E}\left[V_t \mid \mathcal{F}_{t-1}\right] = \tilde{\alpha}_0 + \sum_{s=1}^{q} \tilde{\alpha}_s X_{t-s}^2, \quad \forall t > q.
\end{aligned}
$$

□

7.14.3 Parameter Estimation

7.14.4 Forecasting

7.15 GARCH Processes

Parte II

Element of Time Series Analysis

Capitolo 8

Statistics on Time Series

Let $(x_t)_{t=1}^T \equiv x$ be an N-variate real time series of lenght T.

Definizione 675 *We call the* set of values *of the time series* x*, and we denote it by* $\{x_s\}_{t=1}^S$*, or more briefly by* $x_{\{\}}$*, the subset of* $\mathbb{R}^m$

$$\{x_s\}_{s=1}^S \equiv x_{\{\}} \overset{def}{=} \{x \in \mathbb{R} : x = x_t, \quad t = 1, \ldots, T\}.$$

Osservazione 676 *We clearly have*

$$1 \leqslant S = \left|x_{\{\}}\right| \leqslant T,$$

where $\left|x_{\{\}}\right|$ *is the cardinality of the set* $x_{\{\}}$*.*

Definizione 677 *For any* $s = 1, \ldots, S$*, we call the* frequency *of the value* x_s *in the times series* x *the positive integer* T_s *given by*

$$T_s \overset{def}{=} |\{t \in \{1, \ldots, T\} : x_t = x_s\}|.$$

We call the relative frequency *of the value* x_s *in the time series* x *the positive rational number* F_s *given by*

$$F_s \overset{def}{=} \frac{T_s}{T},$$

where T_s *is the frequency of* x_s *in* x*.*

Osservazione 678 *We have*

$$\textstyle\sum_{s=1}^S T_s = T \quad \textit{and} \quad \sum_{s=1}^S F_s = 1.$$

8.1 Histograms

When the set of values of a time series is large and the frequency of the values is small, it is customary to group the data values along the horizontal axis in disjoint *class intervals* or *bins* and then, with reference to the vertical axis, plot a bar representing the number or the proportion of the data whose value falls in each bin. Such a bar graph plot is called a *histogram*. More specifically, we talk of *frequency* or *relative frequency* histogram according to whether we plot the number or the proportion of the data whose value falls in each bin. The endpoints of a bin are called *bin boundaries*. We will adopt the *left-end inclusion convention* which stipulates that each bin contains its left-end but not its right-end boundary; that is each bin is a closed-open interval. Depending on the actual data distribution and the goals of the analysis, different numbers of bins, or equivalently different bin widths, may be appropriate. The number of bins chosen should be a trade-off between choosing too few large bins, at the cost of losing too much information about the actual data values in a bin, and choosing too many small bins, which would result in a too small number or proportion of the data values in each bin to obtain a discernible pattern. In many cases, the appropriate number of bins may be found only through various attempts. However, there are various useful guidelines and rules of thumb (see e.g. (**?**, ï¿½ 5.6)).

Let $T \geqslant 1$ and let $(x_t)_{t=1}^T \equiv x$ be a real time series. Write $\lceil\cdot\rceil : \mathbb{R} \to \mathbb{Z}$ for the *ceiling function*, which returns the smallest integer greater than its argument,

$$\lceil x \rceil \overset{\text{def}}{=} \min \{z \in \mathbb{Z} : x \leqslant z\}.$$

If the bins' width w is somehow assigned in advance then the number k of bins is determined by the formula

$$k = \left\lceil \frac{\max x - \min x}{w} \right\rceil,$$

where $\max x$ [resp. $\min x$] is the maximum [resp. minimum] data value.

If the bin's width is not assigned in advance a rather common choice of the number k of bins is determined by the *Tukey & Mosteller square-root rule* (1977)

$$k = \left\lceil \sqrt{T} \right\rceil,$$

where n is the size of the data set. Another common choice of the number k of bins is determined by the *Sturges rule* (1926)

$$k = \lceil 1 + \log_2 (T) \rceil,$$

where $\log_2 (\cdot)$ is the base 2 logarithm. The Sturges rule is derived from a binomial distribution and implicitly assumes an approximately normal distribution. The *Teller & Scott rice rule* (1985)

$$k = \left\lceil \sqrt[3]{2T} \right\rceil,$$

is also commonly used as a simple alternative to the Sturges rule. The *Doane rule* (1976)

$$k = \left\lceil 1 + \log_2 (T) + \log_2 \left(1 + |skew\,(x)| \sqrt{\frac{(T+1)(T+3)}{6(T-2)}} \right) \right\rceil,$$

where $skew\,(x)$ is the skewness of the data set x, is a modification of the Sturges rule which attempts to improve the choice of the number k of bins with non-normal data. Another choice of the number k of bins is determined by the *Wichard rule* (2008)

$$k = \left\lceil 1 + \ln (T) + \ln \left(1 + |kurt\,(x)| \sqrt{\frac{T}{6}} \right) \right\rceil,$$

where $\log (\cdot)$ is the natural logarithm and $kurt\,(x)$ is the kurtosis of the data set x. In addition, there are rules aimed to the determination of the bin width w. The *Scott normal reference rule* (1979)

$$w = \frac{3.49 \tilde{s}\,(x)}{\sqrt[3]{T}},$$

where $\tilde{s}\,(x)$ is the biased sample standard deviation of the data set x. The Scott normal reference rule is optimal for random samples of normally distributed data, in the sense that it minimizes the integrated mean squared error of the density estimate. The *Freedman & Diaconis rule* (1981)

$$w = \frac{2IQR}{\sqrt[3]{T}},$$

replacing the quantity 3.49σ of the Scott rule with the interquartile range IQR which is less sensitive than the standard deviation to outliers in data.

(see https://en.wikipedia.org/wiki/Histogram, https://estatistics.eu/what-is-statistics-graph-figures-histogram/).

An histogram is said to be *normal* if it has the following properties:

1. the highest (relative) frequency of data is attained at the middle bin;
2. the histogram is symmetric about the middle bin;
3. moving from the middle bin in either direction, the height decreases in such a way that the entire histogram is bell shaped.

8.2 Order Statistics, Range, Median, Quantiles

Let $T \geqslant 1$ and let $(x_t)_{t=1}^T \equiv x$ be a real time series.

Definizione 679 *We call the* order statistics *of the time series x, and we denote it by $\left(x_{(t)}\right)_{t=1}^T$, or more briefly by $x_{()}$, any non-decreasing reordering of x.*

Let $\left(x_{(t)}\right)_{t=1}^T \equiv x_{()}$ the order statistics of x.

Definizione 680 *For any $t = 1, \ldots, T$, we call the real number $x_{(t)}$ the tth order statistic of the time series x. In particular,*

$$x_{(1)} = \min\{x_1, \ldots, x_T\}$$

is the first order statistic,

$$x_{(T)} = \max\{x_1, \ldots, x_T\}$$

is the Tth order statistic.

Definizione 681 *We call the* range *of the time series x the real number*

$$\operatorname{range}(x) \stackrel{def}{=} x_{(T)} - x_{(1)}. \tag{8.1}$$

Definizione 682 *We call the* median *of the time series x the real number*

$$x_{1/2} \stackrel{def}{=} \begin{cases} x_{\left(\frac{T+1}{2}\right)} & \text{if } T \text{ is odd} \\ \frac{1}{2}\left(x_{\left(\frac{T}{2}\right)} + x_{\left(\frac{T}{2}+1\right)}\right) & \text{if } T \text{ is even} \end{cases}. \tag{8.2}$$

In light of (8.2), when the size T of the time series x is odd [resp. even] the median $x_{1/2}$ of x is the middle value [resp. the average of the two middle values] of the order statistic $x_{()}$ of x. Note that when T is odd the median $x_{1/2}$ is always in the set of values $x_{\{\}}$ of the time series x. On the contrary, when T is even $x_{1/2}$ may not be in $x_{\{\}}$. Note also that the median $x_{1/2}$, which makes use of only one or two middle values of the time series x, is not affected by the extreme values of x. On the contrary, the mean of x, which makes use of the whole data set, is affected by the extreme values. As a consequence, the median is a more accurate measure of the centrality of $x_{\{\}}$ than the empirical mean according to whether the extreme values of x matter. However, when $x_{\{\}}$ is roughly symmetric about the median, the mean attains a value close to the median.

