

**POLITECNICO DI MILANO**

**Scuola di Ingegneria Industriale e dell'Informazione**

**Corso di Laurea Magistrale in Ingegneria Gestionale**



**STRUCTURAL EQUATION MODEL: UN MODELLO STATISTICO PER LE  
SCIENZE ECONOMICHE.**

**CONCETTI, METODOLOGIE E APPLICAZIONE AL PROJECT FINANCE**

**Relatore: Ch.mo Prof. Rocco Roberto MOSCONI**

**Candidato:**

**Francesca RICCIATO**

**Matr. 783363**

**Anno Accademico 2012/2013**



## Sommario

1. INTRODUZIONE.....	6
2. PROJECT FINANCE.....	7
2.1. Definizione.....	7
2.2. Caratteristiche.....	8
2.2.1. Differenze tra project finance e corporate finance.....	8
2.2.2. Settori di impiego.....	11
2.2.3. Attori coinvolti.....	12
2.2.4. Classificazione delle opere.....	13
2.3. Rischi del project finance.....	14
2.4. Bancabilità.....	18
2.5. Il ruolo delle banche.....	19
2.6. Italia.....	21
2.6.1. Normativa.....	21
2.6.2. Il contesto.....	22
2.6.3. Project finance e settore pubblico.....	23
2.7. Vantaggi.....	26
2.8. Svantaggi.....	27
2.9. Conclusioni.....	28
3. PROJECT BONDS.....	29
3.1. Caratteristiche.....	29
3.1.1. Definizione.....	29
3.1.2. Garanzie.....	30
3.2. Rilancio dell'economia.....	31
3.3. Situazione europea.....	32
3.4. Funzionamento.....	33
3.5. Italia e Europa a confronto.....	33
3.6. Le banche.....	34
4. MODELLI SEM.....	36
4.1. Quando utilizzare SEM.....	36
4.2. Caratteristiche.....	37
4.3. Specificazione del modello.....	38

4.3.1.	Procedure di specificazione .....	39
4.3.2.	Variabili latenti e variabili osservabili .....	41
4.3.3.	Rappresentazione grafica.....	42
4.3.4.	Identificazione.....	44
4.3.5.	Disconfirmability .....	45
4.3.6.	Modelli equivalenti .....	45
4.3.7.	Strategie.....	46
4.4.	Stima dei parametri.....	49
4.4.1.	Concetti base nella fase di stima .....	50
4.4.2.	Problemi pratici.....	51
4.4.3.	Modelli: massima verosimiglianza .....	52
4.4.4.	Interpretazione dei parametri stimati .....	55
4.4.5.	Fittare il modello con la matrice di correlazione .....	55
4.4.6.	Metodi di stima alternativi.....	56
4.4.7.	Altri Modelli .....	57
4.4.8.	La funzione di fitting .....	58
4.4.9.	Errori standard .....	58
4.4.10.	Conclusioni .....	58
4.5.	Non normalità .....	59
4.5.1.	Panoramica sulla teoria di stima normale .....	60
4.5.2.	Basi teoriche per il problema .....	61
4.5.3.	Rilevamento degli scostamenti dalla normalità .....	64
4.5.4.	Risultati di studi empirici sulla non normalità .....	65
4.5.5.	Tecniche di stima alternative .....	67
4.5.6.	Ridefinizione di variabili .....	69
4.5.7.	Conclusioni.....	72
4.6.	Estensione: variabili categoriche.....	73
4.6.1.	Il modello .....	74
4.7.	Valutazione del modello.....	77
4.7.1.	Il test $\chi^2$ .....	77
4.7.2.	Performance del test $\chi^2$ .....	78
4.7.3.	Distribuzioni $\chi^2$ non centrate .....	80

4.7.4.	Indici di fit.....	80
4.7.5.	Tipologie di indici .....	81
4.7.6.	Problemi di valutazione degli indici di fit e criteri di selezione .....	84
4.8.	Interpretazione e modifiche.....	85
4.8.1.	Interpretazione .....	85
4.8.2.	Modifiche .....	85
5.	DETERMINAZIONE DELL'ATTRATTIVITÀ DEL PF CON L'UTILIZZO DEI SEM .....	87
5.1.	Attrattività del Project Finance .....	87
5.1.1.	Costi di agenzia .....	87
5.1.2.	Utilizzo di una forte quota di debito.....	88
5.1.3.	Gestione del rischio.....	89
5.2.	Indebitamento.....	89
5.3.	Dimensione del progetto.....	90
5.4.	Settore .....	91
5.5.	Regione.....	92
5.6.	Impegno degli investitori.....	92
5.7.	Modello .....	94
5.8.	Analisi dei dati .....	95
5.8.1.	Database iniziale .....	95
5.8.2.	Preparazione del dataset .....	97
5.9.	Stima.....	99
5.10.	Interpretazione .....	101
5.10.1.	Coefficienti .....	101
5.10.2.	Significatività .....	103
5.11.	Bontà del modello.....	103
5.12.	Conclusioni.....	104
6.	BIBLIOGRAFIA .....	105

# 1. INTRODUZIONE

Il presente lavoro inizia con una descrizione generale del project finance, definito in letteratura economica come “tecnica innovativa di finanziamento” di un progetto valutato dai finanziatori per la sua capacità di generare flussi di cassa atti a rimborsare il debito, piuttosto che sull’affidabilità e capacità di credito dei soggetti che hanno avuto l’idea imprenditoriale di avviare il progetto. Lo strumento è relativamente nuovo e il mercato del project finance, in continua espansione, ha subito rilevanti trasformazioni nel corso dell’ultimo decennio. Mentre alla fine degli anni Novanta la maggior parte delle operazioni era studiata e promossa dal settore privato, oggi si assiste sempre più a iniziative legate a opere pubbliche, sia per sopperire ai sopravvenuti stringenti vincoli di bilancio alla finanza pubblica, sia per promuovere un approccio di stampo imprenditoriale anche nel campo degli investimenti infrastrutturali. Nel primo capitolo si delineano le caratteristiche salienti della finanza di progetto, con riferimento agli aspetti procedurali dell’iniziativa a confronto con la più tradizionale finanza di progetto, ai rischi sottostanti, ai soggetti coinvolti, al ruolo delle banche, alla disciplina normativa vigente nel nostro Paese e un breve dettaglio sull’attività specifica quando viene coinvolto il settore pubblico. Nel secondo capitolo viene introdotto il concetto di project bonds; sebbene questi prodotti finanziari non rientrino nella trattazione econometrica successiva, si ritiene di doverli rendere noti al lettore in quanto parte fondamentale della continua innovazione della finanza di progetto. Dopo aver analizzato gli aspetti giuridici, industriali e finanziari delle operazioni della finanza di progetto, viene svolta la parte rilevante del presente lavoro è dedicata alla descrizione dei modelli SEM, modelli statistici nati dalla unione del metodo della regressione lineare e dell’analisi fattoriale; i modelli vengono inizialmente specificati, in seguito vengono stimati i parametri che devono essere testati rispetto alla loro significatività all’interno della teoria sviluppata, infine è necessario valutare la bontà del modello per poter affermare che questo sia più o meno adatto a spiegare le teorie ipotizzate per poter apportare delle modifiche, o in ogni caso trarne delle conclusioni. Infine, la quarta e ultima parte è dedicata all’applicazione delle tecniche statistiche esposte con lo scopo di studiare il comportamento dello strumento finanziario project finance in Italia. Grazie a una selezione di dati raccolti da Finlombardia attraverso questionari volontari inviati alle banche, si è cercato di enunciare una teoria che valuti l’attrattività del nuovo strumento finanziario.

## **2. PROJECT FINANCE**

### **2.1. Definizione**

In un contesto caratterizzato da mercati sempre più globali, il project finance è uno strumento che vede crescere il suo utilizzo nell'ambito di investimenti produttivi e infrastrutturali, in quanto può rappresentare un modo nuovo ed efficiente di reperire le risorse necessarie alla realizzazione delle opere stesse. La letteratura che tratta questo argomento risulta essere molto ampia; si è così scelto di menzionare la classificazione di P. Nevitt che nel 1989 afferma che "il Project Finance è il finanziamento di una specifica attività economica mediante un'operazione in cui il finanziatore considera i flussi di cassa e gli utili di progetto come garanzia per il rimborso del debito e le attività dell'unità economica come garanzia collaterale". In altre parole si tratta di una forma di finanziamento differente e alternativa alla tradizionale finanza di impresa, caratterizzata da dei livelli di complessità strutturali elevati; infatti questa alternativa è volta alla realizzazione, gestione e finanziamento di progetti complessi che presentano in particolare garanzie di tipo contrattuale piuttosto che reali, con una amministrazione orientata maggiormente ai flussi di cassa e al profitto. Sono infatti proprio i flussi di cassa generati le garanzie dell'investimento stesso. Lo strumento è innovativo in quanto i flussi di cassa visti come garanzia danno una valutazione prettamente qualitativa del progetto; la valutazione riguarda quindi la sostenibilità economica e finanziaria del singolo progetto, ovvero la capacità che questo presenta di generare flussi di cassa durante l'intero ciclo di vita dell'opera in essere. Si ricorda invece che nella finanza di impresa tradizionale la valutazione dell'equilibrio sopra citato non riguarda il nuovo investimento in sé, ma come questo modifica l'equilibrio globale della totalità dei business presenti in azienda. Si stacca quindi il concetto di merito creditizio dei soggetti partecipanti all'investimento e si valuta la bontà del solo progetto. Ciò può accadere poiché il capitale di debito per il finanziamento viene concentrato in un unico soggetto autonomo di riferimento giuridico e finanziario: viene creata una società ad hoc (Special Purpose Vehicle) da uno o più sponsor che diventa la responsabile della realizzazione dell'opera e a cui, di conseguenza, vengono affidati tutti i mezzi finanziari; le operazioni di project finance risultano così essere fuori bilancio per gli sponsor. La conseguenza è quella che i rischi risultano essere il più possibile limitati per l'azionista poiché delimitati all'interno dei confini di una struttura giuridica autonoma, in grado di separare gli sponsor, il progetto e il finanziamento concesso, e creata appositamente per il progetto, non influenzata dalle altre attività presenti nel bilancio degli sponsor che partecipano all'iniziativa.

Cercando di riassumere ciò che è stato detto precedentemente, di seguito vengono presentate le caratteristiche distintive di questo strumento finanziario. Si possono menzionare tre peculiarità della finanza di progetto. La prima caratteristica è data dall'indipendenza legale dagli sponsor, questi ultimi devono provvedere al capitale di equity e le competenze manageriali iniziali necessarie alla fase di avvio del progetto. Per quanto riguarda i rischi è possibile affermare che siano distribuiti in modo equo tra le parti, rendendo migliore la loro gestione e la loro tollerabilità

da parte dei soggetti singoli. Infine i cash flow generati dalla società di progetto devono coprire tutti i costi operativi, la remunerazione del capitale e la restituzione del debito con i rispettivi interessi. Detto ciò si deve far riferimento alla SPV: questa società creata ad hoc è dunque un'unità economica indipendente che deve possedere degli asset capaci di generare dei flussi di cassa che la rendano sostenibile.

## **2.2. Caratteristiche**

### **2.2.1. Differenze tra project finance e corporate finance**

Per meglio comprendere il project finance è opportuno delineare in maniera più decisa le differenze che si presentano tra questo e la corporate finance.

La prima caratteristica da menzionare che distingue le due tipologie di finanziamento è il ring fence, ovvero la separazione giuridica del progetto che porta di conseguenza alla separazione economica dello stesso da tutte le altre attività possedute dai vari promotori. Questa separazione avviene attraverso la creazione di una società di progetto che ha proprio l'obiettivo di conseguire la separazione tra le controparti della parte giuridica e di quella finanziaria. Infatti così facendo risulta possibile tenere separati i capitali destinati al progetto, dagli altri capitali in possesso degli investitori, dei promotori. Operando tale separazione risulta essere facilitata l'attività di monitoraggio. La società di progetto infatti nasce e muore con la concessione dell'opera stessa: viene creata per raccogliere i capitali e svolgere i lavori, in seguito però si occupa della gestione e della raccolta dei flussi di cassa generati. Tutti i capitali e i flussi di cassa risultano così essere di facile allocazione. Al termine del periodo di concessione, la società non ha più alcuna attività da svolgere e per questo motivo non ha più motivo di esistere. Per quanto riguarda i promotori, il ring fence è importante in quanto consente di limitare l'impatto in caso di un eventuale andamento negativo del progetto, che altrimenti porterebbe a delle conseguenze sgradevoli sull'intero bilancio aziendale. In altre parole il rischio di progetto viene circoscritto solamente all'investimento in sé, al capitale sociale presente nella società di progetto e alle eventuali garanzie collaterali presenti. Ma anche le banche ricevono una protezione dal ring fence in quanto protegge i flussi di cassa generati dalla gestione durante la vita dell'opera; i flussi di cassa risultano essere di fondamentale valore in quanto su questi si basano le aspettative di rimborso del debito e di remunerazione del capitale di equity. Infatti nel corporate finance i flussi di cassa generati da un singolo business non possono essere messi in relazione con una sola parte corrispondente di equity e di debito, non sono direttamente riconducibili. La SPV allora ricopre un ruolo fondamentale assumendo i panni di debitore nei confronti delle banche e ad essa si fa riferimento per tutti i doveri e gli obblighi relativi al dato investimento. Di fondamentale importanza è l'assenza di una storia pregressa e quindi di bilanci precedenti che configurano la società come una start up, pertanto i criteri di finanziamento devono essere valutati su flussi di cassa futuri, quindi non certi, come detto precedentemente, e non su elementi di tipo corporate. Tutto ciò favorisce l'avvicinamento di investitori privati nel progetto. La creazione di una SPV è una soluzione tecnica



per isolare giuridicamente i flussi di cassa di un singolo progetto ed è legata a delle ragioni precise: è una soluzione semplice ed efficace allo stesso tempo, non presenta quindi un trade-off tra facilità di implementazione e di gestione e facilità e coerenza allo scopo; permette di non legare il nuovo debito al debito già contratto per altre attività salvaguardando gli indici di bilancio; è una soluzione ideale per progetti di ingenti dimensioni che necessitano la presenza congiunta di più promotori che condividano il rischio totale (caratteristica di fondamentale importanza in particolar modo nei casi in cui la realizzazione avviene al di fuori del paese di origine dei promotori); possibilità di creare società miste in cui possano partecipare sia aziende pubbliche che private, soprattutto per quanto riguarda la realizzazione di opere infrastrutturali. Questa soluzione non è l'unica alternativa, ma sicuramente la più efficace. Infatti lo stesso scopo può essere raggiunto attraverso clausole inserite nei contratti di finanziamento di un particolare investimento all'interno di una normale attività di corporate governance, da evidenziare però come questa alternativa possa presentare delle complessità più o meno ingenti dal punto di vista operativo; risulta comunque essere una soluzione spesso utilizzata.

