

Dottorato di ricerca in: “Scuola di dottorato in scienze sociali ed economiche” curriculum: Scienze statistiche ed analisi del dato sociale

“Sapienza università di Roma”

Ciclo: XXXVIII

Titolo tesi

Polisemia delle crisi nella complessità delle società moderne: una analisi quantitativa per il monitoraggio e la predittività del benessere e della salute dei territori

Lingua: Italiana

Autore:  
Davide Turso

Tutor:  
Prof.ssa Filomena Maggino

Durata Dottorato: 01/11/2022 - 31/10/2025

## TITOLO TESI

# **Polisemia delle crisi nella complessità delle società moderne: una analisi quantitativa per il monitoraggio e la predittività del benessere e della salute dei territori**

## INDICE

INTRODUZIONE		<b>III</b>
CAPITOLO 1	Le crisi nelle società moderne	<b>1</b>
1.1	La crisi nella sua semantica	1
1.2	Brevi riflessioni sugli effetti delle crisi su alcuni dominî del <i>BES</i>	21
CAPITOLO 2	Una metodologia per lo studio dei cambiamenti di un sistema complesso: La teoria delle catastrofi e suoi principî matematici fondamentali	<b>28</b>
2.1	Principî fondamentali e nozioni di topologia	28
2.2	Sistemi dinamici e loro caratteristiche fondamentali	45
2.3	Un possibile utilizzo della teoria delle catastrofi in contesti reali	61

<b>CAPITOLO 3</b>	<b>Alcuni modelli quantitativi per l'analisi delle serie storiche</b>	<b>61</b>
3.1	Nozioni generali sui Processi Stocastici	64
3.2	Alcuni modelli uni-variati per serie storiche	76
3.3	Alcuni modelli multi-variati per serie storiche	92
3.4	Cenni sulle Reti Neurali Artificiali ( <i>ANN</i> )	111
<b>CAPITOLO 4</b>	<b>Analisi quantitativa dei dominî <i>Salute e Benessere Economico</i> per la valutazione del benessere dei territori</b>	<b>119</b>
4.1	Una rappresentazione di analisi predittiva tramite metodi statistici	119
4.2	Applicazione a dati territoriali: il caso di Taranto	130
4.3	Applicazione a dati Nazionali	174
<b>CAPITOLO 5</b>	<b>Conclusioni e considerazioni finali</b>	<b>209</b>
	<b>RIFERIMENTI BIBLIOGRAFICI</b>	<b>213</b>
	<b>RELAZIONE TUTOR</b>	

Il presente documento è distribuito secondo la licenza "Tutti i diritti riservati".

## INTRODUZIONE

Nell'era contemporanea, le crisi rappresentano eventi in grado di modificare profondamente le dinamiche economiche, sociali e territoriali. Esse si manifestano con una rapidità e una complessità tali da influenzare non solo i sistemi economici, ma anche il benessere delle comunità e la qualità della vita dei singoli individui. Comprendere le modalità con cui tali eventi si diffondono ed i loro effetti sui diversi domini del benessere diventa quindi una sfida centrale, sia per la ricerca scientifica sia per la definizione di politiche efficaci.

Questo lavoro di tesi si propone di analizzare le conseguenze delle crisi su alcuni domini del Benessere Equo e Sostenibile (BES), con l'obiettivo di evidenziare le interazioni tra eventi critici e indicatori sociali, economici e territoriali. L'analisi è stata rivolta al territorio di Taranto, per studiare gli effetti locali sul dominio della Salute, e al contesto nazionale, per osservare le dinamiche relative al Benessere Economico.

Nel lavoro si presenterà in maniera generale *“La Teoria delle Catastrofi”* di René Thom, che offre strumenti concettuali per descrivere transizioni improvvise e cambiamenti qualitativi nei sistemi complessi. La teoria di Thom permette di interpretare come variazioni apparentemente piccole nelle condizioni iniziali possano generare conseguenze rilevanti, un'idea che trova riscontro nelle dinamiche osservate nei fenomeni economico-sociali e che offre un ponte concettuale tra la complessità teorica e l'analisi empirica. Dal punto di vista metodologico, la ricerca integra approcci tradizionali di co-integrazione con tecniche di reti neurali artificiali, quest'ultime utilizzate come strumenti ausiliari. Questa combinazione si pone l'obiettivo di cogliere le relazioni di lungo periodo tra le variabili sociali, utilizzando come supporto al metodo analitico principale le reti neurali. In questo modo, il lavoro mira a rappresentare i trend stabili dati dal principio di co-integrazione, con le reti neurali che fungono da ausilio nel colmare delle problematiche che possono nascere dall'implementazione della co-integrazione per via della struttura dei dati utilizzati.

I due casi di studio, l'analisi del dominio della Salute riguardante il territorio di Taranto e del dominio del Benessere Economico a livello nazionale si pone l'obiettivo di cercare relazioni e possibili segnali di vulnerabilità o resilienza, fornendo una base quantitativa ed interpretativa per comprendere come le crisi possano influenzare in modo differenziato le diverse dimensioni del benessere. In sintesi, questo lavoro si propone di coniugare teoria, metodologia e applicazione empirica, creando un filo conduttore tra concetti di sistemi complessi, strumenti quantitativi avanzati e analisi dei fenomeni reali basandosi sulla idea che ha alimentato tutto il lavoro e che verrà esposta in esso. La tesi intende così offrire non solo un quadro generale dei metodi utilizzati, ma anche indicazioni concrete sul ruolo che tali metodi possono assumere per analizzare fenomeni socio-economici a livello locale e nazionale, al fine di poter sviluppare delle policy che mirino al mantenimento ed accrescimento del benessere delle persone.

# CAP. 1

## Le crisi nelle società moderne

### 1.1

#### La crisi nella sua semantica.

Nell'arco del tempo, il concetto di *crisi* ha assunto diversi significati a livello sociale, in ragione dei cambiamenti che le società hanno attraversato nei diversi periodi storici. Una società, intesa in termini di struttura, composizione umana, organizzazione politica, economica, sanitaria e così via, nel Medioevo non poteva essere la stessa anche solo dopo 200 o 300 anni, pur riferendosi al medesimo territorio, né tantomeno nei secoli successivi. Diversi livelli di cultura, di civiltà, di umanità, di popoli provenienti da differenti paesi caratterizzano nel tempo la struttura sociale di uno stesso territorio, generando assetti sempre differenti rispetto a quelli precedenti. Anche il livello di conoscenza tecnologica, in continua evoluzione, influisce sulla conformazione della società di un territorio nei diversi periodi storici. L'insieme di questi fattori ha progressivamente trasformato il significato di *crisi sociale* ed il suo impatto sulla società, a seconda del contesto storico considerato. Si può dire che il concetto di *crisi sociale* abbracci altri concetti in virtù del fatto che una società, anche primordiale, ha sempre avuta una conformazione complessa nel senso che, essa è stata sempre il frutto di una composizione variegata di realtà umane e di conseguenza, una *crisi sociale*, ha sempre avuto il significato di una *crisi* di più realtà che compongono una società umana. Etimologicamente, il termine ha una derivazione greca ed indica una scelta, una decisione. Essa è definita in generale come una perturbazione o improvvisa variazione degli stati di equilibrio della realtà colpita. In una società, una crisi rappresenta, come concetto fulcro, una mancata corrispondenza tra valori umani e modi di vivere e questo rappresenta la conseguenza di sostanziali mutamenti strutturali della stessa società. Il termine viene utilizzato in ogni ambito della società; pertanto, si parla di crisi in ciascun settore che riguardi lo studio dei sistemi e dei comportamenti umani, come ad esempio crisi economico-finanziaria, psicologica, politica e così via. Nonostante l'uso frequente del termine *crisi* nella modernità, non vi è una linea interpretativa chiara dei fenomeni sociali da esso coinvolti, il che contribuisce ad una non semplice e lineare semantica concettuale delle componenti chiave che lo definiscono. Una crisi sociale ha manifestazioni straordinarie percepibili esternamente, irrompe in una comunità sociale sfaldando, a seconda della sua intensità, gli equilibri ed i meccanismi che la regolano. Rappresenta una perturbazione ed alterazione dei processi esistenti del sistema sociale colpito. In una tale situazione, le normali azioni volte a risolvere problemi di funzionamento del sistema sociale vengono meno, diventando non più idonee a risolvere le problematiche emerse. Le ripercussioni di ciò possono essere tali

da minare la esistenza stessa della organizzazione sociale colpita dalla crisi, e questo impone azioni repentine per attuare delle scelte funzionali ad arginare gli effetti negativi da essa generati. Per tutti questi motivi, non vi è una agevole guida di eventi che possa dare una precisa e ripetitiva situazione di crisi, ossia non vi sono delle peculiarità alla manifestazione delle quali si può affermare essere in atto una criticità di un sistema sociale, o almeno esse non sono facilmente definibili. La complessità dell'evento critico, in una società, va studiata ed affrontata in maniera dinamica, ossia seguendo i mutamenti sociali che possono essere repentini o gradualmente, possono colpire l'intero sistema sociale o solamente alcuni suoi sotto sistemi, possono avere natura interna o esterna ad esso, possono avere delle origini spontanee dovute al fisiologico mutamento che colpisce le società nel tempo in maniera non prevedibile, oppure possono derivare da azioni collettive o di una parte degli elementi che la compongono. Una crisi, allora, può essere vista come un evento portatore di un *cambio qualitativo* nei processi che regolano e reggono una società e che mette in contrapposizione situazioni non compatibili tra loro. Tale evento porta con sé una alea di innovazione e di cambiamento nel tessuto sociale, ed allo stesso tempo, l'innovazione può essere il preludio per un mutamento positivista della realtà o portare, in caso estremo, al crollo della stabilità dello stesso sistema sociale. Tuttavia, non si deve necessariamente interpretare ogni situazione di instabilità sociale come l'inizio imminente di una crisi della società. I segnali precedentemente menzionati, infatti, possono manifestarsi con un'intensità insufficiente a costituire il preludio di un cambiamento critico. In alcuni casi, essi rappresentano semplicemente l'effetto di una destabilizzazione momentanea di un sistema sociale, il quale può trovarsi in una fase di transizione destinata a condurre verso una nuova condizione di stabilità. Può dunque trattarsi di una situazione di assestamento del sistema sociale di tipo graduale e fisiologico e soprattutto non netto e drastico, senza conseguenze minatorie per esso. Concetti affini a quello di crisi sono quelli di *conflitto* e di *catastrofe*; quest'ultimo sarà analizzato in modo specifico dal punto di vista matematico nel prosieguo del lavoro. Questi termini sono spesso utilizzati come sinonimi, ma in realtà presentano origini concettuali differenti e danno luogo a situazioni distinte, pur condividendo alcuni elementi comuni, come la discontinuità e l'imprevedibilità dell'esito. Il termine *crisi*, come già accennato all'inizio, ha il significato di aver la facoltà di scegliere, decidere, differenziare. Il conflitto invece, sociologicamente, rappresenta un contrasto, un legame antagonista tra soggetti -singoli o collettivi- in competizione tra loro. Esso si configura, sostanzialmente, come una diade polemogena, ovvero una relazione duale caratterizzata da tensione ed antagonismo tra i soggetti coinvolti. Il conflitto cessa di esistere quando una delle due parti afferma la propria supremazia imponendo i propri dettami all'altra, oppure quando si raggiunge un accordo basato sulla definizione di diritti e doveri reciproci. Talvolta, se il conflitto tra le parti è particolarmente acceso, esso può trasformarsi in crisi oppure può anche rappresentare una fase successiva di una crisi stessa poiché, introducendo la figura del nemico, il conflitto può generare nuove forme di coesione sociale, alimentare un sentimento di opposizione verso un terzo e rafforzare la convinzione che, sconfiggendo tale nemico,

si possa addivenire alla risoluzione della crisi. La crisi, a differenza del conflitto, può anche manifestarsi senza ostilità ed aver fine senza né vincitori e né vinti. Solitamente, essa porta seco situazioni di disorganizzazione, determina dei cambiamenti inattesi e modifica il comune agire sociale, provocando incertezza e squilibrio. Questi ultimi possono protrarsi a lungo, generando smarrimento riguardo alle scelte più appropriate da adottare per contrastare la situazione emersa, limitando notevolmente la capacità previsionale e favorendo la trasformazione della crisi in catastrofe. In una situazione di crisi prolungata possono verificarsi profondi cambiamenti nella società, che nei casi più gravi possono sfociare in rivoluzioni o guerre civili. Tali eventi innescano fasi storiche in cui si consolidano realtà di opposizione profondamente radicate nel tessuto sociale, spesso di lunga durata, note come *cleavages* (sfaldamenti del tessuto sociale), capaci di imprimere una marcata polarizzazione al sistema politico. Inoltre, si può legittimamente supporre che crisi e conflitto condividano, seppur con diversa intensità, alcuni fattori generativi comuni. Tali fattori, in grado di alterare la situazione iniziale di una società, possono essere esogeni, endogeni oppure derivare da una combinazione dei due, tanto nel caso della crisi quanto in quello del conflitto. In entrambi i casi, essi possono manifestarsi sia su scala locale che globale. Per quanto riguarda un altro concetto spesso usato come sinonimo di crisi, ovvero la catastrofe, essa ha avuta particolare applicazione - soprattutto in campo matematico- negli anni Ottanta del secolo scorso grazie alla teoria matematica del filosofo e matematico Renè Thom. Tale teoria veniva applicata a tutti quei sistemi che, al variare di alcuni parametri, subiscono mutamenti repentini e discontinui, determinando una trasformazione della forma preesistente del sistema stesso. Lo stesso Thom, nella sua teoria, evidenzia il forte legame tra crisi e catastrofe, sostenendo che la prima può condurre alla seconda, pur senza che vi sia un determinismo automatico nella trasformazione. Infatti, in molti casi, la crisi può anche riassorbirsi spontaneamente, senza produrre modifiche evidenti nel sistema colpito. Vi sono però delle differenze sostanziali tra i due concetti.

Al contrario di una crisi, che può trovarsi inizialmente in uno stato latente manifestandosi successivamente a seguito di alterazioni dei processi che regolano un sistema, la catastrofe è immediatamente osservabile. Per Thom, la crisi ha carattere soggettivo, nel senso che ogni soggetto può trovarsi in una situazione di crisi, qualora verta in uno stato personale di debolezza dovuto ad una apparente variazione dei meccanismi regolatori del suo essere, tali che vengano - dal soggetto stesso- percepiti come minaccia alla propria esistenza. Dunque, sarà fondamentale l'agire del soggetto stesso per riportarsi in una situazione di equilibrio, attraverso azioni che abbiano una progettualità mirata al perseguimento della re-stabilizzazione dello stato personale pre-crisi. Con riferimento ai sistemi sociali, sussiste una ulteriore differenza tra crisi e catastrofe: la catastrofe è un processo che mina la struttura del sistema che ne è colpito, mentre la crisi ha natura funzionale nel senso che, essa colpisce un sistema nel suo funzionamento ma non è detto che abbia come epilogo la compromissione della struttura del sistema. Sebbene il termine *crisi* venga comunemente usato in una accezione negativa, comportando essa dei cambiamenti e delle rotture nei

meccanismi di funzionamento del sistema che la subisce, tali eventi possono anche implicare un miglioramento ed una crescita del sistema portandolo verso una nuova condizione migliore della precedente. Dal suo stesso significato etimologico, si può intuire che il termine crisi possa essere utilizzato per indicare variazioni di uno stato pre-esistente, applicabile ad una ampia gamma di settori e realtà. Dalla medicina alla storia, dai sistemi fisici, linguistici, dai grandi eventi che hanno colpito il mondo, il termine *crisi* viene utilizzato fundamentalmente per rappresentare quel punto di svolta che, nella realtà colpita, può portare a variazioni positive o negative. Non mancano, nella storia — anche classica — esempi di analisi relative a crisi sociali provocate, ad esempio, da epidemie, in cui l'attenzione è stata rivolta agli effetti della malattia sulla popolazione e alle conseguenze morali, etiche ed organizzative per il sistema sociale colpito. In tutti gli ambiti in cui si affronta una situazione di crisi, un filo interpretativo ricorrente consiste proprio nella ricerca di una dimensione temporale adeguata per compiere delle scelte e metterle in atto, affinché le azioni che ne derivino abbiano come finalità il riposizionamento del sistema in una condizione di equilibrio — uguale a quella precedente o diversa — ma pur sempre di equilibrio. Il concetto di crisi iniziò a essere applicato anche in ambito politico con le prime ribellioni dei popoli europei contro i sistemi monarchici che li mantenevano in condizioni di sudditanza. In tale contesto, il sistema sociale cominciò ad essere concepito come un organismo vivente: una crisi, paragonata all'insorgere di una malattia, richiedeva un intervento tempestivo e mirato, volto a risolverla o ad arginarne gli effetti negativi. Questa lettura cominciò a fungere da linea guida nella gestione delle crisi sociali, basandosi sull'idea che i mutamenti prodotti dalla crisi potessero talvolta rivelarsi necessari — o addirittura fisiologici — per favorire la crescita del sistema stesso, in vista di un nuovo equilibrio capace di migliorare le condizioni di vita degli individui, analogamente a quanto accade nell'evoluzione di un organismo vivente. In una società, la crisi porta con sé un processo di mutamento profondo, conducendo il sistema sociale a un cambiamento sia nella sua struttura che nella cultura che lo permea. Come accennato in precedenza, una crisi non deve necessariamente coincidere con una fase di momentaneo turbamento o smarrimento della società. Vi sono, infatti, casi storici in cui non si è trattato di crisi autentiche, bensì di diatribe — civili, religiose o di altra natura — che, pur protrattesi nel tempo, non hanno determinato cambiamenti fondamentali o sostanziali all'interno della società che le ha attraversate. In una crisi autentica, un ruolo fondamentale è svolto dalla comunicazione tra gli individui. Attraverso il dialogo e lo scambio di informazioni, i membri di una società possono condividere sensazioni, malcontento e la volontà di reagire alle forze negative che permeano il contesto sociale. Questo processo comunicativo favorisce una più o meno rapida diffusione della crisi nei diversi strati della società, contribuendo a generare coesione e coraggio collettivo nel contrastare le sue cause originarie. Una crisi sociale autentica, così, può generare coesione e consapevolezza tra le persone, risvegliando energie vitali che scaturiscono dalla naturale propensione dell'essere umano alla vita e ad una vita migliore, vivibile una volta superata la causa originaria della crisi. Perciò, proprio in quanto capace di

attivare energie nuove, il concetto di crisi si avvicina alla logica dei sistemi viventi, che reagiscono agli agenti esterni quando questi minacciano la loro stessa salute. Una focalizzazione più marcata del concetto di *crisi* in ambito economico iniziò a emergere in maniera significativa nel periodo compreso tra la Grande Guerra e la crisi del 1929 — sebbene già nella prima metà del XVIII secolo il termine fosse stato applicato ai processi economici. In quegli anni si diffuse progressivamente l'idea che una crisi economica rappresentasse una fase di transizione di un sistema da un periodo di espansione ad uno di contrazione, secondo una logica ciclica che si sviluppava lungo un arco temporale definito. Parallelamente, si radicò la convinzione che le civiltà nascano e si sviluppino in quanto capaci di rispondere efficacemente a nuove situazioni emergenti, mentre tendano al declino quando non sono più in grado di affrontare le sfide poste dai mutamenti storici ed economici. È dunque la capacità — o l'incapacità — di far fronte alle sfide a determinare la crescita o il declino di una civiltà, e questo principio si riflette anche nel funzionamento del suo sistema economico. Una civiltà che si trovi ad affrontare profondi cambiamenti socio-culturali può entrare in una fase di crisi, la quale può segnare tanto uno sviluppo quanto un declino. Sul concetto di *crisi* si sono confrontati filosofi, sociologi e pensatori di varia formazione. In particolare, dopo la grande congiuntura economica del 1929, la riflessione filosofica si è concentrata sull'idea che l'umanità, con l'avvento delle scoperte scientifiche moderne, abbia progressivamente smarrito la propria essenza esistenziale: quella concezione dell'essere umano come *ente ontologicamente orientato al senso* che affonda le sue radici nella filosofia greca. Gli studiosi iniziarono ad interrogarsi su cosa le nuove scienze potessero significare per l'umanità in termini di sviluppo dei valori etici, che il genere umano ha da sempre cercato di perseguire per sé stesso. Si osservava infatti un progressivo allontanamento dell'uomo dall'idea che il mondo dovesse essere orientato da principi filosofici classici finalizzati al bene comune, e al contempo si intravedeva una crescente tendenza ad affidare questo compito alle scienze moderne, attratti dagli esiti positivi che esse promettevano di apportare alle società. Questa riflessione non mirava a demonizzare le scienze moderne, bensì a prevenire il possibile errore di confondere il progresso scientifico con il fine ultimo dell'esistenza umana, sostituendo lo sviluppo etico e l'autenticità dell'essere con il solo avanzamento tecnico. Proprio tale sostituzione — lo scambio tra mezzi e fini — veniva interpretata come una crisi delle società moderne, in cui il progresso scientifico rischiava di essere elevato a valore assoluto, anche a scapito del perseguimento del bene dell'umanità. Si cominciava a percepire una tensione crescente tra due visioni contrapposte: da un lato, il perseguimento del progresso scientifico come fine a sé stesso, anche a costo dello sfruttamento indiscriminato delle risorse, umane e naturali, e del sacrificio dei principi fondamentali delle società; dall'altro, la filosofia, con la sua naturale vocazione ad interrogarsi sulla vera essenza dell'essere umano e sul ruolo che la scienza dovrebbe assumere, come strumento funzionale e non dominante, al servizio dell'umanità e del mondo che la circonda. Da questo punto di vista, il concetto di crisi non va inteso come un errore o un male, ma piuttosto come un passaggio

necessario per raggiungere stadi di civiltà migliori rispetto a quelli precedenti. Questa visione positivista della crisi, intesa come punto di svolta verso un miglioramento della società, è presente anche nel pensiero filosofico marxista. Marx infatti considerava la crisi come un fenomeno strutturale del capitalismo: una manifestazione della periodicità del sistema produttivo che, attraverso le sue contraddizioni intrinseche, consente al capitalismo di rinnovarsi e rafforzarsi. In questa ottica, allora, il fulcro della teoria Marxista risiede nel postulare una “ultima crisi generale” del capitalismo, che lo porterebbe definitivamente alla sua fine con i suoi ritmi produttivi disumani, riportando al centro della società non il profitto, bensì l’uomo come protagonista principale, con la sua dignità e centralità nelle vicende del mondo. In questa linea di pensiero rientrava anche l’idea di crisi di un altro intellettuale, Antonio Gramsci, il quale non considerava il concetto dal punto di vista economico, bensì come la rottura di uno stato di equilibrio sociale tra il popolo ed i suoi rappresentanti, che genera una mobilitazione di massa. In tale contesto, l’uomo è spinto dagli eventi ad emanciparsi sentendosi membro di un collettivo con responsabilità e valori morali accomunanti da difendere. Successivamente al pensiero Marxista sul concetto di crisi, si sono avute delle idee che si distaccavano dalla linea economicistica e che consideravano la crisi come un mal funzionamento dei meccanismi regolatori degli elementi che costituiscono una società nella sua complessità, soprattutto in termini di integrazione sociale. Tali crisi, salvo in casi eccezionali, non minavano la sopravvivenza del sistema sociale stesso. In situazioni in cui si verificano eventi di tale portata da innescare una crisi del sistema, quest’ultimo tende ad attivare, per istinto di autoconservazione, meccanismi compensativi finalizzati al ripristino dell’equilibrio, attraverso la rimozione o il contenimento dei fattori negativi che hanno generato la stessa crisi. In tale ottica, i sotto-sistemi sociali possono entrare in condizioni di disequilibrio che finiscono per compromettere la stabilità dell’intero sistema; il cambiamento in un sistema sociale non avviene in modo unitario e coerente, ma è determinato dalle trasformazioni delle singole componenti, spesso in conflitto con fattori destabilizzanti. Ciò evidenzia un cattivo coordinamento tra i diversi sotto-sistemi, rendendo necessario l’avvio di meccanismi correttivi capaci di ristabilire i legami funzionali tra gli elementi che compongono il sistema nella sua complessità. Una crisi, infatti, altera sempre — in misura più o meno marcata — la stratificazione sociale, generando nuove disuguaglianze e accentuando quelle già esistenti. In quest’ottica, il superamento di una lettura esclusivamente economicistica del concetto di crisi consente di costruire un quadro descrittivo e interpretativo più esaustivo, orientato alla promozione del benessere collettivo. Pur considerando la crisi come un fenomeno che investe l’insieme delle componenti di un sistema sociale, non sono mancati studiosi che, nell’era moderna, hanno posto l’accento sulla dimensione individuale della crisi. In particolare, lo sviluppo delle grandi città ha dato origine a fenomeni come l’alienazione e l’anonimato, che hanno reso l’individuo sempre più isolato nel contesto urbano. Tali condizioni, secondo alcuni approcci, sono all’origine di vere e proprie patologie sociali e psicologiche, in grado di generare situazioni di profondo

squilibrio personale, dando luogo a forme di crisi individuali. In tale condizione, tra i principali meccanismi di difesa messi in atto dall'individuo vi è una marcata attenuazione della sensibilità verso le differenze, gli stimoli e le opportunità offerte dalla grande città, con la conseguente tendenza alla chiusura rispetto al mondo esterno. Tuttavia, le grandi città possono anche generare stati di crisi in quegli individui che vivono in modo eccessivo l'esperienza urbana: persone che frequentano salotti intellettuali, alla continua ricerca di nuove esperienze conoscitive e consumistiche inseguendo stimoli diversi. Questa tensione verso il "nuovo" finisce per alimentare una condizione di perenne insoddisfazione, dovuta alla spasmodica aspirazione a realtà sempre differenti, che spesso risultano irraggiungibili. L'individuo scivola così in uno stato di crisi profonda, sentendosi spaesato e straniero nella propria stessa città, come un immigrato in un luogo incapace di offrirgli ciò che possa rispecchiare la sua autentica natura umana. Resta il fatto che, anche una tale situazione può fungere da volano per una nuova fase, un nuovo stato migliorativo innescato dai meccanismi messi in atto dall'individuo, volti al superamento di tale negatività per addivenire ad un nuovo stato, migliore del precedente. Nelle società moderne, la sociologia individua nella crisi sociale una delle principali cause di processi che portano alla formazione di strati sociali differenti. Da un lato, vi sono individui che riescono a vivere e a muoversi con disinvoltura in un contesto sempre più individualizzato e, sotto molti aspetti, umanamente asettico; dall'altro, si trovano coloro che vivono una condizione di sofferenza derivante dal decadimento dei valori, dalla perdita di umanità e dalla carenza di stimoli all'inclusione sociale e lavorativa. Tale condizione genera sfiducia e, nei casi più gravi, può avere conseguenze significative sul benessere psicologico e sociale degli individui. In Italia, questi fenomeni tendono a colpire prevalentemente giovani, donne, immigrati e persone con basso livello di istruzione, con una maggiore concentrazione nelle regioni del Mezzogiorno. Si delinea così, all'interno di una crisi sociale, una sorta di bivio: due direzioni possibili che gli individui possono intraprendere, in base alla propria personalità, sensibilità e alle opportunità a disposizione. Spetta alla sociologia il compito di sviluppare strumenti analitici in grado di cogliere con precisione lo stato della società, nonché di anticiparne le possibili linee evolutive. È evidente, infatti, che una crisi sociale non è soltanto un concetto astratto, ma una questione altamente operativa, poiché comporta implicazioni cruciali per la sopravvivenza della comunità. In un mondo globalizzato, uno stato di crisi profonda può indurre i popoli a necessitare di shock forti — eventi estremi — per riconoscere l'urgenza di un cambiamento radicale, volto a evitare un declino irreversibile della società. Tale necessità di azioni traumatiche, finalizzate a risvegliare le coscienze, potrebbe anche portare ad un "bisogno di guerre", essendo esse eventi estremi e totalizzanti che storicamente hanno dimostrato di essere — purtroppo — capaci di provocare mutamenti profondi e immediati nell'assetto sociale e valoriale delle comunità. Tuttavia, non sempre le scienze sociali, pur disponendo degli strumenti teorici e metodologici adeguati, sono riuscite a prevedere l'insorgere di una crisi in uno specifico ambito sociale. Ciò ha reso necessaria, in diversi casi, una rivalutazione

critica degli stessi strumenti interpretativi. Un esempio emblematico è rappresentato dalla crisi finanziaria degli ultimi decenni, che, con le gravi conseguenze sociali che ha comportate, non fu adeguatamente e prontamente prevista. In particolare, non si riuscì a cogliere appieno la portata delle sue ripercussioni sul tessuto sociale. Quell'evento ha imposto una riflessione profonda sui nuovi legami tra finanza, economia ed etica, evidenziando la necessità di un approccio più integrato e interdisciplinare nell'analisi dei fenomeni complessi. Tuttavia, non sono mancati numerosi studiosi che avevano già condotto analisi operative sul ruolo fondamentale che l'etica dovrebbe assumere nelle questioni economiche, finanziarie, nonché nella gestione, tutela e distribuzione della ricchezza pubblica. Essi evidenziarono come gli Stati occidentali stessero progressivamente orientando le proprie politiche sempre meno verso l'interesse pubblico. Tale fenomeno non fu però riconosciuto come un campanello d'allarme rispetto al rischio di collasso di un sistema basato su forti interessi privati, appartenenti a pochi soggetti dotati di un potere finanziario sconfinato. Spesso una crisi che esplose in uno Stato può far emergere problematiche già presenti, seppur latenti, nel tessuto sociale di un altro Paese, diventando così soltanto un mezzo di manifestazione di tali criticità e non la loro vera causa. Questo fenomeno è dovuto alle profonde connessioni che ormai legano le diverse componenti delle società umane, un fattore ulteriormente accentuato dallo sviluppo e dalla diffusione di massa dei mezzi di comunicazione. Il significato di crisi resta valido, anzi si accentua ulteriormente nelle società moderne, dove le scelte da compiere in condizioni difficili diventano ancora più urgenti per risolvere le problematiche sorte. Ciò è dovuto alla maggiore rapidità con cui una situazione critica si diffonde, nonché al fatto che, nelle società contemporanee, la crisi amplifica il disagio vissuto dall'individuo. La crescente consapevolezza di tali situazioni di disagio, legate a fattori culturali, all'aumento delle informazioni e ad altri elementi, consente di affrontare le crisi in modo più coerente ed efficace, sfruttando al meglio gli strumenti cognitivi e materiali di cui la società dispone. Ciò consente, ancor più, di trasformare una crisi in un'opportunità di crescita, sviluppo e rafforzamento del tessuto sociale, nella consapevolezza che su ogni individuo e su tutte le istituzioni grava l'onere e la responsabilità di compiere azioni orientate all'interesse collettivo. In questo contesto, il punto di riferimento fondamentale nella gestione e risoluzione delle crisi sociali rimane sempre lo Stato nazionale. In un mondo globalizzato, si è notata una polarizzazione sociale definita anche in un dato spazio, e dunque il ruolo dello stato è importante per il senso di appartenenza ad una collettività che esso può dare agli individui, generando coesione e forza nell'affrontare situazioni critiche. Però questa tendenza alla polarizzazione sociale, se mal gestita, può dar vita a situazioni di marginalità sociale per gli individui interessati, e questa dinamica può fungere da volano per l'accentuazione di situazioni negative già in essere in situazioni di crisi sociali. Una crisi economico-finanziaria, ad esempio, potrebbe portare ad una cristallizzazione delle differenze sociali già esistenti, rendendole più pregnanti nello stesso tessuto sociale. Questo fenomeno potrebbe portare ad un'accentuazione delle differenze spaziali all'interno della società, ma il rovescio

della medaglia risiede proprio in quanto si è detto riguardo alla crisi: essa può suscitare una reazione energica da parte degli individui, spingendoli a fronteggiare le nuove condizioni generate dal cambiamento. Di conseguenza, potrebbe emergere una nuova etica sociale, orientata a stili di vita fondati sulla solidarietà, sul bene collettivo e sulla giustizia sociale. In un contesto di crisi sociale, è importante sottolineare che essa colpisca ed influenzi in modo differenziato i vari strati della società. Normalmente, le persone più vulnerabili sono quelle che subiscono le conseguenze più gravi di una crisi, e tra queste va considerata in particolare la “categoria” dei giovani. Questo perché essi rappresentano non solo il contributo attuale alla società, ma anche il suo futuro, essendo coloro che guideranno la società domani. Una società che attraversa una situazione socio-economica e istituzionale critica, caratterizzata da incertezza che mina i valori e l’identità di un popolo, genera forti problematiche per i giovani. Queste si manifestano soprattutto in ambiti come il lavoro, la precarietà del futuro, e riflessioni profonde sull’essenza e il significato stesso della vita. In una situazione simile, spetta a coloro che, anni prima, erano i giovani della società avere la capacità di mettere in atto iniziative mirate a risolvere le criticità emergenti. Questo sarà possibile solo se il periodo della loro giovinezza è stato vissuto in modo sano e stabile, con la necessità che tale condizione si ripeta anche nelle generazioni successive. Se invece, nel momento attuale, gli attori principali di una società non fossero in grado di affrontare adeguatamente le sfide sociali ed economiche, la classe giovanile ne risentirebbe maggiormente. L’incapacità di progettualità sociale potrebbe tradursi in scoramento verso le istituzioni, generando un forte senso di repulsione verso il coinvolgimento politico, alimentato dalla precarietà dei valori, delle condizioni occupazionali e dal decadimento dell’etica. Da studi effettuati, si è osservato che questa situazione potrebbe portare ad una crescita di comportamenti individualistici tra i più giovani. Tuttavia, questo individualismo non va interpretato come un segno di scarsa moralità o di un deficit nel senso civico, ma piuttosto come una nuova modalità di interazione con le mutate condizioni sociali. Si tratta, infatti, di una forma di reazione degli individui, ed in particolare di quelli più giovani, alle nuove situazioni di incertezza, precarietà e ai rischi emergenti. Si può affermare, quindi, che un’efficienza delle istituzioni favorirebbe una maggiore partecipazione alla vita politica, promuovendo un naturale e saluberrimo ricambio generazionale. Questo processo permetterebbe il continuo rinnovo di idee e soluzioni migliorative, trasmesse di generazione in generazione. Da questi concetti fin qui esposti, frutto di uno studio approfondito sulla crisi effettuato nel tempo, si può affermare che essa rappresenti una processualità in cui si individua un livello oltre il quale si manifesta un cambiamento qualitativo nella stessa dinamica. Si genera così un processo lungo e complesso, nel corso del quale si creano situazioni tra loro incompatibili, spesso dovute anche alle innovazioni che, per loro natura, accompagnano una crisi. Tali innovazioni possono, tuttavia, diventare anche fonte di un inasprimento delle incompatibilità già presenti all’interno di una società, in relazione ai suoi valori, alle sue azioni e alle regole che la governano, con il rischio che ciò preannunci un crollo del sistema sociale stesso. In alternativa, a una fase di destabilizzazione innescata

dalla crisi seguirà una fase in cui il sistema sociale colpito riuscirà a raggiungere una nuova condizione di stabilità. Queste dinamiche porteranno comunque ad un assetto diverso rispetto a quello esistente prima della crisi. Il tema della complessità strutturale di una società è fondamentale nello studio delle crisi sociali, ancor più nelle società moderne. In questa prospettiva si fonda l'idea che le società odierne siano caratterizzate da un'enorme quantità di possibilità decisionali, spesso superiori a quelle che la struttura sociale stessa è in grado di gestire ed attualizzare efficacemente. Sarà quindi la capacità del sistema sociale di dominare e governare questa complessità a determinarne la capacità di amministrare con successo gli eventi che lo colpiranno. Questa concezione di complessità implica un legame stretto tra il sistema e l'ambiente in cui esso si sviluppa, con un'evoluzione caratterizzata da dinamiche di instabilità e indeterminatezza, nelle quali contraddizioni e conflitti assumono un ruolo determinante. Le contraddizioni portano con sé una natura di crisi, poiché rappresentano realtà interne al sistema caratterizzate da incompatibilità che diventano uno stimolo per la società a reagire di fronte a situazioni di pericolo e rischio. In un contesto sociale in cui le sue componenti comunicano tra loro, tali contraddizioni permettono di selezionare in modo mirato le azioni da intraprendere per favorire un'interazione proficua tra gli elementi costitutivi della società. Questo processo contribuisce a "semplificare", in un certo senso, la complessità del sistema, consentendo una scelta più efficace e razionale delle alternative disponibili e attuabili nell'ambiente in cui la società si trova ad operare. Sebbene il concetto di crisi sia storicamente legato all'economia, alla politica ed alla giurisprudenza, è fondamentale il contributo che la sociologia può offrire per un'analisi critica delle crisi di un sistema sociale. Quest'ultimo, infatti, non si fonda esclusivamente su principi economici, politici o giuridici, ma ha la sua componente principale negli individui che compongono la società. Per questo motivo, ogni azione dovrebbe essere orientata al benessere delle persone, anche quando ciò comporta direzioni non pienamente convergenti con gli interessi di altre discipline. La sociologia, dunque, rientra pienamente in questo processo di studio e analisi, senza cadere in idee fuorvianti. Sarebbe infatti errato ritenere che una situazione economica, politica o giuridica favorevole garantisca automaticamente una società migliore, o che il benessere collettivo coincida necessariamente con indicatori positivi in questi ambiti. È fondamentale, invece, che il benessere e la felicità degli individui restino il fine primario, mentre le condizioni economiche, politiche e giuridiche devono essere considerate strumenti funzionali a tale obiettivo, e non il centro del lavoro sociale cui tutto deve asservirsi. Non mancano esempi, nella storia, in cui anche in società grandi e d'esempio per tutte le collettività del mondo, i decisori abbiano affrontate le crisi con l'apporto esclusivo di economisti, politologi, giuristi, dando una importanza marginale, o quasi nulla, al consulto dei sociologi per trovare soluzioni alle crisi che le avevano colpite. Questo approccio può forse essere spiegato dal fatto che una crisi viene percepita principalmente come una crisi economica. Di conseguenza, si tende a ritenere necessarie risposte politiche, da formalizzare attraverso strumenti giuridici, per ottenere i risultati auspicati nella gestione della situazione. È quindi

comprensibile che si pensi più facilmente a questi ambiti di studio — economia, politica, diritto — come i più adeguati per affrontare una crisi. Nell'era moderna, ad esempio, solo dopo un po' di anni dalla crisi del 1929, vi furono i primi studi sociologici volti a dare una spiegazione della stessa crisi di tipo sociale, con una penetrazione nelle cause ed effetti di essa che non portò a grandi risultati, ma sicuramente segnò il primo passo verso una inclusione della sociologia nello studio delle crisi delle collettività. In un'ottica contemporanea di analisi del concetto di crisi, è possibile individuare due principali chiavi di lettura. La prima chiave di lettura interpreta la crisi sociale come una conseguenza di fenomeni economico-finanziari. Quando si innescano gravi problematiche di natura finanziaria, il loro protrarsi, unito alle forti interconnessioni commerciali tra i paesi industrializzati, favorisce una rapida diffusione della crisi su scala globale. Questo processo ha effetti diretti sull'economia reale e, nel lungo periodo, genera forti tensioni sociali. Tale approccio è prevalente tra gli economisti, i quali individuano in esso una spiegazione centrale del concetto stesso di crisi. All'interno di questa prospettiva si possono individuare due distinti filoni di pensiero, ciascuno dei quali propone un approccio diverso per interpretare e affrontare il concetto di crisi. Da un lato vi è un orientamento di economisti che reputano siano necessari degli interventi da parte degli stati, interventi che però devono essere momentanei e mirati al solo raggiungimento della soluzione dello stato di crisi. Dall'altro lato, vi sono economisti che interpretano la crisi come un indebolimento e, successivamente, come l'interruzione del processo di crescita economica. Secondo questa visione, la crisi va contrastata attraverso l'intervento di istituzioni competenti — quali governi e banche centrali — mediante misure come la riduzione dei tassi di interesse, l'aumento della spesa pubblica in investimenti e l'immissione di liquidità nel sistema economico, con l'obiettivo di stimolare la domanda e rilanciare i consumi. Entrambi i pensieri sono riferibili, dunque, alla prima chiave di lettura che, come detto precedentemente, vede una crisi esclusivamente sotto un profilo economico-finanziario. La seconda chiave di lettura del concetto di crisi vede chi considera, tale situazione, come la risultante di problematiche innescatesi molto tempo prima della irruzione di essa, problematiche che fondamentalmente sono di tipo energetico. Le due visioni sono divise da una sottile linea interpretativa che contempla la soluzione attraverso un differente modalità di intervento dello stato e degli enti che hanno competenze di tipo economico-finanziario.

Quindi, entrambi gli approcci condividono la critica al modello neoliberista, sostenendo la necessità di un rinnovato intervento statale nella gestione delle crisi, pur se con modalità differenti. Al di là dei diversi approcci adottati dagli studiosi delle crisi che colpiscono le società, si è progressivamente affermata — in particolare nell'ultimo quindicennio — una visione condivisa circa la necessità di non separare il settore della finanza da quello dell'economia reale. Mediante evidenze empiriche, negli ultimi quindici anni si è potuto constatare un forte legame tra il settore finanziario e quello dell'economia reale. Questo legame è il risultato di una profonda evoluzione a livello globale, spinta in larga parte dall'avvento delle nuove tecnologie,

dal potenziamento della capacità di calcolo e dall'ampia accessibilità a tali strumenti da parte di una vastissima platea di persone. Tutte queste opportunità non sono più riservate a pochi, ma risultano oggi fruibili da una platea globale estremamente ampia. Queste nuove opportunità, oggi accessibili a un pubblico molto ampio, hanno avuto come effetto più evidente rispetto al passato quello di connettere strettamente i due sistemi economici — finanziario e reale — al punto che una crisi nel primo genera inevitabilmente ripercussioni significative anche nel secondo.

Tali ripercussioni risultano tanto più gravi quanto più il sistema finanziario evolve attraverso strumenti di ingegneria finanziaria, i quali producono prodotti finanziari sempre più complessi e caratterizzati da un effetto leva elevato. Questo effetto è in grado di amplificare e trasmettere rapidamente, su scala globale, sia le fasi di espansione positiva del sistema sia le crisi, contribuendo ad accrescerne l'impatto sistemico. In tal senso, non sono mancate aspre critiche e vere e proprie demonizzazioni nei confronti dei prodotti dell'industria finanziaria, ritenuti responsabili delle gravi crisi economico-finanziarie verificatesi negli ultimi anni. Tuttavia, queste accuse si sono spesso rivelate infondate e ingiuste, frutto del tentativo di individuare un capro espiatorio anziché analizzare con maggiore obiettività la realtà empirica. In molti casi, infatti, il problema non risiedeva negli strumenti finanziari in sé, ma nell'uso improprio o fraudolento che ne è stato fatto. Un approccio etico e responsabile all'utilizzo degli strumenti finanziari da parte di tutti gli attori del sistema economico-finanziario potrebbe permettere di valorizzarne le finalità originarie e le potenzialità positive, spesso offuscate da comportamenti distorsivi. Non mancano, tuttavia, studiosi che propongono visioni alternative in merito ai rapporti tra i due sistemi economici — finanziario e reale — e che riconoscono nelle potenzialità dell'ingegneria finanziaria un'opportunità per entrambi. Tale prospettiva incoraggia l'analisi dei legami tra i due ambiti e l'identificazione delle leve strategiche da attivare al fine di massimizzare il rendimento degli strumenti finanziari a beneficio dell'intero sistema economico. In qualsiasi contesto ci si trovi, e da qualunque prospettiva si consideri una crisi — sia essa sistemica o localizzata — una dimensione fondamentale da tenere in conto è quella sociologica.

Gli effetti di una crisi, infatti, non si distribuiscono in modo uniforme: le disuguaglianze sociali, accentuatesi con l'avanzare della globalizzazione, producono impatti differenziati tra individui e gruppi, in un contesto in cui tali disparità sono spesso tollerate in misura significativa per l'instaurarsi di una sorta di indifferenza verso le condizioni sociali dei meno fortunati, che spinge ad un disinteressamento nei confronti delle realtà di vita di questi ultimi. Le differenze emergono anche tra le diverse fasce generazionali: i giovani, avendo vissuto i primi anni della loro vita in un contesto di crisi, considerano tale situazione come normale, sviluppando una rassegnazione rispetto alle restrizioni che la crisi comporta nelle molteplici sfere dell'esistenza individuale. È fondamentale che i decisori non considerino una crisi unicamente dal punto di vista economico-finanziario, come spesso fanno gli economisti, ma che valutino con pari attenzione anche le altre dimensioni coinvolte.

Tra queste, infatti, possono esserci realtà culturali, sociali e politiche che non solo possono avere origine da una crisi, ma possono essere a loro volta profondamente influenzate da essa. Solo prendendo in considerazione queste diverse dimensioni, i decisori potranno intraprendere iniziative volte a stimolare lo spirito di reazione degli individui, grazie al quale una società può superare una situazione di crisi. Ciò consentirebbe non solo di uscire dalla crisi stessa, ma anche di fare tesoro degli insegnamenti acquisiti, instaurando una condizione migliore rispetto a quella pre-crisi. Una crisi sociale rappresenta, quindi, un cambiamento negli approcci che gli individui adottano nei confronti della cultura, della politica, dell'ambiente che li circonda e delle loro relazioni interpersonali. Essa colpisce in modo differente gli individui a seconda dei diversi strati sociali. È fondamentale che lo studio delle crisi sociali rimanga un ambito prioritario della sociologia, per evitare che le azioni volte alla loro risoluzione vengano intraprese senza il supporto scientifico di chi ha fatto dello studio delle società la propria missione. Una risoluzione "fai-da-te" delle crisi sociali può infatti spingere gli individui a scegliere soluzioni semplicistiche, quali la chiusura verso la comunità, l'odio verso gli immigrati considerati capri espiatori, la segregazione di intere comunità che si isolano come in un feudo, o la nascita di movimenti tendenti al totalitarismo. Tali reazioni, già osservate nella storia, hanno prodotto risultati negativi e dannosi per le società. Il concetto di crisi è spesso associato a discipline diverse dalla sociologia e, frequentemente, viene considerato pertinente ad altri ambiti scientifici. Tuttavia, sarebbe auspicabile un più ampio coinvolgimento della sociologia nello studio delle crisi. Tale approccio potrebbe dare i migliori risultati se applicato durante la manifestazione stessa della crisi, permettendo di analizzarne gli effetti, la natura e le eventuali traiettorie di trasformazione sociale che ne derivano. Un modello di società che tende a trasformarsi attraverso una crisi non esclude la possibilità di generare un nuovo modello migliore e più efficiente, sia dal punto di vista culturale, delle relazioni interpersonali, professionali e ambientali, sia in termini di una rinnovata umanità che permea i diversi ambiti che compongono la società stessa. La *crisiologia* non può fondarsi esclusivamente sulla finanza e sull'economia, ma dovrebbe integrare la sociologia, che può permeare le analisi e gli studi delle crisi con una prospettiva ed una direzione coerenti con il compito, di grande importanza etica, di governare le stesse. Fino a qualche decennio fa, l'analisi della società e dei suoi stati di crisi veniva condotta principalmente all'interno della prospettiva della nazione e dello Stato. La società era infatti identificata con lo Stato-nazione, e tutte le riflessioni sulla società si sviluppavano in funzione di questo concetto fondamentale. Col tempo, fortunatamente, questa visione ristretta, poco propensa a considerare le società come realtà globali, è andata progressivamente superandosi. Gli studi sociali, e dunque anche l'analisi delle realtà che le caratterizzano o le colpiscono — come le crisi — hanno cambiato prospettiva, assumendo il sistema sociale come un'entità globale, caratterizzata da interazioni, influenze e specificità proprie delle società mondiali, e adottando così un approccio di analisi globale. Questo fenomeno risulta ancora più evidente ai giorni nostri, in un contesto in cui le tecnologie di diffusione di massa —

delle notizie, delle culture, dei livelli di istruzione, del benessere, ecc. — rendono le società sempre più interconnesse e reciprocamente influenzate. Di conseguenza, caratteristiche ed eventi che le coinvolgono non possono essere analizzati, studiati o interpretati senza tenere conto delle interazioni e dei condizionamenti che tali eventi stessi generano all'interno delle società. Tale dinamica è talmente evidente e pervasiva da apparire chiara anche ad un osservatore minimamente attento. Le crisi emerse verso la fine della prima decade degli anni 2000 hanno evidenziato come il mondo moderno, e le società che lo compongono, siano profondamente interconnessi a livello globale, in ogni ambito. Un esempio emblematico è rappresentato dalla crisi finanziaria dei mutui sub-prime: le difficoltà nate negli Stati Uniti si sono propagate rapidamente a livello internazionale, manifestando una diffusione tanto estesa quanto celere. Anche eventi terroristici avvenuti in aree geograficamente lontane hanno prodotto effetti tangibili e immediati su scala globale. In diversi casi, le reazioni delle nazioni non direttamente colpite si sono concretizzate in azioni mirate a prevenire o mitigare le ricadute negative di eventi originatisi altrove. Anche eventi naturali avversi, verificatisi in una specifica area del pianeta, possono produrre effetti negativi — ad esempio di tipo economico — in altre parti del mondo, soprattutto quando compromettono in modo significativo il sistema economico del paese colpito. In tali casi, le ripercussioni si estendono ai partner commerciali, anche se situati a migliaia di chilometri di distanza. Tutto è interconnesso: le società sono strettamente legate tra loro perché il contesto globale in cui nascono e si sviluppano è profondamente mutato rispetto al passato. Di conseguenza, partendo dal presupposto che le società sono oggi interdipendenti su scala globale, è logico ritenere che anche i saperi debbano essere sempre più interconnessi. Pertanto, l'analisi di una società — ed, ancor più, lo studio di una crisi sociale — non può prescindere da un approccio interdisciplinare. Ciò che può apparire come una crisi della sociologia — se si assume la definizione classica del concetto di *crisi* — può in realtà rappresentare un suo approdo a una nuova configurazione, differente da quella precedente, determinata dai profondi mutamenti che il mondo ha vissuto e continuerà a vivere. In questo senso, più che perdere rilevanza, la sociologia può affermare il proprio ruolo cruciale nell'analisi delle crisi contemporanee. La crisi di un sistema sociale globale può essere intesa come un indicatore della gravità di una situazione che coinvolge simultaneamente più paesi e popolazioni. Tali crisi, se non governate, possono condurre a cambiamenti drastici, generare conflitti tra i popoli e, nei casi più estremi, sfociare in guerre su scala mondiale.

Grandi cambiamenti negli equilibri di potere sullo scacchiere internazionale possono condurre le popolazioni ad uno stato di crisi, alimentato dall'incertezza che i nuovi assetti geopolitici inevitabilmente comportano. Le decisioni connesse alle politiche internazionali rivestono un'importanza cruciale nella prevenzione e nella gestione delle crisi che colpiscono popoli e società. Di conseguenza, una crisi interna ai sistemi internazionali può innescare dinamiche di transizione da una situazione pre-crisi a una post-crisi, che, in assenza di una risposta adeguata, rischia di generare effetti deleteri su scala globale, soprattutto in un contesto mondiale incapace di

affrontare cambiamenti drastici e negativi. È inoltre fondamentale sottolineare che una crisi sociale coinvolge una comunità di individui, guidata da attori politici che, nel processo decisionale, non possono prescindere da proprie percezioni, idee personali, visioni etiche e livelli soggettivi di sensibilità. Le soluzioni proposte, pertanto, saranno inevitabilmente influenzate da tali fattori, il che implica che approcci diversi possano condurre a esiti differenti. Ne consegue che ogni crisi sociale troverà risposte fondamentalmente diverse a seconda delle persone che, in un determinato momento storico, si trovano alla guida della collettività. Una struttura solida di relazioni internazionali è fondamentale per garantire una risposta efficace in caso di crisi. Al contrario, una debolezza intrinseca in tale struttura può fare in modo che una crisi generi conseguenze tali da compromettere, anche dopo il suo superamento, la piena ripresa del sistema internazionale. In tali circostanze, diventa ancora più difficile favorire l'emergere di nuovi equilibri globali. Esiste inoltre il rischio concreto che, in seguito a una crisi mal gestita, il sistema internazionale entri in una fase di incertezza prolungata, priva di una visione chiara delle regole e delle relazioni che ne governano il funzionamento, precipitando così in uno stato di disorientamento e instabilità. Il secolo scorso può essere considerato, a buon diritto, come quello in cui si sono verificate le più profonde e significative trasformazioni nelle relazioni internazionali, grazie anche alla creazione di organismi sovranazionali volti alla risoluzione pacifica dei conflitti tra Stati. Tuttavia, accanto a questo fermento istituzionale e alle scoperte tecnologiche che hanno radicalmente trasformato la vita degli individui, il XX secolo è stato anche teatro di due dei più devastanti conflitti della storia dell'umanità. Sarà ricordato, infatti, non solo per l'ingegno e l'innovazione, ma anche per aver posto tali conquiste al servizio della produzione di strumenti di morte e distruzione su una scala mai vista prima. Dopo l'ultima grande guerra si erano consolidati equilibri internazionali nei quali i vincitori detenevano la legittimità di determinare le sorti degli Stati sconfitti. Tuttavia, gli eventi che hanno caratterizzato la fine degli anni '80 del secolo scorso hanno profondamente modificato tali assetti, senza che si sia riusciti a instaurare un nuovo sistema internazionale stabile. Quegli anni possono essere considerati una vera e propria crisi del sistema internazionale, che ha coinvolto inevitabilmente le società e gli individui di tutto il mondo. A differenza della situazione di equilibrio precedente alla crisi, non si è venuto a creare un nuovo equilibrio, né simile né differente, nell'epoca successiva. In conseguenza di ciò, il sistema internazionale non è stato in grado di definire un nuovo ordine, lasciando gli Stati in una condizione di forte vulnerabilità e instabilità, con un aumento significativo del rischio di derive potenzialmente catastrofiche per le società, soprattutto in assenza di organismi regolatori internazionali capaci di dirigere le nazioni verso una visione del mondo condivisa. È come se i popoli avessero cominciato a considerare solo sé stessi, trascurando gli altri come partecipanti di un mondo globale in cui regole, accordi e legami di fratellanza tra individui sembrano non esistere più o, al meglio, conservano importanza solo in funzione di interessi economici.

L'altro viene visto prevalentemente come un mezzo per raggiungere fini economici personali, a discapito della sua dimensione più umana e intima. Una visione che quasi richiama i principî de "La Ricchezza delle Nazioni" ma con la differenza che il bene dell'altro — in questo nuovo pensiero del mondo — può essere sacrificato se questo è funzionale ed utile al mio bene (economico). Una visione di questo tipo, priva di un sistema internazionale stabile, potrebbe risultare funzionale agli individui solo se si fondasse sulla centralità sociologica degli stessi, cioè non su ideologie, interessi o alleanze finalizzate ad un beneficio personale, ma piuttosto sulla condivisione di costumi, tradizioni e culture, sull'inclusione e sul benessere di tutti i popoli, attraverso una distribuzione equa delle risorse che il mondo offre gratuitamente a tutte le persone, senza distinzione. Un esempio evidente di questa dinamica si osserva nello scontro tra le strutture sociali occidentali e quelle dei paesi islamici, le cui popolazioni sono da tempo coinvolte in flussi migratori verso l'Occidente. Nonostante le diverse concezioni dell'essere umano presenti in queste due realtà sociali, fino ad oggi non si è mai tentato seriamente di costruire un'interconnessione basata sui valori reciproci, e in parte comuni, su cui tali società si fondano. L'assenza di approcci basati su tali valori da parte degli Stati — in particolare di quelli più forti ed influenti — unita alla mancanza di una regolamentazione internazionale efficace, rappresenta un evidente fattore di rischio. Tale carenza può facilmente spingere i popoli verso crisi di portata devastante per l'umanità. Inoltre, la forza di questi Stati, se non guidata da principi etici improntati al rispetto assoluto della dignità umana e se priva di istituzioni internazionali solide ed autorevoli, rischia di essere impiegata per fini utilitaristici statali piuttosto che per il bene collettivo globale. Etica, trasparenza e fiducia da parte di coloro che, a vario titolo, decidono le sorti degli Stati dovrebbero rappresentare i pilastri di una società moderna. La loro assenza o una presenza debole costituisce certamente un segnale di rischio, destinato prima o poi a sfociare in una vera e propria crisi sociale. Come evidenziato in apertura, il termine *crisi* implica una reazione ad un evento o ad una situazione che irrompe all'interno di un sistema, provocandone la destabilizzazione. Nel contesto sociale, il concetto rimane invariato, ma un ulteriore fattore di rischio può risiedere nella rassegnazione degli individui poichè, abituati a situazioni di crisi continue e sistematiche, potrebbero progressivamente perdere la volontà e l'atteggiamento reattivo necessari per affrontare le difficoltà sopravvenute. Tale rassegnazione può essere ulteriormente accentuata quando sono le stesse strutture guida della società — come le istituzioni statali, la politica, i sistemi religiosi, le associazioni, le tradizioni e la famiglia — a trovarsi in una fase di declino. E anche qualora si manifestasse una risposta collettiva ad una situazione di crisi, in presenza di sentimenti diffusi di scoramento e rassegnazione all'interno delle società, tale reazione rischierebbe di essere un gesto automatico e obbligato, privo di uno spirito di entusiasmo o di una volontà di cambiamento forte e decisa. Di conseguenza, la condizione collettiva potrebbe rimanere sostanzialmente invariata rispetto a quella pre-crisi, alimentando anzi ulteriormente questo senso di scoramento e rassegnazione, con tutte le ripercussioni negative che ciò comporta sul benessere individuale e

sociale. Scelte compiute in stati mentali ed emotivi caratterizzati da scoraggiamento collettivo possono indurre, negli individui, la formazione di convinzioni secondo cui gli atteggiamenti resilienti nei confronti delle crisi sociali siano sostanzialmente inefficaci. In tale visione, la reazione alla crisi viene percepita come inutile, poiché il sistema appare comunque destinato a permanere in uno stato di negatività. Ne deriva una sfiducia generalizzata nella possibilità che lo sforzo individuale o collettivo possa realmente produrre un cambiamento importante. Tali dinamiche possono facilmente attecchire negli individui, soprattutto in contesti sociali fragili. Proprio per questo motivo, le strutture sociali guida — a cui si faceva riferimento in precedenza — hanno il compito ed il dovere fondamentale di prevenirne anche la più minima insorgenza. Ciò richiede che esse siano costantemente sane, attive e vigili, operando con responsabilità per il benessere collettivo ed individuale. Un ulteriore rischio concreto risiede nel fatto che tali atteggiamenti possano indurre il collettivo a concentrarsi esclusivamente sulla dimensione di insicurezza, debolezza e vulnerabilità generata dallo stato di crisi, piuttosto che orientare le proprie energie verso le opportunità di cambiamento, resilienza e sviluppo che una crisi, se ben gestita, può offrire. Studi sociologici hanno evidenziato come, tendenzialmente, il concetto di crisi venga associato quasi automaticamente a situazioni di difficoltà, incertezza e negatività, piuttosto che essere interpretato come occasione di trasformazione positiva. L'attenzione tende a concentrarsi sugli aspetti disfunzionali, tralasciando le potenzialità di cambiamento, le opportunità di miglioramento e le innovazioni nella vita degli individui e delle collettività che una crisi, per sua natura, può stimolare e portare seco. In tal senso, l'azione politica dovrebbe orientarsi verso interventi capaci di guidare e accompagnare gli individui, attraverso opportune regolamentazioni, verso atteggiamenti di risposta e reazione positiva agli stati di crisi. Tale operato dovrebbe essere supportato da studi e analisi di tipo psicologico e sociale, finalizzati ad attenuare, all'interno della collettività, sentimenti di instabilità, insicurezza, paura e altre condizioni emotive che possano ostacolare la capacità di reazione e l'entusiasmo nel fronteggiare le difficoltà. L'obiettivo dovrebbe essere quello di orientare gli animi collettivi verso una visione costruttiva della crisi, intesa come occasione di cambiamento e miglioramento. Una reale collaborazione tra ambiti diversi — politico, psicologico, sociologico ecc. — potrebbe rafforzare significativamente la capacità di resilienza delle società, permettendo di cogliere appieno le potenzialità positive che possono scaturire da situazioni di crisi sociale, di qualunque natura esse siano. In questo quadro, la politica, la comunità scientifica e tutti coloro che ricoprono ruoli di rilievo intellettuale e culturale dovrebbero sentirsi guidati da un forte volano etico, impegnandosi attivamente — e senza riserve — nel perseguimento del bene collettivo. In virtù delle loro competenze e dei mezzi a loro disposizione, essi hanno la responsabilità di supportare quella parte di società che, per carenza di strumenti, idee o risorse, non è in grado di affrontare autonomamente momenti di crisi. Solo così sarà possibile trasformare la crisi da momento di disgregazione a opportunità di sviluppo e miglioramento del vivere sociale, in piena coerenza con il significato profondo del termine *crisi* più volte richiamato in questo

scritto: ossia la capacità di scegliere, di individuare la via migliore da seguire in un momento di difficoltà, reagendo attivamente anziché arrendersi. Sulla base della definizione di crisi adottata in questo scritto — intesa come momento di scelta in risposta a una situazione avversa — si può generalizzare affermando che molte condizioni di sconfitta, negatività o decadimento culturale delle società siano riconducibili a scelte erranee o inadeguate compiute nel tentativo di affrontare una crisi sopravvenuta. Questa riflessione può essere estesa a qualsiasi ambito, sociale o individuale, proprio in virtù della natura universale e trasversale del concetto stesso di crisi. In un contesto sociale, la transizione attraverso fasi di negatività, conflitto e degrado dei valori appare inevitabile, proprio in virtù della pluralità che caratterizza i suoi componenti. Non è mai esistito, nella storia dell'umanità, un momento in cui tali dinamiche siano state completamente assenti. È infatti l'eterogeneità stessa dei comportamenti, dei valori, del senso civico, del livello di istruzione e della formazione culturale degli individui a costituire un terreno fertile per il manifestarsi di tensioni e criticità. È opinione diffusa tra molti sociologi che, nell'attuale fase storica, le società siano caratterizzate da un generale decadimento dei valori etici e morali, fenomeno imputabile alla frenesia del vivere moderno, all'assoggettamento crescente alla tecnologia e a una logica diffusa del profitto a tutti i costi. In questo contesto, si riscontra una sempre minore capacità collettiva di rispondere in modo adeguato agli stati di crisi. La crisi viene percepita e affrontata prevalentemente nella sua accezione negativa, come evento destabilizzante e disfunzionale, anziché come occasione di cambiamento, rinnovamento e miglioramento, significati che, come già discusso, sono impliciti nella stessa etimologia del termine. L'incapacità di reazione che si osserva in molte società contemporanee è, in larga parte, il risultato di un affidamento quasi totale ai fattori tecnologici e, in secondo luogo, alla ricerca spasmodica di ricchezza. Questa dinamica ha generato una sorta di *scientismo*, trasformando progressivamente gli individui in dipendenti passivi di tali pilastri del vivere moderno. Tale sottomissione comporta un progressivo indebolimento della capacità individuale e collettiva di far fronte alle difficoltà. Abituati ad avere soluzioni immediate grazie alla tecnologia, gli individui faticano a reagire quando essa viene meno. La mancanza di reale padronanza sugli strumenti tecnologici — di cui si dovrebbe essere, per lo più, semplici fruitori — espone le persone a stati di panico e paralisi operativa. Un improvviso blocco dei sistemi tecnologici può quindi compromettere la quotidianità, rivelando una profonda fragilità nella struttura reattiva delle società. Tale situazione si riflette profondamente sugli individui, inducendoli a percepire la crisi esclusivamente nella sua accezione negativa, trascurandone invece il significato autentico di cambiamento positivo e svolta migliorativa che essa dovrebbe comportare. L'incapacità di reagire efficacemente alle situazioni problematiche, derivante dai fattori sopra evidenziati, risulta particolarmente marcata nelle giovani generazioni. Di conseguenza, la capacità di interpretare e affrontare uno stato di crisi in termini costruttivi e positivi dipenderà in larga misura dallo stato dei valori, dal livello culturale e dall'etica che caratterizzano gli individui coinvolti. In questo contesto, la politica assume un ruolo di primaria importanza. Essa non solo deve

essere un'amministratrice saggia del potere, al servizio del bene collettivo, e dimostrare lungimiranza nelle scelte legislative volte a mediare gli interessi dell'intera società, ma deve anche possedere capacità progettuali che tengano conto del contesto sia nazionale sia internazionale. In una prospettiva segnata dalla diffusa sfiducia delle collettività, derivante dalle cause precedentemente esposte, la politica dovrebbe assumere il ruolo fondamentale di catalizzatore di fiducia e assicurazione. Essa è chiamata a promuovere condizioni mentali positive nella popolazione, capaci di aprire e facilitare la strada a risposte resilienti e consapevoli di fronte a situazioni di crisi, che possano interessare la società in modo diretto od indiretto. Ad esempio, uno stato di guerra in un paese lontano può apparentemente sembrare distante e non direttamente rilevante per altri stati. Tuttavia, al di là del dolore e dello sdegno che una guerra dovrebbe suscitare in ogni essere umano, una reazione solidale e consapevole da parte della comunità civile globale potrebbe rappresentare un potente stimolo per i governanti di tutti i paesi a ricercare una soluzione al conflitto. Questo non solo porterebbe benefici diretti alle popolazioni coinvolte, ma influirebbe positivamente anche su chi, apparentemente, non ne è direttamente interessato. Infatti, il benessere diffuso favorisce un ciclo virtuoso di crescita e stabilità, capace di generare effetti positivi a livello globale. Questo ultimo punto potrebbe apparire banale a prima vista, ma un'analisi storica evidenzia in modo eloquente come le scelte adottate in risposta a situazioni di crisi che hanno colpito le società siano state, nel tempo, generatori di nuovi valori che, spesso, tali non erano realmente, poiché mancavano di solide radici etiche. Pertanto, le scelte politiche e sociali generano mode e linee di comportamento all'interno delle società e, se tali scelte non sono guidate da principi sani ed etici, gli usi ed i modelli che ne derivano potrebbero non indurre negli individui stati virtuosi, ma ulteriori crisi di valori. Ciò crea un effetto a catena negativo che può condurre una società a livelli di civiltà inferiori rispetto a quelli antecedenti una crisi. Il ruolo della politica in una società è dunque fondamentale, poiché essa rappresenta l'istituzione deputata a prendere decisioni che influenzino l'intera collettività, non solo nel presente ma anche nel futuro. Una crisi impone necessariamente delle scelte, e tali scelte dovrebbero essere guidate esclusivamente da valori etici, al fine di garantire il bene comune. Tuttavia, situazioni che rischiano di compromettere questa missione primaria della politica sono legate a finalità privatistiche, ovvero all'uso della politica per perseguire interessi personali o di gruppo, mascherando tali azioni con un'apparente attenzione al bene collettivo. Non a caso, nei paesi più poveri del mondo— dove la popolazione vive in condizioni inammissibili e completamente anacronistiche rispetto agli standard della modernità occidentale — le classi dirigenti, ad ogni livello politico, sovente presentano livelli di corruzione elevatissimi. Sembrerebbe quasi come se questi leader utilizzassero la loro posizione di potere per fini personali di arricchimento, lasciando il popolo nella miseria e nella disperazione più profonde. Il rischio da scongiurare, anzi la radice di questi problemi da eliminare, è rappresentato dalla cosiddetta "*teoria dell'azzardo morale*", che mina — anzi, annulla completamente — il potere etico della politica. Di conseguenza, qualunque sia la situazione sociale, di crisi o meno, le scelte politiche si

riducono a logiche utilitaristiche e personalistiche, senza apportare alcun reale beneficio alle popolazioni di quei paesi, aggravandone anzi ulteriormente le condizioni. Tali dinamiche, seppur in forma meno accentuata a causa della diversa composizione sociale dell'Occidente, possono colpire anche gli stati occidentali e, dunque, le società moderne, portandole a livelli di benessere progressivamente inferiori. Questo genera scoramento ed una scarsa, se non nulla, capacità di reazione, alimentando sentimenti di rassegnazione anziché combatterli. È invece necessario infondere negli individui lo spirito di trasformare la paura dell'incertezza — che mina la reattività — in una forza motivante e propulsiva. Anche se si tratta di un compito oggettivamente complesso, la politica dovrebbe assumersi la responsabilità di infondere volontà collettiva e fiducia nella possibilità di individuare, all'interno di una crisi, elementi di rinnovamento e trasformazione, anziché limitarsi a una gestione meramente reattiva o focalizzata esclusivamente sugli aspetti patologici dell'emergenza. In quest'ottica, una formazione ispirata all'etica della responsabilità politica diventa fondamentale affinché le società siano in grado di affrontare con efficacia le criticità e migliorare, nel medio-lungo periodo, la propria condizione. Un approccio etico di questo tipo, oltre a facilitare la gestione delle crisi, può anche contribuire a prevenirne l'insorgenza o a mitigarne gli effetti sistemici. Si parla, quindi, non tanto di eliminare le crisi — evenienza storicamente improbabile — quanto di imparare a leggerle come potenziali momenti di svolta, capaci di generare processi trasformativi che, se adeguatamente gestiti, possono condurre a stati di maggiore equilibrio e benessere rispetto alla fase antecedente.

Tali processi di crisi, se non gestiti in modo strutturato e lungimirante, compromettono profondamente il benessere collettivo, specialmente laddove l'azione istituzionale risulti debole, frammentata o inadeguata in termini sia quantitativi che qualitativi.

## 1.2

### Brevi riflessioni sugli effetti delle crisi su alcuni dominî del *BES*

Le crisi economiche, ambientali, sanitarie ed in generale le crisi sociali presentano una natura intrinsecamente multidimensionale, che si riflette in maniera trasversale su vari livelli della vita individuale e collettiva. Per questo motivo, la loro gestione richiede approcci interdisciplinari e multisettoriali, capaci di integrare competenze provenienti dalle scienze sociali, economiche, ambientali, politiche, nonché la partecipazione attiva dei cittadini stessi, che non sono soltanto destinatari degli interventi, ma anche attori del cambiamento. Le ricadute delle crisi si manifestano in particolare su alcune dimensioni fondamentali del benessere, come la salute, il benessere economico, la qualità dell'ambiente, la coesione sociale, il benessere soggettivo, le quali sono rappresentate efficacemente dai domini individuati nel framework del Benessere Equo e Sostenibile (*BES*). Questi dominî offrono una chiave interpretativa utile per analizzare l'impatto complessivo delle crisi sulla qualità della vita dei cittadini e per orientare le politiche pubbliche verso obiettivi di sviluppo sostenibile ed equo. Come più volte si è detto, le crisi e le emergenze rappresentano situazioni caratterizzate da condizioni di forte disagio, instabilità o deterioramento che compromettono la normale funzionalità dei sistemi sociali, economici, sanitari, ambientali. Esse si manifestano spesso in modo simultaneo ed interconnesso, generando effetti a cascata che amplificano la complessità delle risposte necessarie per fronteggiarle influenzando non solo la qualità della vita degli individui, ma anche i principi di equità intergenerazionale e di sostenibilità collettiva. Le crisi, quando non adeguatamente fronteggiate, tendono ad amplificare le disuguaglianze esistenti ed a compromettere la resilienza dei sistemi sociali, economici, ambientali. In particolare, si osserva un peggioramento delle condizioni di benessere in diversi ambiti chiave. Uno di questi è sicuramente quello della *Salute* di una collettività. Infatti la salute costituisce una delle dimensioni fondamentali del Benessere di una collettività, in quanto riflette non solo lo stato fisico e mentale degli individui, ma anche la capacità dei sistemi sanitari e sociali di garantire condizioni di vita adeguate, accesso alle cure e prevenzione. Le crisi di diversa natura — in particolare quella sanitaria degli ultimi anni — hanno evidenziato con forza quanto questo ambito sia vulnerabile a shock sistemici. L'aumento della mortalità, la riduzione dell'accesso alle cure urgenti, il sovraccarico delle strutture ospedaliere sono solo alcuni degli effetti osservati. La salute degli individui, inoltre, è soggetta ai risvolti derivanti da crisi economiche e sociali le quali possono produrre impatti indiretti ma profondi su di essa. La disoccupazione, l'incertezza economica e l'insicurezza abitativa sono tutti fattori associati all'aumento di condizioni di stress cronico, ansia e depressione, influenzando negativamente il benessere psichico degli individui. Le fasce più vulnerabili della popolazione, come gli anziani, i minori e le persone in condizioni socio-economiche svantaggiate, risultano particolarmente esposte a tali conseguenze. Non meno impattanti sulla salute sono le crisi ambientali,

come quelle legate al cambiamento climatico o all'inquinamento, generano ricadute tangibili sullo stato di salute della popolazione, con l'aumento di patologie respiratorie, cardiovascolari e dell'incidenza di eventi estremi (ondate di calore, alluvioni) che mettono a rischio la vita e la sicurezza fisica degli individui. Nel complesso, il dominio della salute si configura come uno dei più sensibili agli effetti multidimensionali delle crisi, evidenziando la necessità di politiche integrate e resilienti che agiscano sia in ottica preventiva che reattiva. Come prima accennato, i vari ambiti di una vita sociale sono tra loro legati, interconnessi, ancor di più nelle società moderne. Dunque situazioni di crisi che possano colpire uno di essi hanno ripercussioni sugli altri. La salute forse è la sfera sociale che più può essere coinvolta da crisi che si manifestano in altre sfere della vita di una società. Una di queste è certamente il Benessere economico il quale rappresenta una componente centrale del Benessere Equo e Sostenibile, in quanto riflette le condizioni materiali di vita degli individui, il livello di sicurezza economica e la possibilità di accedere a risorse e opportunità. Le crisi economiche, come quelle legate a recessioni finanziarie, instabilità dei mercati del lavoro o shock geopolitici, generano effetti diretti sulla disponibilità di reddito, sull'occupazione e sul potere d'acquisto delle famiglie. L'aumento della disoccupazione, la crescita del lavoro precario e la compressione dei salari reali comportano una riduzione della qualità della vita e un peggioramento delle condizioni economiche soprattutto per le fasce più fragili della popolazione. In parallelo, l'accesso ai servizi essenziali – dalla casa all'istruzione, fino alla sanità – diventa più difficile per ampie porzioni della società, contribuendo ad alimentare disuguaglianze strutturali. Tutto questo non può non avere ripercussioni anche sullo stato di salute delle persone. Ma allo stesso modo, anche la salute impatta sullo stato di benessere economico di un popolo. Infatti, le crisi sanitarie, come anche quelle ambientali, possono a loro volta amplificare le difficoltà economiche: eventi climatici estremi possono danneggiare infrastrutture e attività produttive, riducendo la produttività e la sicurezza alimentare, mentre le pandemie possono paralizzare interi settori economici, come avvenuto nel caso del turismo, della ristorazione e del commercio durante l'emergenza degli ultimi anni. Inoltre, la crescente polarizzazione sociale e la concentrazione della ricchezza accentuate dalle crisi mettono in discussione i principi di equità e giustizia distributiva, minando la coesione sociale e la fiducia nelle istituzioni economiche e politiche. Per questo motivo, il dominio del benessere economico, forse ancor più degli altri, non può essere analizzato isolatamente, ma va considerato all'interno di un quadro integrato in cui le dimensioni della salute, dell'ambiente e della partecipazione sociale interagiscono in modo complesso e reciproco. Non è possibile stilare una gerarchia di priorità tra i domini del BES in termini di importanza assoluta, poiché tutti concorrono e interagiscono nel determinare il benessere collettivo. Tuttavia, alcuni domini presentano una rilevanza tale da poter amplificare e protrarre nel tempo gli effetti delle crisi che dovessero colpirli. Uno di questi è sicuramente il dominio *Ambiente*. Esso nel framework del Benessere Equo e Sostenibile rappresenta una dimensione cruciale per la valutazione della sostenibilità del benessere nel lungo periodo poiché

comprende la qualità dell'aria, delle acque, del suolo, la biodiversità e l'equilibrio degli ecosistemi, oltre alla disponibilità e all'uso delle risorse naturali. Tutte realtà fondamentali a garantire il benessere di un popolo nel tempo. Le crisi ambientali, spesso di natura sistemica e globale, costituiscono una delle minacce più gravi alla stabilità ecologica e al benessere delle generazioni presenti ed anche future. Fenomeni come il cambiamento climatico, l'inquinamento atmosferico, la perdita di biodiversità e la gestione insostenibile delle risorse naturali (ad esempio acqua, suolo e foreste) hanno effetti diretti e indiretti su molteplici aspetti della vita quotidiana. Gli eventi climatici estremi, sempre più frequenti e intensi, mettono a rischio la salute, la sicurezza alimentare, l'abitabilità dei territori e l'equilibrio economico delle comunità più vulnerabili minando profondamente il benessere collettivo. Tali crisi ambientali sono spesso interconnesse con le crisi sanitarie, come nel caso delle zoonosi favorite dalla deforestazione, dalla distruzione degli habitat naturali e degli ecosistemi che alterando l'equilibrio ecologico spinge gli animali verso aree abitate dall'uomo favorendone la trasmissione. Le crisi ambientali hanno una forte connessione anche con le crisi economiche poiché creano nel tempo danni alle infrastrutture, la perdita di produttività agricola e fanno aumentare notevolmente i costi della transizione ecologica. Inoltre, l'iniquità nella distribuzione degli impatti ambientali — che colpiscono più duramente le popolazioni marginalizzate e i territori meno attrezzati — solleva questioni rilevanti di giustizia ambientale e sociale. In quest'ottica, il dominio *Ambiente* non deve essere considerato solo come una componente accessoria del benessere, ma come un vero pilastro, una componente strutturale su cui poggia la sostenibilità dell'intero sistema socio-economico. La capacità delle istituzioni di attuare politiche ambientali efficaci, inclusive e lungimiranti rappresenta quindi un indicatore essenziale del livello di benessere sociale. Le società per loro definizione si basano sul legame sociale, relazioni sociali. Con questo concetto ci si riferisce alla qualità dei legami interpersonali, al capitale sociale, alla fiducia reciproca e al senso di appartenenza ad una comunità. È una dimensione centrale del benessere individuale e collettivo, poiché relazioni solide e reti di supporto sociale contribuiscono non solo al benessere soggettivo, ma anche alla capacità delle comunità di affrontare e superare situazioni di crisi. Le crisi sociali — come l'aumento delle disuguaglianze, l'erosione della fiducia nelle istituzioni, la marginalizzazione di alcuni gruppi sociali — minano in modo diretto la coesione sociale e il senso di solidarietà, favorendo fenomeni di isolamento, conflitto e disgregazione. Nell'ottica dei profondi legami esistenti tra i dominî del BES, anche le crisi economiche e sanitarie possono compromettere la qualità delle relazioni sociali: la precarietà lavorativa e abitativa, l'insicurezza economica e le restrizioni alla mobilità (come avvenuto durante le ultime crisi sanitarie) riducono le opportunità di socializzazione e aumentano il rischio di esclusione. Allo stesso tempo, le relazioni sociali costituiscono una risorsa strategica nelle fasi di crisi, poiché favoriscono la circolazione di informazioni, il mutuo aiuto, la cooperazione tra cittadini e tra istituzioni e società civile. In questo senso, la presenza di reti comunitarie forti può svolgere un ruolo di fattore protettivo, mitigando l'impatto negativo delle crisi su altri

domini, come la salute mentale, il benessere economico e il benessere soggettivo. L'implementazione di interventi mirati alla mitigazione degli effetti di una crisi in un singolo dominio del BES non produce benefici isolati, ma attiva dinamiche intersettoriali che si propagano positivamente agli altri ambiti del benessere. Questa interdipendenza sistemica implica che il rafforzamento di un dominio — come ad esempio quello della salute o dell'istruzione — possa generare effetti moltiplicatori su altri domini, contribuendo alla resilienza complessiva della collettività ed al ripristino di condizioni di equilibrio socioeconomico e ambientale. Ad esempio, politiche efficaci nel contrasto alla povertà e all'esclusione sociale possono avere un impatto diretto sul benessere soggettivo e sulla salute psicofisica della popolazione, con ricadute positive anche in termini di partecipazione culturale e coesione sociale. Dunque, il rafforzamento del capitale sociale attraverso politiche orientate alla partecipazione, all'inclusione, all'accoglienza ed alla promozione del volontariato e dell'associazionismo rappresenta un investimento fondamentale per la resilienza collettiva e per la tenuta democratica delle società in tempi di instabilità. Concetto assai affine a quello delle relazioni sociali è quello del Benessere Soggettivo, il quale riguarda la percezione che gli individui hanno della propria vita, del proprio stato emotivo e della soddisfazione rispetto a condizioni personali, relazionali e ambientali. Tale dimensione, fortemente interconnessa con gli altri domini del BES, rappresenta un indicatore essenziale della qualità della vita, poiché integra valutazioni oggettive con elementi legati alla sfera emotiva e psicologica dell'individuo. Le crisi, di qualunque natura esse siano — economica, sanitaria, sociale, ambientale — hanno un impatto rilevante sul benessere soggettivo, influenzando il senso di sicurezza, la fiducia nel futuro, la speranza e il grado di soddisfazione personale. La perdita del lavoro, la precarietà, l'isolamento sociale, la paura per la propria salute o quella dei propri cari, così come la percezione di vivere in un ambiente degradato o minacciato, generano una sensazione diffusa di incertezza e vulnerabilità che si riflette negativamente sul benessere percepito da ogni singolo individuo. Inoltre, eventi traumatici collettivi come le pandemie, le crisi climatiche o le guerre possono portare ad un deterioramento della salute mentale su vasta scala, con un aumento dei livelli di stress, ansia e depressione. Questo effetto è spesso più marcato tra i giovani, gli anziani e le persone che vivono in condizioni socio-economiche fragili, per i quali le prospettive di miglioramento risultano più limitate e difficili da raggiungere. La dimensione soggettiva del benessere rappresenta quindi non solo un indicatore di sintesi dell'impatto delle crisi sugli individui, ma anche un punto di partenza per l'elaborazione di politiche pubbliche più sensibili alle percezioni e ai bisogni emotivi dei cittadini. In quest'ottica, promuovere ambienti sociali inclusivi, garantire l'accesso ai servizi di supporto psicologico su larga scala e costruire narrazioni collettive che valorizzino la propensione a reagire e la partecipazione attiva, sono strategie fondamentali per rafforzare il benessere soggettivo durante una crisi. Non si poteva non spendere delle parole per parlare di un dominio fondamentale nel presente e per il futuro; il dominio della *Istruzione* e *Cultura*. Esso infatti rappresenta un pilastro strategico del BES, in quanto concorre alla formazione del capitale umano,

alla promozione della cittadinanza attiva e alla costruzione di società resilienti ed inclusive. L'accesso ad un'istruzione di qualità e alla partecipazione culturale è essenziale per garantire pari opportunità, per sviluppare capacità critiche e competenze trasversali, e per rafforzare la coesione sociale. Le crisi economiche, sanitarie e sociali hanno effetti profondi su questo dominio. L'ultima crisi sanitaria, ad esempio, ha reso particolarmente evidente quanto le disuguaglianze educative possano ampliarsi in situazioni di emergenza: la chiusura delle scuole, la didattica a distanza e le disuguaglianze nell'accesso alle tecnologie hanno penalizzato soprattutto gli studenti provenienti da contesti socio-economici svantaggiati. Le conseguenze si sono riflesse non solo sull'apprendimento, ma anche sul benessere psicofisico degli studenti e sulla loro socializzazione. Analogamente, le crisi economiche possono ridurre le risorse destinate ai sistemi educativi e culturali, limitando l'offerta formativa, l'accesso a servizi educativi pubblici e la partecipazione alla vita culturale. In particolare, la fruizione della cultura — intesa non solo come consumo, ma anche come produzione, scambio e partecipazione simbolica — tende a diminuire in contesti di incertezza e precarietà, con ricadute negative sul benessere soggettivo e sulla qualità del tessuto sociale. Cultura e istruzione, tuttavia, rappresentano anche leve fondamentali per affrontare le crisi. Un sistema educativo inclusivo e innovativo può sviluppare nei cittadini le competenze necessarie per interpretare la complessità, adattarsi ai cambiamenti e partecipare attivamente alla costruzione di soluzioni sostenibili. La cultura, da parte sua, svolge un ruolo cruciale nel rafforzare l'identità collettiva, nell'alimentare la creatività e nel promuovere il dialogo interculturale, contribuendo così a ricostruire legami sociali e visioni condivise del futuro. Come più volte specificato, le diverse dimensioni del *Benessere Equo e Sostenibile* qui sommestamente analizzate non possono essere considerate isolate l'una dall'altra. Al contrario, esse sono profondamente interconnesse e interdipendenti, in un sistema complesso in cui gli effetti delle crisi si propagano e si amplificano attraverso molteplici canali. Il deterioramento della salute non solo compromette il benessere individuale ma può avere ripercussioni sul benessere economico, attraverso la riduzione della produttività e l'aumento dei costi sanitari; analogamente, le difficoltà economiche influiscono sulle opportunità educative e sulla qualità delle relazioni sociali. Le crisi ambientali, inoltre, rappresentano un fattore di pressione trasversale che agisce come un moltiplicatore di vulnerabilità, influenzando negativamente la salute pubblica, le risorse economiche e le condizioni di vita delle comunità più fragili. In questo contesto, la resilienza sociale — fondata su relazioni solide, capitale sociale e benessere soggettivo — svolge un ruolo cruciale nel mitigare gli impatti negativi e favorire percorsi di recupero e trasformazione. Il quadro multidimensionale offerto dai domini del *BES* fornisce uno strumento analitico prezioso per comprendere come le crisi economiche, sociali, sanitarie, ambientali o di altro genere impattino in modo articolato sulla qualità della vita delle persone e sulle prospettive di sviluppo delle società. Solo riconoscendo e affrontando queste interconnessioni è possibile progettare politiche integrate, efficaci e orientate alla sostenibilità, che siano in grado di ridurre le disuguaglianze, promuovere la coesione sociale e garantire la

tutela dell'ambiente. In questo senso, il *BES* non rappresenta soltanto un sistema di indicatori, ma un vero e proprio paradigma per orientare la governance pubblica verso obiettivi di benessere equo e sostenibile, capace di affrontare la complessità delle crisi contemporanee e di accompagnare le società lungo percorsi di rinnovamento e resilienza. Può, a tutti gli effetti, essere considerato un documento patrimonio collettivo, in quanto pone al centro dell'azione pubblica l'essere umano, inteso nella sua dimensione individuale e sociale, in armonia con l'ambiente e con il contesto relazionale in cui è inserito. Data la natura interconnessa dei domini del *BES* e del *BesT*, un loro studio richiede metodi statistico-econometrici capaci di cogliere la natura dinamica, interdipendente e spesso non stazionaria degli indicatori coinvolti. In questo lavoro si è voluto sperimentare l'utilizzo dei processi stocastici co-integrati per una analisi di alcuni domini per i motivi che verranno esplicitati nei prossimi capitoli. Intanto si può dire che tali processi rappresentino uno strumento metodologico fondamentale in questo contesto, in quanto consentono di modellare relazioni di equilibrio di lungo periodo tra variabili che, pur mostrando trend e deviazioni temporanee, sono legate da vincoli strutturali di lungo termine. Il documento *BES*, così come quello *BesT*, come detto, sono rappresentati da una serie di indicatori che riflettono fenomeni complessi, interconnessi e dinamici nel tempo. Questi indicatori sono spesso caratterizzati da trend di lungo periodo e da oscillazioni dovute a fattori transitori o shock esogeni. Un aspetto cruciale nell'analisi di tali dati è la presenza di relazioni di equilibrio di lungo termine tra i diversi domini del benessere. Sebbene ciascun indicatore possa presentare un comportamento non stazionario, con tendenze di crescita o declino nel tempo, i domini sono spesso legati tra loro da vincoli strutturali che ne regolano la co-evoluzione. Ad esempio, un miglioramento del benessere economico può correlarsi a un aumento della salute pubblica, mentre la degradazione ambientale può influire negativamente sia sulla salute che sul benessere soggettivo. Ma ancor di più si possono apprezzare i legami interni ad un dominio, ovvero i legami che intercorrono tra i suoi indicatori. I processi stocastici co-integrati costituiscono uno strumento particolarmente efficace per modellare e analizzare queste relazioni, grazie alla loro capacità di individuare e stimare relazioni di equilibrio a lungo termine tra variabili non stazionarie. L'adozione di questi modelli per l'analisi degli indicatori presenta numerosi vantaggi. Ad esempio, essi permettono di separare gli effetti momentanei da quelli strutturali, consentendo di comprendere se un cambiamento osservato è destinato a persistere o a essere solo temporaneo. Ancora, i modelli co-integrati potrebbero anche tener conto delle relazioni simultanee tra domini diversi, migliorando la capacità predittiva rispetto a modelli che considerano gli indicatori singolarmente ai fini predittivi. Grazie alla struttura dinamica, sarebbe possibile analizzare come eventi imprevisti – come crisi economiche, sanitarie, ambientali – influenzino non solo un singolo dominio, ma l'intero sistema di benessere, evidenziando percorsi di squilibrio ed aggiustamento. Inoltre, fornendo una rappresentazione integrata e dinamica dei diversi aspetti del benessere, i modelli co-integrati offrono uno strumento analitico robusto per orientare interventi e strategie di lungo termine,

favorendo un approccio multidimensionale e sostenibile. Essi possono essere anche funzionali allo studio ai fini predittivi, delle differenze territoriali del benessere. Si può certamente dire che i processi stocastici co-integrati rappresentino un approccio metodologico avanzato e adeguato per affrontare la complessità delle dinamiche multidimensionali e temporali del benessere equo e sostenibile, offrendo uno strumento prezioso sia per la comprensione teorica sia per la definizione di politiche pubbliche efficaci. Si può concludere dicendo che tali metodi statistici permettono di analizzare simultaneamente variabili interdipendenti, riflettendo la complessità reale dei fenomeni studiati, permettendo di modellare le relazioni di equilibrio a lungo termine migliorando la qualità delle previsioni, fondamentale per anticipare trend ed identificare segnali precoci di crisi, oppure permette (attraverso l'impiego di serie simulate coerentemente col fenomeno oggetto di studio) di testare l'effetto di interventi e shock, facilitando la pianificazione di azioni efficaci e mirate a livello nazionale e territoriale. Nel prosieguo del lavoro verranno specificate le motivazioni che hanno giustificato l'utilizzo del concetto di co-integrazione nell'ambito degli indicatori interni ad un dominio, studiandone le interconnessioni intra-dominio.