Although it is interesting to consider whether the mean or median is more useful in a particular situation, we never need to restrict ourselves to consider only one of these statistics. They are both important, and should always be computed when a data set is summarized.

Let $T \geqslant 1$ and let $(x_t)_{t=1}^T \equiv x$ be a real time series of length T. Write $\lfloor\cdot\rfloor : \mathbb{R} \to \mathbb{Z}$ for the integer part or floor function, which returns the largest integer not greater than its argument,

$$\lfloor x\rfloor \stackrel{\text{def}}{=} \max\{z \in \mathbb{Z} : z \leqslant x\}.$$

Definizione 683 (Sheldon M. Ross - PennState Eberly College of Science STAT 414/415) *Given any rational number* $p \in (0,1)$*, set* $h \equiv (T+1)\,p$*. We call the* $100pth$ percentile *or* p quantile *of the time series* x *the real number*

$$x_p \overset{def}{=} x_{(\lfloor h \rfloor)} + (h - \lfloor h \rfloor)\left(x_{(\lfloor h \rfloor + 1)} - x_{(\lfloor h \rfloor)}\right). \tag{8.3}$$

In particular:

- *the* $25th$ *percentile or* 0.25*-quantile is also called the* first quartile *and is denoted by* x_{q_1}*;*
- *the* $50th$ *percentile or* 0.50*-quantile is also called the* second quartile*, it coincides with the median, and it is also denoted by* x_{q_2}*;*
- *the* $75th$ *percentile or* 0.75*-quantile is also called the* third quartile *and is denoted by* x_{q_3}*.*

Osservazione 684 *Setting* $p \equiv \frac{k}{T+1}$*, we have*

$$x_p = x_{(k)},$$

for every $k = 1, \ldots, T$*,*

Osservazione 685 *For any* $p \in (0,1)$*, the* $100pth$ *sample percentile* x_p *has at least* $\lfloor Tp \rfloor$ *data of the order statistic less than it and at least* $\lfloor T(1-p) \rfloor$ *data of the order statistic greater than it.*

In particular, the quartiles break up a data set into four parts. About the 25% of the data values are smaller than the first quartile, about the 25% of the data values lie between the first and second quartile, about the 25% are between the second and third quartile, and about the 25% are larger than the third quartile.

Definizione 686 *The* interquartile range*, acronym IQR, is the difference between the first and third quartile. In symbols*

$$IQR = x_{q_3} - x_{q_1}.$$

Esempio 687 *Consider the following list of the average Celsius temperatures recorded in Roma during the first half of September* 2017

$$\left(\begin{array}{ccccccccccccccc} 25.2 & 25.2 & 21.8 & 23.9 & 25.6 & 24.9 & 25.2 & 24.8 & 23.8 & 19.5 & 20.3 & 22.2 & 23.8 & 22.9 & 25.3 \end{array}\right)$$

Determine the quartiles and some percentiles.

Discussion. To compute some sample percentile we have first to consider the order statistic of size $T = 15$, that is

$$\left(\begin{array}{ccccccccccccccc} 19.5 & 20.3 & 21.8 & 22.2 & 22.9 & 23.8 & 23.8 & 23.9 & 24.8 & 24.9 & 25.2 & 25.2 & 25.2 & 25.3 & 25.6 \end{array}\right).$$

Now, setting $p = 1/2$, we have that $(T+1)\,p = 8 = \lfloor (T+1)\,p \rfloor$. Hence, the 50th percentile or second quartile or median is given by the 8th element of the order statistic, that is

$$x_{q_2} \equiv x_{1/2} = x_{(8)} \equiv 23.9.$$

Furthermore, $\lfloor Tp \rfloor = \lfloor 15/2 \rfloor = 7$ and $\lfloor T(1-p) \rfloor = \lfloor 15/2 \rfloor = 7$ In fact, there are 7 elements of the order statistic which are less than 23.9 and 7 elements which are greater than 23.9.

Setting $p = 1/4$, we have that $(T+1)\,p = 4 = \lfloor (T+1)\,p \rfloor$. Hence, the 25th percentile or first quartile is given by the 4th element of the order statistic, that is

$$x_{q_1} = x_{(4)} \equiv 22.2.$$

Furthermore, $\lfloor Tp \rfloor = \lfloor 15/4 \rfloor = 3$ and $\lfloor T(1-p) \rfloor = \lfloor 45/4 \rfloor = 11$. In fact, there are 3 elements of the order statistic which are less than 23.9 and 11 elements which are greater than 23.9.

Setting $p = 1/10$, we have that $(T+1)p = 16/10 > \lfloor (T+1)p \rfloor = 1$ and $(T+1)p - \lfloor (T+1)p \rfloor = 6/10$. Hence, the 10th percentile is given by

$$x_{0.1} = x_{(1)} + \frac{6}{10}\left(x_{(2)} - x_{(1)}\right) = 19.5 + 0.6\,(20.3 - 19.5) = 19.98 \simeq 20.$$

Furthermore, $\lfloor Tp \rfloor = \lfloor 15/10 \rfloor = 1$ and $\lfloor T(1-p) \rfloor = \lfloor 135/10 \rfloor = 13$. In fact, there is 1 element of the order statistic which is less than 20 and 14 elements which are greater than 22.

Setting $p = 1/(T+1) = 0.0625$, we have that $(T+1)p = 1 = \lfloor (T+1)p \rfloor$. Hence, the 6.25th percentile is given by the 1th element of the order statistic, that is

$$x_{1/16} = x_{(1)} = 19.5.$$

Furthermore, $\lfloor Tp \rfloor = \lfloor T/(T+1) \rfloor = \lfloor 15/16 \rfloor = 0$ and $\lfloor T(1-p) \rfloor = \lfloor T(T/(T+1)) \rfloor = \left\lfloor 15^2/16 \right\rfloor = 14$. In fact, there are 0 elements of the order statistic which are less than 19.5 and 14 elements which are greater than 19.5. $\square$

Exercise 688 *With reference to Example 687, consider the following list of the average Celsius temperatures recorded in Roma during the second half of September* 2017

$$\left(\ 25.7 \quad 22.0 \quad 21.2 \quad 20.7 \quad 18.3 \quad 19.5 \quad 20.7 \quad 20.8 \quad 20.6 \quad 21.7 \quad 20.5 \quad 21.5 \quad 22.2 \quad 21.4 \quad 20.6\ \right)$$

Determine the quartiles, the IQR, and some further percentiles of the average Celsius temperatures recorded in Roma throughout September 2017*.*

Note that Definition **??** is just one of the several non equivalent definitions that it is possible to find in the statistical literature. For instance, another definition is the following

Definizione 689 (Jay L. Devore & Kenneth N. Berk - Matlab) *Given any $k = 1, \ldots, T$, we call the* $100\,(k-0.5)\,/T$*th* percentile *or* $(k-0.5)\,/T$ quantile *of the time series x the real number $x_{(k)}$ which occupies the kth place in the order statistics $x_{()}$ of x.*

8.3 QQ-Plots

A QQ-plot, where QQ stands for quantile-quantile, is a graphical method in the Cartesian plane $\mathbb{R}^2$ to compare two probability distributions by plotting their quantiles against each other.

Ler X, Y real random variables with strictly increasing and continuous distribution function $F_X : \mathbb{R} \to \mathbb{R}$ and $F_Y : \mathbb{R} \to \mathbb{R}$, respectively.