Volendo schematizzare le differenze tra project finance e corporate finance è possibile far riferimento alle seguenti voci:

- Garanzie, per il PF si basano esclusivamente sulle attività specifiche del progetto mentre nella corporate finance riguardano l'intero patrimonio del prestatore di fondi e tutte le attività presenti nel suo portafoglio;
- Contabilità, nel PF risulta essere fuori bilancio al contrario della corporate finance; ciò significa che nella corporate finance è possibile fare cross-finanziamento all'interno di tutta l'azienda, mentre nel project finance i flussi di cassa generati da tale business possono essere utilizzati solo nello stesso;
- Valutazione del progetto basata esclusivamente sui flussi di cassa futuri che possono essere generati dal progetto stesso, a differenza del caso di corporate finance in cui si considera la solidità patrimoniale e la redditività di tutta l'impresa;
- Controllo e monitoraggio dei processi risultano essere più semplici da intraprendere nella finanza di progetto in quanto le attività risultano essere dedicate. Diversamente in un'azienda i flussi possono intrecciarsi configurando all'interno dei processi attività che servono business differenti;
- Allocazione dei rischi totalmente differente. Nel corporate finance le società possono operare in business differenti, infatti tipicamente presentano dei portafogli diversificati, così il rischio può essere diversificato attraverso le diverse aree di business; inoltre i rischi comuni possono essere gestiti attraverso prodotti assicurativi. Ciò è necessario in quanto i creditori hanno completa rivalsa sugli sponsor. Nella finanza di progetto la situazione è totalmente diversa in quanto i creditori hanno un ricorso limitato o nullo nei confronti degli sponsor, così l'esposizione finanziaria dei creditori è maggiore;
- Struttura del debito che si basa sui soli asset del progetto e sulle sue caratteristiche specifiche, al contrario della corporate finance che valuta l'intero portafoglio aziendale. In questo modo il debito risulta essere sicuro e tagliato su misura per il progetto. Anche grazie

a questa caratteristica il capitale di debito che può essere reperita attraverso project finance è sicuramente estesa.

Ricapitolando, per finanziare un nuovo progetto le soluzioni possibili da intraprendere sono almeno due: la corporate finance e la finanza di progetto. Con la prima alternativa il nuovo progetto viene inserito all'interno di un bilancio che comprende tutte le attività aziendali, di conseguenza gli asset e i cash flow derivanti da business già esistenti possono essere utilizzati come garanzia per ottenere credito aggiuntivo da parte delle istituzioni finanziarie; infatti esiste un solo debito totale a carico dell'azienda e per ripagarlo si utilizzano i cash-flow generati dalla complessività dei business. Con il project finance lo stesso progetto invece è da intendersi fuori bilancio, quindi i creditori non potranno rivalersi sugli altri cash flow e asset degli sponsor partecipanti all'iniziativa. Avendo ricapitolato la sostanziale differenza presente tra le due forme di finanziamento è possibile compararle: in letteratura vengono identificate almeno due motivazioni che fanno preferire il project finance alla tradizionale corporate finance nell'ambito dei progetti infrastrutturali. La prima motivazione riguarda la mitigazione dei problemi di agenzia che hanno origine dalla relazione principale-agente, ovvero del moral hazard. Proprietà e manager in azienda potrebbero avere obiettivi diversi poiché incentivati in modo differenti, possiedono curve di utilità differenti; nel project finance ciò non avviene: l'obiettivo è unico, ovvero la generazione del flusso di cassa massimo possibile, unico elemento di retribuzione finale. La seconda motivazione riguarda la gestione dei rischi, agevolata dalla stesura di un contratto specifico, su misura per l'iniziativa, che quindi può risultare maggiormente completo. Si può quindi affermare che la finanza di progetto, paragonata alla classica corporate finance, permette di sopportare livelli maggiori di leva finanziaria in quanto, attraverso la realizzazione di una nuova società ad hoc, il nuovo progetto risulta essere fuori bilancio per gli sponsor partecipanti e gli asset vengono utilizzati solamente come collaterali; gli sponsor vengono così isolati dagli eventi coinvolti nella realizzazione e gestione del processo stesso. Tali finanziamenti consentono di mantenere bassi i costi, anche se il capitale di debito risulta essere di grandi dimensioni, senza tra l'altro modificare le condizioni economiche e finanziarie dell'azienda. In generale un progetto comporta alti costi di transazione, composti da legali, tecnici e di assicurazione, inoltre tempo per la valutazione del progetto e la seguente stesura dei contratti in cui vengano esposti i termini delle operazioni, ma anche costi di monitoraggio del processo. In un contesto di project finance questi costi riescono però a essere ridotti attraverso la sua rete di contratti specifici e le sue caratteristiche principali.

Per tutte le caratteristiche sopra esposte, il project financing è un istituto di finanziamento che è divenuto di grande rilevanza nella realizzazione delle opere pubbliche in Italia sul finire degli anni '90. Lo scopo era quello di sopperire alla scarsità di risorse finanziarie e recuperare il divario infrastrutturale dell'Italia rispetto ai maggiori Paesi europei, ovvero Germania, Francia e Inghilterra. Il project finance ha così accelerato il processo di liberalizzazione e privatizzazione dei servizi di pubblica utilità, permettendo l'entrata di operatori privati in comparti pubblici, precedentemente esclusi dalle logiche imprenditoriali. Anche Enti e Istituzioni hanno poi intrapreso questa tecnica finanziaria per superare i propri vincoli di bilancio riuscendo a portare avanti i propri programmi capaci di generare effetti positivi nell'ambito macroeconomico. Nella

pratica il soggetto pubblico dà una sorta di concessione di costruzione e gestione a un privato, dove però è il promotore stesso che si propone all'amministrazione per realizzare le opere di cui questa necessita e non viceversa. Quindi è un sistema indiretto di realizzazione delle opere pubbliche, proprio perché lo Stato non interviene in modo diretto e non partecipa all'esecuzione dei lavori. La caratteristica principale di questo strumento è il fatto che dei soggetti privati, detti promotori o sponsor, propongono un progetto per realizzare un'opera pubblica e si impegnano a finanziarlo personalmente, rifacendosi dei costi sostenuti attraverso la gestione dell'opera stessa dato che darà dei ritorni. Le opere da realizzare di cui un'amministrazione necessita sono presenti in un elenco che viene stilato regolarmente dall'Ente pubblico. Il programma contenente le varie opere è da considerarsi come momento attuativo di identificazione e di quantificazione dei propri bisogni da parte delle amministrazioni. Nella programmazione triennale delle opere pubbliche ognuna ha priorità diversa a seconda della natura del progetto stesso, inoltre le opere realizzabili con capitale privato hanno la priorità rispetto a quelle interamente a carico della pubblica amministrazione, in quanto il rischio viene in parte trasferito. I promotori, a differenza che negli appalti di lavori pubblici, possono essere anche imprenditori privati, o enti finanziatori (es. banche, compagnie di assicurazioni).

### **2.2.2. Settori di impiego**

È appena stato esposto come il project finance sia un metodo per la realizzazione di progetti con finanziamenti di larga scala e a capitale intensivo. Risulta così essere utilizzato per progetti infrastrutturali, industriali o di trasporti, e soprattutto per progetti che abbiano una tecnologia conosciuta a cui si possano associare rischi chiari. Alcuni esempi di progetti realizzati finanziandoli con questa tecnologia sono il canale di Suez, l'Eurotunnel nel canale della Manica e, in Italia, l'ospedale di Mestre. I settori che vedono questa metodologia utilizzata per il finanziamento dei progetti godono fondamentalmente di due caratteristiche: in primo luogo si collocano su mercati caratterizzati da contratti a lungo termine e da compratori finanziariamente solidi, inoltre presentano un basso livello di rischio tecnologico. Queste due caratteristiche sono quelle che rendono appetibile per le Amministrazioni Pubbliche l'utilizzo della finanza di progetto, grazie all'entrata di capitali privati, per la realizzazione di opere infrastrutturali. In molti Paesi il successo di questo strumento finanziario è dovuto a fattori come le performance dell'economia domestica, la politica internazionale nell'affrontare la crisi del credito e l'innovazione degli stessi mercati finanziari. L'evoluzione della finanza di progetto in realtà segue principalmente due trend, uno dei quali in disaccordo con le caratteristiche precedentemente citate. Il primo è l'utilizzo dell'esportazione in mercati emergenti supportati da agenzie di credito situate nel paese di origine perché sembra la via più semplice per colmare il gap infrastrutturale; il secondo è l'utilizzo per finanziare progetti con alto rischio di mercato, come ad esempio quei settori in cui i ricavi non riescono a coprire i costi operativi e di investimento e si necessita di conseguenza della garanzia del settore pubblico.

È possibile stilare un elenco dei settori in cui la finanza di progetto è maggiormente utilizzata:

- Energia e ambiente, con particolare interesse verso gli impianti a energia rinnovabile, che sfruttino l'energia del vento, del moto ondoso, del sole, ma anche impianti di termovalorizzatori che convertano rifiuti solidi urbani in energia elettrica;
- Risorse idriche: impianti di captazione, adduzione, trattamento, potabilizzazione e distribuzione delle acque;
- Trasporti: realizzazione di strade, autostrade, ponti e tunnel, aeroporti, interporti e strutture del trasporto pubblico locale;
- Strutture sportive: realizzazione di impianti sportivi quali piscine, centri polivalenti, ecc;
- Strutture culturali e turistici: realizzazione di centri congressi, recupero di immobili storici, costruzione di musei e biblioteche;
- Opere pubbliche varie: realizzazione di opere di edilizia sanitaria, edilizia penitenziaria, edilizia scolastica, impianti ad uso della Pubblica Amministrazione, caserme, parcheggi.

### **2.2.3. Attori coinvolti**

Provando a fare un po' di chiarezza, vengono riassunti quelli che sono gli attori che interagiscono in un finanziamento con project finance:

- Il promotore, ovvero colui che ha l'idea e genera la proposta. Avuta l'idea, il soggetto la sviluppa fino a formulare il progetto che verrà poi proposto. Nella realtà avviene che ci possano essere più promotori contemporaneamente e che questi solitamente siano anche gli azionisti della SPV, ovvero che ognuno fornisca una quota di capitale;
- Lo sviluppatore, che viene scelto all'interno della lista dei promotori, si occupa di realizzare le prime analisi sulla fattibilità economico-finanziaria del progetto. La sua attenzione deve essere rivolta alla pianificazione delle attività, delineando i tempi e i costi di ognuna. Grazie a questa analisi potrà evidenziare eventuali criticità del piano e arrivare alla stesura di un budget per i costi di sviluppo e di realizzazione;
- Il finanziatore è colui che presta al promotore i fondi necessari alla realizzazione del progetto. Il capitale esterno è reperito sia sotto forma di debito (prestiti da banche commerciali), sia sotto forma di capitale azionario (fondi di investimento), sia in forme ibride. Il reperimento delle risorse utili non è un tema così scontato, soprattutto nel periodo seguente la crisi economica e finanziaria;
- Le assicurazioni che forniscono le garanzie finanziarie o contrattuali allo scopo di ridurre alcuni rischi specifici originati dalla natura stessa del progetto;
- Il consulente finanziario è una figura che fa da intermediario tra coloro che forniranno il capitale (di debito o di credito) e il soggetto che utilizzerà queste risorse per la realizzazione dell'opera. Nel conciliare le due posizioni, il consulente finanziario arriverà a strutturare il Modello Economico Finanziario;
- Il consulente tecnico ha invece il compito di assistere il promotore nella parte tecnica della definizione del progetto. Si ricordi che i promotori possono essere di diversa natura, questo comporta che questi possano sviluppare la totalità del progetto, anche senza l'aiuto di

consulenti esterni. Si pensi ad esempio a una banca che voglia realizzare un pezzo di ferrovia: è banale pensare che questa abbia in sé tutte le competenze per sviluppare la parte finanziaria del progetto, ma anche che dovrà rivolgersi a un consulente tecnico per la parte di realizzazione vera e propria dell'opera.

- La società di progetto a cui spetta il compito di realizzare il progetto scelto, nelle modalità e tempistiche definite. Come già visto questa potrà avere forme giuridiche diverse, portando con sé diritti e doveri connessi.
- I fornitori di beni e servizi che si interfacciano con la società di progetto sia in fase di costruzione, soddisfacendo la domanda di acquisti di macchinari o attrezzature che deriva dalla natura capital intensive del progetto, sia di gestione per un'appropriata operatività. Con queste controparti commerciali è bene valutare la possibilità di strutturare contratti a lungo termine con lo scopo di garantire gli standard desiderati di qualità, riducendo allo stesso tempo la volatilità dei prezzi.

## **2.2.4. Classificazione delle opere**

Nel project finance il progetto viene valutato attraverso la capacità di generare flussi di cassa; da una parte questi devono ricoprire i costi sostenuti durante la fase di realizzazione dell'opera, ma devono anche riuscire a remunerare tutte le parti coinvolti. Infatti i cash-flow sono la fonte primaria per il rimborso del capitale di debito e per la remunerazione del capitale di rischio. Le fonti con cui si finanzia un progetto sono principalmente due: capitale di debito (composto prevalentemente da prestiti bancari e obbligazioni) e capitale di rischio (ovvero l'equity, il capitale dei privati che sottoscrivono le azioni). Queste due tipologie di finanziamento sono molto diverse tra loro. Innanzitutto il capitale di debito prevede un rimborso a scadenza, inoltre garantisce dei ritorni sotto forma di pagamento di interessi o di cedole di importo predefinito (solitamente percentuale), sempre con scadenze prestabilite. Al contrario, il capitale di rischio non è caratterizzato da ritorni sicuri o con pagamenti assicurati ad un certo istante temporale. In altre parole, il rimborso e il pagamento degli interessi del capitale di debito ha priorità rispetto alla remunerazione del capitale di rischio. Allora gli azionisti verranno remunerati solo nel caso in cui, una volta ripagati tutti gli stakeholders coinvolti, avanzerà cassa. In realtà neanche in questo caso è obbligatoria la remunerazione, infatti la decisione sul pagamento dei dividendi avviene di anno in anno solo dopo aver valutato la gestione precedente e pianificato quella futura. In pratica gli utili eccedenti potrebbero essere utilizzati per investimenti futuri e quindi non distribuiti agli azionisti; ciò può accadere in parte percentuale o nella totalità della somma. Per quanto detto, la fase di gestione dell'opera risulta di primaria importanza in quanto principale generatrice dei flussi di cassa. Solo se questi saranno adeguati ci sarà convenienza economica a realizzare il progetto, ma anche solo in questo modo sarà possibile attrarre banche e azionisti.