## CAP. 2

### Una metodologia per lo studio dei cambiamenti di un sistema complesso: La teoria delle catastrofi e suoi principi matematici fondamentali

#### 2.1

#### Principi fondamentali e nozioni di topologia

Nel presente capitolo ci si propone l'obiettivo -con molta modicità di intento data la complessità dell'argomento che richiederebbe un'esposizione più approfondita e specifica- di presentare una teoria matematica sviluppata nel secolo scorso, la cui struttura concettuale risulta applicabile a diversi fenomeni caratterizzati da uno stato di criticità. Tali fenomeni possono riguardare ambiti fisici, naturali, ma anche contesti propri delle collettività umane come crisi economiche, sanitarie o sociali. L'approccio adottato sarà di tipo matematico-statistico, sia in chiave storica che predittiva, con l'intento di individuare, attraverso l'analisi dei dati, quei segnali precoci che consentano di attuare misure preventive volte a mitigare il più possibile le conseguenze di tali eventi. Il concetto di *crisi*, inteso in senso generale -come verrà chiarito nel corso del capitolo- costituisce il fondamento della teoria oggetto di analisi. In particolare, si farà riferimento alla *teoria delle catastrofi* elaborata dal matematico e filosofo francese René Thom, con l'intento di esplorare in essa una possibile rappresentazione, anche matematica, delle situazioni di criticità. La teoria delle catastrofi fece la sua irruzione nel panorama scientifico verso la fine degli anni sessanta del secolo scorso. Essa nacque come tentativo di fornire una spiegazione scientifica alle variazioni significative, ai punti critici e dei mutamenti improvvisi -di tipo qualitativo- dei fenomeni oggetto di studio.

Tale teoria trae origine da studi pregressi effettuati sulle singolarità delle applicazioni continue e differenziabili, e sulla teoria delle biforcazioni dei sistemi dinamici. Nell'intento del suo autore, la teoria delle catastrofi aveva lo scopo di fornire una spiegazione dei fenomeni, ciascuno dei quali rappresentava una forma. Spiegare le forme che tali fenomeni incarnano significava, dunque, spiegare i fenomeni stessi. La realtà che ci circonda può essere vista come un catalogo di forme che nascono, si sviluppano fino a maturazione, interagiscono tra loro e infine muoiono. Thom definisce come *forma* una realtà che non subisce variazioni quando è sottoposta a piccole, ma continue, perturbazioni generate dall'ambiente in cui essa è inserita; in tal senso, essa risulta stabile rispetto a tali sollecitazioni esterne.

Il cammino che una realtà compie per raggiungere uno stato finale -ovvero una forma, secondo la teoria- è detto *creodo*. Questo condurrà sempre allo stesso risultato finale (una forma), indipendentemente dalle perturbazioni piccole ma continue che influenzeranno il fenomeno durante la sua evoluzione. Tale proprietà è fondamentale nella teoria ed è indicata con il termine *stabilità strutturale*.

Essa, inoltre, non esclude la possibilità che uno stesso fenomeno, manifestatosi in un determinato spazio-tempo con un certo risultato finale, possa giungere ad un esito diverso qualora si manifesti in uno spazio-tempo differente, oppure ad un risultato identico o solo approssimativamente simile. Questo può certamente accadere, poiché lo spazio-tempo influisce sui fenomeni e dunque sulle forme. Tuttavia, ciò che non dipende dallo spazio-tempo sono le leggi che regolano la formazione e l'evoluzione dei fenomeni. Esse saranno valide a prescindere dallo spazio-tempo in cui si manifesteranno. Tale concetto è detto *Invarianza delle teorie scientifiche nello spazio-tempo*, oppure *omogeneità dello spazio-tempo*. L'autore, con il termine *catastrofi*, si riferisce a quei casi in cui un fenomeno (ovvero una forma) subisce variazioni in seguito a piccole ma continue perturbazioni provenienti dal contesto in cui è inserito. Si tratta, dunque, di fenomeni privi della proprietà di *stabilità strutturale*. Tali forme, che rappresentano fenomeni instabili, vengono definite *discontinue*. L'idea centrale risiede nel concetto che, ponendo l'attenzione sulle catastrofi, è possibile studiare le forme. Ovvero, analizzando le piccole ma continue perturbazioni che agiscono sulle forme (cioè sui fenomeni) si possono comprendere le forme stesse. Una similitudine utile a comprendere meglio questa idea astratta potrebbe essere la seguente: si immagina un'altura dalla quale si diramano, in direzioni differenti, alcuni solchi che conducono ciascuno ad una diversa zona a valle. A ogni solco corrisponde una specifica area di arrivo, caratterizzata da configurazioni differenti. Si supponga ora di far rotolare, da un punto poco sotto il crinale dell'altura, una grande palla di neve. Questa, una volta giunta a valle, si scioglierà riempiendo d'acqua la zona di arrivo. Il solco che la palla seguirà determinerà il percorso e, di conseguenza, la specifica area in cui essa terminerà il suo tragitto. Anche se la palla venisse lanciata da un punto di partenza leggermente diverso, ma comunque confluenti nello stesso solco, il risultato finale non cambierebbe: la palla giungerebbe comunque nella medesima zona, sciogliendosi e generando lo stesso scenario. Questa situazione ipotetica rappresenta un sistema *strutturalmente stabile*: un *creodo*. Se, tuttavia, la palla venisse fatta partire non da un punto situato sotto il crinale dell'altura, ma esattamente dal crinale stesso, allora anche piccoli spostamenti del punto di partenza potrebbero determinare un incanalamento in un solco diverso, con conseguente percorrenza di un altro tragitto, arrivo in una zona differente a valle e, infine, scioglimento in un'area che darebbe origine a uno scenario del tutto diverso. Ebbene, le linee di confine tra i diversi scenari possibili a valle corrispondono, nella teoria delle catastrofi, a quelle che vengono definite *linee dell'insieme delle catastrofi*. I diversi scenari saranno separati da demarcazioni, ossia confini; in questo contesto, assume grande importanza il concetto di topologia e le proprietà topologiche che caratterizzeranno tali scenari.

Come lo stesso Thom asserì nei suoi scritti, il paradigma alla base delle sue idee era fondamentalmente qualitativo. Egli stesso, nel suo saggio sulla teoria delle catastrofi, propose un esempio per illustrare cosa intendeva con il termine *qualitativo*.

L'esempio consisteva nella scelta tra due teorie possibili,  $\theta_1$  e  $\theta_2$ , per spiegare un fenomeno  $\Phi$  rappresentato da una funzione sperimentale  $g$ , definita da una

equazione del tipo  $y=g(x)$  . Tali due teorie siano rappresentate rispettivamente dalle equazioni  $y=g_1(x)$  ed  $y=g_2(x)$  . Nessuna delle due teorie approssima quantitativamente in modo adeguato la curva sperimentale, anche se la curva  $g_1$  produce un valore della differenza tra la sua area sottostante e quella di  $g$  molto minore del valore della differenza prodotta tra l'area sottostante di  $g_2$  e  $g$  , in un dato intervallo. Questo si rappresenta con la seguente scrittura:

$$\int |g_1 - g| dx \ll \int |g_2 - g| dx$$

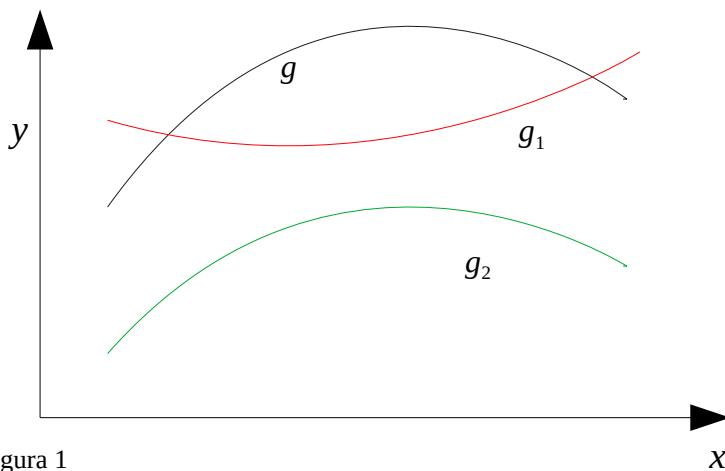


Figura 1

Come si nota dalla Figura 1, la curva  $g_1$  è più precisa quantitativamente per i motivi analitici espressi prima, ma la curva  $g_2$  , pur generando un errore assai maggiore tra la sua area sottostante e quella della funzione sperimentale in esame, ne specifica assai meglio -rispetto a  $g_1$  - il suo andamento qualitativo poichè possiede una forma assai simile alla curva sperimentale dunque, si può concludere che, malgrado un errore quantitativo assai maggiore, la teoria  $\theta_2$  , rappresentata dalla curva  $g_2$  , sia in grado di cogliere meglio le dinamiche sottostanti al fenomeno  $\Phi$  oggetto di studio.

Da tale esempio si evince che, nel lavoro di Thom, centro della spiegazione dei fenomeni è l'aspetto qualitativo nel senso appena visto.

La sua analisi parte dallo studio delle singolarità o criticità delle funzioni lisce, ovvero quelle funzioni nelle quali compaiono solo punti di massimo, di minimo e flessi. In intorni sufficientemente piccoli di tali punti, la approssimazione lineare tramite la retta tangente nelle singolarità non è soddisfacente, dunque servono delle approssimazioni più complesse che si possono ottenere qualora la funzione sia

derivabile un numero elevato di volte; tali funzioni sono dette *lisce* nel senso che esse non presenteranno spigoli.

Tali funzioni vengono classificate come equivalenti qualora abbiano la stessa *qualità* o, più precisamente, lo stesso *tipo topologico*.

Il passaggio da una classe all'altra di funzioni dunque avviene mediante una deformazione delle stesse. Esisteranno punti della deformazione in cui si formeranno o scompariranno nuovi massimi, minimi, flessi e dunque la funzione varierà il proprio andamento qualitativo. Gli insiemi formati da tali punti avranno il compito di classificare le diverse tipologie di funzioni e costituiranno l'insieme delle catastrofi. Sostanzialmente, conoscendo i punti critici o singolari di una funzione ed i rispettivi valori che la funzione assumerà in tali punti, risulterà possibile riconoscerne anche l'andamento qualitativo. Una funzione si deformerà quando varierà il suo tipo topologico. Dunque col nascere e morire di punti singolari si avrà una modificazione della forma preesistente; si parlerà dunque di *morfogenesi* o di creazione delle forme. Oltre a questi concetti matematici, come già sottolineato, la topologia assume un ruolo centrale nella teoria delle catastrofi poiché consente di rappresentare forme, stabilirne le loro uguaglianze o differenze, studiarne le diverse modificazioni.

Poiché essenziali per la rappresentazione della teoria, saranno introdotti alcuni concetti matematici fondamentali. Altri verranno esposti progressivamente, in base alle necessità che emergeranno durante lo sviluppo della teoria. Un concetto matematico fondamentale in questo contesto è quello di *spazio metrico*.

Dato un insieme  $X \neq \emptyset$  con  $X \subset \mathbb{C}_0$ , e dato un funzionale  $d: X \times X \longrightarrow \mathbb{C}_0$  allora il funzionale  $d$  sarà detto *distanza* -e la si indicherà con la scrittura

$d(x, y)$  con  $x, y$  punti appartenenti al prodotto cartesiano  $X \times X$  - se godrà delle seguenti proprietà:

$\forall x, y, z \in X$  dovrà risultare che,

i)  $d(x, y) \geq 0$

ii)  $d(x, y) = 0 \Leftrightarrow x = y$

iii)  $d(x, y) = d(y, x)$  ovvero la distanza è una misura *simmetrica*.

iv)  $d(x, y) \leq d(x, z) + d(y, z)$  questa proprietà, nota come *disuguaglianza triangolare*, prende il nome da alcuni teoremi sui triangoli. Definita la distanza, uno spazio metrico su un insieme  $X$  ed una metrica  $d$  è una coppia  $(X, d)$ , dove

$d$  rappresenta la metrica, come ad esempio quella Euclidea. Pertanto la coppia  $(X, d)$  costituisce uno spazio metrico se soddisferà le proprietà precedentemente enunciate. In uno spazio metrico, la modifica della metrica comporterà una variazione dello spazio metrico stesso; in altre parole, ad ogni metrica corrisponderà uno spazio metrico distinto, pur mantenendo invariato l'insieme su cui è definito. Dato uno spazio metrico  $(X, d)$  ed un numero  $r \in \mathbb{C}_0$ , varrà la seguente relazione:

$d_r(x, y) \cong r \cdot d(x, y)$ . In tal caso, il nuovo spazio metrico risulterà essere una versione scalata dello spazio metrico originale, dove la metrica è stata moltiplicata per un fattore di scala  $r$ . Questa relazione implica che lo spazio metrico, modificato

per un fattore di scala, rimarrà geometricamente invariato rispetto allo spazio metrico iniziale. Dato uno spazio metrico  $(X, d)$ , un punto  $x_0 \in X$  ed un numero  $\varepsilon \in \mathbb{C}_0^{++}$ , si definisce *disco aperto* (oppure intorno fondamentale di base o circolare) di raggio  $\varepsilon$  e centro  $x_0$  il seguente insieme :

$f_{x_0}^\varepsilon = \{x \in X \ni d(x_0, x) < \varepsilon\}$ . A seconda che si lavori in una, due o tre dimensioni l'intorno equivarrà rispettivamente ad un intervallo senza gli estremi, ad un cerchio senza la circonferenza, ed in fine, ad una sfera senza la sua superficie.

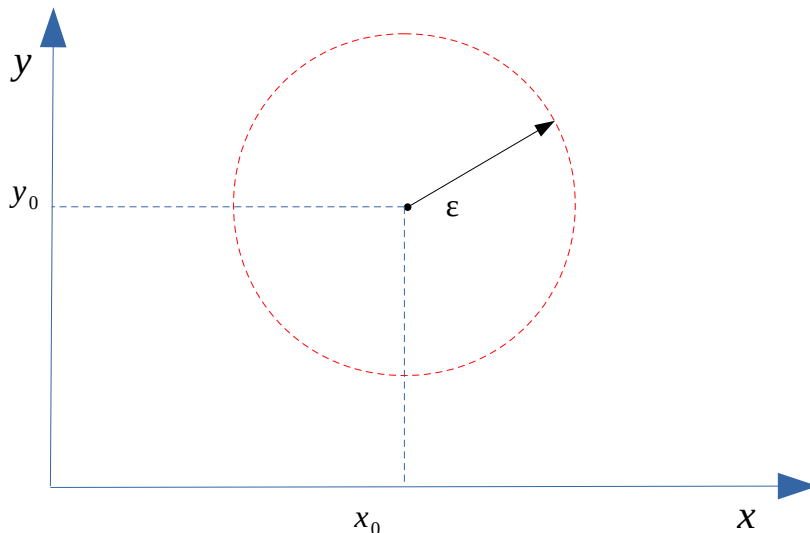


Figura 2 : Esempio di intorno circolare in  $\mathbb{C}_0^2$

Grazie alla nozione di distanza e di intorno fondamentale circolare, si può introdurre il concetto di *funzione continua* nel seguente modo;

Dati due spazi metrici  $(X, d_x)$  e  $(Y, d_y)$  e dato  $f: (X, d_x) \rightarrow (Y, d_y)$  dove  $d_x = d_y$ , il funzionale  $f$  sarà continuo in  $x_0 \in X$  se  $\forall \varepsilon \in \mathbb{C}_0^{++}$

$\exists \delta(\varepsilon) \in \mathbb{C}_0^{++} \ni \forall x \in f_{x_0}^{\delta(\varepsilon)} | f(x) \in f_{f(x_0)}^\varepsilon$  oppure equivalentemente si può scrivere che  $f$  è continua in  $x_0 \in X \Leftrightarrow \forall \varepsilon \in \mathbb{C}_0^{++} \exists \delta(\varepsilon) \in \mathbb{C}_0^{++} \ni f(f_{x_0}^{\delta(\varepsilon)}) \subset f_{f(x_0)}^\varepsilon$ .

Data la definizione di *disco aperto*, si può adesso introdurre quella di *sottoinsieme aperto* di  $(X, d)$ . Dato un insieme  $A \subset X$ , si dice che  $A$  è un *aperto* di  $X$  se  $\forall x \in A \exists f_x^\varepsilon \ni f_x^\varepsilon \subset A$ . Pertanto, possiamo concludere dicendo che un aperto

$A$  è formato dall'unione di infiniti dischi aperti. In simboli possiamo scrivere che,

$$A \text{ aperto} \Leftrightarrow \begin{cases} A = \cup f_x^\varepsilon \\ \text{oppure, } \emptyset \end{cases}$$

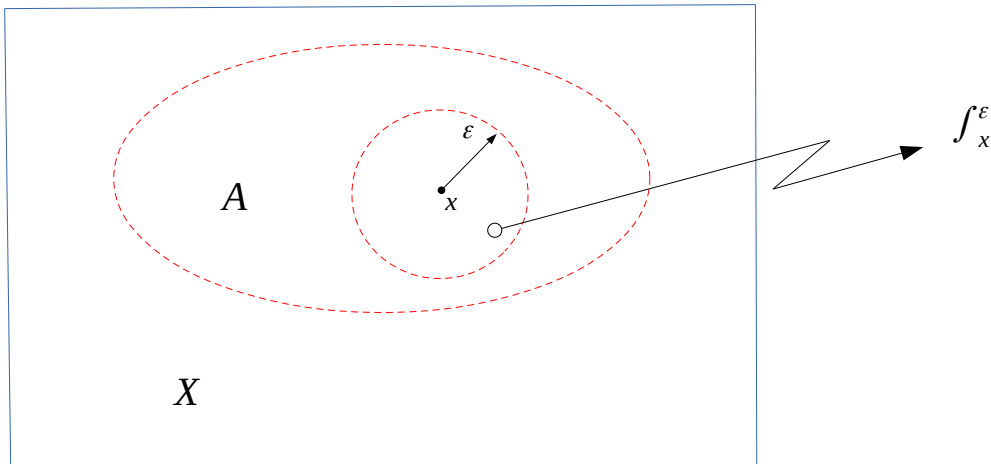


Figura 3: esempio grafico di sottoinsieme aperto  $A$  di  $X$ .

Data la definizione di insieme aperto, si potrà esprimere la proprietà di continuità di una funzione nel seguente modo;

sia una funzione (o funzionale) su due spazi metrici  $f: (X, d_x) \rightarrow (Y, d_y)$ , allora essa sarà continua  $\Leftrightarrow \forall A \subset Y$  aperto risulterà che  $f^{-1}(A) \subset X$  con  $f^{-1}(A)$  aperto.

Ovvero  $f^{-1}(A)$  sarà aperto  $\Leftrightarrow \forall x \in f^{-1}(A) \exists \varepsilon \in \mathbb{C}_0^{++} \ni \int_x^\varepsilon \in f^{-1}(A) \subset X$  dunque, le contro-immagini degli aperti di  $Y$  sono aperti di  $f^{-1}(A) \subset X$ .

Si può affermare che la proprietà di continuità di una funzione dipenderà solo dalla famiglia degli aperti. Alla nozione di famiglia degli aperti è associata la nozione di *funzione aperta*; ovvero una  $f: X \rightarrow Y$  è detta *funzione aperta*  $\Leftrightarrow$  l'immagine di ogni aperto di  $X$  è aperto in  $Y$ . Non sussiste alcuna relazione tra funzione aperta e funzione continua poiché si potranno avere funzioni aperte, ma che non saranno continue, e funzioni continue che risulteranno aperte o meno.

Sia il seguente esempio sui reali con la topologia Euclidea: data la funzione

$$f: \mathbb{C}_0 \rightarrow \mathbb{C}_0^+ \text{ dove } X = \mathbb{C}_0 \text{ aperto e } Y = \mathbb{C}_0^+ = [0, +\infty) \text{ non aperto, essa, pur}$$

$\int_x \quad \int_{x^2}$

risultando una funzione continua, non risulta aperta.

Se esisteranno due spazi metrici  $(X, d) \neq (X, d_1)$  definiti sulla stesso insieme, con metriche diverse (dunque due spazi metrici diversi), ma che daranno luogo alla stessa famiglia degli aperti, allora le funzioni continue nei due spazi coincideranno. Spazi metrici con tale caratteristica sono detti *topologicamente equivalenti*. Una nozione fondamentale nella teoria delle catastrofi è quella di *spazio topologico*.

Definizione di *spazio topologico*:

Data la coppia  $(X, a)$  con  $a \subseteq \wp(X)$  e  $\wp(X)$  insieme delle parti dell'insieme  $X$ , essa formerà uno *spazio topologico* se per  $a$  varranno i seguenti assiomi:

- 1)  $\emptyset, X \in \mathcal{a}$  tutti gli insiemi che si considereranno sono degli aperti, dunque  $\emptyset, X$  sono aperti.
- 2) Dati  $A, B \in \mathcal{a} \Rightarrow A \cap B \in \mathcal{a}$ , più in generale  $\bigcap_{i \in I} A_i \in \mathcal{a}$  dove  $I$  rappresenta una qualunque famiglia degli aperti con  $\#(I) < +\infty$ .
- 3)  $\bigcup_{i \in I} A_i \in \mathcal{a}$  valido  $\forall I$  ovvero valido per qualunque famiglia di aperti con cardinalità finita o anche infinita.

$\mathcal{a}$  è detto *famiglia degli aperti o di topologia su l'insieme X*.

Se considero due particolari insiemi  $\mathcal{a}$  si otterranno due particolari spazi topologici, e precisamente;

- i) se  $\mathcal{a} = \{\emptyset, X\}$  allora  $(X, \mathcal{a})$  è detta *topologia discreta*.
- ii) se considero  $\mathcal{a} = \{\emptyset, X\}$  allora  $(X, \mathcal{a})$  è detta *topologia banale*. Dalla definizione di intorno circolare si evince che esso può essere identificato con un insieme aperto. Tuttavia, un intorno non è necessariamente un insieme aperto: se nella definizione si includerà anche la sua frontiera, si ottiene un *intorno chiuso*.

La definizione di *frontiera* richiede preliminarmente l'introduzione dei concetti di punto interno e punto esterno di un insieme generico.

Dato uno spazio metrico  $(X, d)$  e dato un insieme  $A \subset X$ , preso un generico punto  $x_0 \in A$  si dice che,

- 1)  $x_0$  è punto interno di  $A \Leftrightarrow \exists \varepsilon \in \mathbb{C}_0^{++} \ni \int_{x_0}^\varepsilon \subset A$ . Dunque, dato un certo raggio, ogni intorno circolare di un punto generico di  $A$  sarà tutto contenuto in  $A$ . L'insieme di tutti i punti interni di  $A$  è detto *Interno di A* e lo si indica con  $\overset{\circ}{A}$ .

- 2)  $x_0$  è punto esterno di  $A \Leftrightarrow x_0$  è interno al complementare di  $A$  ovvero  $\Leftrightarrow \exists \varepsilon \in \mathbb{C}_0^{++} \ni \int_{x_0}^\varepsilon \subset A^c \Leftrightarrow \exists \varepsilon \in \mathbb{C}_0^{++} \ni \int_{x_0}^\varepsilon \cap A = \emptyset$

dove il complementare di un insieme generico  $A \subset X$  è dato dall'insieme

$$A^c = X \setminus A = \{x \in X \ni x \notin A\}.$$

L'insieme di tutti i punti esterni di  $A$  è detto *Esterno di A* e lo si indica con  $\overset{\circ}{A}^c$ . Allora  $x_0$  è detto *punto di frontiera di un insieme A* (valevole anche per gli interni)

$\Leftrightarrow x_0$  non è ne interno ne esterno ad  $A$ , oppure equivalentemente

$$x_0 \text{ è punto di frontiera per } A \Leftrightarrow \forall \varepsilon \in \mathbb{C}_0^{++} \exists \int_{x_0}^\varepsilon \ni \begin{cases} \int_{x_0}^\varepsilon \cap A \neq \emptyset \\ \int_{x_0}^\varepsilon \cap A^c \neq \emptyset \end{cases}$$

Tutti i punti di frontiera di  $A$  formano la *frontiera di A* che si indica con  $Fr(A)$ . Si può affermare che, dato un insieme aperto, il suo complementare sarà un chiuso e ad esso apparterrà anche la sua frontiera.

Viceversa, dato un chiuso il suo complementare sarà un aperto. E' importante specificare che, se un insieme non è aperto non implica che esso sia chiuso, ed equivalentemente se non è chiuso che esso risulti un aperto. Infatti, esistono insiemi che non sono né aperti e né chiusi. Un aperto è tale se in suo complementare è chiuso e viceversa. Come esempio si consideri un insieme dato da un intervallo (dunque l'ambiente lavoro è ad una dimensione) del tipo  $[a, +\infty[$  col suo complementare  $[a, +\infty[^c = ]-\infty, a[$  che è un aperto; allora malgrado  $[a, +\infty[$  sia chiuso solo inferiormente, essendo però il complementare di un aperto, allora esso sarà un chiuso. Dato uno spazio topologico  $(X, \mathcal{a})$ , e dato  $A$  aperto, allora si definiscono i chiusi come insiemi complementari degli aperti, ovvero  $\mathcal{c} = \{X \setminus A \mid A \in \mathcal{a}\}$ . Sfruttando le leggi di De-Morgan, e data la dualità tra insieme aperto e suo complementare, si ha che anche i chiusi rispettano gli assiomi di uno spazio topologico. Infatti, partendo dalla definizione di spazio topologico vista prima si ha che le sue proprietà valgono anche per i chiusi poiché,

1)  $X, \emptyset \in \mathcal{a}$  resta invariata poiché tali insiemi non sono inficiati dalla complementarità degli aperti.

2) Dati  $A, B \in \mathcal{a} \Rightarrow A \cap B \in \mathcal{a}$  sfruttando le leggi di De-Morgan si ha che  $A \cap B = A^c \cup B^c \in \mathcal{a}$  e dato che  $A$  e  $B$  sono aperti, i loro complementari saranno dei chiusi.

3)  $\bigcup_{i \in I} A_i \in \mathcal{a}$  valido  $\forall I$  ma da De-Morgan si ha che  $A \cup B = A^c \cap B^c \in \mathcal{a}$  ed essendo valido per una qualunque famiglia di aperti  $I$  allora si può scrivere - sfruttando De-Morgan- che  $\bigcup_{i \in I} A_i = \bigcap_{i \in I} A_i^c \in \mathcal{a}$ .

Dunque, per definire una topologia è indifferente partire dalla famiglia degli aperti o da quella dei chiusi per la dualità esistente tra le due famiglie di insiemi.

Sugli insiemi aperti e chiusi valgono queste regole duali,

1) Dati  $A_i$  chiusi con  $i \in \mathbb{N}$ , allora è falso affermare che  $\bigcup_i A_i$  è chiuso.

2) Dati  $A_i$  aperti con  $i \in \mathbb{N}$ , allora è falso affermare che  $\bigcap_i A_i$  è aperto.

Dire che la 1) e la 2) sono false equivale ad affermare che in generale esse non sono vere, ovvero possono essere o meno verificate. E' facile provare la 1) e la 2) con dei controesempi considerando, ad esempio, l'insieme dei reali con la topologia standard.

Per la 1) si consideri il seguente insieme,  $\{\frac{1}{n}\}_{n \in \mathbb{N}_0}$ . Esso è composto da singoli valori i quali saranno tutti chiusi, poiché i loro complementari saranno degli intervalli aperti. Ma se ne ho in numero infinito, ovvero se  $n \rightarrow +\infty$ , poiché  $\frac{1}{n} \rightarrow 0$  e

poiché  $0 \notin \left\{ \frac{1}{n} \right\}_{n \in \mathbb{N}_0}$  allora la unione di infiniti chiusi non sarà un chiuso poiché zero non appartiene alla successione, ma per avere un chiuso esso dovrebbe rientrare nell'insieme, dunque in questo caso la unione di infiniti chiusi non è un chiuso. Per la

2) si consideri il seguente insieme,  $\bigcap_n ]0, 1 + \frac{1}{n}[$  con  $n \in \mathbb{N}_0$  intersezione di aperti che non è aperto se ne considero un numero infinito poiché,

$$\bigcap_n ]0, 1 + \frac{1}{n}[ = ]0, 1 + \frac{1}{1}[ \cap ]0, 1 + \frac{1}{2}[ \cap ]0, 1 + \frac{1}{3}[ \dots \cap ]0, 1 + \frac{1}{n-1}[ \cap ]0, 1 + \frac{1}{n}[$$

con  $n \rightarrow +\infty$  ho che  $]0, 1 + \frac{1}{n}[ = ]0, 1[$  e dunque la catena di intersezioni

diventa,  $\bigcap_n ]0, 1 + \frac{1}{n}[ = ]0, 2[ \cap ]0, \frac{3}{2}[ \cap ]0, \frac{4}{3}[ \dots \cap ]0, 1 + 0[ = ]0, 1[$  che

non è un aperto. Altri concetti topologici utili sono quelli di *Punto aderente di un insieme* da cui discende la *Chiusura* di un insieme e di *Punto di accumulazione* di un insieme.

Dato un insieme  $A$  e dato uno spazio topologico, un generico punto  $x_0 \in A$  è detto *punto aderente di  $A$*   $\Leftrightarrow \forall \varepsilon \in \mathbb{C}_0^{++} \exists ]x_0, x_0 + \varepsilon[ \cap A \neq \emptyset$  ovvero il punto non è esterno ad  $A$ . L'insieme di tutti i punti aderenti di un insieme è detto *Chiusura di  $A$*  e lo si indica con  $\bar{A}$ . Dalla definizione se ne deduce che tutti i punti interni di un insieme sono anche punti di aderenza per il medesimo insieme.

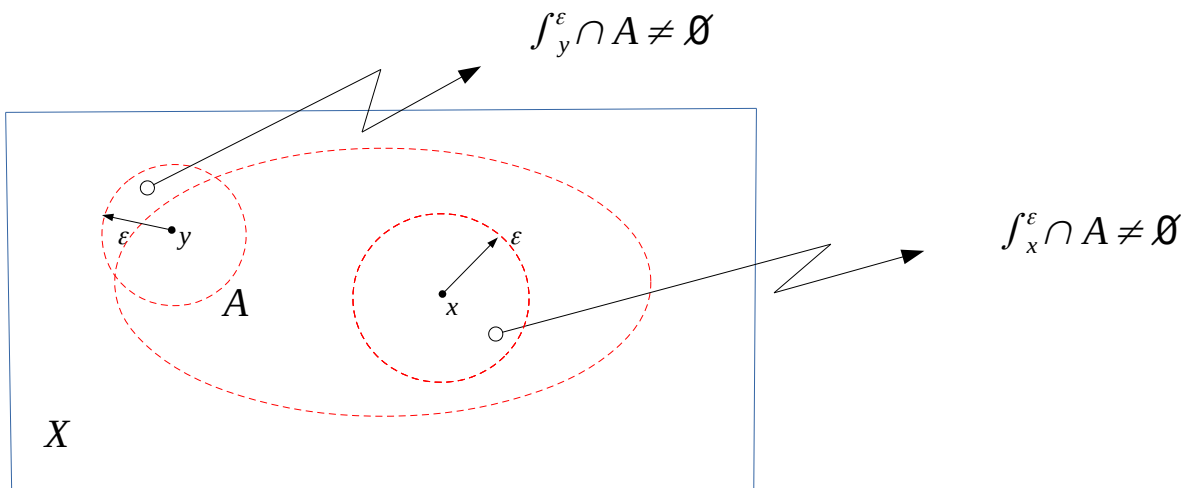


Figura 4:  $x$  ed  $y$  sono di aderenza per  $A$  poiché ne rispettano la definizione. Sono anche punti interni per  $A$ .

Sfruttando la definizione di punto di aderenza si introduce quella di *punto di accumulazione* per un insieme generico  $A$ . Si dice che  $x_0$  è *punto di accumulazione* per  $A$  se esso è punto di aderenza per l'insieme  $A \setminus \{x_0\}$  ovvero

$\Leftrightarrow \forall \varepsilon \in \mathbb{C}_0^{++} \exists f_{x_0}^\varepsilon \ni f_{x_0}^\varepsilon \cap (A \setminus \{x_0\}) \neq \emptyset$  equivalentemente si può scrivere  
 $\Leftrightarrow (f_{x_0}^\varepsilon \setminus \{x_0\}) \cap A \neq \emptyset$  dove  $f_{x_0}^\varepsilon \setminus \{x_0\}$  è detto *intorno bucato*.

Dunque in ogni intorno bucato si troveranno punti di  $A$  distinti dal punto  $x_0$ . L'insieme di tutti i *punti di accumulazione* per un insieme generico  $A$  è detto *Derivato di  $A$*  e lo si indica con  $D(A)$ .

Nella teoria delle catastrofi, come già detto, ogni fenomeno è visto come una forma e dunque, equiparare le forme ad una struttura topologica permette di studiare quelle sfruttando le proprietà topologiche di uno spazio.

Si potrebbe parlare, ad esempio, di due forme uguali considerando i casi in cui, due spazi topologici siano uguali. In tal caso si introduce il concetto matematico che permette di confrontare due spazi topologici e definire se essi sono uguali o meno; si tratta del concetto di *omeomorfismo*.

Dati due spazi topologici sugli insiemi  $X, Y$  diversi dall'insieme vuoto, si definisce *omeomorfismo* la funzione  $f: X \rightarrow Y$  se  $\exists g: Y \rightarrow X \ni X \xrightarrow{f} Y \xrightarrow{g} X \wedge Y \xrightarrow{g} X \xrightarrow{f} Y$  in modo tale che coincidano con la funzione identità, ovvero se  $X \xrightarrow{f} Y \xrightarrow{g} X$  con  $g(y) = f^{-1}(y)$  ed  $Y \xrightarrow{g} X \xrightarrow{f} Y$  con  $f(x) = g^{-1}(x)$ . In

altre parole deve risultare che  $g \circ f = I_x$  ed  $f \circ g = I_y$  dove la funzione  $g$  è la inversa della funzione  $f$  ed il simbolo "o" rappresenta la composizione di funzioni. In topologia le funzioni considerate sono solo quelle continue. Da quanto scritto sopra si può concludere che una funzione  $f$  è un omeomorfismo  $\Leftrightarrow$  essa è continua, biiettiva (ovvero invertibile) e bicontinua (la sua inversa è continua)

$\Leftrightarrow$  essa è continua, biiettiva ed aperta. Per definire se due spazi topologici  $X$  ed  $Y$  sono omeomorfi, assumono importanza le proprietà topologiche di essi, dove con proprietà topologiche si intendono *proprietà invarianti per omeomorfismo*, ovvero proprietà che, se sono possedute da uno spazio, allora le possiederà anche un'altro spazio ad esso omeomorfo.

Se  $X$  possiede una proprietà topologica ed  $Y \cong X \Rightarrow Y$  ha le stesse proprietà topologiche di  $X$ , ovvero  $X$  ed  $Y$  sono omeomorfi.

Un esempio può essere il seguente: siano due spazi topologici  $X, Y$  con topologia standard. Si consideri come proprietà topologica la *connessione*. Un insieme generico  $X$  si dice *connesso per archi* se risulterà che,

$$\forall (x_0, y_0) \in X \exists (g, \Gamma) \ni \left\{ \begin{array}{l} g \in \mathcal{C}([a, b], X) \\ g(a) = x_0 \\ g(b) = y_0 \\ g([a, b]) = \Gamma \in X \end{array} \right. \text{ con } (g, \Gamma) \text{ arco di curva in } X.$$

Questo vuol dire che un insieme sarà connesso se potrò sempre prendere due suoi punti arbitrari ed unirli con un arco di curva, senza che i punti della curva escano dall'insieme. Preso un'intervallo ed un cerchio essi non saranno omeomorfi poiché, se tolgo un punto generico interno all'intervallo esso non sarà più connesso, ma se ne

tolgo la sua immagine dal cerchio esso resterà connesso, dunque a seguito di una variazione nei due spazi, la proprietà di connessione iniziale posseduta da entrambi viene persa da uno dei due; ne consegue che i due spazi non saranno omeomorfi.

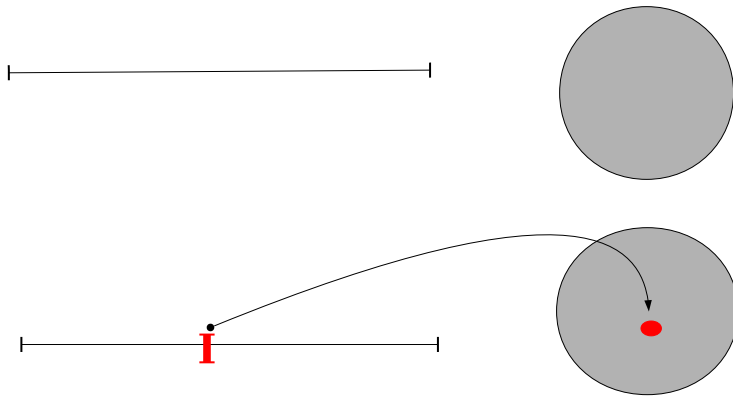


Figura 5: esempio di spazi non omeomorfi.

Nell'ambiente  $\mathbb{C}_0$  tutti gli intervalli aperti saranno tra loro omeomorfi poiché esisterà una funzione continua tale che, ogni punto di un intervallo potrà essere proiettato sul secondo intervallo.

Un esempio grafico può chiarire quanto detto:

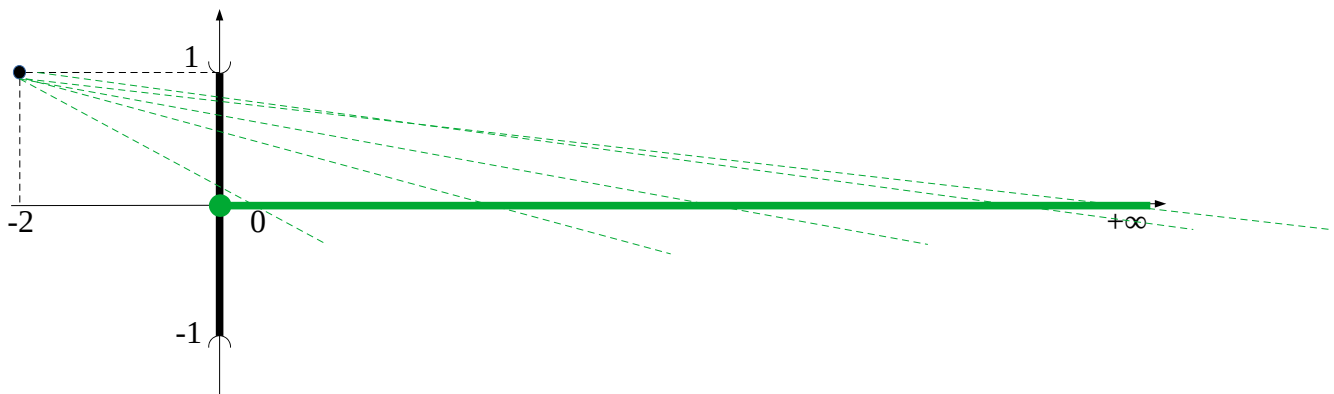


Figura 6: omeomorfismo tra intervalli;  $]0,1[ \cong [0,+\infty)$

Dalla figura 6 si può osservare che, preso un generico intervallo in ordinata -ad esempio  $] -1,1[$  - e considerando un qualunque punto del piano con ordinata compresa tra -1 ed 1, come ad esempio il punto  $(-2,1)$ , proiettando tale punto sull'intervallo in ascisse  $[0,+\infty)$  si avrà che tale proiezione unirà sempre un punto

dell'intervallo in ordinata  $]0,1[$  con un punto dell'intervallo  $[0, +\infty)$  per omeomorfismo. Si possono anche considerare i due intervalli omeomorfi poiché, se a seguito di una esclusione di un punto dell'intervallo in ordinata si perde la proprietà di connessione di esso, proiettando la immagine del punto eliminato sull'intervallo in ascisse, ed eliminandone la proiezione sull'intervallo in ascisse, anche quest'ultimo perderà la proprietà di connessione e dunque, a seguito di una azione sul primo intervallo, il mutamento delle proprietà di questo si otterrà anche sul secondo intervallo. Questo risultato implica che tutti gli intervalli della retta reale sono tra loro omeomorfi. L'idea di rappresentare le forme usando concetti matematici, permette di vedere alcune di esse come uguali o approssimativamente uguali, sfruttando appunto i principii matematici con cui esse vengono descritte e dunque, permette conseguentemente, di studiarne le caratteristiche.

Un esempio significativo è la rappresentazione della *proprietà di stabilità strutturale*, considerata fondamentale affinché una forma possa essere, secondo la teoria, definita come tale. Data una forma  $A$ , essa sarà considerata strutturalmente stabile se esisteranno altre forme  $B$ , sufficientemente vicine ad  $A$  in uno spazio di esistenza  $E$  – rappresentante uno spazio topologico – tali che siano equivalenti ad  $A$  secondo una relazione  $R$ . Affinché una forma sia considerata *strutturalmente stabile*, è sufficiente che la *classe di equivalenza* cui essa appartiene sia un *insieme aperto* nello spazio di esistenza  $E$ . In tal caso, ogni forma sufficientemente vicina ad essa (cioè appartenente allo stesso aperto) risulterà equivalente secondo la relazione  $R$ , e dunque l'intera classe di equivalenza sarà detta *strutturalmente stabile*.

Secondo la teoria, tutte le forme di cui si ha soggettiva contezza e alle quali è stata attribuita una denominazione sono considerate strutturalmente stabili. Ciò accade perché tali forme, immerse in un ambiente che agisce costantemente su di esse con perturbazioni, anche minime, riescono comunque a mantenere inalterata la propria identità. Il fatto che esse non subiscano modificazioni sostanziali a seguito di queste perturbazioni implica che restano all'interno della loro classe di equivalenza, così come definita in precedenza. In altri termini, la loro capacità di resistere ai cambiamenti indotti dall'ambiente conferma la loro stabilità strutturale. In termini topologici, l'insieme delle forme strutturalmente stabili costituisce un aperto nello spazio di esistenza  $E$ . Il complementare di questo aperto nello spazio di esistenza  $E$  corrisponde all'insieme delle *forme instabili*, le quali, anche in presenza di minime perturbazioni, possono subire modificazioni tali da non essere più riconoscibili come appartenenti alla stessa classe di equivalenza. Tali forme non possono essere considerate forme vere e proprie, poiché non possiedono una struttura sufficientemente resistente alle variazioni: per questo motivo, sono definite *informi*.

In topologia, dato un insieme  $A$  e dati  $x, y, z$  suoi generici elementi, si definisce relazione di equivalenza in  $A$  la relazione  $R$  in modo tale che  $\forall x, y, z \in A$  valgano le seguenti tre proprietà:

- i) riflessiva  $\rightarrow x R x$
- ii) simmetrica  $\rightarrow x R y \Rightarrow y R x$
- iii) transitiva  $\rightarrow (x R y) \wedge (y R z) \Rightarrow x R z$

Dunque, quando una relazione  $R$  è di equivalenza per un insieme  $A$ ,  $\forall x, y \in A$  (con  $x, y \in A$ ) si usa denotarla con la scrittura  $x \equiv y \pmod{R}$  che si interpreta come “ $x$  è equivalente ad  $y$  modulo  $R$ ” (o secondo la relazione  $R$ ).

Tale relazione  $R$  genera un sottoinsieme del prodotto cartesiano  $A \times A$  formato da coppie ordinate di elementi di  $A$  che saranno equivalenti secondo la relazione  $R$ .

Dalla relazione di equivalenza scaturisce un altro concetto, quello di classe di equivalenza.

Dato un insieme  $A$  e data una relazione di equivalenza  $R$  tra i suoi elementi, preso un elemento  $a \in A$  allora il sottoinsieme di  $A$  di tutti gli elementi in relazione  $R$  con  $a$  formerà la classe di equivalenza di  $a$  modulo  $R$  e la si indicherà con  $c_a$  oppure  $\dot{a}$ .

Il suo grafo sarà l'insieme  $G_{c_a} = G_a = \{x \in A \ni x R a\}$  essendo stato preso  $a$  come elemento rappresentativo della classe. Ritornando sul concetto di spazio topologico, ad esso si può associare quello di *topologia indotta* sullo spazio iniziale. Si tratta della topologia applicata ad un sottoinsieme dell'insieme della topologia di partenza. Data una topologia  $(X, \mathcal{a})$  e dato  $S \subseteq X$  si definisce una topologia su  $S$ , sottoinsieme di  $X$  (insieme ambiente).

La operazione di inclusione canonica tra spazi topologici è indicata con la scrittura  $S \hookrightarrow X$ . Si potrà allora definire una applicazione  $i$  da  $S$  in  $X$ , ovvero  $S \xrightarrow{i} X$  con la condizione che  $i$  risulti continua.

Dato  $A \subset X$  allora si avrà che  $i^{-1}(A) = i^{-1}(i(S) \cap A)$  con  $i(S) \cap A$  detta *Traccia*. La controimmagine di  $A$  mediante  $i$  sarà pari alla controimmagine della traccia mediante  $i$ . Si tratta degli aperti di  $S$ , appunto i punti che sono controimmagine della traccia e saranno solo quei punti tali che la operazione insiemistica di inclusione sia continua.

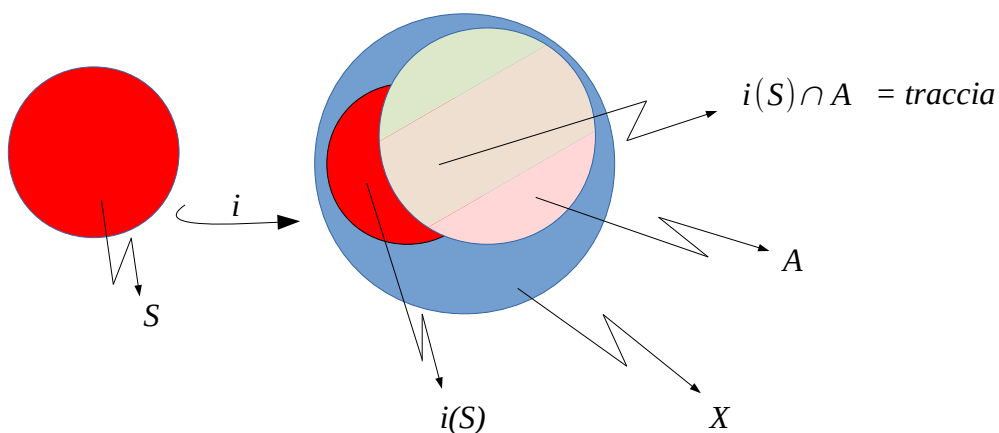


Figura 7: topologia indotta

Dunque la topologia di  $S$  sarà data dall'insieme  $\mathcal{a}_S = \{A \cap i(S), A \in \mathcal{a}\}$  formato da tutti gli insiemi aperti che servono affinché la funzione  $i$  sia continua.

Equivalentemente si può dire che gli aperti della topologia indotta sono le tracce degli aperti dello spazio  $X$ . Si consideri come esempio l'insieme ambiente dato da  $\mathbb{C}_0^2$  e lo spazio indotto dato dall'insieme  $S = \{(x, y) \in \mathbb{C}_0^2 \mid x^2 + y^2 = 1\}$  ovvero una circonferenza di raggio 1 e centrata nell'origine degli assi Cartesiani. Sia poi  $A$ , ad esempio, un generico cerchio aperto che interseca la circonferenza in un luogo qualunque del piano. I punti che formeranno tale intersezione rappresenteranno la traccia degli aperti del piano. Tali intersezioni però potranno anche essere vuote poiché il cerchio aperto potrebbe non intersecare l'insieme  $S$ .

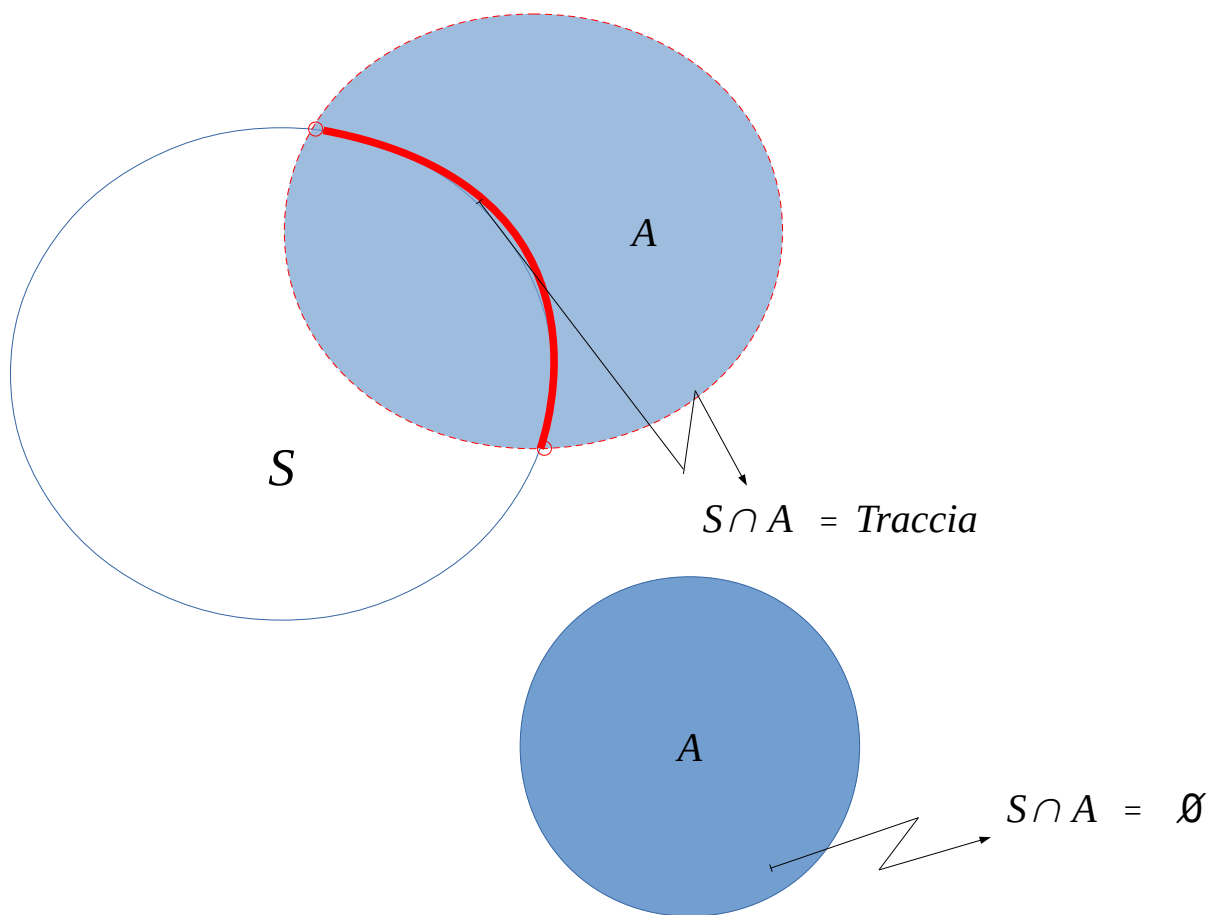


Figura 8: esempio di topologia indotta per spazi nel piano.

Dunque in  $\mathbb{C}_0^2$  le intersezioni tra gli aperti del piano ed  $S$  saranno dati da

$$A \cap S = \emptyset \text{ oppure } A \cap S = \text{Arco di circonferenza aperto.}$$

In questo ambiente, gli aperti nella topologia indotta su  $S$  saranno unione di archi aperti. Nella teoria delle catastrofi, gli aperti rappresentano la regione dello spazio di esistenza in cui una forma è strutturalmente stabile, e dunque può essere considerata propriamente come tale. Il complementare di questi aperti saranno degli insiemi chiusi, i cui elementi vengono considerati *punti di catastrofe*, cioè punti in cui una forma cambia aspetto in modo discontinuo o improvviso. In essi, la struttura figurata si rompe, e la forma cessa di appartenere alla classe di equivalenza stabile: si verifica così una *transizione morfologica*.

Un esempio concreto può essere edificante per la comprensione del significato che i principî della teoria delle catastrofi possono assumere nella spiegazione di fenomeni reali. Un caso emblematico fu proposto da E. C. Zeeman, matematico inglese, il quale contribuì allo sviluppo della teoria insieme a Renè Thom, che ne fu l'ideatore. L'esempio originale aveva delle variazioni rispetto a quello che sarà riportato, ma in questa versione il significato di esso rende ugualmente bene il senso concettuale che si vuol esplicitare. Si considererà il caso della catastrofe detta *grinza*, una tra quelle più rappresentative delle forme esistenti ed idonee a spiegare le stesse.

Sia il caso in cui il fenomeno da studiare rappresenti un sistema in cui vi siano dei parametri dai quali dipenderà un dato risultato finale. Tale caso sia rappresentato dalla capacità creativa di uno studioso dipendente da parametri quali  $C$ =competenza tecnica,  $E$ =entusiasmo ed  $S$ =successo dipendente dai primi due. Il modo in cui, al variare dei parametri  $C$  ed  $E$ , varierà il parametro  $S$  può essere rappresentato dalla catastrofe della *grinza*. Ci si trova in un ambiente tridimensionale con i parametri  $C$  ed  $E$  appartenenti al piano ed  $S$  rappresentante la loro immagine:

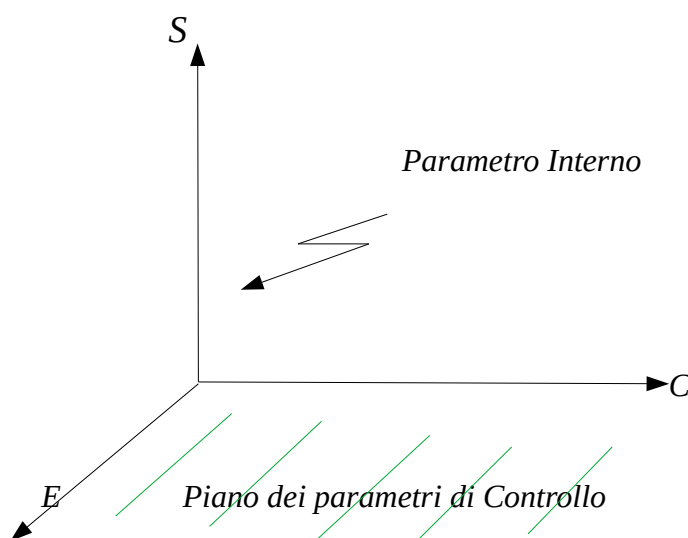


Figura 9: spazio tridimensionale.

L'insieme dei punti del piano dei parametri  $(C, E)$  genereranno dunque un valore  $S$ . Attraverso la grinza si possono trovare in punti in cui il parametro  $S$  ha una evoluzione catastrofica nel senso che, varierà da un livello ad un altro – al variare dei parametri  $E$  ed  $C$  – in maniera brusca e discontinua subendo una sorta di *cut-off*, sia positivo che negativo. I parametri  $C$  ed  $E$  sono detti di *controllo* del sistema, invece il parametro  $S$  è detto *interno*. Di seguito il grafico approssimante una *grinza* sull'esempio riportato;

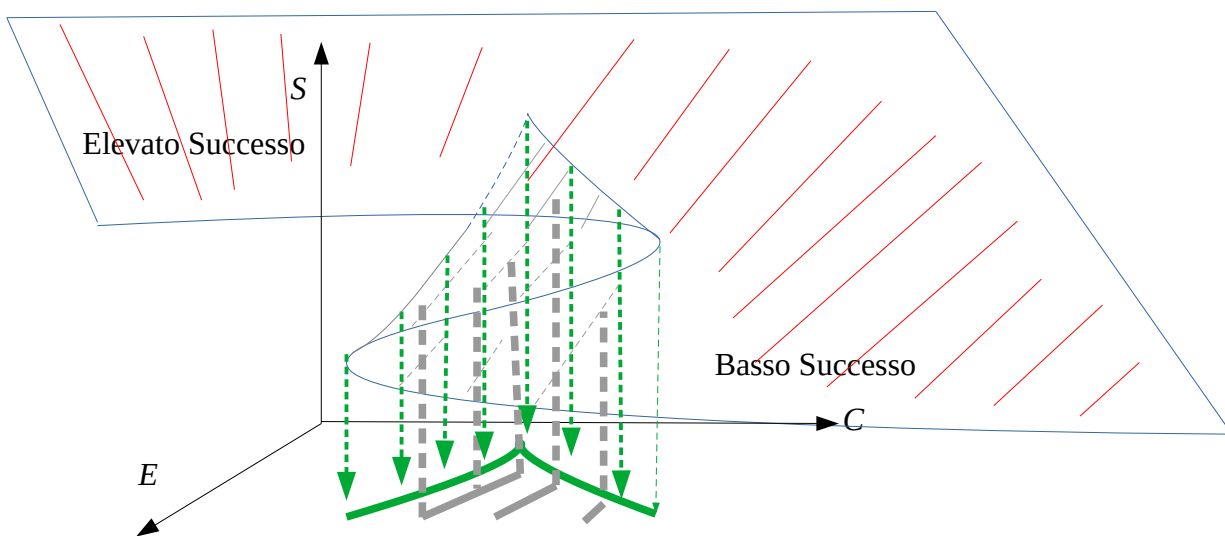


Figura 10: catastrofe della *Grinza*.

Tutta la porzione di piano  $(C, E)$  rappresentata in grigio ed interno alle due curve verdi -che formano una cuspide-, sono i punti a cui corrispondono valori in  $S$  di non accettazione per il fenomeno. Infatti, a ogni punto interno a questa regione corrispondono più valori possibili del parametro  $S$  (successo), il che implica che la relazione tra i parametri  $(C, E)$  ed il risultato  $S$  non è più una funzione univoca, bensì una multifunzione o funzione multivoca. In termini topologici, ciò significa che non è più possibile definire un omeomorfismo tra lo spazio dei parametri e quello dei risultati: manca la corrispondenza biunivoca e continua richiesta. In questa regione, infatti, la superficie del modello della catastrofe si piega su sé stessa, e ad un singolo punto del piano dei parametri corrispondono più punti sulla superficie di risposta, violando così i requisiti per l'invertibilità continua della funzione.

Per questo motivo, la zona grigia è detta *zona di non accettazione*: rappresenta un dominio in cui il sistema perde stabilità e non è più possibile prevedere univocamente il risultato finale.

Invece, i punti del piano rappresentati sulle curve verdi, che delimitano la cuspide, sono detti *punti critici*: a essi corrisponde un brusco cambiamento nello stato del sistema. Nel nostro esempio, questi punti rappresentano il passaggio da uno stato di successo a uno di insuccesso (o viceversa) per lo studioso, anche in presenza di variazioni minime nei parametri  $C$  (competenza) ed  $E$  (entusiasmo). Si tratta dei punti di catastrofe nel senso proprio della teoria: lì il sistema attraversa un cambiamento qualitativo improvviso ed irreversibile. La combinazione tra competenza ed entusiasmo da parte dello studioso incide significativamente sul livello di successo raggiunto. In presenza di un elevato entusiasmo, un incremento delle competenze può generare un salto qualitativo nel livello di successo  $S$ , collocandolo nell'area del dominio ad alto successo, come illustrato in Figura 9. In questo contesto, si osservano variazioni qualitative del comportamento del sistema.

Tuttavia, un aumento dell'entusiasmo non accompagnato da un corrispondente incremento della competenza tecnica può condurre ad una catastrofe: il livello di successo  $S$  subisce un brusco calo, posizionandosi nell'area del basso successo, sempre in riferimento alla Figura 9.

Risultano immediatamente evidenti le caratteristiche dei salti qualitativi del fenomeno, in particolare il fatto che essi si verifichino lungo linee differenti. Le linee di livello, che rappresentano le proiezioni dei valori della variabile  $S$  sulla superficie in cui avvengono i passaggi da Elevato Successo a Basso Successo (e viceversa), non coincidono tra loro, evidenziando così la dipendenza direzionale del comportamento del sistema nei pressi della catastrofe. Questo implica che minime variazioni nei valori delle variabili di controllo  $C$  ed  $E$ , anche se estremamente vicine a quelli precedenti, possano determinare esiti radicalmente diversi in termini di successo. Uno studioso potrebbe infatti passare da un livello di successo elevato ad uno basso (o viceversa), pur in presenza di modifiche apparentemente trascurabili nei parametri. Ciò mette in evidenza come le linee di transizione tra gli stati non coincidano, ma dipendano dalla direzione del cambiamento, ovvero se il sistema sta evolvendo verso un aumento o una diminuzione del successo. Questo comportamento è coerente con i fenomeni di *isteresi*<sup>1</sup> tipici della catastrofe della cuspide. L'insieme di catastrofe costituisce un insieme chiuso, mentre il suo complementare -considerando esclusivamente la regione di accettazione- rappresenta un insieme aperto nel quale il fenomeno si colloca in una condizione di equilibrio, associata a valori elevati o bassi della variabile interna  $S$ . All'interno di questa superficie di equilibrio, è possibile muoversi continuamente da un punto all'altro senza che si verifichino eventi catastrofici, ovvero transizioni discontinue nel valore del parametro interno  $S$ . Lo schema descritto è rappresentativo della maggior parte delle situazioni in cui il fenomeno può essere studiato qualitativamente attraverso gli strumenti offerti dalla teoria delle catastrofi.

---

1 *L'isteresi* è un fenomeno caratterizzante i sistemi non lineari con più stati di equilibrio, in cui la risposta del sistema non è univocamente determinata dai valori istantanei dei parametri di controllo, ma dipende dal percorso seguito nello spazio dei parametri (la storia passata dei parametri di controllo). Nella cuspide, che si genera con la catastrofe della grinzina, questo si traduce in una biforcazione con regioni multi stabili separate da salti lungo i quali si manifestano transizioni discontinue tra stati di stabilità.

Il fenomeno può essere analizzato in funzione di un insieme di parametri di controllo e di variabili interne. Gli stati di equilibrio che il sistema può assumere, al variare di tali parametri, definiscono una superficie (o *varietà*) di equilibrio immersa in uno spazio di dimensione idonea. La proiezione di questa superficie sul piano dei parametri di controllo potrà presentare delle singolarità, ovvero punti in cui il sistema manifesta transizioni qualitative (o *salti*) nello stato, indotte da variazioni continue nei parametri. Tali punti corrispondono alle catastrofi nel senso della teoria di René Thom. L'esempio riportato evidenzia l'evoluzione temporale dello stato di un sistema dinamico, il cui comportamento può essere descritto matematicamente tramite un campo vettoriale definito su uno spazio delle fasi. Ogni punto dello spazio delle fasi rappresenta un possibile stato del sistema, e l'equazione differenziale associata al campo vettoriale ne governa la traiettoria nel tempo. In corrispondenza di alcuni punti dello spazio delle fasi, il vettore associato può annullarsi: tali punti sono detti *punti di equilibrio*, in quanto lo stato del sistema, inizialmente posto in essi, rimane invariato nel tempo.

## 2.2

### Sistemi Dinamici e loro caratteristiche fondamentali.

Per comprendere meglio il comportamento di un sistema dinamico e l'importanza dei punti di equilibrio, è utile considerare un esempio concreto. Un caso classico è il *modello Preda-Predatore*, sviluppato dal chimico Alfred Lotka e dal matematico Vito Volterra, quest'ultimo anche fondatore del CNR. Tale modello descrive l'evoluzione nel tempo di un sistema biologico composto da due specie, predatori e prede, la cui interazione determina l'andamento delle rispettive popolazioni. L'ipotesi fondamentale è che l'ambiente in cui le due specie coesistono abbia un'influenza trascurabile sul sistema, permettendo così di concentrarsi esclusivamente sulle dinamiche interne tra prede e predatori. Sia  $x = x(t)$  il numero delle prede al tempo  $t$ . Si ipotizzi che, qualora tale specie vivesse isolata il suo tasso di crescita assumerebbe un andamento proporzionale al numero totale di individui esistenti; si avrebbe che  $\dot{x} = a \cdot x$ , dunque la popolazione in questione seguirebbe una legge esponenziale poiché  $x(t) = e^{a \cdot t} \Rightarrow \frac{dx(t)}{dt} = e^{a \cdot t} \cdot \frac{d(a \cdot t)}{dt} = a \cdot e^{a \cdot t} = a \cdot x(t)$ .

La presenza dei pesci predatori, però, modificherà la legge di evoluzione del numero delle prede. In tale scenario, si assuma che la dinamica sia descritta dalla seguente equazione:  $\dot{x} = a \cdot x - b \cdot x \cdot y$  dove  $a, b \in \mathbb{C}_0^+$  coefficienti generici ed  $y$  rappresenti la cardinalità iniziale dell'insieme dei predatori.

Il segno “-” davanti al coefficiente  $b$  rappresenta il fatto che i predatori si nutrono delle prede, riducendone conseguentemente il numero.

Sia di contro  $y=y(t)$  il numero di predatori presenti al tempo  $t$ . In assenza di prede, l'evoluzione del numero di predatori potrebbe essere descritta dalla equazione  $\dot{y}=-c \cdot y$  con  $c \in \mathbb{C}_0^+$  coefficiente generico col segno “-” il quale riflette il fatto che, in assenza di prede, la popolazione dei predatori tende a decrescere col tempo poiché, banalmente, in assenza di prede i predatori morirebbero.

Dunque, analogamente al caso delle prede, questa dinamica corrisponde a una legge esponenziale decrescente poiché:

$$y(t)=e^{-c \cdot t} \Rightarrow \frac{dy(t)}{dt}=e^{-c \cdot t} \cdot \frac{d(-c \cdot t)}{dt}=-c \cdot e^{-c \cdot t}=-c \cdot y(t) .$$

Tuttavia, in presenza di prede, la dinamica cambia, e l'evoluzione del numero di predatori potrà essere modellata dalla seguente equazione:

$\dot{y}=-c \cdot y+d \cdot x \cdot y$  dove  $c, d \in \mathbb{C}_0^+$  dove  $d$  rappresenta un ulteriore coefficiente che descrive l'incremento del numero di predatori in funzione dell'interazione con le prede. Il termine  $d \cdot x \cdot y$  rappresenta, invece, il contributo positivo alla crescita della popolazione di predatori dovuto alla attività di predazione. Dunque, la presenza dei predatori influisce sulla evoluzione della popolazione delle prede  $\dot{x}$ , e la presenza delle prede influisce sulla dinamica della popolazione dei predatori  $\dot{y}$ .

Questo reciproco legame tra le due specie può essere descritto dal seguente sistema di equazioni differenziali:

$$\begin{cases} \dot{x}=a \cdot x-b \cdot x \cdot y \\ \dot{y}=-c \cdot y+d \cdot x \cdot y \end{cases} \text{ dove si ricordi che } \dot{x}=\frac{dx(t)}{dt} \text{ ed}$$

$$\dot{y}=\frac{dy(t)}{dt} . \text{ Un possibile approccio per risolvere il sistema consiste nel dividere}$$

membro a membro le due equazioni che lo compongono, ottenendo così una relazione diretta tra le due variabili. Si avrà che:

$$\frac{\dot{y}}{\dot{x}}=\frac{-c \cdot y+d \cdot x \cdot y}{a \cdot x-b \cdot x \cdot y} \Rightarrow \frac{dy}{dx}=\frac{y \cdot (d \cdot x-c)}{x \cdot (a-b \cdot y)} .$$

Questa ultima espressione costituisce un'equazione differenziale lineare omogenea a variabili separabili. Ordinando e risolvendo si avrà che:

$$\frac{(a-b \cdot y)}{y} dy=\frac{(d \cdot x-c)}{x} dx \Rightarrow \left(\frac{a}{y}-b\right) dy=\left(d-\frac{c}{x}\right) dx \text{ integrando si avrà:}$$

$$\int\left(\frac{a}{y}-b\right) dy=\int\left(d-\frac{c}{x}\right) dx \Rightarrow a \cdot \ln(y)-b \cdot y=-c \cdot \ln(x)+d \cdot x+k \Rightarrow$$

$\Rightarrow \ln\left(y^a\right)+\ln\left(x^c\right)=d \cdot x+b \cdot y+k$  con  $k \in \mathbb{C}_0$  costante generica. Proseguendo si avrà  $\ln\left(y^a \cdot x^c\right)=d \cdot x+b \cdot y+k \Rightarrow y^a \cdot x^c=e^k \cdot e^{d \cdot x} \cdot e^{b \cdot y}$  posto  $K=e^k$  ottenendo che  $y^a \cdot e^{-b \cdot y}=K \cdot e^{d \cdot x} \cdot x^{-c}$ . Questa equazione rappresenta l'integrale generale del sistema indicando la relazione implicita tra le due variabili, tuttavia, non

consente di esprimere esplicitamente una variabile in funzione dell'altra. Per superare questo limite e semplificare l'analisi qualitativa del sistema, si introducono due nuove variabili ausiliarie così definite:

$z = y^a \cdot e^{-b \cdot y}$  ed  $w = K \cdot x^{-c} \cdot e^{d \cdot x}$ . Nel problema considerato, l'obiettivo è individuare una coppia di punti  $(x, y)$  tali che si abbia  $z=w$ . Assegnando valori generici ad  $x$  e calcolandone la sua immagine tramite la funzione  $w = K \cdot x^{-c} \cdot e^{d \cdot x}$ , l'uguaglianza  $z=w$  permetterà di calcolare anche un valore di  $z$ .

A tale valore corrisponderanno due valori distinti di  $y$ . Seguendo tale procedimento, si ricaverà una curva chiusa  $H$  che descriverà il comportamento ciclico dell'evoluzione delle due popolazioni –prede e predatori– come mostrato nel grafico sottostante.

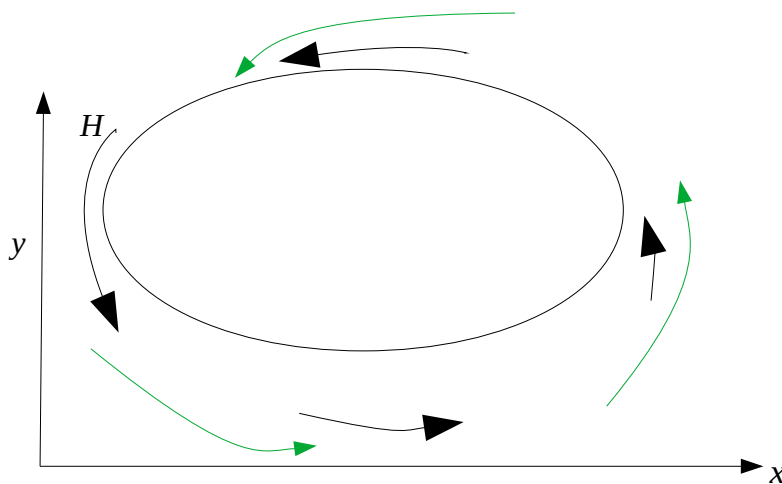


Figura 11: ciclicità della evoluzione delle due popolazioni.

Dunque, per come è strutturato il sistema, la ciclicità delle fluttuazioni delle popolazioni di prede e predatori non dipende da fattori esogeni al modello, bensì da fattori endogeni, riconducibili alla lotta per la sopravvivenza.

Si considerino le popolazioni medie in un periodo di tempo pari a  $T$ . Si definirà che:

$$\bar{x} = \int_0^T x(t) dt \quad \rightarrow \text{Popolazione media delle prede}$$

$$\bar{y} = \int_0^T y(t) dt \quad \rightarrow \text{Popolazione media dei predatori}$$

A causa della ciclicità delle fluttuazioni delle popolazioni, si avrà che:

$$x(0) = x(T), y(0) = y(T) \quad \text{ovvero, il numero di elementi nelle due popolazioni, al}$$

tempo iniziale di un periodo, coinciderà con quello al tempo finale del periodo. Considerando il sistema differenziale iniziale, si dividano le due equazioni rispettivamente per  $x$  ed  $y$  ottenendo il seguente nuovo sistema:

$$\begin{cases} \frac{\dot{x}}{x} = a \cdot \frac{x}{x} - b \cdot \frac{x}{x} \cdot y \\ \frac{\dot{y}}{y} = -c \cdot \frac{y}{y} + d \cdot x \cdot \frac{y}{y} \end{cases} \Rightarrow \begin{cases} \frac{x'}{x} = a - b \cdot y \\ \frac{y'}{y} = -c + d \cdot x \end{cases} \quad \text{integrando e risolvendo le singole}$$

equazioni differenziali si otterranno i valori medi cercati. Dunque si avrà che,

$$\int_0^T \frac{dx(t)}{x} = a \cdot \int_0^T d(t) - \int_0^T b \cdot y(t) dt \Rightarrow \ln(x(t)) \Big|_0^T = a \cdot t \Big|_0^T - b \cdot \int_0^T y(t) dt$$

$$\Rightarrow 0 = a \cdot T - b \cdot \int_0^T y(t) dt \Rightarrow \frac{1}{T} \cdot \int_0^T y(t) dt = \frac{a}{b} \Rightarrow \bar{y} = \frac{a}{b} .$$

Procedendo nello stesso modo si ha che il valore atteso di  $x$  nello stesso periodo di

$$\text{tempo sarà dato da } \int_0^T \frac{dy(t)}{y} = -c \cdot \int_0^T d(t) + \int_0^T d \cdot x(t) dt \Rightarrow$$

$$\Rightarrow \ln(y(t)) \Big|_0^T = -c \cdot t \Big|_0^T + d \cdot \int_0^T x(t) dt$$

$$0 = -c \cdot T + d \cdot \int_0^T x(t) dt \Rightarrow \frac{1}{T} \cdot \int_0^T x(t) dt = \frac{c}{d} \Rightarrow \bar{x} = \frac{c}{d} .$$

Considerando che la attività di pesca messa in atto dall'uomo genera un impatto sul numero medio delle specie in questione, si introdurrà tale ipotesi nel sistema differenziale iniziale. In particolare si inseriranno le due scritte  $\varepsilon \cdot x(t)$  ed  $\varepsilon \cdot y(t)$  che rappresentano rispettivamente la misura in cui diminuiranno le due specie a causa dell'attività di pesca, dove  $\varepsilon \in \mathbf{C}_0^+$  indica la intensità della attività di pesca. Inserendo tale parametro nel sistema differenziale iniziale si avrà il modello

così modificato  $\begin{cases} \dot{x} = (a - \varepsilon) \cdot x - b \cdot x \cdot y \\ \dot{y} = -(c + \varepsilon) \cdot y + d \cdot x \cdot y \end{cases}$ ; procedendo poi al calcolo dei valori

medi delle popolazioni, si otterranno risultati analoghi a quelli precedenti, ma con l'aggiunta del parametro d'intensità della pesca dipendente dalla stessa.

Ciò porterà la seguente modifica nei valori attesi  $\bar{x} = \frac{c + \varepsilon}{d}$  ed  $\bar{y} = \frac{a - \varepsilon}{b}$

confermando che, un aumento della attività di pesca comporterà, data la presenza del segno negativo associato al parametro  $\varepsilon$ , una riduzione dei predatori e conseguentemente un aumento delle prede. Viceversa, una diminuzione dell'attività di pesca determinerà un incremento dei predatori e una riduzione delle prede.

L'esempio descritto rappresenta un caso di come un fenomeno possa essere modellizzato attraverso un sistema dinamico. In generale, un sistema dinamico costituisce uno strumento per descrivere l'evoluzione di un fenomeno nel tempo, sfruttando le potenzialità del calcolo differenziale, in particolare delle equazioni

differenziali e dei sistemi di equazioni differenziali. Dato un fenomeno complesso costituito da più sottosistemi interagenti, e definite le equazioni differenziali che descrivono l'evoluzione di ciascun fenomeno, lavorare su un sistema dinamico significa individuare i punti di singolarità, ossia quei punti in cui i fenomeni cambiano qualitativamente, e successivamente determinarne la tipologia.

Si consideri un generico sistema dinamico semplificato formato da due fenomeni  $x, y$  del tipo  $\begin{cases} \dot{x}=F(x, y, t) \\ \dot{y}=G(x, y, t) \end{cases}$ ; i punti di singolarità (stazionarietà) del sistema saranno

le coppie  $(x^*, y^*) \in \mathbb{C}_0^2 \ni \begin{cases} F(x^*, y^*)=0 \\ G(x^*, y^*)=0 \end{cases}$ . Tali punti rappresentano uno

stato di equilibrio per il sistema, ovvero valori costanti che le variabili del sistema assumono al variare del tempo  $t$ , valori per cui risulta che  $x(t)=x^*, y(t)=y^*$

$\forall t \in \mathbb{C}_0^+$ . Uno dei metodi per stabilire la natura dei punti singolari è quello della linearizzazione del sistema mediante l'utilizzo delle derivate parziali.

Dato un sistema dinamico del tipo  $\begin{cases} \dot{x}=F(x, y, t) \\ \dot{y}=G(x, y, t) \end{cases}$ , si definisce la *matrice*

*Jacobiana* –formata dalle derivate parziali delle funzioni che compongono il sistema– nel seguente modo;

$J = \begin{pmatrix} \frac{\partial F}{\partial x} & \frac{\partial F}{\partial y} \\ \frac{\partial G}{\partial x} & \frac{\partial G}{\partial y} \end{pmatrix}$ . Tale matrice viene calcolata nei punti singolari, ottenendo una

matrice numerica del tipo,  $J(x^*, y^*)$ . A partire da questa si genererà il cosiddetto polinomio caratteristico di una matrice generica  $A$ , ovvero il determinante della matrice  $|A - \lambda \cdot I| = 0$ , dove in questo caso sarà  $A = J(x^*, y^*)$  ed  $I$  una matrice identità della stessa dimensione di  $A$ . Risolvendo il polinomio caratteristico per tutti gli autovalori  $\lambda$ , in relazione ai segni di quest'ultimi e dalla loro parte reale, si avranno le diverse nature dei punti di singolarità (nodi, centri, punti di sella, ecc). L'applicazione del metodo della linearizzazione al modello di Lotka-Volterra consente non solo di individuare i punti di equilibrio, ma anche di determinarne la natura (nodo, sella, centro, ecc.), attraverso l'analisi degli autovalori della matrice *Jacobiana* calcolata in tali punti.

• Definizione dei punti di singolarità:

$$\begin{cases} 0 = a \cdot x - b \cdot x \cdot y \\ 0 = -c \cdot y + d \cdot x \cdot y \end{cases} \text{ lavorando sulle equazioni si avrà } a \cdot x - b \cdot x \cdot y = 0 \Rightarrow$$

$$\Rightarrow x \cdot (a - b \cdot y) = 0 \Leftrightarrow x = 0 \vee y = \frac{a}{b}$$

$$-c \cdot y + d \cdot x \cdot y = 0 \Rightarrow y \cdot (d \cdot x - c) = 0 \Leftrightarrow y = 0 \vee x = \frac{c}{d}$$

ottenendo i due punti di singolarità,  $P^*_1=(0,0)$  e  $P^*_2=(\frac{c}{d}, \frac{a}{b})$

· Determinazione della natura dei punti di singolarità:

Si calcolano le derivate parziali delle due equazioni che compongono il sistema:

$$\frac{\partial F}{\partial x} = a - b \cdot y, \quad \frac{\partial F}{\partial y} = -b \cdot x$$

$$\frac{\partial G}{\partial x} = d \cdot y, \quad \frac{\partial G}{\partial y} = d \cdot x - c$$

Utilizzando le derivate parziali si costruirà la matrice Jacobiana:

$$J = \begin{pmatrix} a - b \cdot y & -b \cdot x \\ d \cdot y & d \cdot x - c \end{pmatrix}.$$

Tale matrice sarà valutata nei due punti singolari trovati ottenendo una matrice

numerica per ogni punto singolare  $J(P^*_1) = \begin{pmatrix} a & 0 \\ 0 & -c \end{pmatrix}$ .

Data la Jacobiana calcolata nel punto, si genererà il polinomio caratteristico al fine di definire la tipologia di punto stazionario (o singolare):

$$| J(P^*_1) - \lambda \cdot I | = 0 \Rightarrow \left| \begin{pmatrix} a & 0 \\ 0 & -c \end{pmatrix} - \lambda \cdot \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right| = 0 \Leftrightarrow$$

$$\Leftrightarrow \left| \begin{pmatrix} a - \lambda & 0 \\ 0 & -c - \lambda \end{pmatrix} \right| = 0 \Rightarrow (a - \lambda) \cdot (-c - \lambda) = 0 \Leftrightarrow$$

$\Leftrightarrow \lambda^2 + \lambda \cdot (c - a) - a \cdot c = 0$  risolvendo la equazione di secondo grado rispetto a  $\lambda$  si otterranno i due autovalori della Jacobiana, e precisamente

$\lambda_1 = a > 0 \vee \lambda_2 = -c < 0$  dunque i due autovalori, nel caso in esame, avranno segno sempre opposto ed il punto di singolarità sarà di *sella*, ovvero punto instabile del sistema dinamico.

Questa classificazione risulta fondamentale per comprendere la dinamica locale del sistema nelle vicinanze dei punti di equilibrio, e fornisce indicazioni sull'evoluzione delle traiettorie nel piano di fase.

Di seguito si riporta una rappresentazione grafica tipica di un *punto di sella* in un sistema dinamico bidimensionale:

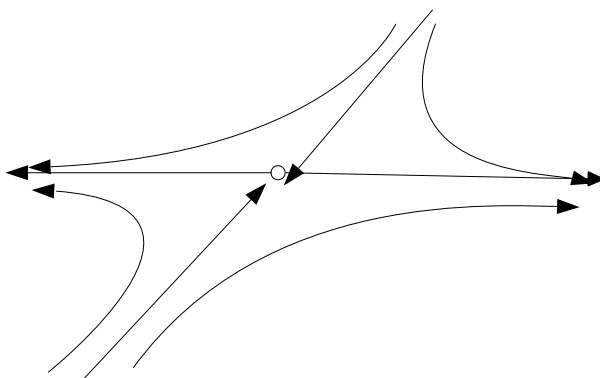


Figura 12: rappresentazione grafica del punto di sella.

Procedendo in modo analogo per il secondo punto di equilibrio  $P^*_2 = (\frac{c}{d}, \frac{a}{b})$  si

otterrà che  $J(P^*_2) = \begin{pmatrix} 0 & -b \cdot c \\ \frac{d \cdot a}{b} & 0 \end{pmatrix}$  e svolgendo i calcoli si giungerà al polinomio

caratteristico  $\lambda^2 + (c \cdot a) = 0$  da cui, risolvendo verso  $\lambda$  si otterranno i seguenti autovalori:  $\lambda_1 = i \cdot \sqrt{c \cdot a} \vee \lambda_2 = -i \cdot \sqrt{c \cdot a}$  ovvero autovalori immaginari puri. In tal caso, il punto di equilibrio è classificato come un *centro*, caratterizzato da *orbite chiuse* attorno al punto stesso. Sebbene la linearizzazione conduca ad un centro, questo metodo non consente di determinare con certezza la stabilità del punto nel sistema non lineare. Tuttavia, nel caso specifico del modello di Lotka-Volterra, ulteriori argomentazioni teoriche permettono di concludere che il centro è stabile, nel senso che le traiettorie restano chiuse e non si allontanano né si avvicinano al punto di equilibrio. In tale sistema vi è una ciclicità della numerosità delle due specie, ciclicità che fa tornare il sistema sempre al punto iniziale, con uno sfasamento temporale della numerosità delle due popolazioni.