Definizione 690 *We call the* QQ-plot of Y against X*, the representation in the Cartesian plane $\mathbb{R}^2$ of the parametric curve $QQ_{X,Y} : (0,1) \to \mathbb{R}^2$ given by*

$$QQ_{X,Y} \overset{def}{=} (x_q, y_q) \quad \forall q \in (0,1)\,,$$

where x_q and y_q fulfill the equations

$$F_X(x_q) = F_Y(y_q) = q.$$

Let $(x_t)_{t=1}^T \equiv x$ and $(y_t)_{t=1}^T \equiv y$ two time series of the same lenght T and let $\left(x_{(t)}\right)_{k=1}^T \equiv x_{()}$ and $\left(y_{(t)}\right)_{k=1}^T \equiv y_{()}$ be the order statistics of x and y, respectively. We agree that x [resp. y] is a time series generated from a reference [resp. tested] probability distribution.

Definizione 691 *We call the time series x [resp. y] the* reference *[resp.* test*] time series.*

Definizione 692 *We call the* QQ-plot of y against x, *the representation in the Cartesian plane $\mathbb{R}^2$ of the parametric point curve $QQ_{x,y} : \{1, \ldots, T\} \to \mathbb{R}^2$ given by*

$$QQ_{x,y}(k) \stackrel{def}{=} (x_t, y_t) \quad \forall t \in \{1, \ldots, T\}.$$

The shape of the pattern of a QQ-plot provides us with several information about the probability distributions which generate the reference and test time series.

Osservazione 693 *Assume that $Y = X$. Then the pattern of the corresponding QQ-plot $QQ_{X,Y}$ lies on the straigth line $y = x$. Conversely, assume that the pattern of the QQ-plot $QQ_{X,Y}$ lies on the straigth line $y = x$, then $Y = X$.*

Osservazione 694 *Assume that the test time series y is drawn by the same distribution generating the reference time series x. Then the pattern of the corresponding QQ-plot $QQ_{x,y}$ is very close to the straigth line $y = x$. Conversely, assume that the pattern of the QQ-plot $QQ_{x,y}$ is very close to the straigth line $y = x$, then the test time series y is likely drawn by the same distribution which generates the reference time series x.*

Osservazione 695 *Assume that $Y = aX + b$, where $a, b \in \mathbb{R}$ with $a \neq 0$. Then the pattern of the corresponding QQ-plot $QQ_{X,Y}$ lies on the straigth line $y = ax + b$. Conversely, assume that the pattern of the QQ-plot $QQ_{X,Y}$ lies on the straigth line $y = ax + b$, then $Y = aX + b$.*

Osservazione 696 *Assume that the test time series y is drawn by a distribution which is a linear transformation of the distribution generating the reference time series x. Then the pattern of the corresponding QQ-plot $QQ_{x,y}$ is very close to a straight line. Conversely, assume that the pattern of the QQ-plot $QQ_{x,y}$ is very close to a straight line, then the test time series y is likely drawn by a linear transformation of the distribution which generates the reference time series x.*

Osservazione 697 *Assume that X and Y belong to the same family of distribution but it is not true that $Y = aX + b$, for some $a, b \in \mathbb{R}$ with $a \neq 0$. Then the differences in the characterizing parameters of X and Y generally do not allow a straight line shape of the pattern of the QQ-plot $QQ_{X,Y}$.*

Osservazione 698 *Assume that the time series x and y are drawn by the same family of distribution but it is not true that the distribution which generates the test time series y is a linear transformation of the distribution generating the reference time series x. Then the differences in the characterizing parameters of the distributions which generate x and y do not generally allow a straight line shape of the pattern of the QQ-plot $QQ_{x,y}$.*

Osservazione 699 *Assume that the distribution of Y is more concentrated [resp. dispersed] than the distribution of X. Then pattern of the the corresponding QQ-plot $QQ_{X,Y}$ is flatter [resp. steeper] than the straigth line $y = x$. Conversely, assume that the pattern of the QQ-plot $QQ_{X,Y}$ is flatter [resp. steeper] than the straigth line $y = x$. Then the distribution of Y is more concentrated [resp. dispersed] than the distribution of X.*

Osservazione 700 *Assume that the test time series y is drawn by a distribution which is more concentrated [resp. dispersed] than the distribution generating the reference time series x. Then the pattern of the corresponding QQ-plot $QQ_{x,y}$ is flatter [resp. steeper] than the straigth line $y = x$. Conversely, assume that the pattern of the QQ-plot $QQ_{x,y}$ is flatter [resp. steeper] than the straigth line $y = x$. Then the test time series y is likely drawn by a distribution which is more concentrated [resp. dispersed] than the distribution generating the reference time series x.*

Osservazione 701 *Assume that the distribution of Y is more concentrated on the left [resp. right] than the distribution of X. Then the pattern of the corresponding QQ-plot $QQ_{X,Y}$ is arched downwards [resp. upwards]. Conversely, assume that the pattern of the QQ-plot $QQ_{X,Y}$ is arched downwards [resp. upwards]. Then the distribution of Y is more concentrated on the left [resp. right] than the distribution of X.*

Osservazione 702 *Assume that the test time series y is drawn from a distribution which is more concentrated on the left [resp. right] than the distribution generating x. Then the pattern of the corresponding QQ-plot $QQ_{X,y}$ is arched downwards [resp. upwards]. Conversely, assume that the pattern of the QQ-plot $QQ_{X,y}$ arched downwards [resp. upwards]. Then the test time series y is likely drawn from a distribution which is more concentrated on the left [resp. right] than the distribution of x.*

Osservazione 703 *Assume that the distribution of Y has lighter [resp. heavier] tails than the distribution of X. Then the pattern of the corresponding QQ-plot $QQ_{X,Y}$ is "S" [resp. reverse "S"]-shaped like the logistic [resp. logit]. Conversely, assume that the pattern of the QQ-plot $QQ_{X,Y}$ is "S" [resp. reverse "S"]-shaped like the logistic [resp. logit]. Then the distribution of Y has lighter [resp. heavier] tails than the distribution of X.*

Osservazione 704 *Assume that the test time series y is drawn from a distribution which has lighter [resp. heavier] tails than the distribution generating x. Then the pattern of the corresponding QQ-plot $QQ_{X,y}$ is "S" [resp. reverse "S"]-shaped like the logistic [resp. logit]. Conversely, assume that the pattern of the QQ-plot $QQ_{X,y}$ is "S" [resp. reverse "S"]-shaped like the logistic [resp. logit]. Then the test time series y is likely drawn from a distribution which has lighter [resp. heavier] tails than the distribution generating x.*

Note that exchanging the test and the reference data sets results in a reversion of the shape of the associated QQ-plot. In general a QQ-plot is more sensitive to deviances from normality in the tails of the data set than the corresponding P-P plot, whereas a P-P plot is more sensitive to deviances near the mean of the distribution than the corresponding QQ-plot. This makes a QQ-plot a better detector of outliers in the test data set that a P-P plot. The identification of outliers in a data set is an important goal. Therefore, QQ-plots are more frequently used than P-P plots to assess the normality of a data set.

8.4 Normal Time Series

A data set is said to be approximately normal if an histogram describing it is approximatively normal.

Osservazione 705 *If a data set is approximately normal, then the sample mean and the sample median of the data set are approximately equal.*

Empirical Rule Assume that a data set is approximately normal with sample mean $\bar{x}_n$ and sample standard deviation s_n. Then the following are true:

1. approximately 68% of the data values lie in the interval $[\bar{x}_n - s_n, \bar{x}_n + s_n]$;
2. approximately 95% of the data values lie in the interval $[\bar{x}_n - 2s_n, \bar{x}_n + 2s_n]$;
3. approximately 99.7% of the data values lie in the interval $[\bar{x}_n - 3s_n, \bar{x}_n + 3s_n]$.