Per quanto riguarda la generazione dei flussi di cassa però bisogna prestare attenzione. Introduciamo una classificazione delle tipologie di opere che avviene solitamente nella prassi operativa:

- Opere calde. Sono progetti in grado di generare intrinsecamente reddito attraverso i ricavi ottenuti da utenza, ovvero dall'uso e dalla disponibilità di un servizio (sia pubblico, sia privato). Il settore privato recupera i costi di investimento, necessari alla realizzazione almeno nell'arco della vita della concessione dell'opera. In questa situazione allora il settore pubblico non ha particolari compiti, infatti si limita a creare le condizioni necessarie per garantire la realizzazione del progetto stesso, facendosi carico delle fasi iniziali di pianificazioni, autorizzazione, indizione di bandi di gara per l'assegnazione delle concessioni, svolgendo le prassi di assistenza per le procedure di autorizzazione. Esempi di opere calde sono i termovalorizzatori, i cimiteri, sistemi per le energie rinnovabili.
- Opere tiepide. I ricavi commerciali sono di per sé insufficienti, allora si necessita di una componente di retribuzione pubblica. Il supporto finanziario può essere sotto forma di integrazione dei ricavi, di contributo finanziario, di cessione di immobili, di valore di subentro. La contribuzione pubblica deve essere tale da costituire (insieme ai cash-flow) dei ritorni adeguati a garantire una convenienza economica per le parti coinvolte. Questo tipo di contribuzione è giustificato dalla generazione di esternalità positive in termini di benefici sociali che il progetto emana nell'arco della sua gestione. Le esternalità positive sono effetti esterni generati dall'attività di consumo o di produzione che influenza la funzione di utilità o di profitto di soggetti diversi da quelli che l'hanno generata. Esempi di queste opere sono gli impianti sportivi, gli acquedotti, le case di riposo.
- Opere fredde. In questo caso vengono generati dall'utenza cash-flow quasi nulli; il concessionario privato allora fornisce direttamente servizi alla Pubblica Amministrazione. La remunerazione che il privato trae è costituita da pagamenti dello Stato, calcolati su base commerciale; questi sono di entità prestabilite e fissi nel tempo detti canoni di disponibilità. Esempi di questa tipologia di opere sono le carceri, gli ospedali, le scuole.

Una seconda classificazione può essere svolta sulle strutture dal punto di vista finanziario, a seconda della tipologia di rivalsa che i soggetti finanziatori hanno sugli azionisti della società di progetto realizzatrice dell'opera:

- Operazioni senza rivalsa, sono operazioni in project finance in cui è esclusa totalmente la rivalsa; in questo caso le banche operano secondo logiche non tradizionali, in quanto si accollano dei rischi simili a quelli imprenditoriali o, in alternativa, sono presenti soggetti terzi che forniscono ulteriori garanzie.
- Operazioni con rivalsa limitata, in cui la rivalsa dei finanziatori sugli azionisti è limitata nel tempo, nell'ammontare o nella qualità.
- Operazioni con rivalsa piena, caso in cui sulla società di progetto cade una rivalsa totale da parte dei finanziatori.

### **2.3. Rischi del project finance**

Uno dei passaggi fondamentali per la costruzione di un'operazione di project finance è l'identificazione dei rischi a cui si va incontro, allo scopo di monitorarli e, dove possibile, mitigarli.

L'entità dei rischi definisce l'esposizione del progetto della variabilità dei suoi risultati, lo scostamento da un valore atteso, sia in modo positivo che negativo. Quello che si vuole fare è capire qual è la probabilità che si verifichino delle conseguenze negative all'interno del business che ne possano modificare i profitti finali. Tutti i progetti sono esposti a dei rischi in quanto cercano di raggiungere degli obiettivi in termini di tempi, costi e performance in un lasso di tempo, caratterizzato da variabilità. Il risk management è inteso come il processo che include tutte le azioni necessari a identificare i rischi per ridurre l'esposizione; è possibile identificare le fasi che lo compongono:

- Planning, momento in cui si definisce l'impostazione del processo;
- Identification, si identificano i rischi che interessano il progetto in oggetto e le loro principali caratteristiche;
- Analysis, dove si stabiliscono le priorità tra i vari rischi, questi vanno studiati in modo tale da stabilirne l'occorrenza e la magnitudo, per definire quali siano i possibili impatti;
- Response planning, momento per definire le azioni necessarie alla mitigazione dei rischi evidenziati, anche in ordine di priorità;
- Monitoring and control, fase molto importante in quanto manifesta il risultato dell'implementazione delle misure stabilite e attuate nelle fasi precedenti.

Nello specifico, all'interno delle operazioni della finanza di progetto sono presenti dei rischi caratteristici. Il risk management è di fondamentale importanza in quanto si ricorda che non è possibile diversificare il rischio attraverso l'attività in business differenti, in quanto la società è creata ad hoc per un unico progetto, e perché i rischi sono divisi tra le parti e non totalmente a carico degli sponsor. Una corretta amministrazione del processo di gestione dei rischi rappresenta quindi un elemento chiave per l'ottimizzazione del rating con cui verranno emesse le obbligazioni con le quali si reperirà ulteriore capitale di debito. Di seguito si riassumono i principali rischi che una SPV che opera con la finanza di progetto deve sopportare.

**Design risk.** Il rischio di design include tutti quei rischi che interessano le carenze nella progettazione o nell'ingegneria del progetto che possono avere effetti sulla generazione di cash-flow. Sono relativi agli eventi come l'approvazione della fattibilità: in seguito a una situazione con rischi troppo elevati è possibile che si intraprendano delle misure correttive che portino al cambiamento del design o degli standard di costruzione. Questa categoria comprende i rischi tecnologici, relativi all'uso di una tecnologia non testata che dalla pratica alla teoria può risultare inapplicabile. Le tecnologie sconosciute richiedono flessibilità, mentre lo scopo della finanza di progetto è di prevedere il più possibile gli avvenimenti futuri per limitare il libero arbitrio dei manager e di evitare l'utilizzo del project finance per scopi diversi.

**Site risk.** Il rischio di sito fa parte della categoria di quei rischi che si riferiscono alla circostanze legate al luogo dove il progetto viene sviluppato. Le questioni principali che lo influenzano sono l'acquisizione dei terreni, la sicurezza del luogo, le caratteristiche geofisiche del suolo, le tematiche ambientali legate all'inquinamento di aria, suolo e acqua, la tradizione culturale del luogo.

**Construction e completion risk.** Il rischio di costruzione e di completamento si assumono essere a carico degli sponsor e dipendono proprio dalla reputazione di quest'ultimi, ma anche dalla natura della tecnologia del progetto. I costi di costruzione, di performance e di ritardo sono gli elementi principali che possono far scostare il risultato dal target prestabilito e che quindi quantificano questo rischio. I costi di costruzione sono fondamentali per la vita finanziaria stessa del progetto poiché proprio su questi si basano tutte le assunzioni finanziarie, quindi i creditori richiedono che siano monitorati perché non possono accettare valori maggiori di quelli anticipati. Per quanto riguarda il ritardo è opportuno ricordare che il ritardo nella costruzione si traduce in un ritardo delle entrate date dalla gestione dell'opera. È importante che questi valori di performance vengano calcolati in modo oggettivo, piuttosto che segnare poi degli scostamenti rispetto a dei valori che sono stati stimati in modo troppo ottimistico inizialmente.

**Force majeure risk.** Il rischio derivante di causa di forza maggiore sono connessi all'interruzione dei lavori a causa di fattori che nessuno può controllare. Si sta facendo riferimento a disastri ambientali, terrorismo, guerre, condizioni metereologiche, incendi. Questi potenziali eventi influenzano diversi aspetti del progetto come i costi lungo il ciclo di vita, delle operations e della manutenzione, l'amministrazione del budget, e così via. È importante evidenziare come in casi di ritardi nella realizzazione dell'opera, si riduce il tempo di ripagamento dell'investimento, di conseguenza diminuisce la probabilità di riuscire a recuperare i costi per poi avere un ritorno economico. Questa considerazione sui tempi dovrebbe portare gli sponsor ad attuare delle stime iniziali plausibili, senza dover poi apportare delle significative correzioni in un secondo momento.

**Revenues risk.** I rischi riguardanti le entrate devono la loro entità dalla disponibilità delle strutture di progetto, dai volumi di mercato e dalle possibili sottoperformance. Le fluttuazioni della domanda possono variare più o meno drasticamente i cash-flow della fase operativa dell'opera. In particolare alti costi fissi implicano un alto rischio, infatti una poca flessibilità nel prezzo finale implica una più o meno elevata variazione della domanda, in relazione all'elasticità di quest'ultima proprio ai prezzi. Per mitigare questo tipo di rischio è possibile stendere contratti con controparti commerciali che si impegnino all'acquisto di una data quantità di prodotto a un prezzo prestabilito. Il volume di domanda, oltre che dal prezzo, può essere influenzato dall'andamento dei prezzi dei prodotti sostitutivi.

**Operating and maintenance risk.** Il rischio operativo e di manutenzione prende in considerazione eventi che potrebbero influenzare il ciclo di vita dei costi. Per esempio ci possono essere costi maggiorati di manutenzione se il volume di produzione aumenta considerevolmente. La gestione operativa condiziona particolarmente il modello finanziario a cui sono particolarmente attenti i creditori che vogliono essere protetti dalle possibili oscillazioni dei cash-flow. I rischi operativi riguardano anche i rischi di produzione, come quelli di design dell'output.

**Performance risk.** Le variabili che influenzano il rischio di performance sono da riferirsi ai costi da sostenersi lungo il ciclo di vita, le performance del progetto, i costi di manutenzione e di operativi. Gli eventi che danno l'entità del rischio sono ad esempio una prematura obsolescenza, la



disponibilità del materiale e della manodopera, reclami da parti terze, questioni di compliance, ecc.

**External risk.** I rischi di natura esterna sono dati dalla instabilità del contesto esterno in cui si sviluppa il progetto. L'ambiente esterno è caratterizzato da un contesto politico, culturale e sociale. Il contesto politico è fonte di instabilità dal punto di vista legislativo, sulle norme di operatività aziendale o sulla tassazione. Cambiamenti di tipo legislativo possono modificare il business della società di progetto, modificando la redditività dello stesso; ma anche il contesto sociale può modificare il business poiché la società si interfaccia con parti terze, come la comunità locale.

**Financial risk.** Con il termine rischio finanziario si vogliono comprendere molte tipologie di rischi, che hanno la caratteristica comune di essere associati alla gestione finanziaria. L'entità della loro magnitudo dipende da vari fattori, quali le fluttuazioni e le variazioni dello spread e degli indici economici, l'inflazione, ma anche possibili operazioni di rifinanziamento che modifichino le condizioni contrattuali precedenti. Il rischio finanziario è presente lungo tutto il ciclo di vita del progetto, dalla fase di iniziativa imprenditoriale a quella di gestione finale. Come esempio si possono portare i seguenti rischi: interest rate risk, Exchange rate risk, inflation risk, funding risk. Il primo è relativo alle possibili variazioni dei tassi di interesse e va considerato con molta attenzione in quanto il progetto si riferisce a un arco di tempo abbastanza lungo da poter vedere diverse variazioni. Il secondo fa anch'esso riferimento alla variazione dei tassi, non più a quelli di interesse, ma a quelli di cambio; questo rischio è presente quando si opera in Paesi diversi da quello di origine, o anche se il progetto si sviluppa su nella sua costruzione e gestione all'esterno dei confini di una sola Nazione. Il terzo riguarda la variazione dei prezzi e dei costi: con l'aumentare dei costi, ad esempio delle materie prime, si potrebbe avere alla fine un profitto minore, dei cash-flow non positivi come quelli preventivati, elemento base della valutazione del progetto. L'ultimo rischio è connesso alla possibilità di non avere il capitale iniziale necessario, con la conseguenza di doverlo reperire sul mercato; il risultato della trattativa dipende anche dalle capacità degli sponsor di relazionarsi ai possessori di capitali.

**Default risk.** Questo rischio prende in considerazione la possibilità che la società di progetto non sia capace di ripagare il debito o, in generale, le controparti fornitrici di risorse finanziarie. Per determinare l'importanza di questo rischio è necessario ricordare che l'iniziativa è fuori bilancio, deve quindi vivere dei soli proventi che riesce a generare. In accordo inoltre con i principi di Basilea.

**Strategic risk.** Il rischio strategico è proprio del business in cui si sviluppa il progetto. Bisogna provare a determinare quali siano i possibili scenari futuri, le prospettive di guadagno e le possibili perdite di utili derivanti da decisioni di business errate. Questo rischio può aumentare con il cambio di proprietà della società di progetto o a fronte di forti conflitti di interesse tra shareholders and stakeholders.

## 2.4. Bancabilità

Nello svolgere l'analisi sulle caratteristiche e le modalità in cui si struttura il project finance si è fatto spesso riferimento alla redditività. Precedentemente è stato affermato che la redditività è una delle voci che consente di valutare la bontà di un progetto, infatti i redditi prodotti nella fase di gestione saranno quei flussi di cassa che permetteranno di coprire i costi sostenuti, rimborsare e ripagare il capitale di debito e remunerare il capitale di rischio. Ciò è vero e di fondamentale importanza, ma in realtà il concetto è più ampio: non è possibile far riferimento solo alla redditività, bisogna considerare anche la bancabilità. Per "bancabilità" del progetto si intende la capacità del progetto di generare, durante la fase di gestione, flussi di cassa positivi in grado di remunerare in modo congruo sia gli azionisti della società di progetto mediante pagamento di dividendi, sia le banche finanziatrici mediante corresponsione di quota capitale e quota interesse sui finanziamenti erogati. Ma anche alla capacità di generare degli elementi interni al progetto che rendono un titolo (assegno, cambiale, ecc.) valido per l'incasso, utilizzabile per lo sconto o per altre forme di accredito presso le banche. I requisiti di cui il progetto necessita per avere bancabilità sono le garanzie che questo possiede, quindi oltre al grado di certezza dei ritorni futuri (come già detto) ci sono delle garanzie collaterali quali l'interesse commerciale o il rating delle controparti aderenti al progetto, insomma elementi che permettono di generare entrate tramite la vendita delle strutture (non nel caso del PF) o attraverso l'uso dei servizi prodotti da parte di terzi. Oltre ai titoli, anche le imprese possono essere o meno bancabili: ciò determina se avranno o meno la possibilità di scontare ad esempio delle fatture presso un istituto di credito. Solo un'adeguata bancabilità permette all'impresa di svolgere in modo efficiente il suo lavoro: l'elemento chiave è la monetizzazione delle fatture, degli assegni, insomma di quegli strumenti che possono diventare banconota liquida e di conseguenza essere rimpiegati nello svolgimento delle operazioni quotidiane. Se invece un'impresa è in difficoltà finanziarie, o nel caso peggiore in liquidazione o in fallimento, allora il suo grado di bancabilità risulta essere fortemente a rischio; ciò significa che non potrà vedere tramutati i suoi titoli in contanti, rischiando così di rimanere paralizzata. Il lavoro della società di progetto è quindi strettamente collegato al rapporto con le banche finanziatrici. Queste vogliono poter verificare l'equilibrio economico e finanziario dell'investimento che viene denominato Piano Economico e Finanziario (PEF). In questo documento è contenuta l'esplicitazione dettagliata dei dati di ingresso che determinano l'equilibrio economico e finanziario del progetto. Solitamente il PEF include i prospetti di conto economico, stato patrimoniale e flusso di cassa del progetto per tutta la durata dell'operazione di project finance; i dati sono quindi da considerarsi previsionali. Per equilibrio economico si intende che i flussi di cassa attualizzati derivanti dai ricavi ottenuti a seguito dell'applicazione di determinate tariffe, siano almeno sufficienti a coprire i costi necessari alla realizzazione del progetto e alla gestione dello stesso in fase di erogazione di prodotti e servizi. L'equilibrio finanziario invece si raggiunge quando quegli stessi flussi di cassa riescono a coprire anche il ripagamento degli esborsi monetari in un arco di tempo determinato, preso in esame dal valutatore.