Di seguito una rappresentazione grafica della ciclicità delle due popolazioni studiate:

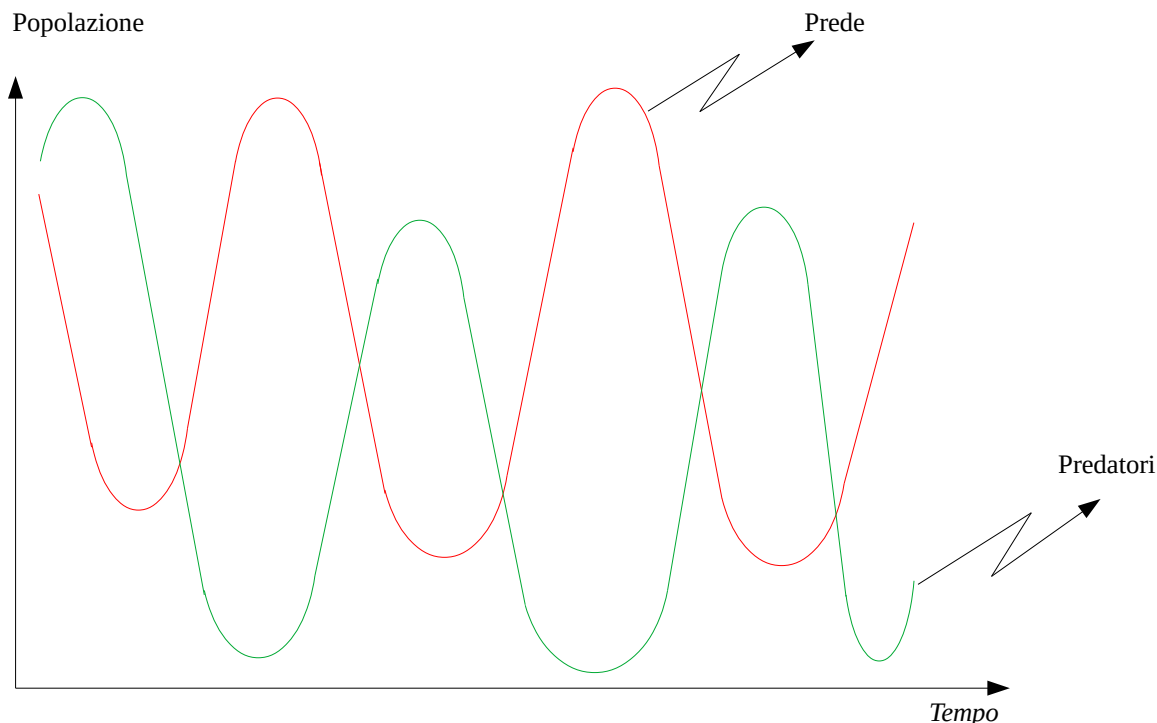
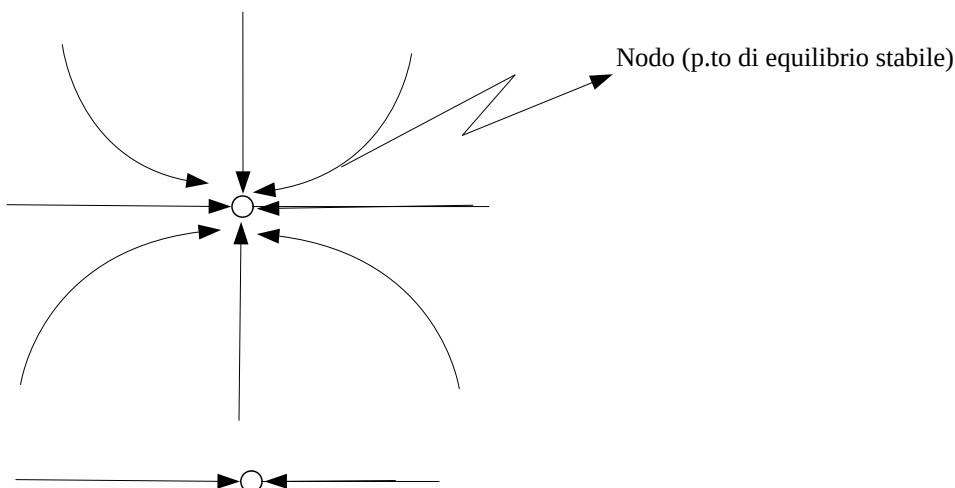
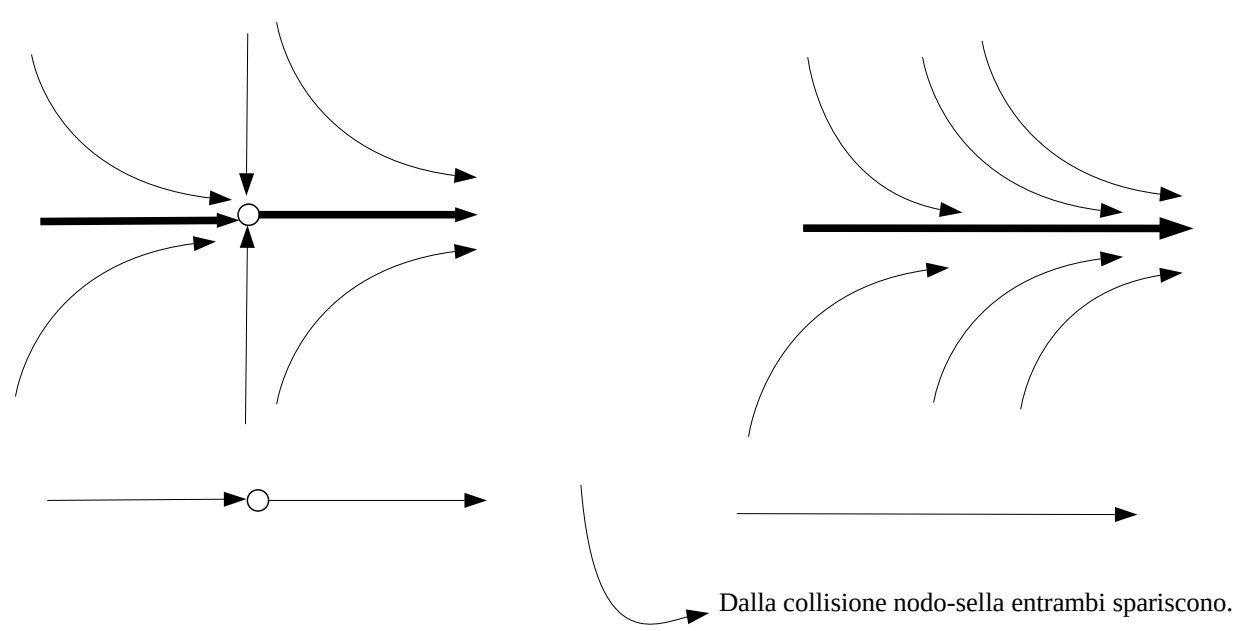
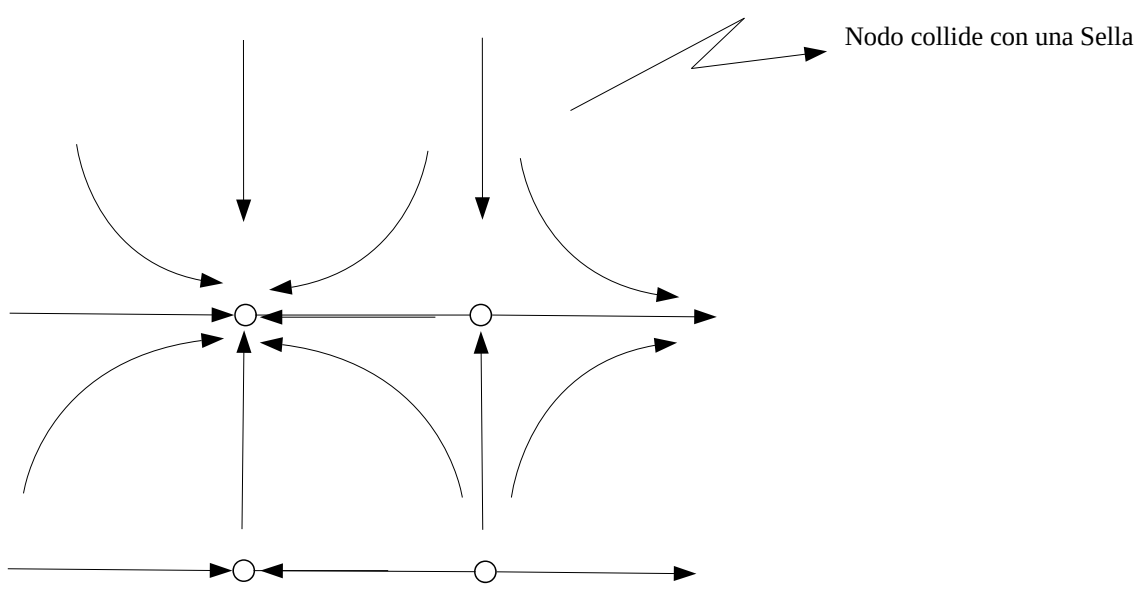


Figura 13: Grafico della ciclicità delle due popolazioni.

I sistemi dinamici, per come sono definiti, dipendono generalmente da uno o più parametri. Al variare di questi parametri (o *perturbazioni*), il comportamento del sistema può modificarsi in modo significativo, passando da condizioni di equilibrio stabile a condizioni instabili, e viceversa. In particolare, variazioni significative di uno o più parametri possono condurre il sistema a fenomeni di *biforcazione*, ovvero a valori critici dei parametri in corrispondenza dei quali si verifica un cambiamento qualitativo nella dinamica del sistema.

Lontano dai punti di biforcazione, gli stati di equilibrio risultano poco sensibili alle variazioni parametriche. Con l'approssimarsi dei parametri verso valori critici di biforcazione, il sistema passerà da situazioni di equilibrio a situazioni di disequilibrio fondendosi con altri stati, viceversa allontanandosi da punti di biforcazione moriranno situazioni di disequilibrio e si formeranno situazioni di equilibrio. Qualitativamente, la biforcazione rappresenta il punto in cui il sistema varia, ovvero passa da un comportamento dinamico a un altro, anche in modo drastico. Ad esempio, considerando le curve di fase di un sistema dinamico dipendente da un singolo parametro, si può osservare che, al variare del valore del parametro, uno stato di equilibrio stabile (come un nodo) può entrare in collisione con uno instabile (come una sella). Questa interazione porta alla fusione dei due punti di equilibrio, i quali si annulleranno a vicenda, generando una transizione qualitativa nel comportamento del sistema: si verifica così una biforcazione nodo-sella. Di seguito si riporta una rappresentazione dei due stati di equilibrio e del loro processo di collisione :





Questo esempio evidenzia come la struttura qualitativa di un sistema dinamico possa subire mutamenti radicali anche in seguito a piccole variazioni di uno o più parametri. Ciò mette in risalto l'importanza dell'analisi biforcazionale nello studio dei sistemi dinamici non lineari, in particolare per la comprensione delle transizioni tra stati di equilibrio. In termini applicativi, qualora uno stato di equilibrio stabile rappresenti una condizione regolatrice di un sistema reale -come un sistema economico, sociale, ecologico ecc.- l'interazione con uno stato instabile (come in una biforcazione nodo-sella) potrebbe causare una transizione improvvisa verso un nuovo stato. In tal caso, una modifica del parametro oltre una soglia critica può comportare la scomparsa della condizione di equilibrio precedentemente stabile, anche in un intorno ristretto, portando il sistema ad un comportamento qualitativamente differente e potenzialmente imprevedibile. Tale fenomeno ha ispirato il termine *teoria delle catastrofi* riguardando essa proprio lo studio delle transizioni improvvise e discontinuità nei sistemi dinamici, pur in presenza di variazioni continue dei parametri. In tale contesto, le *catastrofi* non rappresentano eventi drammatici in senso comune, ma piuttosto – come già esplicitato – cambiamenti qualitativi improvvisi nello stato di un sistema, come nel caso delle biforcazioni.

È importante precisare che la perdita di stabilità di un punto di equilibrio in un sistema dinamico non implichi necessariamente la presenza di una biforcazione. In altre parole, un sistema può passare da uno stato stabile a uno instabile -o viceversa- anche senza che avvenga una collisione tra punti di equilibrio, come avviene nelle biforcazioni nodo-sella. Pertanto, pur essendo le biforcazioni uno strumento fondamentale per l'analisi qualitativa dei sistemi dinamici, esse non esauriscono tutti i meccanismi possibili di cambiamento della stabilità all'interno di un sistema. Per introdurre in modo intuitivo il concetto di biforcazione, è utile partire da un esempio semplice, introducendo prima alcuni concetti ancillari.

Si consideri la famiglia di funzioni quadratiche dipendenti da un parametro  $c$ , definita come  $Q_c(x) = x^2 + c$  dove il pedice  $c$  indica che si tratta di una famiglia di funzioni. Allora  $\forall c \in \mathbb{C}_0$  si definirà una diversa funzione della famiglia. Può essere utile studiare le orbite di punti del dominio che la funzione genera al variare del parametro  $c$ . Con orbita di un punto del dominio si intendono i valori che la funzione assume iterando l'applicazione ad ogni risultato, partendo da un valore del dominio  $x_0$ .

Se, ad esempio, considerassimo la funzione  $f(x) = x^2$  e volessimo calcolare l'orbita delle prime 3 iterazione del valore del dominio  $x_0 = 4$ , allora avremmo la seguente successione,

$$4, f(4) = 16, f^2(4) = f(f(4)) = f(16) = 256, f^3(4) = f(f^2(4)) = 65536$$

ottenuta applicando la funzione quadrato ad ogni risultato. Si noti come, in questo caso, l'orbita cresca rapidamente: si tratta quindi di un punto instabile rispetto alla funzione  $f(x) = x^2$ , dato che le iterazioni divergono.

Il procedimento illustrato sopra può essere esteso a qualunque punto del dominio della funzione, ed il numero di iterazioni può anch'esso essere scelto in modo

arbitrario. In questo contesto semplificato, è possibile stabilire un'interessante analogia con la teoria delle equazioni differenziali, già affrontata in precedenza. Le equazioni differenziali, infatti, forniscono gli strumenti per descrivere l'evoluzione temporale di un sistema continuo, ad esempio la traiettoria di un oggetto in movimento. Analogamente, le iterazioni di una funzione descrivono una dinamica discreta, in cui un valore iniziale (punto) evolve nel tempo in seguito all'applicazione ripetuta della funzione. In entrambi i casi -continuo e discreto- si studia l'evoluzione di uno stato nel tempo, ovvero si analizza il comportamento di una traiettoria (orbitale) in uno spazio di fase. Pertanto, iterare una funzione su diversi punti iniziali del dominio equivale a svolgere un'analisi delle orbite per il sistema dinamico discreto associato. Questo tipo di studio permette di osservare fenomeni come fissità, ciclicità, divergenza, o comportamenti caotici, in funzione del valore del parametro e del punto iniziale. In un sistema dinamico discreto semplificato, se l'iterazione di un funzionale a partire da un punto del dominio genera come orbita sempre lo stesso punto, allora tale punto si definisce *punto fisso* del sistema. L'iterazione della funzione a partire da un punto fisso non modifica lo stato iniziale: il sistema rimane fermo su quel valore. Questo concetto è analogo al concetto di equilibrio nei sistemi dinamici continui. Data la funzione precedente introdotta  $f(x)=x^2$ , il punto  $x=1$  sarà *punto fisso* per essa poiché  $f^n(1)=1 \quad \forall n \in \mathbb{N}$ . L'orbita del punto  $x=1$  risulta quindi costante, e ciò conferma che si tratti di un punto fisso in quanto l'iterazione della funzione non ne modifica il valore iniziale. Esso non si muoverà malgrado l'applicazione iterata della funzione. I punti fissi individuati nel contesto delle iterazioni discrete corrispondono a quelli che, nei sistemi dinamici, vengono definiti *punti di equilibrio*. Si tratta di stati del sistema che rimangono invariati nel tempo. In tali punti, il sistema non evolve: né l'iterazione discreta né l'integrazione continua determinano un cambiamento dello stato iniziale. Essi rappresentano, pertanto, le configurazioni in cui il sistema rimane fermo nel tempo. Se invece, prendessimo il punto  $x=-1$  appartenente al dominio della funzione in esame  $f(x)=x^2$  e ne studiassimo la natura, si otterrebbe che  $f(-1)=(-1)^2=1 \neq -1$ . Dunque, in questo caso, il punto non risulterà *punto fisso* per la funzione data poiché, l'iterazione della funzione ne modificherà il valore iniziale. Un altro concetto fondamentale nell'ambito dei sistemi dinamici può essere introdotto in modo più intuitivo ricorrendo, ancora una volta, ad un sistema semplificato simile al precedente. Sia, ad esempio, il funzionale  $G(x)=\frac{1}{2} \cdot x$ ; si ha che  $G^n(0)=0$

$\forall n \in \mathbb{N}$  dunque il punto  $x=0$  è un punto fisso per il sistema semplificato dato. Un'analisi delle orbite mostra che tutte le traiettorie sono attratte dal punto fisso  $x=0$ . In altri termini, per ogni punto appartenente al dominio della funzione, la rispettiva orbita converge a zero: al crescere del numero di iterazioni, il valore della funzione tende a 0. Questo comportamento si verifica per qualunque punto del dominio considerato. Questa peculiarità consente di definire tale punto fisso come *punto fisso attrattore* per la funzione poiché, esso attrae a sé le orbite di tutti i punti del dominio

della funzione. Generalizzando, data una funzione  $F$  qualunque e dato un punto  $p$  per essa fisso,  $p$  sarà *punto fisso attrattore* per  $F$  se  $\exists ]a,b[$  con  $a,b \in \mathbb{C}_0$  e  $a < p < b$  in modo tale che le orbite di ogni punto dell'intervallo che contiene  $p$  tendano a  $p$  stesso. In simboli si ha che,  $p$  è *punto fisso attrattore* per  $F \Leftrightarrow \forall x \in ]a,b[$  con  $a < p < b \mid F^n(x) \xrightarrow{n \rightarrow +\infty} p$ . Dunque il punto attrattore attirerà a sé tutte le orbite dei punti che saranno sufficientemente vicini ad esso. Per la funzione utilizzata prima  $f(x)=x^2$  il punto  $x=0$  è di attrazione per la funzione per tutti i valori  $x \in ]-1,1[$  poiché essi generano orbite che tendono a 0. L'intervallo  $] -1,1[$  è detto *bacino di attrazione*. Il grafico seguente rappresenta tale concetto:

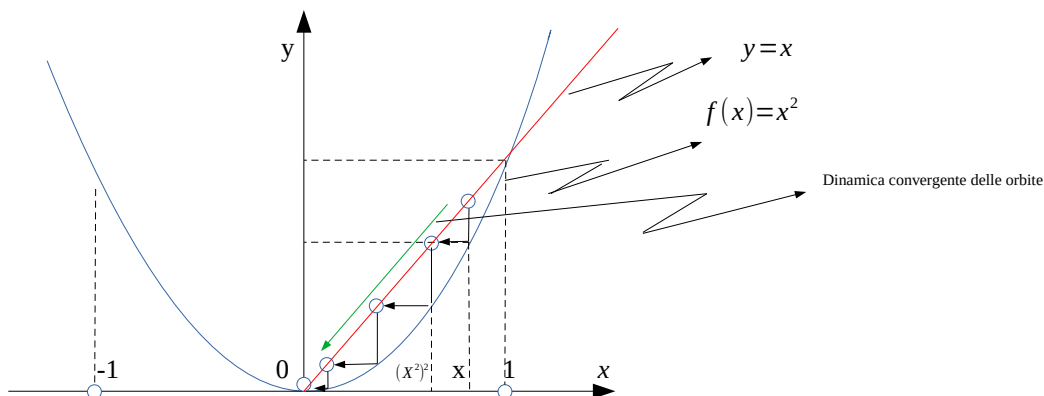


Figura 14: bacino di attrazione ed orbite per  $x \in ]-1,1[$  della funzione  $f(x)=x^2$

Se invece considerassimo la funzione  $f(x)=x^2$  nell'intervallo  $[1, +\infty[$  allora il punto fisso 1 sarà detto di *repulsione* poiché, al contrario del precedente, tutti i valori diversi da 1 appartenenti all'intervallo considerato, genereranno orbite che si allontaneranno velocemente dal punto fisso 1.

Sfruttando le caratteristiche delle orbite si può definire una proprietà importante dei punti fissi, ovvero la *stabilità* ed *instabilità*. Un punto fisso – sia esso attrattore o repulsore – è detto *stabile* per il sistema dinamico se, per ogni piccola perturbazione dei punti vicini ad esso, la nuova orbita generata da queste piccole perturbazioni è molto vicina a quella del punto precedente.

Per gli esempi visti prima, le orbite dei valori del dominio tra -1 ed 1 sono stabili così come lo sono quelle dei valori maggiori di 1 poiché, in entrambi i casi, tutte le orbite dei punti del dominio considerato tendono al punto attrattore per il primo caso e sono esplosive, per il secondo caso. L'orbita generata dal valore 1 della funzione che si sta considerando, invece, sarà *instabile* poiché, tutti i valori interni all'intervallo

$] -1, 1[$  generano orbite convergenti a zero, mentre tutte quelle generate dai punti interni all'intervallo  $[ 1, +\infty[$  saranno esplosive. Ovvero, una orbita di un punto è detta *instabile* se le orbite di punti ad esso adiacenti hanno comportamenti completamente differenti tra loro, come nel caso in esame. Tornando al concetto di biforcazioni, si utilizzerà la funzione quadratica  $Q_c(x) = x^2 + c$  per esplicitarlo al meglio. Come accennato in precedenza, al variare del parametro  $c$  si otterranno differenti funzioni quadratiche, ciascuna delle quali genererà orbite diverse per i punti del dominio. Si assegni il seguente valore arbitrario al parametro  $c$ :  $c = \frac{1}{4}$ . La funzione risultante rappresenterà una parabola con la concavità rivolta verso l'alto ed intersecante l'asse delle ordinate nel punto  $(0, c)$ .

Il grafico della funzione intersecherà la bisettrice  $y = x$  in 2 punti che rappresentano le soluzioni della equazione  $x^2 + c = x \Leftrightarrow x^2 + c - x = 0$  che hanno valori

$$d(c) = \frac{(1 + \sqrt{1 - 4 \cdot c})}{2} \vee q(c) = \frac{(1 - \sqrt{1 - 4 \cdot c})}{2} \text{ e saranno reali per } c < \frac{1}{4}.$$

Dunque per differenti valori del parametro  $c$ , dal quale dipenderanno le due soluzioni, il grafico della funzione giacerà sopra la bisettrice, sarà ad essa tangente, sarà secante nei punti trovati se rispettivamente si avrà che  $c > \frac{1}{4}, c = \frac{1}{4}, c < \frac{1}{4}$ .

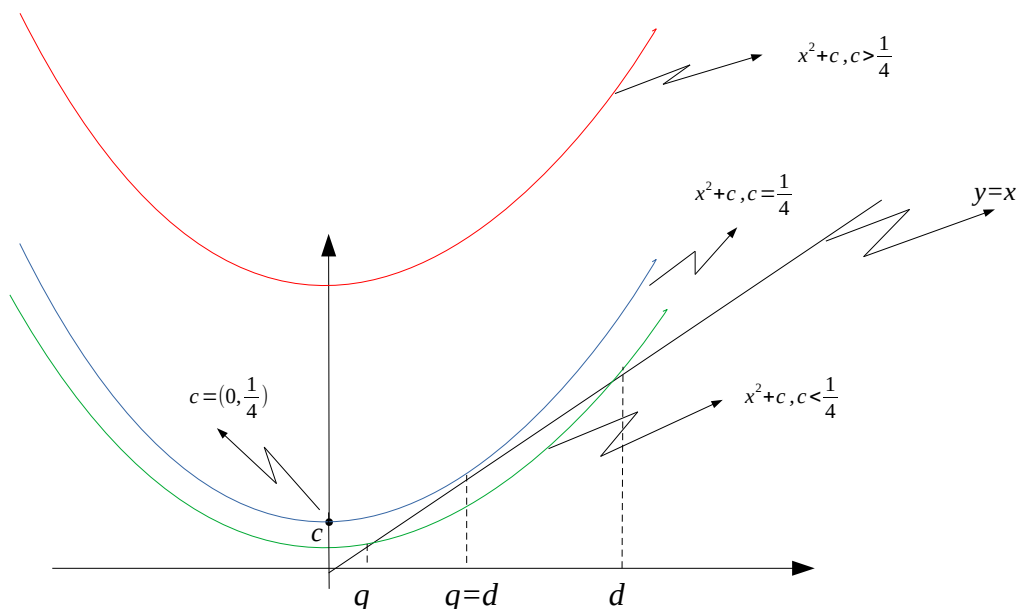


Figura 15: Posizione della  $Q_c(x) = x^2 + c$  al variare del parametro  $c$ .

Le radici  $d$  e  $q$  della funzione in esame, da un punto di vista dinamico, rappresentano dei punti fissi per la stessa funzione. Si può dire che, riducendo il parametro  $c$ , quando esso passa per il valore critico  $\frac{1}{4}$  il sistema dinamico passa da una configurazione con un solo punto fisso reale ad una con due punti fissi reali distinti. Questo fenomeno rappresenta un esempio classico di biforcazione, ovvero un cambiamento qualitativo nella struttura del sistema al variare di un parametro.

In particolare, si verifica una biforcazione del punto fisso  $c = \frac{1}{4}$  poiché un singolo punto fisso -inizialmente reale e doppio- si sdoppia in due punti fissi distinti quando  $c$  scende sotto il valore critico. La configurazione del sistema, quindi, si biforca, dando origine a due comportamenti dinamici differenti. Inoltre, ai fini dell'analisi dinamica, risultano particolarmente rilevanti le orbite che non divergeranno all'infinito e questo avverrà per valori del parametro  $c \in ]-2, \frac{1}{4}[$ . In tal caso, le

orbite generate dai punti della funzione che si otterranno col valore del parametro rientrante nel suddetto intervallo, saranno tutte quelle generate dai valori del dominio appartenenti all'intervallo che chiamiamo  $I_c$ . Ciò vuol dire che, qualora il punto iniziale  $x_0$  appartenga all'intervallo  $I_c$ , la sua orbita resterà intrappolata in un intervallo limitato, intervallo che conterrà tutte le orbite interessanti della funzione, ovvero quelle orbite che non tenderanno ad infinito e che saranno generate dai valori di  $x \in I_c$ . Con semplici passaggi si vede che  $I_c = ]-d, d[$  col parametro

$c \in ]-2, \frac{1}{4}[$ . L'intervallo  $I_c$  ha il pedice  $c$  poiché sta proprio ad indicare la sua

dipendenza dal parametro  $c$ , visto che la soluzione stessa  $d$  è funzione di  $c$  dato che

$d(c) = \frac{(1 + \sqrt{1 - 4 \cdot c})}{2}$ . Si può pensare ad  $I_c$  come ad un bacino di attrazione

per la funzione in oggetto, poiché le orbite dei punti ad esso appartenenti, rimarranno sempre racchiuse in un intervallo limitato senza mai esplodere verso infinito. Il concetto di orbita è strettamente collegato a quello di spazio delle fasi, introdotto in precedenza. In questo contesto, ogni orbita associata ad un punto dell'intervallo di analisi può essere vista come la traiettoria temporale del sistema, ovvero come la sequenza degli stati assunti dal sistema al trascorrere del tempo.

Nei punti in cui si verificano biforcazioni, le orbite risultano tipicamente instabili. Infatti, preso un opportuno intorno del punto di biforcazione, si osserva che valori iniziali arbitrariamente vicini possono generare orbite con comportamenti qualitativamente molto differenti. Questo fenomeno evidenzia l'instabilità intrinseca dei punti di biforcazione: anche piccole variazioni nei dati iniziali o nei parametri del sistema potranno condurre a traiettorie dinamiche profondamente diverse.

Ogni biforcazione rappresenta, dunque, una soglia critica oltre la quale il sistema modifica in modo netto la propria struttura qualitativa. In Figura 12, considerando un

intorno sufficientemente piccolo del punto  $x=1$ , si osserva un comportamento asimmetrico delle orbite. I valori dell'intorno alla sinistra di 1 generano orbite convergenti verso il punto fisso attrattore  $x=0$ , mentre quelli alla destra di 1 producono orbite esplosive. Sebbene le orbite su entrambi i lati mostrino una certa regolarità -convergenza o divergenza-, il punto  $x=1$  risulta instabile, poiché rappresenta una soglia oltre la quale il sistema cambia qualitativamente il proprio comportamento. In corrispondenza di questo punto si verifica una *biforcazione*, ovvero una transizione dinamica che segna una netta discontinuità strutturale del sistema: orbite che si avvicinano al punto fisso da un lato, orbite che esplodono dall'altro. Il sistema, dunque, si biforca, evidenziando una rottura della coerenza dinamica nello spazio delle fasi. Sebbene, ad una prima analisi, la stabilità di un sistema dinamico possa sembrare strettamente connessa al concetto di biforcazione, in realtà i due fenomeni non sono necessariamente legati tra loro. Un sistema dinamico in una configurazione stabile può quindi transitare verso una condizione di instabilità senza che ciò implichi necessariamente la presenza di un fenomeno di biforcazione. Se si considerasse uno stato di equilibrio di un sistema dinamico come il corrispondente stato stazionario di un sistema fisico reale, allora al variare del parametro da cui dipende lo stato di equilibrio, si potrebbero osservare due fenomeni in merito alla riorganizzazione dello stato delle fasi:

1) Al variare del parametro di controllo, può accadere che lo stato di equilibrio perda stabilità e dia origine ad un ciclo limite: il sistema comincia allora ad esibire un comportamento oscillatorio, la cui ampiezza risulta proporzionale alla radice quadrata della criticità, intesa come la differenza tra il valore del parametro ed il corrispondente valore critico di biforcazione. In questa configurazione, la stabilità dell'equilibrio viene trasferita al ciclo limite, mentre il punto di equilibrio precedentemente stabile diventa instabile. Tale perdita di stabilità viene definita *soffice* (o *supercritica*) poiché, per valori piccoli della criticità, il comportamento oscillatorio risultante si discosta in modo graduale e contenuto dallo stato di equilibrio. Di seguito una rappresentazione grafica del fenomeno:

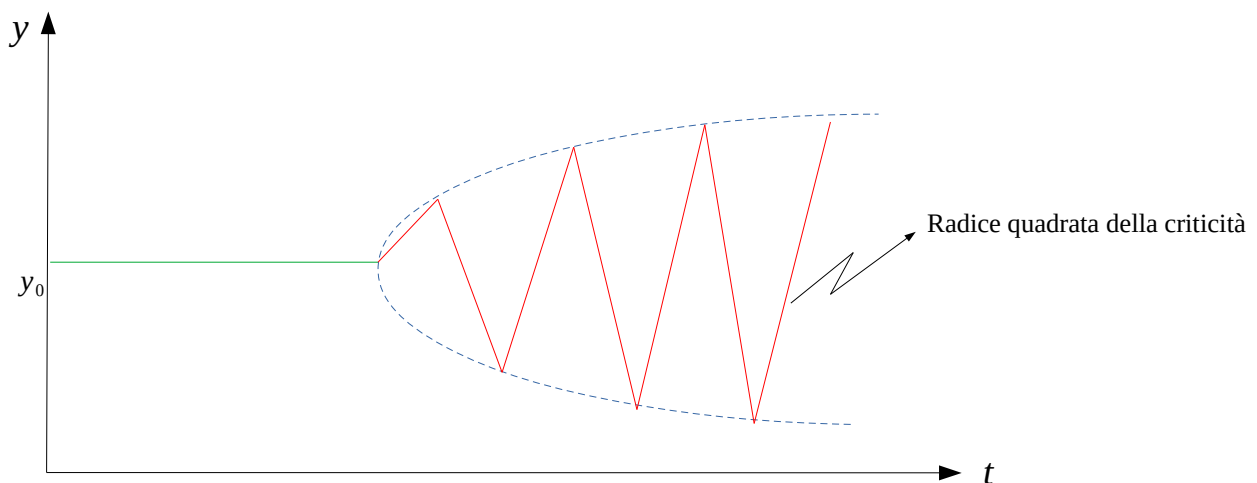


Figura 16: Grafico del fenomeno 1 rappresentante le oscillazioni del parametro dopo la perdita di stabilità.

Nel grafico,  $y$  rappresenta l'asse di valori del parametro, ed  $y_0$  rappresenta il valore di riferimento del parametro del sistema nello stato di equilibrio.

2) Si verifica un collasso di un ciclo limite instabile verso uno stato di equilibrio. Questo avviene in corrispondenza della progressiva riduzione a zero del dominio di attrazione sia del ciclo limite che dell'equilibrio stesso. Alla fine, il ciclo limite scompare, trasferendo l'instabilità al punto di equilibrio, che da stabile diventa instabile.

Tale transizione è detta perdita di stabilità dura (o biforcazione *rigida/subcritica*), in quanto non è accompagnata da una fase intermedia di comportamento ordinato. Le perturbazioni anche minime possono spingere il sistema al di fuori del dominio di attrazione prima ancora che questo si annulli del tutto, portandolo ad un abbandono improvviso dello stato stazionario. Il sistema passa quindi bruscamente ad un nuovo regime dinamico, che può essere a sua volta stazionario, periodico, o caratterizzato da dinamiche più complesse, fino ad arrivare, in certi casi, a un comportamento caotico. In presenza di condizioni di moto che tendono a stabilizzarsi nel tempo, si parla di attrattori, in quanto questi rappresentano dei centri di attrazione per le traiettorie iniziali poste nelle loro vicinanze.

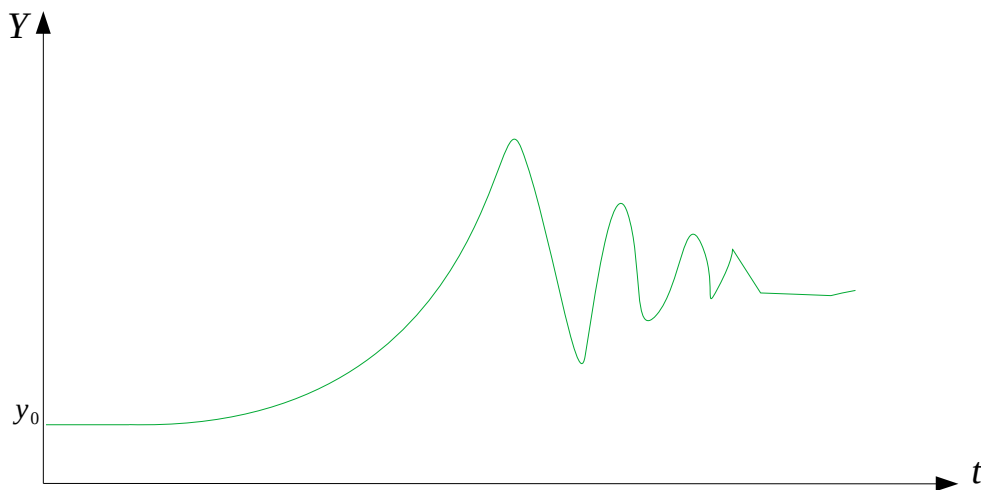


Figura 17: Grafico della fenomeno 2 in cui si ha un salto da uno stato stazionario ad uno di moto.

Dunque, un attrattore può essere considerato un insieme attrattivo nello spazio delle fasi, ovvero un insieme verso cui convergono, nel tempo, le orbite dei punti iniziali appartenenti a un opportuno intorno.

Gli attrattori descrivono quindi il comportamento asintotico del sistema dinamico. La esposizione del concetto di *Biforcazione*, e degli altri trattati durante il capitolo, è assai succinta e non ha la finalità di effettuare, tramite essi, una disamina esaustiva della teoria delle catastrofi di Renè Thom poiché, per la complessità di essa, una esposizione capillare ed esaustiva imporrebbe un intero lavoro improntato su tale argomento. I concetti fin qui esposti hanno l'unico scopo di fornire una panoramica generale della teoria delle catastrofi, utile a inquadrarne i principi fondamentali. Ciò al fine di introdurre il parallelismo con la metodologia statistica impiegata nella ricerca oggetto del presente lavoro, che sarà approfondito nel paragrafo successivo.

### 2.3

#### Un possibile utilizzo della teoria delle catastrofi in contesti reali

Dall'esposizione fin qui condotta della teoria di René Thom emerge uno strumento concettuale potente, ma fortemente teorico, per lo studio dei sistemi complessi. In pratica, un'applicazione rigida e letterale della teoria non risulta sempre fattibile. L'idea centrale su cui essa si fonda deve essere adattata e ri-contestualizzata in funzione delle caratteristiche specifiche del fenomeno complesso oggetto di analisi. Pertanto, la teoria delle catastrofi offre una chiave interpretativa utile, ma richiede un certo grado di flessibilità metodologica per essere impiegata efficacemente nello studio di casi reali. Nel presente lavoro, tale impostazione verrà sviluppata partendo dai principi fondamentali della teoria delle catastrofi ed applicandoli alla metodologia statistica adottata. In particolare, trattandosi di uno studio sul benessere e sulla ricchezza di un territorio secondo le linee guida del documento *BES*, la scelta della metodologia statistica da parte dello scrivente trova giustificazione nel concetto cardine di *processo generatore dei dati*. Il documento *BES* considera il benessere e la ricchezza di una società attraverso una serie di dominî tematici, intesi come categorie concettuali che permettono di misurare le condizioni di benessere di una comunità. Ciascun dominio è definito da un insieme di indicatori selezionati per monitorare aspetti specifici della realtà sociale, ritenuti dagli autori del documento rappresentativi ed adeguati a misurare ed osservare l'evoluzione del benessere collettivo. Alla base della metodologia adottata in questo lavoro vi è, come accennato in precedenza, il concetto fondamentale – tanto semplice quanto potente – di processo generatore dei dati. L'idea di fondo si basa sul presupposto che, qualora per ciascun dominio gli indicatori scelti per rappresentarne la corrispondente variabile latente presentino un forte legame concettuale con essa, si possa allora ipotizzare che tali indicatori siano generati dallo stesso processo sottostante.

In termini statistici, ciò implica la possibilità che tali indicatori risultino co-integrati. La presenza di co-integrazione tra gli indicatori costituirà, pertanto, il fondamento

teorico-metodologico dell'analisi empirica del benessere che verrà dunque condotta attraverso lo studio dei processi stocastici co-integrati.

L'esposizione dettagliata della metodologia sarà oggetto del capitolo successivo; in questa sede è sufficiente chiarire che il parallelismo tra la teoria delle catastrofi ed il concetto di co-integrazione si basa sull'idea di *tipo qualitativo*, elemento centrale della teoria di Thom. In particolare, una volta implementato il sistema co-integrato tra gli indicatori, il concetto di tipo qualitativo sarà applicato alla combinazione lineare stazionaria del II ordine risultante dal sistema co-integrato. Tale combinazione lineare dovrebbe rappresentare, nel modello adottato, la manifestazione empirica del tipo qualitativo che, nella teoria delle catastrofi, costituisce il fondamento del cambiamento strutturale nei sistemi complessi. L'idea alla base dell'integrazione tra la teoria delle catastrofi ed i sistemi co-integrati si fonda sull'assunto che i livelli della combinazione lineare stazionaria, in corrispondenza dei quali si verificano cambi di direzione nelle variabili del sistema, possano essere interpretati come soglie in cui le variabili stesse subiscono un cambio qualitativo. In tal senso, tali livelli assumono, all'interno del sistema co-integrato, il ruolo che l'insieme delle catastrofi assume nella stessa teoria.

Nel grafico riportato sotto (Figura 18) è rappresentata una combinazione lineare stazionaria del II ordine, derivante da un sistema generico co-integrato.

I punti indicati dalle frecce presenti sulle bande superiore, inferiore e centrale del grafico rappresentano i livelli ai quali, tramite l'analisi del vettore di co-integrazione, è possibile determinare se le variabili del sistema assumeranno un segno negativo, positivo oppure se il processo di convergenza si concluderà con il raggiungimento di un punto di equilibrio tra le stesse variabili. Le bande superiori (rosse) ed inferiori (verdi) della combinazione lineare stazionaria sono interpretate come soglie critiche, oltre le quali il fenomeno in esame subisce un cambiamento qualitativo. In corrispondenza di tali livelli, si verifica una variazione concettuale della "forma" del fenomeno, intesa come l'andamento assunto dalle variabili del sistema. Da questi punti in avanti, tale andamento cambia fino a stabilizzarsi al raggiungimento della banda centrale (nera). Di conseguenza, le bande in questione possono essere lette come la rappresentazione dell'*insieme delle catastrofi* nella teoria di Thom, ovvero come quei livelli topologici in cui il fenomeno muta la propria configurazione qualitativa. Questo parallelismo mira a integrare la teoria delle catastrofi con il concetto di co-integrazione stocastica. In corrispondenza di tali livelli critici, le variabili co-integrate subiscono una *catastrofe* nel senso teorico del termine assunto nella teoria, ovvero subiscono una trasformazione qualitativa. Tale trasformazione si verificherà ogni qualvolta la combinazione lineare stazionaria, generata dal sistema co-integrato, raggiungerà tali valori soglia.

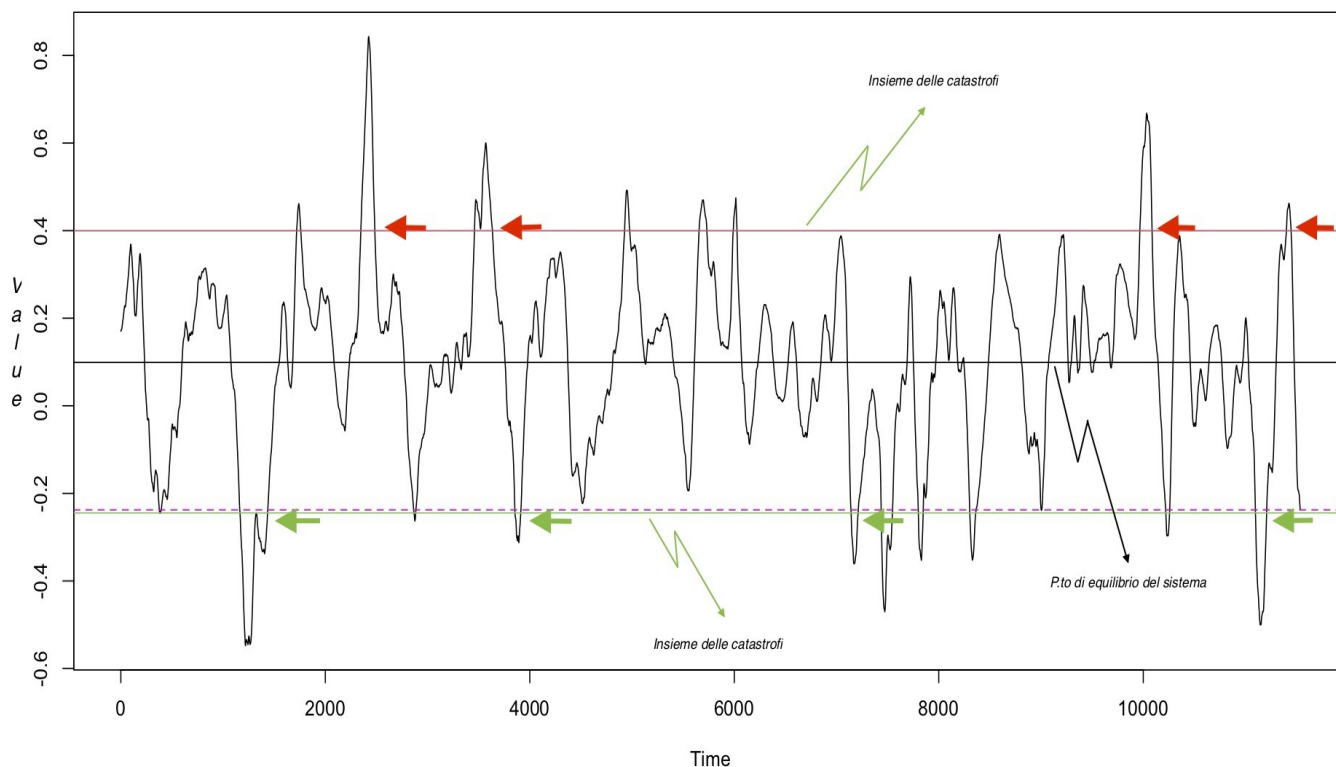


Figura 18: Combinazione lineare stazionaria del II ordine di un sistema co-integrato.

L'obiettivo di questo approccio è consentire un'analisi predittiva degli indicatori appartenenti ai diversi domini, assumendo che essi derivino concettualmente da uno stesso processo generatore dei dati e, pertanto, siano tra loro co-integrati. A tale finalità si affianca un impianto teorico fondato sui principi basilari della teoria delle catastrofi, impiegati in chiave interpretativa per individuare soglie critiche e cambiamenti qualitativi all'interno del sistema.

Le idee alla base di questa interpretazione della struttura e composizione degli indicatori dei domini del BES saranno ulteriormente approfondite e chiarite nella parte applicativa del lavoro, in particolare nel capitolo ad essa dedicato.

## CAP. 3

### Alcuni modelli quantitativi per l'analisi delle serie storiche

#### 3.1

#### Nozioni generali sui processi stocastici

Partendo da un concetto fulcro (di cui parlerò nel prosieguo del capitolo) in esso ho trovata la ispirazione per la scelta della modellistica statistica da utilizzare per effettuare l'analisi dei dati utilizzati per il presente lavoro.

Lo studio statistico di un sistema di traiettorie o serie storiche, non può non considerare preliminarmente una analisi uni-variata di esse, al fine di poterne definire delle proprietà salienti che le caratterizzano. Si inizierà col dire che, una traiettoria o serie storica è considerata una realizzazione di un processo stocastico (p.s.) non noto, il quale rappresenta il corretto processo che ha generato i dati che formano la traiettoria. Si definisce processo stocastico un funzionale  $X(t,s)$ , ovvero una relazione tra un insieme dei tempi  $t \in T$  ed  $s \in S$  insieme delle realizzazioni.

Effettuando l'operazione di bloccare la variabile  $t \in T$  si ottiene un p.s.  $\{y_t\}_{t \in T}$  che, dunque, rappresenta una successione di variabili aleatorie definite tutte sullo spazio di probabilità  $(\Omega, \mathcal{F}, P)$  indicizzate dal tempo.<sup>2</sup>

Di seguito una rappresentazione grafica di un processo stocastico:

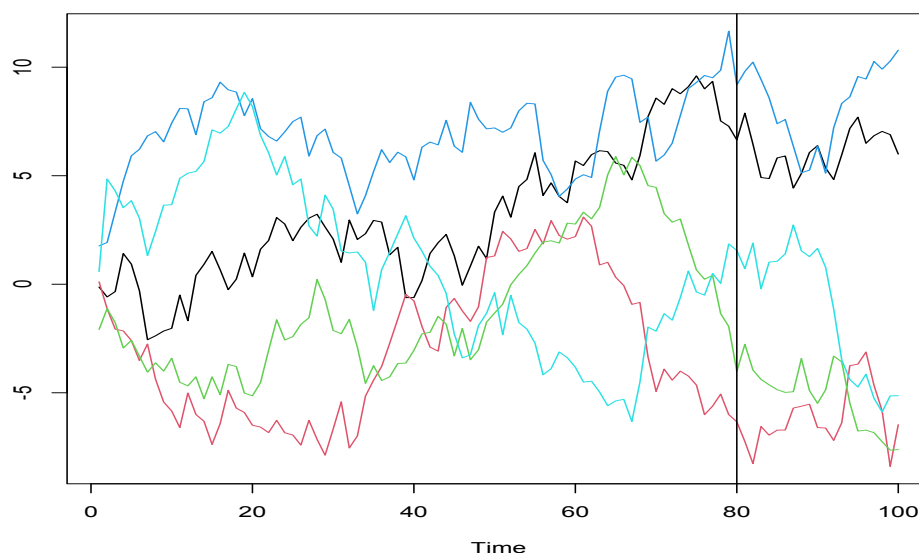


Figura 19: Esempio grafico di un processo stocastico

2. Gli elementi della terna  $(\Omega, \mathcal{F}, P)$  rappresentano rispettivamente, lo spazio campionario  $\Omega$ , ovvero l'insieme dei possibili risultati assumibili dalla variabile aleatoria (v.a.);  $\mathcal{F}$   $\sigma$ -algebra di eventi di  $\Omega$ ;

$P : (\Omega, \mathcal{F}) \rightarrow (\mathbf{C}_0, B)$  funzione probabilità con  $B = \sigma$ -algebra di eventi di  $\mathbf{C}_0$  detta " $\sigma$ -algebra di Borel".

Si parla di processo rispettivamente univariato o scalare e multivariato o vettoriale a seconda che la v.a.  $\{y_t\}_{t \in T}$  sia semplice o multipla.

La descrizione di un p.s. avviene se si conoscono le distribuzioni marginali finite di esso, ovvero, se è nota la distribuzione di probabilità congiunta

$$P(y_{t_1}, y_{t_2}, \dots, y_{t_n}) \text{ per ogni insieme di } n \geq 1 \text{ indici } (t_1, t_2, \dots, t_n) \in T.$$

Si può, però, sintetizzare la legge di probabilità attraverso dei parametri tra cui, in evidenza si ha il valore atteso, la varianza e l'autocovarianza.

La definizione formale di tali indicatori è rappresentata di seguito:

i) Valore atteso:

$$\mu_t = E(y_t)$$

ii) Varianza:

$$\sigma_t^2 = E[(y_t - \mu_t)^2] = E(y_t^2) - \mu_t^2$$

iii) Autocovarianza:

$$\gamma_t(k) = E[(y_t - \mu_t) \cdot (y_{t-k} - \mu_{t-k})] = E(y_t \cdot y_{t-k}) - \mu_t \cdot \mu_{t-k} \text{ con } k = 0, \pm 1, \pm 2, \dots$$

Tale espressione misura la dipendenza tra il processo al tempo  $t$  ed il processo al tempo  $t-k$ . Essa si riduce alla varianza del processo per  $k=0$ .

Effettuare inferenza sulla legge di probabilità di un p.s. o sulle funzioni sopra rappresentate, partendo da una singola realizzazione può essere effettuata se sussistono delle proprietà della legge probabilistica del p.s. stesso.

Tali proprietà si concretizzano in una serie di restrizioni sulla legge di probabilità del processo ed hanno la finalità di ridurre il numero di parametri incogniti al fine di poter fare inferenza avendo a disposizione una singola realizzazione.

Tali restrizioni sono di due tipi:

1) Restrizioni sulla eterogeneità nel tempo del processo.

2) Restrizioni sulla memoria del processo.

Alla prima classe appartengono le condizioni che rendono un processo *stazionario*: si tratta del concetto di *stazionarietà* di un p.s. . Alla seconda quelle di *Indipendenza* e di *Ergodicità*.

Vediamo una veloce panoramica sulle due tipologie di restrizioni.

## RESTRIZIONI SULLA ETEROGENEITA' DEL PROCESSO

### STAZIONARIETA'

La proprietà di stazionarietà riguarda la invarianza della legge di probabilità del p.s. a traslazioni sull'asse dei tempi. Tale proprietà fa sì che la legge di probabilità del p.s., o solo alcuni dei suoi momenti finito dimensionali, siano gli stessi lungo l'asse dei tempi. Ciò permette, in termini inferenziali, di considerare il campione oggetto di analisi come informativo del processo che lo ha generato.

Si evidenziano due tipologie di stazionarietà:

- la stazionarietà del I ordine o forte o in distribuzione, la quale riguarda tutta la distribuzione di probabilità del processo.
- la stazionarietà del II ordine o debole o in covarianza, la quale riguarda i soli momenti primi (valore atteso) i momenti secondi misti ( autocovarianza) ed i momenti secondi (varianza).

Nello specifico, un p.s.  $\{y_{t \in T}\}$  si definisce *stazionario in senso forte* se

$$P(y_{t_1}, y_{t_2}, \dots, y_{t_n}) = P(y_{t_1+h}, y_{t_2+h}, \dots, y_{t_n+h}) \quad \forall \{t_1, t_2, \dots, t_n\} \text{ con}$$

$0 < n < \infty$  e  $\forall h > 0$ , dunque un processo sarà stazionario in senso forte se la sua distribuzione di probabilità sarà invariante a traslazioni sull'asse dei tempi.

Il processo sarà praticamente lo stesso in qualunque istante di tempo. Tale restrizione è assai forte ed è di scarsa utilità per la struttura che hanno empiricamente i processi che rappresentano i fenomeni reali. Allora si potrà far riferimento alla stazionarietà in senso debole. Un p.s.  $\{y_{t \in T}\}$  si definisce stazionario in senso debole se i momenti primi e secondi misti sono finiti e smettono di dipendere dal tempo.

Ovvero se risulterà che:

i)  $\mu_t = \mu$  con  $|\mu| < \infty$

ii)  $\gamma_t(k) = \gamma(k)$  con  $|\gamma(k)| < \infty$  e ciò sia valido  $\forall t, k$

Dunque, tale ultima tipologia di stazionarietà, pone vincoli solo sui momenti primi e secondi del processo e nulla dice sui momenti superiori al secondo, né sulla invarianza temporale del processo.

Per studiare la dipendenza temporale di un p.s. stazionario del II ordine, può essere assai utile l'uso della funzione di autocorrelazione, al fine di eliminare l'influenza del livello di variabilità del processo.

La funzione ha la seguente scrittura:

$$\rho(k) = \frac{\text{Cov}(y_t, y_{t-k})}{\sqrt{\text{Var}(y_t) \cdot \text{Var}(y_{t-k})}} = \frac{\gamma(k)}{\gamma(0)} \quad \text{con } \rho \in [0,1] \text{ e vale } 1, k=0$$

Non vi è una implicazione logica fissa tra le due stazionarietà per via dei diversi vincoli che esse possiedono. In generale si può dire che, se un p.s. è stazionario in distribuzione, allora lo sarà anche in covarianza solo se i momenti primi e secondi esistono finiti. Di contro, un p.s. stazionario in covarianza non implica anche la stazionarietà in distribuzione poiché nulla è imposto, in tal caso, sulla invarianza della distribuzione nel tempo. Infatti, in tal caso, si possono avere delle variazioni sui alcuni momenti superiori al secondo senza che tutti i momenti, se esistono, siano indipendenti dal tempo, vincolo questo necessario affinché un p.s. sia stazionario in distribuzione. Ciò non toglie che vi siano dei casi in cui le due tipologie di stazionarietà coincidano. Un esempio più noto è quello dei processi Gaussiani poiché, essendo tali processi definiti unicamente dal valore atteso, varianza ed autocovarianza, ed essendo tali quantità finite, allora se vale la stazionarietà in distribuzione è automaticamente presente anche la stazionarietà debole.

Di contro, se vale la stazionarietà debole, essendo il processo definito dai soli momenti di cui sopra, ed essendo essi indipendenti dal tempo, allora varrà anche la stazionarietà in senso forte in quanto sarà presente la indipendenza dal tempo di tutta la distribuzione del processo. Di seguito esporrò brevemente alcuni esempi di p.s. fondamentali poiché rappresentano i mattoni per creare dei processi più complessi.

Si considereranno processi a tempo discreto per semplicità, ma i risultati sono riferibili anche a processi a tempo continuo.

## PROCESSO IID

Un primo p.s. fondamentale è il processo IID. Un p.s.  $\varepsilon = \{\varepsilon_t\}_{t \in T}$  che costituisce un sequenza di variabili aleatorie indipendenti ed identicamente distribuite, con valore atteso nullo e varianza costante, è detto p.s. IID e lo si indica con

$IID(0, \sigma_\varepsilon^2)$ . Tale processo sarà stazionario del I ordine per via della proprietà di indipendenza ed identica distribuzione di esso ( tale proprietà verrà esplicitata successivamente). Inoltre, se  $\sigma_\varepsilon^2 < \infty$  il processo sarà stazionario del II ordine essendo il valore atteso sempre nullo, e per la indipendenza ed identica distribuzione la funzione di autocovarianza sarà costantemente nulla poiché,

$$\gamma(k) = E(\varepsilon_k \cdot \varepsilon_{t-k}) = E(\varepsilon_k) \cdot E(\varepsilon_{t-k}) = 0 \quad \text{per } k \neq 0 .$$

## PROCESSO WHITE NOISE

Un p.s.  $\varepsilon = \{\varepsilon_t\}_{t \in T}$  è detto white noise o rumore bianco se esso è costituito da una successione di variabili casuali in-correlate, con valore atteso nullo e varianza finita.

Data la definizione, la sua funzione di auto-covarianza è data da:

$$\gamma = E(\varepsilon_t \cdot \varepsilon_{t-k}) = \begin{cases} \sigma_\varepsilon^2, & k=0 \\ 0, & k \neq 0 \end{cases} \quad \text{con} \quad \sigma_\varepsilon^2 = E(\varepsilon_t^2)$$

Tale p.s. viene indicato con  $WN(0, \sigma_\varepsilon^2)$ .

Esso è più debole in termini di restrizioni poiché è più generale richiedendo solo la in-correlazione tra i suoi elementi e non la indipendenza e la identica distribuzione.

Ricopre una elevata importanza poiché è possibile, partendo da esso, esprimere un generico processo stazionario del II ordine come una particolare sua trasformazione lineare. Tale risultato è conosciuto come teorema di Wold per il quale, se  $y_t$  è un p.s. stazionario del II ordine, sarà sempre possibile ottenere un sua rappresentazione del tipo:

$y_t = \sum_{j=0}^{\infty} \psi_j \cdot \varepsilon_{t-j} + z_t$  con  $\psi_0 \equiv 1$ ,  $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$ ,  $\sum_{j=0}^{\infty} \psi_j^2 < \infty$  questa ultima detta 'condizione di regolarità'.

## PROCESSO RANDOM WALK

Sia un insieme di impulsi  $\{\varepsilon_1, \varepsilon_2, \varepsilon_3, \dots, \varepsilon_t\}$ , si definisce un p.s. *Random Walk (RW)* come  $S_t = \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t$ , avendo posto  $\varepsilon_t \sim IID(0, \sigma_\varepsilon^2)$ ,  $t=0, 1, 2, \dots$ , con la condizione iniziale  $S_0 = 0$ . Tale terminologia deriva dalla interpretazione fisica del moto Browniano poiché, se gli impulsi  $\varepsilon_t$  sono interpretati come l'ampiezza aleatoria dei movimenti di un soggetto vincolato a muoversi lungo una retta, allora  $S_t$  sarà la sua posizione aleatoria occupata al tempo  $t$ . E' possibile scrivere in maniera ricorsiva il processo come  $S_t = S_{t-1} + \varepsilon_t$ ,  $t=1, 2, \dots$ . Si verifica facilmente la non stazionarietà del processo poiché, il valore atteso è dato da  $E(S_t) = E(\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t) = E(\varepsilon_1) + E(\varepsilon_2) + \dots + E(\varepsilon_t) = 0$  dunque è costante e non dipende dal tempo. Per i momenti secondi invece si ha che  $\sigma_{S_t}^2 = E[(S_t - 0)^2] = \sigma_{\varepsilon_1}^2 + \sigma_{\varepsilon_2}^2 + \dots + \sigma_{\varepsilon_t}^2 = \sigma_\varepsilon^2 + \sigma_\varepsilon^2 + \dots + \sigma_\varepsilon^2 = t \cdot \sigma_\varepsilon^2$ . Dunque la varianza del processo dipende da  $t$  ed è esplosiva per  $t$  che diverge. Per la funzione di auto-covarianza si ha che:

$$\gamma(k) = E[(S_t - 0) \cdot (S_{t-k} - 0)] = E\left[\left(\sum_{i=1}^t \varepsilon_i\right) \cdot \left(\sum_{j=1}^{t-k} \varepsilon_j\right)\right] = \dots \stackrel{iid}{=} t \cdot \sigma_\varepsilon^2$$

con  $0 \leq t-k \leq t$  ottenendo una dipendenza temporale che conferma la non stazionarietà del processo. Tale tipologia di RW è detto *simmetrico*.

E' possibile introdurre una variazione per ottenere un c.d. *RW con drift* dato da

$$S_t = \mu + S_{t-1} + \varepsilon_t = \mu + (\mu + S_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = \dots = S_0 + t \cdot \mu + \sum_{j=1}^t \varepsilon_j .$$

In esso, la componente  $\sum_{j=1}^t \varepsilon_j$  del processo prende il nome di *trend stocastico* rappresentando la parte non deterministica e non stazionaria del processo.

La componente  $S_0 + t \cdot \mu$  è denominata *trend deterministico*. Per il calcolo dei

momenti si ha che  $E(S_0 + t \cdot \mu + \sum_{j=1}^t \varepsilon_j) = S_0 + t \cdot \mu$  che dipende sempre dal tempo.

Per la varianza si ha  $E [ S_0 + t \cdot \mu + \sum_{j=1}^t \varepsilon_j - (S_0 + t \cdot \mu) ]^2 = t \cdot \sigma_\varepsilon^2$

dipendente anch'essa da  $t$ . Come per la varianza, anche per la funzione di autocorrelazione, essendo il termine di deriva (drift) una costante, non influisce sui momenti secondi e ciò fa sì che la  $\gamma(k)$  sia la stessa al caso di *RW simmetrico* confermando la non stazionarietà del processo con *drift*.

## RESTRIZIONI SULLA MEMORIA DEL PROCESSO

La nozione di memoria di un processo è legata al livello di dipendenza tra il processo in un dato momento temporale e la propria storia passata. Tanto più il processo è influenzato da eventi manifestatisi nel passato, tanto più esso ha memoria lunga. Di contro, se gli eventi passati non influenzano il processo al tempo corrente, allora il processo non ha memoria; esso non ricorda la sua storia passata ed evolve in maniera indipendente da essa.

Per studiare un p.s. , attraverso parametrizzazioni di dimensioni finite, è necessario che esso abbia una memoria sufficientemente bassa. Una restrizione sulla memoria del processo può essere imposta considerando il concetto di *Indipendenza stocastica*.

## INDIPENDENZA ASINTOTICA

Un p.s.  $\{y_t\}_{t \in T}$  è detto *asintoticamente indipendente* se risulta che,

$$\lim_{h \rightarrow \infty} P(y_{t_1}, y_{t_2}, \dots, y_{t_n}, y_{t_1+h}, y_{t_2+h}, \dots, y_{t_n+h}) \dots$$

$$\dots - P(y_{t_1}, y_{t_2}, \dots, y_{t_n}) \cdot P(y_{t_1+h}, y_{t_2+h}, \dots, y_{t_n+h}) / = 0 \text{ e tale scrittura deve}$$

essere valida  $\forall (t_1, t_2, \dots, t_n) , 0 < n < \infty .$

Tale scrittura ci dice che, se un processo è asintoticamente indipendente, allora, la probabilità congiunta di eventi molto lontani nel tempo tra loro converge al prodotto delle singole probabilità. Tale risultato è legato quindi alla indipendenza stocastica tra eventi per la quale, dati due eventi  $A, B$  essi sono stocasticamente indipendenti  $\Leftrightarrow P(A \cap B) = P(A) \cdot P(B)$ .

In tal caso, il processo sarà anche asintoticamente incorrelato, ammesso che esistano finiti i momenti secondi. La asintotica incorrelazione ci dice che, se  $\sigma^2(y_t) < \infty$  per ogni istante di tempo, allora  $\gamma_t(k) \rightarrow 0$  per  $T$  divergente, con  $k \rightarrow \pm \infty$ . Alla base della indipendenza asintotica (e più in generale per ogni restrizione sulla memoria) vi è l'idea che, se il grado di dipendenza decresce col crescere del tempo, allora si potrà approssimare il processo in oggetto con un altro processo caratterizzato da un numero limitato di parametri.

Per processi debolmente stazionari sarà possibile far riferimento ad un'altra tipologia di restrizioni sulla memoria note come *condizioni di ergodicità*.

### ERGODICITA'

Dato un p.s.  $\{y_t\}_{t \in T}$  è detto *ergodico per il momento*  $E(g(y))$  se risulterà che  $\frac{1}{T} \cdot \sum_{t=1}^T g(y_t) \xrightarrow{p} E(g(y))$  dove  $\xrightarrow{p}$  indica la *convergenza in probabilità*.<sup>3</sup>

Dalla definizione data, ponendo  $g(y) = y$  si ha che, un p.s. è ergodico per il momento primo se la media campionaria converge in probabilità alla media del processo, ovvero se  $\frac{1}{T} \cdot \sum_{t=1}^T y_t \xrightarrow{p} E(y)$ .

Dunque, la ergodicità, è importante ai fini inferenziali sui momenti di interesse di un p.s. In genere, si verificano delle condizioni meno restrittive della in-correlazione asintotica sui p.s. che permettono di ammettere la ergodicità di esso, ed in particolare si dimostra che, un processo debolmente stazionario con funzione di auto-covarianza  $\gamma(k), k=0, \pm 1, \pm 2, \dots$  sarà ergodico per il valore atteso se risulterà che

$$\lim_{T \rightarrow \infty} \frac{1}{T} \cdot \sum_{k=-T}^T \gamma(k) = 0.$$

Tale condizione è detta *condizione di Slutsky* e non implica la incorrelazione asintotica ma vale il viceversa poiché, se  $\gamma(k)$  presenta un andamento oscillante si può verificare il caso in cui vale la condizione di Slutsky in presenza di autocorrelazione asintotica non nulla, invece se vi è in-correlazione asintotica, per come è strutturata la condizione di Slutsky essa varrà sempre.

3. Data un p.s.  $\{y_t\}_{t \in T}$  definito sullo spazio di probabilità  $(\Omega, \mathcal{F}, P)$ , si dice che esso converge in probabilità al p.s.  $y$  se,  $\forall \epsilon \in \mathbf{C}_0^{++} \mid \lim_{t \rightarrow \infty} P\{|y_t - y| < \epsilon\} = 1 \Leftrightarrow \lim_{t \rightarrow \infty} P\{|y_t - y| > \epsilon\} = 0$ .

La finalità di effettuare una analisi di un p.s. è quella di prevederne i futuri andamenti in modo tale da ottenere un vantaggio in termini pratici in riferimento al fenomeno reale che il p.s. vuole replicare.

Tale vantaggio consta nella possibilità di prevedere il futuro andamento del fenomeno oggetto di studio sfruttando le caratteristiche del processo approssimante.

A titolo di esemplificazione, si consideri un p.s.  $\{y_t\}_{t \in T}$  e si immagini che l'analista voglia effettuare una previsione, ad esempio, del valore che il processo assumerà al tempo  $T+1$  basandosi sulla traiettoria delle ultime  $n$  osservazioni. Il set informativo che viene utilizzato ai fini predittivi è dunque dato da

$$y_T, y_{T-1}, \dots, y_{T-n+1} .$$

Dunque, indicato il valore da predire, condizionato al set informativo a disposizione con  $\hat{y}_{T+1/T}$  si avrà che  $\hat{y}_{T+1/T} = g(y_T, y_{T-1}, \dots, y_{T-n+1})$  ovvero, il valore futuro sarà una funzione  $g(\cdot)$  dello stesso set informativo. Nell'effettuare tale previsione si commetterà un errore e dunque, la scelta della funzione sarà effettuata considerando quella per la quale, l'errore di previsione *ex-ante*, risulterà minore.

Se viene scelto come metodo di calcolo dell'errore, l'errore quadratico medio (*EQM*) allora sarà insita una associazione di perdita quadratica all'errore di previsione

$$EQM(\hat{y}_{T+1/T}) = E(y_{T+1} - \hat{y}_{T+1/T})^2 .$$

In tale senso, la funzione  $g(\cdot)$  che minimizza l'errore quadratico medio di previsione è il valore atteso condizionato, dunque si può scrivere che:

$$g(y_T, y_{T-1}, \dots, y_{T-n+1}) = E(y_{T+1} / y_T, y_{T-1}, \dots, y_{T-n+1})$$

e conseguentemente, il predittore a minimo *EQM* sarà dato da:

$$\hat{y}_{T+1/T} = E(y_{T+1} / y_T, y_{T-1}, \dots, y_{T-n+1},)$$

ottenendo l'errore di previsione associato che sarà pari a:

$$\eta_T = y_{T+1} - E(y_{T+1} / y_T, y_{T-1}, \dots, y_{T-n+1}) .$$

Importante è sottolineare che, ove il set informativo fosse rappresentato da tutta la storia passata del processo, ovvero  $I_T = (y_T, y_{T-1}, \dots, y_{T-n+1}, \dots)$  il termine

$\eta_T$  prende il nome di *innovazione*. Dunque si può concludere che, ogni p.s. può essere scomposto in due termini; un termine che rappresenterà il valore atteso condizionato ed in un altro che rappresenterà l'innovazione del processo. Si può scrivere che  $y_T = E(y_T / I_{T-1}) + \eta_T$  .

Si assume che  $\eta_T$  abbia valore atteso condizionato nullo poiché, l'innovazione, rappresenta la variazione che il p.s. assume per fattori casuali e dunque indipendenti dal set informativo disponibile.

Spesso, il valore atteso condizionato può essere un funzionale non semplice da trattare analiticamente e, per tale motivo, si opta per previsori di tipo *lineare*, ovvero rappresentabili tramite combinazione lineare del set informativo a disposizione.

Si avrà allora che,  $\hat{y}_{T+1} = \beta_0 + \beta_1 \cdot y_T + \beta_2 \cdot y_{T-1} + \dots + \beta_n \cdot y_{T-n}$  .

Usando una rappresentazione compatta, si indicherà  $\vec{\beta} = (\beta_0, \beta_1, \beta_2, \dots, \beta_n)$  il vettore riga  $n+1$  dimensionale di costanti, e con  $\vec{y}_T = (1, y_T, y_{T-1}, \dots, y_{T-n})'$  il vettore colonna  $n+1$  dimensionale del set informativo più una costante.

Dunque, in maniera compatta si può scrivere che  $\hat{y}_{T+1} = \vec{\beta} \cdot \vec{y}_T$ . In merito a quanto detto sopra, il fine adesso è quello di trovare un metodo che faccia sì che il vettore  $\vec{\beta}$  sia composto da termini che rendano l'*EQM* il più piccolo possibile. Attraverso la esplicitazione dei calcoli si può mostrare che, il vettore  $\vec{\beta}$  che consente di minimizzare l'*EQM* è quello per il quale è soddisfatta la seguente condizione di ortogonalità:  $E[(y_{T+1} - \vec{\beta} \cdot \vec{y}_T) \cdot \vec{y}_T'] = \vec{0}$ , ovvero è il vettore per il quale viene garantita l'ortogonalità<sup>4</sup> tra l'errore di previsione  $y_{T+1} - \vec{\beta} \cdot \vec{y}_T$  e l'insieme del set informativo  $\vec{y}_T$  sulle cui variabili (detti anche regressori) viene effettuata la previsione. Effettuando i calcoli algebrici (nella espressione della condizione di ortogonalità) per isolare il vettore  $\vec{\beta}$  si ottiene la formula per definire in vettore dei parametri  $\vec{\beta} = [E(\vec{y}_T \cdot \vec{y}_T')]^{-1} \cdot E(\vec{y}_T \cdot y_{T+1})$ . Tale espressione ammetterà soluzione se  $\exists [E(\vec{y}_T \cdot \vec{y}_T')]^{-1} < \infty \Leftrightarrow \exists (\vec{y}_T \cdot \vec{y}_T')^{-1} < \infty$  ovvero, se sarà possibile invertire il prodotto scalare tra i regressori.

L'equazione  $\vec{\beta} \cdot [E(\vec{y}_T \cdot \vec{y}_T')]^{-1} = E(\vec{y}_T \cdot y_{T+1})$  insieme alle equazioni che derivano dai calcoli effettuati imponendo la condizione di primo ordine per minimizzare l'*EQM* sono dette *equazioni normali*. Se i valori attesi di cui sopra sono ignoti, allora sarà possibile sostituirli con i dati campionari ottenendo lo stimatore dei minimi quadrati ordinari (*OLS*) per  $\vec{\beta}$  dato dalla seguente espressione:

$$\vec{\beta} = \left( \frac{1}{T-n+1} \cdot \sum_{t=n}^T \vec{y}_T \cdot \vec{y}_T' \right)^{-1} \cdot \left( \frac{1}{T-n+1} \cdot \sum_{t=n}^T \vec{y}_T \cdot y_{T+1} \right) \quad \text{ed il rispettivo previsore}$$

sarà pari a  $\vec{y}_{T+1}^* = \vec{\beta}' \cdot \vec{y}_T$ . Tale approccio non presuppone delle ipotesi di lavoro di nessun tipo se non la invertibilità della matrice del prodotto scalare tra i regressori. Dunque sarà sempre possibile trovare la migliore approssimazione procedendo algebricamente come visto sopra. Unica condizione resta allora la invertibilità della matrice  $(\vec{y}_T \cdot \vec{y}_T')^{-1}$ . Tale proprietà indica che nessuno dei regressori è superfluo, ovvero nessuno di esso è perfetta combinazione lineare delle altre. Si parla, in tal caso, di ipotesi di *assenza di multicollinearità perfetta*.

Se infine  $y_t$  sarà stazionario del II ordine ed ergodico per i momenti II, allora si avrà che  $\vec{\beta} \xrightarrow{p} \vec{\beta}$  e la previsione basata sullo stimatore *OLS* risulterà a minimo *EQM* tra la classe dei previsori lineari. Nella analisi dei p.s. è spesso utile rappresentare le serie storiche usate in maniera compatta. Questo viene effettuato

<sup>4</sup> Dati due vettori riga  $\vec{x}, \vec{y}$  essi sono detti *ortogonali* se il loro prodotto interno è nullo, ovvero se risulta che  $\vec{x} \cdot (\vec{y}) = 0$ .

facendo ricorso all'operatore ritardo. Tale operatore (funzione) può essere utilizzato per serie storiche a tempo discreto, ma in sostanza viene utilizzato anche per serie storiche a tempo continuo poiché sarà sempre possibile effettuare una discretizzazione della serie storica stessa. Dato un processo al tempo  $t$   $y_t$ , possiamo associare ad esso la variabile ritardata  $y_{t-1}$  sfruttando l'operatore ritardo indicato con il simbolo  $L$  anteposto ad  $y_t$ , ottenendo  $y_{t-1} = Ly_t$ . Si usa il simbolo  $L$  poiché tale operatore ritardo è chiamato in lingua inglese *Lag* oppure *backward shift operator*. Si possono osservare alcune proprietà naturali, per via della stessa definizione, dell'operatore ritardo.

La prima riguarda la possibilità di applicare l'operatore  $L$  diverse volte. Dunque, si potrà scrivere che  $y_{t-1} = Ly_t$ , ma applicando nuovamente l'operatore  $L$  alla scrittura si avrà che  $y_{t-2} = L(Ly_t)$ . Potendo far ciò senza limiti alcuni, allora si può affermare che,  $y_{t-k} = \underbrace{L L (\dots) L}_{k\text{-volte}} L y_t$  ottenendo una traslazione indietro di

$y_t$  di  $k$ -volte che, compattamente, equivale a  $y_{t-k} = L^k y_t$ , con  $k \in \mathbb{Z}_{+0}^*$  e con la convenzione che  $y_t = L^0 y_t$ , ovvero  $L^0 \equiv 1$  equivale al *lag* nullo.

Si noti che si è posto  $k \in \mathbb{Z}_{+0}^*$  ma ponendo un segno negativo di fronte a  $k$  si avrà che l'operatore  $L$  assumerà il seguente significato,

$y_{t+k} = L^{-k} y_t$  per  $k \in \mathbb{Z}_{-0}^*$ . In tal caso si scriverà che  $L^{-k} = F^k$  ottenendo una traslazione di  $y_t$  in avanti di  $k$ -volte.

Si parlerà dell'operatore  $F$  come l'operatore *lead* oppure *forward shift operator*.

La seconda proprietà importante di  $L$  (stesso dicasi per  $F$  vista la sua costruzione) è quella di *linearità*.

Dati due p.s.  $y_t, x_t$  e date due costanti qualunque  $a, b \in \mathbb{C}_0$  varrà la seguente scrittura  $L(a \cdot y_t + b \cdot x_t) = L(a \cdot y_t) + L(b \cdot x_t) = a \cdot Ly_t + b \cdot Lx_t = a \cdot y_{t-1} + b \cdot x_{t-1}$  e più in generale sarà  $L^k(a \cdot y_t + b \cdot x_t) = a \cdot L^k y_t + b \cdot L^k x_t = a \cdot y_{t-k} + b \cdot x_{t-k}$  e stesso procedimento varrà per l'operatore *lead*  $F$ .

Tale ultima proprietà può essere sfruttata attraverso l'utilizzo di polinomi di grado  $k$  dell'operatore ritardo così definiti:

$A(L) = 1 + a_1 L + a_2 L^2 + \dots + a_k L^k = \sum_{j=0}^k a_j \cdot L^j$ ,  $a_0 \equiv 1$  i quali associano una trasformata di una serie  $y_t$  del tipo:

$A(L) y_t = a_0 \cdot y_t + a_1 \cdot y_{t-1} + a_2 \cdot y_{t-2} + \dots + a_k \cdot y_{t-k} = \sum_{j=0}^k a_j \cdot y_{t-j}$  che diviene una serie in  $L$ , se esiste, del tipo  $\sum_{j=0}^{\infty} a_j L^j$  quando  $k \rightarrow \infty$ .

La comodità dell'uso dei polinomi ritardo sta nel fatto che essi, sotto alcune condizioni, risultano invertibili, ovvero sarà possibile trovare un polinomio

$$A(L)^{-1} \ni A(L)^{-1} \cdot A(L) = 1 .$$

Si riporta un piccolo esempio dimostrativo di quanto espresso: sia un caso semplice in cui  $k=1$  e dunque risulti che  $A(L) = 1 + a_1 L$  .

Si ponga  $a_1 = -a$  ottenendo  $A(L) = 1 - aL$  senza perdere in generalità.

Per analogia con una serie geometrica di ragione  $|a| < 1$  si ha che:

$$1 + a + a^2 + \dots = \sum_{j=0}^{\infty} a^j = \frac{1}{1-a} \text{ potendo così definire l'inversa del polinomio in } L$$

$$\text{come } A(L)^{-1} = 1 + aL + a^2 L^2 + \dots = \sum_{j=0}^{\infty} a^j L^j = \frac{1}{1-aL}$$

moltiplicando rispettivamente  $A(L)$  ed  $A(L)^{-1}$  così definiti si ottiene che,

$$\begin{aligned} A(L) \cdot A(L)^{-1} &= (1 - aL) \cdot \sum_{j=0}^{\infty} a^j L^j = \sum_{j=0}^{\infty} a^j L^j - aL \cdot \sum_{j=0}^{\infty} a^j L^j = \sum_{j=0}^{\infty} a^j L^j - \dots \\ &- \sum_{j=0}^{\infty} a^{j+1} L^{j+1} = \sum_{j=0}^{\infty} a^j L^j - \sum_{j=1}^{\infty} a^j L^j = a^0 L^0 + \left( \sum_{j=1}^{\infty} a^j L^j \right) - \sum_{j=1}^{\infty} a^j L^j = a^0 L^0 \equiv 1 . \end{aligned}$$

Tale procedura potrà essere applicata ad un qualunque polinomio di grado  $k > 1$  .

Infatti, per la possibilità di fattorizzare un polinomio di qualunque grado  $k$  nel seguente modo:

$$A(L) = 1 + a_1 L + a_2 L^2 + \dots + a_k L^k = \left(1 - \frac{1}{z_1} L\right) \cdot \left(1 - \frac{1}{z_2} L\right) \dots \left(1 - \frac{1}{z_k} L\right) = \prod_{j=1}^k \left(1 - \frac{1}{z_j} L\right)$$

dove  $z_1, z_2, \dots, z_k$  sono le soluzioni reali e complesse della *equazione caratteristica*

$$A(z) = 1 + a_1 \cdot z + a_2 \cdot z^2 + \dots + a_k \cdot z^k = 0 , \text{ sarà possibile applicare la scrittura}$$

$$A(L)^{-1} = 1 + aL + a^2 L^2 + \dots = \sum_{j=0}^{\infty} a^j L^j = \frac{1}{1-aL} \text{ ad ogni quantità } \left(1 - \frac{1}{z_j} L\right) .$$

Avendo specificato prima, che deve valere la condizione di invertibilità  $|a| < 1$  , per analogia, affinché sia applicabile questa procedura dovrà risultare che:

$$\left| \frac{1}{z_j} \right| < 1 \Leftrightarrow |z_j| > 1 \text{ per ogni valore di } j .$$

Tale ultima condizione ci dice che esisterà l'inversa del polinomio in  $L$  solo se le soluzioni della equazione caratteristica risulteranno in valore assoluto maggiori di 1. Avendo specificato che tali soluzioni possono essere reali e complesse, allora in tal caso, la condizione di invertibilità ci dice che le radici  $z_j$  dovranno risultare esterne al cerchio di raggio unitario sul piano complesso.

Tale condizione è fondamentale per la modellizzazione di molti processi stocastici utilizzati nelle analisi. In molti ambiti di studio come nelle scienze sociali, l'analista si trova nella impossibilità di replicare la situazione oggetto di analisi. In tal modo si dovranno creare delle ipotesi sul processo matematico che ha generato i dati, sulla

base di una sola traiettoria. Nel caso di stazionarietà in senso debole e di ergodicità della traiettoria a disposizione, sarà più agevole svolgere tale compito.

Si possono definire delle misure importanti utili per effettuare delle analisi sul processo generatore dei dati partendo da una traiettoria tra le infinite che esso può generare. Sia allora una traiettoria  $y_1, y_2, \dots, y_T$  di  $T$  elementi generati da un processo generatore dei dati ignoto. Si ipotizzi si voglia avere una stima del *valore atteso*  $\mu = E(y_t)$  e della funzione di *varianza* ed *autocovarianza*  $\gamma(k), k \geq 0$  del p.s. stazionario che ha generato la traiettoria a disposizione.

Tale procedura è un esempio classico di inferenza su alcuni parametri del processo generatore dei dati sulla base una sua realizzazione campionaria.

Uno stimatore utilizzabile per il *valore atteso* del p.s. potrà essere la *media campionaria* ottenendo che  $\hat{\mu} = T^{-1} \cdot \sum_{t=1}^T y_t$ .

Per la stima della *varianza* e della *autocovarianza*  $\gamma(k)$  ignote del p.s. si potrà far ricorso rispettivamente alla funzione di *varianza* ed *autocovarianza campionarie* così rispettivamente definite:

$$\hat{\sigma}^2 = T^{-1} \cdot \sum_{t=1}^T (y_t - \hat{\mu})^2 = T^{-1} \cdot \sum_{t=1}^T y_t^2 - \hat{\mu}^2$$

$$\hat{\gamma}(k) = (T-k)^{-1} \cdot \sum_{t=k+1}^T (y_t - \hat{\mu}_0) \cdot (y_{t-k} - \hat{\mu}_{-k}) \quad \text{dove } \hat{\mu}_0, \hat{\mu}_{-k} \text{ sono pari ad}$$

$$\hat{\mu}_0 = (T-k)^{-1} \cdot \sum_{t=k+1}^T y_t \quad \text{e} \quad \hat{\mu}_{-k} = (T-k)^{-1} \cdot \sum_{t=k+1}^T y_{t-k} = (T-k)^{-1} \cdot \sum_{t=1}^{T-k} y_t, \text{ ma}$$

nel caso di p.s. stazionari, essendo il valore atteso indipendente dal tempo, allora la funzione di *autocovarianza campionaria* potrà essere scritta nel seguente modo

$$\hat{\gamma}(k) = T^{-1} \cdot \sum_{t=k+1}^T (y_t - \hat{\mu}) \cdot (y_{t-k} - \hat{\mu}) \quad \text{con } k=1, \dots, T \text{ ottenendo la funzione}$$

*varianza campionaria* per  $k=0$ .

Effettuando le precedenti considerazioni sulla autocovarianza campionaria per processi stazionari, si può definire la *funzione di autocorrelazione* campionaria come

$$\hat{\rho}(k) = \frac{\sum_{t=k+1}^T (y_t - \hat{\mu}) \cdot (y_{t-k} - \hat{\mu})}{\sqrt{\sum_{t=1}^T (y_t - \hat{\mu})^2 \cdot (y_t - \hat{\mu})^2}} \in [0, 1] \quad \forall k \in \mathbb{N}.$$

Si specifica che, la proprietà di stazionarietà non garantisce che i momenti campionari  $\hat{\mu}; \hat{\gamma}(0), \hat{\gamma}(1), \dots; \hat{\rho}(1), \hat{\rho}(2), \dots$  siano buoni stimatori dei corrispondenti momenti teorici, ma la presenza della proprietà di ergodicità garantisce che essi ne siano degli stimatori consistenti.

In altri termini, all'aumentare della cardinalità campionaria, i momenti campionari tendono ad assumere il valore incognito dei momenti teorici chiamati a stimare.

Dalla definizione di stazionarietà e di ergodicità si può affermare che, se il processo è stazionario ed ergodico per il valore atteso, per  $T \rightarrow \infty$  si avrà che  $\hat{\mu} \xrightarrow{p} \mu$ , ed inoltre se il processo è stazionario ed ergodico per i momenti secondi si avrà che per

$$T \rightarrow \infty \Rightarrow \hat{\gamma}(k) \xrightarrow{p} \gamma(k) \quad \forall k \in \mathbb{N} \quad \text{ma questo implica che } \hat{\rho}(k) \rightarrow \rho(k) \quad \forall k \in \mathbb{N} .$$

Questi concetti sin qui esplicitati sono la base per effettuare una analisi statistica di dati empirici. Nel successivo paragrafo si discuteranno alcuni processi notevoli per la specificazione di modelli uni-variati per traiettorie.

### 3.2

#### Alcuni modelli uni-variati per serie storiche

Partendo da una descrizione probabilistica di un p.s. e dalla analisi statistica delle funzioni principali che lo caratterizzano (funzione di autocorrelazione in primis) si può effettuare il lavoro inferenziale su di esso, per avere una visione quanto più verosimile sui meccanismi dinamici che hanno generato i dati.

Di seguito verranno trattati alcuni p.s. utili per la corretta specificazione di modelli uni-variati (estendibili anche ai multi-variati) nei casi di presenza delle proprietà di stazionarietà ed ergodicità. Un ruolo importante in tal senso è ricoperto dai processi stazionari *ARMA* (Auto Regressive Moving Average) definiti attraverso equazioni lineari dell'operatore ritardo. Caratteristica importante che rendono tali processi utili nelle analisi è che, ogni processo stazionario in covarianza può essere approssimato attraverso un dato modello *ARMA*, tale affermazione dunque fa sì che, per ogni funzione di autocovarianza  $\gamma(k) \ni \lim_{k \rightarrow \infty} \gamma(k) = 0$  esisterà un processo *ARMA* <sub>$\gamma$</sub>

dotato di funzione di autocovarianza  $\gamma_{\gamma}(k) \ni \gamma_{\gamma}(k) = 0$  e ciò sarà valido  $\forall k \in \mathbb{Z}$ . La struttura di tali processi è rappresentata da una particolare combinazione lineare della variabile oggetto di studio e di una componente stocastica, sia dei valori presenti che dei ritardi. Si può esprimere un siffatto processo utilizzando anche solo il valore corrente ed i suoi ritardi o della sola variabile in oggetto, o della sola componente stocastica.

Nel caso in cui si faccia riferimento alle relazioni esistenti tra la variabile ed una componente stocastica ed ai suoi ritardi si avrà il caso dei processi c.d. *M.A.* (Moving Average) o a media mobile. Dunque, in tal caso, dato un p.s.  $\{y_t\}_{t \in T}$  il caso più semplice di processo *M.A.* è dato dalla seguente relazione,  $y_t = \varepsilon_t + \theta_1 \cdot \varepsilon_{t-1}$  detto processo *MA* di ordine 1 (*MA*(1)).

Se le relazioni tra la variabile in oggetto e la componente stocastica sono legate alla storia passata della variabile si avrà il caso dei processi c.d. *A.R.* (Auto Regressive)

ed in tal caso si avrà la seguente relazione  $y_t = \phi_1 \cdot y_{t-1} + \varepsilon_t$  che rappresenta il caso più semplice di processi A.R. detto A.R. di ordine 1 (AR(1)).

Si parla infine di processi misti se si unisce un M.A. ed un A.R. ottenendo un modello A.R.M.A. in cui, il processo al tempo  $t$  è funzione dei valori passati del processo e dei valori presenti e passati della componente stocastica. L'esempio più semplice si ha unendo un AR(1) ed un MA(1) ottenendo un ARMA(1,1) che è rappresentato dalla seguente relazione  $y_t = \phi_1 \cdot y_{t-1} + \theta_1 \cdot \varepsilon_{t-1} + \varepsilon_t$ .

In tali modelli, l'unica parte stocastica è rappresentata dalla variabile  $\varepsilon_t$  che, per collocare tali processi nella classe di quelli stazionari in covarianza, è sufficiente porli delle restrizioni deboli alla sua struttura, ovvero si considera che

$$\varepsilon_t \sim WN(0, \sigma_\varepsilon^2) .$$

Se si aggiunge la proprietà di *Martingale* la classe dei processi ARMA può essere vista come un modello per il valore atteso di  $y$ , condizionatamente alla storia passata<sup>5</sup>. Più in generale, si possono considerare questi modelli considerando un numero di ordine superiore ad 1.