8.5 Sample Mean

Let $T \geqslant 1$ and let $(x_t)_{t=1}^T \equiv x$ be a real time series.

Definizione 706 *We call the (*time) sample mean *of x the real number*

$$\bar{x}_T \overset{def}{=} \frac{1}{T}\sum_{t=1}^{T} x_t.$$

8.6 Sample Central Moments, Variance, Standard Deviation, Skewness, Kurtosis

Let $T \geqslant 1$, let $(x_t)_{t=1}^T \equiv x$ be a real time series, and let $\bar{x}_T$ be the sample mean of x.

Definizione 707 *Given any $n \in \mathbb{N}$ we call the* nth (time) sample central moment *of x the real number*

$$m_n(x) \overset{def}{=} \frac{1}{T}\sum_{t=1}^{T} (x_t - \bar{x}_T)^n.$$

Osservazione 708 *We have*

$$m_1(x) = 0.$$

Definizione 709 *Assume $T \geqslant 2$. We call the* unbiased 2nd (time) sample central moment *of x the positive number*

$$k_2(x) \overset{def}{=} \frac{T}{T-1} m_2(x). \tag{8.4}$$

Definizione 710 *Assume $T \geqslant 2$. We call the* unbiased (time) sample variance *of x the positive number*

$$s^2(x) \overset{def}{=} \frac{1}{T-1}\sum_{t=1}^{T} (x_t - \bar{x}_T)^2. \tag{8.5}$$

Osservazione 711 *We have*

$$s^2(x) = k_2(x).$$

Definizione 712 *We call the* "unbiased" standard deviation *of x the positive number*

$$s_x \overset{def}{=} \sqrt{s_x^2}. \tag{8.6}$$

Definizione 713 *We call the* biased sample variance *of x the positive number*

$$\tilde{s}_x^2 \overset{def}{=} \frac{1}{T}\sum_{t=1}^{T} (x_t - \bar{x}_T)^2.$$

Osservazione 714 *We have*

$$\tilde{s}_x^2 = m_2(x).$$

Osservazione 715 *We have*

$$\tilde{s}_x^2 = \frac{T-1}{T} s_x^2.$$

Definizione 716 *Assume $T \geqslant 3$. We call the* unbiased 3rd sample central moment *of x the real number*

$$k_3(x) \overset{def}{=} \frac{T^2}{(T-1)(T-2)} m_3(x). \tag{8.7}$$

See e.g. Joanes, D.N., Gill, C.A., Comparing Measures of Sample Skewness and Kurtosis, Journal of the Royal Statistical Society. Series D (The Statistician), *Vol. 47, No. 1 (1998), pp. 183-189. see also Cramer, H.,* Mathematical Methods of Statistics, *Princeton University Press, Princeton (1946).*

Definizione 717 *Assume $T \geqslant 3$. Then we call the* "unbiased" sample skewness *of x the real number*

$$skew(x) \overset{def}{=} \frac{k_3(x)}{k_2(x)^{3/2}}. \tag{8.8}$$

Proposizione 718 *We have*

$$skew(x) = \frac{\sqrt{T(T-1)}}{T-2} \frac{m_3(x)}{(m_2(x))^{3/2}} = \frac{T^2}{(T-1)(T-2)} \frac{m_3(x)}{s(x)^3}.$$

Definizione 719 *We call the* biased sample skewness *of x the real number*

$$\widetilde{skew}(x) \overset{def}{=} \frac{m_3(x)}{(m_2(x))^{3/2}}. \tag{8.9}$$

Osservazione 720 *We have*

$$\widetilde{skew}(x) = \frac{T-2}{\sqrt{T(T-1)}} skew(x). \tag{8.10}$$

Definizione 721 *Assume $T \geqslant 4$. We call the* unbiased 4th sample central moment *of x the real number*

$$k_4(x) \overset{def}{=} \frac{T^2\left((T+1)m_4(x) - 3(T-1)m_2^2(x)\right)}{(T-2)(T-3)(T-1)}. \tag{8.11}$$

See e.g. Joanes, D.N., Gill, C.A., Comparing Measures of Sample Skewness and Kurtosis, Journal of the Royal Statistical Society. Series D (The Statistician), *Vol. 47, No. 1 (1998), pp. 183-189. see also Cramer, H.,* Mathematical Methods of Statistics, *Princeton University Press, Princeton (1946).*

Definizione 722 *Assume $T \geqslant 4$. We call the* "unbiased" sample kurtosis *of x the positive number*

$$kurt(x) \overset{def}{=} \frac{k_4(x)}{k_2(x)^2} \tag{8.12}$$

We call the excess unbiased sample kurtosis *of x the positive number*

$$exkurt(x) \overset{def}{=} kurt(x) - 3. \tag{8.13}$$

Proposizione 723 *We have*

$$kurt(x) = \frac{T-1}{(T-2)(T-3)} \left((T+1)\frac{m_4(x)}{m_2(x)^2} - 3(T-1)\right). \tag{8.14}$$

Definizione 724 *We call the* biased sample kurtosis *of x the positive number*

$$\widetilde{kurt}(x) \overset{def}{=} \frac{m_4(x)}{m_2(x)^2}. \tag{8.15}$$

We call the excess biased sample kurtosis *of x the positive number*

$$\widetilde{exkurt}_x \overset{def}{=} \widetilde{kurt}_x - 3. \tag{8.16}$$

Osservazione 725 *We have*

$$\widetilde{kurt}(x) = \frac{(T-2)(T-3)}{(T-1)(T+1)} kurt(x) + 3\frac{(T-1)^2 - (T-2)(T-3)}{(T-1)(T+1)}.$$

8.7 Sample Autocovariance and Autocorrelation Function (SACF)

Let $T \in \mathbb{N}$ such that $T > 1$, and let $(x_t)_{t=1}^{T} \equiv x$ be a real time series.

Definizione 726 *We call the* sample autocovariance function *of x the map $\alpha_x : \{1, \ldots, T-1\} \to \mathbb{R}$ given by*

$$\alpha_x(\tau) \overset{def}{=} \frac{1}{T} \sum_{t=1}^{T-\tau} (x_t - \bar{x}_T)(x_{t+\tau} - \bar{x}_T), \quad \forall \tau \in \{1, \ldots, T-1\}.$$

The real number $\alpha_x(\tau)$ will be called the sample autocovariance *of x referred to the time shift τ*

Definizione 727 *We call the* sample autocorrelation function *of x the map $\rho_x : \{1, \ldots, T-1\} \to \mathbb{R}$ given by*

$$\rho_x(\tau) \overset{def}{=} \frac{\alpha_x(\tau)}{\alpha_x(0)}, \quad \forall \tau \in \{1, \ldots, T-1\},$$

where $\alpha_x(0)$ is the sample variance of x. The real number $\rho_x(\tau)$ will be called the sample autocorrelation *of x referred to the time shift τ.*

8.8 Sample Partial Autocorrelation Function (SPACF)

Introduce the vector

$$\rho_x^{(\tau)} \equiv (\rho_x(1), \ldots, \rho_x(\tau))^\intercal$$

and the matrix

$$\mathrm{P}_x(\tau) \equiv \begin{pmatrix} 1 & \rho_X(1) & \cdots & \rho_x(\tau-2) & \rho_x(\tau-1) \\ \rho_x(1) & 1 & \cdots & \rho_x(\tau-3) & \rho_x(\tau-2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_x(\tau-2) & \rho_x(\tau-3) & \cdots & 1 & \rho_x(1) \\ \rho_x(\tau-1) & \rho_x(\tau-2) & \cdots & \rho_x(1) & 1 \end{pmatrix}.$$

Definizione 728 *We call the* sample partial autocorrelation function *of x the map $\phi_x : \{1, \ldots, T-1\} \to \mathbb{R}$ given by*

$$\phi_x(\tau) = \phi_{\tau,\tau},$$

where $\phi_{\tau,\tau}$ is the τth entry of the vector $\phi^{(\tau)} \equiv (\phi_{\tau,1}, \ldots, \phi_{\tau,\tau})^\intercal$ which solves the equation

$$\mathrm{P}_x(\tau)\, \phi^{(\tau)} = \rho_x^{(\tau)}.$$

The real number $\phi_x(\tau)$ will be called the sample partial autocorrelation *of x referred to the time shift τ.*

Osservazione 729 *We have*

$$\phi_x(\tau) = \frac{\det(\mathrm{P}_x^*(\tau))}{\det(\mathrm{P}_x(\tau))},$$

where the matrix $\mathrm{P}_x^*(\tau)$ has the same columns of the matrix $\mathrm{P}_x(\tau)$ but the τth column which is replaced by the vector $\rho_x^{(\tau)}$.