Come espresso finora, risulta di notevole importanza analizzare quali siano i flussi di cassa del progetto lungo tutto il suo ciclo di vita. L'analisi dei cash-flow deve permettere di rappresentare il profilo finanziario di un investimento, in modo da poter scegliere le modalità di finanziamento che maggiormente si adatti alla situazione in analisi. Per determinare la convenienza dell'investimento bisogna essere in grado di calcolare i flussi di cassa che costituiscono il margine e per far ciò il passo iniziale è quello di calcolare la differenza monetaria tra ricavi e costi, a questa somma andranno poi sottratti altri costi operativi di natura finanziaria, in seguito verrà sottratto anche l'ammontare della somma necessaria al pagamento delle imposte e delle tasse sugli utili di progetto. Il calcolo dei flussi di cassa deve essere svolto in riferimento ad un periodo di tempo predeterminato. Nella prima fase, quella di costruzione, i flussi di cassa saranno sicuramente negativi, in quanto i ricavi risultano essere nulli e alla mancanza di entrate di cassa si associano anche gli esborsi necessari alla realizzazione dell'opera. Questo periodo caratterizzato da un cash-flow totale negativo deve trovare un'adeguata copertura da parte delle banche commerciali che forniscono capitale di debito e degli sponsor che forniscono l'equity. Terminata la fase di realizzazione, l'opera è pronta per l'erogazione dei prodotti e dei servizi: la fase operativa di produzione genera i ricavi destinati a coprire i costi sostenuti. Quindi la curva dei ricavi inverte il suo andamento, andando verso flussi di cassa sempre più positivi. Il tempo in cui i ricavi riescono a coprire i costi è detto pay back time e risulta essere tanto più breve quanto più i ricavi risultano essere consistenti. Superato il punto di pay back l'azienda acquisisce le risorse necessarie al ripagamento dei finanziamenti contratti; ciò avviene in fase di conclusione del progetto. Estinti i debiti finanziari e in seguito possibile i costi operativi e infine alla remunerazione degli sponsor, che hanno aderito con equity, attraverso il pagamento dei dividendi.

## 2.5. Il ruolo delle banche

Le banche sono un attore molto importante nella realizzazione di un progetto in project finance. Queste infatti si interfacciano con la società sin dall'inizio dell'iniziativa e permettono la realizzazione dell'iniziativa; inoltre intervengono nel progetto con quote di capitali ingenti, allora il prestito potrebbe risultare molto complesso. Per questi motivi i ruoli da ricoprire all'interno di un finanziamento in project finance sono numerosi, soprattutto quando a causa di quote davvero elevate più istituti di credito devono collaborare. I ruoli principali ricoperti sono: l'asservatore, il financial advisor, l'arranger; poi ci sono anche altri ruoli all'interno del sindacato di prestito. Si vogliono affrontare i principali ruoli in modo più dettagliato.

**Osservatore.** La norma non specifica direttamente i contenuti di un'attività di asservazione, l'Autorità di Vigilanza è quindi dovuta intervenire per dettare delle linee guida. L'attività viene definita come un accertamento che deve essere effettuato dall'istituto di credito sulla proposta, in luogo dell'amministrazione, con fini pubblicistici. In altre parole deve verificare la coerenza del piano economico e finanziario precedentemente determinato in riferimento ad alcuni elementi prestabiliti, quali:

- Prezzo massimo che l'amministrazione può corrispondere;

- Prezzo minimo che il concessionario deve corrispondere allo scopo della costituzione o del trasferimento dei diritti;
- Eventuale canone;
- Percentuale di lavori da appaltare a terzi in modo obbligatorio;
- Tempistiche di esecuzione dei lavori per l'avvio della gestione;
- Livello minimo di qualità dei prodotti o dei servizi erogati;
- Tariffe da applicare, possibilità di modifiche delle stesse e livello massimo.

In una modifica del testo i contenuti dell'attività sono stati ridotti, affermando che l'elenco iniziale non aveva una natura assoluta, ma che potesse variare a seconda dei casi e delle loro caratteristiche. Secondo questa nuova interpretazione, l'ambito dell'asseverazione può essere anche concordato tra l'istituto di credito ed il promotore sempre che sia accertata la coerenza degli elementi tipici del piano del punto di vista economico-finanziario. Da segnalare inoltre che l'asservatore non ha l'obbligo di contrarre il finanziamento del progetto per il quale sta valutando il piano economico e finanziario, al contrario è normativamente distinto il momento asservativo e il finanziamento stesso.

**Financial advisor.** Nella fase iniziale, di sviluppo dell'idea imprenditoriale, è necessaria una figura di consulenza bancaria per poter formulare l'offerta finanziaria. L'advisor ha il compito di preparare uno studio di fattibilità per verificare la bancabilità del progetto, in caso in cui questo venga finanziato attraverso project finance. L'analisi deve essere svolta sulle variabili tecniche, economiche, fiscali e finanziarie. In altre parole sono proprio le banche a dover stabilire se un progetto è in grado o meno di generare flussi di cassa tali da rendere il progetto vantaggioso, ovvero se sono in grado da coprire il valore dell'investimento sostenuto. Il financial advisor viene remunerato attraverso due componenti: la prima è quella fissa e pagata periodicamente per l'attività di analisi e di redazione di un documento; la seconda è un premio da corrispondere qualora il ruolo di arranger sia stato affidato a un'altra banca.

**Arranger.** Il financial advisor svolge solitamente anche l'attività di arranging, che consiste nell'ottenere il permesso a strutturare e gestire il contratto di finanziamento da parte della SPV. L'arranger allora ha il compito di prendere contatto con altre banche per reperire tutte le risorse necessarie alla realizzazione del progetto; in seguito svolge anche il ruolo di coordinatore tra tutti gli istituti coinvolti. Essendo presenti diversi istituti le commissioni sono organizzati secondo la seguente struttura: l'arrangement fee è la commissione riconosciuta dagli sponsor alla banca arranger a fronte del finanziamento concesso; l'underwriting fee invece è la commissione che la banca arranger corrisponde alle altre banche che sono presenti nel pool di finanziamento, con quote in proporzione all'ammontare del debito che hanno emesso.

Ci sono poi altri ruoli all'interno di un prestito sindacato, quali il lead manager, il manager e il co-manager; tutte queste hanno in comune la caratteristica di prendere parte al finanziamento erogando una parte del debito. La caratteristica invece che le contraddistingue è rappresentata proprio dall'importo di partecipazione.

## **2.6. Italia**

### **2.6.1. Normativa**

Questa nuova disciplina della finanza di progetto, tradotta poi in nuovi strumenti finanziari come i project bond che verranno trattati più avanti, viene introdotta in Italia per la prima volta l'11 novembre 1998 con la legge n. 415 (cd. legge Merloni-ter). Questa legge nasce con un duplice obiettivo: da una parte si vuole frenare la spesa pubblica senza rinunciare alla realizzazione di opere necessarie e fondamentali per lo sviluppo socio-economico del Paese, dall'altra si vuole agevolare le imprese fornendogli una modalità alternativa di finanziamento per la realizzazione di opere, rendendo non particolarmente vincolante il capitale proprio iniziale. Questo settore è sicuramente diverso dagli altri: contraddistinto da forte burocrazia e ritorni di cassa lenti, non riesce ad attrarre capitale privato, rendendo l'impiego di quest'ultimo solo parziale. Con il PF si cerca proprio di aumentare il numero di opere realizzate senza capitale pubblico, o almeno con una percentuale minore. Nel settore pubblico la legge prevedeva che la costruzione e la gestione dell'opera fosse data in concessione a favore di un soggetto che si occuperà dell'intera vita dell'opera sul suolo di proprietà pubblica. In altre parole in cambio del terreno e degli utili generati dall'opera nella fase di gestione, il soggetto privato si accolla le spese di realizzazione. La norma prevedeva anche un contributo alla realizzazione da parte dell'Amministrazione stessa nel caso in cui la realizzazione dell'opera fosse stata particolarmente onerosa.

Alla legge n. 415/1998 ha fatto seguito la legge del 1 agosto 2002 n. 166 (cd. legge Merloni-quater). In questa ridefinizione della legge viene ampliato il numero dei potenziali soggetti promotori aumentando le tipologie che possono aderirvi. Furono così incluse le Camere di Commercio e le Fondazioni Bancarie. Inoltre viene apportata una modifica in merito al limite temporale della durata per il quale un'opera così realizzata rimanga in concessione del privato; in particolare il limite temporale viene abolito.

Un'altra rivisitazione della legge è avvenuta in seguito alle osservazioni formulate dalla Commissione Europea nel 2004 nella procedura d'infrazione n. 2001/2182. Suddette osservazioni facevano riferimento a delle difformità tra la normativa italiana e quella europea nell'ambito degli appalti pubblici. La legge venne così nuovamente modificata nel 2005: le modifiche riguardavano principalmente l'avviso pubblico con cui le amministrazioni committenti indicano quali opere possono essere realizzate attraverso capitale privato.

Infine è stato redatto il Codice dei contratti pubblici (decreto legislativo n. 163 del 2006, in vigore dal 1 luglio 2006) che ha abrogato tutte le leggi precedenti. Nei suoi articoli il Codice contiene la disciplina italiana del project finance e raccoglie tutte le norme della contrattazione pubblica. La finanza di progetto ha in seguito subito una riforma con il decreto legislativo n. 152 del 2008 in cui si è cercato di mettere in relazione questa disciplina con la crisi economica in modo da cercare una via per la ripresa. In realtà la riforma non presenta un procedimento generale per operare in regime di finanza di progetto, definisce però le caratteristiche che questi progetti devono

presentare per essere definiti tale: strumento per la realizzazione delle opere pubbliche, regolato però principalmente dall'autonomia privata.

In sostanza si può notare come anche in questo caso, ogni Stato avesse iniziato un suo percorso in questo ambito in cui, ovviamente, apparivano divergenze rispetto agli altri Paesi della comunità europea; quest'ultima in seguito è intervenuta a uniformare la questione dando delle linee guida che devono essere recepite a livello nazionale.

## **2.6.2. Il contesto**

Nell'impostazione classica di questa disciplina è presente una ripartizione del rischio corretta tra i soggetti che portano capitale di rischio o equity (ovvero i promotori) e quelli che forniscono capitale di debito o prestiti obbligazionari (le banche). Questa caratteristica è proprio uno dei cardini del project finance. In realtà in Italia ciò non avviene: il rischio è prevalentemente a carico del promotore. Infatti chi fornisce capitale di debito risulta essere coperto da eventuali circostanze negative, protetto da una serie di rischi di cui ancora si fa carico il promotore in modo iniquo. Questo motivo potrebbe bastare a spiegare lo scarso successo che il project finance nel nostro Paese, nonostante la corposa normativa presente a riguardo. La non condivisione in modo opportuno del rischio fa sì che le aziende siano riluttanti a intraprendere questo tipo di progetto. La tematica non è banale, infatti soprattutto in un contesto come quello attuale, risulta di fondamentale importanza la gestione del rischio, che in un certo senso significa anche condivisione, cosa che nel nostro Paese non accade.

La finanza di progetto, e in generale i Partenariati Pubblico Privato (PPP), però rimangono oggetto di attenzione sempre crescente in Italia e a livello europeo. Approfondendo in modo sintetico il PPP è possibile definirlo come una vasta gamma di modelli dove il settore pubblico collabora con quello privato, ma non risulta essere definito né a livello nazionale, né a livello comunitario. Le forme di collaborazione sono tutte quelle in cui una Pubblica Amministrazione affida ad un soggetto privato l'attuazione di un progetto con l'obiettivo di realizzare delle opere pubbliche o che siano di pubblica utilità, affidandone anche la gestione dei relativi servizi. Tornando alla crescente attenzione che si delinea nei confronti del project finance è possibile affermare che ciò accade perché rimangono persistenti le difficoltà dell'Amministrazione Pubblica nell'ambito finanziario e in contemporanea aumentano i vincoli di bilancio imposti dall'Unione Monetaria. Quindi dato che non si può prescindere dall'utilizzo dei capitali privati nella gestione pubblica, si necessita di modalità di finanziamento diverse, più efficienti, che siano migliori della situazione in cui sia lo Stato in modo diretto a dover recuperare le risorse; solo così è possibile conciliare l'obiettivo di risanamento del bilancio con quello di colmare il divario di infrastrutture presente in relazione agli altri Stati dell'Unione Europea.

### 2.6.3. Project finance e settore pubblico

Quando il project finance è utilizzato per la realizzazione di progetti desiderati dall'Amministrazione Pubblica, il percorso di realizzazione percorre un iter distintivo che viene esposto nel seguito. Per arrivare all'obiettivo finale di stendere un contratto tra l'amministrazione e il promotore sono necessarie 3 fasi:

**Fase preliminare.** Con la revisione annuale della programmazione triennale, le amministrazioni devono pubblicare un avviso in cui vengano elencate quali opere pubbliche programmate possano essere realizzabili attraverso l'impiego totale o parziale di capitali privati, ovvero a carico dei promotori. Una volta che un soggetto promotore abbia individuato nell'elenco triennale di una determinata amministrazione un'opera di suo interesse, può presentare all'amministrazione stessa entro il 30 giugno dell'anno in corso la sua proposta alla realizzazione di lavori pubblici o di pubblica utilità, tramite contratti di concessione. La proposta deve essere caratterizzata da elementi distintivi, ovvero deve contenere:

- uno studio di inquadramento territoriale e ambientale della zona dove si vuole realizzare l'opera;
- uno studio di fattibilità in cui indicare anche le finalità, gli obiettivi e le modalità di esercizio del servizio;
- il progetto preliminare;
- la bozza di convenzione del contratto che regola i rapporti tra il promotore e la pubblica amministrazione (es. piano tariffario, manutenzione dell'opera ecc.);
- piano economico-finanziario, in cui viene evidenziato il tempo necessario affinché l'opera inizi a dare dei ritorni positivi, ovvero quando questa diventerà remunerativa. In altre parole si vuole calcolare il break-even point, ossia il punto di pareggio: prima di questo istante gli utili servono a ripagare la costruzione dell'opera, quindi non possono essere distribuiti per remunerare le parti, cosa che invece può avvenire dopo questo istante temporale;
- spese sostenute per formulare la proposta, infatti nella proposta il promotore chiede e individua nel caso in cui non sarà poi l'aggiudicatario le spese che ha sostenuto per poter partecipare al bando. Queste non possono essere superiori al 2,5% dell'importo totale dell'investimento.