In particolare, per i modelli MA, si consideri un p.s.  $\{y_t\}_{t \in T}$  ottenuto come funzione di un  $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$  ottenendo dunque  $y_t = f(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots)$ .

Considerando il funzionale  $f$  lineare nei primi  $q$  ritardi di  $\varepsilon_t$  si ha che, un p.s. a *Media Mobile* di ordine  $q$ , (MA( $q$ )), è tale se soddisfa la seguente equazione alle differenze  $y_t = \varepsilon_t + \theta_1 \cdot \varepsilon_{t-1} + \theta_2 \cdot \varepsilon_{t-2} + \dots + \theta_q \cdot \varepsilon_{t-q}$  dove  $q=1,2,\dots,k$  e  $k \in \mathbb{N}$ .

Sfruttando l'operatore ritardo si può rappresentare un MA( $q$ ) nel seguente modo

$$y_t = \Theta(L)\varepsilon_t \text{ con } \varepsilon_t \sim WN(0, \sigma_\varepsilon^2) \text{ avendo posto il polinomio dell'operatore}$$

$$\text{ritardo } \Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q = 1 + \sum_{j=1}^q \theta_j L^j .$$

Dalle premesse fatte si può dire che un MA( $q$ ) è stazionario in covarianza poiché  $E(y_t) = E(\varepsilon_t + \theta_1 \cdot \varepsilon_{t-1} + \theta_2 \cdot \varepsilon_{t-2} + \dots + \theta_q \cdot \varepsilon_{t-q}) = 0$  dunque il valore atteso non dipende dal tempo; per la funzione di autocovarianza si ha che

$$\gamma(k) = E(y_t \cdot y_{t-k}) = E\left[\varepsilon_t \cdot \left(\sum_{j=0}^q \theta_j L^j\right) \cdot \left(\sum_{j=0}^q \theta_{j-k} L^{j-k}\right)\right] = \begin{cases} \sigma_\varepsilon^2 \cdot \sum_{j=0}^{q-k} \theta_j \cdot \theta_{j+k}, & k \leq q \\ 0, & k > q \end{cases}$$

avendo posto  $\theta_0 \equiv 1, L^0 \equiv 1$ . Se  $k=0$  si ha la varianza del processo

$$\sigma_y^2 = \gamma(0) = \sigma_\varepsilon^2 \cdot (\theta_0^2 + \theta_1^2 + \dots + \theta_q^2) = \sigma_\varepsilon^2 \cdot \left(1 + \sum_{j=1}^q \theta_j^2\right) .$$

5 Dato un p.s.  $\{\varepsilon_t\}_{t \in T}$  e dato il set informativo al tempo  $t$   $I_{t-1} = (\varepsilon_{t-1}, \varepsilon_{t-2}, \dots)$  se valgono le condizioni  $E(\varepsilon_t / I_{t-1}) = 0$ ,  $E(\varepsilon_t^2 / I_{t-1}) < \infty$  il p.s. è detto *Differenza di Martingala* (MDS). Inoltre, se nel set informativo vi è un ulteriore insieme, diciamo  $X$ , di informazioni allora il p.s. è detto *processo MDS condizionatamente ad X*. Questo implica che, un MDS è un processo il cui valore atteso, condizionato ad un set informativo, risulta nullo. Dunque il set informativo a disposizione non è utile per prevedere il valori futuri del processo stesso.

La funzione di autocorrelazione assumerà dunque la seguente forma

$$\rho(k) = \begin{cases} 1, & k=0 \\ \frac{\sum_{j=0}^{q-k} \theta_j \cdot \theta_{j+k}}{(1 + \sum_{j=1}^q \theta_j^2)}, & 0 < k \leq q \\ 0, & k > q \end{cases}$$

col processo che risulterà correlato sino al ritardo  $q$ .

Si evince dunque che, esistono il valore atteso e la funzione di autocovarianza ed inoltre non dipendono dal tempo ma dal solo la  $k$ , confermando la proprietà di stazionarietà. Prima di portare un semplice esempio, è utile specificare delle proprietà fondamentali della funzione di auto-covarianza.

La prima è stata già evidenziata sopra e riguarda il fatto che la funzione di autocovarianza coincide con la varianza del processo se  $k=0$ .

La seconda riguarda la simmetria di essa e discende dalla stessa sua definizione, dunque si ha che  $\gamma(k) = \gamma(-k)$  qualunque sia  $k$  tra i valori ammissibili.

Terza ed ultima proprietà deriva dal contesto di uno spazio vettoriale ed è legata al prodotto scalare tra vettori<sup>6</sup>, dove nel nostro caso, i vettori sono rimpiazzati dalla funzione di autocovarianza. Tale proprietà prende il nome di *disuguaglianza di Chauchy-Schwartz* e può essere così esplicitata,

$|\gamma(k)| \leq \sqrt{\gamma(t) \cdot \gamma(s)}$  dove  $t$  ed  $s$  sono due istanti temporali diversi per i quali vale che  $s - t = k, s > t$ .

Si riporta di seguito un esempio di *p.s.* costruito a partire da un *WN* applicando un semplice filtro lineare come una media aritmetica:

Sia  $\{y_t\}_{t \in T}$  un *p.s.* allora esso sarà un *MA(1)* se, ad esempio, vale che

$$y_t = \frac{1}{2} \cdot \varepsilon_t + \frac{1}{2} \cdot \varepsilon_{t-1} \quad \text{con} \quad \varepsilon \sim WN(0, \sigma_\varepsilon^2)$$

Considerando le proprietà di un White

Noise è agevole calcolare i momenti per verificare la proprietà di stazionarietà.

$$\text{Si ha che} \quad E(y_t) = E\left(\frac{1}{2} \cdot \varepsilon_t + \frac{1}{2} \cdot \varepsilon_{t-1}\right) = E\left(\frac{1}{2} \cdot \varepsilon_t\right) + E\left(\frac{1}{2} \cdot \varepsilon_{t-1}\right) = 0$$

$$\gamma(0) = \sigma_{y_t}^2 = E[(y_t - 0) \cdot (y_{t-0} - 0)] \stackrel{inc}{=} E(y_t^2) = E\left(\frac{1}{2} \cdot \varepsilon_t + \frac{1}{2} \cdot \varepsilon_{t-0}\right)^2 = \frac{1}{2} \cdot \sigma_\varepsilon^2$$

$$\gamma(k) = E[(y_t - 0) \cdot (y_{t-k} - 0)] = E\left[\left(\frac{1}{2} \cdot \varepsilon_t + \frac{1}{2} \cdot \varepsilon_{t-1}\right) \cdot \left(\frac{1}{2} \cdot \varepsilon_{t-1} + \frac{1}{2} \cdot \varepsilon_{t-2}\right)\right] \stackrel{inc}{=} \frac{1}{4} \cdot \sigma_\varepsilon^2$$

---

6 Dati due vettori  $\vec{v}, \vec{w}$  appartenenti ad uno spazio vettoriale, si definisce *prodotto scalare* tra i due vettori l'operazione  $\langle \vec{v}, \vec{w} \rangle = |\vec{v}| \cdot |\vec{w}| \cdot \cos(\theta)$  dove  $|\cdot|$  rappresenta il modulo di ogni vettore, e  $\theta$  l'angolo tra essi compresi. Il risultato di tale operazione tra vettori sarà uno scalare. Più in generale, la disuguaglianza di Cauchy-Schwartz in uno spazio vettoriale assume la forma  $\langle \vec{v}, \vec{w} \rangle \leq |\vec{v}| \cdot |\vec{w}|$ .

essendo  $k=1$ . Con analoghi passaggi si mostra che  $\gamma(k)=0, k>1$  ; allora si conclude che la funzione di autocovarianza assume i seguenti valori,

$$\gamma(k) = \begin{cases} \frac{1}{2} \cdot \sigma_\varepsilon^2, & k=0 \\ \frac{1}{4} \cdot \sigma_\varepsilon^2, & k=1 \\ 0, & k>1 \end{cases} \quad \text{ovvero il p.s. definito tramite un filtro lineare di un WN è}$$

stazionario in covarianza. Con semplici passaggi si può mostrare che la funzione di autocorrelazione assume i seguenti valori,

$$\rho(k) = \begin{cases} 1, & k=0 \\ \frac{1}{2}, & k=1 \\ 0, & k>1 \end{cases} .$$

Tale semplice esempio mostra come, partendo da un processo con una struttura che lo rende privo di memoria, applicando un banale filtro lineare come la media aritmetica, si costruisce un nuovo processo che possiede una struttura di dipendenza seriale, anche se si limita al solo *lag*1. Ma si potrebbe avere lo stesso risultato complicando un po' la situazione. Tale motivo è alla base dell'importanza che tale processo (WN) definito nelle pagine precedenti assume nello studio delle traiettorie di fenomeni reali.

Utilizzando un filtro lineare causale infinito di un WN, si ottiene la naturale generalizzazione di un  $MA(q)$  la quale generalizzazione prende il nome di *Processo lineare generale* ed è indicato con  $MA(\infty)$ . Dunque, dato un p.s. stazionario

$\{y_t\}_{t \in T}$ , esso è detto  $MA(\infty)$  se soddisfa la equazione alle differenze

$$y_t = \varepsilon_t + \psi_1 \cdot \varepsilon_{t-1} + \dots + \psi_q \cdot \varepsilon_{t-q} + \dots = \sum_{j=0}^{\infty} \psi_j \cdot \varepsilon_{t-j} \quad \text{con le condizioni} \quad \psi_0 \equiv 1 \quad \text{ed}$$

$|\psi_j| < \infty, j=1,2,\dots$ . Tale specificazione, avendo in se una serie infinita, non ha un interesse diretto per l'analisi dei dati ma è solo funzionale alla rappresentazione in maniera unitaria di diversi processi lineari a memoria finita.

Uguualmente, però, si deve dare un senso alla serie infinita di variabili aleatorie che compare in essa. Si può, brevemente dire che, l'ambiente di lavoro è costituito da variabili aleatorie con varianza finita dette anche variabili aleatorie *quadrato - integrabili*, ed avendo un WN varianza finita, esso appartiene a tale spazio. Tale spazio rappresenta uno *spazio di Hilbert* e viene indicato col simbolo  $L_2$ . Tale spazio è un oggetto matematico in cui valgono gli stessi risultati della analisi ordinaria, con la considerazione di alcune proprietà precipue se si considerano come suoi elementi delle variabili aleatorie. Allora, se  $\{y_t\}_{t \in T}$  è una successione di variabili aleatorie (o equivalentemente di p.s.) nello spazio  $L_2$ , si dice che la successione converge alla variabile aleatoria limite  $y$  se converge in media quadratica

ad essa, ovvero se  $E(y_t - y)^2 \xrightarrow{t \rightarrow \infty} 0$  che equivalentemente può essere scritto come  $y_t \xrightarrow{qm} y$ .

Considerando la rappresentazione di un  $MA(\infty)$ , tale convergenza alla variabile limite del caso  $y_t$  si ottiene se viene rispettata la condizione di *assoluta sommabilità*

$\sum_{j=0}^{\infty} |\psi_j| < \infty$  ottenendo che  $\sum_{j=0}^{\infty} \psi_j \cdot \varepsilon_{t-j} \xrightarrow{qm} y_t$ . Tale condizione implica che il  $MA(\infty)$  sia *stazionario ed ergodico* per il valore atteso, ed inoltre permette di calcolare agevolmente i momenti di interesse del processo poiché è possibile scambiare il segno di valore atteso (Integrale) con quello di una serie.

Compattamente, un  $MA(\infty)$  può essere rappresentato come  $y_t = \Psi(L) \cdot \varepsilon_t$  dove  $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$  e  $\Psi(L) = \sum_{j=0}^{\infty} \psi_j \cdot L^j$ .

I momenti primi e secondi sono calcolabili nel seguente modo:

$E(y_t) = E(\varepsilon_t \cdot \sum_{j=0}^{\infty} \psi_j \cdot L^j) = \dots = 0$  dunque, il valore atteso risulta indipendente dal tempo. Inoltre, essendo la funzione di autocorrelazione

$\gamma(k) = E[(y_t - 0) \cdot (y_{t-k} - 0)] = E(y_t \cdot y_{t-k}) = E[(\sum_{j=0}^{\infty} \psi_j \cdot \varepsilon_{t-j}) \cdot (\sum_{j=0}^{\infty} \psi_j \cdot \varepsilon_{t-k-j})] = \dots$   
 $\dots = \sum_{j=0}^{\infty} \sum_{s=0}^{\infty} \psi_j \cdot \psi_s \cdot E(\varepsilon_{t-j} \cdot \varepsilon_{t-k-s}) = \dots = \sigma_\varepsilon^2 \cdot \sum_{j=0}^{\infty} \psi_j \cdot \psi_{j+k}$ , allora, anch'essa risulta indipendente dal tempo e dipendente dal solo lag  $k$ .

Se  $k=0$  si ottiene la varianza del processo data da  $\sigma_\varepsilon^2 \cdot \sum_{j=0}^{\infty} \psi_j^2 < \infty$  poiché si può dimostrare che, dalla condizione di assoluta sommabilità scaturisce la convergenza di  $\sum_{j=0}^{\infty} \psi_j^2$ . Sempre da tale condizione scaturisce la proprietà della funzione di auto-

covarianza che dice che  $\sum_{k=0}^{\infty} |\gamma(k)| < \infty$  e dunque il processo risulterà ergodico per

il momento primo. Affianco ai modelli  $MA(q)$ , della classe dei modelli  $ARMA$ , vi sono i modelli  $AR(p)$  detti *Modelli Autoregressivi di ordine  $p$* .

A differenza dei modelli a media mobile, i modelli  $AR(p)$  hanno una struttura di dipendenza seriale tra i livelli differenti poiché, i valori correnti della traiettoria assumono le vesti di variabile risposta di un modello lineare nel quale i regressori sono rappresentati dai valori ritardati della traiettoria stessa osservata (la serie storica). Vi è dunque una relazione tra, il valore del processo al tempo attuale ed i suoi più recenti valori passati.

Questo fa sì che le correlazioni lineari siano introdotte attraverso la medesima struttura del modello.

Si introducono tali classe di modelli dicendo che, dato un p.s.  $\{y_t\}_{t \in T}$  stazionario, un modello  $AR(p)$  è dato dalla esistenza di un funzionale lineare tale che

$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p}, \varepsilon_t)$ , ovvero il p.s. dipende (a meno di un termine di disturbo dato dal WN) dalla sua storia passata. Essendo il funzionale lineare, allora si può definire un  $AR(p)$  dicendo che, un processo  $\{y_t\}_{t \in T}$  stazionario è detto *Autoregressivo di ordine p (AR(p))* se soddisfa la seguente equazione alle differenze

$$y_t = \phi_1 \cdot y_{t-1} + \phi_2 \cdot y_{t-2} + \dots + \phi_p \cdot y_{t-p} + \varepsilon_t .$$

Anche per questi modelli è possibile adottare una scrittura compatta attraverso l'operatore ritardo ottenendo la seguente equazione stocastica alle differenze

$$\Phi(L)y_t = \varepsilon_t \quad \text{con} \quad \Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p = 1 - \sum_{j=1}^p \phi_j L^j \quad \text{e dove si ha}$$

che  $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$ .

Per i modelli  $AR(p)$  vi è, a differenza dei modelli  $MA(q)$ , un vincolo sui parametri affinché il modello risulti stazionario. Precisamente, un  $AR(p)$  sarà stazionario del II ordine se tutte le soluzioni della equazione caratteristica  $\Phi(z) = 0$  del polinomio dell'operatore ritardo, saranno esterne al cerchio unitario sul piano complesso.

Solo in tal caso esisterà un processo  $AR(p)$  stazionario che soddisferà la equazione alle differenze che rappresenta il processo stesso e lo si potrà esplicitare come un  $MA(\infty)$  nel seguente modo:

$$\Phi(L)y_t = \varepsilon_t \Rightarrow y_t = \Phi(L)^{-1} \cdot \varepsilon_t = \frac{\varepsilon_t}{1 - \phi_1 L - \dots - \phi_p L^p} = \dots = \sum_{j=0}^{\infty} \psi_j \cdot \varepsilon_{t-j} = \Psi(L) \cdot \varepsilon_t$$

avendo posto  $\Phi(L)^{-1} = \sum_{j=0}^{\infty} \psi_j \cdot \varepsilon_{t-j} = \Psi(L) \cdot \varepsilon_t$  con  $\psi_j = \phi^j, j=0,1,2,\dots$  in

coerenza con la scrittura di un  $MA(\infty)$  di cui sopra.

Per mostrare la correttezza di tale rappresentazione si consideri un modello autoregressivo minimale, ovvero con  $p=1$ . Si avrà allora che il processo sarà del tipo

$y_t = \phi_1 \cdot y_{t-1} + \varepsilon_t$  in tal caso, nella rappresentazione di cui sopra di un autoregressivo di ordine  $p$  si porrà  $\Phi(L) = 1 - \phi_1 L$  ottenendo la corrispondente

equazione caratteristica  $1 - \phi_1 \cdot z = 0$  che ammette come soluzione  $z = \frac{1}{\phi_1}$ .

Per la stazionarietà del processo, tale soluzione deve risultare in valore assoluto maggiore di uno, e ciò sarà possibile solo se  $|\phi_1| < 1$ . Dunque si potrà scrivere

che  $y_t = \phi_1 \cdot y_{t-1} + \varepsilon_t \Rightarrow y_t - \phi_1 \cdot y_{t-1} = \varepsilon_t \Rightarrow y_t = \frac{1}{1 - \phi_1 L} \cdot \varepsilon_t$ .

Sfruttando la rappresentazione di una serie geometrica di ragione minore di 1, si potrà

dire che  $y_t = \frac{1}{1 - \phi_1 L} \cdot \varepsilon_t = (1 + \phi_1 L + \phi_1^2 L^2 + \dots) \cdot \varepsilon_t = \left( \sum_{j=0}^{\infty} \phi_1^j L^j \right) \cdot \varepsilon_t$  essendo

$|\phi_1| < 1$  affinché il processo risulti stazionario. La condizione di assoluta

sommabilità vista in precedenza per i modelli  $MA(\infty)$  si concretizza in questo caso nel vincolo che  $\sum_{j=0}^{\infty} |\phi_1|^j < \infty$ .

Questo implica che il processo sarà stazionario del II ordine e causale nel senso che, esso dipenderà dalla storia passata delle innovazioni  $\varepsilon_t$ .

Un  $AR(1)$  è, dunque, rappresentabile come una serie infinita dei valori passati di  $WN$  con pesi che decrescono esponenzialmente, ovvero come un  $MA(\infty)$ .

Si evidenzia il caso limite di un  $AR(1)$  in cui risulti  $|\phi_1| = 1$ . In tale situazione, il processo non risulterà più stazionario per la presenza di una radice unitaria. Tale caso limite corrisponde al noto processo stocastico detto *Random Walk* che dunque, non sarà stazionario. Tornando ad un  $AR(1)$ , è agevole provare che esso sia stazionario.

Infatti, il valore atteso è dato da  $E(y_t) = E\left(\frac{1}{1 - \phi_1 L} \cdot \varepsilon_t\right) = \frac{E(\varepsilon_t)}{1 - \phi_1 L} = 0$ , dunque indipendente dal tempo. Dalla struttura del processo si avrà che

$$E(y_t \cdot y_{t-k}) = E\left[\left(\sum_{j=0}^{\infty} \phi_1^j \cdot \varepsilon_{t-j}\right) \cdot \left(\sum_{j=0}^{\infty} \phi_1^j \cdot \varepsilon_{t-k-j}\right)\right] = \dots = \phi_1^k \cdot \left(\sum_{j=0}^{\infty} \phi_1^{2 \cdot j}\right) \cdot \sigma_\varepsilon^2 = \frac{\phi_1^k \cdot \sigma_\varepsilon^2}{1 - \phi_1^2}$$

Allora, la funzione di autocovarianza sarà data da 
$$\gamma(k) = \begin{cases} \frac{\sigma_\varepsilon^2}{1 - \phi_1^2}, & k=0 \\ \frac{\phi_1^k \cdot \sigma_\varepsilon^2}{1 - \phi_1^2}, & k \neq 0 \end{cases}$$

dove, con  $k=0$  si ottiene la varianza. Dalla scrittura della funzione di auto-covarianza si vede come essa sia indipendente dal tempo e dipenda dal solo lag  $k$  rispettando, così, la ulteriore condizione affinché il processo sia stazionario.

Da ciò segue che 
$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \frac{\frac{\phi_1^k \cdot \sigma_\varepsilon^2}{1 - \phi_1^2}}{\frac{\sigma_\varepsilon^2}{1 - \phi_1^2}} = \frac{\phi_1^k \cdot \sigma_\varepsilon^2}{1 - \phi_1^2} \cdot \frac{1 - \phi_1^2}{\sigma_\varepsilon^2} = \phi_1^k$$
 e fa sì che tale  $p.s.$

sia *asintoticamente incorrelato* poiché la  $\rho(k)$  decade a zero esponenzialmente in ragione di  $k$  avendo imposta la condizione  $|\phi_1| < 1$ .

Più precisamente, se  $0 < \phi_1 < 1$  la funzione di autocorrelazione decade esponenzialmente a zero ma in maniera regolare. Se invece  $-1 < \phi_1 < 0$  essa decade esponenzialmente a zero ma in modo oscillatorio<sup>7</sup>. Quindi, il processo non può essere considerato a memoria lunga. E' utile esplicitare anche un  $AR(2)$  per poter meglio giustificare la rappresentazione  $MA(\infty)$  dei modelli auto-regressivi, prima di introdurre la rappresentazione  $AR(\infty)$  di un  $ARMA(p,q)$ .

<sup>7</sup> Il caso  $\phi_1 = 0$  riduce banalmente il processo  $y_t = 0 \cdot y_{t-1} + \varepsilon_t = \varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$  ovvero ad un  $W.N.$

Per aumentare la ricchezza strutturale e la varietà dei comportamenti della funzione di autocorrelazione, si possono aggiungere ulteriori variabili autoregressive ritardate alla relazione lineare.

Così, si può definire un processo  $AR(2)$  aggiungendo una seconda variabile come segue  $y_t = \phi_1 \cdot y_{t-1} + \phi_2 \cdot y_{t-2} + \varepsilon_t$ , ed anche in questo caso sarà possibile rappresentare il processo in termini dell'operatore ritardo come

$$\Phi(L)y_t = \varepsilon_t \text{ avendo cura di porre } \Phi(L) = 1 - \phi_1 L - \phi_2 L^2 \text{ ed ottenendo}$$

$$(1 - \phi_1 L - \phi_2 L^2) \cdot y_t = \varepsilon_t \Rightarrow y_t = \frac{1}{1 - \phi_1 L - \phi_2 L^2} \cdot \varepsilon_t .$$

Come eseguito per il processo  $AR(1)$ , anche in questo caso si deve perseguire l'obiettivo di esplicitare una soluzione stazionaria del processo, con le condizioni che ne garantiscano l'esistenza. Per far ciò si rappresenterà il processo  $AR(2)$  come un  $MA(\infty)$ , ovvero si stimeranno i parametri  $\psi_j$  affinché sia vera una rappresentazione del tipo  $y_t = (1 + \psi_1 L + \psi_2 L^2 + \dots) \cdot \varepsilon_t$  ottenendo allora che

$$y_t = \frac{1}{1 - \phi_1 L - \phi_2 L^2} \cdot \varepsilon_t = (1 + \psi_1 L + \psi_2 L^2 + \dots) \cdot \varepsilon_t .$$

Potendo trattare le espressioni in  $L$  alla stregua dei polinomi, l'eguaglianza tra le due rappresentazioni del processo, sarà vera solo se saranno uguali i due polinomi a destra ed a sinistra del segno di uguaglianza, ossia:

$$\frac{1}{1 - \phi_1 L - \phi_2 L^2} = (1 + \psi_1 L + \psi_2 L^2 + \dots) \text{ oppure equivalentemente se risulterà che}$$

$$1 = (1 - \phi_1 L - \phi_2 L^2) \cdot (1 + \psi_1 L + \psi_2 L^2 + \dots) .$$

Sviluppando il prodotto si avrà che:

$$1 = 1 + \psi_1 L + \psi_2 L^2 + \dots - \phi_1 L - \phi_1 \cdot \psi_1 L - \phi_1 \psi_2 L^2 - \dots - \phi_2 \cdot \psi_3 L^3 - \dots .$$

Tale prodotto di infiniti termini può essere risolto sfruttando sempre le proprietà dei polinomi, ed esattamente, affinché il polinomio di grado nullo con 1 come unico coefficiente sia uguale al polinomio a destra della uguaglianza, abbisogna che tutti i coefficienti del polinomio a sinistra siano uguali ai coefficienti di pari grado del polinomio a destra del segno di uguaglianza. Ma il grado dei coefficienti del polinomio a sinistra è 0 per ogni coefficiente e dunque, al netto della costante, tutti i coefficienti del polinomio a destra dell'uguaglianza saranno posti pari a 0 poiché deve risultare che  $\psi_1 - \phi_1 = 0$ ,  $\psi_2 - \phi_1 \cdot \psi_1 - \phi_2 = 0, \dots$ . Procedendo in questo modo si arriva alla seguente relazione ricorsiva, che rappresenta una equazione alle differenze  $\psi_0 = 1$ ,  $\psi_1 = \phi_1$ ,  $\psi_j = \phi_1 \cdot \psi_{j-1} + \phi_2 \cdot \psi_{j-2}$  con  $j = 2, 3, \dots$ , la quale ha, come soluzione, la successione dei coefficienti  $\{\psi_j\}_{j \in \mathbb{N}}$  voluti.

Come per i processi  $MA(\infty)$  anche per i processi  $AR(2)$  la stazionarietà sarà garantita dalla proprietà di assoluta sommabilità della successione dei coefficienti. Ma si dimostra che esiste una condizione sufficiente affinché sia garantita la stazionarietà, senza dover risolvere esplicitamente la ricorsione scritta sopra.

Precisamente, si definisce la equazione caratteristica del polinomio  $1 - \phi_1 L - \phi_2 L^2$  come la equazione algebrica in  $z$  e funzione dei parametri  $\phi_1, \phi_2$  come segue,  $1 - \phi_1 \cdot z - \phi_2 \cdot z^2 = 0$  ottenendo il risultato secondo cui, la successione  $\{\psi_j\}_{j \in \mathbb{N}}$  sarà assolutamente convergente se e solo se le radici  $z_1, z_2$  (eventualmente complesse e coniugate) risulteranno esterne al cerchio unitario sul piano complesso, ovvero se esse risulteranno maggiori di 1 in valore assoluto. Tale restrizione sulle radici può essere espressa ponendo dei vincoli sui coefficienti del polinomio  $1 - \phi_1 L - \phi_2 L^2$  e precisamente, se i coefficienti

$$\phi_1, \phi_2 \text{ saranno tali da rispettare le condizioni } \begin{cases} \phi_1 + \phi_2 > -1 \\ \phi_1 - \phi_2 < 1 \\ |\phi_2| < 1 \end{cases} \text{ allora il processo}$$

$AR(2)$ , rappresentato come un  $MA(\infty)$ , risulterà stazionario in covarianza.

Nei processi della classe  $ARMA(p, q)$  sin'ora definiti, si è assunto che il valore atteso fosse nullo. Si può, tuttavia, rimuovere tale assunzione senza alterare la proprietà di dipendenza temporale di essi. Per i modelli  $MA(q)$  si può generalizzare il caso di una media non nulla modificando la scrittura, esplicitata in funzione dell'operatore ritardo, come segue  $y_t = \mu + \Theta(L)\varepsilon_t$  ottenendo come unica differenza che la speranza matematica del processo non sarà nulla bensì risulterà pari a  $\mu$ .

Invece, per i modelli  $AR(p)$ , si può esprimere la rappresentazione di essi ( in termini dell'operatore ritardo) considerando una speranza matematica non nulla nel seguente modo  $\Phi(L)(y_t - \mu) = \varepsilon_t$  ed effettuando i calcoli con l'operatore  $L$  si ha che

$$y_t - \mu = \phi_1 \cdot (Y_{t-1} - \mu) + \dots + \phi_p \cdot (Y_{t-p} - \mu) + \varepsilon_t .$$

Considerando la rappresentazione  $MA(\infty)$  sopra descritta si avrà che, l'introduzione del termine costante  $\mu$  modificherà solo la speranza matematica del processo che sarà pari proprio a  $\mu$ , senza andare ad inficiare il calcolo delle funzioni di autocovarianza ed autocorrelazioni per le quali il processo continuerà ad essere stazionario.

Per i modelli  $AR(p)$  definiti sopra, ponendo il valore atteso del processo diverso da zero, le funzioni dei momenti di interesse, ovvero valore atteso e funzione di varianza autocovarianza ed autocorrelazione, sono ottenibili attraverso manipolazioni algebriche che si basano fondamentalmente sui procedimenti visti per i casi con  $p=1$  e  $p=2$  con la differenza d'avere una maggiore complessità computazionale.

I processi  $AR(p)$  e  $MA(q)$  possono essere annidati all'interno di un modello più generale che prende il nome di *Processo Autoregressivo a media mobile di ordine  $p, q$*  ( $ARMA(p, q)$ ). Tale procedura è finalizzata ad assicurare la massima flessibilità di comportamento per un processo lineare a memoria finita.

Tale processo può essere visto come un  $AR(p)$  in cui però, la componente stocastica, non è rappresentata da  $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$  bensì da un  $MA(q)$ .

Utilizzando la notazione usata sin'ora, un  $ARMA(p,q)$  potrà essere rappresentato come  $y_t = \phi_1 \cdot y_{t-1} + \phi_2 \cdot y_{t-2} + \dots + \phi_p \cdot y_{t-p} + u_t$  dove, per compattezza, si è posto  $u_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$  con la condizione che  $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$ .

Anche per tali modelli, è possibile esplicitare la loro rappresentazione sfruttando i polinomi dell'operatore ritardo per la componente, sia auto-regressiva che per quella a media mobile. Dunque un p.s. sarà un  $ARMA(p,q)$  se soddisfa la seguente equazione

$y_t = \phi_1 \cdot y_{t-1} + \dots + \phi_p \cdot y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$  esplicitabile tramite l'operatore  $L$  come  $\Phi(L)y_t = \Theta(L)\varepsilon_t$  con la solita condizione  $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$  ed avendo posto  $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$  e  $\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q$ .

La rappresentazione  $MA(\infty)$  di  $y_t$  sarà  $y_t = \Psi(L) \cdot \varepsilon_t = \sum_{j=0}^{\infty} \psi_j \cdot \varepsilon_t$ .

I coefficienti di  $\Psi(L)$  si ottengono come visto precedentemente, ovvero ponendo

$$\Psi(L) = \frac{\Theta(L)}{\Phi(L)} = \frac{1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q}{1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p} \Rightarrow \Phi(L) \cdot \Psi(L) = \Theta(L)$$

ed esplicitando i prodotti uguagliando i coefficienti dei polinomi a destra e sinistra della uguaglianza, ottenendo la funzione ricorsiva per il calcolo di essi.

Si può mostrare che la stazionarietà di  $y_t$  dipenderà dalla sola componente  $AR(p)$  e che il processo risulterà stazionario se e solo se le radici dell'equazione caratteristica del polinomio  $\Phi(L) = 0$  saranno esterne al cerchio di raggio unitario sul piano complesso. In tal caso sarà ottenibile una rappresentazione in forma  $MA(\infty)$  del processo  $ARMA(p,q)$ . Inoltre, tale risultato sarà valido solo se le equazioni caratteristiche  $\Phi(z) = 0$ ,  $\Theta(z) = 0$  non avranno radici in comune.

Per capire questo ultimo asserto si consideri un processo  $ARMA(1,1)$  del tipo  $(1 - \phi_1 L) \cdot y_t = (1 + \theta_1) \cdot \varepsilon_t$ . Se le equazioni caratteristiche dei due polinomi in  $L$  avessero radici uguali in valore assoluto, ovvero se risultasse che  $\phi_1 = -\theta_1$  si

avrebbe che  $y_t = \frac{1 + \theta_1}{1 - \phi_1} \cdot \varepsilon_t$  ma essendo  $\phi_1 = -\theta_1$  allora  $\frac{1 + \theta_1}{1 - (-\theta_1)} = 1$  ed il

processo si ridurrebbe ad  $y_t = \varepsilon_t$  che si sa essere un  $WN(0, \sigma_\varepsilon^2)$  stazionario per qualunque valore di  $\phi_1$ . Tale parametrizzazione va evidentemente evitata poiché, essendo gli operatori  $(1 - \phi_1 L), (1 + \theta_1)$  uguali per via delle radici comuni, essi semplificandosi non permettono di poter decidere, sulla base dei dati, se il processo generatore dei dati sia un  $WN(0, \sigma_\varepsilon^2)$  o un  $ARMA(1,1)$ .

Si può dire che, se in un processo  $ARMA(1,1)$  risulterà che  $\phi_1 = -\theta_1$ , allora esso sarà *sovraparametrizzato* nel senso che, per il principio dei rasoi di Occam, sarà meglio, a parità di potere esplicativo, utilizzare una modellizzazione più parsimoniosa come un White Noise (equivalente ad un  $ARMA(0,0)$ ) rispetto ad un  $ARMA(1,1)$ . Si introdurrà ora, una proprietà importante dei processi della classe  $ARMA(p,q)$  partendo da un  $MA(1)$ .

Come per i processi  $AR(1)$  si è vista la procedura per una rappresentazione a media mobile di ordine infinito così, per i processi a media mobile, ci si può chiedere se è possibile rappresentarli come un processo autoregressivo di ordine infinito. Tale concetto riguarda la proprietà fondamentale di invertibilità dei processi della classe  $ARMA(p,q)$ .

Sia un processo  $MA(1)$  del tipo  $y_t = \varepsilon_t + \theta_1 \cdot \varepsilon_{t-1} \Rightarrow y_t = (1 + \theta_1 L) \cdot \varepsilon_t \Rightarrow (1 + \theta_1 L)^{-1} \cdot y_t = \varepsilon_t$  interpretando, senza modificare la sostanza della scrittura,  $(1 + \theta_1 L)^{-1} = (1 - \theta_1(-L))^{-1}$  si è scritto il polinomio in  $L$  iniziale come un altro polinomio in  $-L$  di ragione  $\theta_1$  dunque esprimibile come una serie geometrica di ordine infinito, ovvero  $(1 + \theta_1 L)^{-1} = 1 + \eta_1(-L) + \eta_2(-L)^2 + \dots$  con l'assegnazione  $\eta_j = (-\theta_1)^j$  con  $j = 1, 2, \dots$  e la condizione iniziale che  $\eta_0 \equiv 1$ .

Possiamo, dopo queste manipolazioni algebriche, rappresentare il processo come  $(1 + \eta_1(-L) + \eta_2(-L)^2 + \dots) \cdot y_t = \varepsilon_t$  ottenendo una somma infinita di variabili aleatorie che, come specificato nelle precedenti pagine, ha senso solo se converge in media quadratica ad una variabile limite e ciò si verifica solo se vale la condizione di assoluta sommabilità verificata solo se  $|\theta_1| < 1$ . Un processo  $MA(1)$  della classe  $ARMA(p,q)$  per il quale  $|\theta_1| < 1$  è detto *invertibile* nel senso che, è possibile invertire l'operatore che definisce la parte a media mobile per dare una rappresentazione autoregressiva del processo, di ordine infinito.

La condizione di invertibilità, in generale, risolve ogni problema di identificazione dei modelli per gli  $ARMA(p,q)$ .

Essa sarà garantita per i processi  $ARMA(p,q)$  se le radici del polinomio caratteristico della componente a media mobile del processo, risulteranno esterne al cerchio di raggio unitario sul piano complesso. In tal caso, la proprietà di invertibilità permetterà una rappresentazione  $AR(\infty)$  del processo, dunque, esso sarà funzione lineare dei suoi valori ritardati a meno di un termine  $WN$  e si potrà prevedere l'andamento futuro del processo sulla base della sua storia passata.

Ciò permetterà di identificare una specifica funzione di auto-covarianza e di autocorrelazione in maniera univoca eliminando così problemi di identificazione del modello. Si genererà in tal modo, una corrispondenza biunivoca tra lo spazio parametrico del processo e lo spazio delle funzioni di auto-covarianza ed autocorrelazione. Un semplice esempio può ben mostrare tale concetto.

Siano i seguenti processi  $MA(1)$ ;  $y_t = \varepsilon_t + \theta_1 \cdot \varepsilon_{t-1}$  con  $|\theta_1| < \infty$  ed

$$\varepsilon_t \sim WN(0, \sigma_\varepsilon^2), \quad \tilde{y}_t = \tilde{\varepsilon}_t + \tilde{\theta}_1 \cdot \tilde{\varepsilon}_{t-1} \quad \text{con} \quad |\tilde{\theta}_1| < \infty \quad \text{ed} \quad \tilde{\varepsilon}_t \sim WN(0, \tilde{\sigma}_\varepsilon^2).$$

Per il primo processo si ha che la funzione di autocovarianza assumerà ai  $lag0$  ed  $lag1$  i valori rispettivamente  $\gamma(0) = (1 + \theta_1) \cdot \sigma_\varepsilon^2$ ,  $\gamma(1) = \theta_1 \cdot \sigma_\varepsilon^2$  e sarà nulla per tutti gli altri  $lag$ . Per il secondo processo, imponendo le restrizioni (azione fattibilissima)  $\tilde{\theta}_1 = \frac{1}{\theta_1}$ ,  $\tilde{\sigma}_\varepsilon^2 = \sigma_\varepsilon^2 \cdot \theta_1^2$  si avrà che la funzione di auto-covarianza al  $lag0$  e  $lag1$  sarà indistinguibile da quella del primo processo poiché risulterà

$\tilde{\gamma}(0) = (1 + \tilde{\theta}_1^2) \cdot \tilde{\sigma}_\varepsilon^2$  ,  $\tilde{\gamma}(1) = \tilde{\theta}_1^2 \cdot \tilde{\sigma}_\varepsilon^2$  ma considerando le restrizioni si avrà

$$\tilde{\gamma}(0) = \left(1 + \frac{1}{\theta_1^2}\right) \cdot \sigma_\varepsilon^2 \cdot \theta_1^2 = \frac{1 + \theta_1^2}{\theta_1^2} \cdot \sigma_\varepsilon^2 \cdot \theta_1^2 = (1 + \theta_1^2) \cdot \sigma_\varepsilon^2 \text{ con}$$

$\tilde{\gamma}(1) = \frac{1}{\theta_1} \cdot \sigma_\varepsilon^2 \cdot \theta_1^2 = \sigma_\varepsilon^2 \cdot \theta_1$  e come detto, le funzioni di auto-covarianza dei due

processi saranno indistinguibili e non sarà possibile, sulla base di esse, identificare correttamente il processo generatore dei dati.

Tuttavia, essendo il primo processo invertibile sarà rappresentabile come un  $AR(\infty)$  al contrario del secondo processo. In tal caso, la invertibilità per i processi della famiglia  $ARMA(p,q)$  stazionari, garantisce una rappresentazione autoregressiva di ordine infinito e permette di identificare lo specifico processo con una ed una sola funzione di autocovarianza eliminando i problemi di identificazione.

I modelli della famiglia  $ARMA(p,q)$  così come definiti sin'ora, non possono essere considerati come modelli per il valore atteso condizionato di  $y$  poiché, la componente stocastica del processo è un  $WN$ , sul cui valore atteso condizionato nessuna condizione è stata imposta.

Se introduciamo una diversa struttura per la componente stocastica, è possibile considerare a tutti gli effetti i processi della famiglia  $ARMA(p,q)$  modelli per il valore atteso condizionato.

Allora, sia un processo  $ARMA(1,1)$  del tipo  $y_t = \phi_1 \cdot y_{t-1} + \varepsilon_t + \theta_1 \cdot \varepsilon_{t-1}$  dove  $\varepsilon_t \sim IID(0, \sigma^2)$  con la condizione  $E(\varepsilon_t / I_{t-1}) = 0$  ovvero, il valore atteso condizionato al set informativo disponibile la tempo  $t$  della componente stocastica del processo, è nullo.

In tal caso, dato il set informativo disponibile al tempo  $t$ , si avrà che, le misure

$$E(y_t / I_{t-1}) = E(\phi_1 \cdot y_{t-1} + \varepsilon_t + \theta_1 \cdot \varepsilon_{t-1} / I_{t-1}) = \phi_1 \cdot y_{t-1} + \theta_1 \cdot \varepsilon_{t-1} \text{ e}$$

$$Var(y_t / I_{t-1}) = Var(\phi_1 \cdot y_{t-1} + \varepsilon_t + \theta_1 \cdot \varepsilon_{t-1} / I_{t-1}) = Var(\varepsilon_t / I_{t-1}) = \sigma_\varepsilon^2$$

rappresenteranno rispettivamente il valore atteso condizionato e la varianza attesa condizionata del processo, al tempo  $t$ .

Si può ben notare come, imponendo la condizione alla componente stocastica

$\varepsilon_t \sim IID(0, \sigma^2) \wedge E(\varepsilon_t / I_{t-1}) = 0$  le misure del valore atteso e della varianza del  $p.s.$  (condizionate) siano assai diverse da quelle non condizionate esplicitate precedentemente. Un ulteriore processo notevole da richiamare ha una struttura simile ad un  $MA(\infty)$  con una differente condizione sugli elementi che lo compongono.

Precisamente, dato un  $p.s.$   $y = \{y_t\}_{t \in T}$  esso è definito *Processo Lineare* se

ammetterà una rappresentazione  $MA(\infty)$  in termini di  $\varepsilon_t \sim IID(0, \sigma_\varepsilon^2)$  del tipo

$$y_t = \sum_{j=0}^{\infty} \psi_j \cdot \varepsilon_{t-j} = \Psi(L) \varepsilon_t \text{ dove } \Psi(L) = \sum_{j=0}^{\infty} \psi_j L^j \text{ con la condizione che}$$

$$\sum_{j=0}^{\infty} \psi_j \cdot z^j < \infty \text{ per } |z| < 1 + \delta, \delta \in \mathbf{C}_0^{++}.$$

Tale ultima condizione è legata alla rappresentazione dei processi integrati VAR di cui si parlerà dettagliatamente nel prosieguo del capitolo. Confrontando tale rappresentazione  $MA(\infty)$  rispetto a quella esplicitata precedentemente, si nota subito che essa è funzione di impulsi  $IID$  e non più di un  $WN$ , dunque essa sarà stazionaria anche in senso forte. Sin qui si sono esposti, in linea di massima, le principali rappresentazioni di traiettorie derivanti da un processo generatore dei dati, funzionali alla esplicitazione dei modelli che saranno usati per lo studio dei dati reali di questo lavoro. Empiricamente, una volta definito il modello idoneo alla rappresentazione del processo generatore dei dati, si deve avere un metodo per la stima dei parametri d'interesse. Ad esempio, per i modelli  $ARMA(p,q)$  esplicitati, tale attività consta nella stima del vettore dei parametri della componente autoregressiva  $\phi_i$  con  $i=1,2,\dots,p$ , della componente a media mobile  $\theta_j$  con  $j=1,2,\dots,q$  del valore atteso  $\mu$  e della varianza  $\sigma_\varepsilon^2$ .

Per la versatilità rispetto ai metodi statistici utili a tal fine, si introduce il metodo di stima della *Massima Verosimiglianza (ML)* utilizzabile per i modelli  $ARMA(p,q)$  e per qualunque modello parametrico<sup>8</sup>. Il metodo della *Massima Verosimiglianza* fu introdotto da Sir Ronald Aymer Fisher nel 1912, e gode di grande applicazione grazie ad una proprietà fondamentale che fu specificata dallo stesso autore, ovvero la proprietà di *efficienza asintotica*.<sup>9</sup>

Dunque, si ipotizzi di aver specificato un modello statistico per la rappresentazione di una data traiettoria  $(z_1, z_2, \dots, z_T)$ , descritto dalla funzione di densità di probabilità congiunta  $f(z_1, z_2, \dots, z_T; \vec{\theta})$  con  $\vec{\theta}$  vettore di  $p$  parametri incogniti che assumono valore in uno spazio parametrico  $\Theta \subseteq \mathbf{C}_0^p$ .

In caso di indipendenza stocastica degli impulsi  $z_t$  con  $t=1,2,\dots,p$

---

8 Nella introduzione di un modello analitico finalizzato alla rappresentazione statistica di un fenomeno d'interesse, si possono avere due principali situazioni di mancanza di conoscenza: la prima riguarda la presenza di una conoscenza parziale della funzione che ha generati i dati  $f(x; \vec{\theta})$  (dove  $\vec{\theta} = \{\theta_1, \theta_2, \dots, \theta_k\}$  è il vettore dei parametri d'interesse). In tal caso, si conosce la forma analitica della funzione ma non si conoscono alcuni o tutti i valori dei parametri che caratterizzano la funzione. Si parla in questo caso di "inferenza statistica parametrica". La seconda situazione si ha quando non si conosce la forma analitica della funzione generatrice dei dati; si parla in tal caso di "inferenza statistica non parametrica" anche se, la denominazione più appropriata è quella di "inferenza statistica libera da distribuzione o distribution free". Una terza situazione, intermedia tra le precedenti, si ha quando si specificano alcune assunzioni sul modello come, ad esempio, l'ipotesi sulla famiglia a cui appartiene il p.s. ma non si dice nulla sulla sottofamiglia di appartenenza, ovvero nulla si dice sulla funzione di massa o di densità di probabilità della traiettoria. Si parla in tal caso di "inferenza statistica semi-parametrica" poiché la forma analitica del modello è esplicitata solo parzialmente.

9 Uno stimatore è detto *asintoticamente efficiente* o *consistente in media quadratica* se il suo errore quadratico medio tende a zero al crescere della numerosità campionaria, ovvero dato  $MSE = E[(T - \theta)^2]$  errore quadratico medio dello stimatore  $T$  di  $\theta$ , allora  $T$  è stimatore asintoticamente efficiente o consistente del parametro d'interesse se e solo se risulterà che  $\lim_{n \rightarrow +\infty} MSE(T_n) = \lim_{n \rightarrow +\infty} E[(T_n - \theta)^2] = 0$  dove  $T_n$  indica la dipendenza dalla numerosità campionaria dello stimatore  $T$ .

la fattorizzazione della  $f(z_1, z_2, \dots, z_T; \vec{\theta})$  risulterà facilitata e sarà rappresentabile come  $f(z_1, z_2, \dots, z_T; \vec{\theta}) = \prod_{t=1}^T f(z_t; \vec{\theta})$  con  $f(z_t; \vec{\theta})$  funzione di densità marginale dei singoli impulsi.

Nel caso in cui non sia presente la proprietà di indipendenza stocastica tra i vari impulsi, la funzione di densità congiunta potrà essere fattorizzata nel seguente modo:

$f(z_1, z_2, \dots, z_T; \vec{\theta}) = \prod_{t=2}^T f(z_t / z_{t-1}, z_{t-2}, \dots, z_1; \vec{\theta}) \cdot f(z_1; \vec{\theta})$  con la espressione  $f(z_t / z_{t-1}, z_{t-2}, \dots, z_1; \vec{\theta})$  che rappresenta la funzione di densità della variabile aleatoria  $z_t$  condizionatamente alla propria storia passata.

Tale ultima fattorizzazione è legata al concetto della probabilità totale di eventi qualunque e precisamente alla c.d. *formula della catena* della probabilità congiunta riguardante eventi, ma che può essere applicata anche a v.c. e di cui si riporta una breve dimostrazione per chiarezza di esposizione, riferendo il tutto ad eventi.

Dato uno spazio campionario di eventi  $\Omega$  e dati due eventi  $(A, B) \in \Omega$  si definisce probabilità condizionata di A dato B come  $P(A/B) = \frac{P(A \cap B)}{P(B)}$  con

$P(B) > 0$ ; da questo si ottiene che  $P(A \cap B) = P(A/B) \cdot P(B)$  ovvero, la probabilità congiunta di eventi è data dal prodotto della probabilità condizionata tra i due eventi per la probabilità dell'evento condizionante. Tale approccio può essere esteso ad un numero di eventi maggiore di due utilizzando la *formula della catena* e precisamente, data un numero finito di eventi  $A_j$  per  $j=1, 2, \dots, n$  si ha che, sfruttando la probabilità condizionata tra eventi, la probabilità congiunta di essi sarà data da  $P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) \cdot P(A_2/A_1) \cdot \dots \cdot P(A_n/A_1 \cap A_2 \cap \dots \cap A_{n-1})$  valida quando  $P(A_1 \cap A_2 \cap \dots \cap A_{n-1}) > 0$ .

Se applichiamo questo procedimento a v.c. anziché ad eventi si ottiene la formula scritta prima poiché:

$$P(z_1, z_2, \dots, z_T; \vec{\theta}) = P(z_T / z_{T-1}, z_{T-2}, \dots, z_1) \cdot P(z_{T-1} / z_{T-2}, z_{T-3}, \dots, z_1) \cdot \dots \cdot P(z_2 / z_1) \cdot P(z_1) = P(z_1) \cdot \prod_{t=2}^T P(z_t / z_{t-1}, z_{t-2}, \dots, z_1; \vec{\theta})$$

considerando la funzione densità di probabilità  $f$  delle v.c. e scambiandola con la massa di probabilità  $P$  si giunge alla dimostrazione della fattorizzazione seguente:

$$f(z_1, z_2, \dots, z_T; \vec{\theta}) = \prod_{t=2}^T f(z_t / z_{t-1}, z_{t-2}, \dots, z_1; \vec{\theta}) \cdot f(z_1; \vec{\theta}) .$$

Al netto della giustificazione della fattorizzazione della densità congiunta, il principio della *Massima Verosimiglianza* si fonda sulla idea di considerare la funzione densità di probabilità della traiettoria osservata, come funzione del vettore dei parametri del tipo  $L(\vec{\theta}) = f(z_1, z_2, \dots, z_T; \vec{\theta})$ .

In sostanza, la funzione di verosimiglianza è uguale alla funzione densità di probabilità congiunta di *variabili casuali*, a meno di un fattore di proporzionalità  $k=k(z_1, z_2, \dots, z_T)$  indipendente dal vettore dei parametri.

Si avrà che  $L(\vec{\theta})=k(z_1, z_2, \dots, z_T) \cdot f(z_1, z_2, \dots, z_T; \vec{\theta}) \propto f(z_1, z_2, \dots, z_T; \vec{\theta})$  pur giungendo alle medesime conclusioni, date le due definizioni di *verosimiglianza*.

Concettualmente essa rappresenta, dato un generico vettore dei parametri  $\vec{\theta}$ , la probabilità che si manifesti la specifica traiettoria osservata. Il problema dunque consta del trovare il vettore dei parametri d'interesse affinché sia massima la probabilità di manifestazione di essa. Questo ragionamento conduce alla definizione di *stimatore di massima verosimiglianza*. Si definisce, dunque, lo *stimatore di massima verosimiglianza*, indicandolo con la notazione  $\hat{\theta}_{ML}$ , come il vettore  $\vec{\theta} \in \vec{\Theta}$  che massimizza la funzione di verosimiglianza.

Si tratta di trovare l'elemento dello spazio parametrico  $\vec{\Theta}$  che massimizzi la probabilità di realizzazione della specifica traiettoria osservata.

Lo stimatore *ML* (notazione abbreviata) consta, dunque, della soluzione del problema di massimo  $L(\hat{\theta})=\max_{\vec{\theta}} L(\vec{\theta})$ . Nel caso la funzione di verosimiglianza sia

derivabile due volte, il vettore  $\hat{\theta}$  che risolve il problema di massimo, soddisferà la equazione  $L'(\vec{\theta})=\frac{\partial}{\partial \vec{\theta}} L(\vec{\theta})=\vec{0}$  e la condizione della matrice *Hessiana*

$$\frac{\partial^2}{\partial \vec{\theta} \partial \vec{\theta}} L(\vec{\theta}) \text{ che dovrà risultare definita negativa}^{10}.$$

Questa ultima condizione, data la forma funzionale della verosimiglianza, sarà sostanzialmente sempre garantita assicurando l'esistenza di un massimo unico che corrisponderà all'unica soluzione dell'equazione  $L'(\vec{\theta})=\vec{0}$ .

Una metodologia, che semplifica il calcolo del punto di ottimo, è quella di calcolare la derivata prima della trasformata logaritmica della funzione di verosimiglianza e la relativa matrice *Hessiana*. Si avrà che  $l(\theta)=\ln L(\theta)$  sarà la logaritmica della funzione di verosimiglianza ed il problema di ottimo sarà  $\ln L(\hat{\theta})=\max_{\vec{\theta}} \ln L(\vec{\theta})$

ovvero, la stima di massima verosimiglianza del parametro  $\theta$  sarà la soluzione della equazione  $\frac{\partial}{\partial \vec{\theta}} \ln L(\vec{\theta})=\vec{0}$  e che nel punto  $\vec{\theta}=\hat{\theta}$  soddisfi la condizione

$\frac{\partial^2}{\partial \vec{\theta} \partial \vec{\theta}} \ln L(\vec{\theta}) < 0$  ovvero che la matrice delle derivate seconde sia definita negativa.

10 Questo significa che  $L'(\hat{\theta})=\frac{\partial}{\partial \vec{\theta}} L(\vec{\theta})|_{\vec{\theta}=\hat{\theta}}$  sarà nulla e che  $L''(\hat{\theta})=\frac{\partial^2}{\partial \vec{\theta} \partial \vec{\theta}} L(\vec{\theta})|_{\vec{\theta}=\hat{\theta}}$  sarà definita negativa.

La quantità  $\frac{\partial}{\partial \vec{\theta}} \ln L(\vec{\theta}) = \vec{0}$  e detta *score*, e si evidenzia che la sua soluzione soddisferà anche l'equazione  $L'(\vec{\theta}) = \frac{\partial}{\partial \vec{\theta}} L(\vec{\theta}) = \vec{0}$  poiché si avrà che

$$l'(\vec{\theta}) = \frac{\partial}{\partial \vec{\theta}} \ln L(\vec{\theta}) = \frac{1}{L(\vec{\theta})} \cdot \frac{\partial}{\partial \vec{\theta}} L(\vec{\theta}) = \frac{L'(\vec{\theta})}{L(\vec{\theta})} .$$

Rappresentando la  $L(\vec{\theta})$  una probabilità, per gli assiomi di Kolmogorov sarà  $L(\vec{\theta}) > 0$  e dunque  $l'(\vec{\theta}) = \vec{0} \Leftrightarrow L'(\vec{\theta}) = \frac{\partial}{\partial \vec{\theta}} L(\vec{\theta}) = \vec{0}$  ovvero la soluzione del problema di massimo

$$L(\hat{\vec{\theta}}) = \max_{\vec{\theta}} L(\vec{\theta}) .$$

Si riportano brevemente delle fondamentali proprietà campionarie, finite ed asintotiche, che rendono lo stimatore *ML* molto utilizzato nei problemi inferenziali per via della sua duttilità.

Una prima proprietà importante è la *Proprietà di Equivarianza*. Essa afferma che, se

$g(\cdot)$  è un funzionale derivabile rispetto al vettore dei parametri  $\vec{\theta}$ , allora la

stima *ML* di  $g(\vec{\theta})$  sarà  $g(\hat{\vec{\theta}})$ . Tale risultato è importante poiché assicura che, noto lo stimatore *ML* sarà automaticamente anche noto lo stimatore *ML* di una

qualunque funzione di  $\hat{\vec{\theta}}$ . Conducendo, quindi, una analisi di *Massima Verosimiglianza* su diverse ri-parametrazioni del modello statistico, si addiverrà ai medesimi risultati. Una ulteriore importante proprietà dello stimatore *ML* è quella di *efficienza*. Precisamente, se esiste uno stimatore efficiente, lo stimatore *ML* ne è funzione di esso. La efficienza considerata è quella di Rao-Cramer legata al *limite di Rao-Cramer* che trova un minimo assoluto per la matrice di varianza/covarianza di uno stimatore corretto. Nello specifico si dimostra che, per ogni stimatore corretto

$$\tilde{\vec{\theta}} \text{ varrà che } \text{Var}(\tilde{\vec{\theta}}) > \mathbf{I}_T(\vec{\theta})^{-1} = -E\left(\frac{\partial^2 l(\vec{\theta})}{\partial \vec{\theta} \partial \vec{\theta}}\right)^{-1} \text{ dove } \mathbf{I}_T(\vec{\theta})^{-1} \text{ è detta}$$

*Matrice Informativa di Fisher*. Con tale specifica, allora, la proprietà in questione afferma che, se esiste uno stimatore con varianza minima, lo stimatore *ML* ne è funzione. Si riportano, in fine, le proprietà asintotiche dello stimatore *ML* che lo rendono assai utilizzato.

Nello specifico, nel caso di processi stazionari ed ergodici, lo stimatore *ML* risulta consistente, asintoticamente normale ed efficiente, ovvero risulterà che,  $\hat{\vec{\theta}} \xrightarrow{p} \vec{\theta}$  per

la consistenza;  $\sqrt{T} \cdot (\hat{\vec{\theta}} - \vec{\theta}) \xrightarrow{d} N(\vec{0}, \mathbf{I}_\infty(\vec{\theta})^{-1})$  per la normalità asintotica dove

$$\mathbf{I}_\infty(\vec{\theta}) = \lim_{T \rightarrow \infty} \frac{\mathbf{I}_T(\vec{\theta})}{T}$$

rappresenta la *matrice di informazione asintotica di Fischer* la quale esplicita la varianza dello *score*.

Tale matrice misura la curvatura della funzione di verosimiglianza a livello dello stimatore *ML*.

Una derivata seconda bassa dello *score* indica una bassa curvatura e dunque, le variabili del sistema sono di scarsa informazione per il vettore dei parametri  $\vec{\theta}$  da cui dipende la *funzione di verosimiglianza*.

Le proprietà suddette dello stimatore *ML* lo rendono efficiente poiché, non esisterà nessuno stimatore che avrà varianza minore.

I concetti sin qui brevemente esposti, sono riferibili ai casi di traiettorie univariate, ovvero situazioni in cui il fenomeno studiato non entra in relazione con altre variabili. Sono stati esposti le principali metodologie di stima dei parametri d'interesse, utilizzati nella prassi statistica.

Il presente lavoro non è finalizzato ad una esposizione dettagliata ed esaustiva delle metodologie di stima adottate in statistica, motivo questo per il quale si sono viste solo le principali e più usate di esse. Si vuole specificare che, nella letteratura, le metodologie di stima sono assai numerose anche se, fondamentalmente, si basano su quelle qui succintamente esposte.

Nel prossimo paragrafo si affronterà l'approccio multivariato della analisi statistica, situazione assai più simile alla realtà con la quale la materia si deve confrontare e su cui deve espletare la sua funzione.

### 3.3

#### Alcuni modelli multi-variati per serie storiche

Le metodologie affrontate nelle righe precedenti assumono che, i fenomeni a cui esse vengono applicate, evolvano esclusivamente in funzione della propria storia passata, senza che vi siano interazioni con altri fenomeni che fanno parte del sistema oggetto di studio. Tale assunzione risulta, evidentemente, assai poco veritiera poiché nella realtà, ogni fenomeno viene a trovarsi in un sistema nel quale vi sono legami, interazioni, dipendenze, causalità e quant'altro, con altri fenomeni (variabili) del sistema e tutto questo è legato alla complessità di cui i sistemi reali sono permeati. Dunque, i modelli uni-variati, mal si prestano all'analisi di siffatti sistemi. Nel caso multi-variato, invece, tali caratteristiche vengono correttamente modellizzate al fine di poterle interpretare ai fini dell'analisi.

Molti concetti visti nel caso uni-variato non variano a meno di una maggiore complessità computazionale. Il concetto fondamentale di stazionarietà resta lo stesso per la stazionarietà forte e debole con l'aggiunta che, per la stazionarietà forte, l'invarianza a traslazioni sull'asse dei tempi riguarderà la legge di probabilità *G*-variata

del processo stocastico  $\vec{y}_t = (y_{1,t}, y_{2,t}, \dots, y_{G,t})$  dove il pedice  $1, 2, \dots, G$  rappresenta la manifestazione della *i*-esima v.a. al tempo *t*.

Per la stazionarietà debole invece, l'indipendenza dal tempo del valore atteso del processo riguarderà un vettore di valori attesi *Gx1*-dimensionale e l'indipendenza dal

tempo della funzione di covarianza riguarderà una matrice di covarianze  $G \times G$ -dimensionale.

Precisamente si avrà che, il vettore di valori attesi sarà rappresentato come

$$\mu(t) = E(\vec{y}_t) = \begin{bmatrix} E(y_{1,t}) \\ E(y_{2,t}) \\ \vdots \\ E(y_{G,t}) \end{bmatrix} \text{ e la matrice di covarianze sarà data da}$$

$$\Gamma_t(k) = \text{Cov}(\vec{y}_t, \vec{y}_{t-k}) = \begin{bmatrix} \text{Cov}(\vec{y}_{1,t}, \vec{y}_{1,t-k}) & \text{Cov}(\vec{y}_{1,t}, \vec{y}_{2,t-k}) & \dots & \text{Cov}(\vec{y}_{1,t}, \vec{y}_{G,t-k}) \\ \text{Cov}(\vec{y}_{2,t}, \vec{y}_{1,t-k}) & \text{Cov}(\vec{y}_{2,t}, \vec{y}_{2,t-k}) & \dots & \text{Cov}(\vec{y}_{2,t}, \vec{y}_{G,t-k}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(\vec{y}_{G,t}, \vec{y}_{1,t-k}) & \text{Cov}(\vec{y}_{G,t}, \vec{y}_{2,t-k}) & \dots & \text{Cov}(\vec{y}_{G,t}, \vec{y}_{G,t-k}) \end{bmatrix} .$$

Tale matrice assume il solito significato della covarianza e nello specifico, gli elementi sulla diagonale principale indicano le autocovarianze univariate delle componenti del p.s.  $\vec{y}_t$  in esame, mentre gli elementi sulle parti triangolari della matrice indicano quale è la misura delle interrelazioni lineari, ai vari ritardi, delle componenti di  $\vec{y}_t$ .

Va sottolineata una lieve differenza tra la funzione di autocovarianza nel caso univariato ed in questo. Ovvero, la stazionarietà debole fa sì che, nel caso univariato,

$$y(k) = y(-k) .$$

Per quanto riguarda il caso multivariato, invece, rappresentando l'elemento generico  $i,j$  di  $\Gamma(k)$  la covarianza tra  $y_i$  al tempo  $t$  e  $y_j$  al tempo  $t-k$ , nulla ci garantisce che tale elemento generico sia uguale all'elemento  $i,j$  di

$$\Gamma(-k) \text{ che rappresenta la covarianza tra } y_i \text{ al tempo } t \text{ e } y_j \text{ al tempo } t+k.$$

In generale risulterà che  $\Gamma(k) \neq \Gamma(-k)$  e la relazione corretta sarà

$$\Gamma(k) = \Gamma(-k)' \quad ^{11}.$$

Anche per la proprietà di ergodicità, nel caso multivariato, il concetto resta lo stesso al caso univariato con la unica differenza formale che, essa, viene definita in termini di una matrice di momenti  $E(g(\vec{y}_t))$  e richiederà la convergenza della somma

$$\frac{1}{T} \cdot \sum_t g(y_t) \text{ al corrispondente momento teorico.}$$

Per completezza si esplicita la notazione dell'operatore ritardo nel caso multivariato che, come per i concetti appena esposti sopra, resta fondamentalmente uguale al caso uni-variato a meno di una maggiore dimensionalità che può apparentemente rendere la notazione più farraginoso.

Precisamente, dato un processo  $G$ -dimensionale  $y_t$  la sua dinamicità può essere rappresentata sfruttando l'operatore ritardo come segue:

11 Data una matrice qualunque  $A$ , la matrice  $A'$  ne rappresenta la sua trasposta, formata invertendo le righe con le colonne. Se la matrice iniziale è quadrata la sua trasposta sarà sempre quadrata, se invece sarà di dimensione  $m \times n$  la trasposta avrà dimensione  $n \times m$ .

$$L \mathbf{y}_t = L \begin{bmatrix} y_{1,t} \\ \dots \\ y_{G,t} \end{bmatrix} = L \begin{bmatrix} Ly_{1,t} \\ \dots \\ Ly_{G,t} \end{bmatrix} = \begin{bmatrix} y_{1,t-1} \\ \dots \\ y_{G,t-1} \end{bmatrix} \quad \text{restando valide le proprietà specificate nel caso}$$

univariato come, ad esempio, l'utilizzo dell'esponente dell'operatore ritardo per aumentare il lag- $k$  del processo, e la proprietà di linearità che sarà così rappresentata; dati due *p.s.* multivariati  $\vec{y}_t, \vec{x}_t$  si avrà che, l'operatore  $L$  legato ad una loro combinazione lineare sarà pari

$$L(\mathbf{A} \cdot \vec{y}_t + \mathbf{B} \cdot \vec{x}_t) = \mathbf{A} \cdot L \vec{y}_t + \mathbf{B} \cdot L \vec{x}_t = \mathbf{A} \cdot \vec{y}_{t-1} + \mathbf{B} \cdot \vec{x}_{t-1} \quad \text{dove } \mathbf{A}, \mathbf{B} \text{ sono matrici numeriche } G \times G\text{-dimensionali.}$$

Proseguendo su questa linea, anche il polinomio nell'operatore ritardo sarà uguale al caso univariato con una dimensionalità aumentata, ovvero, esso sarà dato da

$$\mathbf{A}(L) = \mathbf{I} + \mathbf{A}_1 L + \mathbf{A}_2 L^2 + \dots + \mathbf{A}_g L^g \quad \text{dove, come nel caso univariato, l'esistenza della inversa } \mathbf{A}(L)^{-1} \text{ è garantita nel caso in cui, le soluzioni dell'equazione caratteristica } \det \mathbf{A}(z) = \vec{0} \text{ siano tutte esterne al cerchio unitario sul piano complesso.}$$

Essendo i modelli che verranno utilizzati nel presente lavoro, si espliciteranno in modo essenziale i modelli *ARMA* multivariati che, in tal caso, prendono la denominazione di modelli *VARMA* o anche modelli *ARMA-vettoriali*.

Preliminarmente, si definisce un *W.N.* multivariato una sequenza di variabili aleatorie  $\epsilon_t = (\epsilon_{1,t}, \epsilon_{2,t}, \dots, \epsilon_{G,t})$  non auto-correlate e con valore atteso nullo, ovvero risulterà che  $E(\epsilon_t) = \vec{0}$  e la funzione di auto-covarianza sarà pari a

$$\Gamma_t(K) = \begin{cases} \Sigma, & k=0 \\ \vec{0}, & k \neq 0 \end{cases} .$$

Definito un *W.N.* multivariato allora si può dire che, un *p.s.*-multivariato è un *ARMA(p,q)*-vettoriale o *VARMA* se risulterà stazionario e se soddisferà la equazione ricorsiva  $\mathbf{y}_t = \mathbf{c} + \Phi_1 \cdot \vec{y}_{t-1} + \Phi_2 \cdot \vec{y}_{t-2} + \dots + \Phi_p \cdot \vec{y}_{t-p} + \vec{\epsilon}_t + \Theta_1 \cdot \vec{\epsilon}_{t-1} + \dots + \Theta_q \cdot \vec{\epsilon}_{t-q}$  dove

$$\vec{\epsilon}_t \sim WN(\vec{0}, \Sigma) . \quad \text{Tale espressione può essere rappresentata compattamente come}$$

$$\Phi(L) \vec{y}_t = \vec{c} + \Theta(L) \vec{\epsilon}_t \quad \text{dove le scritture } \Phi(L) = \mathbf{I} - \Phi_1 L - \Phi_2 L^2 - \dots - \Phi_p L^p \text{ e}$$

$\Theta(L) = \mathbf{I} + \Theta_1 L + \Theta_2 L^2 + \dots + \Theta_q L^q$  rappresentano i polinomi degli operatori ritardo in  $\Phi$  e  $\Theta$ .

L'equazione ricorsiva rappresentante il processo *VARMA* sarà un sistema di  $G$  equazioni in  $(p+q) \cdot G^2$  parametri incogniti con  $G$  medie e  $\frac{G \cdot (G+1)}{2}$  varianze e

covarianze di  $\Sigma$ . La proprietà di stazionarietà del processo è garantita se l'equazione caratteristica  $\det(\Phi(z)) = \vec{0}$  ammette soluzioni esterne al cerchio unitario sul piano complesso e se non esistono radici comuni con l'equazione

$$\det(\Theta(z)) = \vec{0} .$$

Come nel caso univariato, anche nel caso multivariato sarà possibile riparametrizzare il processo VARMA come un MA( $\infty$ ) multivariato e, qualora valesse la condizione di invertibilità  $\Theta(L)^{-1}$  allora si potrebbe rappresentare il processo come un AR( $\infty$ ).

Benchè statisticamente tali risultati abbiano valenza tecnica, nella realtà essi sono difficilmente applicabili per studiare fenomeni d'interesse e ciò è dovuta alla eccessiva parametrizzazione che questi modelli hanno.

In aggiunta, la presenza della componente MA riguardante la componente stocastica, rende assai difficile dare una interpretazione del modello legata al fenomeno oggetto di studio. Per ovviare a ciò, si predilige una struttura dinamica di tipo solo autoregressivo denominato modello *autoregressivo vettoriale* di ordine  $p$  oppure semplicemente VAR( $p$ ). In sostanza i processi di tipo VAR( $p$ ) sono un caso particolare della generalizzazione VARMA poichè considerano la componente a media mobile nulla, ovvero pongono  $q=0$ . Ma anche procedendo in questo modo, dovendo a volte considerare un numero elevato di variabili che abbiano il potere di spiegare bene il fenomeno, dunque non tralasciando nessuna variabile utile a far ciò, ci si potrebbe trovare di fronte ad un modello VAR( $p$ ) di dimensione enorme, non differendo molto da un VARMA. Allora sarà necessario effettuare delle scelte, porre delle restrizioni sui parametri, considerare solo delle variabili dette endogene tralasciando le restanti poichè considerate generate separatamente e non dal modello stesso.

Dunque, dato un fenomeno oggetto di studio, si consideri un generico vettore di variabili rilevanti  $M \times 1$  dimensionale  $\vec{w}_t = (w_{1,t}, w_{2,t}, \dots, w_{M,t})$  un vettore di costanti  $\vec{c}$  di dimensione  $M \times 1$ ,  $p$  matrici di parametri  $A_i$  di dimensioni  $M \times M$  con  $i=1, 2, \dots, p$  dove  $p$  rappresenta il numero di ritardi del modello, e dato un vettore  $\vec{\epsilon}_t$  di dimensioni  $M \times 1$  di una componente stocastica *White Noise*, si potrà rappresentare il fenomeno attraverso un VAR( $p$ ) del tipo

$$\vec{w}_t = A_1 \cdot \vec{w}_{t-1} + A_2 \cdot \vec{w}_{t-2} + \dots + A_p \cdot \vec{w}_{t-p} + \vec{c} + \vec{\epsilon}_t .$$

Tale modello potrebbe, come detto prima, risultare ancora assai dispendioso in termini computazionali poichè potrebbero essere contemplate in esso troppe variabili anche non rilevanti ai fini dello studio. Si ipotizzi allora che esistano delle variabili  $\vec{y}_t, \vec{z}_t, \vec{h}_t$  che rappresentino rispettivamente le variabili endogene da spiegare nel modello, le variabili esogene non spiegate nel modello ma che vengono incorporate perché utili a spiegare parte della variabilità delle  $\vec{y}_t$ , ed in fine le restanti fonti di variabilità del fenomeno che potrebbero essere non osservabili o considerate non rilevanti. Dunque si è interessati solo alle variabili  $\vec{z}_t$  per spiegare il fenomeno.

Si dovrà dunque procedere ad una giustificazione plausibile al fine di non considerare nel modello le variabili  $\vec{h}_t$ . Considerando un sistema dinamico in  $t=1, 2, \dots, T$  si potrà definire la legge di probabilità congiunta in termini di una funzione densità di probabilità del tipo  $f(\vec{y}_1, \vec{z}_1, \vec{h}_1, \dots, \vec{y}_T, \vec{z}_T, \vec{h}_T; \vec{\theta})$  fattorizzandola tramite il prodotto della densità marginale di  $y$  e  $z$  ed la densità condizionata di  $h$  rispetto ad  $y$  e  $z$ , ovvero  $f(\vec{y}_1, \vec{z}_1, \dots, \vec{y}_T, \vec{z}_T; \theta_1) \cdot f(\vec{h}_1, \dots, \vec{h}_T / \vec{y}_1, \vec{z}_1, \dots, \vec{y}_T, \vec{z}_T; \theta_2)$  dove  $i$

parametri  $\theta_1, \theta_2$  delle due distribuzioni sono funzioni dell'insieme dei parametri  $\Theta$  della distribuzione congiunta di tutte le variabili del sistema.

Questa fattorizzazione potrebbe, però, avere in se delle variabili che conducono ad una maggiore complessità computazionale senza portare ad una maggiore informazione del fenomeno. Dunque si potrebbe considerare il solo sotto modello marginale trascurando l'informazione contenuta nelle variabili  $h$  poiché non rilevante. Questa possibilità, in linea di massima, potrà essere considerata se si appureranno le seguenti condizioni, ovvero

i) l'imposizione di vincoli sull'insieme dei parametri  $\theta_1$  non implichi anche l'imposizione di vincoli su  $\theta_2$  e viceversa.

Ciò si traduce nel fatto che l'insieme dei parametri  $(\theta_1, \theta_2) \in \Theta_1 \times \Theta_2$  saranno *liberi di variare*, ovvero saranno detti *variation free*.

ii) i parametri d'interesse saranno solo funzione dell'insieme  $\theta_1$  ovvero  $T = \tau(\theta_1)$ .

Tali condizioni esprimono rispettivamente l'idea che, il modello condizionato non fornisca informazioni aggiuntive sui parametri del modello marginale, e che l'attenzione sia, ai fini informativi del fenomeno oggetto di studio, focalizzabile esclusivamente su di una funzione dei parametri del modello marginale.

La verifica simultanea di tali condizioni fa sì che, il considerare il solo modello marginale delle variabili  $\vec{y}_t$  e  $\vec{z}_t$  sia sufficiente per i parametri d'interesse e l'insieme delle variabili  $\vec{z}_t$  sia semplicemente ancillare ai medesimi parametri d'interesse.

Questo processo, detto di *marginalizzazione*, permette di ridurre l'insieme dei parametri a quelli che hanno una valenza informativa per il fenomeno oggetto di studio, riducendo il processo inferenziale ai soli parametri del modello marginale escludendo quelli del modello condizionato.

Il problema che sorge alla luce di queste osservazioni risiede nel capire, in quale modo poter arrivare alla conclusione che alcune variabili, siano ancillari e non fondamentali ai fini informativi nel modello usato.

In aiuto, per far ciò, torna utile il concetto di *esogenità* di variabili, ovvero, il considerare tali variabili come date e non generate dal modello.

La letteratura individua tre diverse tipologie di esogenità a seconda del ruolo che il modello statistico assume nella ricerca.

Ai fini del presente lavoro si espliciteranno le prime due poiché l'ultima è fondamentalmente legata allo studio di fenomeni economico-finanziari anche se, con opportuni aggiustamenti, può essere adottata anche per altre tipologie di fenomeni.

Il filo conduttore comune delle diverse definizioni di *esogenità* risiede nel legame che esse hanno con l'insieme dei parametri d'interesse.

## ESOGENITA' DEBOLE

Il primo tipo di *esogenità* assume il ruolo di inferenza sui parametri d'interesse di un dato modello usato per analizzare il fenomeno oggetto di studio.

Sia un vettore  $\vec{w}_t = \begin{pmatrix} \vec{y}_t \\ \vec{z}_t \end{pmatrix}$  di variabili considerate rilevanti per il fenomeno oggetto di studio.

Sia una funzione densità congiunta che esprima la evoluzione del fenomeno

$$f(\vec{y}_1, \vec{z}_1, \dots, \vec{y}_T, \vec{z}_T; \theta_1) = \prod_{t=1}^T f(\vec{y}_t, \vec{z}_t / \vec{y}_{t-1}, \vec{z}_{t-1}, \dots, \vec{y}_0, \vec{z}_0; \theta_1)$$

dove l'insiemi  $\vec{y}_0, \vec{z}_0$  rappresentino le condizioni iniziali del fenomeno osservabili e non stocastiche tali che  $f(\vec{y}_1, \vec{z}_1 / \vec{y}_0, \vec{z}_0; \theta_1) = f(\vec{y}_1, \vec{z}_1; \theta_1)$ , e sia una fattorizzazione della suddetta funzione del tipo

$$\prod_{t=1}^T f(\vec{y}_t / \vec{z}_t, \vec{y}_{t-1}, \vec{z}_{t-1}, \dots, \vec{y}_0, \vec{z}_0; \lambda_1) \times \prod_{t=1}^T f(\vec{z}_t / \vec{z}_{t-1}, \vec{y}_{t-1}, \dots, \vec{y}_0, \vec{z}_0; \lambda_2)$$

in cui gli insiemi dei parametri  $\lambda_1, \lambda_2$  sono una generica trasformazione dell'insieme  $\theta_1$ . In tale fattorizzazione, nella prima produttoria la variabile  $y$  dipende, oltre che dalla sua storia passata, anche dal valore presente e dalla storia passata di  $z$ . Nella seconda produttoria invece, la variabile  $z$  dipende dalla sua storia passata e dalla sola storia passata della variabile  $y$ .

Tale fattorizzazione fatta della funzione densità congiunta viene definito *taglio sequenziale* se risulterà che gli insiemi  $(\lambda_1 \times \lambda_2) \in \Lambda_1 \times \Lambda_2$  ovvero sono *variation free*. In tal caso, non vi sarà nessun vincolo o legame tra i due insiemi dei parametri e se risulterà anche che i parametri d'interesse saranno soltanto una funzione dell'insieme  $\lambda_1$  allora il modello condizionato sarà sufficiente per i parametri d'interesse e si potrà usare solo questo per condurre l'analisi tralasciando la seconda produttoria della precedente fattorizzazione.

Date queste premesse, si dirà che il processo  $\vec{z}_t$  è detto debolmente esogeno per un insieme di parametri d'interesse  $\tau$  se la fattorizzazione di cui sopra è un taglio sequenziale e se risulterà che i parametri d'interesse sono una funzione dei soli parametri del modello condizionato, ovvero  $T = \tau(\lambda_1)$ .

Allora, se varrà la condizione di esogenità delle variabili  $\vec{z}_t$ , basare un processo inferenziale per l'insieme  $T = \tau(\lambda_1)$  esclusivamente sul modello condizionato di  $\vec{y}_t$  non implicherà perdita di informazione per il fenomeno oggetto di studio ed in tal caso, le variabili  $\vec{z}_t$  saranno considerate fisse e non spiegate dal modello.

Questo implicherà che l'inferenza effettuata sul solo modello condizionato condurrà ad una riduzione della complessità computazionale poiché riuscirà a dare gli stessi risultati in termini informativi, con uno sforzo computazionale inferiore rispetto ad una inferenza effettuata sul modello completo.

## CAUSALITA' ED ESOGENITA' FORTE

La seconda tipologia di *esogenità* riguarda la funzione di predittività che il modello usato dovrà assolvere nello studio del fenomeno sotto analisi.

Nel primo tipo di esogenità, si è visto come la variabile  $y$  sia stata fatta dipendere dai suoi valori passati e dai valori passati e presenti di  $z$ , ma non si è considerata la possibilità che  $z$  dipenda dai valori passati di  $y$  e che dunque non sia possibile prevedere  $y$  considerando  $z$  fissa.

Tale problema non si pone se  $z$  non dipende dalla storia passata di  $y$ ; in tal caso si potrà scomporre il processo predittivo nella parte relativa a  $z$  condizionatamente alla sua storia passata, ed in una parte relativa alla previsione di  $y$  sulla base del modello condizionato.

Tale scomposizione del processo predittivo è assicurata dalla condizione detta *causalità di Granger* o *Granger causalità*, ovvero  $y$  non causa  $z$  se il set informativo di  $y$  non migliora la previsione di  $z$ .

Si nota la totale indipendenza di tale condizione dall'insieme dei parametri d'interesse, che possono essere qualunque, e la sola dipendenza dalla distribuzione congiunta delle variabili del sistema.

In termini di funzione densità di probabilità si può rappresentare la *Granger causalità* dicendo che,  $y$  non causa  $z$  nel senso di Granger se e solo se  $f(\vec{z}_t / \vec{y}_{t-1}, \vec{z}_{t-1}, \dots, \vec{y}_1, \vec{z}_1; \lambda_2) = f(\vec{z}_t / \vec{z}_{t-1}, \vec{z}_{t-2}, \dots, \vec{z}_1; \lambda_2)$ .

Si noti che, dalla suddetta definizione, la *Granger causalità* non implica né è implicata dalla esogenità debole poiché, quest'ultima, dipende dalla scelta dei parametri d'interesse da cui la *Granger causalità* ne è indipendente essendo essa, invece, una condizione non vincolata dal modello statistico scelto (fattore che invece influenza la *esogenità debole*) ma riguardante il processo che ha generati i dati.

Tali due condizioni considerate simultaneamente sulle variabili di interesse, conducono al concetto di *esogenità forte*, la quale afferma che, dato l'insieme delle variabili  $z$ , se si verifica la condizione di esogenità debole su di esso e le variabili che lo compongono sono non causate da  $y$  secondo Granger, allora si dice che l'insieme delle variabili  $z$  è fortemente esogeno per l'insieme dei parametri di interesse.

Da quanto detto si può concludere che, una variabile di interesse può essere considerata esterna al modello ai fini della stima dei parametri d'interesse ma, contemporaneamente, possa essere causata dalle variabili endogene ma potrà esser vero anche il viceversa, ovvero non essere causata dalle endogene e non per questo risultare debolmente esogena.

Tutti questi concetti si basano sull'assunzione che ci si trovi di fronte alla conoscenza della funzione densità di probabilità congiunta nota a priori.

Non essendo questo lavoro finalizzato a giustificare questa ultima osservazione, si accennerà in modo succinto a questo problema dando delle linee guida minimali.

Partendo dal fenomeno oggetto di analisi, l'utilizzo di un dato modello rispetto ad un altro è legato al concetto di *corretta identificazione di un modello*, e precisamente, se alla base del modello stimato vi è un paradigma teorico corretto riguardante il

fenomeno sotto analisi, allora il significato della analisi inferenziale sarà univoco nel senso che, la relazione stimata non sarà ottenibile con un modello diverso da quello identificato. Se invece la relazione ottenuta col modello specificato può essere ottenuta con altri modelli teorici, allora il modello specificato non è correttamente identificato. Dunque, alla base di una corretta identificazione di un modello, vi è l'utilizzo di un corretto paradigma teorico che regoli le relazioni tra le variabili che entrano in gioco per l'analisi del fenomeno oggetto di studio.

Se, ad esempio, si stimasse un modello in cui un parametro potesse essere espressione della variabilità di una componente del fenomeno ma, allo stesso tempo, potesse esserlo anche per un'altra componente del fenomeno, allora non si sarebbe in grado di dare un significato univoco tra il modello specificato ed i dati utilizzati. In questo caso il modello sarebbe del tutto inutilizzabile poiché non permetterebbe di dare un senso chiaro alla variabilità del fenomeno e sarebbe non consono a spiegare il paradigma teorico alla base della definizione delle relazioni tra le variabili che lo rappresentano. Detto questo, considerando allora un dato modello scelto per l'analisi, esso si baserà sulla funzione di probabilità congiunta delle variabili che entrano in gioco nel fenomeno studiato.

Tale funzione sarà rappresentata da un insieme di parametri ignoti, appartenenti ad uno spazio parametrico.

Considerando un campione di  $T$  numerosità delle suddette variabili, si potrà esprimere la funzione di probabilità come fatto sin'ora, ovvero una funzione del tipo  $f(\vec{y}_1, \vec{z}_1, \dots, \vec{y}_T, \vec{z}_T; \theta)$  indicizzata all'insieme dei parametri d'interesse con le fattorizzazioni viste prima.

Considerando tale campione, la funzione di verosimiglianza, come già detto, avrà la caratteristica di essere proporzionale<sup>12</sup> alla suddetta funzione di densità a meno di una costante moltiplicativa, dunque si avrà che

$L(\theta) \propto f(\vec{y}_1, \vec{z}_1, \dots, \vec{y}_T, \vec{z}_T; \theta)$ . Ai fini di una corretta identificazione del modello statistico da usare, ci si aspetterà che, a diversi valori dei parametri di interesse dello spazio parametrico di riferimento, corrispondano diversi valori della funzione di densità di probabilità e, conseguentemente, della rispettiva funzione di verosimiglianza. Dati dunque due insiemi dei parametri di interesse  $(\theta, \theta^*) \in \Theta$  con la condizione che  $\theta \neq \theta^*$ , un modello sarà *identificato* se risulterà che:

$$f(\vec{y}_1, \vec{z}_1, \dots, \vec{y}_T, \vec{z}_T; \theta) \neq f(\vec{y}_1, \vec{z}_1, \dots, \vec{y}_T, \vec{z}_T; \theta^*) \Leftrightarrow L(\theta) \neq L(\theta^*)$$

Invece, dati  $(\theta, \theta^*) \in \Theta$  con  $\theta \neq \theta^*$ , se si avrà che  $L(\theta) = L(\theta^*)$  allora si dirà che i due insiemi di parametri sono *osservazionalmente equivalenti*.

Si può sfruttare questa ultima definizione dicendo che, un modello statistico per delle variabili di interesse  $y, z$  sarà correttamente identificato se  $\exists(\theta, \theta^*) \in \Theta$  con  $\theta \neq \theta^* \ni L(\theta) = L(\theta^*)$  ovvero, se gli insiemi dei parametri di interesse non sono *osservazionalmente equivalenti*.

12 Date due variabili  $y, x$  esse saranno dette proporzionali se esisterà una relazione del tipo  $y = k \cdot x$  con  $k \neq 0$ .

Altri modi di rappresentare la proporzionalità tra variabili sono  $y \sim x$  oppure  $y = \text{cost}(x)$ .

Questo accade poiché, se esistono insieme di parametri *osservazionalmente equivalenti*, sulla base della funzione di verosimiglianza non si sarà in grado di distinguere a quale insieme dei parametri è riferita l'inferenza e quindi, il modello non risulterà correttamente identificato.

Allora, quando si procede ad identificare il paradigma concettuale che regola le relazioni tra le variabili oggetto di analisi, ovvero quando si procede alla *specificazione di un modello*, tale specificazione dovrà garantire che il modello risultante sia anche *correttamente identificato*, diversamente si dovrà rivedere il framework concettuale sul quale si è basata la specificazione del modello.

Tutti questi concetti si vanno ad inserire in un insieme di azioni che, a seconda dello studio che si sta effettuando, l'analista andrà ad eseguire e grazie alle quali si potranno raggiungere risultati tali che si potrà definire un modello statistico idoneo a far emergere il più possibile le vere leve che regolano il fenomeno studiato. Nello studio di fenomeni dinamici, la definizione e la verifica di alcune caratteristiche delle traiettorie dei dati che descrivono il fenomeno, sono di importanza fondamentale per poter effettuare una corretta analisi.

Si esporranno di seguito le caratteristiche dinamiche che i fenomeni reali hanno ed alcuni modelli che verranno utilizzati per lo studio del fenomeno oggetto del presente lavoro. Una classe di processi dinamici a cui si è già accennato è quella dei processi  $VAR(p)$ , introdotti per sopperire alla complessità computazionale che si riscontra nei modelli  $ARMA(p,q)$  multivariati ovvero  $VARMA(p,q)$ .

Sia un processo  $VAR(p)$   $\vec{w}_t = \mathbf{A}_1 \cdot \vec{w}_{t-1} + \mathbf{A}_2 \cdot \vec{w}_{t-2} + \dots + \mathbf{A}_p \cdot \vec{w}_{t-p} + \vec{c} + \vec{\epsilon}_t$  visto precedentemente.