8.9 Population Central Moments, Variance, Standard Deviation

Let $T \geqslant 1$ and let $(x_t)_{t=1}^T \equiv x$ be a real time series. Assume that the time series x is a realization of a weakly stationary process $(X_t)_{t\in\mathbb{T}} \equiv X$ with mean μ_X.

Definizione 730 *Given any $n \in \mathbb{N}$, we call the nth* population central moment *of x the real number*

$$\mu_n(x) \stackrel{def}{=} \frac{1}{T}\sum_{t=1}^{T}(x_t - \mu_x)^n .$$

Osservazione 731 *In general,*

$$\mu_1(x) \equiv \mu(x) \neq 0.$$

Definizione 732 *We call the* population variance *of x the real number*

$$\sigma_x^2 \stackrel{def}{=} \frac{1}{T}\sum_{t=1}^{T}(x_t - \mu_X)^2 . \tag{8.17}$$

We call the population standard deviation *of x the positive number*

$$\sigma_x \stackrel{def}{=} \sqrt{\sigma_x^2}. \tag{8.18}$$

Osservazione 733 *We have*

$$\sigma_x^2 = \frac{T-1}{T}s_x^2 \quad \text{and} \quad \sigma_x = \sqrt{\frac{T-1}{T}}s_x.$$

Capitolo 9

Time Series Models

Fixed any $N, T \in \mathbb{N}$, let $(x_t)_{t=0}^{T} \equiv x$ be an N-variate real time series of lenght T and let $(X_t)_{t\in\mathbb{T}} \equiv X$ be a stochastic process on a probability space Ω with time set $\mathbb{T}$ and state space $\mathbb{R}^N$, which is a "good model" for x (see 6.2). In general, for a simpler inference of the structure of X, it is common practice to postulate that X can be separated into some components.

- A *seasonal component* $(s_t)_{t\in\mathbb{T}} \equiv s$, which is a deterministic time function accounting for possible regular fuctuations shown by the path of the time series x in specific sub-periodsof $\mathbb{T}$, usually weeks, months, quarters, and years, of the observation period, generally due to environmental changes.
- A *mean* component $(m_t)_{t\in\mathbb{T}} \equiv m$, which is another deterministic time function accounting for a possible regular dynamics shown by the path of the time series x during the observation period, generally not attributable to seasonal changes. In particular, we speak of a *cyclic* component when the time series exhibits regular fluctuations in occasional sub-periods of the observation period; we speak of a *positive* or *negative trend* component according to whether the time series exhibits an increasing or decreasing pattern.
- A *noisy component*, $(N_t)_{t\in\mathbb{T}} \equiv N$, which is a stochastic process accounting for the erratic fluctuations shown by the path of the time series x during the observation period. In turn, the process N may contain a stochastic stationary trend, for instance when N is an autoregressive process, or N may contain a stochastic non-stationary trend, for instance when N is a random walk.

Note that while a time series may not contain a seasonal or mean component. It always contains a noisy component.

The most common decomposition of X is of additive type. That is we write

$$X_t = s_t + m_t + N_t, \tag{9.1}$$

for every $t \in \mathbb{T}$, and assume that N_t has mean 0. Nevertheless, especially with reference to models for time series in financial markets, it is often considered a multiplicative decomposition of the type

$$X_t = s_t \times m_t \times N_t, \tag{9.2}$$

for every $t \in \mathbb{T}$, under the assumption that N_t has mean 1. Note that we can always pass from a multiplicative to an additive [resp. from an additive to a multiplicative] decomposition by a logarithm [resp. exponential] transformation.

Loosely speaking, when dealing with a time series x, a rather standard procedure requires that we try to "guess" the structure of the deterministic components s and m. Possibly after having transformed the time series x into another time series $(\tilde{x}_t)_{t=0}^{T} \equiv \tilde{x}$ by some non linear transformation, typically a

Box-Cox transformation, which could be the logarithm or the square root transformation. Therafter, if the transformed time series $\tilde{x}$ presents some stochastic trend we try to remove it by differencing $\tilde{x}$ a number of times, until we obtain a detrended residual. In absence of a deterministic mean or seasonal component, we can then try to model the transformed series $\tilde{x}$ by a ARIMA(p,d,q) process, where d is the number of differences that we needed to obtain the detrended residual. On the other hand, a deterministic mean or a seasonal component in the transformed time series $\tilde{x}$ can be removed by a time-regression, for instance a linear or a quadratic time-regression, or by smoothing and deseasonalizing procedures, for instance moving averages and seasonal averages. Once the deterministic mean and seasonal components, m and s are found, with reference to a possible additive [resp. multiplicative] decomposition, we subtract the sum of m and s from $\tilde{x}$ [resp. divide $\tilde{x}$ by the product of the components s and m]. What remains is the noisy component of $\tilde{x}$ for which we will try to find a good model N, typically an ARMA(p,q) or GARCH(p,q) process, depending on the properties of the noisy component of $\tilde{x}$. Of course, there are countless cases which cannot completely dealt with this procedure, nevertheless it represents an undoubtedly useful general modus operandi.

9.1 Box Cox Transformations

Fixed any $T \in \mathbb{N}$, let $(x_t)_{t=0}^{T} \equiv x$ be a positive time series of lenght T

Definizione 734 *We call the Box-Cox transformed of x with exponent parameter λ the time series $(y_t(\lambda))_{t=0}^{T} \equiv y(\lambda)$ given by*

$$y_t(\lambda) \overset{def}{=} \begin{cases} \frac{x_t^{\lambda}-1}{\lambda} & \text{if } \lambda \neq 0 \\ \ln(x_t) & \text{if } \lambda = 0 \end{cases} \qquad \forall t = 0, \ldots, T.$$

Fixed any $T \in \mathbb{N}$, let $(x_t)_{t=0}^{T} \equiv x$ be a real time series of lenght T.

Definizione 735 *We call the Box Cox transformed of x with exponent parameter λ_1 and shift parameter λ_2 the time series $(y_t(\lambda_1,\lambda_2))_{t=0}^{T} \equiv y(\lambda_1,\lambda_2)$ given by*

$$y_t(\lambda_1,\lambda_2) \overset{def}{=} \begin{cases} \frac{(x_t+\lambda_2)^{\lambda_1}-1}{\lambda_1} & \text{if } \lambda_1 \neq 0,\ \lambda_2 > \min_{t=0,\ldots,T}\{x_t\} \\ \ln(x_t+\lambda_2) & \text{if } \lambda = 0,\ \lambda_2 > \min_{t=0,\ldots,T}\{x_t\} \end{cases} \qquad \forall t = 0, \ldots, T.$$

Note that the shift parameter λ_2 plays no other role than allowing the exponential transformation.

9.2 Backshift and Difference Operators

Let $(X_t)_{t\in\mathbb{Z}} \equiv X$ be a stochastic process on a probability space Ω with time set $\mathbb{T} \equiv \mathbb{Z}$ and state space $\mathbb{R}^N$.