L'Amministrazione ha poi quattro mesi di tempo per effettuare una valutazione di fattibilità della proposta, tenendo conto della qualità dell'opera, dei tempi di realizzazione e delle tariffe che applicherà il soggetto aggiudicatario per la gestione dell'opera stessa, ma anche dell'interesse pubblico che questa riveste per la comunità. È prevista la possibilità di operare delle modifiche da concordare tra promotore e Amministrazione con l'obiettivo di mantenere l'equilibrio economico-finanziario della programmazione triennale. Nel caso in cui per la stessa opera siano state presentate più domande, si procederà inizialmente con una valutazione comparativa (ovvero la proposta migliore verrà individuata attraverso un bando di gara) e in seguito si svolgerà la

valutazione totale solo della proposta migliore. Una volta che la proposta è stata approvata bisogna procedere ad individuare chi effettivamente realizzerà e gestirà l'opera. Questo avviene appunto attraverso la fase di gara.

**Fase di gara.** Questa fase si compone di due momenti diversi: la gara pubblica e la procedura negoziata. Durante la gara pubblica l'amministrazione ha lo scopo di individuare due soggetti che concorreranno insieme al promotore all'aggiudicazione, che avverrà poi attraverso la procedura negoziata. I soggetti che competeranno con il promotore vengono scelti esaminando tutte le offerte pervenute e ponendole a confronto con quella del promotore stesso, che fa da base d'asta. La gara viene aggiudicata secondo il criterio dell'offerta economicamente più vantaggiosa. Alla gara non partecipa il promotore, il quale parteciperà solo alla procedura negoziata. A seguito della gara pubblica ha luogo la procedura negoziata. Gli attori di questa fase sono i promotori e i soggetti risultati migliori nella gara precedente. Lo scopo di questa negoziazione è l'individuazione dell'offerta migliore dal punto di vista economico. Durante questa fase i soggetti sopra citati si confrontano tra di loro attraverso una procedura del tutto informale, per stabilire chi offre le condizioni migliori all'Amministrazione. Bisogna però precisare che il promotore gode di una sorta di diritto di prelazione, rispetto agli altri competitori, sull'aggiudicazione della gara (legge n. 166/2002). Quindi può accadere che il promotore adegui la sua proposta a quella del soggetto vincitore della negoziazione, ritenuta più conveniente dall'Amministrazione. In questo modo il promotore si aggiudicherà ugualmente il project financing. In realtà questo diritto è stato abrogato nel 2007 con una modifica apportata al d.lgs. 163/2006.

Non sempre hanno luogo entrambe le fasi di gara; questo dipende dal numero di soggetti interessanti alla realizzazione di una data opera, ma anche dalla ammissibilità delle loro offerte. Nel caso non venissero presentate proposte e la gara pubblica vada deserta, non si necessiterà della procedura negoziata e l'aggiudicatario il promotore, il quale è vincolato a tutti i termini della sua proposta. Invece, nel caso in cui ci fossero solo uno o due partecipanti alla gara pubblica, questi verranno automaticamente ammessi alla procedura negoziata, una volta verificata la completezza e la qualità delle loro offerte.

A questo punto risulta necessario stendere un contratto tra le parti, ovvero l'aggiudicatario e l'Amministrazione Pubblica. Il contratto di concessione è di costruzione e gestione; la bozza di convenzione è l'offerta che il promotore ha presentato in sede di presentazione alla proposta. Nel contratto devono essere riportate la modalità di costruzione dell'opera e la gestione della stessa; l'opera in oggetto potrà essere sfruttata per un tempo predeterminato proprio in questa fase di contrattazione. Bisogna porre attenzione a determinare in maniera adeguata questa variabile, perché determinante nel calcolo dei cash flow attesi e di conseguenza della remunerazione delle parti coinvolte. In questo documento sono anche descritte le tempistiche con cui l'aggiudicatario della gara dovrà presentare alla controparte (l'Amministrazione) le tempistiche con cui si concluderanno la presentazione del progetto finale e del progetto esecutivo, nonché i termini di realizzazione dell'opera.



**Fase di costruzione e gestione.** Un'attenta analisi va fatta su ciò che concerne la realizzazione e la gestione dell'opera; infatti in queste fasi non troviamo le stesse dinamiche di un tradizionale appalto. Vediamole nel dettaglio. Nel project finance la direzione dei lavori è totalmente a carico del soggetto aggiudicatario, a differenza di ciò che avviene nei contratti di appalto. Qui l'amministrazione, infatti, non è responsabile dei lavori ma ha solo un ruolo di vigilanza che gli permette di verificare in ogni momento l'avanzamento dei lavori e la conformità dell'opera al progetto stesso. Se da un lato l'amministrazione può fare vigilanza, dell'altro è nell'interesse stesso dell'aggiudicatario rispettare modalità e termini di realizzazione per arrivare il prima possibile a poter sfruttare l'opera in oggetto, perché è da questa fase che arriveranno le entrate di cassa a fronte di tutti gli esborsi sostenuti. La costruzione della società di progetto in realtà avviene durante il bando di gara per l'assegnazione di un project financing, dove l'aggiudicatario potrà scegliere di costruire una società in forma di società per azioni, a responsabilità limitata, ma anche consortile. Durante la fase di realizzazione e di gestione la società di progetto così creata si sostituirà in toto all'aggiudicatario nei rapporti con l'Amministrazione, infatti diventa a tutti gli effetti concessionaria. La società di progetto, al di fuori di meccanismi di approvazione o autorizzazione, diventa a tutti gli effetti concessionaria, subentrando nel rapporto di concessione dell'aggiudicatario. Passando alla fase di gestione, si può notare come anche qui le modalità siano state definite a priori in fase contrattuale, tra queste sono comprese anche le tariffe da applicare all'utenza, ma anche i modi e termini per effettuare eventuali rialzi delle stesse. Deve essere previsto anche un piano di manutenzione dell'opera da utilizzare durante tutta la gestione. È necessario che anche questo punto sia presente nel contratto tra le controparti, infatti alla termine della concessione la struttura verrà riconsegnata all'amministrazione, che quindi deve essere tutelata contro una gestione che renda l'opera obsoleta e inutilizzabile. Un'altra caratteristica nel project finance è il pagamento di un canone annuale di gestione al proprietario dell'opera, che viene calcolato come percentuale sul fatturato ottenuto. Nella maggior parte dei casi la proprietà dell'opera realizzata è dell'ente pubblico, ma questo non sempre avviene. Infatti la normativa prevede la separazione tra gestione (società di progetto) e proprietà (Amministrazione Pubblica); quest'ultima però potrebbe prendere la decisione di privatizzarla attraverso la vendita, anche allo stesso concessionario che ne detiene la gestione. Risulta di fondamentale importanza determinare con attenzione la durata della concessione per la gestione in modo da riuscire a remunerare il capitale privato, a pagare i canoni di gestione annuale allo Stato e garantire una certa redditività. Il tempo di concessione viene definito utilizzando come istante iniziale la data di conclusione dei lavori prevista nel progetto. Una volta terminato il periodo di concessione, si conclude il periodo di gestione dell'aggiudicatario che dovrà restituire l'opera all'Amministrazione Pubblica, o in ogni caso al suo proprietario. Quest'ultimo allora potrà decidere come procedere, in alternativa può assumere lui stesso la gestione diretta dell'opera, indire un'altra gara d'appalto per rinnovare la concessione a dei soggetti privati, oppure vendere la proprietà al concessionario attuale.

## 2.7. Vantaggi

Dopo aver illustrato quali sono le caratteristiche del project finance, è possibile passare all'analisi di quali possano essere i vantaggi che vengono generati si procederà all'analisi con riferimento ai diversi soggetti coinvolti. In particolare si vedranno quali sono i vantaggi per la Pubblica Amministrazione e quali per i soggetti concessionari.

I vantaggi per l'Amministrazione si possono riassumere in due macro categorie, ovvero la quantità e la qualità delle opere realizzate.

**Quantità.** Come già detto, le opere di infrastruttura sono necessarie per lo sviluppo sociale ed economico di un Paese perché permettono lo svolgimento di numerose attività attraverso la distribuzione di servizi primari, con alta accessibilità e basso costo. Il discorso risulta essere valido sia da un punto di vista economico (pensando alle aziende come fruitori di tali servizi), sia da un punto di vista sociale (pensando ai cittadini). Essendo opere di così grande importanza, l'Amministrazione non può prescindere dalla loro realizzazione, tenendo presente che maggiori saranno le infrastrutture, maggiori saranno le possibilità di sviluppo; non sempre però le risorse necessarie allo svolgimento dei lavori sono di facile reperimento. Il project finance permette di collegare questi due aspetti: le Amministrazioni possono realizzare opere di notevole interesse per la collettività limitando l'impatto sul bilancio pubblico e senza assumersi il rischio finanziario o di mercato. Questi infatti vengono trasferiti sui privati che si occuperanno della realizzazione fisica del progetto. In questo modo si amplia il numero di opere che verranno costruite, perché si è potuto accrescere il numero di soggetti che possono realizzarle.

**Qualità.** C'è però un altro vantaggio fondamentale, ovvero la qualità dei lavori svolti. Intuitivamente è possibile comprendere che ciò accade solo pensando al fatto che un mercato aumenta la sua efficienza con il crescere dei soggetti partecipanti (si pensi ai mercati concorrenziali). Infatti, nella fase di gara i vari soggetti competono per aggiudicarsi la concessione, che verrà affidata a quel soggetto che avrà presentato l'offerta più vantaggiosa. Ma non solo questo garantisce una maggiore qualità. La redditività della società di progetto dipende dalle modalità e dai tempi di realizzazione, allora questa sarà incentivata a svolgere i lavori nel miglior modo possibile e nel rispetto delle tempistiche; l'obiettivo è quello di massimizzare l'utile, lavorando su entrambi gli aspetti: minimizzazione dei costi attraverso un'attenta gestione e massimizzazione dei ricavi attraverso i ricavi da utenza. È evidente che prima si completa l'opera, prima inizieranno ad arrivare i ritorni per gli investitori.

Anche per il privato però i vantaggi sono particolarmente rilevanti. Innanzitutto bisogna ricordare che il progetto si presenta sempre come entità autonoma, quindi non viene incorporato nel bilancio societario. Ciò permette di poter svolgere dei progetti limitando l'impatto sul bilancio nel caso di fallimento; tematica fondamentale in quanto potrebbe portare ad una modifica del rating aziendale e di conseguenza modificare quelle che sono le sue condizioni di accesso al credito, in termini di capitale ricevuto e di costo del denaro. Essendo parte esterna al bilancio, i finanziamenti per questi progetti non vanno a modificare la leva finanziaria globale dell'impresa, evitando il

peggioramento degli indici di indebitamento. Il tema è rilevante in quanto per la realizzazione di tali progetti le somme di debito di richiesto non sono trascurabili rispetto all'indebitamento totale d'impresa. In questo modo si ha accesso a nuovi finanziamenti con condizioni di mercato, ragionevoli. Indebitarsi così tanto significa portare la leva finanziaria a livelli molto alti (fino a giungere al 70-90% sui mezzi propri). Si è prima detto che gli indici di bilancio non peggiorano perché l'indebitamento è fuori bilancio, ma inoltre si possono godere di molte agevolazioni fiscali essendo gli interessi passivi deducibili dal reddito. Un ultimo ma non meno rilevante vantaggio è dato dalla possibilità di creare società ad hoc con altri soggetti: ciò permette di mettere in comune le risorse, siano esse di natura finanziaria o di natura conoscitiva con l'obiettivo di realizzare progetti più innovativi, più rischiosi, ma anche più redditizi. Le possibilità di collaborazione sono molto ampie, infatti possono interessare sia soggetti pubblici che privati; inoltre il numero di imprese coinvolte non è prestabilito.

## **2.8. Svantaggi**

Dopo aver analizzato i vantaggi di questa disciplina, è opportuno soffermarsi sugli svantaggi che presenta. Ciò risulta essere indispensabile per poter avvicinarsi e gestire al meglio un progetto di project finance.

Innanzitutto bisogna notare che le fasi da svolgere necessitano di tempo, per questo motivo si avrà un allungamento dei tempi di avvio dei lavori per la realizzazione dell'opera. Per poter burocratizzare il sistema le fasi vanno svolte con determinate modalità, spostando ovviamente in avanti la data di inizio lavori rispetto al caso in cui sia l'ente stesso a presiedere la realizzazione. Si può inoltre osservare che maggiore sarà il numero dei promotori, maggiore sarà il tempo necessario per valutare le proposte; stesso discorso vale per la fase di gara pubblica perché i ogni soggetto si presenterà con offerta diversa e la comparazione di queste richiederà uno grosso sforzo riconducibile a un numero maggiore di ore per arrivare ai vincitori. Tutti questi allungamenti potrebbero quindi sommarsi a una procedura che già di per sé non risulta essere snella; ciò ricordando che le tempistiche hanno un impatto economico non banale sul progetto e sui ritorni che esso può produrre. Oltre ad essere lunga, la procedura risulta essere molto strutturata; questa burocratizzazione infatti necessita di costi per poter funzionare, preferibilmente in modo efficiente. Il discorso svolto sui tempi e sui costi sembra fornire motivazioni valide alla non attuazione di questo tipo di progetto, bisogna però ricordare che questi possono essere recuperati più che abbondantemente con un corretto esercizio della società di progetto. Da ricordare inoltre che i soggetti partecipanti al progetto sono molti e di natura diverse, questo fa sì che i rischi siano di difficile identificazione e, di conseguenza, non sarà semplice poterli allocare. Infine bisogna notare che la struttura uscente dalla fase di contrattazione risulta essere molto rigida, quindi potrebbero verificarsi perdite di tempo e di denaro a fronte della realizzazione di una modifica. D'altronde il contratto di concessione deve contenere al suo interno tutte le caratteristiche con cui si realizzerà e si gestirà l'opera; proprio perché si ha un gran numero di parti interessate, nulla deve essere lasciato alla discrezionalità di uno o dell'altro soggetto. Ciò però viene pagato con

l'allungamento dei tempi, ad esempio nell'attesa dell'approvazione di una modifica, e la perdita di soldi, dovuto al rallentamento dei lavori o alle opere di correzione necessarie.

## **2.9. Conclusioni**

È possibile dunque riassumere affermando che il project finance è un finanziamento di lungo termine, in cui la garanzia è rappresentata dai flussi di cassa che il progetto genera durante la fase di gestione. Il project finance per le sue caratteristiche è stato particolarmente utilizzato dal settore pubblico per realizzare opere che altrimenti rimarrebbero solo dei progetti in quanto troppo costose. Nella realizzazione dell'opera prevede il coinvolgimento (totale o parziale) di soggetti privati che devono trovare autonomamente le risorse finanziarie necessarie allo svolgimento dei lavori, sia impegnando capitale proprio, sia ottenendo capitale di debito. L'iniziativa di questo nuovo strumento finanziario è avvenuta in seguito alla necessità di colmare un gap infrastrutturale senza dover utilizzare delle risorse finanziarie proprie; il piano permette di creare opere indispensabili allo sviluppo del Paese impiegando le risorse disponibili sul mercato dei capitali, senza dimenticare che la proprietà dell'opera in oggetto rimane della Pubblica Amministrazione. Essendo un'iniziativa aperta a vari tipi di soggetti, i quali possono presentare offerte per realizzare l'opera di loro interesse, il mercato risulta essere molto ampio. Un mercato ampio è un mercato aperto dove maggiore sarà il numero di competitori e di potenziali entranti, maggiore sarà l'efficienza. La competizione influenza l'efficienza del mercato stesso, in particolare si otterrà che la realizzazione venga effettuata con un livello dei costi sicuramente inferiore di quello che si avrebbe in monopolio, portando ad un conseguente aumento degli utili.