Esso può essere rappresentato nel seguente modo:  $\mathbf{A}(L)\vec{w}_t = \vec{c} + \vec{\epsilon}_t$  dove

$$\mathbf{A}(L) = \mathbf{I}_M - \sum_{i=1}^p \mathbf{A}_i L^i \text{ con } L \text{ che rappresenta il solito operatore ritardo, le } \mathbf{A}_i$$

sono matrici di coefficienti  $M \times M$  e  $\vec{c}, \vec{\epsilon}_t$  sono rispettivamente insiemi di componenti deterministiche e di variabili casuali che, in questo caso, sono considerate *i.i.d.* con varianza  $\Lambda = \sigma_\epsilon^2$ . Se tale ultima componente avrà una distribuzione di tipo Gaussiano allora si dirà che il  $VAR(p)$  è Gaussiano.

Un esempio molto usato nella pratica di componenti deterministiche sono i *trend deterministici*. Una proprietà importante dei processi  $VAR(p)$  è quella di essere chiusi (stabili) a trasformazioni lineari invertibili delle variabili, questo si traduce nel fatto che il processo  $\vec{w}_t$  visto prima, continuerà ad essere un  $VAR(p)$  anche se si effettuerà una sua trasformazione lineare del tipo  $\vec{w}_t^* = \mathbf{D} \cdot \vec{w}_t$  con  $\mathbf{D}$  matrice quadrata invertibile. Se tale proprietà risulterà vera, allora, la scelta del sistema di coordinate sarà indifferente e questo si tradurrà nel fatto che, la scelta delle variabili oggetto di studio  $\vec{w}_t^*$  o  $\vec{w}_t$  non muterà il senso della analisi a patto che tra le variabili vi siano delle relazioni lineari.

A questa proprietà è legato il concetto di *co-integrazione* tra processi stocastici, concetto potente statisticamente per effettuare previsioni di traiettorie e che sarà discusso in dettaglio in questo capitolo poiché sarà la metodologia statistica utilizzata per effettuare la parte applicativa, in unione ad un altro metodo di analisi.

Per esporre il concetto di co-integrazione si esporrà brevemente prima quello di *integrazione* a cui si è già accennato in precedenza.

Si consideri il processo  $ARMA(p,q)$  visto precedentemente nella sua rappresentazione  $MA(\infty)$  del tipo  $y_t = \mu + \Psi(L) \cdot \varepsilon_t$  dove il valore atteso del processo non è considerato nullo bensì è costante e pari a  $\mu$ .

Come detto, l'assenza di radici unitarie rende il processo invertibile e dunque stazionario. Se in tale rappresentazione, però, modifichiamo il valore atteso del processo rendendolo non costante, si può notare una differenza di proprietà del processo. Si ipotizzi dunque che  $y_t = \mu + \delta \cdot t + \Psi(L) \cdot \varepsilon_t$  sia il nuovo processo in cui, il valore atteso è stato rimpiazzato dalla funzione  $\mu + \delta \cdot t$  deterministica ed evolutiva nel tempo, che per semplicità consideriamo essere un trend lineare.

Un processo siffatto si definisce *trend-stationary* e rappresenta una prima tipologia di non stazionarietà; esso è non stazionario in covarianza e ciò dipende esclusivamente dalla componente deterministica. Il concetto non cambierebbe se si fosse scelto un trend deterministico rappresentato da una funzione polinomiale di grado superiore al primo. La componente di disturbo del processo, ovvero la sua componente stocastica, fluttua intorno al trend deterministico. Oltre a questa tipologia di non stazionarietà ve ne è una seconda rappresentata da processi a radici unitarie come  $(1-B) \cdot y_t = \mu + \psi(B) \cdot \varepsilon_t$ . Dato che  $(1-L) \cdot y_t = y_t - Ly_t = y_t - y_{t-1}$  allora si potrà scrivere che  $y_t - y_{t-1} = \mu + \psi(B) \cdot \varepsilon_t$  e ricordando che  $\psi(B) \cdot \varepsilon_t$  è un processo  $ARMA(p,q)$  stazionario del secondo ordine, si potrà dire che il processo a radice unitaria  $(1-B) \cdot y_t = \mu + \psi(B) \cdot \varepsilon_t$  è stazionario in covarianza nelle sue differenze. Un tale processo è detto *difference-stationary*.

Evitando di entrare nei dettagli poiché non è la finalità del presente lavoro, la specificazione di queste due tipologie di stazionarietà è legata alla metodologia attraverso la quale rendere un processo stazionario, quando non lo è.

Per i processi *trend-stationary* tale risultato si ottiene sottraendo al processo il suo trend deterministico.

Per i processi *difference-stationary*, la presenza di radici unitarie fa sì che la differenziazione del processo lo trasformi in un processo stazionario.

Le differenze che rendono un processo stazionario possono anche essere di grado superiore al primo. In generale, un processo le cui differenze di ordine  $d$  lo rendono stazionario è detto *Integrato di ordine  $d$*  e lo si indica con  $y_t \sim I(d)$ .

Se la differenziazione di ordine  $d$  genera un processo stazionario del tipo  $ARMA(p,q)$  allora si dirà che il processo è *autoregressivo* di ordine  $p$  *integrato* di ordine  $d$  ed a *media mobile* di ordine  $q$  e si scriverà che

$$y_t = ARIMA(p, d, q) \text{ dove } d=1,2,\dots$$

Un esempio tipico di processo integrato è il processo *Random Walk* con o senza deriva (drift) visto precedentemente.

Sulla base di esso si può rappresentare un piccolo sistema di serie storiche integrate di ordine  $d$  generico, che godano di una caratteristica particolare; ovvero, essere tra loro *co-integrate*. Allora, siano  $X, Y$  due soggetti, sia  $t=0,1,2,\dots$  una misura temporale. Si ipotizzi che i due soggetti partano insieme da uno stesso luogo e che la distanza da tale luogo alle loro case, in linea d'aria, sia data rispettivamente da  $z_t$  per  $X$  e  $s_t$  per  $Y$ , considerando comune la direzione delle loro abitazioni.

Partendo dallo stesso luogo e nello stesso istante  $t=0$ , le distanze tra loro ed il punto di partenza sarà nullo per entrambi, ovvero  $z_0 = s_0 = 0$ .

Si indichino con  $u_t, v_t$  gli spostamenti di  $X$  e  $Y$  nell'istante  $t$ .

Si assuma che  $X$  segua un *RW* unidimensionale non simmetrico ma che esista un *drift* o *deriva* pari a  $\phi$ . La distanza di  $X$  dal luogo di partenza in ogni istante sarà dunque pari a  $z_t = z_{t-1} + \phi + u_t$ . Si consideri, adesso, che il soggetto  $X$  accompagni sottobraccio  $Y$  in modo tale che, prima di ogni passo, lo attiri a se facendo sì che il punto di partenza di  $Y$ , in ogni istante  $t$ , sia proporzionale a quello di  $X$ , ovvero la distanza di  $Y$  dal luogo di partenza ad ogni istante  $t$  sarà data da  $s_t = \gamma \cdot z_{t-1} + v_t$  dove  $\gamma$  è la costante di proporzionalità tra  $Y$  ed  $X$ .

Se, ad esempio,  $\gamma=1$  in ogni istante temporale  $t$  entrambi i soggetti si muovono dalla stessa posizione di partenza di  $X$  pari a  $z_{t-1}$ .

Si ponga  $\vec{\epsilon}_t = (u_t, v_t)$  e che  $\epsilon_t \sim IID(0, \Lambda)$  dove  $\Lambda = [\sigma_{u,v}]$  rappresenta la matrice di varianza/covarianza rispettivamente di  $u_t, v_t$  positiva definita con  $\sigma_u^2, \sigma_v^2$  rispettivamente la varianza di  $u$  e  $v$ <sup>13</sup>.

Il sistema può essere formalizzato in forma matriciale nel seguente modo

$$\begin{pmatrix} z_t \\ s_t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \gamma & 0 \end{pmatrix} \cdot \begin{pmatrix} z_{t-1} \\ s_{t-1} \end{pmatrix} + \begin{pmatrix} \phi \\ 0 \end{pmatrix} + \begin{pmatrix} u_t \\ v_t \end{pmatrix}$$
 ed effettuando le rispettive assegnazioni si avrà

che, il sistema diventerà  $\vec{w}_t = A_1 \cdot \vec{w}_{t-1} + \vec{\mu} + \vec{\epsilon}_t$  dove  $\vec{w}_t = \begin{pmatrix} z_t \\ s_t \end{pmatrix}$ ,

$$A_1 = \begin{pmatrix} 1 & 0 \\ \lambda & 0 \end{pmatrix}, \quad \vec{w}_{t-1} = \begin{pmatrix} z_{t-1} \\ s_{t-1} \end{pmatrix}, \quad \vec{\mu} = \begin{pmatrix} \phi \\ 0 \end{pmatrix} \text{ ed } \vec{\epsilon}_t = (u_t, v_t) \text{ già assegnato}$$

precedentemente. Tale semplice sistema definisce un sistema *VAR(1)* ed il sistema  $\vec{w}_t$  è detto *rappresentazione AR*.

A tale sistema si può associare  $A(L) = I - A_1 L$  detto polinomio autoregressivo del sistema  $\vec{w}_t$ .

Si noti che il processo in esempio sia non-stazionario dato che le sue componenti seguono un *RW* non simmetrico.

13 Data una matrice quadrata  $M$  di ordine  $n$  e dato un vettore riga  $\vec{x}$  di dimensione  $n$  (matrice  $1 \times n$ ) si dice che  $M$  è definita positiva se  $\forall \vec{x} \in \mathbb{C}_0^n$  risulterà che  $\vec{x} \cdot M \cdot \vec{x}' > 0$  con  $\vec{x} \neq \emptyset$ ;  $\vec{x} \cdot M \cdot \vec{x}' > 0$  sarà dunque un numero positivo.

Concentrando l'attenzione sulla strada percorsa da  $X$ , ovvero sulla equazione

$z_t = z_{t-1} + \phi + u_t$ , questa può essere rappresentata ricorsivamente ottenendo che

$$z_t = \sum_{i=1}^t (u_i + \phi) + z_0 ; \text{ essendo } z_0 = 0 \text{ ed effettuando la seguente assegnazione si}$$

ha che  $z_t = \sum_{i=1}^t (u_i + \phi) = g_t$  dove  $g_t$  segue un RW con drift dato da  $t \cdot \phi$  poiché

$$g_t = \sum_{i=1}^t u_i + t \cdot \phi .$$

Tale processo è non-stazionario in covarianza poiché il suo momento secondo è dipendente da  $t$ , infatti  $\sigma_{g_t}^2 = t \cdot \sigma_u^2$ .

Dalla struttura del processo  $g_t$  si può dedurre una proprietà di un RW e precisamente si può dire che, nel tempo, un RW domina una componente stazionaria come un *i.i.d.* in termini di varianza, infatti, se si crea una combinazione lineare dei processi qui definiti  $g_t, v_t$  del tipo  $g_t + \alpha \cdot v_t$  dove  $\alpha \in \mathbb{C}_0$ , con varianza  $V(g_t + \alpha \cdot v_t) = t \cdot \sigma_u^2 + \alpha^2 \cdot \sigma_v^2 + \alpha \cdot \sigma_{u,v}$ , si nota subito che, per  $t \rightarrow \infty$  la varianza della componente  $g_t$  non stazionaria domina quella del processo *i.i.d.*.

Si può dimostrare che una passeggiata aleatoria, sia essa simmetrica o con *drift*, sia un processo che può generare dei trend stocastici e per tale motivo essa è anche detta *trend stocastico*. Tornando al sistema  $\vec{w}_t$ , sulla base della rappresentazione del processo  $z_t$  si può esplicitare il secondo processo del sistema come

$$s_t = \gamma \cdot z_{t-1} + v_t = \gamma \cdot g_t - \gamma \cdot \phi + (v_t - \gamma \cdot u_t) .$$

Questo è possibile poiché essendo  $z_t = z_{t-1} + \phi + u_t$  ed avendo dimostrato che  $z_t = g_t$  allora si ha che  $z_{t-1} = z_t - (\phi + u_t)$  e sostituendo si ottiene che  $s_t = \gamma \cdot [g_t - (\phi + u_t)] + v_t = \gamma \cdot g_t - \gamma \cdot \phi + (v_t - \gamma \cdot u_t)$ .

Queste manipolazioni algebriche, lasciano invariati i processi facenti parte del sistema, ma mettono in evidenza una caratteristica comune agli stessi processi, e precisamente evidenziano il fatto che entrambi condividono nella loro struttura una componente comune dalla quale essi dipendono; la componente *non-stazionaria*  $g_t$ . Tale componente viene infatti definita *trend comune*.

Essendo  $g_t$  un processo  $I(1)$  ovvero integrato di ordine 1, è per costruzione non-stazionario.

Il sistema derivante  $w_t$  è detto perciò *co-integrato* ad indicare che tale componente integrata è comune ai processi del sistema.

E' possibile dare una rappresentazione compatta del sistema

$$\vec{w}_t = \begin{pmatrix} 1 \\ \gamma \end{pmatrix} \cdot g_t + \begin{pmatrix} 0 \\ -\gamma \cdot \phi \end{pmatrix} + \tilde{\epsilon}_t \text{ del tipo } \vec{w}_t = \vec{\xi} \cdot g_t + \vec{k} + \tilde{\epsilon}_t \text{ dove le assegnazioni fatte}$$

sono evidenti considerando il sistema stesso.

Tale ultima rappresentazione è detta a *trend comuni* ed essa scompone il processo  $\vec{w}_t$  in una parte non stazionaria comune  $g_t$  ed in una parte stazionaria  $\vec{k} + \vec{\epsilon}_t$  rappresentante un processo *IID*.

La componente stazionaria  $\vec{\xi} \cdot g_t$  è detta *permanente*, mentre quella stazionaria  $\vec{k} + \vec{\epsilon}_t$  è detta *transitoria*.

Nella rappresentazione fatta sopra, il processo  $g_t$  rappresenta la passeggiata casuale che accomuna i due soggetti  $X$  ed  $Y$ .

La componente  $\vec{\xi} = \begin{pmatrix} 1 \\ \lambda \end{pmatrix}$  viene chiamata *matrice dei caricamenti del trend comune* e specifica i pesi coi quali il processo  $g_t$  influenza ciascuna variabile del sistema  $\vec{w}_t$ . Per la dominanza in varianza di una componente non-stazionaria su una stazionaria discussa prima, il trend comune  $g_t$  del sistema domina la componente stazionaria nel lungo periodo, allora per cardinalità campionaria assai grande si ha che  $s_t \approx \lambda \cdot z_t$  e questo si traduce nella caratteristica che, i livelli dei processi  $z_t, s_t$ , tenderanno a disporsi lungo la retta  $\vec{\xi} \cdot c$  dove  $c \in \mathbb{C}_0$ . L'insieme  $\vec{\xi}$  viene definito per questo *attrattore* del sistema.

A questo punto, si può mettere in evidenza sfruttando questo piccolo esempio, una caratteristica di tali processi, ovvero quella di essere, come detto all'inizio dell'esempio, co-integrati.

Infatti, si è visto come i singoli processi che rappresentano il sistema siano processi integrati poiché non stazionari.

Adesso, si consideri la matrice  $2 \times 1$   $\vec{\xi} = \begin{pmatrix} 1 \\ \lambda \end{pmatrix}$  che ha *rango* 1.

Essa potrà allora essere annullata pre-moltiplicandola per un qualunque vettore riga (matrice  $1 \times n$ )  $n$ -dimensionale del tipo  $\vec{\xi}'_{\perp} = a \cdot (\lambda, -1)$  con  $a \in \mathbb{C}_0$ , ottenendo  $\vec{\xi}'_{\perp} \cdot \vec{\xi} = 0$ .

Effettuando la operazione di pre-moltiplicazione del sistema  $\vec{w}_t$  con  $\vec{\xi}'_{\perp}$  si avrà che  $\vec{\xi}'_{\perp} \cdot \vec{w}_t = \vec{\xi}'_{\perp} \cdot (\vec{\xi} \cdot g_t) + \vec{\xi}'_{\perp} \cdot (\vec{k} + \vec{\epsilon}) = 0 \cdot g_t + \vec{\xi}'_{\perp} \cdot (\vec{k} + \vec{\epsilon})$ .

Il risultato di tale operazione, lecita matematicamente, sarà il processo  $\vec{\xi}'_{\perp} \cdot (\vec{k} + \vec{\epsilon})$  stazionario in covarianza.

Nel caso dei due soggetti portato per questo esempio, tale processo sarà dato da  $\vec{\xi}'_{\perp} \cdot (\vec{k} + \vec{\epsilon}) = a \cdot (\gamma \cdot \phi + \gamma \cdot u_t - v_t)$  palesemente stazionario in covarianza per le assunzioni fatte sulle componenti stocastiche  $u_t$  e  $v_t$ .

Questo risultato è caratteristica precipua dei processi *co-integrati*, ovvero *esisterà sempre almeno una* (ma in realtà ne esisteranno infinite) *combinazione lineare stazionaria in covarianza non banale delle variabili del sistema*.

I coefficienti non banali della combinazione lineare stazionaria sono detti *vettori di co-integrazione*.

Questo risultato genera il fulcro della forte utilità che tali processi hanno statisticamente, ossia effettuare ottime previsioni delle variabili del sistema sfruttando la caratteristica dei processi stazionari in covarianza, ovvero il ritorno al valore di equilibrio dato dal loro valore atteso.

Come detto, esisteranno infinite combinazioni lineari stazionarie che, nell'esempio portato sopra, si otterranno al variare di  $a \in \mathbb{C}_0$  e dunque esisteranno infiniti vettori di *co-integrazione*.

In generale, si può indicare l'insieme dei vettori di cointegrazione come lo spazio vettoriale  $\mathbf{B} = \{ \beta_i \exists' \beta_i' \cdot \vec{\xi} = 0 \}$  dove  $\beta_i'$  è un vettore riga<sup>14</sup>.

Nel nostro esempio, si ha che  $\beta_i' = a \cdot (\gamma, -1)$  ed al variare del valore  $a$  si avrà un vettore di co-integrazione i-esimo diverso.

La rappresentazione dello spazio di *co-integrazione*  $\mathbf{B}$  sarà possibile perciò, nel caso portato come esempio, utilizzando una sola base rappresentata da  $(\gamma, -1)$  ed ogni vettore di co-integrazione sarà ottenibile attraverso una combinazione lineare di essa al variare dello scalare  $a \in \mathbb{C}_0$ .

La rappresentazione dei sistemi cointegrati come  $\vec{w}_t = \mathbf{A}_1 \cdot \vec{w}_{t-1} + \vec{\mu} + \vec{\epsilon}_t$  evidenzia la presenza di processi integrati in un  $VAR(p)$ .

Esistono altre rappresentazioni di tali sistemi diverse da quella  $VAR$  che hanno la finalità di evidenziare delle diverse caratteristiche di tali sistemi.

Una possibile di queste è quella a *Correzione dell'Errore* (*Error Correction Mechanism*) ottenibile nel seguente modo.

Sia  $\vec{w}_t = \mathbf{A}_1 \cdot \vec{w}_{t-1} + \vec{\mu} + \vec{\epsilon}_t$  e si sottragga al termine  $\vec{w}_t$  e  $\vec{w}_{t-1}$  il termine

$\vec{w}_{t-1}$  ottenendo  $\vec{w}_t - \vec{w}_{t-1} = \mathbf{A}_1 \cdot \vec{w}_{t-1} - \vec{w}_{t-1} + \vec{\mu} + \vec{\epsilon}_t$ . Adesso, dato che

$\mathbf{A}_1 \cdot \vec{w}_{t-1} - \vec{w}_{t-1} = \vec{w}_{t-1} \cdot [ \begin{pmatrix} 1 & 0 \\ \gamma & 0 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} ] = \vec{w}_{t-1} \cdot [ \begin{pmatrix} 0 & 0 \\ \gamma & -1 \end{pmatrix} ]$  si ottiene che

$\Delta \vec{w}_t = (\mathbf{A}_1 - \mathbf{I}) \cdot \vec{w}_{t-1} + \vec{\mu} + \vec{\epsilon}_t = \begin{pmatrix} 0 & 0 \\ \gamma & -1 \end{pmatrix} \cdot \vec{w}_{t-1} + \vec{\mu} + \vec{\epsilon}_t$  e ponendo

$\Pi = \begin{pmatrix} 0 & 0 \\ \gamma & -1 \end{pmatrix}$  la rappresentazione più compatta sarà data da

$\Delta \vec{w}_t = (\Pi) \cdot \vec{w}_{t-1} + \vec{\mu} + \vec{\epsilon}_t$ , con la matrice  $\Pi$  detta *matrice d'impatto* associata all'operatore ritardo  $\mathbf{A}(L)$ .

Una osservazione fattibile sulla *matrice d'impatto* è che avendo la prima riga nulla, il suo rango sarà minore della sua dimensione e sarà pari ad 1, dunque sarà di rango ridotto. Essa potrà allora essere scomposta nel prodotto dei vettori  $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$  e

$(\gamma, -1)$  ovvero in vettori che hanno rango 1.

<sup>14</sup> La esplicitazione di uno spazio vettoriale esula dalla finalità di questo lavoro e non può essere inserita come nota vista la lunghezza d'esposizione, si rimanda dunque ad un qualunque testo di algebra lineare per una delucidazione di tale oggetto matematico.

Non sfugge il fatto che  $(\gamma, -1)$  è la base scelta per  $\mathbf{B}$ .

Una siffatta situazione in cui vi è una *matrice d'impatto* di rango ridotto delle sue dimensioni è una precipuità dei sistemi *VAR(p) co-integrati*.

Questo concetto sarà meglio esplicitato nelle prossime righe di questo paragrafo.

Per la fattorizzazione effettuata della matrice d'impatto  $\Pi = \begin{pmatrix} 0 & 0 \\ \gamma & -1 \end{pmatrix}$  si può porre

$\vec{\alpha} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$  come fattore che precede la base  $\vec{\beta} = (\gamma, -1)$  ottenendo la rappresentazione del sistema  $\Delta \vec{w}_t = (\Pi) \cdot \vec{w}_{t-1} + \vec{\mu} + \vec{\epsilon}_t$  nella nuova forma

$\Delta \vec{w}_t = \vec{\alpha} \cdot \vec{\beta}' \cdot \vec{w}_{t-1} + \vec{\mu} + \vec{\epsilon}_t$  la quale prende il nome di *rappresentazione a correzione dell'errore* o anche *Error Correction Mechanism*.

Attraverso piccole assegnazioni di variabili, tale denominazione può essere motivata riscrivendo  $\Delta \vec{w}_t = \vec{\alpha} \cdot \vec{\beta}' \cdot \vec{w}_{t-1} + \vec{\mu} + \vec{\epsilon}_t$  come  $\Delta \vec{w}_t = \vec{\alpha} \cdot (\vec{\beta}' \cdot \vec{w}_{t-1} - \vec{\mu}_0) + \vec{\tau} + \vec{\epsilon}_t$  avendo posto  $\vec{\mu} = -\vec{\alpha} \cdot \vec{\mu}_0 + \vec{\tau}$  con  $\vec{\mu}_0 = E(\vec{\beta}' \cdot \vec{w}_{t-1})$ .

Tale scrittura, non modifica il sistema di partenza per via delle assegnazioni fatte nel rispetto degli assunti iniziali e mette in evidenza lo scostamento della combinazione lineare stazionaria  $\vec{\beta}' \cdot \vec{w}_{t-1}$  dal proprio valore atteso  $\vec{\mu}_0$  considerando il termine  $(\vec{\beta}' \cdot \vec{w}_{t-1} - \vec{\mu}_0)$  come il valore di equilibrio o obiettivo a cui il sistema deve tendere.

Dunque la *rappresentazione a correzione dell'errore* mette in evidenza la reazione di  $\Delta \vec{w}_t$  a deviazioni dall'obiettivo  $(\vec{\beta}' \cdot \vec{w}_{t-1} - \vec{\mu}_0)$  o *errore*, per mezzo dei termini di aggiustamento  $\vec{\alpha}$ .

Per chiarire tali concetti, si consideri il caso dei due soggetti che partono da un punto e si dirigono verso le proprie abitazioni.

In tale situazione, il termine  $(\vec{\beta}' \cdot \vec{w}_{t-1} - \vec{\mu}_0)$  equivale ad  $\gamma \cdot z_t - s_t$  ovvero la distanza tra i due soggetti che tendono a riportare a zero avendo detto che  $A$  attira a se  $B$  prima di ogni passo.

La rappresentazione a *correzione dell'errore* mette in risalto allora quali delle variabili in gioco siano sensibili allo scostamento dall'obiettivo.

Nell'esempio considerato, avendo  $\vec{\alpha} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$  si capisce che solo l'individuo  $B$  reagisca agli scostamenti  $(\vec{\beta}' \cdot \vec{w}_{t-1} - \vec{\mu}_0)$  avendo detto che esso viene attirato dagli spostamenti di  $A$  ad ogni passo che  $A$  stesso compie.

Le tre rappresentazioni di un processo co-integrato, pur essendo equivalenti tra loro, fanno emergere le diverse caratteristiche che tali processi possono mettere in evidenza, ed in particolare per la rappresentazione autoregressiva vettoriale, l'attenzione viene posta sul fatto che essa contempra processi integrati, per la rappresentazione a trend comuni si evidenzia la presenza di processi comuni non stazionari in covarianza all'interno di un sistema co-integrato, ed infine la

rappresentazione a correzione dell'errore evidenzia il meccanismo attraverso il quale il sistema effettua degli aggiustamenti delle variabili verso un valore di equilibrio.

Effettuando un confronto tra processi co-integrati e non co-integrati si possono notare delle peculiarità che li differenziano nelle loro traiettorie, e precisamente, nei processi co-integrati le traiettorie che li compongono si intersecano un numero elevato di volte procedendo lungo un tragitto non stazionario comune mentre nei processi non co-integrati si nota la persistenza di tragitti non comuni tra le traiettorie, ed inoltre esse si intersecheranno un numero assai limitato di volte.

Ancora, prese per semplificazione due traiettorie, se esse risulteranno co-integrate, si noterà che la combinazione lineare data dalla differenza tra le due serie, risulterà essere un processo stazionario in covarianza. Se esse risulteranno non co-integrate, invece, la combinazione lineare data dalla loro differenza avrà caratteristiche non-stazionarie e così sarà per ogni altra combinazione lineare di esse.

Infine, si nota che traiettorie co-integrate tendono a muoversi lungo una retta rappresentata dall'attrattore di cui si è parlato in precedenza, ma a differenza delle altre due caratteristiche le quali sono precipue dei processi co-integrati, quest'ultima può essere osservata anche in processi che non risultano co-integrati poiché, le loro traiettorie possono disporsi lungo una qualche retta pur se esse seguono degli andamenti che non hanno delle relazioni di co-integrazione.

Si può dunque asserire, sulla base di queste piccole caratteristiche, che la peculiarità dei sistemi co-integrati consiste nella esistenza di combinazioni lineari stazionarie in covarianza delle variabili che li compongono.

In molti fenomeni reali, da quelli economici a quelli fisici, passando per quelli sociali ecc. si nota la presenza di situazioni in cui si possono evincere delle chiare manifestazioni dei concetti di relazioni di equilibrio di medio lungo periodo e di scostamento delle variabili, di tipo transitorio o permanente, da tale valore di equilibrio. Tali fenomeni possono essere abbinati tra loro formulando delle piccole considerazioni, e precisamente, considerando dei processi integrati, quando essi raggiungono certi livelli, non vi è statisticamente nessuna certezza sul fatto che possano ritornare al punto di partenza o che possano far ritorno ad i livelli passati, pertanto tali processi hanno la caratteristica che, ogni loro spostamento risulterà essere permanente. Per i processi autoregressivi stazionari, invece, oscillando essi intorno al proprio valore atteso, gli scostamenti da esso che si manifesteranno saranno considerati meramente transitori. Infine, la esistenza di relazioni di co-integrazione (dette anche *relazioni lineari*) tra processi non stazionari vanno considerate alla stregua di situazioni di equilibrio tra le variabili oggetto di analisi, e dunque, gli eventuali scostamenti da tali situazioni di equilibrio, essendo processi stazionari, vanno considerati di natura meramente transitoria. Focalizzando l'attenzione sul concetto di co-integrazione, le sue relazioni, considerandole a pieno come relazioni di equilibrio, permettono di ben specificare, in un processo VAR, i movimenti di lungo periodo che le variabili assumono, potendo sfruttare il tutto per fini predittivi.

Tornando alla trattazione dei processi VAR( $p$ ) iniziata precedentemente, si è visto come un modello VAR di ordine  $p$  sia invariante rispetto a trasformazioni lineari

invertibili delle variabili (ciò, come già precisato, non è più valido se si effettuano delle trasformazioni non lineari delle variabili).

Questo concetto permette di facilitare i calcoli dei coefficienti utili alla sua specificazione. Ad esso è legata la possibilità di rappresentare un processo  $VAR(p)$  in termini di un processo più semplice  $VAR(1)$ . Allora, dato un generico processo

$VAR(p)$  del tipo  $A(L)\vec{w}_t = \vec{c} + \vec{\epsilon}_t$  dove  $A(L) = I_M - \sum_{i=1}^p A_i L^i$  con  $A_i$  matrici

$M \times M$  di coefficienti,  $L$  il solito operatore ritardo delle variabili, ed  $\vec{c}, \vec{\epsilon}_t$  sono rispettivamente insiemi di componenti deterministiche e di variabili casuali che, in questo caso, sono considerate *i.i.d.* con varianza  $\Lambda = \sigma_\epsilon^2$ .

Sfruttando l'invarianza a trasformazioni lineari di un  $VAR(p)$  si può rappresentare un tale processo nella cosiddetta *forma compagna* attraverso una ridefinizione della variabili del sistema. Precisamente, si ponga il vettore  $\vec{c}$  di dimensione  $M \times 1$  come

$\vec{c} = \mathbf{u} \cdot \vec{d}_t$  dove  $\mathbf{u}$  è una matrice di dimensioni  $M \times S$  e  $\vec{d}_t$  vettore  $S \times 1$ , entrambi

formati da componenti deterministiche. Inoltre, si ponga  $\mathbf{v}_t = \begin{pmatrix} \vec{\epsilon}_t \\ \vec{0} \end{pmatrix}$  composto da un

processo *iid*  $M$ -variato con valore atteso nullo e matrice di *varianza-covarianza*  $\Lambda = \sigma_\epsilon^2$  e da un vettore nullo di dimensioni  $1 \times (M \cdot p - M)$ .

Si ridefinisca il vettore delle variabili  $\vec{w}_t$  come:  $\vec{w}_t^* = \begin{pmatrix} \vec{w}_t \\ \vec{w}_{t-1} \\ \vdots \\ \vec{w}_{t-p+1} \end{pmatrix}$  ed inoltre si

definisca la seguente matrice di dimensioni  $M \cdot p \times M \cdot p$  come segue:

$$\mathbf{J} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \dots & \mathbf{A}_p \\ \mathbf{I}_M & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_M & \dots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{I}_M & \mathbf{0} \end{bmatrix}$$
 detta *matrice compagna* formata dalle matrici dei coefficienti

$\mathbf{A}_i$ , da matrici identità e da matrici nulle.

Considerando tali ridefinizioni, si potrà rappresentare il vettore delle variabili  $\vec{w}_t^*$  come un processo  $VAR(1)$  nel seguente modo  $(\mathbf{I} - \mathbf{J}(L)) \cdot \vec{w}_t^* = \vec{c} = \mathbf{u} \cdot \vec{d}_t + \mathbf{v}_t$  alla stregua di quanto già precedentemente detto sulla scrittura di un  $VAR(1)$ . Tale rappresentazione di un  $VAR(P)$  è detta *forma compagna* ed ha la finalità di facilitare i calcoli necessari alla definizione delle misure del processo.

Il vettore delle variabili iniziali  $\vec{w}_t$  può essere estrapolato facendo ricorso alla selezione delle prime  $M$  righe del vettore delle variabili ridefinite  $\vec{w}_t^*$  facendo ricorso alla cosiddetta *matrice di selezione*  $\mathbf{R} = (\mathbf{I}_M, \mathbf{0}_{M \times M \cdot p - M})'$  di dimensioni  $M \cdot p - M \times M \cdot p - M$  nel seguente modo  $\vec{w}_t = \mathbf{R}' \cdot \vec{w}_t^*$ .

Tale manipolazione ha la finalità di facilitare i calcoli per le soluzioni della equazione caratteristica  $|\mathbf{A}(z)|=0$  poiché si può dimostrare che, tali soluzioni, possono essere calcolate mediante il calcolo degli autovalori della matrice *compagna*.

Considerando il legame tra la matrice compagna e le soluzioni della equazione caratteristica  $|\mathbf{A}(z)|=0$  appena visto, si può definire la condizione di stazionarietà del VAR dicendo che, il processo risulterà stazionario se gli autovalori della matrice compagna  $\mathbf{J}$  saranno minori dell'unità in valore assoluto.

Tornando al concetto di co-integrazione, nel caso illustrato come esempio si è considerato un sistema composto da due variabili legate da una relazione di co-integrazione. Tale sistema è stato rappresentato mediante un modello a correzione dell'errore (*ECM, Error Correction Model*).

Qualora si estenda l'analisi ad un insieme di più variabili, la rappresentazione del modello assume una forma vettoriale, nota come *Vector Error Correction Model (VECM)*. L'impianto teorico alla base del *VECM* rimane invariato: anche nel caso multivariato il modello descrive la dinamica di breve periodo delle variabili in funzione dello scostamento di lungo periodo dall'equilibrio, con la sola differenza che il *VECM* consente di trattare simultaneamente più di 2 variabili con i rispettivi meccanismi di aggiustamento.

La stima del *VECM* potrà avvenire con il metodo degli *OLS* e della *ML* a seconda della gaussianità o meno delle innovazioni  $\vec{\epsilon}_t=(u_t, v_t)'$ . In particolare, sotto l'assunzione di innovazioni gaussiane presenti nella rappresentazione a trend comuni, la stima per massima verosimiglianza (come nel metodo di *Johansen*) restituisce stime efficienti. In assenza di normalità, è comunque possibile ottenere stime consistenti applicando il metodo dei minimi quadrati ordinari (*OLS*) alle equazioni del *VECM*, sebbene si incorra in perdita di efficienza.

Si è specificato precedentemente che, se il sistema risulta essere co-integrato, la

matrice di impatto  $\Pi=\begin{pmatrix} 0 & 0 \\ \gamma & -1 \end{pmatrix}$  può tranquillamente essere fattorizzata come

prodotto tra il vettore dei caricamenti o *loading* del sistema ed i coefficienti del vettore di co-integrazione. In generale, il numero di relazioni di co-integrazione di un sistema formato da 2 o più variabili sarà dato dal rango  $r$  della matrice d'impatto.

Si potranno avere le seguenti situazioni:

i)  $\text{rango}(\Pi)=r=0 \rightarrow \Delta \mathbf{w}_t \sim \text{VAR}(p-1)$  stazionario

ii)  $\text{rango}(\Pi)=r=G \rightarrow \mathbf{w}_t \sim \text{VAR}(p)$  stazionario senza unit roots

iii)  $0 < \text{rango}(\Pi) < G \rightarrow \mathbf{w}_t$  il processo risulta avere  $r$  relazioni di co-integrazione e  $G-r$  trend comuni

$G$  rappresenta la dimensione della matrice d'impatto pari a  $G \times G$ . Nel caso di una matrice derivante da un *ECM* si avrà che  $G=2$ , se la matrice riguarderà un *VECM* allora  $G>2$ . Il rango di co-integrazione potrà essere stimato attraverso il test *traccia* oppure attraverso il test del *massimo autovalore*. Tali test si basano sul rapporto di  $ML$ , il quale è rappresentato da  $\delta = \frac{Max L / H_0}{Max L} \in [0,1]$  .

Le distribuzioni dei test non seguono una distribuzione chi-quadro ( $\chi^2$ ) . Esse sono state tabulate per diversi valori di  $G$  ed  $r$  con differenti componenti deterministiche quali, costanti, trend o nulla della specificazione del *VECM*. La specificazione deterministica del *VECM* ha impatto sul valore dei livelli critici delle statistiche test e dovrà rispecchiare una coerenza con le traiettorie dei dati. Tali test necessitano di un grado di integrazione delle variabili inserite a sistema pari ad 1 e di un numero di variabili sufficientemente elevato per garantire un numero sufficiente di *lag* nel modello *VAR*.

### 3.4

## Cenni sulle Reti Neurali Artificiali (ANN)

In questa sezione si fornisce una breve descrizione delle reti neurali artificiali (*ANN*, *Artificial Neural Networks*), utilizzate nel presente lavoro come strumenti ausiliari a supporto dell'analisi dei sistemi co-integrati.

Le *ANN* rappresentano una classe di modelli di apprendimento automatico (*Machine Learning*) ispirati al funzionamento dei sistemi neuronali biologici e si caratterizzano per la capacità di approssimare funzioni non lineari complesse. Senza entrare nel dettaglio tecnico, vengono di seguito richiamati i principi fondamentali relativi alla loro struttura, al processo di apprendimento e alle principali architetture di rete, al fine di contestualizzarne l'impiego nell'ambito della presente analisi empirica.

Le reti neurali artificiali (d'ora in avanti *ANN*) rappresentano una classe di modelli statistici e computazionali ispirati al funzionamento del cervello umano, in particolare al modo in cui il cervello umano elabora le informazioni attraverso un elevato numero di neuroni biologici interconnessi.

Analogamente ai sistemi biologici, le *ANN* sono costituite da una rete di unità elementari di elaborazione o *nodi* (neuroni artificiali), organizzate in strati e interconnesse tra loro, con l'obiettivo di rappresentare e risolvere problemi complessi che risultano difficilmente affrontabili mediante approcci statistici tradizionali.

A titolo esemplificativo, il cervello umano è costituito da circa 85 miliardi di neuroni, un numero significativamente superiore rispetto a quello osservato in altre specie viventi; si consideri, ad esempio, che un moscerino possiede soltanto circa 100 mila neuroni. Nonostante tale ispirazione biologica, anche una rete neurale di dimensioni elevate rispetto alle capacità computazionali attuali non sarebbe in grado di riprodurre fedelmente il funzionamento del cervello di uno scarafaggio. Dunque, le *ANN* devono pertanto essere considerate come modelli euristici o approssimazioni funzionali, in grado di rappresentare alcuni aspetti del comportamento dei sistemi biologici, ma non di emularli integralmente. Ciò nonostante, le reti neurali artificiali rappresentano uno strumento di supporto decisionale di grande efficacia, grazie alla loro capacità di modellare e rappresentare in modo flessibile un'ampia varietà di fenomeni complessi del mondo reale. I primi veri tentativi di riproduzione artificiale del funzionamento del cervello umano risalgono alla metà del XX secolo, periodo in cui la comunità scientifica iniziò a interrogarsi sulla possibilità che una macchina artificiale, pur caratterizzata da una struttura estremamente più semplice rispetto a quella di un organismo biologico — persino di uno scarafaggio — potesse riprodurre o prevedere il comportamento di un essere vivente in modo efficace.

Tale riflessione rappresentò uno dei primi passi verso lo sviluppo dei modelli di rete neurale artificiale. I primi tentativi di replicare i meccanismi di elaborazione del cervello umano nella risoluzione di problemi si basavano sull'apprendimento di semplici funzioni logiche, quali AND ed OR. Con il progressivo e continuo aumento delle capacità di calcolo dei computer — iniziato nella seconda metà del XX secolo e

tuttora in corso — tali studi hanno conosciuto una crescita esponenziale, portando allo sviluppo di reti neurali sempre più complesse e strutturate, capaci di affrontare problemi di natura via via più articolata.

Tra le principali applicazioni si annoverano i sistemi di riconoscimento del parlato, della scrittura e delle immagini, nonché le tecnologie di automazione intelligente, come i sistemi di guida autonoma per automobili e droni.

Ulteriori ambiti di impiego riguardano la modellazione del clima, la dinamica dei fluidi, nonché l'analisi di fenomeni sociali, economici e finanziari, problemi di classificazione, la predizione numerica, rilevamento di pattern, il tutto anche in modalità senza supervisione<sup>15</sup>.

In merito al modo in cui le ANN interpretano le informazioni, diciamo che esso è assimilabile al modo in cui il sistema neuronale biologico interpreta le informazioni che riceve dall'esterno.

Infatti, è possibile creare un diagramma di rete col quale definire una relazione tra i segnali di input ricevuti dai dendriti (definiti come variabili  $x$ ) ed il segnale di output prodotto dal neurone artificiale (definito come variabile  $y$ ).

Analogamente a quanto avviene nel neurone biologico, ciascun segnale in ingresso (la variabile  $x$ ) viene ponderato da un peso sinaptico (i pesi  $w$ ) che ne determina l'importanza relativa.

I segnali pesati vengono quindi combinati nel corpo del neurone artificiale ed il risultato è trasformato mediante una funzione di attivazione  $f$ .

Un singolo neurone artificiale con  $n$  ingressi (dendriti) può essere rappresentato come:

$$y(x) = f\left(\sum_{i=1}^n w_i \cdot x_i\right) \quad \text{dove } x_i, w_i, y \text{ rappresentano rispettivamente le variabili di}$$

input con i rispettivi pesi sinaptici e l'output della rete, funzione delle variabili  $x$ .

Questa struttura costituisce l'unità elementare con cui vengono costruite le reti neurali, utilizzate per modellare relazioni complesse tra insiemi di dati.

Esistono diverse tipologie di ANN, ma ognuna di esse può essere definita in termini delle seguenti caratteristiche:

- i) Una funzione di attivazione  $f$  che trasforma il segnale netto di input di un neurone in segnale di output che viene trasmesso alla rete
- ii) Una topologia o architettura di rete che descrive il numero di neuroni del modello, il numero di strati (*layer*) e le loro interconnessioni

---

15 Si parla di apprendimento supervisionato quando, durante la fase di addestramento, il modello statistico o computazionale in genere dispone di un insieme di osservazioni del fenomeno studiato per le quali siano noti i valori di output o di target. In tal caso, l'obiettivo dell'apprendimento è quello di stimare un funzionale che, a partire dalle variabili di input, predica nel modo più accurato possibile i valori noti di output, consentendo così di valutare la bontà della previsione attraverso il confronto tra valori stimati e quelli osservati. Invece, l'apprendimento è detto non supervisionato quando non sono disponibili valori di output o target del fenomeno con i quali effettuare un confronto dei risultati generati dal modello applicato.

- iii) Un algoritmi di addestramento con il quale si specifica l'impostazione dei pesi di connessione, con lo scopo di inibire o eccitare i neuroni proporzionalmente al segnale di input

In merito a queste caratteristiche, è opportuno soffermarsi brevemente su alcune delle principali varianti rilevanti per l'implementazione delle reti neurali artificiali.

Per quanto riguarda la funzione di attivazione, essa ha la finalità di rappresentare, in termini matematici, il meccanismo con cui il neurone artificiale elabora i segnali in ingresso e decide se trasmettere o meno l'informazione agli altri neuroni della rete. In analogia con il neurone biologico, la funzione di attivazione consente di modellare il processo per cui i segnali di input, opportunamente pesati, vengono sommati e confrontati con una soglia di attivazione. Se la somma ponderata supera tale soglia, il neurone si attiva e trasmette il segnale in uscita; diversamente, rimane inattivo. Questo tipo di meccanismo è formalizzato nella cosiddetta *funzione di attivazione a soglia*. Tra le funzioni più semplici si annovera la *funzione a step unitario*, concettualmente simile alla funzione segno, nella quale il neurone trasmette il segnale di output solo quando la somma pesata degli input raggiunge o supera un determinato valore di soglia (generalmente pari a zero). Sebbene di interesse storico e teorico, tali funzioni risultano poco utilizzate nelle reti moderne, poiché la loro natura non differenziabile limita l'applicazione di algoritmi di apprendimento basati sulla discesa del gradiente. Una delle alternative più utilizzate alla *funzione di attivazione a step unitario* è la funzione *sigmoide*, ed in particolare la *sigmoide logistica*.

A differenza della funzione *step unitario*, la *sigmoide* è una funzione continua nell'intero intervallo di input e restituisce valori compresi nell'intervallo  $[0,1]$ . Tale funzione è inoltre derivabile in tutto il suo intervallo di input, il che facilita la creazione di algoritmi efficienti per ottimizzare le reti. Quello che principalmente distingue tali funzioni è il valore che esse possono assumere, ovvero, l'intervallo del segnale di output. Gli intervalli delle funzioni diverse dalla funzione a *step unitario* possono assumere la forma  $(0,1)$ ,  $(-1,1)$ ,  $(-\infty, \infty)$ .

Tra le funzioni menzionate, ve ne sono alcune che presentano una pendenza maggiore in prossimità dell'origine, dove il neurone risulta più sensibile alle variazioni di input, mentre tende ad assumere valori prossimi a 0 o 1 per input molto negativi o molto positivi. Tale fenomeno è detto *saturazione*. Questo fa sì che vi sia una sorta di restrizione dei valori di input in un intervallo di output molto limitato. Un esempio di tali funzioni è quello della *sigmoide*.

Per questa loro caratteristica, esse vengono spesso denominate come *funzioni squashing*, letteralmente *funzioni di schiacciamento*. Una funzione di recente utilizzo è la *funzione rectifier*. I nodi che utilizzano tale funzione nelle reti sono detti *ReLU* (*Rectifier Linear Unit*). Essa è definita in modo tale da assumere il valore zero, se i valori di input sono minori di zero, ed assumere il valore stesso dell'input se esso è maggiore o uguale a zero. E' una sorta di bisettrice del primo e terzo quadrante in cui però, per valori in ascisse minori di zero, l'ordinata è identicamente uguale a zero.

Tale funzione può generare problemi con alcuni software di calcolo risultando indefinita la sua derivata per  $x=0$ . Per questo, è stata creata una sua approssimazione curva chiamata *softplus* o equivalentemente *SmoothReLU*.

Tale funzione è così definita  $y(x)=\log(1+e^x)$  ed assume valori sempre più prossimi a zero per valori di input minori di zero, e si approssima sempre di più agli input per valori di questi uguali o maggiori di zero.

In merito invece alla struttura o topologia di una rete neurale si può dire che essa costituisce un elemento cruciale per determinarne la capacità di apprendimento e la complessità computazionale.

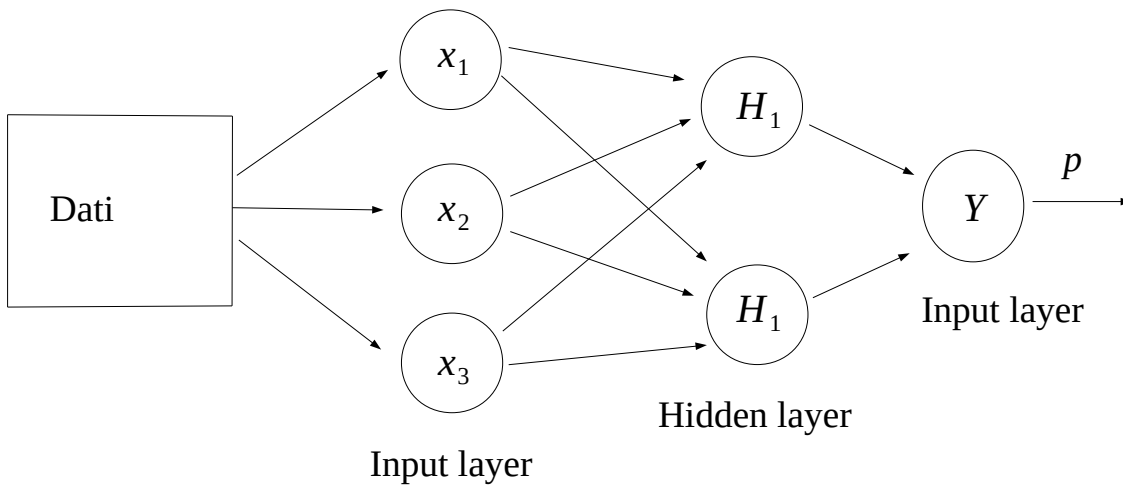
Essa riguarda principalmente il numero di strati (*layer*), la possibilità di retro-propagare l'informazione attraverso la rete ed il numero di neuroni presenti in ciascuno strato. Per quanto riguarda i *layer*, in una rete artificiale se ne distinguono tre tipologie:

- 1) *Input layer*: rappresenta lo strato di ingresso e riceve i dati grezzi provenienti dall'esterno del sistema. Ogni neurone di questo strato corrisponde tipicamente ad una variabile esplicativa del modello
- 2) *Hidden layer (strati nascosti)*: sono gli strati intermedi in cui avviene l'elaborazione dei segnali provenienti dall'*input layer*. Attraverso le funzioni di attivazione, essi consentono alla rete di rappresentare relazioni complesse e non lineari. Il numero e la dimensione di questi strati determinano in gran parte la capacità di generalizzazione e la profondità della rete
- 3) *Output layer*: costituisce lo strato di uscita e fornisce il risultato finale dell'elaborazione, che potrà assumere la forma di una classe o di un valore a seconda del problema analizzato

La presenza di strati nascosti caratterizza le cosiddette reti multilayer (*Multi-Layer Neural Networks*), nelle quali i neuroni sono organizzati in più livelli successivi di elaborazione. Tali reti risultano generalmente pienamente connesse, ovvero ogni nodo di uno strato è collegato a tutti i nodi dello strato successivo, permettendo così una propagazione completa dell'informazione attraverso la rete.

A titolo esemplificativo, nella seguente figura è riportata la rappresentazione schematica di una rete neurale di base, composta da un unico strato nascosto.

## SCHEMA RETE NEURALE CON UNO STRATO HIDDEN



Nella struttura della rete neurale rappresentata in figura si può osservare come i segnali vengano trasmessi da uno strato all'altro seguendo una direzione ben definita indicata dalle frecce: ogni neurone di uno strato è connesso, in questo unico verso, a tutti i neuroni dello strato successivo sino allo strato di *output* il quale, tramite una sua funzione di attivazione, genera una predizione finale indicata in figura con  $p$ . Le reti in cui il flusso dell'informazione procede solo in avanti, senza *retroconnessioni*, sono denominate *reti ad avanzamento* (*Feed-Forward Neural Networks*). Quando una rete di questo tipo comprende più di uno strato nascosto, si parla di rete neurale profonda (*Deep Neural Network, DNN*), ed il processo di apprendimento associato prende il nome di *deep learning*. Accanto alle reti ad avanzamento esistono anche le *reti ricorrenti* (*Recurrent Neural Networks, RNN*), o *feedback networks*, nelle quali i segnali possono muoversi anche all'indietro attraverso connessioni cicliche. Queste architetture risultano particolarmente indicate per la modellizzazione di fenomeni dinamici o sequenziali — come l'elaborazione del linguaggio, del parlato o di serie temporali — e richiedono, in genere, risorse computazionali più elevate, talvolta disponibili solo su hardware dedicato o su infrastrutture *cloud*. Le reti ad avanzamento, tuttavia, risultano già ampiamente sufficienti per rappresentare e modellizzare un'ampia gamma di fenomeni reali, e ne costituiscono la topologia standard.

Queste reti sono spesso indicate anche con il nome di *Perceptron Multi-Layer* (*Multi-Layer Perceptron, MLP*). Nella definizione della struttura di una rete neurale, il numero di nodi negli strati di *input* e di *output* è determinato automaticamente in funzione delle caratteristiche del problema considerato: il primo dipende dal numero

di variabili esplicative o regressori presenti nei dati di ingresso, mentre il secondo è fissato dal numero di variabili da modellare o, nel caso di problemi di classificazione, dal numero di classi da distinguere.

La configurazione degli strati nascosti (*hidden layer*), invece, è lasciata alla discrezionalità dell'utente, che può definire sia il numero di strati sia il numero di nodi presenti in ciascuno di essi. Tale scelta influenza direttamente la capacità di apprendimento e generalizzazione della rete, costituendo uno degli aspetti più rilevanti nella progettazione dell'architettura. Non esiste una regola generale per determinare il numero ottimale di *layer* nascosti e di neuroni in ciascuno di essi. Tale scelta dipende da diversi fattori, tra cui il numero di nodi di input, la natura dei dati di addestramento, nonché la complessità intrinseca del fenomeno oggetto di studio. In linea generale, un aumento della complessità della rete — intesa come maggiore profondità o numero di connessioni — consente di modellizzare fenomeni via via più articolati, poiché una rete più densa e profonda è in grado di apprendere strutture più complesse dai dati di training. Tuttavia, un numero eccessivo di strati o di neuroni nascosti può condurre a fenomeni di *overfitting*, riducendo la capacità della rete di generalizzare correttamente su dati non osservati.

Inoltre, architetture particolarmente complesse comportano un incremento significativo del carico computazionale, rendendo spesso necessario l'impiego di hardware specializzato o di infrastrutture di calcolo ad alte prestazioni. Una buona prassi consiste pertanto nell'adottare l'architettura più semplice possibile che consenta comunque un adattamento soddisfacente ai dati di validazione o test. È stato dimostrato, in particolare, che anche una rete neurale con un solo strato nascosto e un numero sufficiente di neuroni è in grado di fungere da approssimatore universale. Ciò implica che una rete di tale tipo può approssimare qualunque funzione continua su un insieme *compatto* con un grado di precisione arbitrariamente elevato, entro i limiti di un dato intervallo di errore. Una rete neurale, in maniera analoga agli esseri viventi — ed in particolare al cervello umano — mira a riprodurre il processo di apprendimento dall'esperienza. Durante la fase di addestramento, la rete apprende a partire dai dati di input, modificando progressivamente l'intensità delle connessioni sinaptiche artificiali tra i neuroni: tali connessioni si rafforzano o si indeboliscono in risposta agli errori commessi, in modo simile a quanto avviene nei meccanismi biologici di apprendimento. Questo principio ha ispirato lo sviluppo delle reti neurali con *retropropagazione dell'errore* (*error backpropagation*), un metodo di apprendimento che, sebbene risulti computazionalmente più oneroso rispetto ad altri approcci, presenta numerosi vantaggi. In particolare, tali reti possono essere impiegate sia per compiti di classificazione sia per problemi di previsione numerica, dimostrandosi capaci di modellizzare schemi complessi che possono sfuggire agli algoritmi tradizionali.

Un ulteriore punto di forza di esse risiede nella loro capacità di apprendere relazioni non lineari anche in assenza di ipotesi specifiche sulle proprietà strutturali dei dati. Tuttavia, esse presentano anche alcuni limiti intrinseci: oltre all'elevato costo computazionale, sono soggette a fenomeni di *overfitting* sui dati di addestramento e

producono modelli di tipo *black-box* molto complessi la cui interpretabilità risulta spesso limitata o quasi impossibile. L'algoritmo di retropropagazione dell'errore (*backpropagation algorithm*) si basa su un processo iterativo che si sviluppa attraverso numerosi cicli, detti epoche, ciascuna delle quali comprende due fasi fondamentali. Poiché la rete, all'inizio dell'addestramento, non possiede alcuna conoscenza a priori, i pesi sinaptici vengono inizializzati in maniera casuale. L'algoritmo procede poi con l'aggiornamento iterativo di tali pesi fino al raggiungimento di un criterio di arresto, che può dipendere, ad esempio, dal numero massimo di epoche, dal valore minimo dell'errore o dalla stabilizzazione della funzione di costo. Ogni epoca di addestramento si articola in due fasi principali: Una fase all'indietro in cui l'output ottenuto viene confrontato con i valori target presenti nel dataset di addestramento. La differenza tra valore previsto e valore reale genera un errore, che viene propagato all'indietro attraverso la rete. In questo modo, i pesi delle connessioni vengono modificati in modo proporzionale al contributo di ciascun neurone all'errore complessivo, con l'obiettivo di ridurre progressivamente l'errore totale.

Il meccanismo di aggiornamento dei pesi si basa su una tecnica di ottimizzazione numerica nota come *discesa del gradiente*. In particolare, l'algoritmo utilizza la derivata parziale della funzione di attivazione di ciascun neurone per determinare il gradiente dell'errore rispetto ai pesi. Tale gradiente indica la direzione e l'intensità della variazione dell'errore al variare di un determinato peso. Affinché ciò sia possibile, le funzioni di attivazione devono essere differenziabili nell'intero dominio di interesse. La variazione dei pesi è controllata da un parametro chiamato tasso di apprendimento che determina la velocità con cui la rete corregge i propri pesi. Un valore elevato del tasso di apprendimento accelera la convergenza, ma può causare instabilità numerica o far sì che l'algoritmo rimanga intrappolato in minimi locali. In sintesi, nel processo di generazione degli output, la rete neurale elabora i dati di ingresso attraverso una serie di trasformazioni non lineari distribuite su più livelli. In ciascun *layer*, ogni neurone calcola una combinazione lineare dei segnali ricevuti dagli strati precedenti, ponderata dai relativi pesi sinaptici. Il risultato di tale combinazione viene poi trasformato da una funzione di attivazione non lineare, che consente alla rete di catturare relazioni complesse e non lineari tra le variabili. L'insieme di queste trasformazioni successive definisce un insieme funzionale complessa tra lo spazio degli input e quello degli output, rappresentabile come  $\hat{y} = G(x; W)$  dove  $x$  rappresenta i valori di input,  $W$  i pesi della rete e  $G(\cdot)$  rappresenta la funzione appresa. Durante la fase di apprendimento, l'algoritmo di ottimizzazione aggiorna iterativamente i parametri  $W$  per minimizzare una funzione di errore, fino ad ottenere una rappresentazione funzionale in grado di riprodurre con buona accuratezza i valori attesi. Tuttavia, la profondità di questa architettura e la non linearità interna rendono difficile la comprensione esplicita del contributo di ciascun neurone o peso alla decisione finale.

Questo conferisce alle reti neurali il carattere di modelli *black-box*, in cui il processo di inferenza non è immediatamente interpretabile, sebbene altamente efficace sul

piano predittivo. L'esposizione dei metodi di analisi utilizzati in questo lavoro ha avuto l'obiettivo di fornire una rappresentazione generale e sufficientemente esaustiva della loro struttura concettuale ed operativa, senza tuttavia entrare nei dettagli più tecnici, che esulano dagli scopi principali di questo lavoro. La scelta di tale impostazione risponde all'esigenza di contestualizzare gli strumenti teorici impiegati — in particolare i processi stocastici *co-integrati* e le *reti neurali artificiali* — evidenziandone i principi fondamentali e la logica di funzionamento, più che la formalizzazione analitica completa.

Data la complessità e l'ampiezza delle metodologie descritte, un'analisi pienamente esaustiva richiederebbe un lavoro specifico interamente dedicato a tale scopo. Nel presente contesto, esse sono state trattate in misura sufficiente a comprendere il loro ruolo operativo e la funzione interpretativa all'interno delle analisi empiriche sviluppate nel prossimo capitolo.

## CAP. 4

Analisi quantitativa dei domini *Salute e Benessere Economico* per la valutazione del benessere dei territori.

### 4.1

#### Una rappresentazione di analisi predittiva tramite metodi statistici

Il nucleo di questo lavoro consiste nell'applicazione delle metodologie statistiche descritte nel capitolo precedente, in particolare dell'approccio predittivo basato sull'analisi delle traiettorie di sistemi co-integrati e sull'impiego di reti neurali artificiali. La scelta di tali metodologie è motivata da diverse considerazioni. Per quanto riguarda l'utilizzo dei sistemi stocastici co-integrati, l'interesse è nato dalla riflessione sulla loro principale proprietà, che si riscontra anche nella struttura dei domini del BES. Pur non essendo prassi scientifica consolidata applicare tali modelli alla tipologia di dati analizzata in questo studio, la loro adozione è motivata dalla coerenza tra i principi teorici su cui si basano ed il framework concettuale alla base delle traiettorie dei singoli indicatori dei domini Bes. Si partirà osservando che, nell'ambito dello studio del BES, è emersa una proprietà ricorrente all'interno dei suoi domini: gli indicatori appartenenti ad un dato dominio risultano strettamente legati al concetto che il dominio stesso intende rappresentare attraverso essi.

Riflettendo sulla proprietà fondamentale dei sistemi co-integrati, si è immediatamente individuato un parallelismo con questa relazione concettuale: in entrambi i casi, vi è un legame strutturale tra le componenti osservabili (gli indicatori o le serie storiche) ed un'entità teorica sottostante (il concetto del dominio o il processo generatore dei dati). In un sistema co-integrato, infatti, le serie storiche che ne fanno parte condividono una caratteristica essenziale: tutte derivano dallo stesso processo generatore dei dati. Concettualmente significa che l'evoluzione nel tempo delle variabili del sistema è generata da forze comuni sottostanti. Invece, tecnicamente, dire che “delle traiettorie derivino da un comune processo generatore dei dati” significa affermare che il modello matematico alla base delle serie co-integrate è comune. Questo impianto teorico rappresenta il fulcro concettuale di questo approccio e ne spiega l'elevata capacità predittiva.

Affinché ciò sia possibile, è necessario che ciascuna traiettoria appartenente al sistema co-integrato risulti integrata.

Sebbene questa proprietà debba essere sempre verificata empiricamente, è noto che molte serie storiche — relative a fenomeni finanziari, economici, sociali, ecc. — presentino effettivamente un comportamento integrato. La stazionarietà, infatti, rappresenta una condizione piuttosto rara in questo tipo di dati, il che rende l'ipotesi della presenza di processi integrati una scelta metodologicamente fondata.

Verificata la natura integrata dei processi in esame, la sfida principale nell'ambito dei sistemi co-integrati consiste nell'individuare combinazioni di serie storiche che presentino effettiva co-integrazione, condizione necessaria per l'applicazione del modello e per valorizzarne le capacità predittive.

In ambito finanziario, ad esempio, la ricerca di strumenti co-integrati si concentra tipicamente su asset che presentino affinità strutturali, come l'appartenenza allo stesso settore economico, la quotazione su mercati comuni, una capitalizzazione comparabile ecc. Tali criteri di selezione mirano ad identificare serie storiche la cui dinamica evolutiva possa derivare da uno stesso processo stocastico sottostante.

Ma, una volta individuata la relazione di co-integrazione tra le serie, si presenta frequentemente un ulteriore problema: la persistenza di tale co-integrazione nel tempo. Infatti, anche dopo aver accertata la natura co-integrata delle traiettorie ed implementato il sistema co-integrato, non è garantito che la relazione di co-integrazione permanga nel tempo. Al contrario, è spesso osservabile un indebolimento progressivo di tale legame, il che compromette la capacità predittiva del modello. Con il venir meno della co-integrazione, infatti, il modello perde di validità e affidabilità nelle sue previsioni. Ebbene, alla luce delle caratteristiche dei modelli co-integrati, nello studio dei dominî del BES, si è notato che essi posseggono delle caratteristiche che risolvono questi due problemi legati alle serie storiche co-integrate. Più precisamente, in un sistema co-integrato — come già accennato — la proprietà centrale risiede nel fatto che tutte le serie storiche coinvolte condividono un medesimo processo generatore dei dati. Questa osservazione è stata trasferita allo studio dei dominî del BES attraverso il seguente ragionamento: se la variabile latente associata ad un determinato dominio è correttamente specificata dai suoi indicatori osservabili, allora tra questi indicatori e la variabile latente stessa sussiste un legame profondo. Come precedentemente detto, l'evoluzione nel tempo delle variabili del sistema può essere considerata come generata da forze comuni sottostanti; ciò giustifica il profondo legame esistente tra la variabile latente alla base di un dominio ed i suoi indicatori. Tale legame è stato assimilato a quello che unisce le traiettorie in un sistema co-integrato, ovvero la condivisione di un comune processo matematico generatore. Allora, la variabile latente può, a tutti gli effetti, essere considerata come il motore concettuale che genera e giustifica gli indicatori ad essa associati. Dal punto di vista statistico, questo motore concettuale assume il ruolo cruciale di corretto processo generatore dei dati, rappresentando dunque il concetto sottostante che dà origine alle traiettorie dei singoli indicatori.

In altri termini, la funzione matematica che ha generati i dati — ovvero gli indicatori relativi al dominio considerato — viene, in questo caso, fatta coincidere con il concetto rappresentato dalla variabile latente alla base del dominio stesso.

Questo principio, semplice ma statisticamente potente ai fini predittivi, costituisce la base del concetto di co-integrazione, come già evidenziato. Le ipotesi formulate finora permettono di risolvere simultaneamente i due principali problemi associati ai sistemi co-integrati, precedentemente discussi.

Infatti, poiché gli indicatori sono legati in modo strutturale alla variabile latente che rappresenta il dominio di riferimento, l'eventuale esistenza di una relazione di co-integrazione può essere garantita per costruzione. Essendo gli indicatori alimentati dalla stessa variabile latente, è possibile far coincidere quest'ultima con il processo generatore dei dati, assicurando così sempre la natura co-integrata delle traiettorie degli indicatori del dominio.

Inoltre, trattandosi di un legame strutturale, esso non perderà di validità nel tempo, in quanto tale relazione rimane invariata, essendo intrinseca alla stessa costruzione degli indicatori all'interno di un dominio. Queste semplici ma fondamentali considerazioni sugli indicatori di un dominio costituiscono il fulcro del concetto di co-integrazione tra traiettorie. Alla luce di ciò, si è scelto di testare empiricamente la validità statistica dei concetti esposti, utilizzando le serie storiche degli indicatori relativi ai domini *Salute*, con riferimento al territorio della provincia di Taranto, e del dominio *Benessere Economico*, riferito al territorio nazionale. Questa scelta è stata motivata dalla situazione critica — dal punto di vista della salute — che caratterizza questa zona e la popolazione che vi risiede. Negli ultimi decenni, a partire dalla costruzione dell'impianto siderurgico, si sono contrapposte due realtà profondamente divergenti. Da un lato, l'introduzione del polo industriale ha modificato radicalmente il tessuto socio-economico locale, trasformando molti abitanti del territorio da agricoltori ad operai metallurgici, offrendo così nuove opportunità occupazionali ed un potenziale miglioramento delle condizioni di vita. Tale opportunità, tuttavia, può essere considerata effimera, in quanto contestualizzata in un periodo storico in cui l'agricoltura rappresentava un settore economico svolto in condizioni quasi disumane — una realtà ben diversa da quella odierna, caratterizzata da modalità produttive più sostenibili e tutelate. In quel contesto, le dure condizioni lavorative nel settore agricolo spinsero molte persone ad orientarsi verso un impiego industriale, che offriva, almeno in apparenza, un miglioramento della qualità della vita. Oggi, però, il miglioramento tanto auspicato non ha prodotto i risultati sperati. Il settore agricolo ha infatti compiuto importanti progressi, sia in termini di condizioni operative che di riconoscimento economico a livello nazionale, rendendo evidente il divario rispetto ai benefici — spesso solo temporanei — apportati dal settore siderurgico in termini di benessere reale e duraturo per la popolazione locale. A questi concetti si contrappone un aspetto fondamentale e drammaticamente attuale: la tutela della salute pubblica, che è stata gravemente compromessa dalla presenza dell'impianto siderurgico. La salute della popolazione residente in questo territorio è stata infatti messa a rischio dalle emissioni e dalle scorie prodotte dall'attività industriale, sollevando interrogativi critici sull'effettivo miglioramento delle condizioni di vita garantito

dalla presenza del polo siderurgico nel corso degli ultimi decenni. Tornando ora all'approccio metodologico adottato in questo studio, come già accennato nelle righe precedenti, accanto all'analisi della co-integrazione si è scelto di utilizzare anche le *Reti Neurali Artificiali*, valorizzandone la versatilità applicativa esclusivamente in chiave strumentale.

L'utilizzo delle reti neurali, infatti, non è finalizzato ad uno studio teorico delle stesse, bensì al loro impiego come strumento predittivo all'interno del sistema co-integrato, in particolare per stimare una componente fondamentale dei modelli co-integrati e potenziarne così la capacità previsionale.

Pur essendo modelli di tipo *black-box*, le *reti neurali artificiali (ANN)* si sono rivelate particolarmente adatte a questo studio, poiché l'obiettivo non è fornirne una spiegazione statistica dettagliata, ma utilizzare le previsioni generate come strumenti funzionali per stimare la componente di interesse nei modelli co-integrati. I dati impiegati provengono da rilevazioni ufficiali effettuate dall'ISTAT e riguardano le variabili chiave presenti nel documento BesT (Benessere Equo e Sostenibile dei Territori), relativo ai singoli territori nazionali, e del BES relativo al territorio nazionale. Questo lavoro si propone di dimostrare come, grazie alla struttura specifica dei domini del BES (e del BesT), il concetto di co-integrazione possa risultare particolarmente appropriato per l'analisi delle serie storiche di tali indicatori. A tal fine, con un intento essenzialmente dimostrativo, si è scelto di includere nello studio quegli elementi e caratteristiche che le serie storiche degli indicatori dovrebbero possedere per essere più coerenti con il quadro teorico proposto. Si afferma quanto sopra poiché, alla luce delle rilevazioni effettuate sugli indicatori del BES e del BesT, la loro numerosità campionaria, pur non compromettendo l'apparato concettuale finora esposto, limita significativamente la potenza predittiva dei modelli basati su tale quadro teorico.

Essendo questo un lavoro di ricerca sperimentale, dunque, verrà esposta l'idea progettuale volta a superare la problematica legata alla limitata numerosità delle serie storiche degli indicatori. Tale proposta, avendo la finalità di cercare di risolvere una problematica tecnica, sarà logicamente ed operativamente implementabile e non si limiterà ad un mero costrutto teorico. In assenza di questa concretizzazione, l'intero impianto metodologico esposto rischierebbe di configurarsi come una semplice rappresentazione concettuale con scarsa applicazione pratica. Proseguendo allora, si evidenzia come il problema che si presenta nella applicazione dei modelli co-integrati in questa sperimentazione è legato proprio alla dimensionalità campionaria, poiché potrebbe risultare poco proficuo l'utilizzo di tali metodi con serie storiche brevi come quelle utilizzate. Tuttavia si è voluto ugualmente procedere con la sperimentazione,

effettuando delle considerazioni sui risultati riferite alla dimensionalità campionaria. In relazione a tali considerazioni, si è successivamente voluto simulare una ipotetica realtà nella quale le rilevazioni riferite ai singoli indicatori siano effettuate su un time-frame più fine, e precisamente si è considerato un time-frame pari ad una settimana, andando così a generare delle traiettorie “aumentate” — in termini di dimensionalità campionaria — degli stessi indicatori. Questa rilevazione più fine, in termini temporali, è stata simulata poiché, nella realtà, sarebbe realizzabile tramite sistemi automatici di acquisizione dei dati relativi a specifici indicatori sociali. Tale simulazione è stata ottenuta implementando le serie storiche aumentate degli indicatori attraverso la generazione di valori random appartenenti agli intervalli delle rilevazioni annuali, presi a coppia e considerando delle misure di distribuzione (media e varianza) simili a quelli delle serie originali. In tal modo si è riusciti a replicare delle serie storiche degli indicatori studiati con dimensionalità campionaria maggiore, ma con caratteristiche simili alle serie originali, potendo dunque riprodurre una coerente situazione nella quale il time-frame di rilevazione possa risultare più fine. Partendo dalle congetture precedentemente formulate riguardo alla struttura degli indicatori dei domini BES e BesT, si è proceduto ad implementare un ipotetico sistema co-integrato composto dalle serie storiche relative al dominio *Salute*, riferito al territorio di Taranto, ed al dominio *Benessere Economico* riferiti al territorio nazionale. La struttura dei dati ufficiali utilizzati comprende un numero di osservazioni annuali limitata, riferite agli indicatori citati per i quali è stata garantita una continuità temporale. Va immediatamente ribadito che una tale struttura dati non è propriamente ideale per la modellizzazione tramite modelli co-integrati; tuttavia, pur riconoscendo i limiti insiti nella numerosità campionaria, si è voluto comunque verificare l'eventuale presenza di una relazione di co-integrazione tra le serie sulla base di una certa solidità delle ipotesi teoriche che hanno guidata questa sperimentazione.

I risultati ottenuti non hanno respinto le ipotesi formulate, considerando tutte le limitazioni dovute alla struttura del campione. Si precisa che, un arco temporale in termini di anni, eccessivamente lungo potrebbe apparire inappropriato per un'analisi temporale, nel senso che potrebbe includere dati riferiti ad una realtà storica e sociale ormai distante dal contesto attuale, generati quindi da processi diversi da quelli contemporanei. Tuttavia, considerando che in ambito sociale i tempi di aggiustamento per osservare gli effetti di determinate iniziative sono generalmente lunghi, tale arco temporale risulta adeguato in termini di ampiezza.

In generale, la scelta della dimensionalità campionaria di una serie storica è di cruciale importanza statistica, poiché essa implica due esigenze contrapposte: da un lato, è necessario disporre di un campione sufficientemente ampio per garantire un numero adeguato di gradi di libertà, statistiche affidabili e risultati asintotici; dall'altro, il campione deve essere omogeneo, nel senso che i dati in esso contenuti appartengano ad uno stesso processo generatore, che in ambito sociale può corrispondere allo stile di comportamento degli agenti sociali o degli individui coinvolti. Si intuisce che queste due esigenze devono essere saggiamente bilanciate poiché, allungando il campione studiato si soddisferebbe la prima esigenza ma, se non si effettuasse una attenta analisi si potrebbe incorrere nel grave errore di inserire in un unico campione dei dati che in realtà appartengono a campioni differenti poiché rispecchiano, riferendoci all'ambito sociale, stili di comportamento, realtà sociali strutturalmente diversi. Se ad esempio, si volesse studiare un legame tra consumo e reddito partendo dagli anni a cavallo con la grande guerra e gli anni 2000, si rischierebbe di incorrere in gravi errori di valutazione poiché, in tale arco temporale si sono susseguiti dei comportamenti di consumo completamente differenti passando da un consumo di mera sussistenza, tipico del primo dopoguerra, agli anni del consumo legato al boom economico sino ad arrivare ai giorni nostri, in cui vi è un consumo legato all'opulenza ed al superfluo. È naturale aspettarsi che, in un intervallo temporale tanto ampio, i legami causali tra le variabili si modifichino significativamente, poiché i fenomeni sociali ed economici sottostanti evolvono nel tempo, rispecchiando mutamenti strutturali nei comportamenti degli individui, nelle istituzioni e nel contesto socioeconomico complessivo. Chiarito quanto sopra, si procede col dire che, i risultati ottenuti dalla sperimentazione non hanno rigettate le ipotesi formulate: i valori delle statistiche test si sono infatti collocati entro soglie ottimali di non rifiuto dell'ipotesi di esistenza di un certo rango di co-integrazione tra le traiettorie degli indicatori appartenenti ai dominî analizzati. Prima di proseguire con l'esposizione dei risultati ottenuti attraverso l'applicazione delle metodologie precedentemente elencate, verrà fornita una breve spiegazione delle tecniche adottate, già ampiamente esposte nel capitolo precedente. In particolare, si chiariranno le motivazioni teoriche e pratiche che hanno guidato la scelta di tali metodologie, nonché il significato da attribuire ai risultati prodotti, al fine di garantire una corretta interpretazione dell'intera analisi. Si comincerà sottolineando che la finalità principale dell'impiego del concetto di co-integrazione, e della conseguente implementazione di un sistema co-integrato, risiede nella sua elevata capacità predittiva. In termini operativi, una volta verificata l'esistenza di una relazione di co-integrazione tra le serie storiche analizzate ed accertata la natura co-integrata del

sistema, è possibile sfruttare tale caratteristica attraverso due elementi strettamente connessi:

- 1) Il vettore di co-integrazione, derivante dal sistema stimato
- 2) La combinazione lineare stazionaria del  $II$  ordine, generata dallo stesso vettore

Queste due componenti costituiscono il nucleo tecnico attraverso cui è possibile modellare l'equilibrio di lungo periodo tra le variabili e formulare previsioni attendibili in presenza di serie storiche non stazionarie.

Il vettore di co-integrazione rappresenta l'intensità con cui ciascuna delle traiettorie che compongono il sistema interagisce con le altre. Attraverso una combinazione lineare delle serie storiche e dei corrispondenti coefficienti contenuti nel vettore di co-integrazione, si ottiene una combinazione lineare stazionaria del ordine  $II$ . Quest'ultima rappresenta il nucleo informativo del modello co-integrato e costituisce l'elemento da cui è possibile derivare previsioni sulle variabili del sistema. In altri termini, è proprio la presenza di una combinazione lineare stazionaria — risultato diretto della relazione di co-integrazione — che consente di sfruttare il modello ai fini predittivi, in presenza di variabili singolarmente non stazionarie. In particolare, i segnali di movimento delle serie storiche co-integrate sono determinati dal comportamento di *mean-reversion* della combinazione lineare. Quando tale combinazione si discosta significativamente dal proprio valore medio (il suo valore di equilibrio), raggiungendo livelli estremi superiori o inferiori, tende a riavvicinarsi alla media nel tempo. Questo movimento di ritorno rappresenta il meccanismo correttivo che caratterizza i sistemi co-integrati. Il vettore di co-integrazione svolge un ruolo cruciale nell'indicare la direzione dei movimenti che ciascuna serie storica intraprenderà durante la fase di *mean-reversion*, in funzione del segno e dell'intensità dei coefficienti che lo compongono. Pertanto, a seconda della posizione iniziale della combinazione lineare rispetto al proprio equilibrio e del contributo di ciascuna variabile alla deviazione, è possibile inferire l'andamento futuro delle singole serie. Ad esempio, qualora la combinazione lineare stazionaria intraprenda un movimento di *mean-reversion* a partire da un livello superiore rispetto al proprio valore atteso — o valore di equilibrio — allora:

- le variabili associate ai coefficienti positivi nel vettore di co-integrazione tenderanno a diminuire
- mentre quelle associate ai coefficienti negativi tenderanno ad aumentare

fino a quando la combinazione lineare non si sarà riportata verso il proprio equilibrio.

Viceversa, se il processo di *mean-reversion* ha origine da un livello inferiore, allora:

- le variabili associate ai coefficienti positivi tenderanno a crescere
- mentre quelle associate ai coefficienti negativi tenderanno a decrescere

seguendo la stessa logica di ri-equilibrio. In entrambi i casi, le serie storiche si muoveranno secondo direzioni coerenti con il segno degli elementi del vettore di co-integrazione, mentre l'intensità del loro aggiustamento dipenderà dal valore assoluto dei rispettivi coefficienti, che rifletteranno il contributo di ciascuna variabile alla dinamica complessiva del sistema. Si deve specificare che l'implementazione del sistema co-integrato effettuata sulle serie originali, pur restituendo statistiche test che non rigettano l'ipotesi di presenza di un certo rango di co-integrazione — come già accennato in precedenza — presenta alcune criticità. In particolare, la combinazione lineare generata dal sistema non risulta perfettamente stazionaria. Tuttavia, trattandosi di un sistema co-integrato, è lecito considerarla stazionaria, in quanto tale caratteristica è intrinseca ai sistemi co-integrati; si può quindi imputare la mancata perfetta stazionarietà alla ridotta dimensionalità campionaria. Inoltre, sul piano predittivo, essa manifesta un'ulteriore problematica: tende infatti a passare da un estremo all'altro del proprio percorso in maniera eccessivamente repentina, rendendo difficile sfruttare efficacemente il vettore di co-integrazione.

Tale criticità, tuttavia, verrà meno quando si implementerà il sistema utilizzando le serie aumentate: in tal caso, la combinazione lineare assumerà un andamento più regolare nel senso che, una volta raggiunto un estremo (superiore o inferiore), tenderà a ritornare verso il proprio valore atteso in modo più graduale, consentendo così un utilizzo proficuo della co-integrazione ai fini predittivi. Dunque, la forza predittiva legata al concetto di co-integrazione risiede proprio nel movimento di *mean-reversion* tipico dei processi stazionari.

Infatti, le variabili del sistema dovranno obbligatoriamente tendere verso un livello di convergenza, garantito dal ritorno della combinazione lineare verso il proprio valore di equilibrio. Per quanto riguarda le reti neurali invece, queste verranno impiegate prevalentemente per la previsione della componente di un sistema co-integrato che potrebbe risultare problematica. In particolare, ci si riferisce alla combinazione lineare delle variabili co-integrate. Come già accennato, vista la struttura dei dati utilizzati, la combinazione lineare derivante dal processo di co-integrazione potrebbe avere delle caratteristiche non ideali per sfruttarne appieno il potenziale predittivo. Per ovviare a questa possibile problematica, si implementerà una rete neurale semplice, basata sull'algoritmo di *back-propagation*. La rete sarà alimentata da dati di input ottenuti dalla differenza prima di una combinazione lineare delle variabili del

sistema. Tale combinazione lineare sarà ottenuta con coefficienti pari ai valori della varianza spiegata, ottenuta attraverso una analisi delle componenti principali (*PCA*). Tale procedura avrà la finalità di predire l'andamento della combinazione lineare derivante dal processo di co-integrazione, al fine di ridurre al minimo la problematica derivante da un possibile scarsa qualità della stessa, attribuibile alla struttura dei dati campionari utilizzati, come già discusso nelle righe precedenti. La procedura descritta potrebbe inizialmente apparire poco duttile, in quanto prevede la stima — per sua natura approssimativa — di una componente che verrà successivamente utilizzata per effettuare previsioni su variabili di interesse. Tuttavia, considerando che tale componente è già nota a priori nella sua struttura — in quanto derivante da una combinazione lineare di variabili co-integrate — e presentando caratteristiche di stazionarietà, questo approccio non risulta incoerente. Infatti, il comportamento di un processo stazionario è, per definizione, più prevedibile e meno soggetto a deviazioni strutturali, rendendo legittimo l'impiego di una previsione su tale componente come strumento di supporto all'intero processo predittivo. Inoltre, la finalità predittiva della combinazione lineare viene attivata solo quando quest'ultima si trova in prossimità dei suoi estremi superiori-inferiori. Saranno solo questi i punti ai quali la previsione tramite reti neurali della stessa combinazione lineare avrà una funzione ancillare, andando così a restringere il campo di previsione tramite rete neurale della stessa traiettoria. L'uso sinergico di queste due metodologie dovrebbe consentire di effettuare una azione predittiva il più precisa possibile sugli indicatori oggetto di analisi. Ciò permetterebbe al decisore di intraprendere le azioni necessarie per affrontare al meglio, a livello sociale, l'evoluzione futura della variabile latente rappresentata dai dati degli indicatori. Nell'effettuare l'analisi dei dati al fine di investigare la presenza di una qualche possibile relazione di co-integrazione tra le variabili, si è partiti mettendo a sistema le traiettorie degli indicatori presi a coppia. Si è passato poi ad implementare il sistema considerando delle triple, sino ad arrivare ad una implementazione con il numero massimo di variabili utili, in relazione al numero massimo di indicatori che si è riuscito ad estrapolare dal dataset, per ogni singolo dominio studiato. Ebbene, procedendo in tal modo, si è notato come, implementando il sistema co-integrato con due sole variabile — considerando tutte le possibili coppie — non si raggiungevano sempre dei risultati con valenza statistica, in termini di valori delle statistiche test per la stima del rango di co-integrazione. Però, con l'aumento del numero di variabili coinvolte nell'analisi attraverso l'implementazione di un sistema formato da terne, si è osservato un netto miglioramento nei valori delle statistiche test per la verifica della ipotesi nulla di esistenza di co-integrazione. Sebbene alcuni di questi valori si collocassero ancora nelle regioni di rifiuto di tale

ipotesi, l'entità del miglioramento ha giustificato l'estensione dell'analisi a sistemi con il numero massimo di variabili disponibili. Inoltre la co-integrazione derivante da un sistema implementato con triple risultava ridondante in termini di relazioni tra le variabili, rispetto al sistema implementato con 4 variabili. In questo contesto, i test hanno restituito valori pienamente compatibili con l'esistenza di una qualche relazione di co-integrazione, rientrando nelle soglie per non rifiutare l'ipotesi nulla. Questo risultato ha rafforzato l'idea che l'aumento del numero di variabili considerate contribuisca a far emergere legami di lungo periodo tra le stesse, suggerendo la presenza di una relazione di co-integrazione con valenza statistica. Tale evidenza ha fornito ulteriore conferma al quadro teorico inizialmente ipotizzato, che ha guidato l'intera sperimentazione. Inoltre, il miglioramento progressivo delle statistiche test con l'incremento delle dimensioni del sistema potrebbe essere interpretato alla luce del concetto di complessità: infatti, in un sistema complesso, le variabili non si comportano in maniera indipendente, bensì risultano fortemente interconnesse, e le loro dinamiche sono in larga misura determinate dalle reciproche interazioni. Ne consegue che, in contesti di tale natura, le relazioni significative — come quelle di co-integrazione — possono emergere soltanto quando l'intero sistema viene considerato ed analizzato nella sua complessità, piuttosto che attraverso analisi parziali.

Tale proprietà è infatti caratterizzante dei sistemi complessi poiché li differenzia sia dai sistemi semplici sia da quelli complicati. La differenziazione dei suddetti sistemi dipenderà, dunque, dalle intensità dei legami esistenti tra le variabili di cui sono composti. Tale affermazione trova fondamento nella stessa struttura di tali sistemi. Infatti, in un sistema semplice, una generica variabile *i-esima* sarà accoppiata debolmente alle altre variabili. La sua dinamica potrà essere assimilata a quella di un elemento isolato e risulterà debolmente perturbata dall'azione degli altri elementi. Questo comportamento tende a manifestarsi nella maggior parte delle variabili, portando il sistema a configurarsi come un insieme di sottosistemi costituiti da elementi singoli che evolvono nel tempo in modo quasi indipendente. Nei sistemi complicati, invece, data una *i-esima* componente, essa dipenderà prevalentemente da elementi appartenenti a sorgenti di flusso esterne e sarà soltanto debolmente legata alle restanti componenti del sistema. Questo accadrà per la maggior parte delle componenti. Anche in questo caso, il sistema potrà essere scomposto in diversi sottoinsiemi che saranno controllati da tante fonti esterne quante saranno le componenti del sistema che interagiranno direttamente con l'ambiente esterno. Diversamente, per i sistemi complessi, come già accennato, una *i-esima* componente sarà accoppiata molto intensamente con le restanti componenti del

sistema e la sua dinamica sarà determinata in larga misura dalle interazioni interne al sistema. Questa caratteristica inoltre accomunerà la maggior parte delle variabili del sistema o un numero molto rilevante di esse. Dunque, il controllo della dinamica non sarà esercitato dall'esterno del sistema bensì dall'interno, poiché saranno le interazioni tra gli elementi interni a generare un effetto predominante rispetto a quelle con l'ambiente esterno. In una tale struttura si verifica il fenomeno dell'autorganizzazione, processo in cui la dinamica delle componenti trae il suo controllo dall'interno del sistema stesso.

Alla luce di queste proprietà si può affermare che esista una netta differenza tra gli stati semplice e complicato di un sistema, rispetto a quello complesso. Nei primi due casi, sarà possibile identificare distintamente le singole componenti del sistema, consentendone una descrizione in termini riduzionisti. Nel caso dei sistemi complessi, invece, tale approccio risulterà impraticabile a causa delle intense interazioni esistenti tra le variabili, le quali non potranno essere isolate senza comprometterne la visione globale del sistema. Come poc'anzi accennato, condizioni che giustifichino la presenza di un sistema complesso potrebbero essere individuate, per analogia, anche nello studio di un dominio, inteso a tutti gli effetti come un piccolo sistema le cui componenti sono rappresentate dai singoli indicatori. Tale sistema potrà essere studiato nella sua interezza, e solo così facendo sarà possibile far emergere le relazioni significative tra le sue componenti, al fine di ottenere una visione globale del sistema stesso.