Definizione 736 *We call the* identity operator *the operator I which transforms the process $(X_t)_{t\in\mathbb{Z}}$ in the process $(IX_t)_{t\in\mathbb{Z}}$ given by*

$$IX_t \overset{def}{=} X_t, \quad \forall t \in \mathbb{Z}. \tag{9.3}$$

Definizione 737 *We call the* 1st-order backshift operator *the operator B which transforms the process $(X_t)_{t\in\mathbb{Z}}$ in the process $(BX_t)_{t\in\mathbb{Z}}$ given by*

$$BX_t \overset{def}{=} X_{t-1}, \quad \forall t \in \mathbb{Z}. \tag{9.4}$$

For any fixed $n \in \mathbb{N}$ such that $n > 1$, we call the nth-order backshift operator *the operator B^n which transforms the process $(X_t)_{t\in\mathbb{Z}}$ in the process $(B^n X_t)_{t\in\mathbb{Z}}$ given by*

$$B^n X_t \overset{def}{=} BB^{n-1}X_t, \quad \forall t \in \mathbb{Z},$$

where $B^1 \equiv B$ and $B^0 \equiv I$.

Osservazione 738 *Fixed any $n \in \mathbb{N}$ such that $n > 1$, we have*

$$B^n X_t = X_{t-n},$$

for every $t \in \mathbb{Z}$.

Definizione 739 *We call the* first order difference operator *the operator Δ which transforms the process $(X_t)_{t\in\mathbb{Z}}$ in the process $(\Delta X_t)_{t\in\mathbb{Z}}$ given by*

$$\Delta X_t \overset{def}{=} X_t - X_{t-1}, \quad \forall t \in \mathbb{Z}. \tag{9.5}$$

Fixed any $n \in \mathbb{N}$ such that $n > 1$, we call the nth-order difference operator *the operator Δ^n which transforms the process $(X_t)_{t\in\mathbb{Z}}$ in the process $(\Delta^n X_t)_{t\in\mathbb{Z}}$ given by*

$$\Delta^n X_t \overset{def}{=} \Delta\Delta^{n-1}X_t, \quad \forall t \in \mathbb{Z}, \tag{9.6}$$

where $\Delta^1 \equiv \Delta$ and $\Delta^0 \equiv I$. We call the n-lag difference operator *the operator Δ_n which transforms the process $(X_t)_{t\in\mathbb{Z}}$ in the process $(\Delta_n X_t)_{t\in\mathbb{Z}}$ given by*

$$\Delta_n X_t \overset{def}{=} X_t - X_{t-n}, \quad \forall t \in \mathbb{Z},$$

where $\Delta_1 \equiv \Delta$.

Osservazione 740 *We have*

$$\Delta X_t = (I - B)\, X_t, \tag{9.7}$$

for every $t \in \mathbb{Z}$.

Proposizione 741 *Fixed any $n \in \mathbb{N}$, we have*

$$\Delta^n X_t = (I - B)^n\, X_t = \sum_{k=0}^{n} (-1)^k \binom{n}{k} B^k X_t, \tag{9.8}$$

for every $t \in \mathbb{Z}$.

Proof. Equation (9.8) is clearly true for $n = 1$. Assume it is true for some $n \in \mathbb{N}$, such that $n > 1$, and consider the case $n + 1$. By virtue of the inductive hypothesis, we can write

$$\begin{aligned}
\Delta^{n+1} X_t &= \Delta\Delta^n X_t = (I - B) \sum_{k=0}^{n} (-1)^k \binom{n}{k} B^k X_t \\
&= \sum_{k=0}^{n} (-1)^k \binom{n}{k} B^k X_t - B \left(\sum_{k=0}^{n} (-1)^k \binom{n}{k} B^k X_t \right) \\
&= \sum_{k=0}^{n} (-1)^k \binom{n}{k} B^k X_t - \sum_{k=0}^{n} (-1)^k \binom{n}{k} B^{k+1} X_t \\
&= X_t + \sum_{k=1}^{n} (-1)^k \binom{n}{k} B^k X_t + \sum_{k=0}^{n-1} (-1)^{k+1} \binom{n}{k} B^{k+1} X_t + B^{n+1} X_t \\
&= X_t + \sum_{k=1}^{n} (-1)^k \binom{n}{k} B^k X_t + \sum_{k=1}^{n} (-1)^k \binom{n}{k-1} B^k X_t + B^{n+1} X_t \\
&= X_t + \sum_{k=1}^{n} (-1)^k \left(\binom{n}{k} + \binom{n}{k-1} \right) B^k X_t + B^{n+1} X_t \\
&= X_t + \sum_{k=1}^{n} (-1)^k \binom{n+1}{k} B^k X_t + B^{n+1} X_t \\
&= \sum_{k=0}^{n+1} (-1)^k \binom{n+1}{k} B^k X_t.
\end{aligned}$$

This proves that Equation (9.8) is also true in the case $n + 1$. Thanks to the Induction Principle, we can conclude that Equation (9.8) is true for every $n \in \mathbb{N}$. □

Osservazione 742 *Fixed any $n \in \mathbb{N}$ such that $n > 1$, we have*

$$\Delta_n X_t = (I - B^n) X_t, \tag{9.9}$$

for every $t \in \mathbb{Z}$.

9.3 Dealing with Mean and Seasonal Component

Let $(X_t)_{t\in\mathbb{Z}} \equiv X$ be a stochastic process on a probability space Ω with time set $\mathbb{T} \equiv \mathbb{Z}$ and state space $\mathbb{R}^N$.

Proposizione 743 *Assume that X has no additive seasonal component, that is we can write*

$$X_t = m_t + N_t, \tag{9.10}$$

for every $t \in \mathbb{Z}$, where $(m_t)_{t\in\mathbb{Z}}$ and $(N_t)_{t\in\mathbb{Z}}$ are the mean and noisy component of X, respectively. Moreover, assume that for some $x_0, \beta \in \mathbb{R}^m$ we have

$$m_t = x_0 + \beta t, \tag{9.11}$$

for every $t \in \mathbb{Z}$. Then the differenced process $(\Delta X_t)_{t\in\mathbb{Z}}$ fulfills

$$\Delta X_t = \beta + \Delta N_t, \tag{9.12}$$

for every $t \in \mathbb{Z}$, where $(\Delta N_t)_{t\in\mathbb{Z}}$ is the noisy component of $(\Delta X_t)_{t\in\mathbb{Z}}$. More generally, assume that for some $n \in \mathbb{N}$ and $x_0, \beta_1, \ldots, \beta_n \in \mathbb{R}^N$ we have

$$m_t = x_0 + \sum_{k=1}^{n} \beta_k t^k, \tag{9.13}$$

for every $t \in \mathbb{Z}$. Then the nth-order differenced process $(\Delta^n X_t)_{t\in\mathbb{Z}}$ fulfills

$$\Delta^n X_t = n!\beta_n + \Delta^n N_t, \tag{9.14}$$

for every $t \in \mathbb{Z}$, where $(\Delta^n N_t)_{t\in\mathbb{Z}}$ is the noisy component of $(\Delta^n X_t)_{t\in\mathbb{Z}}$.

Note that Proposition 743 presents a method to remove the mean component from the stochastic process X, when the mean component is modeled by a polynomial of order n, but does not deal with the problem of the determination of the coefficients of the polynomial. This can be done by a standard least square procedure.

Let $(x_t)_{t=1}^{T} \equiv x$ an N-variate real time series of lenght T which is modeled by X. Then, estimates $\hat{x}_0, \hat{\beta}_1, \ldots, \hat{\beta}_n$ of the coefficients $x_0, \beta_1, \ldots, \beta_n$ in Equation (9.13) can be obtained by the minimization of the error function

$$\mathrm{Err}\left(x_0, \beta_1, \ldots, \beta_n\right) \stackrel{\text{def}}{=} \frac{1}{2} \sum_{t=1}^{T} \left\| x_t - m_t \right\|^2.$$

Note also that the order n of the polynomial which models the mean component cannot be estimated, but has to be choosen ex ante, for example by inspecting the time plot of x.