## **3. PROJECT BONDS**

Gli sponsor che si apprestano a realizzare opere in project finance hanno la necessità di reperire risorse sul mercato, oltre al loro capitale e ai finanziamenti concessi. Il finanziamento di un'impresa è fondamentalmente di due tipi: capitale di debito e capitale di rischio. Il capitale di rischio è il capitale proprio degli azionisti, dei proprietari della società stessa, ovvero degli sponsor in questo caso. Il capitale di debito invece è di soggetti terzi che durante il periodo di esistenza del prestito si vedranno pagati degli interessi e, a scadenza, avranno il rimborso del totale valore nominale. Tra le forme di capitale di debito si trovano i titoli obbligazionari.

### **3.1. Caratteristiche**

#### **3.1.1. Definizione**

Nel caso specifico della società operanti in project finance, le obbligazioni emesse sono proprio i project bonds. Questi strumenti finanziari vengono emessi esclusivamente da soggetti privati coinvolti nella realizzazione di un progetto che solitamente riguarda opere infrastrutturali (in particolare nel settore dei trasporti, dell'energia e della banda larga). I soggetti emittitori possono essere le società di progetto e le società titolari di concessioni, autorizzazioni o contratti. L'obiettivo della nascita di queste particolari obbligazioni è proprio quello di coinvolgere più capitali privati possibili nelle opere pubbliche senza utilizzare le classiche fonti di finanziamento, quali i bilanci statali e il credito bancario, che oggi non hanno la capacità di assicurare la copertura di tutte le risorse finanziarie necessarie. Questi bonds hanno in particolare due caratteristiche: sono strumenti standardizzati e liquidi. Le caratteristiche di standardizzazione e di liquidità li rendono molto appetibili sul mercato, anche grazie alla possibilità di poterli acquistare e vendere con facilità. In Italia, per realizzare le opere di utilità pubblica, i project bonds sono in questo momento l'unico strumento davvero appetibile in quanto le risorse per gli investimenti sono stati drasticamente ridotti e questi bonds sono a impatto zero sul debito pubblico.

I project bonds presentano una novità, ovvero la capacità di contribuire al sostentamento della fase più rischiosa dell'intero progetto, ovvero l'avvio; infatti è proprio questa fase che si necessita di molto capitale per poter dare inizio ai lavori e al contempo l'opera non genera alcun flusso di cassa positivo. Ciò avviene in contrasto a quello che accade per gli strumenti tradizionali che raccolgono liquidità per lo sviluppo di infrastrutture già in esercizio e con i flussi di cassa generati dall'utilizzo (come pedaggi, canoni, ecc) con cui ripagano le obbligazioni. In sostanza mentre con gli strumenti tradizionali si è in grado di finanziare solamente i progetti brownfield, con i project bonds è possibile dare credito anche alla realizzazione dei greenfield, inoltre è possibile sbloccare progetti fermi o terminare opere incompiute, a causa della mancanza dei fondi liquidi. Nello specifico sono definiti brownfield tutti quei siti in cui le costruzioni sono risalenti al passato e non

vengono utilizzate totalmente o solo parzialmente nel presente, a seguito del degradamento o dell'abbandono della gestione. Per greenfield invece si intendono aree libere da costruzioni o vincoli preesistenti in cui i progettisti hanno la libertà di prendere scelte tecniche e impiantistiche, ovviamente rimanendo in accordo con le normative vigenti.

Un'altra complicanza alla reperibilità del capitale nella fase di avvio è che, oltre alla sicurezza di avere inizialmente dei cash flow negativi, nella fase di realizzazione esiste ancora incertezza in riferimento alla presenza di una serie di rischi, quali possono essere di progettazione, di costruzione, relative alle tempistiche (al rispetto delle scadenze), di ritardo magari dovute a iter burocratici troppo lenti (presenti sicuramente in Italia), di abbandono, di variazioni nelle voci a budget come costi non previsti o di inflazione e tassi di cambio (se si opera con controparti fuori dall'Unione Monetaria), ma anche di variazione della normativa che possono cambiare le condizioni dell'intero contesto. È quindi possibile notare come il profilo di rischio di questi progetti risulti essere abbastanza elevato: per questo motivo i project bonds sono riservati esclusivamente a investitori qualificati, che abbiano le conoscenze necessarie a valutare la situazione, delineando il corretto profilo di rischio del progetto nel suo insieme. In questo modo verranno finanziati prima i progetti migliori, quelli che hanno maggiori possibilità di riuscita e di generazione futura di cash flow, ovvero quelli che risultano essere economicamente e finanziariamente più attraenti. Solo dopo, eventualmente, verranno finanziati gli altri. La presenza di addetti sul mercato fa sì che venga stilata una sorta di graduatoria della bontà dei progetti, ciò permette l'allocazione efficiente delle risorse.

Il direttore generale della Cassa Depositi e Prestiti da una parte riconosce la validità dello strumento, ma risulta essere cauto sull'esprimersi a riguardo della loro applicazione in Italia. Il project finance potrebbe incontrare difficoltà nei progetti greenfield poiché caratterizzati da eccessiva burocratizzazione in riferimento alla fase iniziale, ovvero quella riguardante la pianificazione e la progettazione antecedente la messa in cantiere dell'intervento. In Europa invece la legislazione è meno vincolante, presentando un iter meno dispendioso: per questo motivo si potrebbe avere un utilizzo dei project bond più ampio.

### **3.1.2. Garanzie**

Di conseguenza a quanto scritto finora, quindi in riferimento alla pericolosità dei suddetti strumenti, sono state sviluppate specifiche garanzie: i wrap. Queste sono state introdotte dal decreto del 7 agosto 2012. In sostanza, a seguito della valutazione del progetto e del soggetto realizzatore, vengono rilasciate da vari enti, quali banche, intermediari finanziari, imprese di assicurazioni autorizzate, Cassa Depositi e Prestiti e Bei. Da una parte si necessita della valutazione del soggetto: si vuole determinare il suo rischio di credito, ovvero il rischio che non sia in grado di ripagare il suo debito; il rischio di credito è connesso sia al pagamento degli interessi, sia al rimborso a scadenza del credito. Dall'altra bisogna porre attenzione all'opera stessa: è necessario comprendere quale sia la sostenibilità economico-finanziaria degli investimenti impiegati, ovvero quanto fruttano in termini di rendimento le risorse proprie (questa informazione è possibile

dedurla partendo dal calcolo del ROE), ma bisogna anche tenere conto della redditività ipotizzabile dell'opera stessa, quindi della sua capacità di generare flussi positivi di cassa attraverso l'utilizzo. Tutti questi giudizi devono essere il più oggettivi possibili, infatti anche piccoli errori di previsioni iniziali possono portare a grandi scostamenti nel futuro dando luogo anche a grandi perdite. Le garanzie rivestono un ruolo molto importante nella diffusione di questi strumenti perché di fatto ne modificano l'attrattività. La presenza di forte garanzie infatti avrà un ruolo positivo sugli investitori che saranno così maggiormente propensi ad acquistare questi strumenti finanziari, preferendo tra tutti quelli più sicuri. Il mercato è quindi condizionato da questi segnali. È possibile affermare che, oltre ad essere una tutela per i soggetti sottoscrittori, le garanzie siano un premio di merito rilasciate da soggetti terzi ai quali viene trasferito in carico il rischio operativo, a quei progetti che presentano un grado elevato di sostenibilità degli investimenti.

Come ogni cosa, le garanzie presentano vantaggi e svantaggi. Lo svantaggio è dato dal fatto che queste costituiscono un costo non indifferente per gli emittenti delle stesse e andando a incidere in modo negativo sulla redditività degli strumenti a cui vengono legate; i ritorni saranno quindi meno cospicui e di conseguenza si ridurrà il guadagno. Il vantaggio però è altrettanto significativo, infatti abbassando il profilo di rischio, il numero dei potenziali investitori aumenta ampliando significativamente il mercato e la sua domanda. Il risultato finale è quello di avere un accesso più ampio ai capitali privati, accompagnato da una riduzione del rischio. Tutto ciò risulta essere in accordo con le linee guida per un'uscita meno lenta dalla crisi attuale. Si vuole tuttavia ridurre l'onerosità economica degli strumenti per chi vi vuole accedere, rendendo le garanzie meno rigide; allo stato attuale, secondo il decreto sopra citato, le garanzie devono essere "esplicite, irrevocabili, incondizionate e stipulate in forma scritta". Ciò può essere attuato perché sussiste una condizione necessaria: gli investitori che hanno la possibilità di sottoscrivere questi strumenti sono solo quelli istituzionali, che quindi non richiedono una particolare protezione. Questa possibile implementazione contiene però al suo interno una forte minaccia: se si dovesse cedere alla tentazione di aprire al mercato retail l'offerta di tali strumenti per aumentarne la diffusione, la non capacità di alcuni soggetti non qualificati di valutare correttamente i profili di rischio potrebbe indurre a fenomeni speculativi che porterebbero al finanziamento di progetti meno sicuri ma con redditività maggiore, a scapito dei progetti migliori. Questa possibile evenienza andrebbe a ripetere quello che è precedentemente successo con altri strumenti finanziari; il problema è di forte rilevanza poiché i comportamenti degli investitori si vanno a riflettere sull'economia reale. Un modo corretto di ampliare maggiormente la domanda, favorendo la diffusione dei project bonds potrebbe essere quello di aumentare il periodo temporale soggetto ad agevolazioni fiscali, che ad oggi sono previste solo per quei bonds emessi entro i tre mesi dall'entrata in vigore del decreto. In ogni caso l'eliminazione di precedenti condizioni restrittive ha reso più agevole l'utilizzo.

## **3.2. Rilancio dell'economia**

Per le considerazioni finora fatte, i project bonds possono essere utili alla ripresa dell'economia e alla conseguente uscita dalla crisi in modo concreto. Realizzare le infrastrutture significa da una

parte creare nell'immediato dei posti di lavoro, la creazione di flussi di cassa prossimi, ma anche una base futura su cui appoggiare l'economia del Paese. Questo potrà avvenire se gli operatori nel mercato riusciranno a coglierne le potenzialità valutandoli in alternativa ad altri strumenti finanziari e se comprenderanno la possibilità di diversificare il proprio portafoglio acquistando questi strumenti diversi di medio/lungo periodo con una remunerazione sufficientemente garantita. Ovviamente ci saranno soluzioni positive in relazione ai soggetti che se ne avvarranno per la realizzazione dei loro progetti; infatti solo se quest'ultimi saranno validi, sostenibili e remunerativi, gli investitori vorranno possederli all'interno del loro portafoglio.

Le infrastrutture necessarie sono sia a livello nazionale sia a livello europeo. Non si può negare come l'Italia necessiti al più presto di infrastrutture avanzate che facciano da scheletro allo sviluppo generale dell'economia: l'obiettivo è quello di fornire alle aziende operanti sul territorio una serie di servizi ai quali deve risultare facile accedervi con dei costi contenuti, provando anche ad attirare imprese estere che siano interessate a estendersi su nuovi territori. Non bisogna però dimenticare quanto la costruzione di infrastrutture sia complessa proprio da un punto di vista progettuale; questo avviene proprio in seguito alla natura del territorio italiano, ricco di diversità morfologiche al suo interno. Anche a livello europeo però bisogna fare ancora tanto. Quello che si denota ancora oggi è che a fianco di un'unione dal punto di vista monetario, non si appoggia un'unione economica. Se è vero che una delle problematiche è l'assenza di una normativa unica a livello superiore, anche le infrastrutture a livello sovranazionale non sono molte. Infine, per quanto riguarda le garanzie, è necessario sottolineare una criticità derivante dalla natura dei soggetti che le emettono. In precedenza si sono fatti alcuni esempi di soggetti in grado di emettere le suddette garanzie, tra questi la Cassa Depositi e Prestiti e la Sace. Entrambi sono soggetti riconducibili allo Stato. Se da una parte garantiscono maggiore affidabilità rispetto a quelle dei privati, dall'altra se venissero attivate bisognerebbe fare ricorso a quei fondi pubblici che non si volevano toccare per non coinvolgere il debito pubblico. Il project finance nasce con l'esigenza di realizzare progetti senza utilizzare risorse finanziarie pubbliche, che però con un eventuale intervento di CDP e Sace si andrebbe ad usufruirne come una sorta di sussidio pubblico implicito. Ci si troverebbe così in un paradosso: è necessario quindi un generale ripensamento delle procedure amministrative, che in Italia risultano essere caratterizzate dall'assenza di tempi e regole certe per la fase di esecuzione.

### **3.3. Situazione europea**

Come precedentemente accennato, questi strumenti finanziari saranno rilasciati a livello europeo. Un ruolo fondamentale è stato assunto dalla Bei, la banca europea per gli investimenti. Questa dovrebbe farsi carico della parte delle obbligazioni con il rischio più elevato, in modo da alzare il rating di questi strumenti e di conseguenza abbassarne il costo. La vendita sarà poi effettuata a investitori istituzionali e privati. Per comprendere meglio cosa siano i project bonds è possibile metterli a confronto con gli eurobond: quest'ultimi non riguardano progetti specifici ma mettono in comune il debito complessivo dell'eurozona, al contrario dei primi che vengono emessi per il finanziamento di progetti infrastrutturali.



All'incirca un anno fa a Bruxelles è stato trovato un accordo politico che ha visto la nascita di una fase pilota, atta all'introduzione di questi nuovi strumenti nel mercato finanziario europeo. L'iniziativa era in cantiere da almeno un anno e può essere considerata, con una visione ottimistica, come il primo passo di un bilancio comune. La fase pilota prevede di coprire il biennio 2012-2013 e l'utilizzo di 230 milioni di euro provenienti dal bilancio comunitario. L'obiettivo è quello di creare un effetto moltiplicatore che porti ad un investimento totale, che coinvolgerà attori privati, pari a 4,5 miliardi di euro circa. Ai soggetti investitori si è già fatto riferimento in precedenza e sono fondi pensione, compagnie assicurative e altre grandi investitori istituzionali. Queste risorse sono destinate alla realizzazione di progetti nel settore dei trasporti, dell'energia e della banda larga, soprattutto di interesse europeo, infatti molto spesso i Governi investono in quelle che sono le reti interne, ma per uno sviluppo dell'economia comunitaria bisogna ancora fare tanto a livello sovranazionale: la maggiore integrazione è ritenuta uno strumento indispensabile per uscire dalla crisi del debito. Allora vennero individuati alcuni progetti che sposavano proprio questa causa, ma la vera incognita e scopo della fase pilota era verificare che questi strumenti avrebbero effettivamente attirato l'attenzione degli investitori presenti sul mercato.