Sulla base di queste definizioni si può ritenere che le relazioni di co-integrazione all'interno di un dominio tenderanno ad emergere man mano che nel sistema si inseriranno più indicatori. Si ribadisce che i risultati dovranno essere sempre letti alla luce della qualità dei campioni utilizzati, fatto che non tenderà a minare l'apparato concettuale esposto ma che servirà a giustificare la qualità stessa dei risultati ottenuti. Formulate queste brevi motivazioni e specificazioni alla base della scelta dei metodi computazionali adottati nel presente lavoro, si procederà adesso alla implementazione degli stessi.

## 4.2 Applicazione a dati territoriali: il caso di Taranto

Questo paragrafo sarà dedicato allo studio sperimentale dei dominî selezionati, applicando le metodologie illustrate sin'ora.

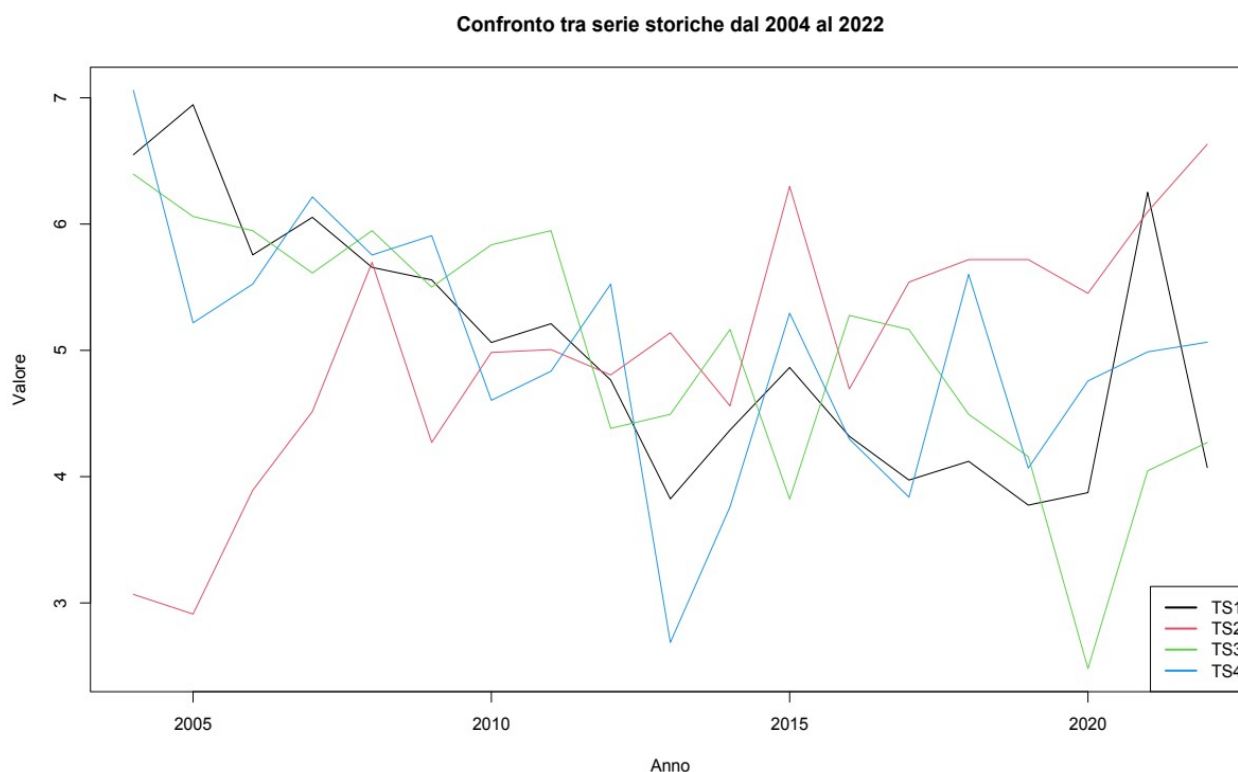
I dati analizzati sono stati estratti dai dataset ufficiali del BES e del BesT. Sono stati scelti i dominî *Benessere Economico* per l'analisi a livello nazionale, mentre, per il territorio della città di Taranto, l'attenzione è stata rivolta al dominio *Salute*. A causa della disomogeneità dei dati in termini di numerosità delle rilevazioni, sia il numero di indicatori utilizzati sia quello delle relative osservazioni potrebbe variare da dominio a dominio. Questa specificazione non influisce sui risultati dell'analisi, ma è riportata al solo scopo di giustificare le eventuali differenze nella numerosità campionaria riscontrabili tra i vari indicatori. L'analisi sperimentale prende avvio dal dominio *Salute*, con riferimento al territorio della città di Taranto.

Gli indicatori selezionati per lo studio sono quattro, e precisamente:

- 1) Mortalità evitabile (0-74 anni) tassi standardizzati
- 2) Mortalità per demenza e malattie sistema nervoso (65 anni e più)
- 3) Mortalità per tumore (20-64 anni) tassi standardizzati
- 4) Mortalità infantile (decessi per 1000 nati vivi residenti)

Per tutti e quattro gli indicatori, le rilevazioni coprono un arco temporale compreso tra il 2004 e il 2022, con un time-frame annuale. Nella rappresentazione grafica, gli indicatori sono contrassegnati rispettivamente con le sigle TS1, TS2, TS3 e TS4.

## GRAFICO DELLE RILEVAZIONI



Da una prima analisi grafica, tenendo conto della limitata numerosità campionaria, emerge una caratteristica tipica dei processi co-integrati: le traiettorie delle variabili incluse nel sistema tendono infatti a intrecciarsi nel tempo, suggerendo un'evoluzione congiunta. Dall'osservazione grafica delle traiettorie emerge anche la natura integrata delle variabili, considerando, tra l'altro, che traiettorie stazionarie nei fenomeni sociali siano piuttosto rare. Per verificare formalmente tale proprietà — su ciascuna variabile — si procederà all'implementazione del test di Augmented Dickey-Fuller (*ADF*) finalizzato a rilevare l'eventuale presenza di radici unitarie. Il test *ADF* consente di verificare se una serie storica sia caratterizzata da stazionarietà o meno. In particolare:

- l'ipotesi nulla assume la presenza di una radice unitaria, e quindi la non stazionarietà della serie
- l'ipotesi alternativa implica la stazionarietà della serie temporale

Il test viene condotto stimando un modello di regressione che include, se necessario, una costante, un trend deterministico e termini auto-regressivi per correggere una eventuale autocorrelazione seriale.

La decisione circa la stazionarietà viene assunta confrontando la statistica test con appositi valori critici, derivanti dalla distribuzione della statistica sotto l'ipotesi nulla. La struttura del test *ADF* implementato in R risulta della forma

$$\Delta y_t = \alpha + \gamma \cdot y_{t-1} + \sum_{i=1}^p \beta_i \cdot \Delta y_{t-i} + \epsilon_t \quad \text{dove } \Delta y_t \text{ rappresenta la differenza prima}$$

della serie storica,  $\alpha$  è una costante,  $y_{t-1}$  è il livello al *lag*1 della serie,  $\Delta y_{t-i}$  sono le differenze dei ritardi,  $\beta_i$  sono i coefficienti delle differenze prime ritardate ed  $\epsilon_t \sim WN(0, \sigma_\epsilon^2)$ . L'ipotesi nulla testa che  $\gamma=0$  che equivale a non rifiutare la ipotesi di presenza di radice unitaria. Per le singole serie il suddetto test non mostra evidenze per rifiutare la presenza di radici unitarie, si tratta dunque di traiettorie non stazionarie in covarianza. Di seguito i test per ogni singola serie storica, con il *lag* definito dalla regola di *Schwert* che indica come valore massimo quello così definito:

$$p_{max} = 12 \cdot \left( \frac{T}{100} \right)^{0.25} .$$

## TS1

Augmented Dickey-Fuller Test

data: TS1

Dickey-Fuller = -0.020752, Lag order = 4, p-value = 0.99

alternative hypothesis: stationary

Messaggio di avvertimento:

In `adf.test(TS1, k = 4)` : p-value greater than printed p-value

## TS2

Augmented Dickey-Fuller Test

data: TS2

Dickey-Fuller = -2.0874, Lag order = 4, p-value = 0.5391

alternative hypothesis: stationary

## TS3

### Augmented Dickey-Fuller Test

data: TS3

Dickey-Fuller = -2.771, Lag order = 4, p-value = 0.3104

alternative hypothesis: stationary

## TS4

### Augmented Dickey-Fuller Test

data: TS4

Dickey-Fuller = -0.7192, Lag order = 4, p-value = 0.9567

alternative hypothesis: stationary

Il numero di ritardi utilizzato per il test *ADF* è pari a 4, valore ritenuto adeguato per evitare risultati distorti e garantire l'assenza di autocorrelazione dei residui utilizzando un principio di parsimonia riducendo una eccessiva perdita di gradi di libertà alla luce della composizione del test, ed anche in considerazione del numero massimo di ritardi suggerito dalla regola di *Schwert* (pari a 7 nel presente caso). La natura integrata – e dunque non stazionaria – delle serie temporali analizzate consente di procedere alla verifica di un'eventuale relazione di co-integrazione, condizione che può sussistere tra variabili non stazionarie ma che evolvono congiuntamente nel tempo. La verifica della presenza di cointegrazione è condotta mediante la procedura di Johansen, basata sul test del massimo autovalore, che consente di determinare il numero di relazioni di equilibrio di lungo periodo tra le variabili considerate. Tale scelta, rispetto all'approccio di Engle-Granger, trova giustificazione nella natura multi-variata del sistema oggetto di analisi, permettendo di identificare il rango di co-integrazione in un contesto di sistema e di evitare l'arbitrarietà connessa alla selezione di una specifica variabile dipendente. Nella pagina successiva, si riporta la diagnostica di tale procedura.

## PROCEDURA PER VERIFICA RANGO DI CO-INTEGRAZIONE

```
#####  
# Johansen-Procedure #  
#####
```

Test type: maximal eigenvalue statistic (lambda max) , with linear trend

Eigenvalues (lambda):

```
[1] 0.8055018 0.5129092 0.3913229 0.3332009
```

Values of teststatistic and critical values of test:

	test	10pct	5pct	1pct
r <= 3	6.89	6.50	8.18	11.65
r <= 2	8.44	12.91	14.90	19.19
r <= 1	12.23	18.90	21.07	25.75
r = 0	27.83	24.78	27.14	32.14

Eigenvectors, normalised to first column:

(These are the cointegration relations)

	ts.TS1..l1	ts.TS2..l1	ts.TS3..l1	ts.TS4..l1
ts.TS1..l1	1.0000000	1.000000	1.0000000	1.0000000
ts.TS2..l1	1.5604868	1.346462	-0.8675337	0.4556241
ts.TS3..l1	0.1765571	1.725758	-2.6414473	-0.1594254
ts.TS4..l1	-0.4157537	-1.885209	-4.2134914	-0.6355068

Weights W:

(This is the loading matrix)

	ts.TS1..l1	ts.TS2..l1	ts.TS3..l1	ts.TS4..l1
ts.TS1..d	-0.0729193	-0.1089294	0.125517137	-0.4879972
ts.TS2..d	-1.0023335	-0.0892314	0.004860626	-0.2361173
ts.TS3..d	0.1321978	-0.4347735	0.025757760	1.5570490
ts.TS4..d	-0.5877549	0.3212804	0.126083554	-0.4594043

```
> as.integer(mean(IND$selection))
```

```
[1] 2
```

```
> adf.test(Comb. Lin.);
```

Augmented Dickey-Fuller Test

data: Comb. Lin

Dickey-Fuller = -3.8575, Lag order = 2, p-value = 0.0316

alternative hypothesis: stationary

Da una prima analisi si può constatare la presenza di un rango di co-integrazione dato il valore della statistica test pari a 12.23 inferiore ai valori soglia, e si può rifiutare la ipotesi nulla di assenza di co-integrazione ad un livello di significatività del 5 per cento. Tali evidenze suggeriscono l'esistenza di un trend comune di lungo periodo tra le quattro variabili considerate, indicativo di un processo generatore condiviso che guida l'evoluzione congiunta delle serie. Nella idea del presente

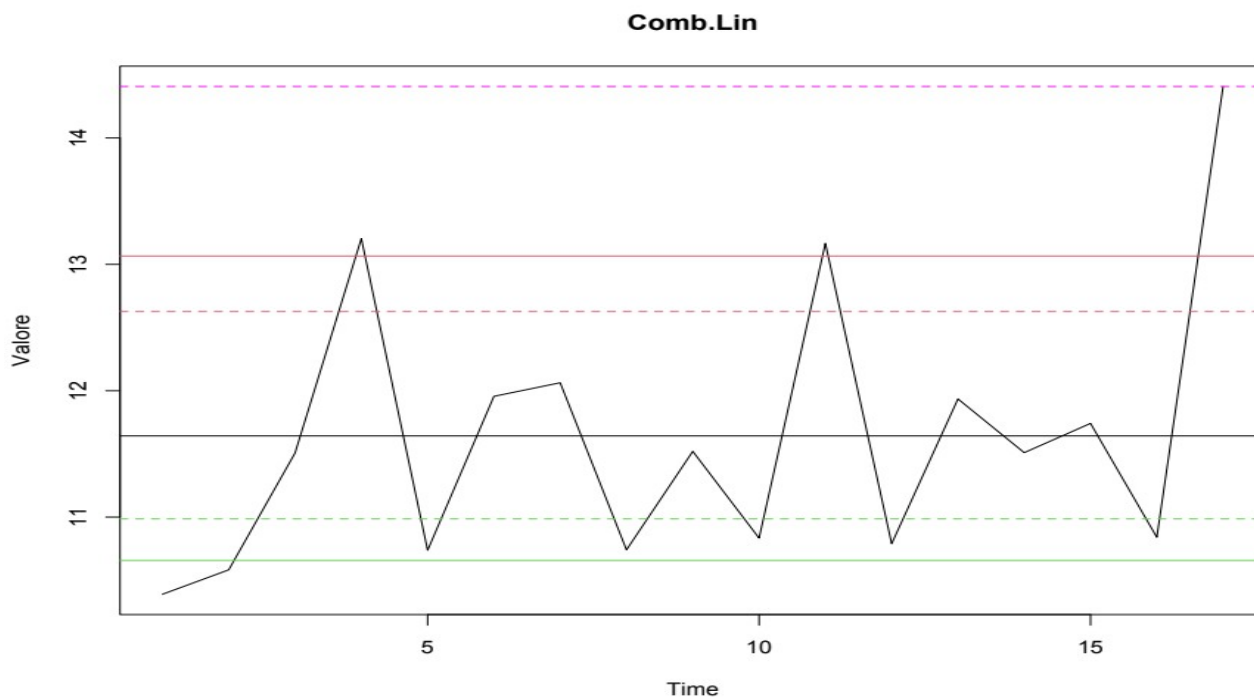
lavoro, tale processo generatore dei dati condiviso viene fatto coincidere con la variabile latente sottostante al dominio di riferimento degli indicatori in esame, il che potrebbe giustificare la co-integrazione osservata tra le traiettorie. L'analisi del vettore di co-integrazione consente di approfondire la natura della convergenza verso l'equilibrio, specificando il modo in cui ciascuna variabile partecipa alla dinamica di aggiustamento nel lungo periodo. I risultati della diagnostica del modello verranno ora interpretati criticamente, tenendo conto della struttura del dataset, in particolare della numerosità campionaria, che potrebbe condizionare l'affidabilità delle stime e la robustezza delle inferenze senza minare quello che è l'apparato su cui si fonda il framework concettuale del lavoro. Si inizierà col dire che, il numero ottimale di ritardi per il modello VAR è risultato pari a 2; di conseguenza, il modello VECM è specificato con un ritardo pari a 1 nelle variabili in forma differenziata. Dallo studio degli autovalori si evince come la relazione più forte tra le variabili sia quella che si ottiene dal primo vettore di co-integrazione il quale resta la principale fonte descrittiva della co-integrazione. Come già accennato, dalle statistiche test si può evincere la presenza di una sola relazione di co-integrazione con valenza essendo il valore della statistica test per  $r=1$  inferiore ai valori critici ad ogni livello di significatività. Inoltre viene rifiutata l'ipotesi di assenza di co-integrazione ad un livello di significatività del 5 per cento. Si può concludere che, pur considerando delle traiettorie con limitata dimensionalità campionaria, la procedura propende per una evidenza della esistenza di una relazione di lungo periodo tra le serie. Il vettore di co-integrazione, che rappresenta la relazione di equilibrio di lungo periodo, consente di interpretare il contributo relativo delle variabili affinché la loro combinazione lineare risulti stazionaria. In particolare, per la variabile TS2, il coefficiente stimato è pari a 1.56048, indicando un peso significativo e positivo nella determinazione dell'equilibrio. Questo implica che eventuali variazioni in TS2 richiederanno aggiustamenti compensativi da parte delle altre variabili, al fine di mantenere la relazione di co-integrazione e quindi il sistema in equilibrio. Il coefficiente associato alla variabile TS3 è pari a 0.17655, suggerendo un contributo contenuto alla relazione di equilibrio. Infine, per la variabile TS4, il coefficiente stimato è pari a -0.41575, indicando una maggiore rilevanza rispetto a TS3 in termini di apporto alla relazione di lungo periodo. Inoltre, il segno negativo del coefficiente suggerisce che TS4 si aggiusti in direzione opposta rispetto alle altre variabili: ad esempio, se TS1, TS2 e TS3 tendono a diminuire, TS4 tenderà ad aumentare, e viceversa. Questo però dipenderà anche dal segno del suo coefficiente di *loading*. In sintesi, dall'analisi del vettore di co-integrazione emerge che la variabile TS2, avendo il coefficiente maggiore in valore assoluto, rappresenta il contributo più rilevante al mantenimento

della relazione di equilibrio. Le restanti variabili mostrano un contributo più contenuto alla relazione di equilibrio di lungo periodo, con la particolarità che TS4 si dovrebbe muovere in direzione opposta rispetto alle altre componenti del sistema. Per quanto riguarda la matrice dei coefficienti di aggiustamento (*loading matrix*), la sua analisi consente di valutare la velocità con cui ciascuna variabile reagisce alle deviazioni dall'equilibrio definito dal vettore di co-integrazione. In altre parole, essa quantifica il grado di aggiustamento delle variabili nel riportare il sistema verso la relazione di lungo periodo. Poiché si è assunto che la prima relazione di co-integrazione sia quella rilevante, l'analisi si concentrerà esclusivamente sul primo vettore colonna della matrice dei coefficienti di aggiustamento, in quanto unico associato alla relazione di equilibrio individuata. Precisamente, quando la relazione di equilibrio è violata, le componenti del vettore dei pesi indica la velocità con cui, in differenza, ogni variabile si aggiusta per riportare il sistema in equilibrio. Il segno dei coefficienti indica la tendenza della variabile di riferimento, ad aumentare o diminuire al passo successivo, in relazione al segno assunto dall'errore della combinazione lineare al medesimo istante. Nel caso in esame, la variabile TS2 ha il coefficiente più grande — in valore assoluto — pari ad  $-1.002333$  ed indica che la variabile reagirà più fortemente delle altre alle deviazioni dalla condizione di equilibrio del sistema. Dunque se la combinazione lineare risulterà lontana dal suo valore di equilibrio, la variabile TS2 reagirà subendo l'aggiustamento maggiore. Inoltre, essendo il valore minore di zero, se la combinazione lineare si troverà sopra il suo valore di equilibrio — come in questo caso —, TS2 reagirà riducendosi maggiormente in differenza, rispetto alle altre variabili. La variabile TS4 è la seconda più reattiva con coefficiente negativo pari a  $-0.58775$ . La TS3 ha un peso positivo e limitato, pari a  $0.13219$  e reagisce in maniera opposta alle variabili con coefficiente negativo. Infine, la variabile TS1 contribuisce all'aggiustamento ma in maniera esigua rispetto alle altre variabili avendo un coefficiente pari a  $-0.072919$ . In conclusione, quando il sistema si discosterà dalla sua condizione di equilibrio di lungo termine, la variabile TS2 correggerà il disequilibrio per prima con la maggiore velocità, conducendo la combinazione lineare verso la condizione di equilibrio. La variabile TS4, visto il segno del suo coefficiente e dell'errore di lungo periodo, potrà fungere da disturbo alla correzione dello stesso errore. La variabile TS3 si aggiusterà in senso opposto alle altre variabili con una intensità limitata, e la variabile TS1 risponderà al disequilibrio con una velocità inferiore a tutte le altre variabili. In sintesi, si può affermare che, quando la combinazione lineare si discosta dall'equilibrio in eccesso, le variabili con coefficienti negativi nella matrice dei coefficienti di aggiustamento tenderanno a correggere l'errore riducendosi in termini

di variazioni, contribuendo così al ritorno verso l'equilibrio, ad esclusione della variabile TS4 che potrà fungere da disturbo alla correzione. Al contrario, le variabili con coefficienti positivi si muoveranno in direzione opposta, incrementando l'errore e quindi non contribuendo alla correzione. In generale, il contributo fornito dalle variabili facenti parte di un sistema co-integrato, dipenderà dal segno dei coefficienti di loading, dal segno dei coefficienti del vettore di co-integrazione e dal segno assunto dall'errore in quel dato istante. Si specifica che una possibile bassa significatività dei valori dei *loading*, in contrasto col significato dato dai loro rispettivi valori assoluti, potrebbe esser dovuto in parte alla dimensionalità campionaria ed in parte ai metodi con cui essi sono stimati nei diversi pacchetti software. Al netto di tali puntualizzazioni, dalla procedura stimata sono derivate dei valori di previsione in linea con le indicazioni date dal sistema e ciò può certamente essere indice del fatto che le discrepanze non hanno grande impatto sulla idea di fondo. In definitiva, il quadro emerso consente di formulare considerazioni interpretative su una possibile configurazione sottostante al sistema sociale analizzato, così come delineato dal modello implementato. Dai risultati ottenuti, la variabile TS2 è emersa come quella con il maggiore impatto sul meccanismo di correzione dell'errore dell'intero sistema. In particolare, ciò implica che, nel caso in cui la variabile latente *Salute* si discosti dal proprio equilibrio di lungo periodo, sarà proprio TS2 – corrispondente all'indicatore “*Mortalità per demenza e malattie del sistema nervoso (65 anni e più)*” – a contribuire in misura preponderante al ripristino dell'equilibrio. Le restanti variabili, in particolare TS1, eserciteranno un'influenza più contenuta sul processo di aggiustamento, suggerendo una loro maggiore inerzia rispetto alle variazioni di equilibrio, oppure una possibile esposizione a limitate dinamiche esogene non catturate dal modello stimato. L'implementazione di un modello volto a descrivere l'evoluzione di un indice sintetico costruito a partire dai suddetti indicatori consente di identificare quali variabili influenzino maggiormente l'andamento dell'indice e in quale direzione esse si muoveranno per determinare un dato cambiamento nel valore complessivo. Ad esempio, se un modello predittivo prevede una diminuzione dell'indice sintetico *Salute*, allora, tramite il modello co-integrato sviluppato, sarà possibile analizzare in che modo le variabili coinvolte interagiranno tra loro per generare la dinamica prevista. A tale scopo, verrà implementata una rete neurale che utilizza come input gli stessi indicatori alla base del sistema co-integrato. Inoltre, come già accennato, verrà costruita una seconda rete neurale finalizzata a prevedere la combinazione lineare stazionaria derivante dal modello co-integrato. Quest'approccio si pone l'obiettivo di superare i limiti imposti dalla ridotta numerosità campionaria, che comprometterebbe un utilizzo diretto e

affidabile della combinazione lineare ottenuta con metodi tradizionali. Si riporterà adesso il grafico della combinazione lineare derivante dal sistema:

### GRAFICO COMBINAZIONE LINEARE

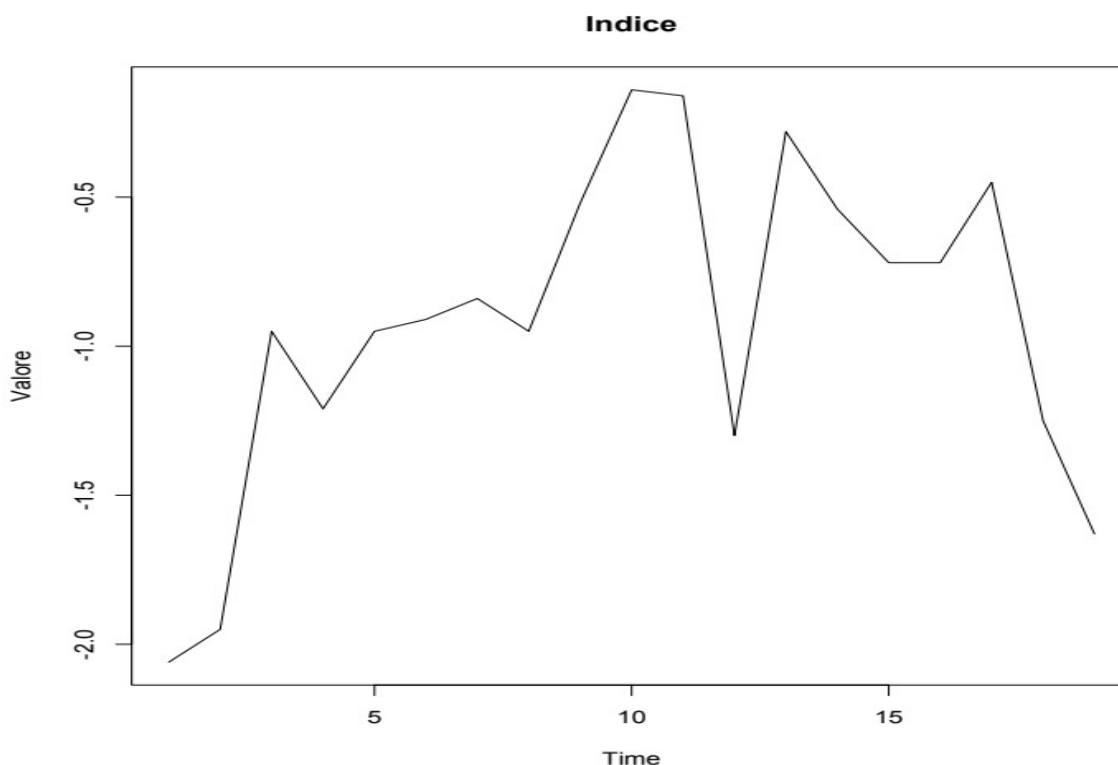


Come già evidenziato dal test di Dickey-Fuller aumentato (*ADF*), la combinazione lineare ottenuta dal sistema co-integrato mostra comportamento stazionario, confermato anche da un'analisi grafica preliminare. Tuttavia, un'eventuale assenza di evidenza di stazionarietà non avrebbe dovuto destare particolare sorpresa. Infatti, nonostante la co-integrazione implichi teoricamente la stazionarietà della combinazione lineare, la limitata numerosità campionaria può impedire l'emersione chiara del movimento di ritorno verso la media (*mean-reversion*) tipico delle serie stazionarie. In tali condizioni, il comportamento osservato potrebbe risultare simile a quello di una serie non stazionaria, sia nei test formali sia nella rappresentazione grafica. La conformazione attuale del sistema suggerirebbe la presenza di un aggiustamento verso il basso della relazione di equilibrio, consentendo una possibile previsione degli indicatori in sistema sfruttando le componenti analizzate sin'ora. Data però la bassa numerosità campionaria, vi è una limitata utilità della combinazione stazionaria che si manifesta attraverso oscillazioni brusche e frequenti, che ostacolano l'identificazione di un chiaro comportamento di mean-reversion.

Questa problematica verrà meno quando si implementerà il sistema sulla base di una ipotetica situazione in cui le rilevazioni dei valori degli indicatori vengano effettuate con una cadenza più fine. In quella sede si procederà a testare la bontà delle previsioni che si potranno effettuare utilizzando la combinazione lineare che scaturirà dal sistema implementato, con campioni delle serie più numerosi. In questo lavoro, avendo trattato in modo ancillare la teoria delle catastrofi di René Thom, si propone un parallelismo tra alcuni concetti chiave di tale teoria e la combinazione lineare derivante dal sistema co-integrato. In particolare, si è scelto di interpretare le bande verdi e rosse (tratteggiate e non) presenti nel grafico della combinazione lineare, alla stregua di ciò che, nella teoria delle catastrofi, viene definito “*insieme delle catastrofi*”, ovvero l’insieme dei punti in cui una forma cambia qualitativamente il proprio stato. In questa prospettiva, tali livelli rappresentano soglie critiche oltre le quali il sistema studiato subisce un cambiamento qualitativo, coinvolgendo simultaneamente tutte le variabili che lo compongono. In corrispondenza di questi punti, le variabili tendono a spostarsi verso nuovi livelli, dando luogo a una transizione strutturale dell’intero sistema. Inoltre, all’interno di questo contesto interpretativo, le bande tratteggiate sono considerate come piccole perturbazioni del sistema: esse influenzano le dinamiche, ma non determinano un cambiamento qualitativo rilevante. Al contrario, le bande continue rappresentano concettualmente l’effettivo “*insieme delle catastrofi*” secondo la teoria di René Thom, identificando i livelli critici nei quali si verifica una transizione strutturale nel sistema, ovvero un cambiamento qualitativo delle sue dinamiche. Per rendere la analisi più inclusiva nell’ambito dello studio dei dominî del BES, si è voluto inserire una ipotetica previsione dell’indice sintetico del dominio oggetto di studio. Tale previsione sarà effettuata implementando una rete neurale basata sull’algoritmo di Back-Propagation. Come variabili esplicative, sono stati utilizzati esclusivamente gli indicatori appartenenti al medesimo dominio, che costituiscono anche i componenti del relativo indice sintetico. Questa scelta metodologica consente di integrare la previsione dell’indice con quella dei singoli indicatori che lo compongono. In sostanza, l’analisi combinata mira sia a stimare l’evoluzione dell’indice sintetico, sia a interpretarne la variazione attraverso il contributo previsto dei singoli indicatori. Ad esempio, qualora si preveda un aumento dell’indice, sarà possibile determinare in che misura tale crescita sia attribuibile alle dinamiche dei singoli indicatori. La bontà della previsione dell’indice sintetico è valutata tramite l’indice di correlazione tra i valori osservati e quelli previsti. L’indice sintetico del dominio Salute riferito al territorio di Taranto è stato implementato usando l’algoritmo del minimo, dopo aver effettuato la normalizzazione degli indicatori attraverso il metodo della standardizzazione. Il

risultato grafico è quello riportato nella figura seguente. Un incremento del valore dell'indice è associato ad un miglioramento della variabile latente sottostante, mentre una diminuzione del valore segnala un peggioramento della stessa.

GRAFICO INDICE SALUTE

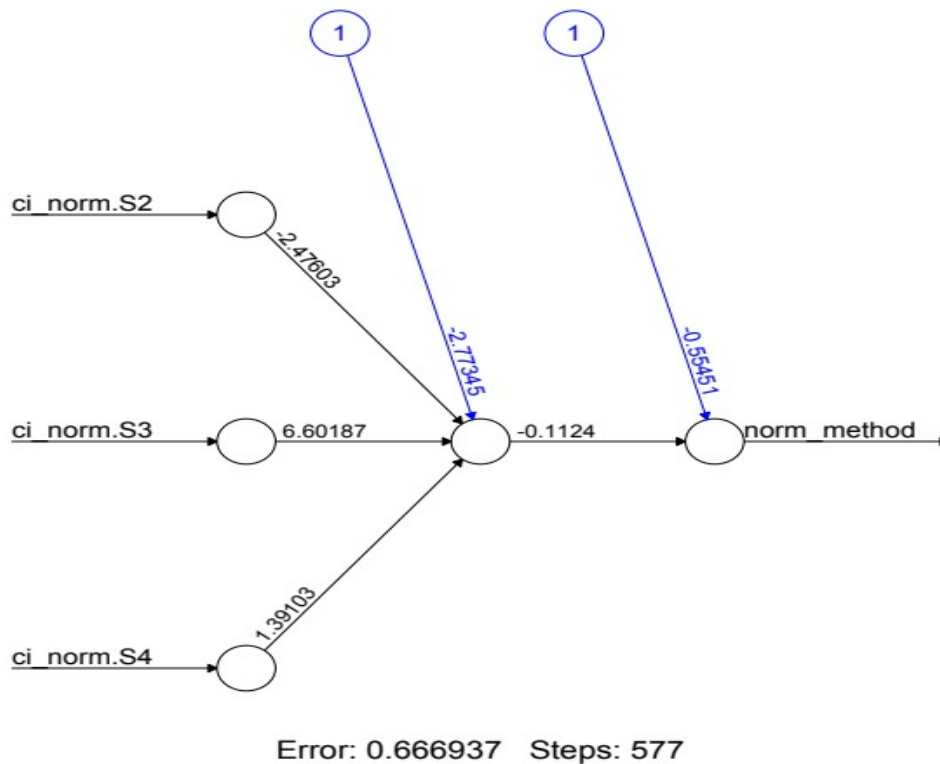


Ai fini predittivi non sarà importante tanto il valore che assumerà l'indice in termini quantitativi bensì il suo movimento in termini qualitativi, ovvero la direzione del suo movimento (rialzista o ribassista). Nel periodo di riferimento, l'indice ha mostrato una dinamica di miglioramento continuo, seguita — negli ultimi cinque periodi — da una tendenza al ribasso. Si procederà ora alla costruzione di una rete neurale, basata sugli ultimi tre indicatori (TS2, TS3, TS4) del dominio considerato, al fine di stimare il valore che l'indice potrà assumere al passo temporale successivo.

La rete implementata è una classica rete neurale di tipo *Multilayer Perceptron* (MLP) con un singolo strato nascosto. È stato sufficiente un modello semplice per ottenere una buona approssimazione dell'indice.

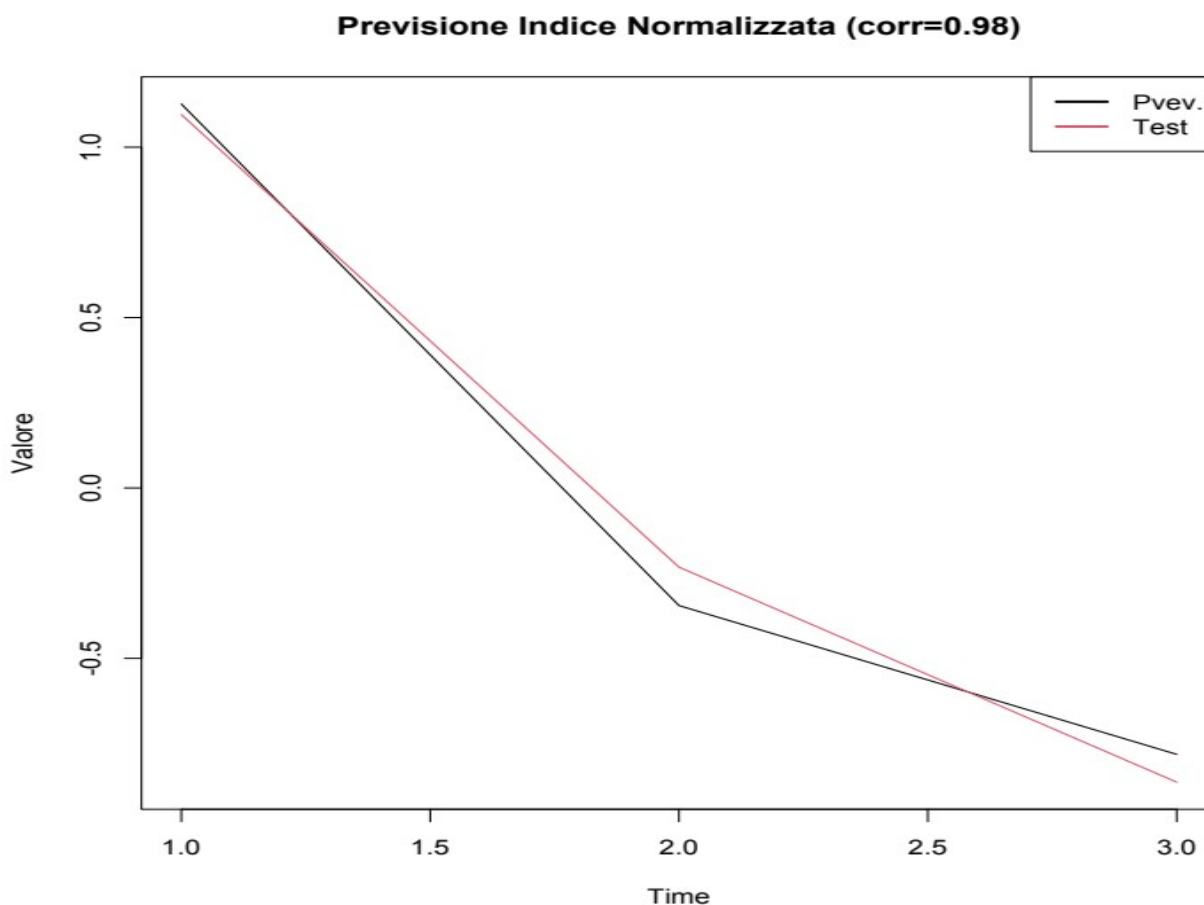
Di seguito si riporta il grafico della rete implementata:

## PLOT RETE IMPLEMENTATA CON MINIMO ERRORE



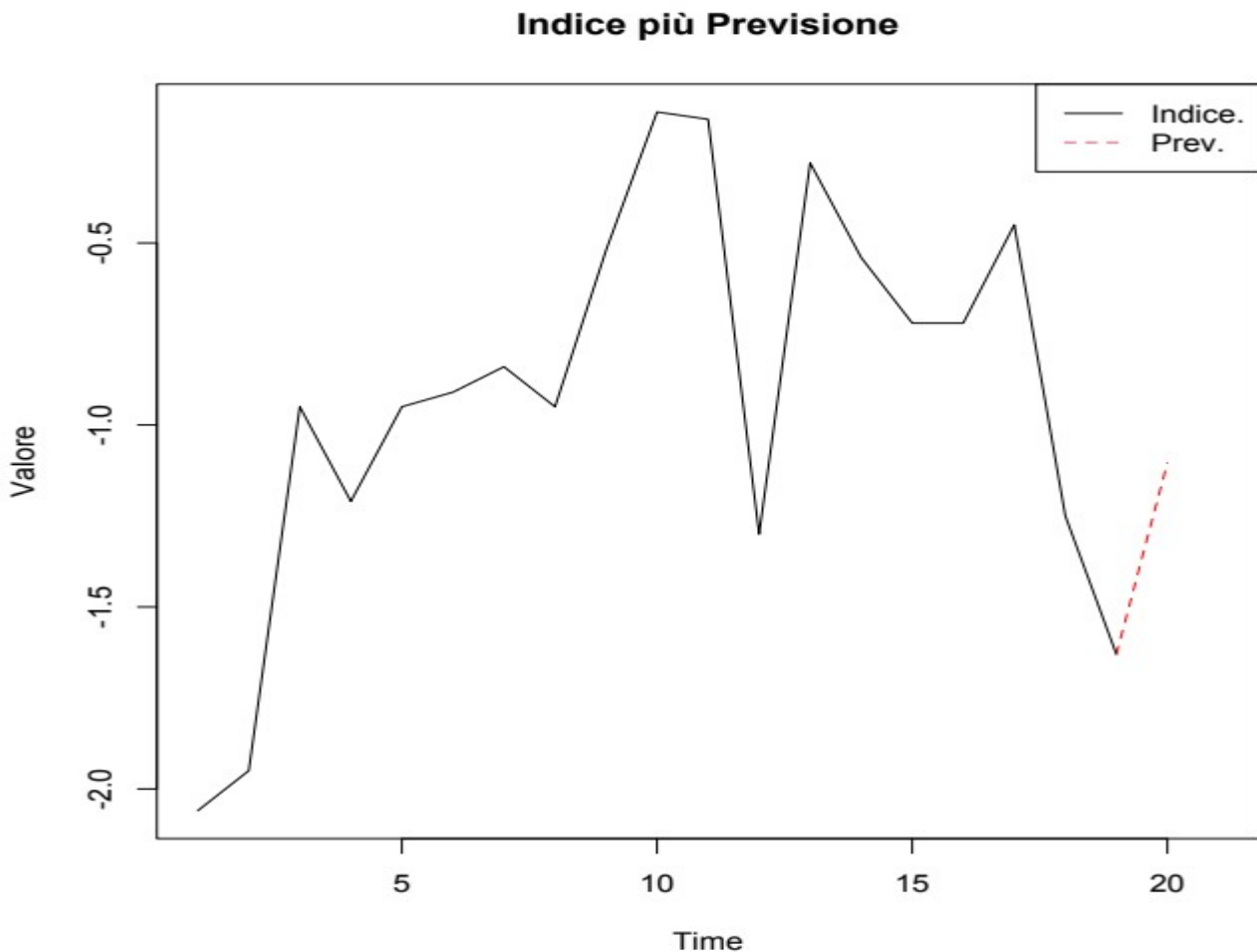
L'errore minimo generato dalla rete è pari a 0.666937 ed è stato raggiunto alla iterazione numero 577. Di seguito si presenta il grafico della diagnostica effettuata sul data-set di test (costituito dalle ultime tre osservazioni) con l'obiettivo di valutare la capacità di adattamento della rete, addestrata sul data-set di training (composta dalle prime sedici osservazioni).

## PLOT BONTA' ADATTAMENTO



Dal grafico si può vedere la qualità dell'adattamento sulla serie dei dati di test che genera un coefficiente di correlazione che si approssima ad un valore di 0.98, calcolato col metodo di *Pearson*. Le rilevazioni usate (dati test ed approssimazioni) sono state standardizzate per un migliore impatto grafico della bontà di adattamento. L'analisi prosegue con la stima del valore che l'indice potrà assumere allo step successivo. A tal fine è stata utilizzata la funzione *compute* (packages *neuralnet*) che, sulla base della rete implementata, restituisce una previsione pari a  $-1,10422$ . Di seguito si presenta il grafico dell'indice osservato, accompagnato dalla previsione generata per lo step successivo, corrispondente al valore che l'indice potrà assumere per l'anno successivo.

## GRAFICO INDICE E PREVISIONE



La rete, sulla base dell'andamento storico dell'indice, genera una previsione in crescita rispetto all'ultimo valore osservato. Sulla base di questa indicazione, il sistema co-integrato suggerisce che tale variazione qualitativa al rialzo sarà determinata da modifiche negli indicatori, secondo le combinazioni specificate dal vettore di co-integrazione. La velocità di aggiustamento verso l'equilibrio sarà invece determinata dai coefficienti di *loading* associati ai singoli indicatori. Nell'impostazione proposta in questo lavoro, il passo successivo consisterebbe nell'implementazione di una rete neurale finalizzata alla previsione della combinazione lineare associata al sistema co-integrato, al fine di comprendere in che modo gli indicatori interagiranno per generare la variazione stimata dell'indice. Nel caso in esame, poiché la combinazione lineare stazionaria generata dal sistema co-integrato presenta attualmente valori prossimi agli estremi superiori della sua distribuzione, è ragionevole ipotizzare che la variazione al rialzo dell'indice sarà indotta da una convergenza verso il basso del meccanismo di correzione dell'errore. Verrà comunque implementata una rete neurale per prevedere l'andamento della combinazione lineare, con l'obiettivo di illustrare il funzionamento dell'apparato

predittivo proposto. Tuttavia, la limitata disponibilità di osservazioni non consente al momento di eseguire una simulazione di previsione mediante il troncamento delle serie storiche.

Tale esercizio sarà condotto in una fase successiva, una volta che l'intero schema predittivo sarà applicato a serie storiche aumentate, costruite in coerenza con le traiettorie delle serie originali.

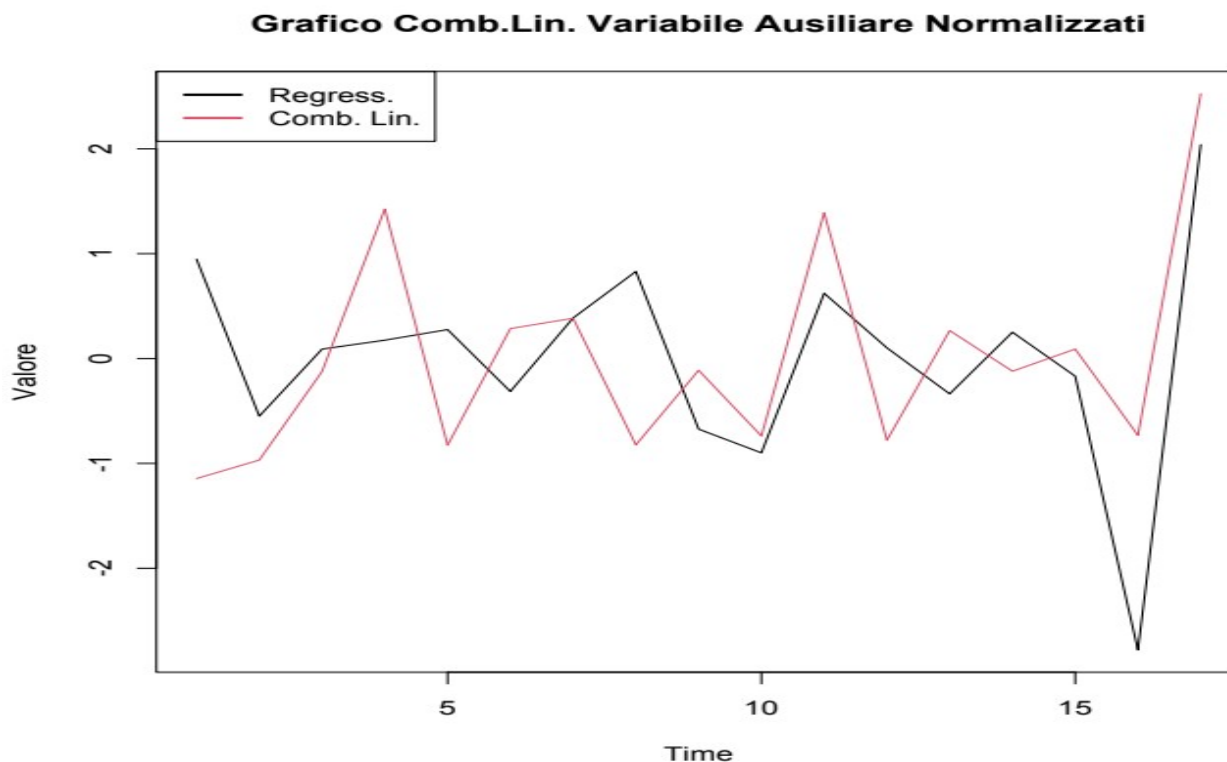
Poiché è necessario disporre di una variabile che funga da regressore per l'addestramento della rete neurale destinata alla previsione della combinazione lineare derivante dal modello co-integrato, si procederà alla costruzione di una serie storica ausiliaria appositamente definita a tale scopo.

La serie ausiliaria sarà costruita come differenza prima di una combinazione lineare degli indicatori del dominio, i cui coefficienti saranno ottenuti dalla varianza spiegata di un'Analisi delle Componenti Principali (ACP) condotta sugli stessi indicatori. Di seguito si presenta il grafico della combinazione lineare stazionaria derivante dal sistema co-integrato, affiancato a quello della variabile ausiliaria costruita mediante i coefficienti riportati di sotto:

#### COEFFICIENTI VARIANZA SPIEGATA

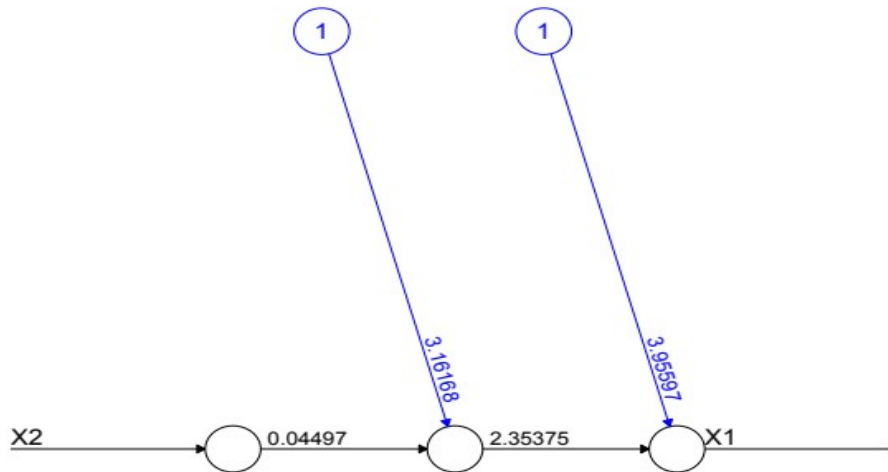
PC1	PC2	PC3	PC4
0.84164	0.11747	0.02701	0.01388

#### GRAFICO COMB. LIN. CON REG. AUSILIARE



Di seguito si mostra la rete neurale implementata per la previsione dell'andamento futuro della combinazione lineare.

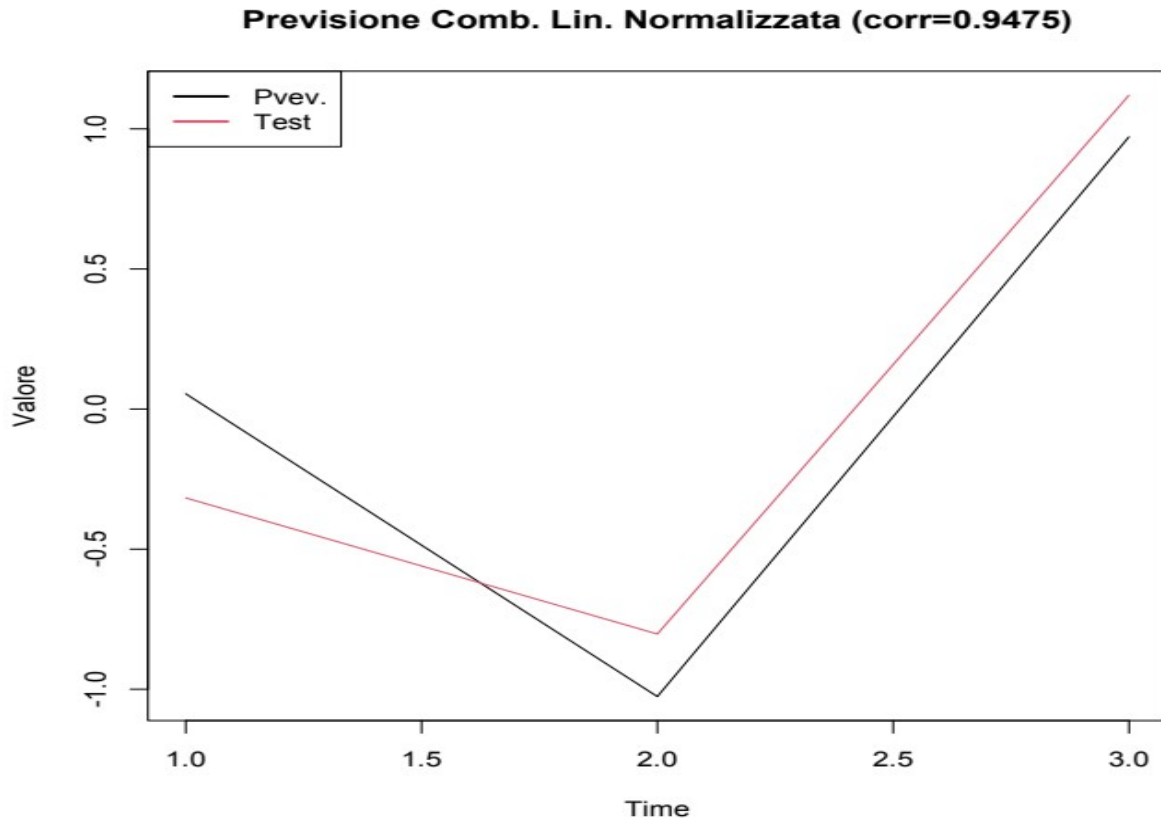
### PLOT RETE IMPLEMENTATA CON MINIMO ERRORE



Error: 5.301459 Steps: 41

L'errore minimo ottenuto dalla rete è pari a 5,301459 ed è stato raggiunto all'iterazione numero 41. Di seguito si presenta il grafico della diagnostica condotta sul data-set di test (costituito dalle ultime tre osservazioni), con l'obiettivo di valutare la capacità di adattamento della rete addestrata sul data-set di training, composto dalle prime quattordici osservazioni.

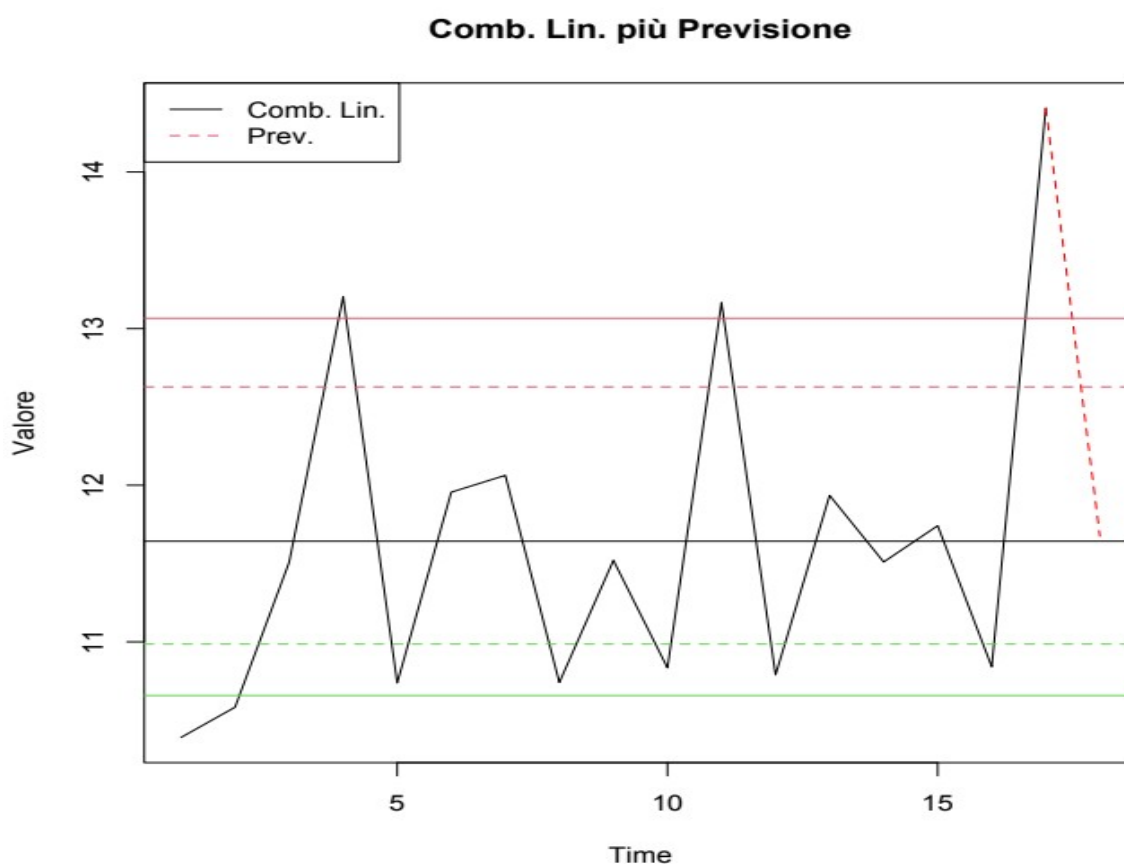
## PLOT BONTA' ADATTAMENTO



Anche in questo caso, dal grafico si osserva una buona qualità di adattamento sul data-set di test, con un coefficiente di correlazione pari a circa 0.95, calcolato secondo il metodo di *Pearson*. Le rilevazioni utilizzate (sia i dati di test sia le approssimazioni) sono state standardizzate, al fine di migliorarne la leggibilità grafica dell'accuratezza del modello.

L'analisi prosegue con la stima del valore che la combinazione lineare potrà assumere allo step successivo. A tal fine è stata utilizzata nuovamente la funzione *compute* che, sulla base della rete implementata, restituisce una previsione al ribasso. Di seguito si presenta il grafico della combinazione lineare ottenuta dal modello di co-integrazione, affiancato dalla previsione generata per lo step successivo.

## GRAFICO COMB. LIN. E PREVISIONE



Come previsto, dato l'elevato livello raggiunto dalla combinazione lineare, la stima evidenzia una tendenza alla convergenza verso il basso del suo valore. Una volta completato l'intero sistema predittivo, sarà possibile formulare previsioni sull'andamento complessivo del fenomeno oggetto di studio. In particolare, è stata ottenuta una previsione dell'indice che indica un aumento al passo successivo. Data la natura co-integrata degli indicatori, tale crescita sarà determinata dalla convergenza verso il basso del meccanismo di correzione dell'errore e le variabili considerate dovrebbero contribuire alla variazione osservata nel modo indicato precedentemente.

L'apparato predittivo esposto si fonda sulla idea espressa all'inizio di questo capitolo, secondo cui, per costruzione, in seno agli indicatori di un dominio del BES e del BesT possa esistere co-integrazione tra gli stessi indicatori.

L'utilizzo delle reti neurali si è rivelato funzionale a superare la scarsa duttilità della combinazione lineare derivante dal processo di co-integrazione. Tale rigidità è imputabile alla limitata numerosità campionaria tipica di questa tipologia di dati, che non consente di sfruttare appieno il potenziale predittivo del concetto stesso di co-integrazione.

Inoltre, le reti neurali sono state impiegate per effettuare previsioni sull'indice di riferimento del dominio oggetto di analisi.

La limitata numerosità campionaria degli indicatori non ha consentito di implementare la struttura predittiva su serie troncate, al fine di verificarne la bontà. L'unica modalità di validazione possibile consiste nell'attendere la disponibilità dei dati relativi al periodo successivo. Ciò equivale a testare la metodologia direttamente sul campo, senza la possibilità di retrocedere nel tempo cristallizzando un periodo passato da utilizzare come presente, così da applicare la procedura e confrontare l'adattamento del modello a dati già disponibili ma non inseriti nel modello. Data la natura sperimentale di questo lavoro, si è scelto di considerare un contesto reale caratterizzato da una disponibilità di dati sensibilmente maggiore, al fine di verificare l'efficacia dell'apparato predittivo implementato in presenza di una più ampia numerosità campionaria. Questa proposta non rappresenta soltanto un esercizio tecnico, ma una possibilità concretamente realizzabile, in quanto la rilevazione degli indicatori con una cadenza temporale più fine rispetto a quella annuale risulta perfettamente attuabile, date le caratteristiche stesse degli indicatori. Infatti, tali indicatori rilevano l'andamento di fenomeni sociali che presentano variazioni anche nel corso dell'anno, ovvero entro l'intervallo temporale attualmente utilizzato per la rilevazione dei dati. Pertanto, sarebbe possibile effettuare una rilevazione di questi fenomeni su un intervallo temporale significativamente più breve dell'anno, grazie a sistemi automatizzati di raccolta dati. Sulla base di questa possibilità, si è deciso di simulare una situazione ipotetica (ma realizzabile) in cui la frequenza di rilevazione sia settimanale, considerando che nella pratica potrebbe, magari, essere attuabile una cadenza anche infrasettimanale.

Si procederà, dunque, all'implementazione del sistema — precedentemente applicato alle serie originali — anche sulle serie aumentate generate.

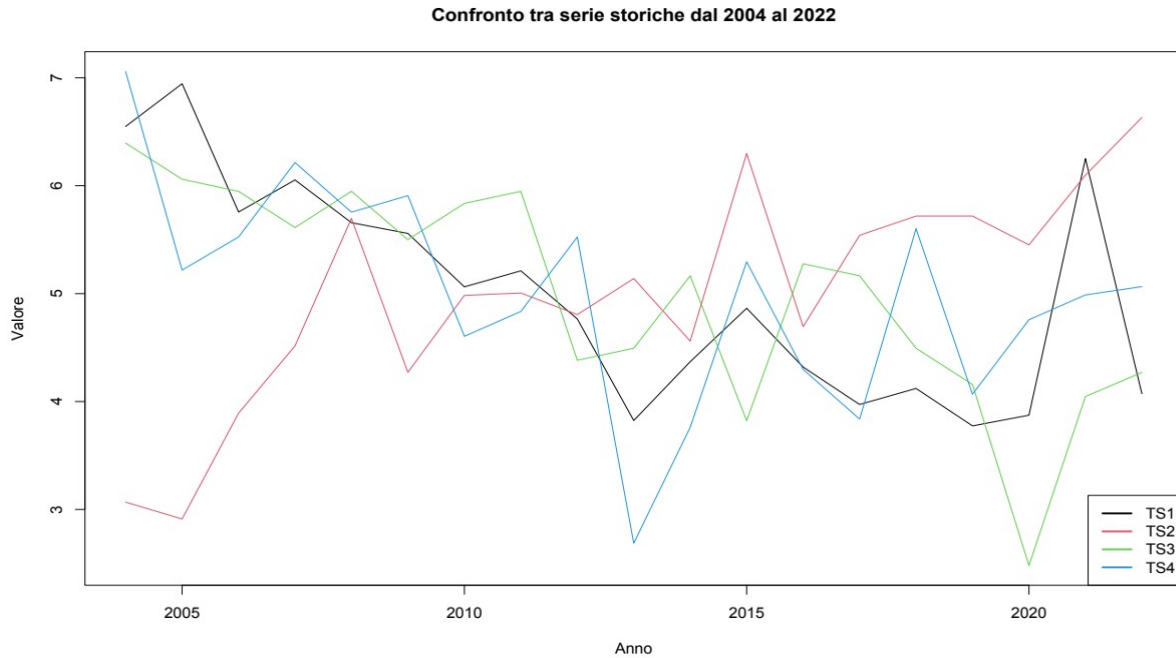
La generazione delle serie aumentate è stata effettuata considerando le statistiche di media e varianza delle rilevazioni degli indicatori, analizzati a coppie, e producendo valori casuali compresi tra ciascun valore e il successivo, mantenendo le stesse caratteristiche statistiche (media e varianza) della coppia di riferimento.

Le serie aumentate sono state generate ipotizzando una situazione in cui il time-frame di rilevazione dei dati fosse settimanale anziché annuale.

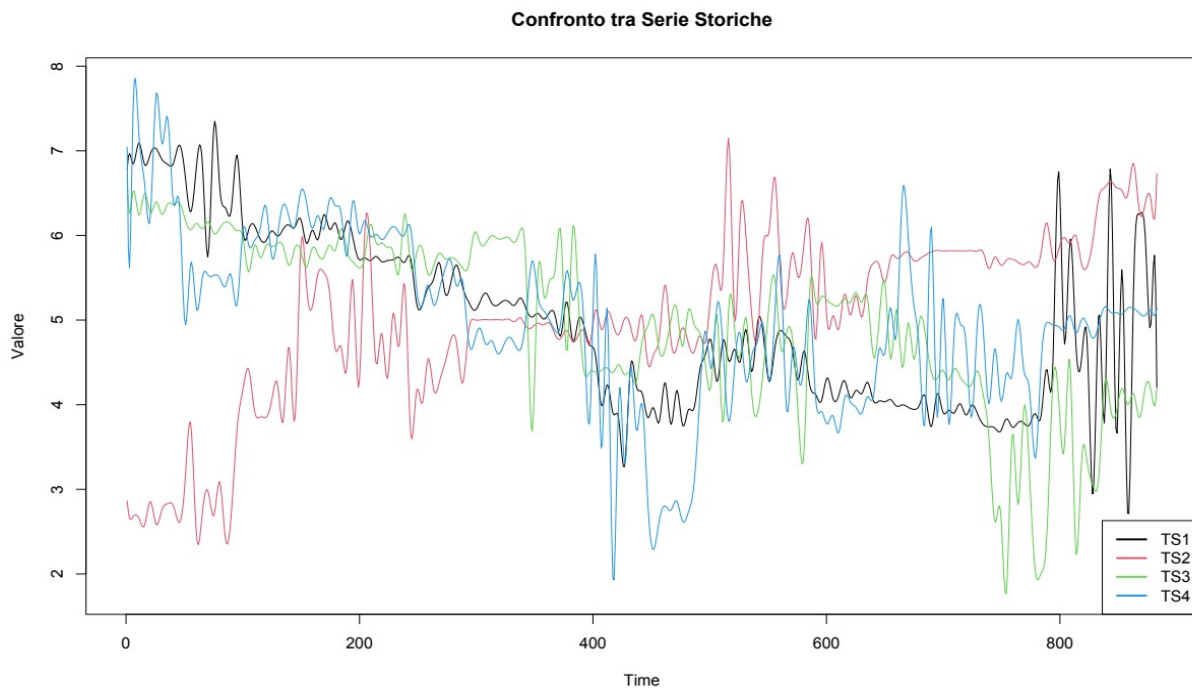
Questa simulazione ha permesso di ottenere traiettorie con una numerosità campionaria più elevata, ma con caratteristiche molto simili a quelle delle serie originali. L'obiettivo, infatti, è quello di mantenere il più possibile invariate le proprietà delle serie storiche originarie, così da poter applicare alle serie aumentate gli stessi concetti e strumenti utilizzati per le serie originali, dalle quali esse derivano. Inoltre, considerata l'invarianza sostanziale delle proprietà delle traiettorie, si può ragionevolmente ipotizzare che anche le serie aumentate, analogamente a quelle originali, possano risultare generate da un medesimo processo generatore dei dati, potendole considerare pertanto co-integrate. Si inizierà presentando i grafici delle serie originali e di quelle aumentate, al fine di evidenziare graficamente la sostanziale

similarità tra le due rappresentazioni. Le etichette assegnate alle serie aumentate coincidono con quelle utilizzate per le serie originali.

## GRAFICO SERIE ORIGINALI



## GRAFICO DELLE SERIE STORICHE AUMENTATE



Da una prima analisi grafica emerge una sostanziale similarità in termini di traiettorie, il che consente di ipotizzare che l'aumento della numerosità campionaria non abbia alterato le caratteristiche stocastiche delle serie originali. Come nell'analisi precedentemente condotta, anche in questo caso si procederà a verificare la non stazionarietà delle serie aumentate — già osservabile a livello grafico — mediante il test *ADF*, utilizzando un numero di ritardi determinato, anche in questo caso, secondo la regola di *Schwert*.

## TS1

### Augmented Dickey-Fuller Test

data: TS1

Dickey-Fuller = -1.7886, Lag order = 16, p-value = 0.6678

alternative hypothesis: stationary

## TS2

### Augmented Dickey-Fuller Test

data: TS2

Dickey-Fuller = -1.7109, Lag order = 16, p-value = 0.7007

alternative hypothesis: stationary

## TS3

### Augmented Dickey-Fuller Test

data: TS3

Dickey-Fuller = -1.9417, Lag order = 16, p-value = 0.603

alternative hypothesis: stationary

## TS4

### Augmented Dickey-Fuller Test

data: TS4

Dickey-Fuller = -1.8347, Lag order = 16, p-value = 0.6483

alternative hypothesis: stationary

I valori dei  $p$ -value confermano la natura non stazionaria delle quattro traiettorie analizzate. È pertanto possibile ipotizzare l'esistenza di una relazione di co-integrazione tra le serie. A tal fine, si procederà ora con l'implementazione del sistema co-integrato tramite la procedura di *Johansen*, basata sul test del massimo autovalore.

## PROCEDURA PER VERIFICA RANGO DI CO-INTEGRAZIONE

```
#####
# Johansen-Procedure #
#####
```

Test type: maximal eigenvalue statistic (lambda max) , with linear trend

Eigenvalues (lambda):

```
[1] 0.048324218 0.019654281 0.014673082 0.001848263
```

Values of teststatistic and critical values of test:

	test	10pct	5pct	1pct
r <= 3	1.62	6.50	8.18	11.65
r <= 2	12.90	12.91	14.90	19.19
r <= 1	17.33	18.90	21.07	25.75
r = 0	43.24	24.78	27.14	32.14

Eigenvectors, normalised to first column:

(These are the cointegration relations)

	ts.TS1..l1	ts.TS2..l1	ts.TS3..l1	ts.TS4..l1
ts.TS1..l1	1.00000000	1.00000000	1.00000	1.0000000
ts.TS2..l1	0.61590316	-0.52269380	-37.99211	-0.5638250
ts.TS3..l1	0.05648728	-1.66553097	-12.51940	0.2310044
ts.TS4..l1	-0.70409248	0.03976334	-46.79348	-0.4223451

Weights W:

(This is the loading matrix)

	ts.TS1..l1	ts.TS2..l1	ts.TS3..l1	ts.TS4..l1
ts.TS1..d	-3.230321e-03	-0.0004651635	9.950070e-06	-9.764169e-05
ts.TS2..d	-4.880092e-04	0.0005305807	6.949755e-06	1.360065e-04
ts.TS3..d	3.140425e-05	0.0007407303	8.333763e-07	-1.195105e-04
ts.TS4..d	2.488152e-03	-0.0001634078	1.737224e-05	-8.397614e-06

```
> as.integer(mean(IND$selection))
```

```
[1] 10
```

```
> adf.test(Comb. Lin.);
```

Augmented Dickey-Fuller Test

data: Comb. Lin.

Dickey-Fuller = -4.658, Lag order = 15, p-value = 0.01

alternative hypothesis: stationary

Messaggio di avvertimento:

In adf.test(Comb. Lin.) : p-value smaller than printed p-value

Anche in questo caso, ed in modo ancora più evidente, un'analisi preliminare consente di individuare la presenza di un rango di co-integrazione, poiché il valore della statistica test pari a 17.33 risulta inferiore alle soglie critiche. Inoltre, l'ipotesi di assenza di co-integrazione può essere rifiutata ad ogni livello di significatività con una statistica test pari a 43.24.

I risultati ottenuti suggeriscono l'esistenza di un trend di lungo periodo condiviso tra le quattro variabili, coerente con l'ipotesi di un processo generatore comune che ne determina l'evoluzione. Considerando che le serie aumentate provengono da dati originari co-integrati, si può concludere che tale proprietà sia stata mantenuta, confermando la coerenza del processo stocastico tra le due versioni delle serie. Le serie aumentate sono state generate al fine di simulare una rilevazione ipoteticamente più fine. Dal momento che i dati fanno sempre riferimento al dominio *Salute*, è lecito considerarli come manifestazione empirica del concetto latente che il dominio mira a rappresentare. Anche in questo caso, dunque, le serie possono essere interpretate come espressione di un unico processo generatore sottostante, riconducibile al concetto di *Salute*.

Il modello VAR implementato per l'analisi presenta un numero di ritardi pari a 10; di conseguenza, il VECM risulta essere di ordine 9 per quanto riguarda le variabili in forma differenziata. È importante precisare, anche in questo caso, che eventuali discrepanze in termini di affidabilità delle stime e di robustezza delle inferenze rispetto ai valori ritenuti ottimali devono essere interpretate alla luce della qualità e della natura dei dati utilizzati. L'attenzione va posta, in particolare, sulla capacità complessiva del metodo proposto di cogliere le dinamiche del fenomeno in esame. In altri termini, qualora si disponesse di dati osservati con una frequenza di rilevazione più fine e rappresentativi dell'effettiva evoluzione del fenomeno, la metodologia adottata sarebbe presumibilmente in grado di generare risultati ancora più robusti. Il fatto che, anche con dati simulati in modo da riprodurre il più fedelmente possibile la realtà, si ottengano risultati statisticamente validi, può essere considerata come una conferma della possibile solidità dell'approccio proposto. L'esame dei risultati ottenuti mediante la procedura di *Johansen* conferma che, anche in questo caso, il primo vettore di co-integrazione, associato al massimo autovalore, rappresenta la relazione di co-integrazione dominante e la principale componente descrittiva della dinamica di lungo periodo. Dall'analisi del vettore di co-integrazione emerge che la variabile TS4 è quella che contribuisce maggiormente al mantenimento della relazione di equilibrio, presentando un coefficiente pari a -0.70409. Il segno negativo del coefficiente indica che il suo aggiustamento avviene in direzione opposta rispetto a quello delle altre variabili. Di conseguenza, eventuali variazioni in TS4

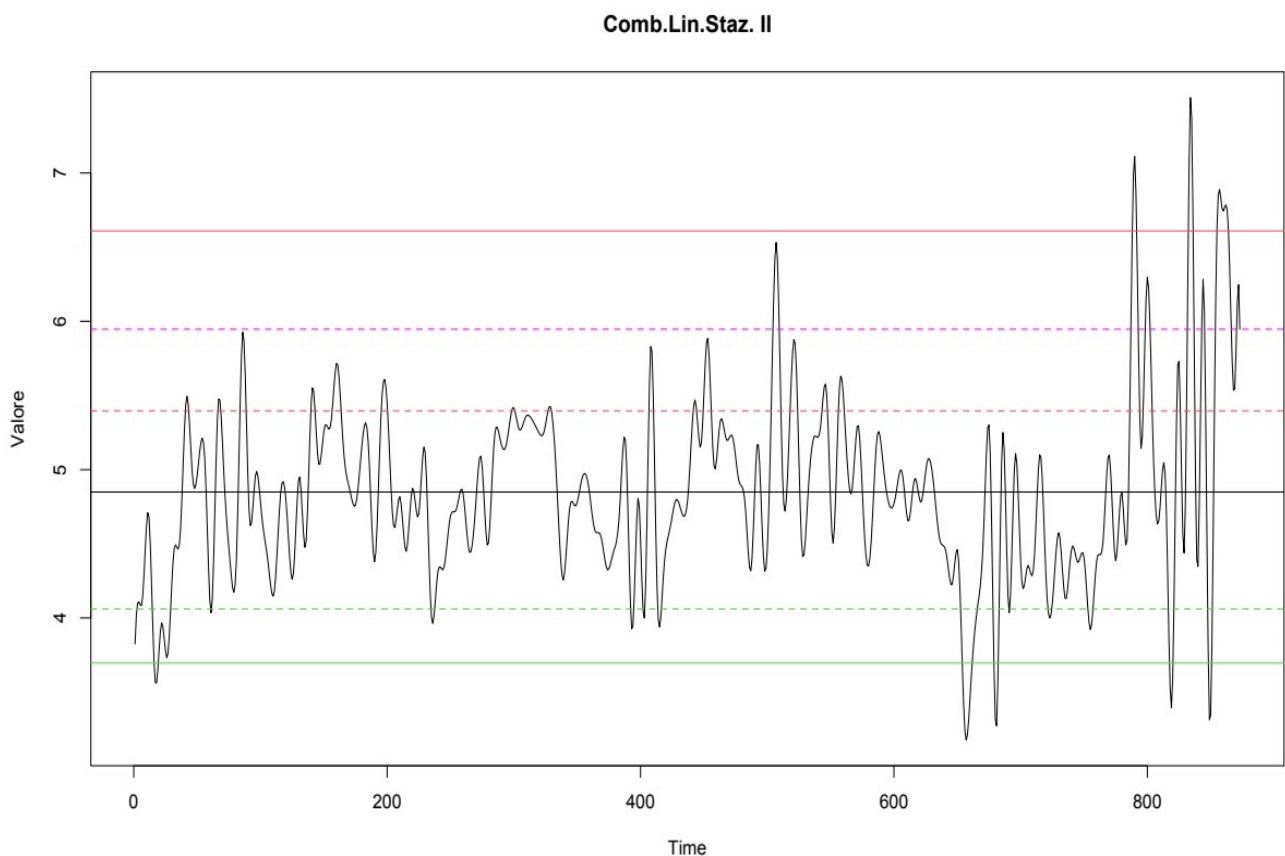
richiederanno movimenti compensativi, e di segno contrario, da parte delle restanti variabili del sistema.

La seconda variabile che maggiormente contribuisce al mantenimento della relazione di equilibrio è la variabile TS2 avendo un coefficiente di pari a 0.6159. Dunque insieme alla variabile TS4 saranno le due componenti del sistema che più contribuiranno all'equilibrio del sistema, anche se con versi opposti. Sebbene meno rilevante rispetto alle altre componenti, la variabile TS3 potrebbe fungere da disturbo al mantenimento dell'equilibrio di lungo periodo, con un coefficiente pari a 0.05648. Il segno positivo potrebbe indicare che il suo movimento è orientato nello stesso verso di quello di TS1 e TS2. Si ricorda che TS1 presenta, per convenzione, un coefficiente pari a 1, trattandosi della variabile sulla quale è stata effettuata la normalizzazione del vettore di co-integrazione. In sintesi, sulla base dei valori e dei segni dei coefficienti di co-integrazione, le variabili TS1, TS2 e TS3 si dovrebbero muovere nella stessa direzione in termini di contributo alla relazione di equilibrio di lungo periodo, mentre la variabile TS4 agirà in direzione opposta, assumendo un ruolo correttivo nel sistema. Il tutto però dovrà essere studiato alla luce dei valori e dei segni dei coefficienti di *loading*. Passando ad analizzare proprio la matrice dei *loading* riferiti alle variabili del sistema, si può vedere come quelle che partecipano più velocemente alla correzione della relazione di equilibrio sono la variabile TS1 e TS4 avendo rispettivamente coefficienti pari a -0.00323 e 0.002488. La variabile TS1, stando al segno dell'errore di lungo periodo, tenderà ad effettuare un movimento che condurrà alla correzione dell'errore della relazione di equilibrio avendo il coefficiente negativo. La variabile TS4 invece, con un coefficiente positivo tenderà ad effettuare un movimento al rialzo, contribuendo a ridurre l'errore. Proseguendo con l'analisi degli aggiustamenti di breve periodo, si osserva che, dopo TS1 e TS4, la variabile con la maggiore velocità di aggiustamento risulti essere TS2 con un valore del coefficiente di *loading* pari a -0.000488. Sebbene il valore sia sensibilmente inferiore rispetto ai primi due, la sua direzione negativa suggerisce che TS2 interviene comunque, seppur in modo più contenuto, per contrastare eventuali deviazioni dal trend di lungo periodo. Tale comportamento potrebbe riflettere una dinamica più lenta di aggiustamento della variabile o, in alternativa, una minore sensibilità rispetto agli scostamenti dalla relazione di equilibrio di lungo periodo. In entrambi i casi, TS2 risulterebbe meno reattiva nel breve periodo, contribuendo in modo più marginale al riequilibrio del sistema. Infine, la variabile TS3 presenta un coefficiente di *loading* estremamente contenuto pari a 0,0000314 indicando una velocità di aggiustamento trascurabile rispetto agli scostamenti dalla relazione di equilibrio. Il segno positivo del coefficiente suggerisce che, essendo il sistema in

presenza di un errore positivo di equilibrio, la variabile tende a muoversi nella stessa direzione dello scostamento, contribuendo poco o addirittura andando a disturbare il meccanismo di correzione. Comunque, dato il valore assoluto del suo *loading*, la sua dinamica appare sostanzialmente neutra nel breve periodo. Questa evidenza può supportare l'interpretazione di TS3 come variabile la cui evoluzione sembrerebbe non essere guidata direttamente dal processo di aggiustamento interno definito dalla relazione di co-integrazione.

Dai risultati ottenuti, si può concludere che la variabile TS4 emerge come la componente con il maggiore impatto sul meccanismo di correzione dell'errore dell'intero sistema. Le restanti variabili, in particolare TS3, mostrano invece una influenza più contenuta sul processo di aggiustamento, evidenziando una maggiore inerzia rispetto alle variazioni nella relazione di equilibrio, oppure una possibile esposizione a deboli dinamiche esogene non pienamente catturate dal modello stimato. Tale comportamento risulta tendenzialmente coerente con quanto osservato anche nel sistema stimato a partire dalle serie non aumentate. Proseguendo nell'analisi, si riporterà di seguito il grafico della combinazione lineare derivante dal sistema co-integrato appena implementato, utile per una valutazione visiva della stabilità della relazione di lungo periodo.

## GRAFICO DELLA COMBINAZIONE LINEARE



Dall'osservazione della combinazione lineare derivante dal sistema stimato emerge chiaramente una componente stazionaria di ordine II, confermando la natura co-integrata delle variabili. Tale evidenza è supportata anche dal test di *Dickey-Fuller* aumentato (*ADF*) condotto direttamente sulla serie combinata che, dato il *p-value*, permette di rifiutare l'ipotesi nulla di presenza di radice unitaria a tutti i livelli di significatività confermando la stazionarietà della relazione di equilibrio di lungo periodo. Di seguito viene riportato il test *ADF* effettuato sulla serie combinata:

Augmented Dickey-Fuller Test

data: Comb. Lin.

Dickey-Fuller = -4.658, Lag order = 15, p-value = 0.01

alternative hypothesis: stationary

Messaggio di avvertimento:

In adf.test(Comb. Lin.) : p-value smaller than printed p-value

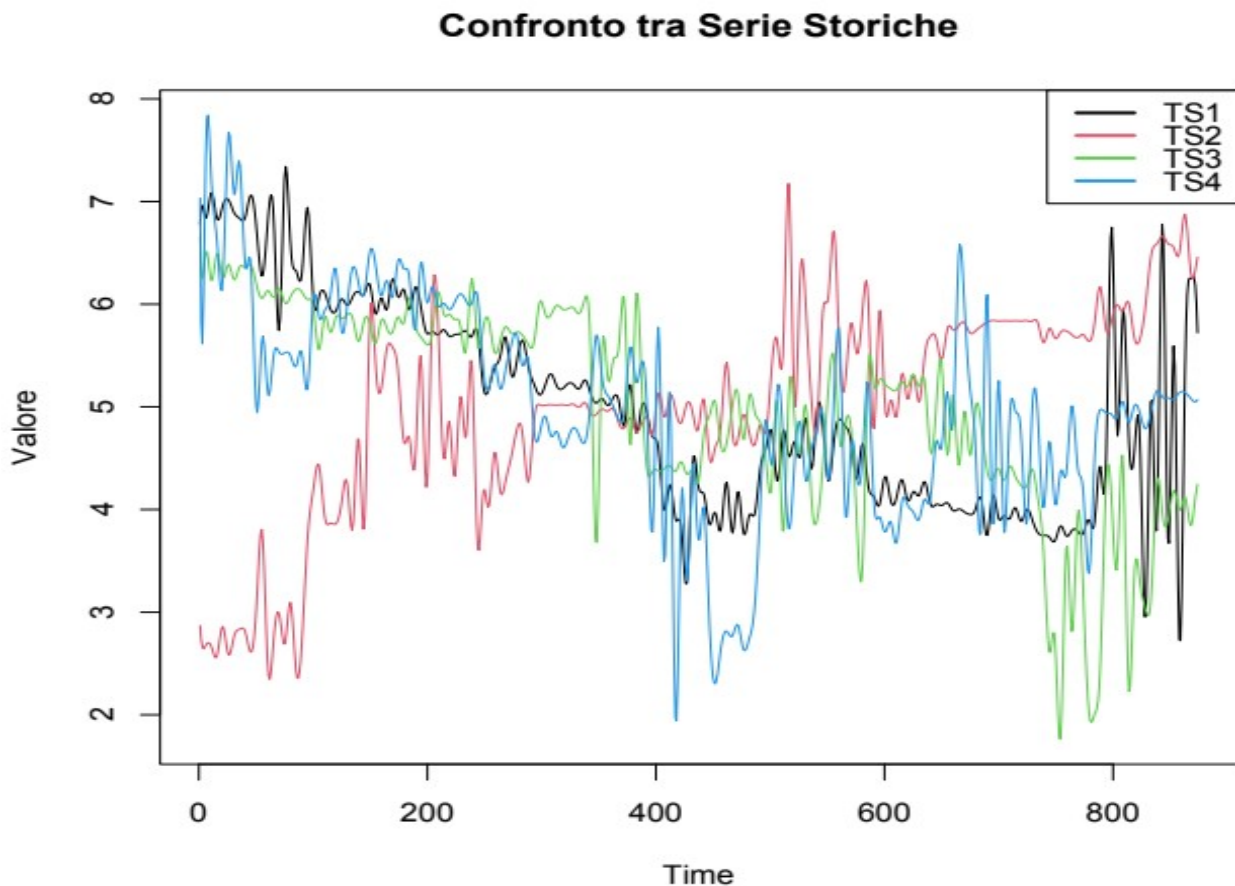
Nell'analisi condotta in precedenza sulle serie non aumentate, si era evidenziato come la limitata numerosità campionaria degli indicatori potesse compromettere la capacità della combinazione lineare di offrire una base sufficientemente flessibile per la formulazione di previsioni attendibili sulle traiettorie future.

Per ovviare a questa criticità, era stata implementata una rete neurale, con l'obiettivo di stimare l'andamento atteso della combinazione lineare al passo successivo. In questo nuovo contesto, invece, la combinazione lineare stimata appare più stabile e regolare, mostrando una struttura che ne favorisce l'utilizzo a fini predittivi anche senza ricorrere a tecniche di apprendimento automatico. Alla luce della maggiore stabilità e regolarità mostrata dalla combinazione lineare stimata, non si procederà all'implementazione di una rete neurale per la previsione dell'andamento futuro, ritenendo che la dinamica osservata sia già sufficientemente informativa per eventuali utilizzi predittivi. Avendo ora a disposizione una maggiore numerosità campionaria, si potrà procedere alla validazione del sistema predittivo implementato, estendendolo alle serie troncate. Questo permetterà di cristallizzare un momento passato e di riportare l'analisi indietro nel tempo, come se l'evento passato si stesse verificando in questo preciso istante. Dal grafico della combinazione lineare precedente si osserva che il sistema aveva dato un segnale di convergenza verso il valore di equilibrio circa una decina di rilevazioni prima.

Si implementerà quindi il sistema utilizzando le serie troncate a quel momento, eliminando le ultime dieci rilevazioni e inserendo nel modello solo i dati disponibili dieci rilevazioni prima. Così facendo, ci si riporta a quel preciso istante, quando le

rilevazioni successive non erano ancora note, cristallizzando quel momento passato nel presente. Si inserisce ora il grafico delle serie troncate.

## GRAFICO SERIE STORICHE TRONCATE



Il troncamento delle traiettorie, come detto, ci riporta a quel preciso istante; di conseguenza, ogni analisi effettuata su tali dati non fa altro che riportarci indietro nel tempo, replicando fedelmente la disponibilità e la tipologia di dati presenti in quel momento.

Di seguito la diagnostica della procedura di *Johansen* sul sistema implementato con le serie rappresentate nel grafico.

## PROCEDURA PER VERIFICA RANGO DI CO-INTEGRAZIONE

```
#####  
# Johansen-Procedure #  
#####
```

Test type: maximal eigenvalue statistic (lambda max) , with linear trend

Eigenvalues (lambda):

```
[1] 0.041255725 0.018225688 0.009372594 0.005246555
```

Values of teststatistic and critical values of test:

	test	10pct	5pct	1pct
r <= 3	4.54	6.50	8.18	11.65
r <= 2	8.14	12.91	14.90	19.19
r <= 1	15.89	18.90	21.07	25.75
r = 0	36.40	24.78	27.14	32.14

Eigenvectors, normalised to first column:  
(These are the cointegration relations)

	ts.TS1..l1	ts.TS2..l1	ts.TS3..l1	ts.TS4..l1
ts.TS1..l1	1.0000000	1.0000000	1.0000000	1.0000000
ts.TS2..l1	0.5834031	-0.9295678	0.4490118	2.28062296
ts.TS3..l1	0.1851376	-1.3927434	-0.3786500	-2.09055840
ts.TS4..l1	-0.8385700	-0.6919203	0.2362945	0.08823816

Weights W:

(This is the loading matrix)

	ts.TS1..l1	ts.TS2..l1	ts.TS3..l1	ts.TS4..l1
ts.TS1..d	-0.0016417544	-6.834898e-05	-0.0005826895	4.641700e-05
ts.TS2..d	-0.0007786681	6.128882e-04	-0.0001292179	-8.083328e-05
ts.TS3..d	-0.0004852584	5.693676e-04	0.0002675885	9.960667e-05
ts.TS4..d	0.0026641350	3.656382e-04	-0.0006272017	1.039789e-05

```
> as.integer(mean(IND$selection))
```

```
[1] 10
```

```
> adf.test(Comb.Lin.);
```

Augmented Dickey-Fuller Test

data: Comb. Lin.