9.3.1 Moving Averages

Let $(X_t)_{t\in\mathbb{Z}} \equiv X$ be a stochastic process on a probability space Ω with time set $\mathbb{T} \equiv \mathbb{Z}$ and state space $\mathbb{R}^N$.

Definizione 744 *Fixed any $q \in \mathbb{N}$, we call the* two side sample moving average *or* rolling average *or* running average *of order q the process $\left(\overleftrightarrow{M}_t\right)_{t\in\mathbb{Z}}$ given by*

$$\overleftrightarrow{M}_t\left(q\right) \stackrel{def}{=} \frac{1}{2q+1} \sum_{s=-q}^{q} X_{t+s}, \quad \forall t \in \mathbb{Z}. \tag{9.15}$$

Definizione 745 *Fixed any $q \in \mathbb{N}$, we call the* one side sample moving average *or* rolling average *or* running average *of order q the process $\left(\overrightarrow{M}_t\right)_{t\in\mathbb{Z}}$ given by*

$$\overrightarrow{M}_t\left(q\right) \stackrel{def}{=} \frac{1}{q+1} \sum_{s=0}^{q} X_{t-s}, \quad \forall t \in \mathbb{Z}. \tag{9.16}$$

Osservazione 746 *We have*

$$\overrightarrow{M}_{t+1}\left(q\right) = \overrightarrow{M}_t\left(q\right) + \frac{1}{q+1}\left(X_{t+1} - X_{t-q}\right), \tag{9.17}$$

for every $t \in \mathbb{Z}$.

Proposizione 747 *Assume that X has no additive seasonal component, that is we can write*

$$X_t = m_t + N_t, \tag{9.18}$$

for every $t \in \mathbb{Z}$, *where* $(m_t)_{t\in\mathbb{Z}}$ *and* $(N_t)_{t\in\mathbb{Z}}$ *are the mean and noisy component of* X, *respectively. Moreover, assume that for some* $q \in \mathbb{N}$ *the mean component is* locally linear, *that is for some* $x_0(q), \beta(q) \in \mathbb{R}^m$ *we have*

$$m_s = x_0(q) + \beta(q)\, s, \tag{9.19}$$

for every $s \in [t-q, t+q]$. *Then the two side moving average process* $\left(\overleftrightarrow{M}_t(q)\right)_{t\in\mathbb{Z}}$ *fulfills*

$$\overleftrightarrow{M}_t(q) = m_t + \frac{1}{2q+1}\sum_{s=-q}^{q} N_{t+s}, \tag{9.20}$$

for every $t \in \mathbb{Z}$.

Proof. A straightforward computation yields

$$\overleftrightarrow{M}_t(q) = \frac{1}{2q+1}\sum_{s=-q}^{q} X_{t+s} = \frac{1}{2q+1}\sum_{s=-q}^{q}(m_{t+s}+N_{t+s}) = \frac{1}{2q+1}\sum_{s=-q}^{q} m_{t+s} + \frac{1}{2q+1}\sum_{s=-q}^{q} N_{t+s}. \tag{9.21}$$

Now, by virtue of Equation (9.19), we can write

$$\begin{aligned}\frac{1}{2q+1}\sum_{s=-q}^{q} m_{t+s} &= \frac{1}{2q+1}\sum_{s=-q}^{q}(x_0(q)+\beta(q)(t+s))\\ &= \frac{1}{2q+1}\left(\sum_{s=-q}^{q} x_0(q) + \sum_{s=-q}^{q}\beta(q)\,t + \sum_{s=-q}^{q}\beta(q)\,s\right)\\ &= x_0(q)+\beta(q)\,t. \end{aligned} \tag{9.22}$$

Combining (9.21) and (9.22) the desired result follows. □

Corollary 748 *The two side moving average process* $\left(\overleftrightarrow{M}_t(q)\right)_{t\in\mathbb{Z}}$ *is an unbiased estimator of the mean component* $(m_t)_{t\in\mathbb{Z}}$ *of* X.

Proposizione 749 *Assume that* X *has no additive seasonal component, that is we can write*

$$X_t = m_t + N_t, \tag{9.23}$$

for every $t \in \mathbb{Z}$, *where* $(m_t)_{t\in\mathbb{Z}}$ *and* $(N_t)_{t\in\mathbb{Z}}$ *are the mean and noisy component of* X, *respectively. Moreover, assume that for some* $q \in \mathbb{N}$ *the mean component is* locally linear, *that is for some* $x_0(q), \beta(q) \in \mathbb{R}^m$ *we have*

$$m_s = x_0(q) + \beta(q)\, s, \tag{9.24}$$

for every $s \in [t-q, t]$. *Then the one side sample moving average process* $\left(\overrightarrow{M}_t(q)\right)_{t\in\mathbb{Z}}$ *fulfills*

$$\overrightarrow{M}_t(q) = m_t - \frac{q}{2}\beta(q) + \frac{1}{q+1}\sum_{s=0}^{q} N_{t-s}, \tag{9.25}$$

for every $t \in \mathbb{Z}$.

Proof. By a straightforward computation,

$$\overrightarrow{M}_t(q) = \frac{1}{q+1}\sum_{s=0}^{q}(m_{t-s}+N_{t-s}) = \frac{1}{q+1}\sum_{s=0}^{q} m_{t-s} + \frac{1}{q}\sum_{s=0}^{q} N_{t-s}. \tag{9.26}$$

On the other hand, by virtue of Equation (9.24),

$$
\begin{aligned}
\frac{1}{q+1}\sum_{s=0}^{q} m_{t-s} &= \frac{1}{q+1}\sum_{s=0}^{q}\left(x_0(q)+\beta(q)(t-s)\right)\\
&= \frac{1}{q+1}\left(\sum_{s=0}^{q} x_0(q)+\sum_{s=0}^{q}\beta(q)\,t-\sum_{s=0}^{q}\beta(q)\,s\right)\\
&= x_0(q)+\beta(q)\,t-\frac{1}{q+1}\frac{q(q+1)}{2}\beta(q)\\
&= x_0(q)+\beta(q)\,t-\frac{q}{2}\beta(q)\,. \qquad (9.27)
\end{aligned}
$$

Combining (9.26) and (9.26), Equation (9.25) follows.

Proposizione 750 *Assume that $m=1$ and X has no multiplicative seasonal component, that is we can write*

$$X_t = m_t \times N_t, \qquad (9.28)$$

for every $t\in\mathbb{Z}$, where $(m_t)_{t\in\mathbb{T}}$ and $(N_t)_{t\in\mathbb{Z}}$ are the mean and noisy component of X, respectively. Moreover, assume that for some $x_0,\beta\in\mathbb{R}$ we have

$$m_t = x_0\exp(\beta t)\,, \qquad (9.29)$$

Then the logarithm process $(\ln(X_t))_{t\in\mathbb{Z}}$ fulfills

$$\ln(X_t) = \ln(x_0)+\beta t+\ln(N_t) \qquad (9.30)$$

for every $t\in\mathbb{Z}$, where $(\ln(N_t))_{t\in\mathbb{Z}}$ is the noisy component of $(\ln(X_t))_{t\in\mathbb{Z}}$.

Note that after the logarithm trasform which yields Equation (9.30) estimates $\hat{x}_0$ and $\hat{\beta}$ of x_0 and β can be obtained by a standard least square procedure.