### **3.4. Funzionamento**

In Italia il Governo Monti ha definito tre regole per il funzionamento dei project bond. Queste sono state introdotte con il Decreto Sviluppo con lo scopo di favorire la nascita e la diffusione di questi nuovi strumenti finanziari, coinvolgendo i capitali privati nella realizzazione di infrastrutture. Entrando nello specifico:

- aliquota sugli interessi attivi pari al 12,5% come per i titoli di Stato, applicabile però esclusivamente ai project bond emessi nei tre anni successivi al 26 agosto 2012, data dell'entrata in vigore del decreto (il regime fiscale agevolato si intende poi applicabile alla totalità del prestito obbligazionario); il regime sarà applicato anche a società di progetto già esistenti che utilizzeranno i project bond come rifinanziamento del debito;
- deducibilità degli interessi passivi per le società progetto come per quelle quotate, ovvero gli oneri fiscali verranno sottratti dal reddito complessivo ottenendo una riduzione della base imponibile su cui verrà calcolata l'imposta attraverso l'aliquota;
- imposta di registro in misura fissa; l'imposta è dovuta al conseguente registro degli atti giuridici presso l'agenzia delle entrate e la sua misura risulta in questo caso essere di importo fisso piuttosto che proporzionale al valore dell'oggetto in atto.

### **3.5. Italia e Europa a confronto**

Le obbligazioni italiane possono essere garantite dal sistema finanziario stesso, da fondazioni e da fondi privati; ciò risulta possibile sin dall'avvio della gestione dell'infrastruttura da parte del concessionario, senza che ci siano oneri per il debito pubblico. Proprio grazie alla garanzia di enti

pubblici (come la CDP) possono assicurare costi ridotti, analoghi a quelli sostenuti per i bond europei. In Europa si vogliono utilizzare gli stanziamenti di bilancio per finanziare i progetti in modo da rafforzare il rating del credito: un miglioramento del credito dovrebbe portare al coinvolgimento di capitali da altre fonti. Per i bond europei l'organo a fare da garante è la Bei, responsabile dell'utilizzo delle risorse UE, che risulta essere in grado di abbassare notevolmente il costo di emissione.

Sia a livello nazionale, sia a livello internazionale, questi titoli vengono emessi direttamente dalle società di progetto costituite per la realizzazione dell'opera stessa. Il finanziamento è quindi diretto: gli investitori forniscono capitale di debito alle società che lo remunereranno e lo rimborseranno a scadenza.

Il settore energetico è quello che in Italia sembra essere maggiormente adatto all'utilizzo di questi strumenti, in conseguenza del grande lavoro svolto dall'Authority di settore. L'aspettativa è che ciò che avvenuto nell'energia possa avvenire anche nel settore dei trasporti. L'iniziativa italiana si è sviluppata in parallelo a quella europea, infatti il project bond italiano può integrarsi con l'obbligazione europea promossa dalla Bei. Nella pratica la banca per gli investimenti potrebbe accollarsi il 20% del rischio totale di progetto, lasciando il restante 80% del finanziamento ai capitali privati dei singoli Stati: nulla sembra vietare che questo 80% venga finanziato attraverso project bond nazionali emessi secondo le varie normative interne vigenti.

In Italia i primi project bond saranno messi sul mercato dalla Cassa Depositi e Prestiti, in realtà non in modo diretto, ma a sostegno di alcune iniziative presenti all'interno dell'istituto stesso. Le operazioni prescelte sembrerebbero essere state individuate tra le richieste di finanziamento per infrastrutture presentate precedentemente alla Cassa. La scelta è stata presa a seconda delle caratteristiche del progetto e della sua adattabilità al finanziamento attraverso project bond piuttosto che con i prestiti tradizionali. Del Fante già precedentemente infatti aveva annunciato che la CDP stava lavorando "sia come soggetto facilitatore delle prime emissioni, che come possibile sottoscrittore". I progetti prescelti fanno parte del settore energetico e di quello autostradale; in riferimento a quest'ultimo sembra siano stati preferiti progetti già iniziati (i cosiddetti brownfield) rispetto a quelli che devono partire da zero. In realtà per poter dare il via alla vera e propria emissione mancano alcuni passi. Innanzitutto la Cassa vorrebbe coinvolgere la Bei allo scopo di aumentare il rating di questi nuovi titoli per renderli più appetibili agli investitori e più economici per le società che avranno l'esigenza di ricorrere a questo tipo di finanziamento. Inoltre è al lavoro con le principali banche del sistema italiano per il collocamento e l'eventuale sottoscrizione di una parte di queste nuove emissioni obbligazionarie in arrivo.

### **3.6. Le banche**

Se l'obiettivo dei project bond è quello di finanziare le imprese nella realizzazione di opere infrastrutturali, anche le banche hanno dall'altra un vantaggio. Le banche sono molto attive sul project financing e vorrebbero spostare queste competenze sui project bond. Questo

trasferimento permetterebbe alle banche di ridurre la loro esposizione nel settore, infatti attraverso il rifinanziamento dei progetti attraverso queste obbligazioni si libererebbero risorse da poter essere indirizzate su nuovi progetti. Il punto di vista della Bei risulta essere più in linea con l'obiettivo centrale dei project bond, ovvero mobilitare le risorse necessarie alla realizzazione delle opere in oggetto, coinvolgendo i capitali privati presenti sul mercato. Si è stimato un fabbisogno tra il 2014 e il 2020 solo per le reti europei vicino ai 1000 miliardi di euro.

## 4. MODELLI SEM

I modelli di equazioni strutturali sono una metodologia statistica utilizzata in diversi ambiti di ricerca; in modo particolare nelle scienze sociali, comportamentali ed educative, ma anche da economisti e ricercatori di marketing. In altre parole i modelli SEM forniscono ai ricercatori un approccio globale che permette la verifica e la quantificazione di teorie sostantive. La loro applicazione risulta essere così vasta perché consentono di modellare un'ipotesi causale, sia in modo confermativo, sia esplorativo. La loro formulazione deriva dall'unione di altre due metodologie statistiche: l'analisi fattoriale e la regressione lineare. L'analisi fattoriale è una tecnica che ha lo scopo di spiegare un fenomeno, una variabile, attraverso una serie di fattori che ne indicano il comportamento; in altre parole attraverso delle variabili osservabili si vuole osservare una o più non osservabili. La regressione lineare invece si propone di prevedere il comportamento di una variabile dipendente da altre variabili indipendenti; la relazione si può trovare analizzando i dati passati che mostrano il legame di dipendenza.

### 4.1. Quando utilizzare SEM

Questi modelli prevedono un duplice utilizzo, a seconda dello scopo e degli obiettivi da ricercare. La modellazione può essere di tipo confermativo o esplorativo; entrambe vengono esposte in seguito.

L'utilizzo risulta essere di tipo confermativo quando alla base sussiste un'ipotesi di comportamento delle variabili. In altri termini si inizia con uno studio dell'ambito in oggetto per delineare quali siano i risultati attesi della variabile che si vuole osservare, in relazione ad altre che vengono considerate influenti. Quindi stabilita la variabile di interesse, sul cui comportamento si vuole prestare attenzione, si cerca di determinare se esistano delle variabili che ne possano spiegare in modo diretto o indiretto il mutare del valore. Facendo un esempio pratico si potrebbe ipotizzare che l'intelligenza degli studenti universitari possa essere implicitamente misurata attraverso la media dei voti degli esami sostenuti e il numero di crediti acquisiti in un dato intervallo temporale: a medie maggiori dovrebbe corrispondere quozienti intellettivi maggiori. In questo caso allora i modelli vengono utilizzati per rappresentare la conoscenza (ipotetica) sui fenomeni studiati; vengono applicati solitamente a teorie già esistenti o in ogni caso già proposte che spiegano il comportamento di alcuni fenomeni. Una volta che la teoria è stata ipotizzata e che sono state individuate le variabili interessate nel fenomeno in oggetto, si possono raccogliere i dati sui quali verrà utilizzato il modello. Quindi i SEM utilizzati in queste situazioni servono a testare la bontà della teoria sviluppata e per questo la modalità viene chiamata di conferma.

Una seconda modalità di utilizzo è invece volta allo sviluppo della teoria stessa. Al contrario del caso precedente, viene preso un insieme di dati corrispondenti a diverse variabili e su questo insieme vengono ripetute diverse applicazioni dei SEM. L'obiettivo è quello di esplorare le

potenziali relazioni in essere tra le variabili coinvolte nello studio. Questa modalità di procedere prevede che alla base non vi sia nessuna teoria disponibile, o comunque che questa sia ancora in forma iniziale, grezza, e di conseguenza ancora sotto inchiesta. In seguito alle capacità di chiarificazione o di sviluppo di teoria, la modalità di utilizzo dei SEM in questo ambito viene definita come esplorativa. Quando si utilizzano dei modelli in questa modalità bisogna prestare molta attenzione: il set (campione) di dati viene preso all'interno di una popolazione studiata. Allora i dati che si otterranno dal modello vanno letti con grande cautela; infatti solo una volta che i risultati saranno replicati su un altro campione della stessa popolazione, potranno essere considerati maggiormente affidabili. La preoccupazione deriva dal fatto che il modello viene fittato su un campione dato, allora sicuramente sarà valido per quel set, ma non è dato sapere se rimarrà valido all'interno di tutta la popolazione; è possibile che i risultati ottenuti da molte ripetizioni di SEM sullo stesso campione, possano aver accumulato dei fattori casuali che hanno portato ai risultati stessi. Ciò impedisce di poter estendere le relazioni trovate dal campione alla popolazione, ovvero non è possibile generalizzare i risultati oltre il campione prescelto.

## 4.2. Caratteristiche

I modelli SEM presentano delle caratteristiche che vengono ora discusse.

I costrutti ipotetici interessati in questi modelli sono concepiti in termini non direttamente misurabili. Le variabili, facenti parte di questi costrutti, che non possono essere misurate vengono nominate variabili latenti. Infatti a queste non è associabile un valore numerico, ma è possibile individuare alcuni fattori sottostanti che ne determinano il valore implicito in modo causale, definendo così il costrutto. Esempi di variabili latenti sono l'intelligenza, la motivazione, l'abilità di scrittura o di lettura, ma anche i comportamenti socio-economici e di conseguenza le decisioni che i soggetti prendono nel contesto economico e finanziario.

I modelli tengono solitamente conto delle eventuali inesattezze che possono essere presenti nella misura di tutte le variabili osservate, in particolar modo nelle variabili indipendenti (i predittori e le esplicative). La presa in considerazione dell'imprecisione delle misure è possibile attraverso l'inserimento di una misura di errore. Le misure infatti sono per loro natura fallibili: la misura è il risultato di un processo (la misurazione) che risulta essere variabile a seconda delle caratteristiche del processo stesso. Se si fa riferimento ad esempio alla misurazione di un intervallo di tempo che intercorre tra due eventi successivi misurato da un operatore con cronometro, risulta evidente che i valori che si avranno saranno diversi a seconda dell'operatore incaricato, ma anche della precisione dello strumento di misura, ovvero il cronometro. Allora si comprende come il termine di errore sia necessario all'interno della stima. A riguardo dei termini di errore, il parametro da stimare è la varianza e ciò avviene quando il modello viene fittato; il loro valore atteso sarà infatti sempre uguale a zero se questi sono effettivamente casuali. Con il fit dei dati si stimano tutti i parametri, sia quelli di relazione tra le variabili, sia le varianze degli errori. Sugli errori è possibile anche effettuare test di ipotesi quando questi rappresentano affermazioni sostanzialmente significative sulle variabili di errore o le loro relazioni con altri parametri. Ciò distingue i SEM dai

modelli classici dove l'errore preso in considerazione è solamente costituito dalla varianza della variabile dipendente. Infine, i modelli SEM sono di solito fittati su matrici di correlazione tra le variabili osservate. La correlazione è una grandezza che prevede di spiegare come in una coppia di variabili ci possano essere delle tendenze di un dato comportamento, ma non è indice di un vero rapporto di causa-effetto. Ad esempio un correlazione di  $-1$  tra due variabili indica che a seconda del valore assunto da una prima variabile, per la seconda ci si aspetta un valore uguale e opposto (ciò può essere approssimato graficamente con una retta con coefficiente angolare pari a  $-1$ ). I modelli possono essere utilizzati per gestire molte variabili contemporaneamente. Sono modelli di grandi dimensioni, descritti da diverse equazioni. Di conseguenza il numero di parametri da stimare risulterà elevato.

Nell'esecuzione dei modelli SEM esistono dei passi da seguire. Si elencheranno di seguito le fasi che verranno poi descritte con maggior dettaglio in seguito. Il metodo econometrico suggerisce un approccio ai modelli costituito da:

- specificazione del modello,
- stima dei parametri,
- valutazione del fit,
- interpretazione, modifica e conclusioni.

La trattazione che segue fa riferimento principalmente a Hoyle (1995) e a Kline (2010).

### 4.3. Specificazione del modello

Si vuole iniziare la trattazione definendo cosa sia il modello e cosa significhi specificazione. È possibile definire un modello come un'affermazione statistica che dichiara una relazione esistente tra variabili; può essere di differenti forme e applicato in diversi contesti. La specificazione di un modello è l'esercizio dello stendere formalmente le equazioni che lo descrivono e risulta di particolare importanza nell'approccio ai modelli SEM. Infatti, nessuna analisi può essere svolta senza aver ipotizzato un modello di relazione tra le variabili coinvolte nello studio. In questa fase è compresa la formulazione del set di parametri, fondamentali per determinare il tipo di relazioni esistenti (assunti in SEM come costanti). Un'altra importante considerazione che deve essere fatta in questa fase è l'identificazione. Questo aspetto riguarda la corrispondenza tra l'informazione da stimare, ovvero i parametri, e l'informazione su cui verranno stimati, ovvero la varianza e la covarianza osservate.

Entrando nello specifico si necessita di introdurre degli elementi di classificazione. Come detto in precedenza una delle più importanti caratteristiche dei modelli SEM è quella di poter lavorare con variabili latenti, oltre a quelle osservabili. Si vogliono quindi evidenziare le differenze che presentano queste due tipologie di variabili per poterle trattare nel modo corretto. In seguito si illustrerà un metodo grafico per la rappresentazione delle variabili, delle relazioni tra loro esistenti e dei parametri.

### 4.3.1. Procedure di specificazione

Un modello di equazioni strutturali lineari viene ipotizzato come uno schema di relazioni lineari esistenti in un gruppo di variabili. L'intento di molti modelli è quello di fare previsione in modo significativo e allo stesso tempo parsimonioso all'interno di un campione di variabili misurabili per osservarne le relazioni presenti. Bisogna tener presente che nel migliore dei casi i modelli riescono a ottenere una approssimazione molto buona della realtà, che però non è esattamente il valore effettivo. Nei modelli SEM i fenomeni vengono rappresentati con relazioni semplici, che poi si convertono in equazioni lineari, ma la loro vera natura risulta essere sicuramente più complessa. Tenendo conto di questa approssimazione è comunque possibile ottenere dei buoni modelli capaci di interpretare in modo abbastanza preciso la realtà oggetto di studio. Quindi è possibile ottenere un output del modello che sia abbastanza vicino ai dati osservati, che danno luogo a una buona prestazione nell'obiettivo di interpretare la realtà se compresa nel modo corretto. In sostanza è possibile solo concludere se un modello sia o meno plausibile, però possono esistere sempre altri modelli ugualmente buoni, o addirittura migliori, per spiegare la situazione oggetto di studio. L'insieme dei possibili modelli alternativi è molto vasto; quindi il modello realizzato risulta essere solamente uno dei tanti modelli plausibili possibili, capaci di rappresentare la struttura prodotta dai dati osservati.