Dickey-Fuller = -4.4158, Lag order = 15, p-value = 0.01

alternative hypothesis: stationary

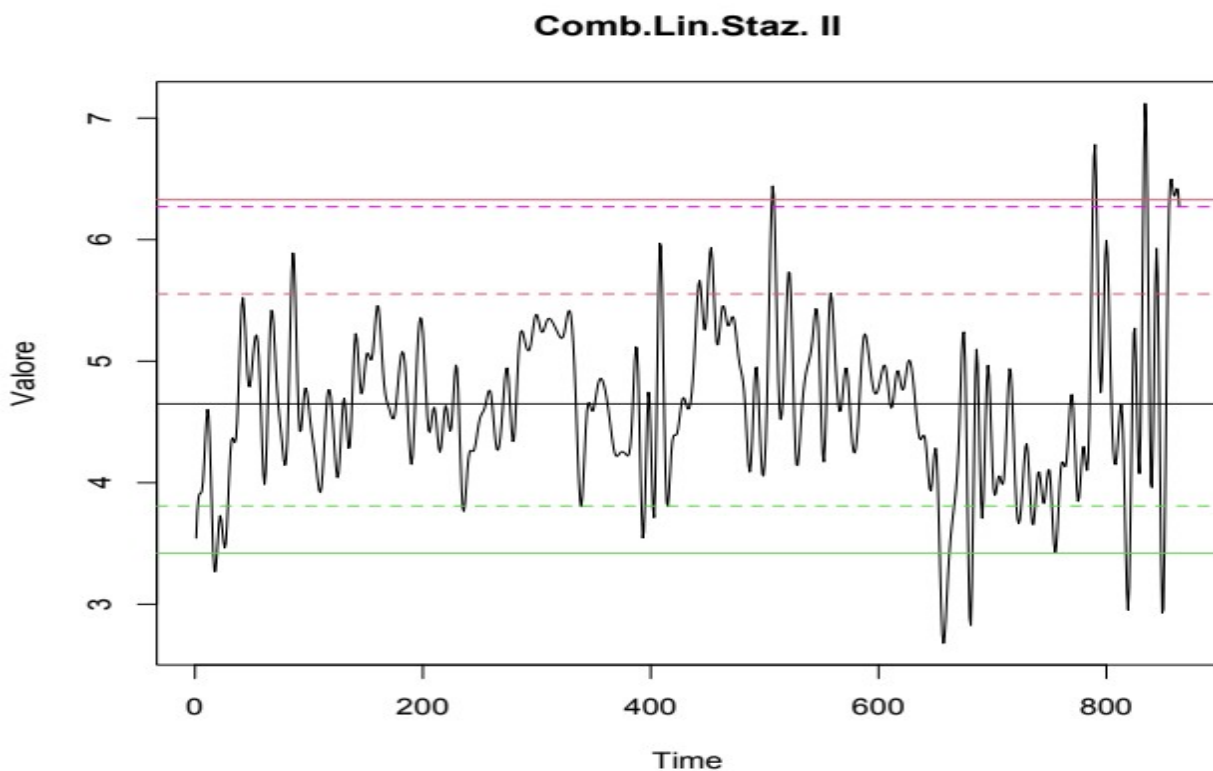
Messaggio di avvertimento:

In adf.test(Comb. Lin.) : p-value smaller than printed p-value

Come evidenziato dalla procedura di *Johansen*, il rango di co-integrazione del sistema è pari ad 1, il che indica l'esistenza di una sola relazione di equilibrio di lungo periodo tra le quattro variabili analizzate. Il primo autovalore è il più elevato tra quelli stimati, e ciò suggerisce che la corrispondente combinazione lineare, ovvero il primo vettore di co-integrazione, rappresenti la relazione più forte e significativa

nel lungo periodo<sup>16</sup>. In questa relazione, le variabili TS1, TS2 e TS3 presentano coefficienti positivi, contribuendo in modo concorde al mantenimento dell'equilibrio, mentre TS4 ha un coefficiente negativo, agendo quindi in senso opposto. In termini relativi, TS1 è la variabile con il contributo più rilevante, seguita da TS4 (con effetto contrario) e da TS2. TS3 mostra invece un'influenza più contenuta. Analizzando i coefficienti di *loading* (matrice dei pesi), si osserva che le variabili TS1, TS2 e TS3 presentano coefficienti negativi, contribuendo all'aggiustamento del sistema verso l'equilibrio nel breve periodo. La variabile TS4 mostra la maggiore velocità di reazione (in valore assoluto) e presenta un coefficiente positivo. Alla luce del segno positivo dell'errore di lungo periodo, essa subirà un aumento dei suoi corsi contribuendo in modo preponderante alla correzione dell'errore. Si inserirà adesso il grafico della relazione di equilibrio derivante dal vettore di co-integrazione:

### GRAFICO COMBINAZIONE LINEARE STAZIONARIA



Come si osserva dal grafico, la relazione di equilibrio ha raggiunto un estremo superiore, indicato dalla linea continua rossa, che rappresenta una distanza pari a 2,6 volte la deviazione standard della combinazione lineare rispetto al suo valore medio.

<sup>16</sup> Nella procedura di *Johansen*, gli autovalori stimati indicano la forza delle relazioni di co-integrazione individuate. L'autovalore più elevato corrisponde al vettore di co-integrazione che spiega la componente stazionaria più significativa del sistema.

In accordo con quanto accennato in precedenza, e nell'ottica della teoria delle catastrofi, tale soglia viene fatta coincidere con l'insieme delle catastrofi previsto dalla teoria stessa. Questo livello segnala un punto critico in cui il sistema può subire un cambiamento qualitativo, passando da uno stato di disequilibrio a uno di nuovo equilibrio. Il modello implementato, considerando che la relazione di equilibrio di lungo periodo si trova attualmente al di sopra del suo valore medio, e tenuto conto dei coefficienti del vettore di co-integrazione, suggerisce che nei passi successivi il sistema tenderà a muoversi verso una condizione di equilibrio. Tale dinamica sarà guidata dalle interazioni tra le variabili, che evolveranno in base ai coefficienti del vettore di co-integrazione ed ai rispettivi loading, riflettendo i meccanismi di aggiustamento di lungo e breve periodo. Precisamente, le variabili TS1, TS2, TS3 subiranno una discesa dei loro corsi nei passi successivi contribuendo, dati i segni dei loro coefficienti di loading e dato il segno dell'errore, al movimento di aggiustamento nel breve periodo con la maggiore intensità apportata da TS1 e poi rispettivamente da TS2 e TS3. La variabile TS4, si muoverà al rialzo con una velocità in differenza superiore alle altre tre variabili contribuendo anche lei al movimento di aggiustamento. Al momento di tale rilevazione, le variabili (non normalizzate) assumono rispettivamente il seguenti valori:

TS1=21.1

TS2=32.54

TS3=8.44

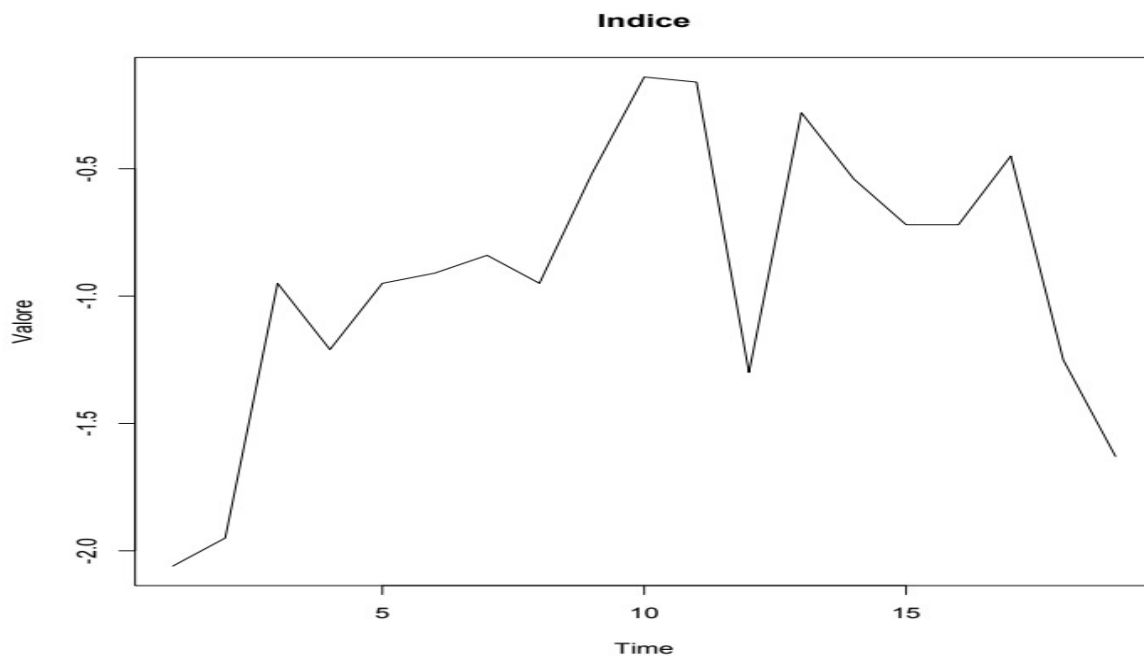
TS4=3.64

Dal quadro, ai passi successivi si dovrebbero trovare valori più bassi per le prime tre variabili, ed un valore maggiore per la quarta variabile. I tassi di variazione maggiori dovrebbero subirli la variabile TS4 e TS1 seguite da TS2 e TS3.

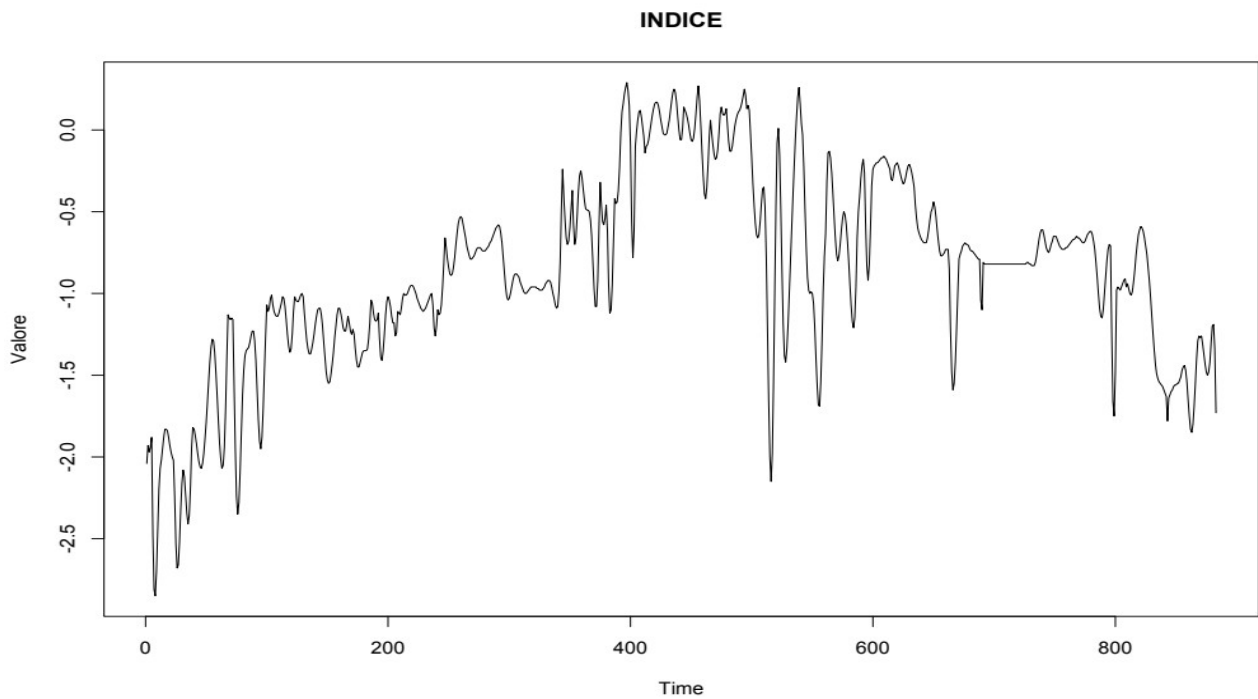
Si procederà adesso al calcolo dell'indice sintetico delle serie troncate utilizzate per questa implementazione. Il metodo usato sarà lo stesso utilizzato per la costruzione dell'indice sulla base delle serie originali.

Di seguito verrà presentato, in primo luogo, il grafico comparativo dell'indice costruito sulle traiettorie originali e su quelle aumentate, al fine di valutarne la similarità in termini di andamento e struttura. Successivamente, si mostrerà il grafico dell'indice ottenuto a partire dalle serie troncate, impiegate nell'implementazione del modello co-integrato illustrato poc'anzi.

## GRAFICO INDICE SERIE ORIGINALI

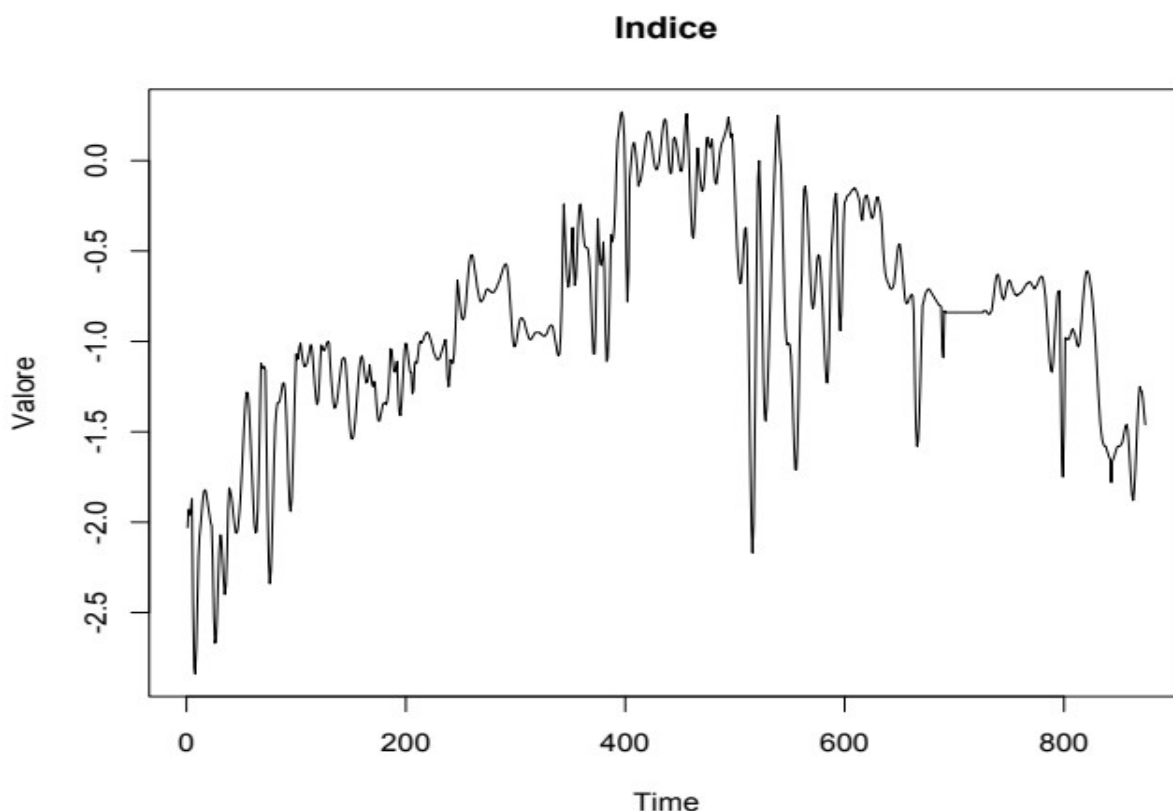


## GRAFICO INDICE SERIE ALLUNGATE



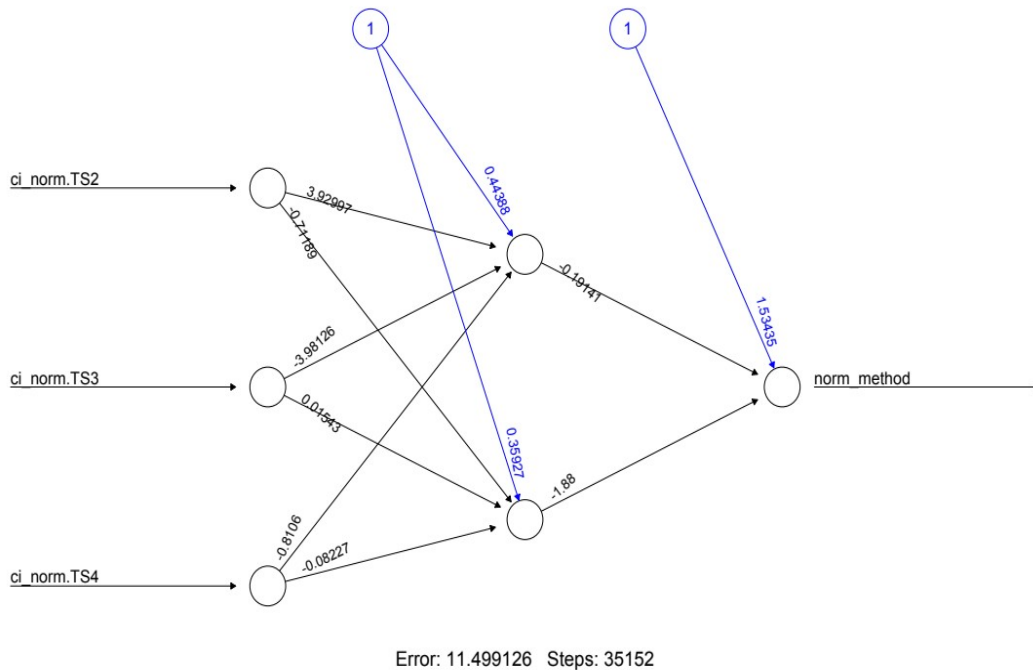
Dalle traiettorie si può ben apprezzare come l'andamento sia uguale in entrambi gli indici a riprova che il metodo per aumentare la numerosità campionaria non ha inficiate le caratteristiche di fondo degli indicatori.

### GRAFICO INDICE SERIE TRONCATE



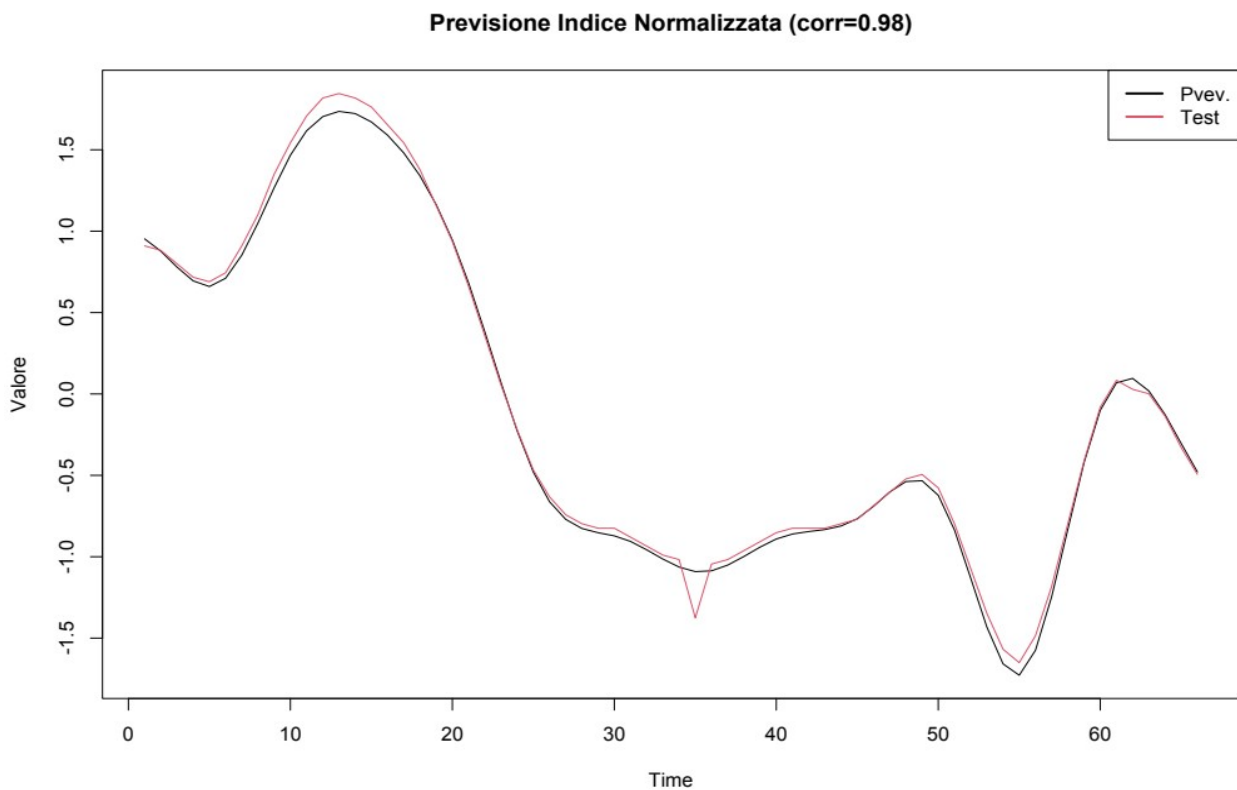
Il grafico riportato rappresenta l'indice sintetico calcolato utilizzando le serie storiche troncate alla nona ultima rilevazione. Esso fornisce una fotografia dell'indice così come si sarebbe potuto osservare sulla base dei dati disponibili in quell'istante temporale, offrendo quindi una rappresentazione "cristallizzata" del sistema in quel punto dell'orizzonte temporale. Verrà ora implementata una rete neurale artificiale al fine di effettuare una previsione sull'andamento futuro dell'indice sintetico. L'obiettivo è comprendere in che modo l'interazione tra le serie, modellata attraverso il sistema co-integrato, influenzi l'evoluzione dell'indice: se tale interazione condurrà ad un incremento, ad una diminuzione o al mantenimento di uno stato di stabilità. Anche in questa fase si adotterà una rete neurale della stessa tipologia impiegata in precedenza per l'indice calcolato sulle serie originali. La principale differenza consiste nell'utilizzo di due neuroni nascosti, al fine di cogliere eventuali relazioni più complesse generate dall'interazione tra le variabili.

## PLOT RETE IMPLEMENTATA CON MINIMO ERRORE



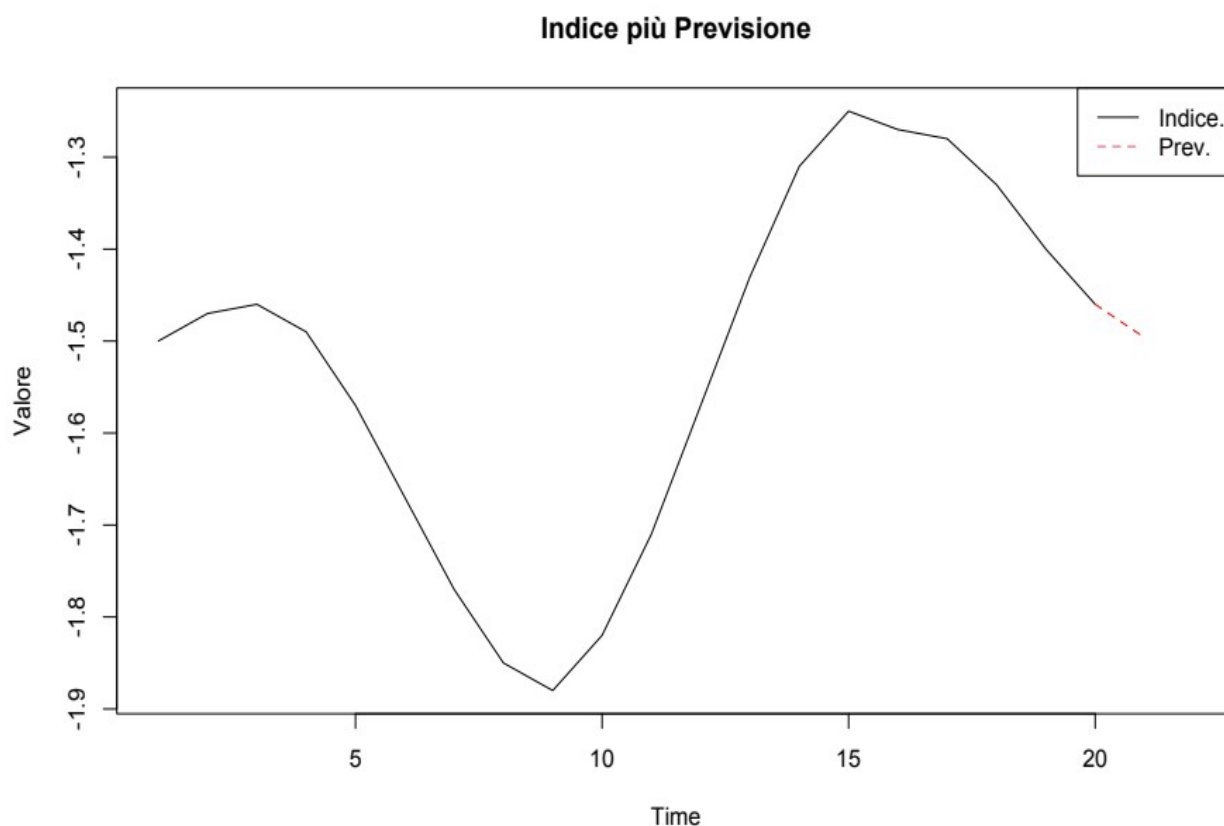
L'errore minimo ottenuto durante il processo di addestramento della rete neurale è pari a 11.499126 ed è stato raggiunto all'iterazione numero 35152. Di seguito si riporta il grafico relativo alla diagnostica eseguita sul data set di test, costituito dalle ultime 68 osservazioni, con l'obiettivo di valutare la capacità di generalizzazione del modello addestrato. Il data set di training, su cui è stata condotta la fase di apprendimento, include le prime 806 osservazioni della serie storica.

## PLOT BONTA' ADATTAMENTO



Dal grafico si può vedere la qualità dell'adattamento sulla serie dei dati di test che genera un coefficiente di correlazione che si approssima ad un valore di 0.98, calcolato col metodo di Pearson. Le rilevazioni (dati test ed approssimazione) rappresentate dal grafico sono state, anche questa volta, standardizzate per avere un impatto visivo più chiaro ed efficace della capacità approssimante della rete. Di seguito verrà rappresentato il grafico dell'indice delle ultime 20 osservazioni con il valore previsto dalla rete.

## GRAFICO INDICE E PREVISIONE



In figura è riportato il grafico dell'indice osservato, affiancato dalla previsione relativa allo step temporale successivo, ovvero la stima del valore che l'indice potrà assumere nella prossima rilevazione.

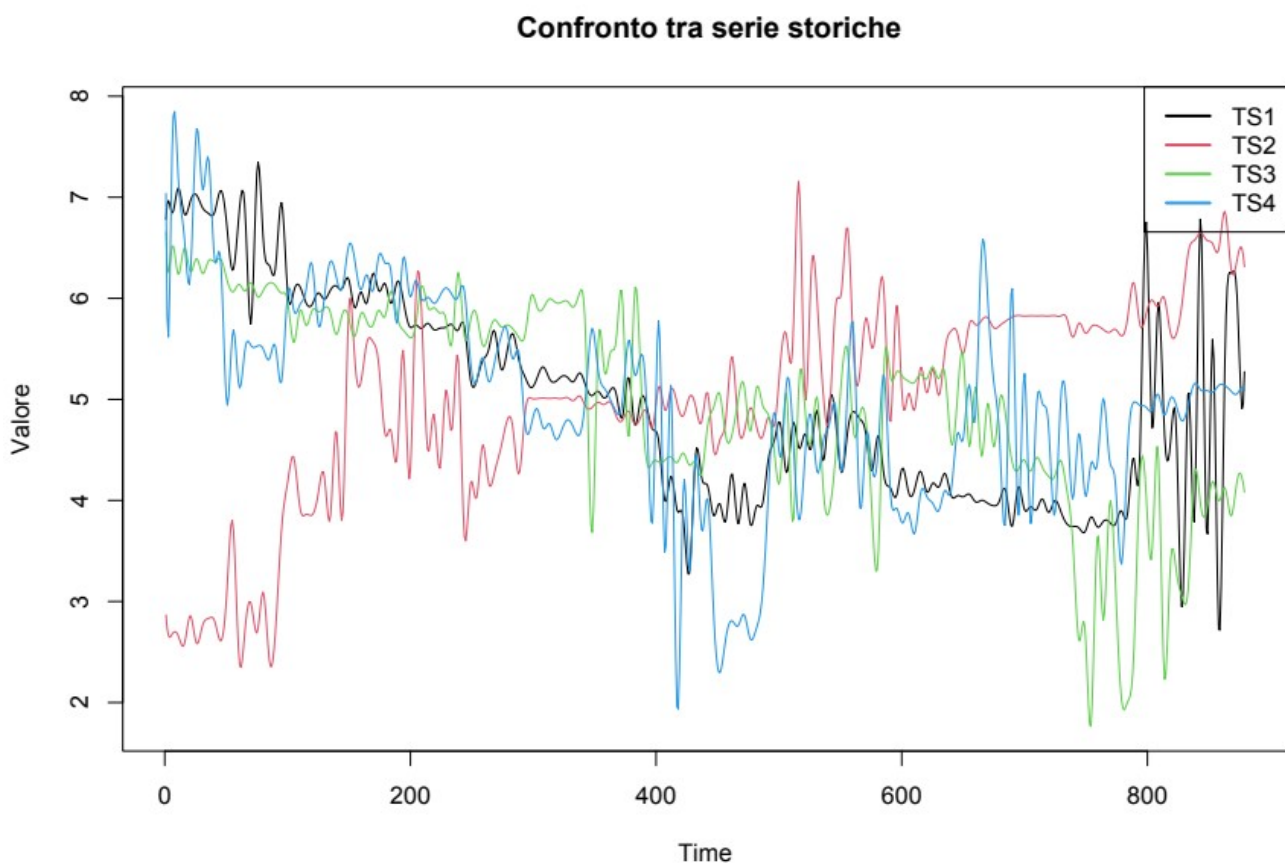
La rete neurale ha stimato per il prossimo periodo un valore in diminuzione, pari a -1.49. Questo risultato suggerisce che, sulla base delle dinamiche apprese durante l'addestramento e delle relazioni co-integrate tra le variabili, il sistema tenderà a muoversi in direzione dell'equilibrio di lungo periodo, generando una riduzione del valore dell'indice sintetico. Considerando che un aumento dell'indice rappresenta un miglioramento del livello di benessere in ambito sanitario, la previsione ottenuta implica una contrazione del concetto di *Salute* nel periodo successivo, segnalando un potenziale peggioramento della situazione complessiva misurata dagli indicatori.

L'intero apparato analitico descritto in precedenza dovrà essere implementato ad ogni nuova rilevazione, nel momento in cui saranno disponibili dati aggiornati. Questo approccio permette di mantenere una previsione dinamica e adattiva dell'evoluzione del sistema.

In questa sede, tuttavia, si procederà direttamente all'analisi del quinto periodo successivo all'ultima osservazione utilizzata nelle serie troncate. In corrispondenza di tale istante, si osserva un movimento al rialzo della combinazione lineare generata dal

sistema co-integrato. Questo comportamento viene interpretato come una lieve deviazione dal percorso di correzione verso l'equilibrio di lungo periodo, rappresentando una fase di temporanea instabilità. In tale contesto, il meccanismo di aggiustamento subisce un piccolo shock, che ne sospende momentaneamente la tendenza alla convergenza. Adesso si procederà alla verifica della qualità delle previsioni generate dal sistema co-integrato implementato con le nuove serie. Si omettono i test *ADF* poiché la natura integrata non viene inficiata dalla manipolazione del troncamento effettuato essendo di natura limitata rispetto alla totalità dei dati iniziali. Si procederà direttamente con l'implementazione della procedura di *Johansen* con le serie a cui sono state aggiunte le 5 rilevazioni successive.

## GRAFICO SERIE STORICHE TRONCATE SUCCESSIVE



L'andamento delle traiettorie nel tempo evidenzia chiaramente l'effetto dell'aggiunta delle cinque osservazioni successive alle serie storiche precedentemente troncate. Tale estensione temporale permette di visualizzare l'evoluzione del sistema e di analizzare come le nuove informazioni ne influenzino la dinamica.

Di seguito la procedura di *Johansen* per la verifica della presenza di co-integrazione.

## PROCEDURA PER VERIFICA RANGO DI CO-INTEGRAZIONE

```
#####  
# Johansen-Procedure #  
#####
```

Test type: maximal eigenvalue statistic (lambda max) , with linear trend

Eigenvalues (lambda):

```
[1] 0.036679830 0.017379017 0.006805958 0.005867735
```

Values of teststatistic and critical values of test:

	test	10pct	5pct	1pct
r <= 3	5.11	6.50	8.18	11.65
r <= 2	5.93	12.91	14.90	19.19
r <= 1	15.24	18.90	21.07	25.75
r = 0	32.47	24.78	27.14	32.14

Eigenvectors, normalised to first column:

(These are the cointegration relations)

	ts.TS1..l1	ts.TS2..l1	ts.TS3..l1	ts.TS4..l1
ts.TS1..l1	1.0000000	1.000000	1.0000000	1.00000000
ts.TS2..l1	0.7193749	-1.173594	-0.5088002	0.42794887
ts.TS3..l1	0.3833267	-1.299244	0.5884520	-0.61464187
ts.TS4..l1	-0.9550186	-1.126717	0.0481086	-0.05680797

Weights W:

(This is the loading matrix)

	ts.TS1..l1	ts.TS2..l1	ts.TS3..l1	ts.TS4..l1
ts.TS1..d	-0.0009325920	-4.109146e-05	-2.196045e-04	-0.0003403624
ts.TS2..d	-0.0008718507	4.995452e-04	9.237010e-05	-0.0003048705
ts.TS3..d	-0.0006612721	3.564893e-04	-8.469356e-05	0.0006223936
ts.TS4..d	0.0020516735	4.982930e-04	-1.065740e-04	-0.0002886033

```
> as.integer(mean(IND$selection))
```

```
[1] 10
```

```
> adf.test(Comb. Lin.);
```

Augmented Dickey-Fuller Test

data: Comb. Lin.

Dickey-Fuller = -4.1403, Lag order = 15, p-value = 0.01

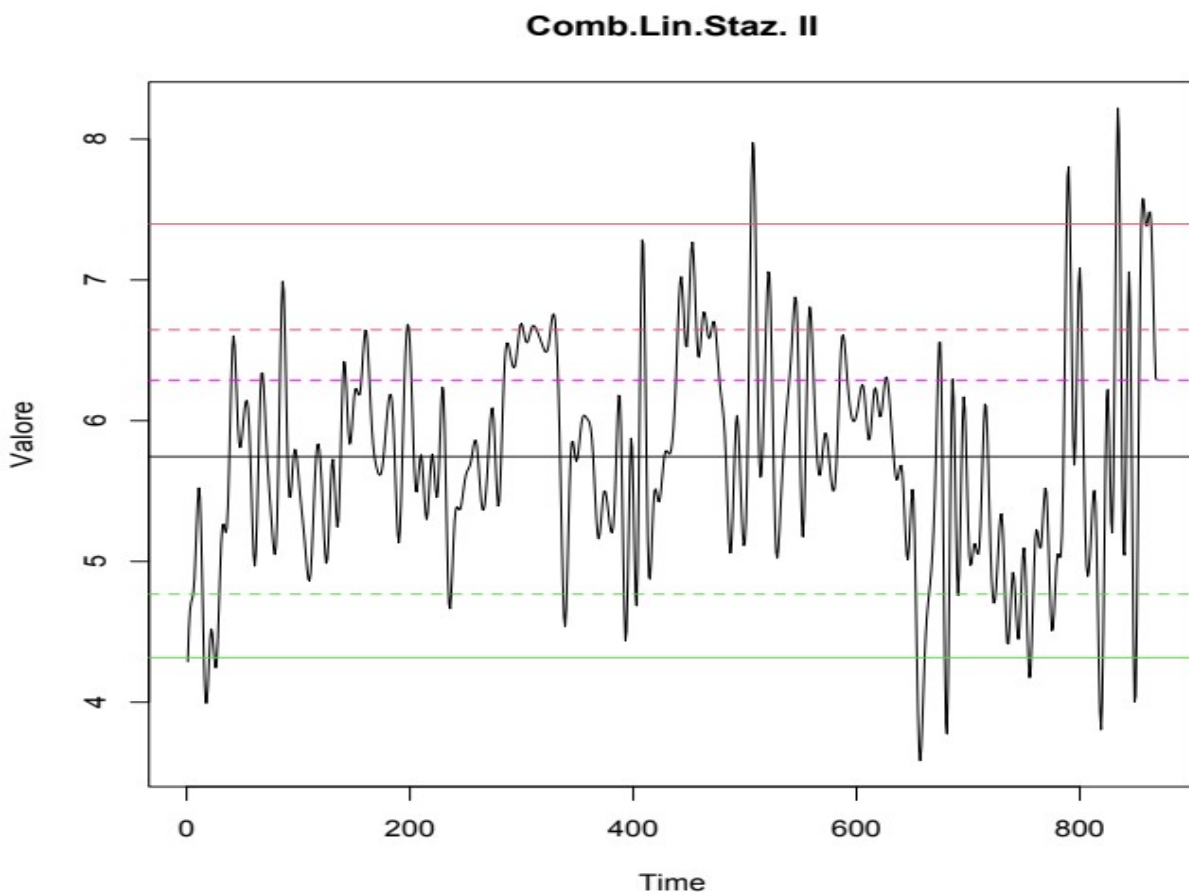
alternative hypothesis: stationary

Messaggio di avvertimento:

In adf.test(Comb. Lin.) : p-value smaller than printed p-value

L'applicazione della procedura di *Johansen* evidenzia un rango di co-integrazione pari ad 1, come indicato dalle statistiche test. L'analisi degli autovalori suggerisce che sia il primo vettore di co-integrazione a rappresentare più adeguatamente l'equilibrio di lungo periodo. La stabilità dei coefficienti della relazione di co-integrazione, rispetto alla implementazione precedente, suggerisce che il sistema mantenga una struttura di equilibrio coerente nel tempo, nonostante l'introduzione di nuovi dati. Ciò è indicativo di un meccanismo di aggiustamento che continua ad operare nel lungo periodo e che potrebbe confermare, per costruzione, l'ipotesi di stabilità temporale della relazione di co-integrazione nei domini BES e BesT ipotizzata all'inizio. In un contesto applicativo, questa stabilità può essere interpretata come evidenza della validità del modello nel rappresentare le dinamiche di lungo periodo delle componenti del sistema. Si inserirà di seguito il grafico della combinazione lineare derivante dal sistema co-integrato implementato:

### GRAFICO COMBINAZIONE LINEARE STAZIONARIA



Il grafico evidenzia la convergenza, verso il valore di equilibrio, osservata nei cinque periodi successivi all'ultima analisi effettuata.

Prima di procedere alla verifica del livello raggiunto dall'indice, previsto in discesa cinque periodi precedenti, si analizzeranno i valori attuali delle variabili, confrontandoli con quelli osservati cinque rilevazioni prima, quando la relazione di equilibrio di lungo periodo ne aveva indicate le direzioni evolutive. Più precisamente, il sistema aveva prevista una dinamica discendente per le variabili TS1, TS2 e TS3, ed una dinamica ascendente per TS4. Di seguito si riportano i valori rilevati all'epoca (già evidenziati in precedenza) e quelli attuali, al fine di verificarne la coerenza con le indicazioni fornite dalla relazione di equilibrio.

TS1 precedente: 21.1      TS1 ultimo: 20.27

TS2 precedente: 32.54      TS2 ultimo: 32.02

TS3 precedente: 8.44      TS3 ultimo: 8.333763

TS4 precedente: 3.64      TS4 ultimo: 3.66

Dall'analisi dei valori più recenti, confrontati con quelli registrati al momento del segnale, emerge che tutte e quattro le variabili hanno seguita la direzione indicata dalla relazione di lungo periodo. Si procederà adesso al calcolo delle variazioni percentuali assunte dalle quattro variabili.

TS1:  $\frac{(-21.1+20.27)}{21.1} = -0.0394$  la variabile ha avuta una discesa dei corsi del 3.94%

TS2:  $\frac{(-32.54+32.02)}{32.54} = -0.016$  la variabile ha avuta una discesa dei corsi del 1.6%

TS3:  $\frac{(-8.44+8.33)}{8.44} = -0.013$  la variabile ha avuta una discesa dei corsi del 1.3%

TS4:  $\frac{(-3.64+3.66)}{3.64} = 0.0055$  la variabile ha avuta una salita dei corsi dello 0.55%

Nei cinque periodi considerati, la variabile TS1 è risultata quella che ha subita la maggiore flessione dei suoi corsi. TS2 ha mostrato la secondo flessione più rilevante, seguita da TS3. Si può ipotizzare che in termini di velocità di aggiustamento, le

previsioni fornite dal sistema si sono rivelate sostanzialmente coerenti con i dati osservati, ad eccezione della variabile TS4, che si è confermata la meno reattiva. Tale discrepanza, tuttavia, non compromette la validità complessiva delle previsioni generate dal modello co-integrato implementato, permettendo di confermare una certa solidità delle ipotesi alla base dell'intera sperimentazione. Si procederà adesso al calcolo approssimato della variazione attesa che ci si poteva aspettare sfruttando i coefficienti *loading*, quelli del vettore di co-integrazione ed il segno dell'errore presenti al momento del segnale, per effettuare un confronto tra la variazione percentuale delle variabili effettivamente ottenuta e quella attesa sfruttando tali informazioni. Per far questo è stata adottata una procedura iterativa semplificata basata esclusivamente sulla componente di lungo periodo del *VECM* secondo la seguente regola ricorsiva:  $y_{t+h} = y_{t+h-1} + \alpha_i \cdot (\beta'_i \cdot y_{t+h-1})$  dove  $h=1,2,3,4,5$  sono i periodi successivi in cui viene effettuata la previsione,  $\alpha_i, i=1,2,3,4$  rappresenta il coefficiente di loading della *i-esima* variabile e  $\beta_i, i=1,2,3,4$  il coefficiente del vettore di co-integrazione della *i-esima* variabile con  $\beta_1 \equiv 1$ . A partire dal termine di errore di co-integrazione stimato al tempo  $t$  (istante in cui il sistema genera il segnale), ogni variabile è stata aggiornata iterativamente assumendo un aggiustamento proporzionale al proprio coefficiente di loading  $\alpha_i, i=1,2,3,4$ . Il processo è stato ripetuto per i 5 periodi successivi, aggiornando ad ogni passo i livelli previsti e ri-computando l'errore. Si riportano i valori intermedi delle previsioni ottenute con la procedura iterativa basata sulla componente di lungo periodo del modello *VECM*. I primi quattro valori previsti sono utilizzati come stime intermedie per valutare l'andamento delle serie, mentre il quinto valore rappresenta la previsione finale da confrontare con l'osservazione reale a cinque periodi di distanza. Questo confronto consente di valutare la bontà della metodologia adottata, verificando la coerenza della previsione con i dati effettivamente osservati.

	TS1	TS2	T3	T4	
t+1	5.7131	6.4253	4.2471	5.0860	Valori Intermedi
t+2	5.7033	6.4207	4.2442	5.1019	
t+3	5.6935	6.4160	4.2413	5.1178	
t+4	5.6838	6.4114	4.2384	5.1336	
t+5	5.6741	6.4068	4.2355	5.1493	Valore di Confronto

Questi valori rappresentano la dinamica di aggiustamento secondo questa procedura semplificata, che hanno subito le 4 variabili. I valori sono standardizzati. Si procederà al confronto con i valori realmente avuti de-standardizzando i risultati per capire in

che percentuale tale procedura di stima si è discostata dai valori realmente manifestatisi.

#### VALORI REALI VARIABILI

TS1 : 20.27

TS2 : 32.02

TS3 : 8.333763

TS4 : 3.66

#### VALORI STIMATI VARIABILI

TS1: 21.013

TS2: 32.39

TS3: 8.42

TS4: 3.72

Si procederà adesso al calcolo dell'errore percentuale commesso dalla implementazione di questa semplice modalità di previsione del valore puntuale delle serie storiche analizzate. Si sono avuti i seguenti errori:

$$TS1: \frac{|(20.27 - 21.013)|}{20.27} = 0.0366$$

$$TS2: \frac{|(32.02 - 32.39)|}{32.02} = 0.0115$$

$$TS3: \frac{|(8.333763 - 8.42)|}{8.333763} = 0.0103$$

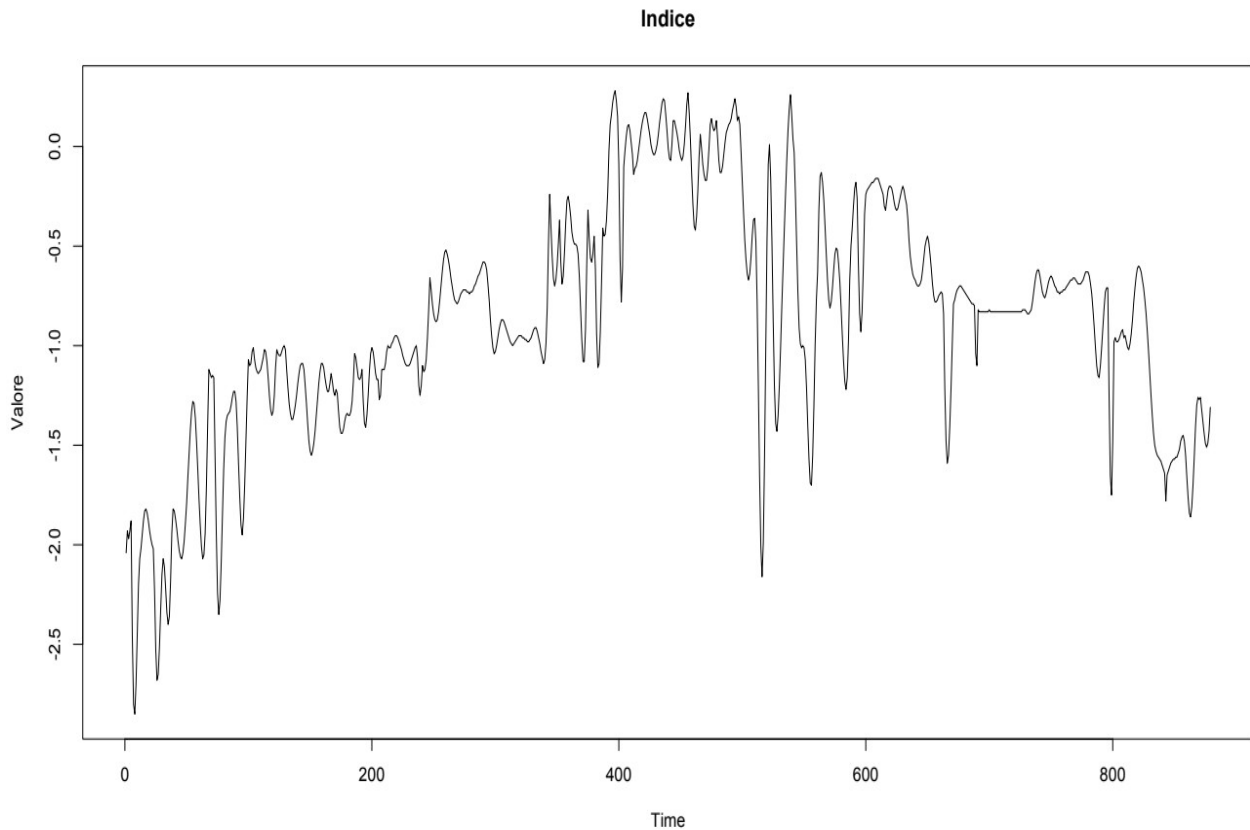
$$TS4: \frac{|(3.66 - 3.72)|}{3.66} = 0.0164$$

La procedura iterativa basata esclusivamente sulla componente di lungo periodo del *VECM* ha restituito previsioni che si sono rivelate coerenti nella direzione e contenute nell'errore finale calcolato rispetto al valore osservato nell'ultima rilevazione. Nonostante l'assenza delle dinamiche a breve termine, l'errore medio percentuale si è mantenuto entro limiti accettabili (tra 1% e 3.6%), mostrando l'efficacia euristica del solo termine di correzione  $\alpha \cdot EC_{t-1}$  nel tracciare il rientro verso l'equilibrio. Nelle prime 3 previsioni il valore predetto è risultato superiore al valore effettivo

rilevato dopo 5 periodi sottovalutando la reale entità del movimento discendente. Nell'ultima previsione invece vi è stata una sopravvalutazione del movimento rialzista. Tale comportamento è coerente con la natura della procedura, che non incorpora le dinamiche a breve termine, né eventuali shock esogeni, ma si limita a modellare il processo di aggiustamento verso la relazione di equilibrio di lungo periodo. Si può ragionevolmente osservare che, qualora la previsione fosse stata aggiornata passo dopo passo, inserendo progressivamente i nuovi dati osservati e ricalcolando di volta in volta i vettori di co-integrazione e i coefficienti di *loading*, la procedura avrebbe potuto restituire stime ancora più precise e puntuali, pur mantenendosi nella sua struttura semplificata. Questo approccio avrebbe infatti consentito al modello di adattarsi dinamicamente alle lievi variazioni nella struttura di lungo periodo che possono emergere con l'ampliamento del campione, migliorando l'allineamento tra le previsioni ed i valori osservati.

Per concludere l'analisi, si procederà alla costruzione di un indice sintetico del dominio oggetto di studio, basato sull'aggregazione delle serie storiche considerate. In particolare, l'indice verrà calcolato utilizzando le serie storiche aggiornate, ovvero comprensive delle cinque osservazioni aggiuntive rispetto al troncamento iniziale effettuato per la stima del sistema co-integrato precedente. L'obiettivo di questa fase finale dell'analisi è verificare l'andamento assunto dall'indice sintetico nel periodo di riferimento, ovvero quello successivo all'aggiunta delle ultime cinque osservazioni, e valutare in che modo ciascuna delle variabili analizzate — collegate da una relazione di co-integrazione — abbia contribuito a tale evoluzione. In particolare, si intende comprendere come le dinamiche individuali delle serie storiche abbiano inciso sulla traiettoria dell'indice, in coerenza con le forze di lungo periodo rilevate dal sistema *VECM*. Ciò permette di evidenziare non solo la co-movimentazione strutturale tra gli indicatori, ma anche il ruolo specifico di ciascuno nella determinazione della tendenza aggregata osservata.

## GRAFICO INDICE



Si osserva come, in seguito alla previsione del valore dell'indice sintetico tramite rete neurale effettuata allo step successivo, l'indice abbia effettivamente mostrato una flessione. Tuttavia, per esigenze di sintesi, non si è proceduto all'aggiornamento del modello ad ogni nuova osservazione. In un contesto applicativo reale, la previsione dovrebbe essere invece aggiornata passo dopo passo, integrando progressivamente i nuovi dati disponibili e combinandola con le previsioni fornite dal sistema *VECM*. Nell'arco delle ultime cinque rilevazioni, l'indice sintetico relativo al dominio *Salute* dell'area di Taranto ha mostrata un'evoluzione complessivamente positiva: è passato da un valore di  $-1.46$  (registrato cinque periodi prima) ad  $-1.31$  nell'ultima rilevazione, dopo aver toccato un minimo intermedio di  $-1.51$ . Questo risultato conferma che, nella simulazione ipotetica costruita sui dati originali, il sistema ha sperimentato un progressivo miglioramento, trainato dalla dinamica congiunta degli indicatori considerati. Tale rialzo risulta coerente con il comportamento di convergenza descritto in precedenza nell'analisi del sistema co-integrato. In particolare, come indicato dal sistema co-integrato implementato precedentemente, nelle ultime cinque rilevazioni si è osservata una riduzione nei valori di tre indicatori: la mortalità evitabile (0-74 anni), la mortalità per demenza e malattie del sistema nervoso (65 anni e oltre) e la mortalità per tumore (20-64 anni). Al contrario,

l'indicatore relativo alla mortalità infantile (decessi per 1.000 nati vivi residenti) ha registrato un limitato incremento. Nonostante l'andamento di quest'ultimo indicatore, la dinamica complessiva della variabile latente *Salute* ha subita una evoluzione positiva trainata dalla contrazione dei primi tre indicatori, mentre l'aumento della mortalità infantile è stata di entità limitata per la quale non ha avuto un impatto negativo tale da invertire la tendenza favorevole osservata. La conclusione dell'apparato interpretativo-predittivo di questo lavoro si fonda sull'ipotesi che, all'interno degli indicatori dei domini BES e BesT, si manifesti un'evoluzione co-integrata. Identificato questo framework concettuale, e sfruttando in maniera strumentale le reti neurali, è possibile prevedere l'andamento futuro della variabile latente associata ad un determinato dominio.

Tale previsione si basa sia sull'evoluzione stimata degli indicatori, ottenuta tramite l'analisi di co-integrazione, sia sulla capacità del modello di apprendere le dinamiche interne al sistema. In questo modo, è possibile non solo anticipare il comportamento della variabile latente, ma anche comprendere come i singoli indicatori contribuiscano alla sua evoluzione nel tempo. Nel caso di studio, si può ritenere plausibile che, con le future rilevazioni, la relazione di equilibrio tenderà a raggiungere il proprio livello di equilibrio, con le traiettorie delle variabili che proseguiranno nel processo di aggiustamento individuato dal sistema co-integrato. La stabilità nel tempo dei valori chiave del test di co-integrazione consente di ritenere verosimile che le variabili si muoveranno nella direzione indicata dal segnale principale. Considerando la polarità degli indicatori, si può inoltre ipotizzare che, quando la relazione di lungo periodo raggiungerà il suo livello di equilibrio, l'indice avrà raggiunto un livello superiore a quello precedentemente osservato, poiché la convergenza al ribasso dei primi tre indicatori rappresenta un forte segnale di miglioramento del concetto di Salute nella zona studiata.

## 4.3

### Applicazione a dati Nazionali

Si procederà ora con l'implementazione della procedura sul dominio *Benessere Economico* nazionale.

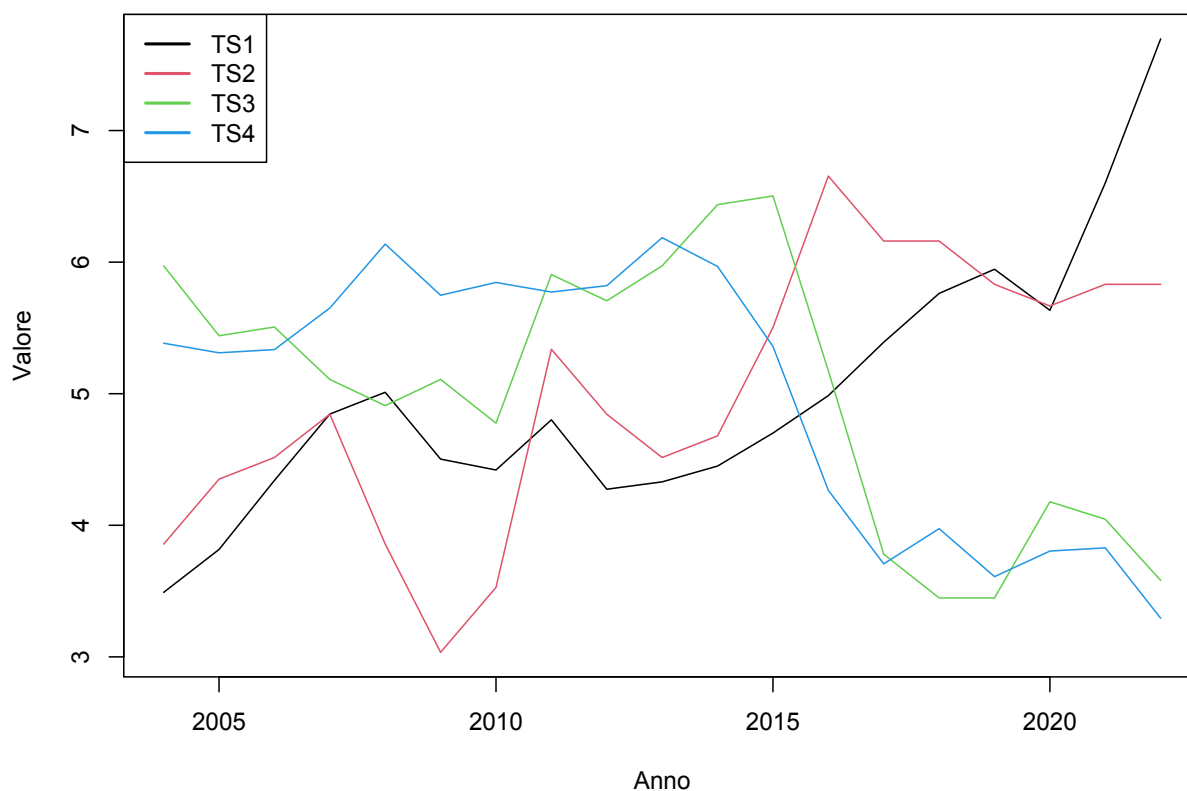
Gli indicatori selezionati per l'analisi sono quattro contrassegnati anch'essi rispettivamente con le sigle TS1, TS2, TS3 e TS4 e coprono un arco temporale compreso tra il 2004 e il 2022, con un time-frame annuale.

- 1) Reddito medio procapite valori assoluti in euro
- 2) Rischio di povertà valori percentuali
- 3) Grave deprivazione abitativa valori percentuali
- 4) Grave difficoltà ad arrivare a fine mese valori percentuali

La scelta di questi indicatori è stata guidata dalla disponibilità di rilevazioni continuative sull'intero periodo di osservazione, compreso tra il 2004 ed il 2022, analogamente a quanto avvenuto per gli indicatori analizzati del dominio *Salute*. Sebbene le etichette utilizzate per identificare gli indicatori in questa analisi coincidano con quelle impiegate nello studio relativo al precedente dominio, ciò non compromette in alcun modo la correttezza e la coerenza metodologica del presente lavoro trattandosi di una scelta volta unicamente ad evitare un appesantimento della notazione attraverso ulteriori etichettature. Si procederà innanzitutto alla visualizzazione grafica delle serie temporali degli indicatori originali (standardizzati), ovvero quelle non sottoposte ad allungamento.

## GRAFICO DELLE RILEVAZIONI

Confronto tra serie storiche dal 2004 al 2022



Dal punto di vista grafico, si osserva una possibile relazione di co-integrazione tra le serie, evidenziata dal tipico andamento intrecciato delle traiettorie. Inoltre, l'apparente natura integrata delle serie sarà opportunamente verificata mediante l'applicazione del test *ADF*. La struttura del test adottata è quella implementata in ambiente R, già descritta nell'analisi precedente. Anche la determinazione del *lag* ottimale segue il criterio proposto da *Schwert* che suggerisce il valore massimo, in continuità con quanto applicato in precedenza.

## TS1

### Augmented Dickey-Fuller Test

data: TS1

Dickey-Fuller = -0.27206, Lag order = 4, p-value = 0.9848

alternative hypothesis: stationary

## TS2

### Augmented Dickey-Fuller Test

data: TS2

Dickey-Fuller = -1.7227, Lag order = 4, p-value = 0.678

alternative hypothesis: stationary

## TS3

### Augmented Dickey-Fuller Test

data: TS3

Dickey-Fuller = -1.0941, Lag order = 4, p-value = 0.9067

alternative hypothesis: stationary

## TS4

### Augmented Dickey-Fuller Test

data: TS4

Dickey-Fuller = -1.0775, Lag order = 4, p-value = 0.9092

alternative hypothesis: stationary

Nel test è stato utilizzato un *lag* pari a 4, selezionato in riferimento al valore massimo consentito dalla regola di *Schwert*, che in questo caso risulta essere pari a 7.

La natura integrata – e dunque non stazionaria – delle serie temporali analizzate consente di procedere alla verifica di un'eventuale relazione di co-integrazione. La verifica della presenza di co-integrazione viene condotta — anche in questo caso — tramite la procedura di *Johansen*, basata sul test del massimo autovalore. Di seguito la diagnostica della procedura:

## PROCEDURA PER VERIFICA RANGO DI CO-INTEGRAZIONE

```
#####  
# Johansen-Procedure #  
#####
```

Test type: maximal eigenvalue statistic (lambda max) , with linear trend

Eigenvalues (lambda):

```
[1] 0.90277345 0.51844224 0.37560614 0.06706903
```

Values of teststatistic and critical values of test:

	test	10pct	5pct	1pct
r <= 3	1.18	6.50	8.18	11.65
r <= 2	8.01	12.91	14.90	19.19
r <= 1	12.42	18.90	21.07	25.75
r = 0	39.62	24.78	27.14	32.14

Eigenvectors, normalised to first column:

(These are the cointegration relations)

	ts.TS1..l1	ts.TS2..l1	ts.TS3..l1	ts.TS4..l1
ts.TS1..l1	1.000000	1.0000000	1.00000000	1.00000000
ts.TS2..l1	-1.070911	0.7085150	-0.7275134	-0.52906718
ts.TS3..l1	2.154343	0.2325805	1.2073848	0.02252531
ts.TS4..l1	-2.485692	0.8966911	-0.2571313	0.09691463

Weights W:

(This is the loading matrix)

	ts.TS1..l1	ts.TS2..l1	ts.TS3..l1	ts.TS4..l1
ts.TS1..d	0.6021153	-0.1387225	-0.01267334	0.0136168
ts.TS2..d	0.2692135	-0.4989217	0.28045239	0.2467411
ts.TS3..d	-0.2782386	-0.2602833	-0.17793034	0.2028024
ts.TS4..d	-0.1562275	-0.2727279	-0.20150857	-0.1299914

```
> as.integer(mean(IND$selection))
```

```
[1] 2
```

```
> adf.test(Comb. Lin.);
```

Augmented Dickey-Fuller Test

data: Comb. Lin.

Dickey-Fuller = -3.0383, Lag order = 2, p-value = 0.1768

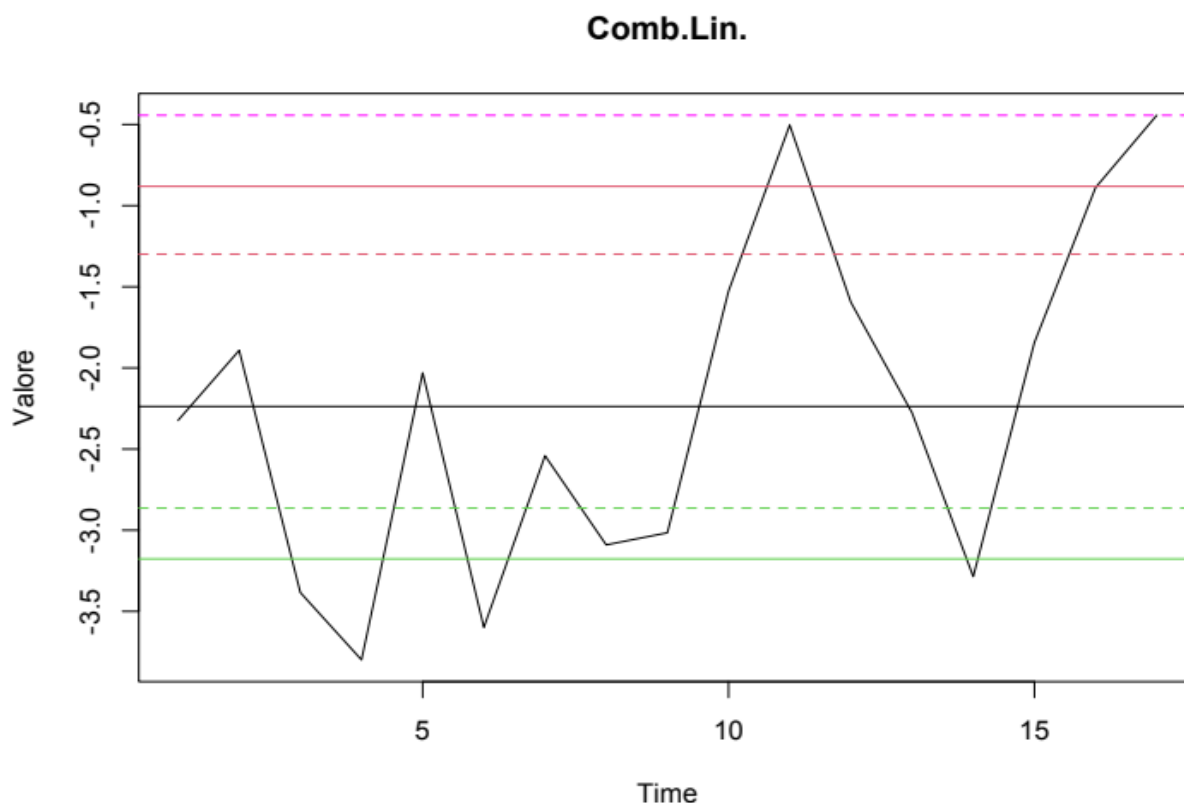
alternative hypothesis: stationary

Dallo studio degli autovalori si evince, anche questa volta, come la relazione più forte tra le variabili sia quella che si ottiene dal primo vettore di co-integrazione. L'applicazione del test di Johansen ha permesso di stimare un rango di co-integrazione pari ad 1. L'ipotesi nulla di assenza di co-integrazione viene infatti rifiutata a favore dell'esistenza di almeno una relazione di lungo periodo tra le serie considerate. Tuttavia, in questo caso, la combinazione lineare derivante dal sistema

non risulta stazionaria. Come già accennato in precedenza, ciò può essere attribuito alla ridotta numerosità campionaria, che potrebbe ostacolare l'evidenza del movimento di ritorno verso la media (*mean reversion*) tipico delle serie stazionarie. Quando l'analisi verrà estesa alle serie aumentate, emergerà con maggiore evidenza come la combinazione lineare risulti essere un processo  $I(0)$ , ovvero stazionaria. Il primo vettore co-integrante, normalizzato rispetto alla prima variabile, evidenzia una relazione stabile di lungo periodo tra gli indicatori, suggerendo la presenza di una dinamica di equilibrio comune. L'analisi dei coefficienti del vettore mostra che la variabile TS3 presenta segno positivo, mentre TS2 e TS4 hanno segno negativo, indicando un andamento opposto nella composizione della relazione di equilibrio rispetto a TS3 e TS1. In termini assoluti, TS4 risulta la variabile con il peso maggiore nella determinazione dell'equilibrio di lungo periodo, seguita da TS3. Anche TS1 e TS2 apportano un contributo significativo, sebbene relativamente inferiore. Per quanto riguarda i coefficienti di aggiustamento (*loading*), la variabile TS1 risulta quella con il coefficiente più grande — in valore assoluto — pari ad 0.6021153 ad indicare che la variabile reagirà più fortemente delle altre alle deviazioni dalla condizione di equilibrio del sistema. Dunque se la combinazione lineare risulterà lontana dal suo valore di equilibrio, la variabile TS1 reagirà subendo l'aggiustamento maggiore nel breve periodo. Tuttavia, dato il segno positivo del *loading*, associato al segno dell'errore e del coefficiente di co-integrazione, TS1 potrebbe muoversi in maniera tale da ostacolare temporaneamente il ritorno all'equilibrio. La seconda variabile più reattiva risulta essere la TS3 con un coefficiente di *loading* pari a -0.2782386. Il segno negativo indica che la variabile contribuirà sempre al processo di aggiustamento. La variabile TS2 ha coefficiente *loading* pari a 0.2692135 ed essa contribuirà alla correzione dell'errore della relazione di equilibrio di lungo periodo. Infine, la variabile TS4 è quella con il *loading* più basso (-0.1562275 ) ad indicare che essa reagirà meno velocemente delle altre nella correzione dell'errore. Alla luce delle altre informazioni, il suo movimento risulterà essere di disturbo al movimento correttivo. Anche in questo caso, come per lo studio relativo al dominio precedente, si evidenzia che la possibile bassa significatività dei coefficienti di *loading* potrebbe essere attribuibile alle stesse motivazioni già discusse, ovvero la limitata dimensione campionaria e le differenze nei metodi di stima adottati dai vari pacchetti software. Nonostante questo, la procedura di stima ha comunque prodotto valori predittivi coerenti con le indicazioni fornite dal sistema, il che suggerisce che tali discrepanze potrebbero non incidere in modo rilevante sulla validità complessiva del modello. In definitiva, il quadro emerso consente di formulare considerazioni interpretative anche per il dominio in esame, suggerendo l'esistenza di una possibile configurazione

latente del sistema sociale analizzato, così come catturata dal modello implementato. Si riporterà adesso il grafico della combinazione lineare derivante dal sistema:

### GRAFICO COMBINAZIONE LINEARE

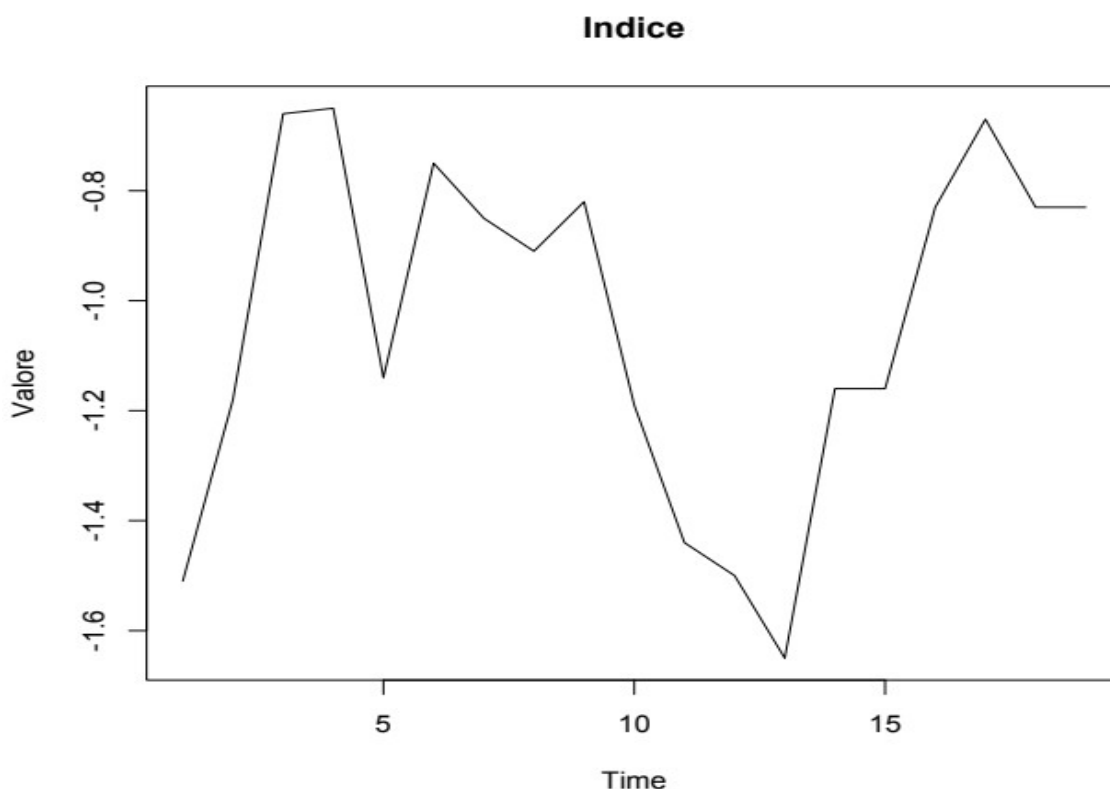


Come già evidenziato dal test di Dickey-Fuller aumentato (*ADF*), la combinazione lineare ottenuta dal sistema co-integrato non mostra comportamento stazionario, confermato anche da un'analisi grafica preliminare. Tuttavia, tale assenza di evidenza di stazionarietà non deve destare particolare sorpresa. Infatti, si è prima evidenziato come, nonostante la co-integrazione implichi teoricamente la stazionarietà della combinazione lineare, la limitata numerosità campionaria può impedire l'emersione chiara del movimento di ritorno verso la media (*mean-reversion*) tipico delle serie stazionarie. In tali condizioni, il comportamento osservato — come in questo caso — potrebbe risultare simile a quello di una serie non stazionaria, sia nei test formali sia nella rappresentazione grafica. Malgrado ciò si evidenzia ugualmente come, la conformazione attuale del sistema suggerirebbe la presenza di un aggiustamento verso il basso della relazione di equilibrio, consentendo una possibile previsione degli indicatori in sistema sfruttando le componenti analizzate.

Anche in questo caso, sono state inserite delle bande rosse e verdi, tratteggiate e continue, che rappresentano concettualmente l'insieme delle catastrofi. Nella configurazione attuale, la rottura verso il basso della banda rossa continua indica il punto in cui il sistema subisce una variazione radicale, passando da una situazione di disequilibrio di lungo periodo ad una condizione di equilibrio, che si realizza al raggiungimento della banda nera centrale. Questo livello rappresenta il *creodo* della teoria delle catastrofi, ovvero il punto in cui il sistema raggiunge una configurazione di equilibrio stabile. In questa prospettiva, tali livelli rappresentano soglie critiche oltre le quali il sistema studiato subisce un cambiamento qualitativo, coinvolgendo simultaneamente tutte le variabili che lo compongono. In corrispondenza di questi punti, le variabili tendono a spostarsi verso nuovi livelli, dando luogo ad una transizione strutturale dell'intero sistema. All'interno di questo contesto interpretativo, le bande tratteggiate sono considerate come piccole perturbazioni del sistema: esse influenzano le dinamiche, ma non determinano un cambiamento qualitativo significativo. Al contrario, le bande continue identificano i livelli critici nei quali si verifica una transizione strutturale nel sistema, ovvero un cambiamento qualitativo delle sue dinamiche che lo porterà al raggiungimento del punto di equilibrio. Si inserirà adesso il grafico dell'indice sintetico riguardante il dominio *Benessere Economico*. Il metodo usato è lo stesso applicato nello studio del dominio *Salute* riferito al territorio di Taranto.

Anche in questo caso, data la polarità degli indicatori, un incremento del valore dell'indice è associato ad un miglioramento della variabile latente sottostante, mentre una diminuzione del valore segnala un peggioramento della stessa. Dunque, non sarà rilevante tanto il valore che assumerà l'indice in termini quantitativi bensì il suo movimento in termini qualitativi, ovvero la direzione del suo movimento (rialzista o ribassista).

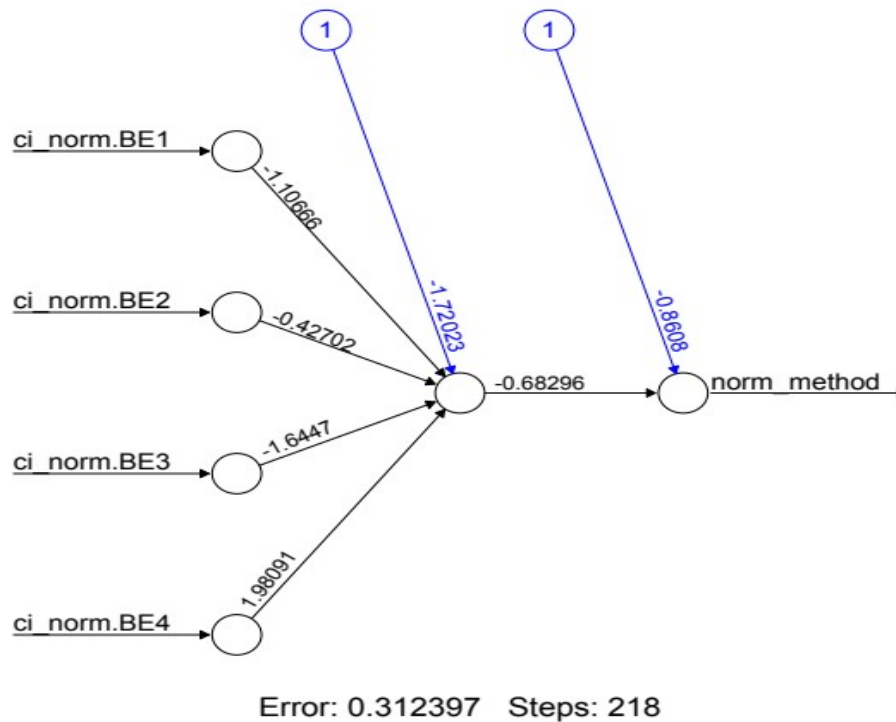
## GRAFICO INDICE BENESSERE ECONOMICO



L'analisi grafica dell'indice evidenzia una coerenza tra il modello e la dinamica economico-sociale del sistema. In particolare, si nota la coerenza tra l'andamento dell'indice degli ultimi 3 o 4 periodi e l'andamento della combinazione lineare negli stessi periodi di riferimento. Inoltre si può aggiungere che, dato il livello elevato su cui si colloca la combinazione lineare, si potrebbe prevedere una sua discesa al passo temporale successivo.

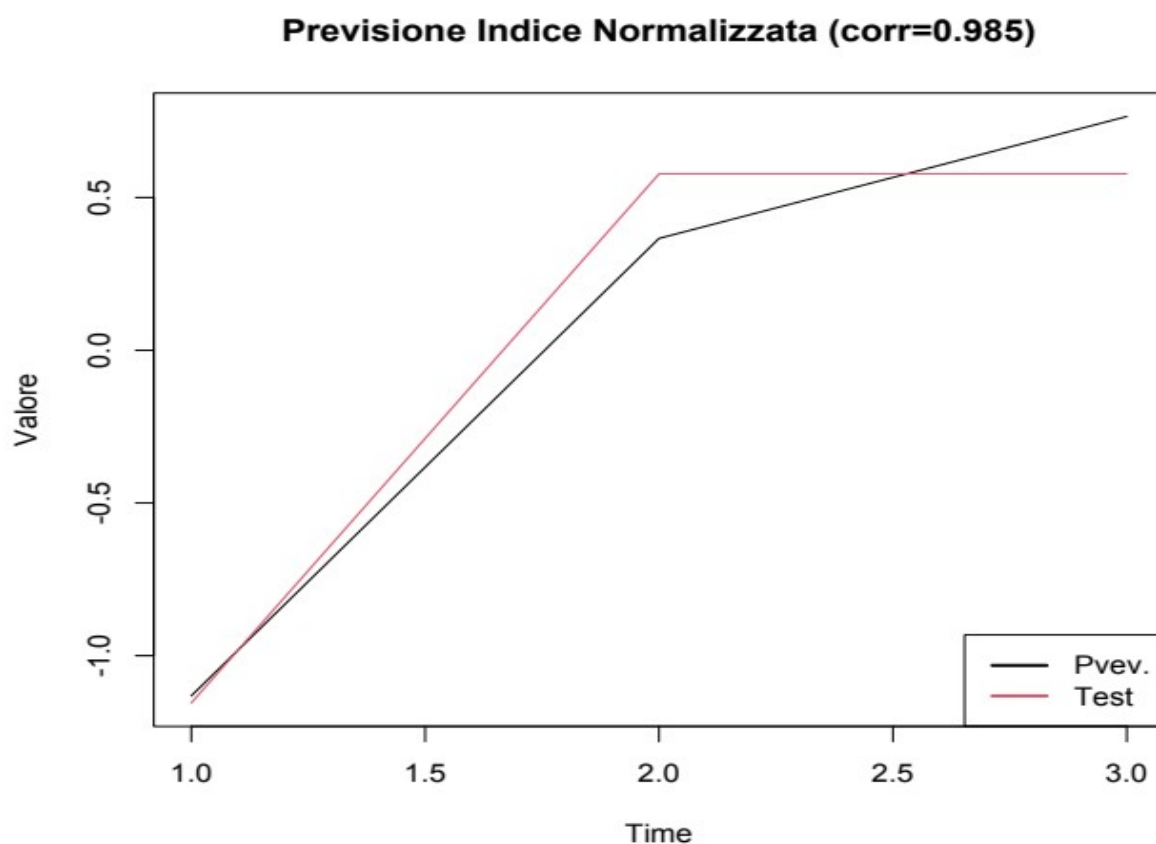
Considerando i segni dei coefficienti del vettore di co-integrazione associati alle variabili del sistema, si può dedurre che il livello di benessere economico del sistema analizzato tenderà a diminuire nel prossimo periodo, in linea con le previsioni fornite dalle variabili che compongono il sistema co-integrato. Di seguito viene rappresentato il grafico della rete neurale implementata per modellare l'andamento futuro dell'indice sintetico.

## PLOT RETE IMPLEMENTATA CON MINIMO ERRORE



L'errore minimo generato dalla rete è pari a 0.312397 ed è stato raggiunto alla iterazione numero 218. Di seguito si presenta il grafico della diagnostica effettuata sul data-set di test (costituito dalle ultime tre osservazioni) con l'obiettivo di valutare la capacità di adattamento della rete, addestrata sul data-set di training (composta dalle prime sedici osservazioni).

## PLOT BONTA' ADATTAMENTO



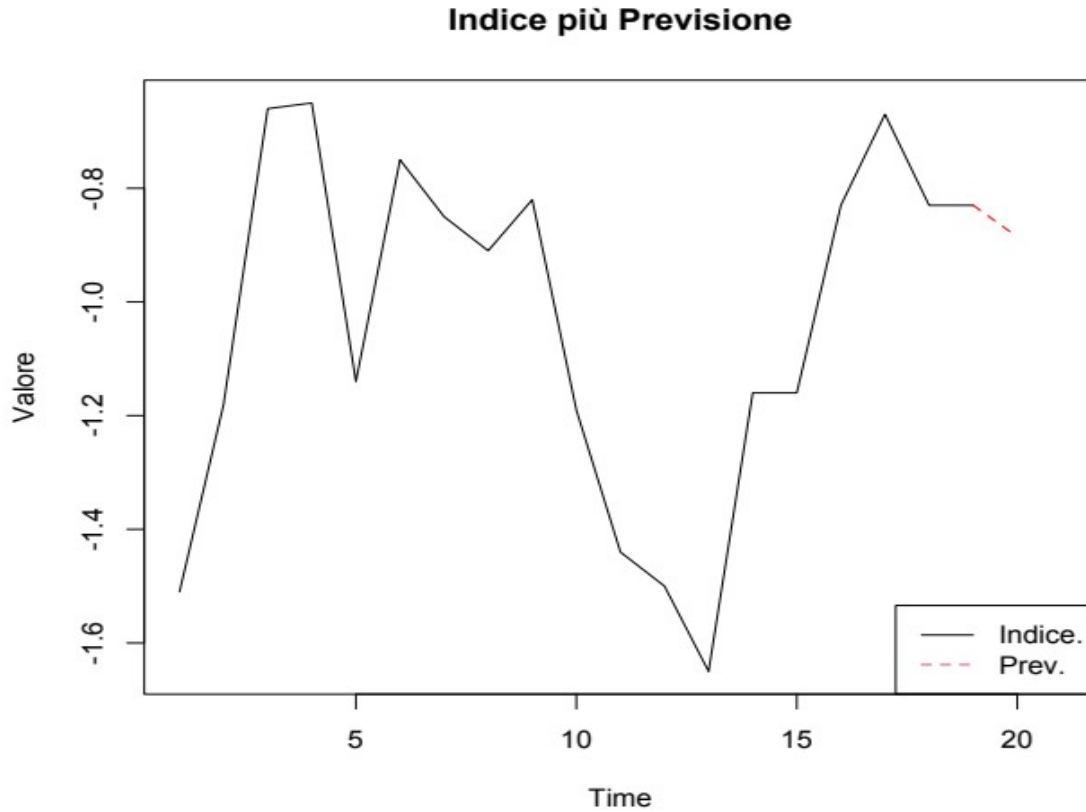
Dal grafico si osserva, anche in questo caso, un'elevata qualità dell'adattamento della rete sulla serie dei dati di test, che si traduce in un coefficiente di correlazione pari a circa 0,985, calcolato secondo il metodo di *Pearson*.

Le osservazioni utilizzate (dati di test e relative approssimazioni) sono state standardizzate al fine di migliorare la rappresentazione grafica della bontà dell'adattamento.

L'elevato grado di accuratezza delle reti implementate sin'ora — pur nella loro semplicità — è attribuibile alla bassa complessità delle traiettorie degli indici considerati. Al contrario, nel caso di dinamiche più articolate, reti neurali più complesse non sempre producono risultati di qualità paragonabile.

L'analisi prosegue con la stima del valore che l'indice potrà assumere nello step temporale successivo, come fatto per il dominio *Salute*.

## GRAFICO INDICE E PREVISIONE



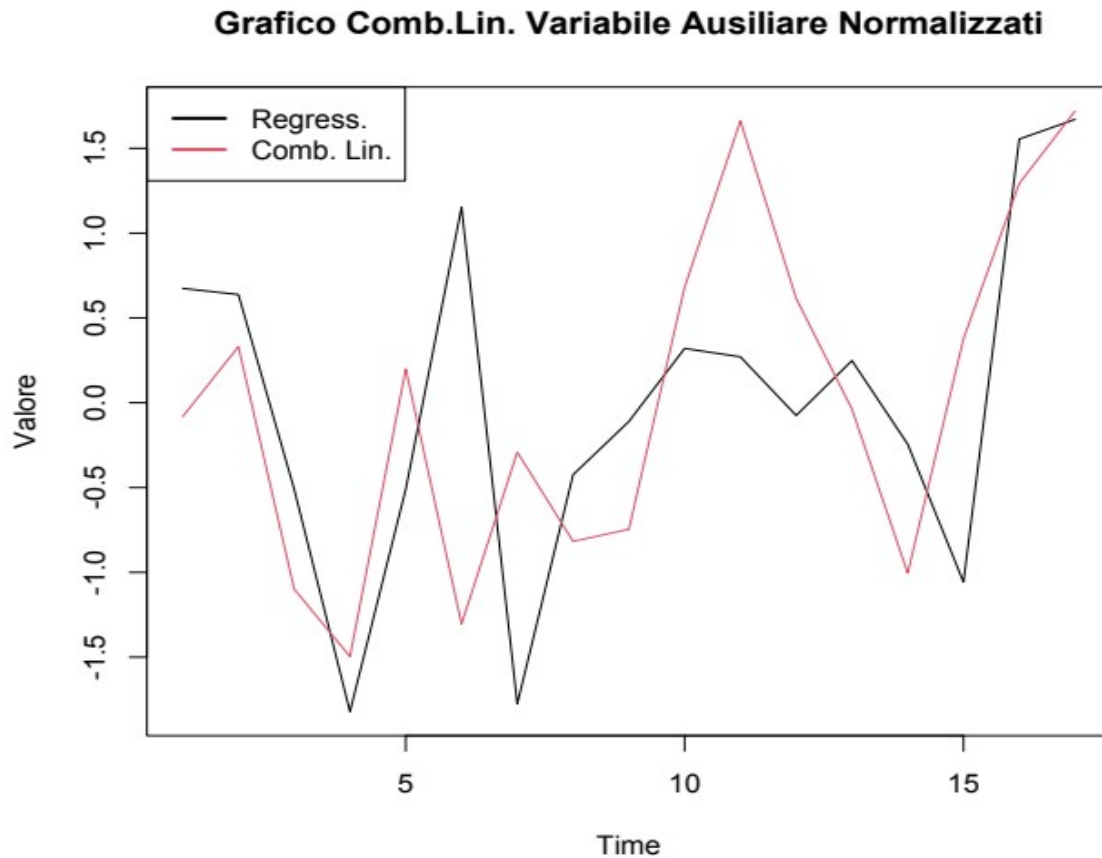
La rete, sulla base dell'andamento storico dell'indice, genera una previsione in discesa rispetto all'ultimo valore osservato. Sulla base di questa indicazione, anche in questo caso il sistema co-integrato suggerisce che tale variazione qualitativa al ribasso sarà determinata da modifiche negli indicatori, secondo le combinazioni specificate dal vettore di co-integrazione. La velocità di aggiustamento verso l'equilibrio sarà invece determinata dai coefficienti di *loading* associati ai singoli indicatori. Dato il livello della combinazione lineare, come già accennato prima, se ne può dedurre un futuro ribasso. Ugualmente verrà implementata una rete neurale per fare previsione sulla combinazione lineare utilizzando la tecnica proposta per lo studio del dominio precedente.