Proposizione 751 *Assume that X has no additive mean component, that is we can write*

$$X_t = s_t + N_t, \qquad (9.31)$$

for every $t\in\mathbb{Z}$, where $(s_t)_{t\in\mathbb{Z}}$ and $(N_t)_{t\in\mathbb{Z}}$ are the seasonal and noisy component of X, respectively. Moreover, assume that for some $p\in\mathbb{N}$ we have

$$s_t = s_{k \pmod p}, \qquad (9.32)$$

for every $t\in\mathbb{Z}$ and $k=1,\dots,p$, and

$$\sum_{t=1}^{p} s_t = 0. \qquad (9.33)$$

Then the p-lag differenced process $(\Delta_p X_t)_{t\in\mathbb{Z}}$ fulfills

$$\Delta_p X_t = \Delta_p N_t,$$

for every $\in\mathbb{Z}$, where $(\Delta_p N_t)_{t\in\mathbb{Z}}$ is the noisy component of $(\Delta_p X_t)_{t\in\mathbb{Z}}$.

Definizione 752 *The positive integer p introduced in Proposition 751 is called the* period *of the seasonal component $(s_t)_{t\in\mathbb{T}}$.*

Proposizione 753 *Assume that $m = 1$ and X has both additive seasonal and mean component, that is we can write*

$$X_t = s_t + m_t + N_t, \tag{9.34}$$

for every $t \in \mathbb{Z}$, where $(s_t)_{t\in\mathbb{Z}}$, $(m_t)_{t\in\mathbb{Z}}$, and $(N_t)_{t\in\mathbb{Z}}$ are the seasonal, mean, and noisy component of X, respectively. Let $p \in \mathbb{N}$ be the period of the seasonal component, let $(x_t)_{t=1}^{T} \equiv x$ be a real time series of lenght T which is modeled by X, and let $(n_t)_{t=1}^{T} \equiv n$ the time series of the realizations of $(N_t)_{t\in\mathbb{Z}}$ corresponding to x. Write

$$x_{j,k} \equiv x_{p(j-1)+k} \quad \text{and} \quad n_{j,k} \equiv n_{p(j-1)+k}, \quad \forall j = 1, \ldots, \lfloor T/p \rfloor, \ k = 1, \ldots, p,$$

where $\lfloor\cdot\rfloor : \mathbb{R} \to \mathbb{Z}$ is the ceiling function. Define

$$\hat{c}_j \overset{def}{=} \frac{1}{p} \sum_{k=1}^{p} x_{j,k}, \quad \forall j = 1, \ldots, \lfloor T/p \rfloor.$$

Then the real numbers

$$\hat{s}_k \overset{def}{=} \frac{1}{\lfloor T/p \rfloor} \sum_{j=1}^{\lfloor T/p \rfloor} \left(x_{j,k} - \hat{c}_j \right), \quad \forall k = 1, \ldots, p,$$

are an estimate of the seasonal component $(s_t)_{t\in\mathbb{Z}}$ in the jth period $j = 1, \ldots, \lfloor T/p \rfloor$ and

$$\hat{n}_{j,k} \overset{def}{=} x_{j,k} - \hat{c}_j - \hat{s}_k, \quad \forall j = 1, \ldots, \lfloor T/p \rfloor, \ k = 1, \ldots, p,$$

are an estimate of the noisy component.

Proposizione 754 *Assume that $N = 1$ and X has both additive seasonal and mean component, that is we can write*

$$X_t = s_t + m_t + N_t, \tag{9.35}$$

for every $t \in \mathbb{Z}$, where $(s_t)_{t\in\mathbb{Z}}$, $(m_t)_{t\in\mathbb{Z}}$, and $(N_t)_{t\in\mathbb{Z}}$ are the seasonal, mean, and noisy component of X, respectively. Let $p \in \mathbb{N}$ be the period of the seasonal component and let $(x_t)_{t=1}^{T}$ be a real time series of lenght T which is modeled by X. Distinguish two cases:

1. *p is odd, that is $p = 2q + 1$ for some $q \in \mathbb{N}$;*
2. *p is even, that is $p = 2q$ for some $q \in \mathbb{N}$.*

In the first case, consider the estimators

$$\overleftrightarrow{M}_t(q) \overset{def}{=} \frac{1}{2q+1} \sum_{s=-q}^{q} X_{t+s}, \quad \forall t = p+1, \ldots, p\left(\lfloor T/p \rfloor - 1\right),$$

the estimates

$$\hat{c}_t \equiv \frac{1}{2q+1} \sum_{s=-q}^{q} x_{t+s} \equiv \frac{1}{p} \left(x_{t-\left(\frac{p-1}{2}\right)} + \cdots + x_t + \cdots + x_{t+\left(\frac{p-1}{2}\right)} \right), \quad \forall t = p+1, \ldots, p\left(\lfloor T/p \rfloor - 1\right),$$

and set

$$\mu_k \overset{def}{=} \begin{cases} \frac{1}{\lfloor T/p \rfloor - 1} \sum\limits_{j=2}^{\lfloor T/p \rfloor} \left(x_{k+p(j-1)} - \hat{c}_{k+p(j-1)} \right) & \text{if } k = 1, \ldots, \frac{p-1}{2} \\ \frac{1}{\lfloor T/p \rfloor - 1} \sum\limits_{j=1}^{\lfloor T/p \rfloor - 1} \left(x_{k+p(j-1)} - \hat{c}_{k+p(j-1)} \right) & \text{if } k = \frac{p+1}{2}, \ldots, p \end{cases}.$$

In the second case, consider the estimator

$$\overleftrightarrow{M}_t(q) \stackrel{def}{=} \frac{1}{2q}\left(\frac{1}{2}X_{t-q} + \sum_{s=-q+1}^{q-1} X_{t+s} + \frac{1}{2}X_{t+q}\right), \quad \forall t = p+1, \ldots, p\left(\lfloor T/p \rfloor - 1\right),$$

the estimates

$$\hat{c}_t \equiv \frac{1}{2q}\left(\frac{1}{2}x_{t-q} + \sum_{s=-q+1}^{q-1} x_{t+s} + \frac{1}{2}x_{t+q}\right) \equiv \frac{1}{p}\left(x_{t-\frac{p}{2}} + \cdots + x_t + \cdots + x_{t+\frac{p}{2}}\right), \quad \forall t = p+1, \ldots, p\left(\lfloor T/p \rfloor - 1\right),$$

and set

$$\mu_k \stackrel{def}{=} \begin{cases} \frac{1}{\lfloor T/p \rfloor - 1} \sum_{j=2}^{\lfloor T/p \rfloor} \left(x_{k+p(j-1)} - \hat{c}_{k+p(j-1)}\right) & \text{if } k = 1, \ldots, \frac{p}{2} \\ \frac{1}{\lfloor T/p \rfloor - 1} \sum_{j=1}^{\lfloor T/p \rfloor - 1} \left(x_{k+p(j-1)} - \hat{c}_{k+p(j-1)}\right) & \text{if } k = \frac{p}{2}+1, \ldots, p \end{cases}.$$

In both cases we have that

$$\hat{s}_k \stackrel{def}{=} \mu_k - \frac{1}{p}\sum_{\ell=1}^{p} \mu_\ell, \quad \forall k = 1, \ldots, p$$

constitutes an estimate of the seasonal component of x in the period p. More generally, an estimate $(\hat{s}_t)_{t=1}^T$ of the seasonal component of x is given by

$$\hat{s}_t = \hat{s}_k, \ \text{s.t. } t = k \pmod{p}$$

for every $t = 1, \ldots, T$ and every $k = 1, \ldots, p$.

Corollary 755 *Under the assumptions of Proposition (754), let $(\hat{s}_t)_{t=1}^T$ be an estimate of the seasonal component of x. Then the time series $(y_t)_{t=1}^T$ given by*

$$y_t \stackrel{def}{=} x_t - \hat{s}_t, \quad \forall t = 1, \ldots, T$$

can be modeled by a stochastic process $(Y_t)_{t\in\mathbb{Z}}$ which has no additive seasonal component.

Bibliografia

[1] Karatzas, I.; Shreve, S. E., Brownian Motion and Stochastic Calculus, 2nd Ed. (Graduate Texts in Mathematics; 113). New York, Springer-Verlag 1996.