Di seguito si presenta il grafico della combinazione lineare stazionaria derivante dal sistema co-integrato, affiancato a quello della variabile ausiliaria costruita mediante i coefficienti derivanti da una *ACP* riportati di sotto:

## COEFFICIENTI VARIANZA SPIEGATA

PC1	PC2	PC3	PC4
0.7786	0.1416	0.05831	0.02156

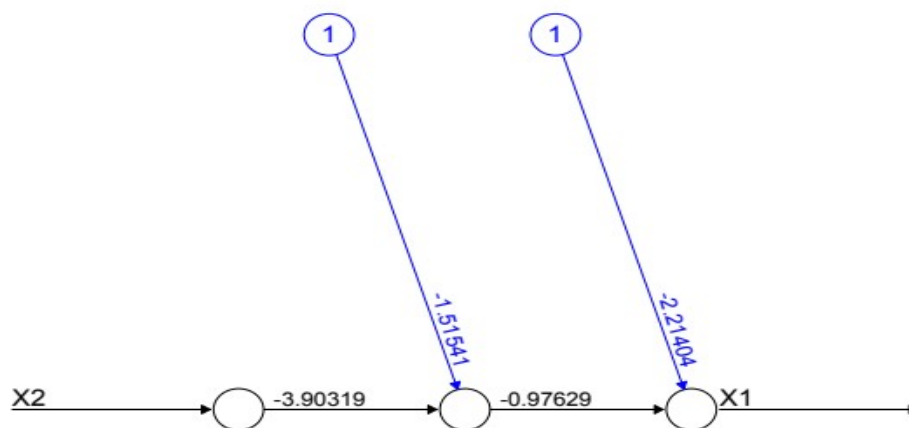
## GRAFICO COMB. LIN. CON REG. AUSILIARE



Si osservi come una semplice combinazione lineare delle variabili, ottenuta attraverso un'analisi delle componenti principali, abbia generato una traiettoria informativa e utile ai fini regressivi. Tale traiettoria è stata impiegata per l'implementazione di una rete neurale finalizzata alla previsione dell'andamento futuro della combinazione lineare derivante dal sistema co-integrato.

Di seguito si mostra la rete neurale implementata per la previsione dell'andamento futuro della combinazione lineare.

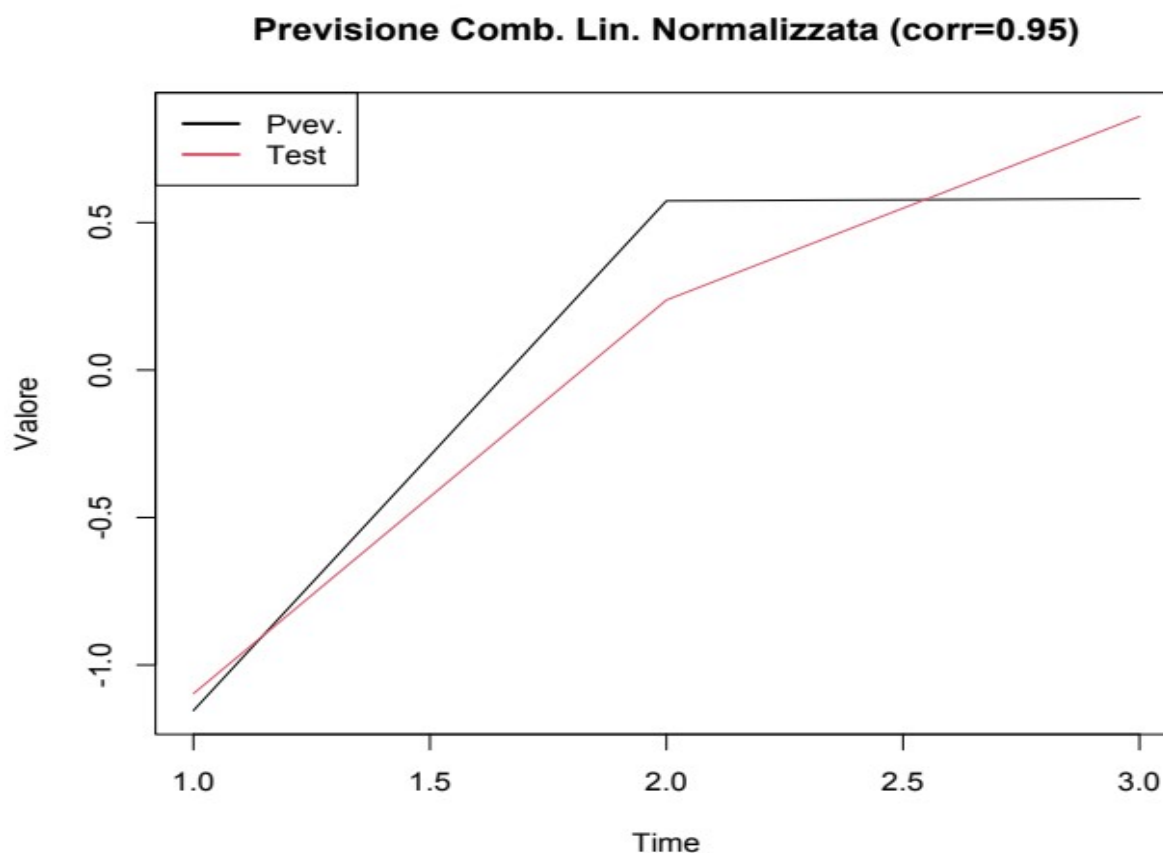
## PLOT RETE IMPLEMENTATA CON MINIMO ERRORE



Error: 5.257743 Steps: 95

L'errore minimo ottenuto dalla rete è pari a 5.257743 ed è stato raggiunto all'iterazione numero 95. Di seguito si presenta il grafico della diagnostica condotta sul data-set di test (costituito dalle ultime tre osservazioni), con l'obiettivo di valutare la capacità di adattamento della rete addestrata sul data-set di training, composto dalle restanti osservazioni.

## PLOT BONTA' ADATTAMENTO

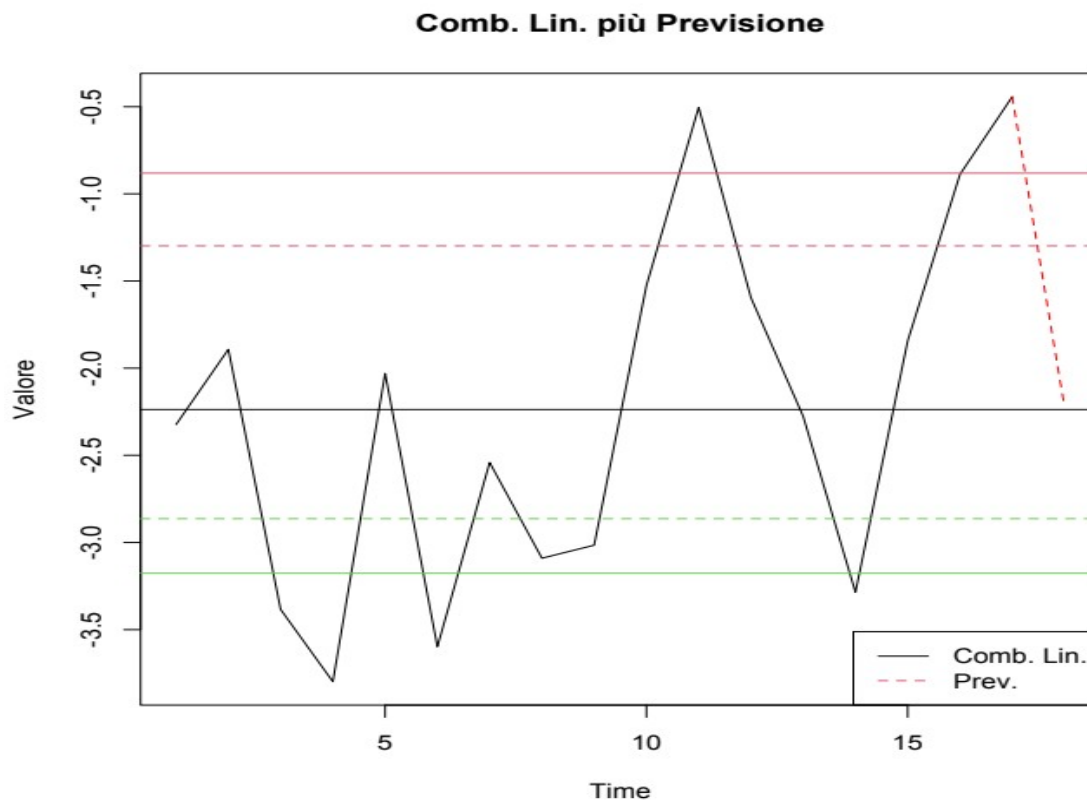


Dal grafico si osserva una buona qualità di adattamento sul data-set di test, con un coefficiente di correlazione pari a circa 0.95, calcolato secondo il metodo di *Pearson*. Le rilevazioni utilizzate sono state standardizzate come sempre al fine di migliorarne la leggibilità grafica dell'accuratezza del modello.

L'analisi prosegue con la stima del valore che la combinazione lineare potrà assumere allo step successivo. A tal fine è stata utilizzata sempre la funzione *compute* che, sulla base della rete implementata, restituisce una previsione al ribasso. La previsione della combinazione lineare è in linea con l'ultimo valore assunto da essa ed in linea con la previsione effettuata dell'indice.

Di seguito si presenta il grafico della combinazione lineare ottenuta dal modello di co-integrazione, affiancato dalla previsione generata per lo step successivo.

## GRAFICO COMB. LIN. E PREVISIONE



Con questi risultati si può procedere alla costruzione di scenari previsivi per il sistema. Considerando il livello della combinazione lineare, la direzione prevista della stessa ed i coefficienti del vettore di co-integrazione riferiti agli indicatori inseriti a sistema, si prevede che la variabile TS1 (Reddito medio pro capite) allo step successivo subirà una riduzione come anche la variabile TS3 (Grave deprivazione abitativa). Le variabili TS2 (Rischio di povertà) e TS4 (Grave difficoltà ad arrivare a fine mese) subiranno invece un aumento dei loro corsi. La TS4 come detto precedentemente, sarà quella che guiderà l'evoluzione del sistema per via del coefficiente del vettore di co-integrazione più alto in valore assoluto. La variabile TS3 apporterà il contributo successivo maggiore nella determinazione dell'equilibrio, seguita dalle variabili TS1 e TS2. Queste previsioni adesso saranno riviste alla luce dei valori dei *loading*. Allora, per quanto riguarda la velocità di aggiustamento, la variabile TS1 sarà quella che subirà l'aggiustamento maggiore in differenza. Considerando il suo coefficiente di co-integrazione e di *loading* positivo ed il segno dell'errore della combinazione lineare, essa tenderà a muoversi verso la direzione dell'errore fungendo da disturbo al movimento correttivo. I suoi corsi sono previsti al rialzo. La variabile TS2 si muoverà in differenza con una certa velocità e dato il segno del suo *loading* e di co-integrazione parteciperà al meccanismo di correzione

dell'errore in maniera maggiore di tutte le altre variabili. Il suo movimento sarà di salita allo step successivo. La variabile TS3 è la seconda in termini di velocità del movimento correttivo e si muoverà a favore della correzione dell'errore. I suoi corsi saranno ribassisti. La TS4 è quella con la minore velocità di aggiustamento e dati i segni del suo *loading*, del suo coefficiente nel vettore di co-integrazione e dato il segno dell'errore, essa tenderà a disturbare il processo di correzione dell'errore. I suoi corsi saranno ribassisti.

Anche questa volta, utilizzando le serie originali per implementare il sistema, la limitata numerosità campionaria degli indicatori non ha consentito di sperimentare la struttura predittiva su serie troncate, al fine di verificarne la bontà. L'unica modalità di validazione possibile consiste nell'attendere la disponibilità dei dati relativi al periodo successivo, come specificato per il dominio *Salute*. Ciò equivale a testare la metodologia direttamente sul campo, senza la possibilità di retrocedere nel tempo così da applicare la procedura e confrontare l'adattamento del modello a dati già disponibili ma non inseriti nel modello. Come per il dominio *Salute* analizzato precedentemente, si procederà all'implementazione del sistema — applicato alle serie originali — anche sulle serie aumentate generate usando la metodologia utilizzata per l'analisi dello stesso dominio *Salute*.

Anche in questo caso, la tecnica usata per generare delle serie storiche di numerosità campionaria maggiore non ha inficiate le caratteristiche delle serie dalle quali sono state derivate, lasciando sostanzialmente inalterate le proprietà delle traiettorie originali. Dunque, anche le serie aumentate possono essere ragionevolmente considerate come generate dallo stesso processo generatore dei dati, potendo ipotizzare per questo la esistenza di co-integrazione tra le stesse.

Si inizierà presentando i grafici delle serie originali e di quelle aumentate, al fine di evidenziare graficamente la sostanziale similarità tra le due rappresentazioni. Le etichette assegnate alle serie aumentate coincidono con quelle utilizzate per le serie originali per evitare di appesantire le notazioni rischiando di generare confusione.

Così facendo si cercherà di creare un filo continuo tra l'analisi effettuata sulle serie originali e questa, senza rottura tra le due.

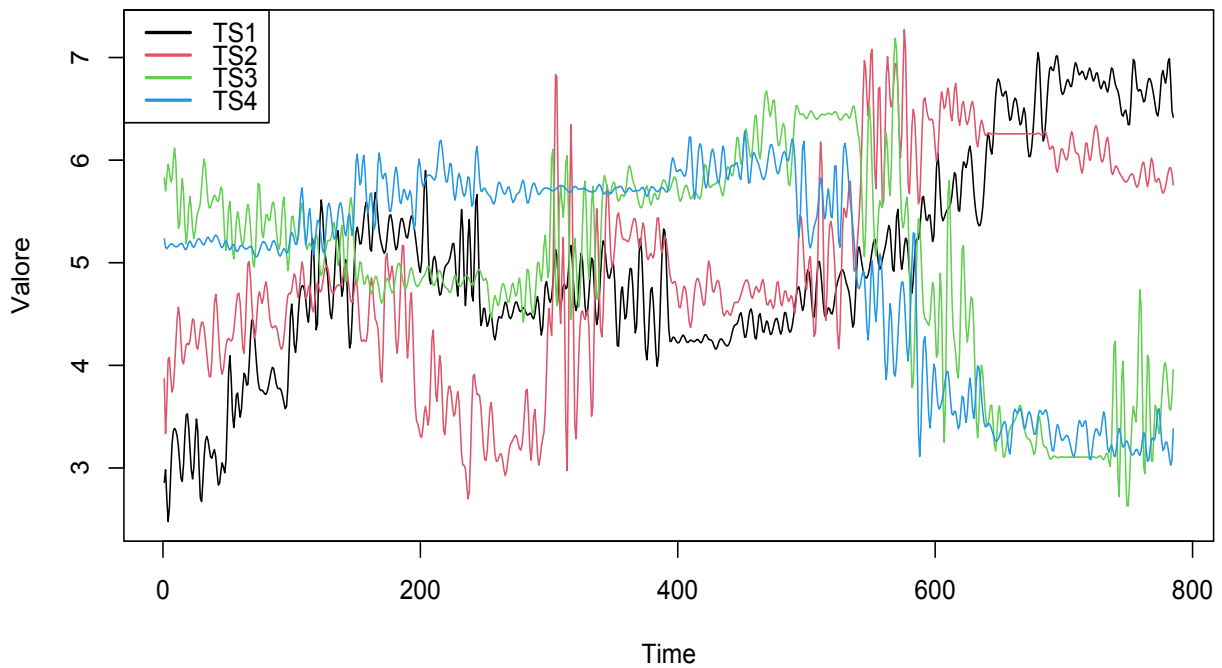
# GRAFICO SERIE ORIGINALI

## Confronto tra serie storiche dal 2004 al 2022



# GRAFICO DELLE SERIE STORICHE AUMENTATE

## Confronto tra Serie Storiche Aumentate



L'analisi grafica suggerirebbe l'esistenza di una sostanziale similarità in termini di traiettorie, il che consente di ipotizzare che l'aumento della numerosità campionaria non abbia alterato le caratteristiche stocastiche delle serie originali.

Anche in questo caso si procederà a verificare la non stazionarietà delle serie aumentate — già osservabile a livello grafico — mediante il test *ADF*, utilizzando un numero di ritardi determinato, anche in questo caso, secondo la regola di *Schwert*.

## TS1

### Augmented Dickey-Fuller Test

data: TS1

Dickey-Fuller = -1.4933, Lag order = 16, p-value = 0.7928

alternative hypothesis: stationary

## TS2

### Augmented Dickey-Fuller Test

data: TS2

Dickey-Fuller = -1.869, Lag order = 16, p-value = 0.6338

alternative hypothesis: stationary

## TS3

### Augmented Dickey-Fuller Test

data: TS3

Dickey-Fuller = -1.0881, Lag order = 16, p-value = 0.9245

alternative hypothesis: stationary

## TS4

### Augmented Dickey-Fuller Test

data: TS4

Dickey-Fuller = -1.4483, Lag order = 16, p-value = 0.8118

alternative hypothesis: stationary

Dai *p-value* si ha conferma della natura non stazionaria delle quattro traiettorie analizzate. È pertanto possibile ipotizzare l'esistenza di una relazione di co-

integrazione tra le serie. Si procederà ora con l'implementazione del sistema co-integrato tramite la procedura di *Johansen*, basata sul test del massimo autovalore.

## PROCEDURA PER VERIFICA RANGO DI CO-INTEGRAZIONE

```
#####  
# Johansen-Procedure #  
#####
```

Test type: maximal eigenvalue statistic (lambda max) , with linear trend

Eigenvalues (lambda):

```
[1] 0.055304784 0.019185821 0.011783917 0.001774199
```

Values of teststatistic and critical values of test:

```
      test 10pct 5pct 1pct  
r <= 3 | 1.38 6.50 8.18 11.65  
r <= 2 | 9.19 12.91 14.90 19.19  
r <= 1 | 15.01 18.90 21.07 25.75  
r = 0 | 44.09 24.78 27.14 32.14
```

Eigenvectors, normalised to first column:  
(These are the cointegration relations)

	ts.TS1..l1	ts.TS2..l1	ts.TS3..l1	ts.TS4..l1
ts.TS1..l1	1.000000	1.000000	1.000000	1.000000
ts.TS2..l1	-1.579311	2.160714	-0.01965614	-0.59279996
ts.TS3..l1	1.814840	2.9611799	0.10924433	-0.03987992
ts.TS4..l1	-1.811493	-0.4074805	0.81158791	-1.91810966

Weights W:

(This is the loading matrix)

	ts.TS1..l1	ts.TS2..l1	ts.TS3..l1	ts.TS4..l1
ts.TS1..d	-0.002771978	-8.626103e-05	-0.0035216096	-5.431931e-05
ts.TS2..d	0.008590903	-5.706170e-04	-0.0010207520	-2.771259e-04
ts.TS3..d	-0.005333899	-1.119731e-03	0.0013003695	4.561138e-06
ts.TS4..d	0.003288527	-1.944525e-04	-0.0007436839	2.392838e-04

```
> as.integer(mean(IND$selection))
```

```
[1] 10
```

```
> adf.test(Comb. Lin.);
```

Augmented Dickey-Fuller Test

data: Comb. Lin.

Dickey-Fuller = -6.5253, Lag order = 9, p-value = 0.01

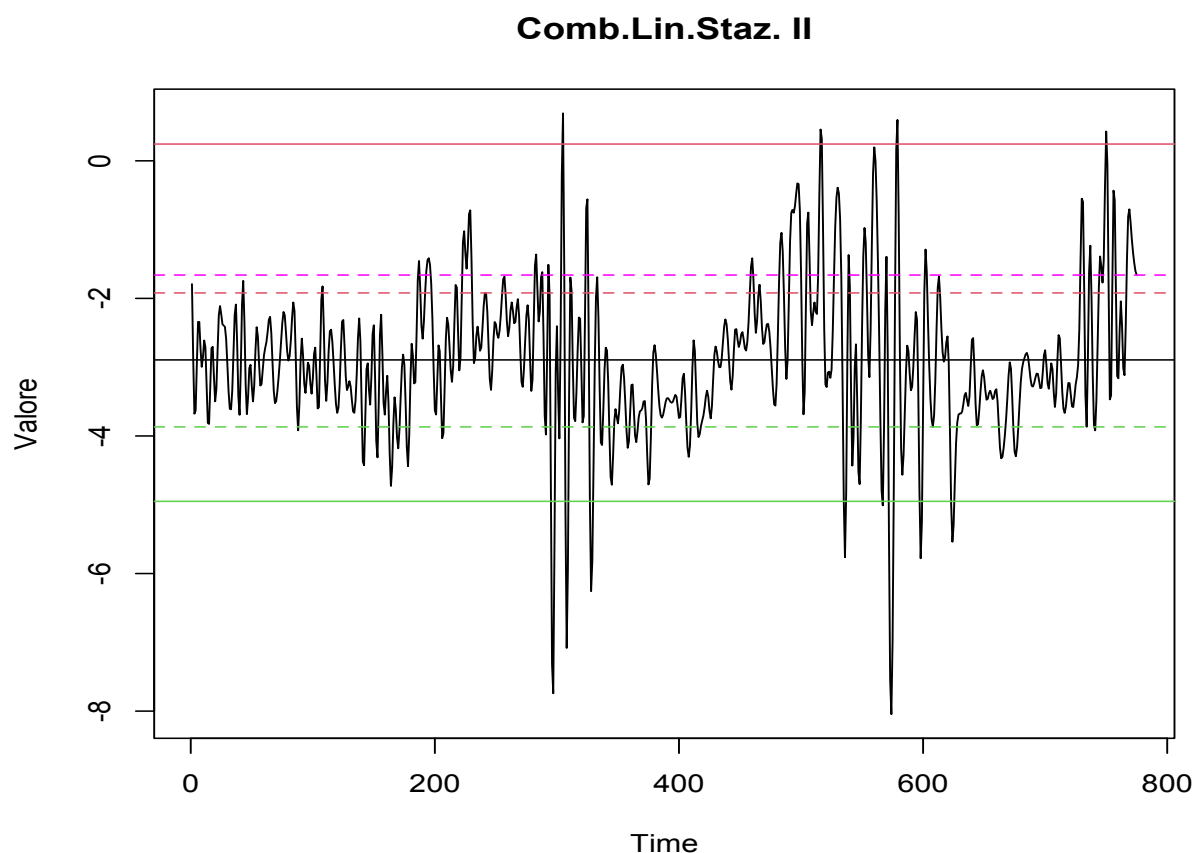
alternative hypothesis: stationary

Messaggio di avvertimento:

In adf.test(Comb. Lin.) : p-value smaller than printed p-value

Un'analisi preliminare consente di individuare la presenza di un rango di co-integrazione, poiché il valore della statistica test risulta inferiore alle soglie critiche. Inoltre, l'ipotesi di assenza di co-integrazione può essere rifiutata ad ogni livello di significatività. I risultati ottenuti suggeriscono nuovamente l'esistenza di un trend di lungo periodo condiviso tra le quattro variabili, coerente con l'ipotesi di un processo generatore comune che ne determina l'evoluzione. Considerando che le serie aumentate provengono da dati originari co-integrati, si può concludere che tale proprietà sia stata mantenuta, confermando la coerenza del processo stocastico tra le due versioni delle serie. E' possibile anche in questo caso esprimere le stesse considerazioni che hanno riguardato l'analisi del precedente dominio, ovvero considerare le serie come espressione di un unico processo generatore sottostante, riconducibile al concetto di *Benessere Economico*. È importante precisare, ancora una volta, che eventuali discrepanze in termini di affidabilità delle stime e di robustezza delle inferenze rispetto ai valori ritenuti ottimali devono essere interpretate alla luce della qualità e della natura dei dati utilizzati. L'attenzione va posta, in particolare, sulla capacità complessiva del metodo proposto di cogliere le dinamiche del fenomeno in esame. In altri termini, qualora si disponesse di dati osservati con una frequenza di rilevazione più fine e rappresentativi dell'effettiva evoluzione del fenomeno, la metodologia adottata sarebbe presumibilmente in grado di generare risultati ancora più robusti. Il fatto che, anche con dati simulati in modo da riprodurre il più fedelmente possibile la realtà, si ottengano risultati statisticamente validi, potrebbe essere considerata come una conferma della possibile solidità dell'approccio proposto. Si inserirà adesso il grafico della combinazione lineare derivante dal sistema. Si potrà constatare, anche graficamente, la natura stazionaria della combinazione lineare, già evidenziata dal test *ADF*.

## GRAFICO DELLA COMBINAZIONE LINEARE

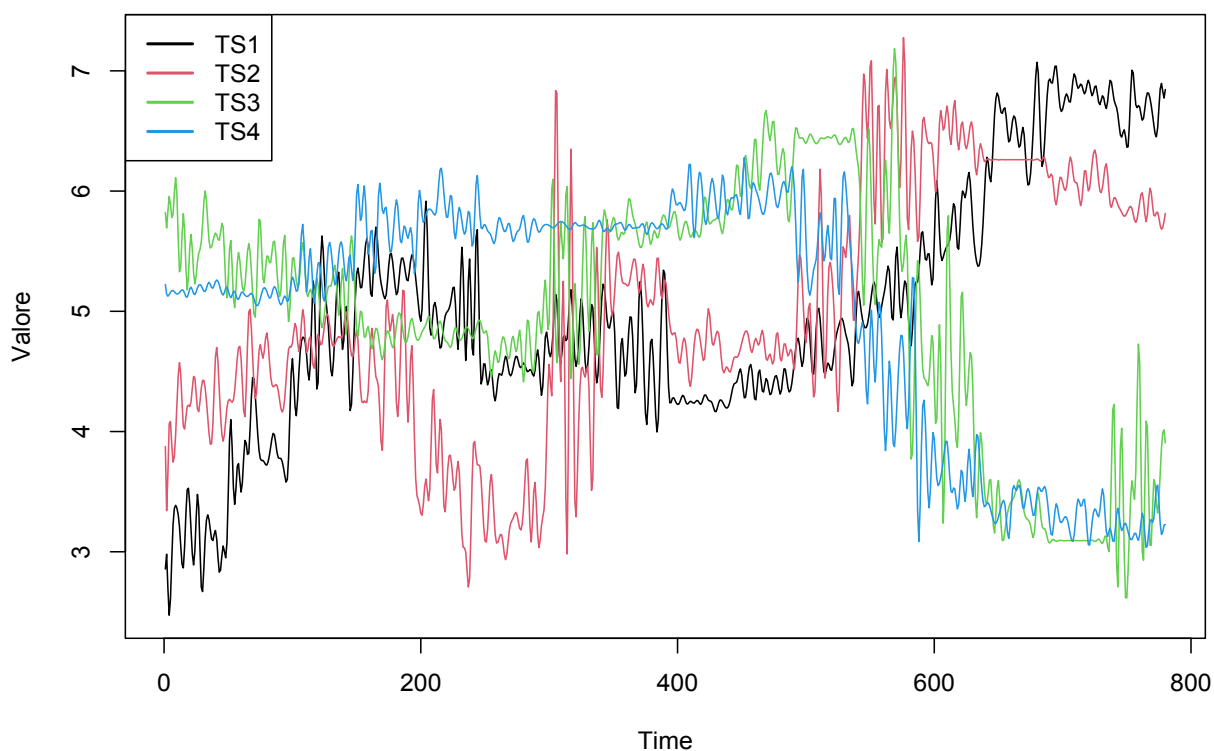


È evidente la stazionarietà della traiettoria, dalla quale si osserva il tipico comportamento di *mean-reversion* caratteristico dei processi stazionari.

Si procederà ora al troncamento delle serie temporali, al fine di retrocedere nel tempo fino ad individuare il punto in cui la combinazione lineare inverte la propria tendenza, segnando un cambiamento nella direzione del movimento. A partire da quel punto, verrà implementata la procedura predittiva al fine di confrontare la direzione suggerita dal sistema con quella effettivamente osservata nei periodi successivi. Tali periodi, pur essendo disponibili, non saranno inclusi nel modello a causa del troncamento delle serie. Dal grafico della combinazione lineare precedente si osserva che il sistema aveva dato un segnale di convergenza verso il valore di equilibrio circa 6 rilevazioni prima. Le serie verranno troncate alla sesta ultima rilevazione considerando quell'istante come il momento attuale. In altre parole, si cristallizza l'analisi in quell'istante, ignorando le ultime 5 osservazioni.

## GRAFICO SERIE STORICHE TRONCATE

Confronto tra Serie Storiche Troncate



Il troncamento delle traiettorie, come detto, ci riporta a quel preciso istante; di conseguenza, ogni analisi effettuata su tali dati non fa altro che riportarci indietro nel tempo, replicando fedelmente la disponibilità e la tipologia di dati presenti in quel momento. Di seguito la diagnostica della procedura di Johansen sul sistema implementato con le serie rappresentate nel grafico.

## PROCEDURA PER VERIFICA RANGO DI CO-INTEGRAZIONE

```
#####  
# Johansen-Procedure #  
#####
```

Test type: maximal eigenvalue statistic (lambda max) , with linear trend

Eigenvalues (lambda):

```
[1] 0.0568257815 0.0201845636 0.0119639950 0.0005445217
```

Values of teststatistic and critical values of test:

```
test 10pct 5pct 1pct  
r <= 3 | 0.42 6.50 8.18 11.65  
r <= 2 | 9.27 12.91 14.90 19.19  
r <= 1 | 15.70 18.90 21.07 25.75  
r = 0 | 45.05 24.78 27.14 32.14
```

Eigenvectors, normalised to first column:

(These are the cointegration relations)

	ts.TS1..l1	ts.TS2..l1	ts.TS3..l1	ts.TS4..l1
ts.TS1..l1	1.000000	1.00000000	1.00000000	1.00000000
ts.TS2..l1	-1.512627	2.20276066	-0.03912755	-0.8181572
ts.TS3..l1	1.801700	2.31084283	0.06097484	-0.1466824
ts.TS4..l1	-1.764651	0.04025843	0.84543538	-2.0024387

Weights W:

(This is the loading matrix)

	ts.TS1..l1	ts.TS2..l1	ts.TS3..l1	ts.TS4..l1
ts.TS1..d	-0.002918748	-0.0001054998	-0.0033715314	-4.387737e-05
ts.TS2..d	0.009091682	-0.0007795325	-0.0008833157	-1.350838e-04
ts.TS3..d	-0.006206294	-0.0011511484	0.0015612464	-1.411400e-05
ts.TS4..d	0.002879563	-0.0002607328	-0.0007358797	1.329194e-04

```
> as.integer(mean(IND$selection))
```

```
[1] 10
```

```
> adf.test(Comb. Lin.);
```

Augmented Dickey-Fuller Test

data: Comb. Lin.

Dickey-Fuller = -6.6029, Lag order = 9, p-value = 0.01

alternative hypothesis: stationary

Messaggio di avvertimento:

In adf.test(Comb. Lin.) : p-value smaller than printed p-value

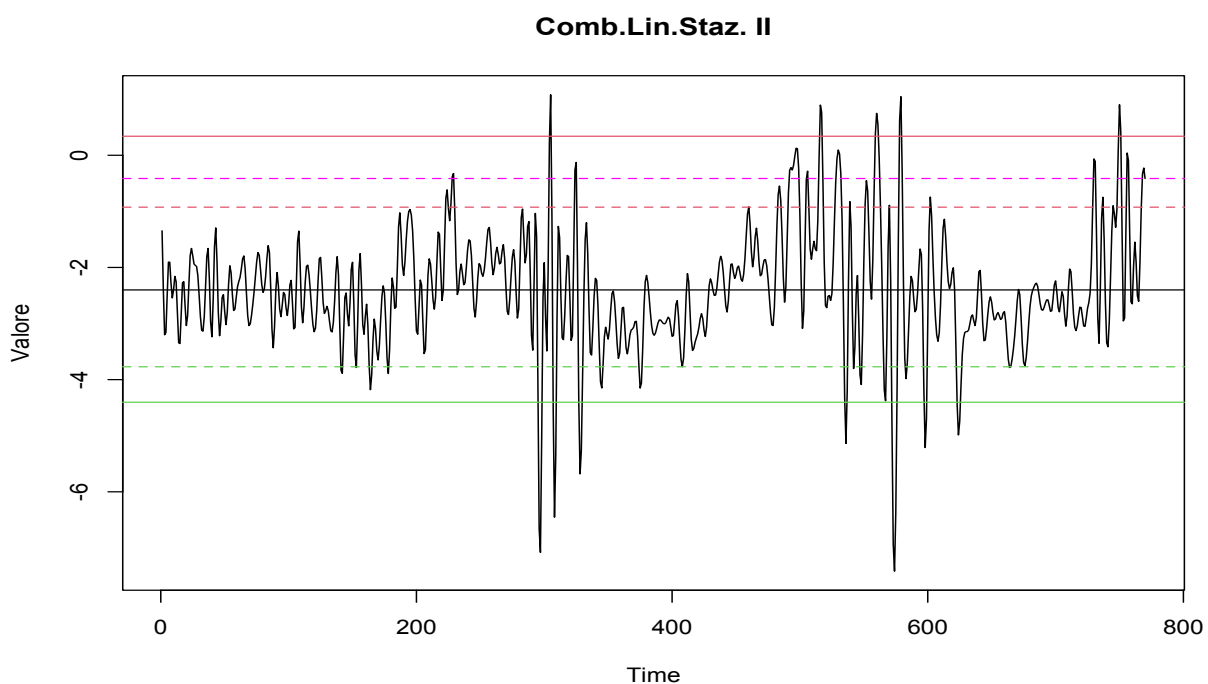
Attraverso l'applicazione della procedura di Johansen, si è riscontrato un rango di co-integrazione pari ad 1, indicando l'esistenza di una sola relazione di equilibrio di lungo periodo tra le quattro variabili analizzate. Il primo autovalore risulta essere il più elevato tra quelli stimati, suggerendo che il relativo vettore di co-integrazione rappresenta la combinazione lineare più significativa in termini di co-movimento di

lungo periodo. All'interno di tale relazione, la variabile TS3 presenta il coefficiente più elevato e di segno positivo, contribuendo in modo diretto al mantenimento dell'equilibrio. Al contrario, TS2 e TS4 mostrano coefficienti negativi, indicando un'influenza opposta alle variabili con coefficiente positivo (TS3 e TS1). In termini relativi, TS3 emerge come la variabile con l'apporto più rilevante alla relazione di co-integrazione, seguita da TS4 (in senso opposto) e da TS2. La variabile TS1 mostra invece un impatto più contenuto rispetto alle altre.

Analizzando i coefficienti di *loading* (matrice dei pesi), si osserva che le variabili TS1 e TS3 presentano coefficienti negativi, contribuendo all'aggiustamento del sistema verso l'equilibrio nel breve periodo alla luce dei segni dei loro coefficienti e di quelli del vettore di co-integrazione. La variabile TS2 mostra la maggiore velocità di reazione (in valore assoluto), presenta un coefficiente positivo, indicando un comportamento che, visto il segno del coefficiente del vettore di co-integrazione, tende a rafforzare il processo di aggiustamento.

La variabile TS4 possiede il coefficiente di *loading* più piccolo in valore assoluto ed il suo segno positivo, associato al segno negativo del suo coefficiente di co-integrazione, indica che anch'essa apporterà un contributo al meccanismo di correzione dell'errore. In conclusione tutte e 4 le variabili — dati i segni dei loro coefficienti di co-integrazione e di *loading*, e dato il segno positivo dell'errore di lungo periodo — contribuiranno al meccanismo di correzione dell'errore di lungo periodo. Si inserirà adesso il grafico della relazione di equilibrio di lungo periodo derivante dal vettore di co-integrazione:

## GRAFICO COMBINAZIONE LINEARE STAZIONARIA



Anche in questo contesto, come già evidenziato dal test *ADF*, si conferma la natura stazionaria della relazione di lungo periodo. La combinazione lineare ha iniziato un movimento di inversione dopo aver registrato un massimo relativo due rilevazioni fa. Attualmente, essa si trova in prossimità della banda rossa tratteggiata, livello interpretato — secondo la teoria delle catastrofi — come una piccola perturbazione del sistema. I livelli assunti dalle variabili in questo istante temporale sono i seguenti:

TS1= 19160

TS2=20.03

TS3=6.103

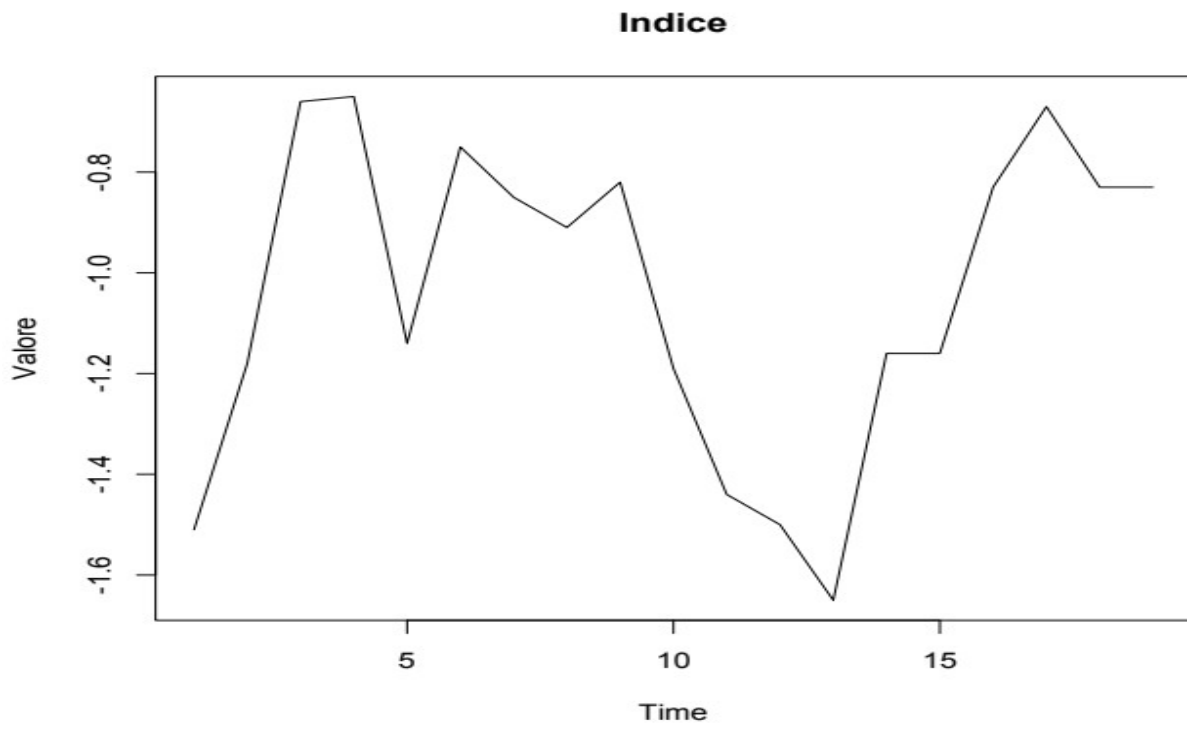
TS4=8.62

In precedenza si è specificato quali siano le unità di misura delle variabili analizzate che, ai fini della verifica della bontà previsionale del metodo implementato, vengono considerate nella loro forma non standardizzata. I valori verranno quindi confrontati al momento del segnale e successivamente, senza attribuire rilevanza alle unità di misura. Al contrario, il modello — come in ogni implementazione — è stato stimato su variabili standardizzate, al fine di consentire un confronto coerente tra gli output delle diverse procedure diagnostiche.

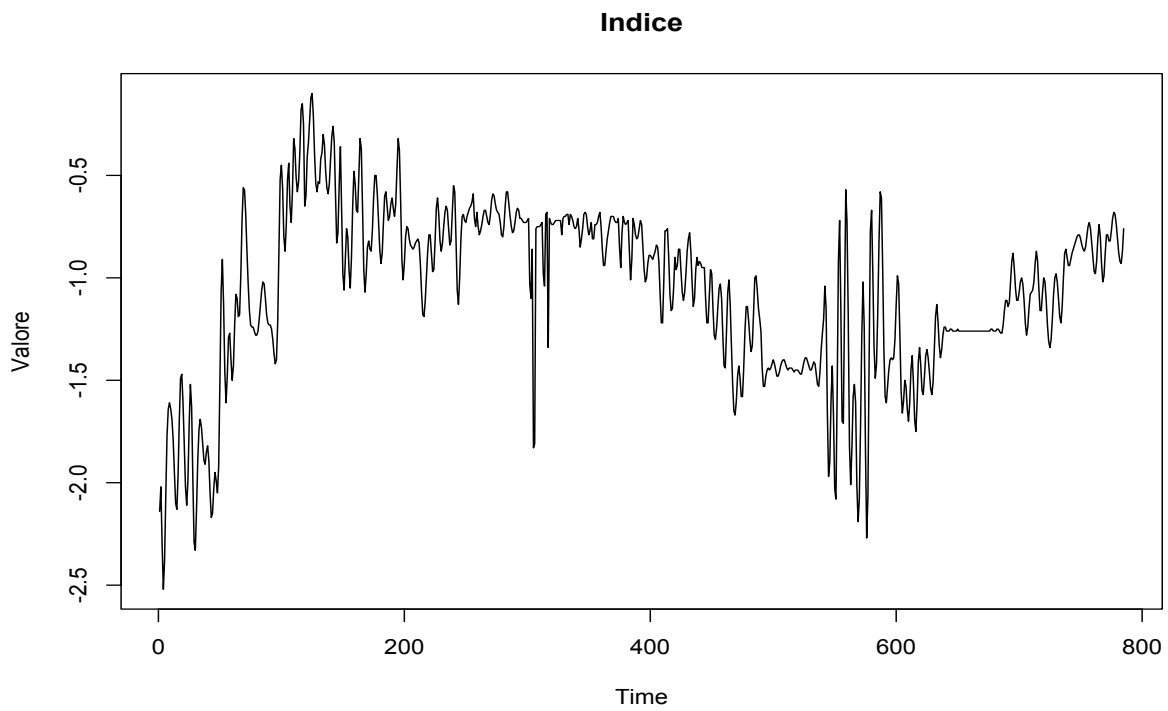
Dal quadro emergerebbe che, ai passi successivi, si dovrebbero trovare valori più bassi per la variabili TS1 e TS3, ed un valore maggiore per TS2 e TS4. I tassi di variazione maggiori dovrebbero subirli la variabile TS2 e TS3 seguite da TS1 e TS4.

Di seguito verrà presentato il grafico comparativo dell'indice costruito sulle traiettorie originali e su quelle aumentate, al fine di valutarne la similarità in termini di andamento e struttura. Successivamente, si procederà al calcolo dell'indice sintetico delle serie troncate utilizzate per l'ultima implementazione del modello. Il metodo usato sarà lo stesso utilizzato per la costruzione dell'indice sulla base delle serie originali.

## GRAFICO INDICE SERIE ORIGINALI

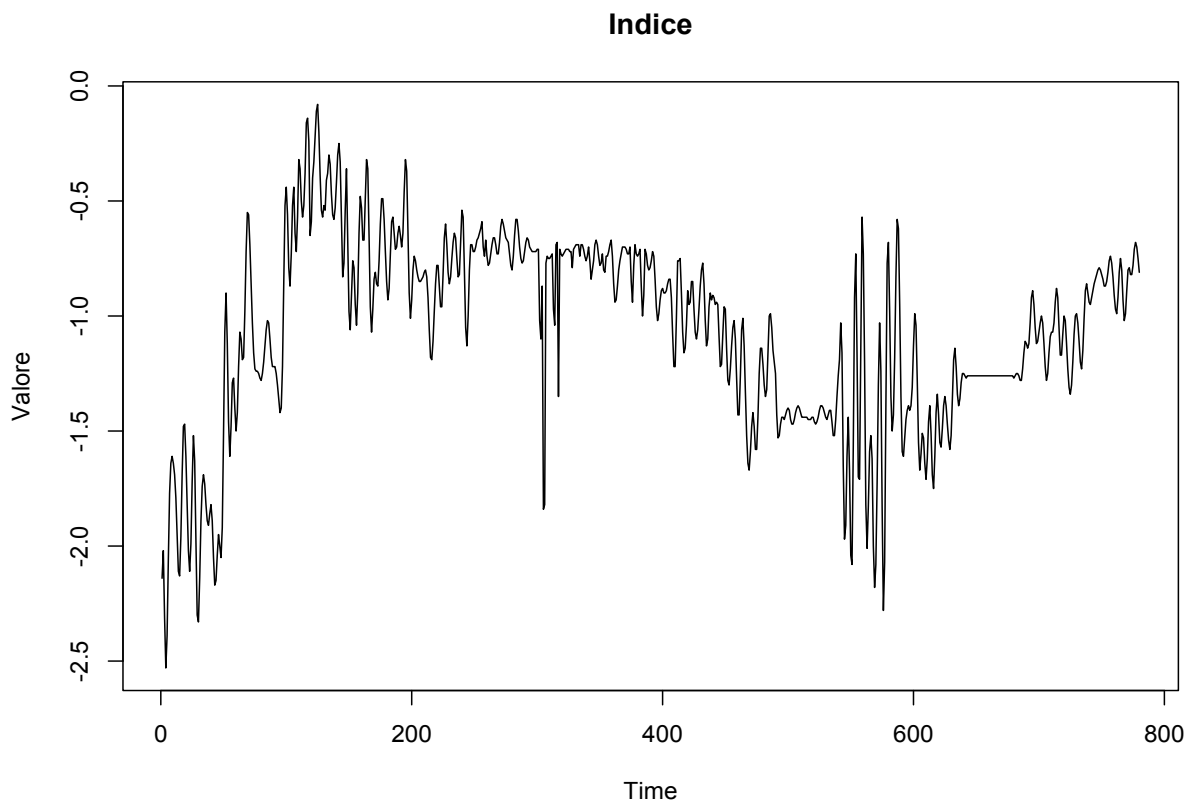


## GRAFICO INDICE SERIE ALLUNGATE



Dalle traiettorie si può ben apprezzare come l'andamento sia uguale in entrambi gli indici a riprova che il metodo per aumentare la numerosità campionaria non ha inficiate le caratteristiche di fondo degli indicatori.

### GRAFICO INDICE SERIE TRONCATE

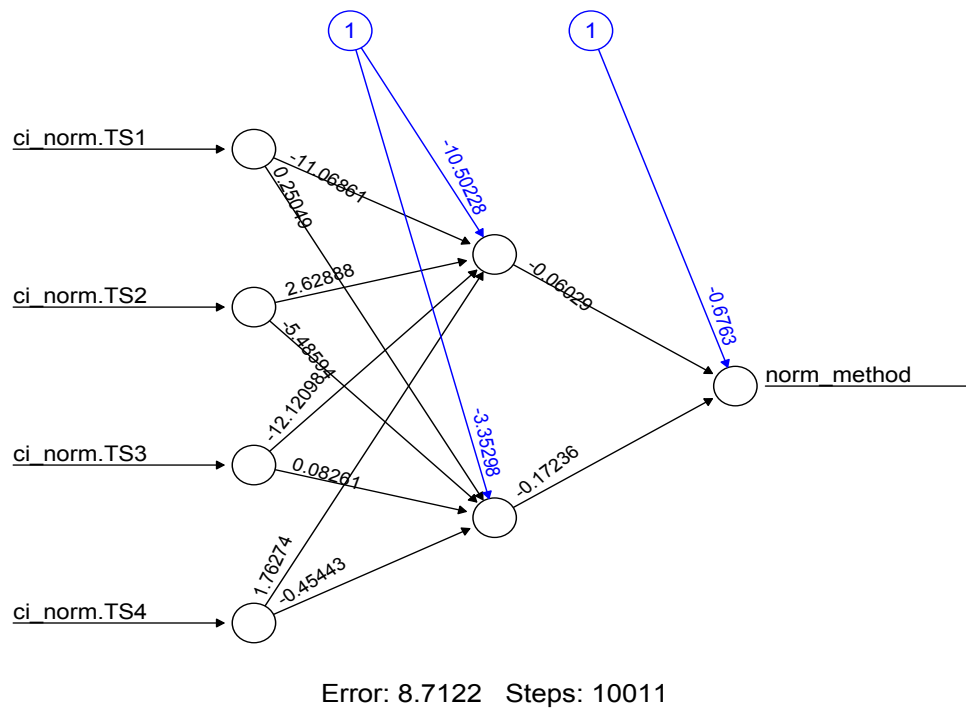


Il grafico riportato rappresenta l'indice sintetico calcolato utilizzando le serie storiche troncate alla sesta ultima rilevazione. Esso fornisce una fotografia dell'indice così come si sarebbe potuto osservare sulla base dei dati disponibili in quell'istante temporale, offrendo quindi una rappresentazione cristallizzata del sistema in quel punto dell'orizzonte temporale.

Verrà ora implementata una rete neurale artificiale al fine di effettuare una previsione sull'andamento futuro dell'indice sintetico. L'obiettivo, anche questa volta, è quello di comprendere in che modo l'interazione tra le serie, modellata attraverso il sistema co-integrato, influenzi l'evoluzione dell'indice.

Si adotterà una rete neurale della stessa tipologia impiegata in precedenza per l'indice calcolato sulle serie originali. La principale differenza consiste nell'utilizzo di due neuroni nascosti, al fine di cogliere eventuali relazioni più complesse generate da una dimensionalità campionaria maggiore.

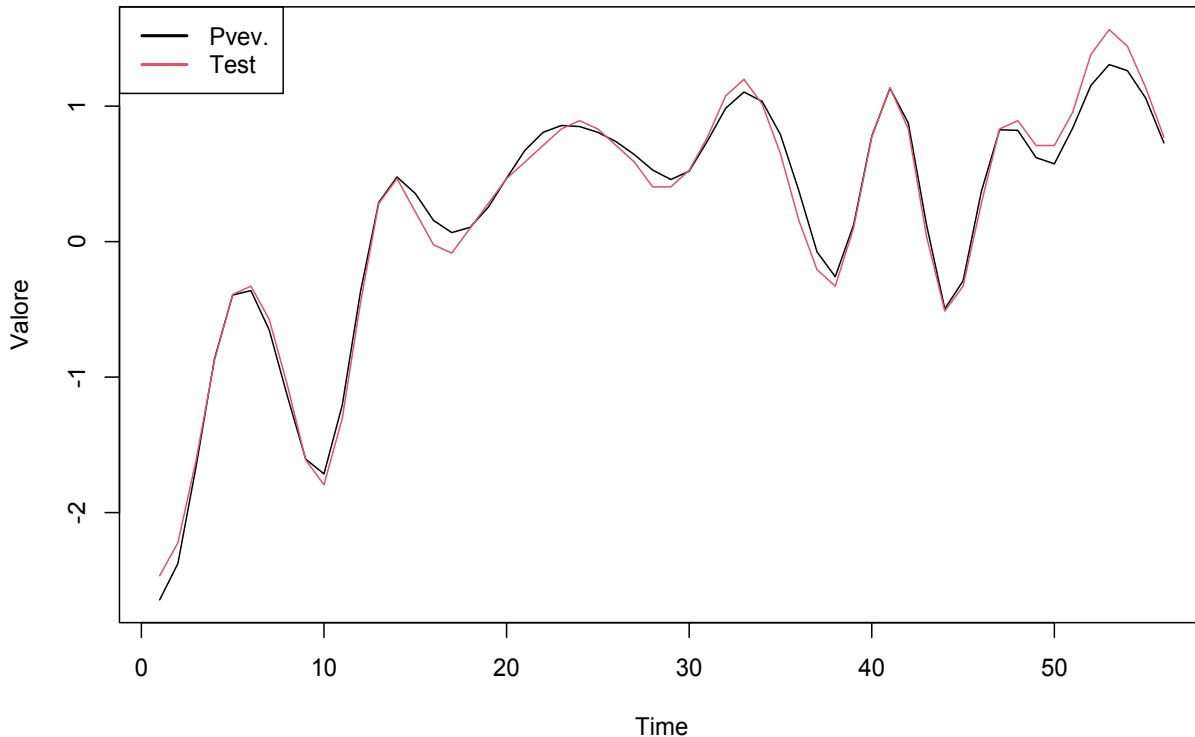
## PLOT RETE IMPLEMENTATA CON MINIMO ERRORE



L'errore minimo ottenuto durante il processo di addestramento della rete neurale è pari a 8.7122 ed è stato raggiunto all'iterazione numero 10011. Di seguito si riporta il grafico relativo alla diagnostica eseguita sul data set di test, costituito dalle ultime 65 osservazioni, con l'obiettivo di valutare la capacità di generalizzazione del modello addestrato. Il data set di training, su cui è stata condotta la fase di apprendimento, include le restanti osservazioni della serie storica.

## PLOT BONTA' ADATTAMENTO

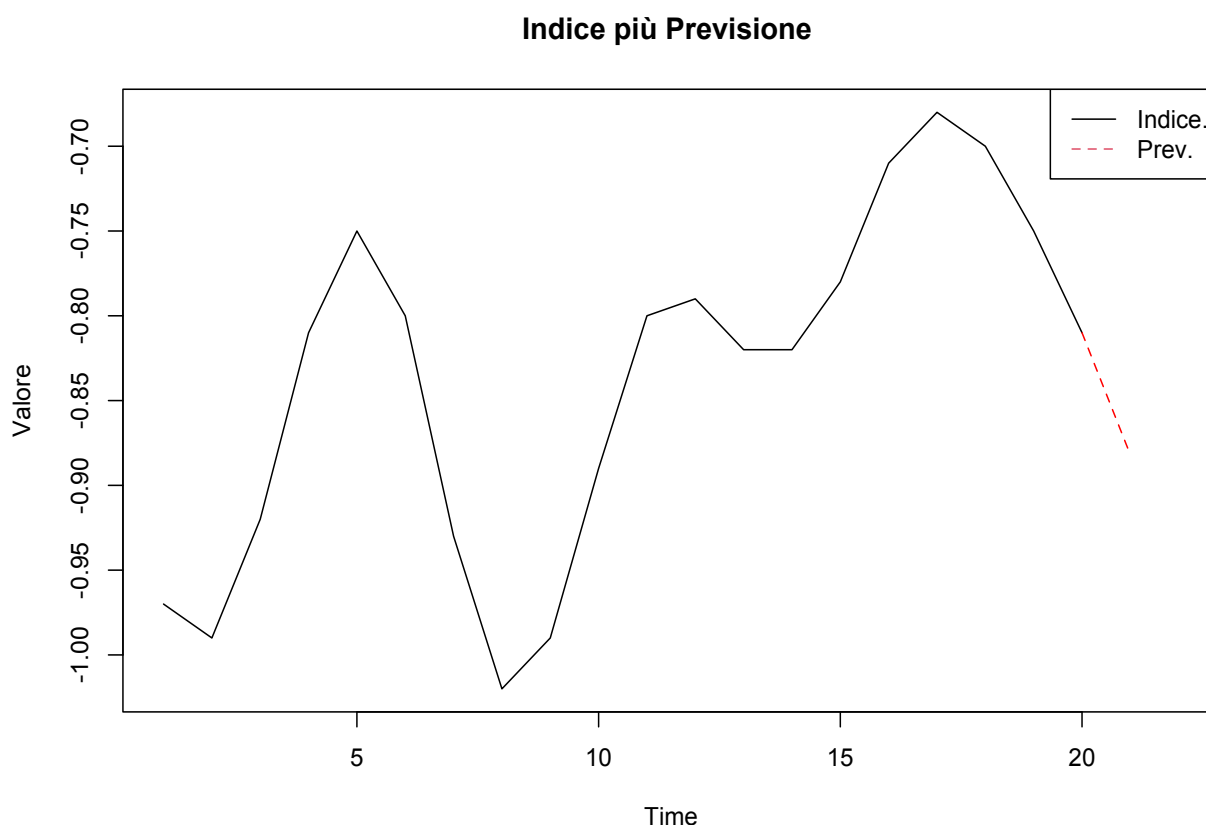
Previsione Indice Normalizzata (corr=0.987)



Dal grafico si può vedere la qualità dell'adattamento sulla serie dei dati di test che genera un coefficiente di correlazione che si approssima ad un valore di 0.987, calcolato col metodo di *Pearson*. Le rilevazioni (dati test ed approssimazione) rappresentate dal grafico sono state standardizzate per avere un impatto visivo più chiaro ed efficace della capacità approssimante della rete.

Di seguito verrà rappresentato il grafico dell'indice delle ultime 20 osservazioni con il valore previsto dalla rete. Si noterà come la previsione restituita sarà di un calo per lo step successivo, in linea con l'andamento che si nota dall'indice calcolato sulle serie allungate e non troncate.

## GRAFICO INDICE E PREVISIONE



La rete neurale ha stimato per il prossimo periodo un valore in diminuzione, pari a -0.882. Questo risultato suggerisce che, sulla base delle dinamiche apprese durante l'addestramento e delle relazioni co-integrate tra le variabili, il sistema tenderà a muoversi in direzione dell'equilibrio di lungo periodo, generando una riduzione del valore dell'indice sintetico. Considerando che un aumento dell'indice rappresenta un miglioramento del livello di benessere in ambito economico, la previsione ottenuta implica una contrazione del concetto di Benessere Economico nel periodo successivo, segnalando un potenziale peggioramento della situazione complessiva misurata dagli indicatori.

L'intero apparato analitico descritto in precedenza dovrà essere implementato ad ogni nuova rilevazione, nel momento in cui saranno disponibili dati aggiornati. Questo approccio permette di mantenere una previsione dinamica e adattiva dell'evoluzione del sistema. In questo studio si omettono le verifiche intermedie e si testa la metodologia inserendo tutte le restanti rilevazioni, arrivando sino alla ultima rilevazione utile.

I valori delle diagnostiche ed i grafici considerando tutte le rilevazioni disponibili sono quelli ottenuti applicando la metodologia alle serie allungate iniziali, prima del troncamento alla sesta ultima rilevazione.

Si riportano di seguito i valori delle serie non standardizzate al momento del troncamento e quelli delle serie a cui sono stati aggiunti i restanti valori.

TS1 precedente: 19160      TS1 ultimo: 18909

TS2 precedente: 20.03      TS2 ultimo: 20.003

TS3 precedente: 6.103      TS3 ultimo: 6.16

TS4 precedente: 8.62      TS4 ultimo: 9.072

Dalla procedura di *Johansen* implementata, attraverso i valori di output in essa presenti e considerando il segno positivo dell'errore della relazione di lungo periodo, si era prevista una discesa per la variabile TS1 con un contributo attivo nella riduzione dell'errore. Dall'ultimo risultato disponibile emerge come la variabile TS1 sia diminuita. Per la variabile TS2 la previsione era stata fatta al rialzo con un apporto positivo alla riduzione dell'errore della relazione di lungo periodo. Vedendo l'ultimo valore assunto dalla variabile si appura come questa sia rimasta sostanzialmente invariata poiché ridotta dello 0.13%. La variabile TS3 aveva ricevuta una previsione al ribasso con un contributo attivo alla riduzione dell'errore. L'ultimo valore denota invece un salita. In fine, per la variabile TS4 era stato previsto un rialzo dei suoi corsi con un contributo attivo nella correzione dell'errore di lungo periodo. La variabile in questione, dall'ultimo valore disponibile, si vede come effettivamente abbia subito un rialzo dei corsi di entità considerevole. Si precisa che, vedendo il grafico della combinazione lineare del sistema implementato con tutte le rilevazioni, esso non abbia raggiunto il valore di equilibrio ma si trovi quasi a metà strada tra il livello a cui il sistema ha generato il segnale di inversione ed il valore di equilibrio. Si può verosimilmente ritenere che, al raggiungimento di detto ultimo valore, dato il carattere ipotizzato co-integrato per struttura del dominio, i valori dovrebbero seguire i corsi definiti dalla procedura.

Comunque, dalla verifica effettuata, il sistema si è comportato sostanzialmente in linea con le previsioni effettuate.

Di seguito si calcoleranno le variazioni percentuali subite dalle variabili nel periodo di riferimento:

TS1:  $\frac{(-19160+18909)}{19160} = -0.013$  la variabile ha avuta una discesa dei corsi del -1.3%

TS2:  $\frac{(-20.03+20.003)}{20.03} = -0.0013$  la variabile ha avuta una discesa dei corsi del -0.13%

TS3:  $\frac{(-6.103+6.16)}{6.103} = 0.0093$  la variabile ha avuta una salita dei corsi dello 0.93 %

TS4:  $\frac{(-8.62+9.072)}{8.62} = 0.052$  la variabile ha avuta una salita dei corsi del 5.2%

Si procederà adesso al calcolo approssimato della variazione attesa che ci si poteva aspettare sfruttando i coefficienti *loading* e quelli del vettore di co-integrazione per effettuare un confronto tra la variazione percentuale delle variabili effettivamente ottenuta e quella attesa sfruttando i *loading* e la relazione di co-integrazione. La procedura è identica a quella utilizzata per effettuare la stessa verifica nello studio del dominio *Salute*.

I primi quattro valori previsti sono utilizzati come stime intermedie per valutare l'andamento delle serie, mentre il quinto valore rappresenta la previsione finale da confrontare con l'osservazione reale a cinque periodi di distanza.

	TS1	TS2	T3	T4	
t+1	6.8183	5.8052	3.8964	3.2017	Valori Intermedi
t+2	6.8166	5.8106	3.8928	3.2034	
t+3	6.8148	5.8162	3.8890	3.2051	
t+4	6.8130	5.8219	3.8850	3.2069	
t+5	6.8111	5.8278	3.8810	3.2088	Valore di Confronto

Questi valori rappresentano la dinamica di aggiustamento secondo questa procedura semplificata, che hanno subito le 4 variabili. I valori sono standardizzati. Si procederà al confronto con i valori realmente avuti de-standardizzando i risultati per capire in che percentuale tale procedura di stima si è discostata dai valori realmente manifestatisi.

## VALORI REALI VARIABILI

TS1 : 18909

TS2 : 20.003

TS3 :6.16

TS4 : 9.072

## VALORI STIMATI VARIABILI

TS1: 19153

TS2: 20.04

TS3: 6.05

TS4:8.47

L'errore percentuale commesso dalla procedura per ogni variabile è il seguente:

$$\text{TS1: } \frac{|(-18909+19153)|}{18909} = 0.012$$

$$\text{TS2: } \frac{|(-20.003+20.04)|}{20.003} = 0.0018$$

$$\text{TS3: } \frac{|(-6.16+6.05)|}{6.16} = -0.017$$

$$\text{TS4: } \frac{|(-9.72+8.47)|}{9.072} = -0.066$$

Le previsioni effettuate tramite questa semplice procedura iterativa ha generati valori in linea con la direzione assunta dalle variabili con un errore percentuale che va da un minimo dello 0.18% ad un massimo del 6.6%.

In particolare, la previsione ha sottostimata la discesa per la variabile TS1 e TS2 rispettivamente del 1.2% per la prima variabile e dello 0.18% per la seconda. Per le variabili TS3 e TS4 invece la previsione ha sovrastimata la discesa rispettivamente del 1.7% e del 6.6%. Anche in questo caso un commento sulla procedura iterativa implementata non può che essere positivo alla luce della semplicità della stessa.

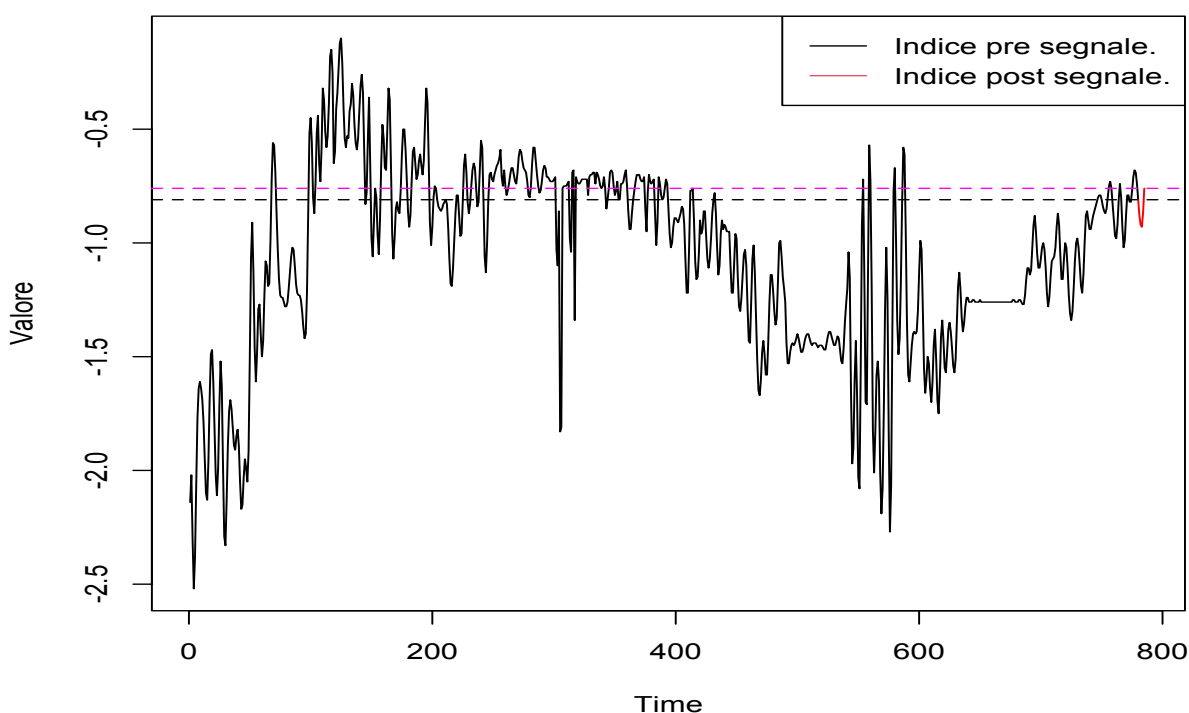
Va aggiunto anche in questo caso il commento che, qualora la previsione fosse stata aggiornata passo dopo passo, inserendo progressivamente i nuovi dati osservati e ricalcolando di volta in volta i vettori di co-integrazione ed i coefficienti di *loading*, la procedura avrebbe potuto restituire stime ancora più precise e puntuali, pur mantenendosi nella sua struttura semplificata.

Per concludere l'analisi, si procederà alla costruzione di un indice sintetico del dominio oggetto di studio, basato sull'aggregazione delle serie storiche considerate. In particolare, l'indice verrà calcolato utilizzando le serie storiche aggiornate, ovvero comprensive delle cinque osservazioni aggiuntive rispetto al troncamento iniziale effettuato per la stima del sistema co-integrato precedente.

In maniera identica allo studio condotto sul dominio precedente, l'obiettivo di questa fase finale dell'analisi è verificare l'andamento assunto dall'indice sintetico nel periodo di riferimento, ovvero quello successivo all'aggiunta delle ultime cinque osservazioni, e valutare in che modo ciascuna delle variabili analizzate — collegate da una relazione di co-integrazione — abbia contribuito a tale evoluzione. In particolare, si intende comprendere come le dinamiche individuali delle serie storiche abbiano inciso sulla traiettoria dell'indice, in coerenza con le forze di lungo periodo rilevate dal sistema *VECM*. Ciò permette di evidenziare non solo la co-movimentazione strutturale tra gli indicatori, ma anche il ruolo specifico di ciascuno nella determinazione della tendenza aggregata osservata.

## GRAFICO INDICE

### Indice più Previsione



L'indice denota un leggero rialzo del benessere economico nelle 5 rilevazioni successive al segnale di inversione dato dalla relazione di lungo periodo del sistema

co-integrato. La differenza è data dalla distanza tra la linea nera tratteggiata e quella magenta presenti sul grafico. Si consideri che, vedendo la distanza tra il livello raggiunto dalla combinazione lineare stazionaria ed il proprio valore di equilibrio, considerando i segni degli indicatori con le loro polarità, si può prevedere che, quando la relazione di equilibrio di lungo periodo avrà effettuato il suo movimento di mean-reversion, con le dinamiche previste degli indicatori, l'indice subirà una flessione andando ad evidenziare una riduzione del Benessere Economico della società a cui gli indicatori si riferiscono. Alla luce dei risultati ottenuti dalla metodologia sperimentata si può concludere che, le variabili oggetto di studio hanno manifestata una dinamica dettata dalle relazioni di co-integrazione che ben si adattano alla evoluzione che l'indice ha avuta nel periodo di sperimentazione. Tali risultati hanno trovato conferma sia nel dominio *Salute* riferito al territorio di Taranto, sia nel dominio *Benessere Economico* su scala nazionale. Nel prossimo ed ultimo capitolo verranno presentate ulteriori riflessioni e considerazioni interpretative relative alle analisi svolte, ai limiti delle metodologie adottate e ai possibili sviluppi dell'approccio proposto, con particolare riferimento a un impiego più articolato degli strumenti statistici utilizzati. Inoltre, verrà approfondita la ratio alla base della scelta metodologica, evidenziando come essa possa favorire l'estensione dell'applicazione anche ad altri fenomeni di natura sociale.

## CAP. 5

### Conclusioni e considerazioni finali

Il lavoro presentato nasce dall'idea che il concetto di co-integrazione possa essere intrinseco negli indicatori dei domini *BES* e *BesT*, per loro stessa natura e costruzione. I domini *BES* e *BesT* rappresentano categorie tematiche utilizzate per valutare diversi aspetti del benessere socio-economico ed ambientale delle società, attraverso un insieme di indicatori specifici che riflettono varie dimensioni di uno sviluppo sostenibile. Studi condotti in Europa hanno applicato con successo la co-integrazione per analizzare le relazioni tra indicatori appartenenti a domini differenti, quali ad esempio economia, energia, ambiente, dimostrando l'esistenza di relazioni di lungo periodo tra queste variabili.

Tuttavia, non sono stati finora condotti approfondimenti sull'applicazione di questo concetto agli indicatori appartenenti ad uno stesso dominio. È proprio da questa osservazione che prende le mosse il presente lavoro, basato sulla convinzione che, una corretta specificazione degli indicatori di un dominio, possa implicare, per struttura, la presenza di co-integrazione. I risultati ottenuti nel presente lavoro sembrerebbero confermare l'ipotesi iniziale, mostrando che gli indicatori all'interno di uno stesso dominio *BES* e *BesT* presentino effettivamente una struttura di co-integrazione significativa. Questo risultato suggerisce che tali indicatori condividano un percorso comune di evoluzione nel lungo periodo, rafforzando l'idea che essi rappresentino manifestazioni diverse di uno stesso fenomeno sottostante. L'applicazione delle tecniche di co-integrazione a livello di singolo dominio potrebbe costituire così un nuovo contributo alla letteratura, ampliando le prospettive metodologiche per l'analisi degli indicatori del benessere e sviluppo sostenibile. In particolare, questa impostazione permetterebbe di migliorare la modellizzazione e la comprensione delle dinamiche interne ai domini, offrendo strumenti robusti per la valutazione e le scelte di policy. Inoltre, i risultati ottenuti potrebbero aprire ulteriori possibilità per l'implementazione di modelli previsionali accurati e coerenti, basati su relazioni di lungo periodo tra indicatori strettamente legati tra loro statisticamente. Tutto questo potrebbe favorire un miglior monitoraggio del benessere e delle sue variazioni nel tempo, con ricadute positive sia in ambito della conoscenza del fenomeno sia in quello applicativo. Si evidenzia comunque, la necessità di ulteriori studi che approfondiscano queste dinamiche in altri contesti geografici e temporali, nonché la possibile estensione delle tecniche utilizzate ad indicatori di natura diversa, per validare ulteriormente e generalizzare le evidenze emerse. Intanto, si può dire che, nel presente lavoro, la implementazione di un modello co-integrato, partendo dalla idea di un comune processo generatore dei dati all'interno degli indicatori di un dominio, può essere considerato come un punto di partenza per l'applicazione di tali tecniche alle loro traiettorie, e questo alla luce dei risultati ottenuti. In entrambi i domini analizzati è emerso come la co-integrazione sia effettivamente presente e

caratterizzi gli indicatori di ciascuno di essi. La natura dei dati disponibili ha verosimilmente influenzato, in modo non favorevole, la significatività statistica di alcuni parametri del modello; tuttavia, tale fenomeno deve essere interpretato non come indice di una scarsa applicabilità dei modelli utilizzati, bensì come diretta conseguenza della struttura specifica dei dati ufficiali disponibili per questi fenomeni. In altri termini, qualora si potesse disporre di dati più dettagliati — in termini di ampiezza campionaria e dunque di una più fine frequenza temporale delle osservazioni — l'idea di partenza, che ha già mostrati risultati soddisfacenti anche su dati strutturalmente limitati, potrebbe condurre ad esiti ancora più solidi e significativi. In realtà, gli aspetti appena menzionati possono essere considerati i veri limiti di questa sperimentazione.

La scarsità dei dati, sia in termini di dimensionalità delle traiettorie sia di continuità temporale, rappresenta infatti la principale causa delle eventuali problematiche riscontrabili nella significatività statistica di alcuni parametri del modello proposto. A conferma di ciò, la parte del lavoro in cui le metodologie sono state applicate a dati simulati — generati con una dimensionalità campionaria nettamente superiore ma mantenendo le caratteristiche delle serie originali — ha prodotto risultati nettamente migliori nei valori dei test. Tale miglioramento deve essere interpretato anche alla luce del fatto che i dati simulati, pur coerenti con la struttura delle traiettorie reali, riflettono certamente una rappresentazione più completa e regolare del fenomeno osservato ma non sono rilevazioni reali.

È dunque ragionevole ritenere che, disponendo di dati reali più numerosi e ravvicinati nel tempo, la metodologia proposta restituirebbe risultati ancora più solidi ed accurati, sempre a condizione che le ipotesi teoriche sul fenomeno siano valide. Questo sarebbe dovuto alla maggiore coerenza tra il fenomeno oggetto di analisi ed i dati che lo rappresentano tale da far emergere al meglio il legame di co-integrazione tra gli indicatori. La capacità predittiva di tali modelli si fonda essenzialmente sulla stazionarietà della rappresentazione *VECM*. Per questo motivo, una volta avviato il processo di *mean reversion* — ossia il ritorno verso il proprio valore di equilibrio — la convergenza delle traiettorie inserite nel sistema lungo le direzioni indicate dal vettore di cointegrazione può essere considerata un evento quasi certo. Tale evidenza è già stata riscontrata nell'applicazione del metodo proposto nel presente lavoro su serie aumentate, sia per il dominio *Salute* sia per il dominio *Benessere Economico*. La capacità di prevedere l'andamento delle traiettorie inserite nel sistema, con un errore di previsione sui valori puntuali estremamente contenuto, conferma l'efficacia predittiva del modello rispetto alla tipologia di fenomeno analizzato. Va inoltre sottolineato che tale previsione è stata ottenuta mediante una procedura iterativa che utilizza i coefficienti del vettore di co-integrazione e del vettore dei *loading* disponibili al momento del segnale generato dal sistema, senza che tali parametri siano stati aggiornati ad ogni step di previsione in corrispondenza della disponibilità di nuovi dati. L'implementazione della procedura ad ogni nuova disponibilità di dati renderebbe l'analisi più coerente e consentirebbe di ottenere stime ancora più precise, migliorando in modo significativo la capacità predittiva del

modello. Il fulcro concettuale che ha guidato l'intero lavoro può essere ricondotto al presupposto, più volte richiamato, dell'esistenza di una relazione di co-integrazione, implicita nella costruzione degli indicatori dei domini *BES* e *BesT*. Tale relazione costituisce l'idea cardine attorno alla quale si è sviluppata l'intera analisi. Per quanto riguarda l'impiego delle reti neurali, esse sono state introdotte come strumento ausiliario nell'applicazione del metodo, sia agli indicatori originali sia a quelli simulati. In particolare, nell'implementazione del modello su serie originali, la limitata numerosità campionaria ha reso necessario l'utilizzo di una rete neurale al fine di prevedere l'andamento della combinazione lineare stazionaria derivante dal sistema co-integrato. Diversamente, nell'applicazione del modello su serie aumentate, la rete neurale non si è resa necessaria, poiché la maggiore numerosità campionaria ha consentito di prevedere l'evoluzione della combinazione lineare stazionaria senza il ricorso a strumenti ausiliari. La rete è stata successivamente impiegata anche per la previsione dello stato futuro dell'indice sintetico, sia su dati originali sia su dati amplificati. La qualità del *forecast* ottenuto ha evidenziato come le reti neurali possano rappresentare un efficace strumento di supporto per la previsione dell'indice di riferimento del dominio analizzato, indipendentemente dalla configurazione campionaria considerata. L'inserimento di nuove osservazioni ha, infatti, confermato la previsione, almeno in termini direzionali, dell'indice sintetico.

Tale informazione aggiuntiva ha consentito di contestualizzare ed, in un certo senso, di giustificare ulteriormente le previsioni prodotte dal sistema co-integrato. Ad esempio, una previsione di contrazione dell'indice, data la polarità degli indicatori, può essere interpretata in modo coerente con le variazioni previste dei singoli indicatori: conoscendo in anticipo con un buon livello di probabilità il segno atteso dell'evoluzione dell'indice ed il significato attribuito alle sue variazioni, diviene possibile comprendere e motivare più compiutamente la direzione delle dinamiche previste. L'utilizzo delle reti neurali, nel caso del modello implementato con serie originali e di quello implementato con serie allungate, ha avuta come unica differenza il numero di neuroni nascosti impiegati nel caso delle serie allungate. Infatti, in tale ultima configurazione, il numero di neuroni *hidden* è stato uguale a 2 anziché ad 1, come nel caso del modello basato su serie originali, è comunque rimasto uguale il numero di *layer nascosti* pari ad 1. La funzione di attivazione impiegata è stata la stessa per entrambe le reti. Si tratta della funzione *SoftPlus* che è la versione approssimata della funzione *ReLU* (Rectified Linear Unit). La funzione in esame è della forma  $\log_{10}(1+e^x)$  ed utilizzabile col packages *neuralnet*, adoperato per le applicazioni presentate in questo lavoro. Come specificato nel capitolo precedente, non deve sorprendere il fatto che, pur avendo impiegato una struttura di rete neurale relativamente semplice, siano stati ottenuti risultati di elevata qualità in termini predittivi. Tale evidenza è riconducibile alle caratteristiche delle serie sulle quali si è implementata la rete neurale: strutture di dati più complesse richiedono infatti architetture di rete più articolate per poter raggiungere livelli di accuratezza predittiva comparabili con quelli ottenuti con i dati analizzati in questo lavoro. Un'ulteriore

considerazione che si ritiene opportuno formulare riguarda la motivazione per cui l'utilizzo della co-integrazione può risultare particolarmente proficuo a fini predittivi. Come già evidenziato, tali processi vengono impiegati in diversi ambiti di analisi che coinvolgono fenomeni caratterizzati da un'evoluzione dinamica tale da giustificare l'applicazione. Sono state inoltre discusse le principali problematiche che possono emergere nell'utilizzo di queste metodologie, le quali problematiche, in determinate condizioni, possono ridurre o persino annullare la capacità predittiva della co-integrazione. L'idea alla base del presente lavoro assume rilievo proprio nell'intento di eliminare, o almeno mitigare, tali criticità, rafforzando così l'efficacia predittiva complessiva del metodo adoperato. Infatti, considerando il *framework* concettuale che ha ispirato il presente lavoro ed i risultati ottenuti nell'implementazione del sistema predittivo proposto, si può ritenere che fenomeni di qualunque natura, qualora le traiettorie delle diverse realtà che li rappresentano possano essere concettualmente ricondotte ad un comune processo generatore dei dati, allora contemplino per loro stessa struttura il concetto di co-integrazione. In conclusione, il presente lavoro ha cercato di mostrare come la co-integrazione possa essere interpretata non solo come un vincolo di equilibrio di lungo periodo tra indicatori, ma anche come una caratteristica intrinseca delle traiettorie che descrivono il fenomeno analizzato. Considerando il *framework* concettuale proposto ed i risultati ottenuti nell'implementazione del sistema predittivo, si può ritenere che, in presenza di fenomeni la cui dinamica possa essere — per costruzione — ricondotta ad un comune processo generatore dei dati, l'applicazione di sistemi co-integrati conduca a risultati statisticamente significativi e stabili nel tempo.

Tale evidenza suggerisce che la metodologia sviluppata non si limiti ai domini analizzati ma che possieda una generalizzazione potenziale verso altri fenomeni che condividano caratteristiche strutturali analoghe. Inoltre, i risultati predittivi del modello, confermati dall'applicazione sia a dati originali sia a dati aumentati, potrebbe aprire prospettive interessanti per ulteriori sviluppi applicativi, offrendo la possibilità di anticipare con affidabilità l'evoluzione di traiettorie complesse in contesti diversi da quelli considerati nel presente studio.

## RIFERIMENTI BIBLIOGRAFICI

### *Nota introduttiva:*

*I testi e gli articoli consultati elencati di seguito sono stati utilizzati esclusivamente come riferimento per l'implementazione della modellistica statistica impiegata nel presente lavoro. L'analisi, i risultati e le conclusioni qui esposti derivano interamente dalle idee e dal lavoro personale dell'autore. Alcune parti di sintesi teorica, come la teoria delle catastrofi o il concetto di crisi, hanno tratto solo spunto dalla bibliografia consultata; le interpretazioni ed i commenti riportati sono frutto dell'elaborazione personale dell'autore. L'analisi è stata condotta in ambiente R utilizzando diversi pacchetti open-source per la manipolazione e visualizzazione dei dati.*

- Jones, E., Harden, S., & Crawley, M. J. (2023). *The R Book* (3rd ed.). WILEY.
- Lantz, B. (2020). *Machine Learning con R*. APOGEO.
- Botto, M., & Raganelli, V. (2021). *Data visualization in R* (1st. ed.). In riga edizioni.
- Ciaburro, G. (2016). *Guida alla programmazione con R*. Edizioni Create Space.
- Hastie, T., Tibshirani, R., & Friedman, J. (2008). *The Elements of Statistical Learning: Data mining, inference, and prediction* (2nd ed.). Springer.
- Botto, M., & Raganelli, V. (2021). *Programmazione ed analisi statistica con R* (1st ed.). In riga edizioni.
- Liberati, F. (2023). *Fondamenti di Programmazione in Linguaggio R*. (1st ed.) STREETLIB.
- López, C. P. *Modelli Econometrici con Strumenti di Apprendimento Automatico. Esempi con R*.
- Cammarata, S. (1997). *Reti neurali. Dal perceptron alle reti caotiche e neuro-fuzzy*. (2nd ed.) ETASLIBRI.
- Ciaburro, G., & Venkateswaran, B. (2017). *Neural Networks with R* (1st ed.). Packt Publishing.
- Russell, S., & Norving, P. (2021). *Intelligenza Artificiale Vol.1 e 2* (4th ed.). Pearson.
- Mazziotta, M., & Pareto, A (2020). *Gli indici Sintetici*. G. Giappichelli Editore.
- Bertuglia, S. C., & Vaio, F. (2011). *Complessità e modelli* (1st ed.). Bollati Boringhieri.
- Morin, E. (2017). *La sfida della complessità*. Editoriale Le Lettere.
- Morin, E. (2021). *Per una teoria della crisi*. Armando Editore.
- Morin, Ed. (2022). *Svegliamoci*. MIMESIS EDIZIONI.
- Brunnermeier, M. K. (2024). *La società resiliente*. Il Mulino.
- D'Apice, V., & Ferri, G. (2009). *L'INSTABILITÀ FINANZIARIA: DALLA CRISI ASIATICA AI MUTUI SUBPRIME*. (1st ed.). Carocci.

- Hobsbawm, E. J.(2025). *Il secolo Breve 1914-1991*. (1st ed.) BUR I Pilastrri. Rizzoli.
- Tonietti, T. M. (1993). *Catastrofi. Il preludio alla complessità*. Edizioni Dedalo.
- Vladimir, A. (1990). *Teoria delle Catastrofi*. (1st ed). Bollati Boringhieri.
- Thom, R. (1980). *Stabilità strutturale e morfogenesi. Saggio di una teoria generale dei modelli*. Einaudi.
- Guerraggio, A., & Nastasi, P. (2005). *Renè Thom e la teoria delle catastrofi*. CENTRO ELUISI Università Bocconi
- Thom, R. (1980). *Parabole e catastrofi: intervista su matematica, scienza e filosofia* (a cura di Giulio Giorello e Simona Morini). Il Saggiatore, Milano.
- Devaney, R. L. (1990). CAOS E FRATTALI. *Matematica dei sistemi dinamici e applicazioni al calcolatore*. ADDISON-WESLEY.
- Corbetta, P. (2021). *La ricerca sociale: metodologia e tecniche*. Vol. I e IV (2nd ed.) Il Mulino.
- Gardini, A., Cavaliere, G., Costa, M., Fanelli, L., & Paruolo, P. (2000). *Econometria*. Volume primo. FrancoAngeli.
- Gardini, A., Cavaliere, G., Costa, M., Fanelli, L., & Paruolo, P. (2007). *Econometria*. Volume secondo. FrancoAngeli.
- Fusco, N., Marcellini, P., & Sbordone, C.(1996). *Analisi Matematica due*. Liguori Editore.
- Fritsch S, Guenther F, Wright M (2019). *\_neuralnet: Training of Neural Networks\_*. R package version 1.44.2, <https://CRAN.R-project.org/package=neuralnet>.
- Bernhard Pfaff (2008). VAR, SVAR and SVEC Models: Implementation Within R Package vars. Journal of Statistical Software 27(4). URL <https://www.jstatsoft.org/v27/i04/>.
- Pfaff, B. (2008) Analysis of Integrated and Cointegrated Time Series with R. Second Edition. Springer, New York. ISBN 0-387-27960-1
- R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Shumway, R. H, & Stoffer, D. S.(2011). *Time Series,Analysis and Its Applications: With R Examples*.(3rd ed.) Springer.
- Johansen S. (1995). *Identifying restrictions of linear equations with applications to simultaneous equations and cointegration*. Journal of Econometrics, 69, 111-132.
- Engle, R. F. & Granger, C. W. J. (1987). *Co-integration and error correction: Representation, estimation and testing*. *Econometrica*, 55(2), 251-276.
- Johansen, S. (1991). *Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models*. *Econometrica*, 59(6), 1551-1580.

- Johansen, S. (1992). *Determination of Cointegration Rank in the Presence of a Linear Trend*. *Oxford Bulletin of Economics and Statistics*, 54(3), 383-397.
- Johansen, S. & Juselius, K. (1990). *Maximum Likelihood Estimation and Inference on Cointegration – with Applications to the Demand for Money*. *Oxford Bulletin of Economics and Statistics*, 52, 169-210.
- ISTAT. (2013). *BES – Benessere equo e sostenibile 2013*. Istituto Nazionale di Statistica. <https://www.istat.it/it/archivio/bes>
- ISTAT. (2016). *BES – Benessere equo e sostenibile 2016*. Istituto Nazionale di Statistica. <https://www.istat.it/it/archivio/bes>
- ISTAT. (2022). *BES – Benessere equo e sostenibile 2022*. Istituto Nazionale di Statistica. <https://www.istat.it/it/archivio/bes>