Il set di dati necessari nella specificazione di un modello SEM può essere diviso in due categorie di variabili: variabili osservate e variabili latenti; per la loro definizione si legga più avanti. Date le variabili, il modello è definito attraverso un insieme di relazioni lineari tra le variabili coinvolte. Queste relazioni possono essere di due tipologie: direzionali e non direzionali. Le prime vogliono indicare che una variabile possa essere direttamente influenzata da un'altra; le seconde indicano solamente che esiste correlazione tra due variabili senza saper indicare quale possa essere il segno della relazione stessa. Essenzialmente specificare un modello consiste proprio nell'individuare le variabili di interesse e le relazioni esistenti tra queste. Sicuramente non è necessario identificare relazioni per ogni coppia di variabili, infatti si suppone che molte coppie di variabili non siano associabili da alcun tipo di effetto diretto e che quindi non abbiano alcun tipo di relazione. Le relazioni direzionali andranno poi ad assumere un valore numerico, manifestato dal valore del coefficiente stimato. Nel caso di relazioni direzionali il valore potrà essere interpretato come avviene nei modelli di regressione lineare, dove i coefficienti risultano essere i pesi da applicare ad ogni variabile indipendente per poter spiegare l'andamento di una variabile dipendente; per quanto riguarda invece relazioni non direzionali il numero è indicatore della covarianza esistente tra due variabili (o correlazione se queste risultano essere standardizzate). Questi valori numerici di cui si è discusso sono i parametri del modello, la cui stima risulta essere l'obiettivo principale nelle applicazioni dei modelli SEM. È necessario ricordare però un'altra distinzione tra tipologie di variabili: endogene ed esogene. Sono variabili endogene quelle che ricevono almeno una freccia direzionale, che sono quindi influenzate direttamente da un'altra variabile presente nel modello; è possibile che queste esercitino a loro volta delle influenze dirette su altre variabili, ma non è necessario. Una variabile esogena, al contrario, non riceve alcuna freccia direzionale da nessuna delle altre variabili nel sistema che si sta considerando; quindi queste variabili possono essere

coinvolte in relazioni non direzionali oppure essere collegate ad altre variabili in modo non direzionale. È importante specificare che le variabili endogene non devono essere viste come perfettamente spiegabili attraverso le variabili che le influenzano direttamente, infatti queste sono influenzate anche da un termine di errore che non può essere considerato utilizzando esclusivamente le relazioni direzionali presenti nel sistema. Questo errore può essere scomposto a sua volta in due componenti, quella random e quella causale. Una parte dell'errore non può essere spiegata dai dati del modello e quindi risulta essere non prevedibile, quindi random; esiste però una parte dell'errore che risulta essere sistematica proprio perché parte del modello stesso. Si potrebbe affermare quindi che l'errore sia una variabile latente, non essendo una variabile osservabile e esogena, perché non influenzata direttamente da alcuna variabile.

In generale un modello prevede la formulazione di equazioni lineari tra set di variabili latenti, ognuna misurata attraverso una serie di indicatori. Le relazioni che legano le variabili latenti ai loro indicatori sono solitamente di tipo direzionale: la freccia parte dalla variabile latente e arriva agli indicatori. I parametri associati a questo tipo di relazione sono equivalenti ai fattori dell'analisi fattoriale: coefficienti regressivi che rappresentano con relazioni lineari l'influenza dei vari fattori rispetto a una variabile da misurare. La correlazione osservata tra gli indicatori è spiegata proprio dalla dipendenza di questi da una stessa variabile a cui sottendono, ovvero quella latente che si vuole misurare, e cresce all'aumentare di questa dipendenza. Più i fattori sono legati alla variabile latente, maggiore sarà la loro correlazione. Da notare che gli indicatori delle variabili latenti sono variabili endogene in quanto subiscono l'influenza di un'altra variabile nel sistema, infatti ricevono una freccia direzionale in ingresso. Infine anche gli indicatori sono soggetti a un termine di errore che può essere considerato in modo analogo a ciò che avviene nell'analisi fattoriale, dove rappresenta la parte di indicatori che non sono stati conteggiati nei fattori presenti. In riferimento alle relazioni tra variabili osservabili e latenti, è importante notare che le prime non sono indicatori delle seconde, infatti possono essere sole nel modello sia come variabile endogena che esogena. Un modello può anche contenere relazioni tra variabili latenti sia di tipo non direzionale, che di tipo direzionale (come ad esempio quando una delle due variabili latenti è il termine di errore).

È opportuno a questo punto definire la natura di tutti i parametri nei modelli SEM. Innanzitutto, ad ogni variabile esogena (sia questa osservabile, latente o termine di errore) è associata una varianza: tutte le varianze di questa tipologia di variabili sono da considerare come parametri del modello e quindi da stimare. Bisogna prestare attenzione in quanto anche le variabili endogene hanno una varianza, ma questa non è considerata un parametro nei modelli SEM, è piuttosto da considerare come implicita nelle altre variabili e può influenzare il modello. Queste infatti possono essere definite algebricamente in una funzione di varianza di variabili esogene. In altre parole la varianza di una variabile endogena non è un parametro del modello, ma una funzione di altri parametri del modello stesso. Analizzate le varianze, è possibile continuare con l'analisi delle covarianze. Anche in questo caso le covarianze sono parametri del modello solo se riguardano variabili esogene, infatti solo variabili esogene possono presentare relazioni di tipo non direzionale. Non è possibile per l'appunto specificare un'associazione non direzionale in cui vengano considerate delle variabili endogene, perché queste sono proprio le riceventi di frecce



direzionali derivanti da altre variabili. Inoltre, come per la varianza, è possibile esprimere la covarianza di una variabile endogena come funzione di altri parametri del sistema. Infine, tutti gli effetti direzionali specificati nel modello costituiscono la terza categoria di parametri. Come precedentemente visto gli effetti direzionali possono collegare diverse tipologie di variabili, si ricordano ad esempio relazioni tra variabili latenti, tra variabili latenti e i loro indicatori, tra gli errori e le variabili a cui sono associati, ecc. riassumendo si identificano tre classi di parametri: le varianze delle variabili esogene, le covarianze tra variabili esogene, i pesi rappresentanti delle relazioni lineari direzionali tra variabili. Tutti questi parametri sono definiti come parametri liberi, che significa che il loro valore non è conosciuto e deve essere stimato, oppure come parametri fissi per i quali invece è prestabilito un valore in origine e di conseguenza non devono essere stimati dal modello. Il valore dei parametri fissi è stabilito in relazione alle necessità del modello; un importante requisito che può essere necessario fissare in un modello è la scala con cui si misurerà la variabile latente, componente fondamentale perché queste risultano essere non direttamente misurabili.

### **4.3.2. Variabili latenti e variabili osservabili**

Innanzitutto, si definiscono variabili osservabili quelle variabili per le quali è possibile effettivamente misurare o registrare un dato, un valore, su un campione di soggetti; ad esempio l'altezza, il risultato di un test, le risposte ad un questionario. Il concetto si può anche esprimere come il manifestarsi di determinate prestazioni e per questo motivo vengono anche chiamate variabili manifeste. Quando si fa riferimento a una variabile osservabile si sottintende che questa è stata effettivamente misurata da qualcuno durante il processo di raccolta dei dati. In opposizione alle variabili osservabili si presentano le variabili latenti. Al contrario delle variabili manifeste, le variabili latenti non possono essere oggetto di un processo di misurazione, ma sono piuttosto costrutti di interesse in un contesto di studio; come esempio si riportano la cultura, la capacità matematica, l'ansia, lo stato sociale. La loro caratteristica è proprio quella di non poter essere individuate con un numero attraverso la misurazione perché non direttamente osservabili; possono però essere individuate delle proxy di queste utilizzando specifici strumenti quali test, questionari, ecc. Quindi le proxy sono gli indicatori dei costrutti latenti; in altre parole si può affermare che costituiscono una misura indiretta di qualcosa che non può a sua volta essere misurato. Allora significa che per i costrutti latenti sopra indicati è possibile determinare delle proxy: lo stato sociale può essere misurato attraverso il reddito, gli anni di istruzione, la tipologia di occupazione; allo stesso modo la capacità matematica può essere valutata attraverso l'abilità in geometria, algebra e trigonometria. Inoltre è opportuno osservare che le proxy delle variabili latenti sono a loro volta variabili e nello specifico sono variabili osservabili, in quanto è possibile misurarle direttamente. Ci si potrebbe chiedere quanto siano affidabili le proxy al fine di valutare il valore di una variabile latente. In effetti queste possono essere più o meno affidabili e ciò dipende sia se la loro presenza è davvero di effettivo interesse nel costrutto, ma anche dal numero di proxy coinvolte. Banalmente se una variabile latente ha un solo indicatore è possibile che questo non porti troppa informazione, quindi si potrebbe pensare che come proxy sia inaffidabile. In seguito a

questa considerazione si può affermare che è raccomandabile utilizzare più di una proxy per ogni variabile latente che si vuole studiare, in modo da ottenere più informazioni e di conseguenza un quadro più completo ed affidabile. Ciò deve avvenire con criterio perché la raccolta dei dati è un processo oneroso, allora è bene determinare poche proxy che portino un contributo importante, piuttosto che tante proxy ognuna di piccola importanza. Nella realtà inoltre ci possono essere casi in cui un singolo indicatore è sufficiente a valutare una variabile latente: si pensi ad esempio a un test per determinare l'intelligenza dove il punteggio ottenuto (variabile osservata) è un più che sufficiente indicatore per valutare l'intelligenza di una persona (variabile latente). Per quanto detto finora, qualsiasi variabile che in un campione di studio appartenente ad una popolazione di interesse non abbia realizzazioni osservate può essere considerata latente.

### 4.3.3. Rappresentazione grafica

Risulta essere di grande utilità nello sviluppo di un modello SEM tracciare un diagramma che raffiguri le variabili coinvolte e le relazioni tra queste. Si rappresenta così con un diagramma il modello in esame; il diagramma viene denominato Path Diagram e utilizza una particolare notazione grafica che verrà descritta di seguito. L'insieme delle equazioni matematiche che definisce il modello può essere considerato equivalente a questo schema grafico, quindi è un modo alternativo e di più immediata comprensione per presentare un modello SEM. È possibile affermare che questo strumento sia in grado di migliorare la comprensione di questi modelli (spesso anche piuttosto ampi) e di conseguenza la comunicazione tra personale e ricercatori di varia provenienza, ma soprattutto risulta essere la base per la creazione di programmi specializzati che possano implementare e testare modelli SEM, attraverso un file di comando corretto. Si mostrerà quale risulta essere la notazione grafica per i differenti tipi di variabile, di relazioni, degli errori e dei parametri da stimare.

Riportata l'importanza della rappresentazione grafica in supporto delle equazioni matematiche, è possibile continuare con una descrizione della notazione. Innanzitutto risulta necessario individuare le variabili di interesse nel caso studio; oltre a individuarle e poi necessario comprendere se queste siano osservabili o latenti perché a seconda della loro natura verranno trattate in modo diverso e quindi devono essere rappresentate in modo diverso. Nello specifico: le variabili osservabili vengono rappresentate all'interno di un quadrato o di un rettangolo, mentre quelle latenti all'interno di circonferenze o di ellissi.

Tra le variabili esistono poi delle relazioni e anche queste vanno rappresentate: variabili latenti e variabili osservate sono collegate con lo scopo di mostrare quelle che sono le proposizioni ottenute dallo studio di un dato fenomeno, di una data teoria, su cui si sta ponendo l'attenzione. Solitamente proprio il legame esistente tra queste due tipologie di variabili sono lo scopo dello studio. In un diagramma sono rappresentate graficamente attraverso delle linee continue che possono essere a senso unico o a doppio senso a seconda della relazione esistente. Nello specifico le frecce a senso unico esprimono il concetto secondo il quale la variabile all'estremità della freccia risulta essere spiegata attraverso la variabile che si trova all'altro capo della linea. Si potrebbe

pensare che la seconda è causa della prima e le relazioni potrebbero quindi essere viste come rapporti causali, in realtà non è detto che sia così: per esempio a volte le variabili latente sono considerate come causa dei loro indicatori, o voler dire che i valori delle variabili osservate variano proprio in conseguenza di una variabile latente corrispondente. È opportuno astenersi dal ipotizzare relazioni causali nei modelli SEM; un'eccezione potrebbe essere fatta per una variabile che precede temporalmente un'altra e che quindi potrebbe essere individuata come una causa della seconda, ma in questi modelli si è già detto che il tempo non viene solitamente preso in considerazione. Ma il concetto di causalità è molto forte e di conseguenza si necessitano di molte informazioni per poter affermare con sicurezza che questo esista. Tornando alla notazione, le linee continue possono essere bidirezionali e in questo caso indicano che esiste correlazione tra le variabili legate da queste, ovvero che tra esse esiste un legame che non può essere visto come direzionale nel modello. Solitamente relazioni bidirezionali vengono rappresentate con linee curve alle cui estremità sono presenti delle frecce. Questa notazione formata da frecce permette di fare una distinzione tra variabili dipendenti e indipendenti. Le prime sono quelle racchiuse in caselle che ricevono almeno una freccia monodirezionale da un'altra variabile del modello; le seconde sono quelle caselle da cui partono i percorsi unidirezionali. A seguito della classificazione espressa, nel passare da grafico e equazioni, le variabili dipendenti si troveranno a sinistra dell'uguale e quelle indipendenti a destra. I modelli SEM risultano essere più generali dei modelli di regressione, infatti in questo caso una variabile può essere allo stesso tempo dipendente perché regredita su altre variabili e indipendente perché predittore di un'altra variabile.

Anche gli errori devono essere riportati nella rappresentazione grafica del modello da analizzare. Come detto in precedenza, i SEM tengono conto sia degli errori nella previsione, ovvero dell'inesattezza che si può ottenere nella determinazione del valore delle variabili latenti dato dall'analisi dei fattori sottostanti, sia delle imprecisioni che contengono le variabili osservate. Allora ogni variabile presenterà una freccia con verso in entrata alla cui estremità ci sarà un termine di errore. Il termine d'errore viene riportato all'interno di un cerchio o di un'ellisse, perché come detto precedentemente viene considerato come una variabile latente in quanto non risulta possibile misurarlo.

Infine nel diagramma è necessario rappresentare quelli che sono i parametri del modello. I parametri risultano essere dei valori incogniti che descrivono come si muove una variabile rispetto a un'altra. La notazione propone di rappresentarli a mezzo di asterischi proprio accanto alle frecce unidirezionali e bidirezionali che legano le variabili del modello tra loro. A differenza dei modelli classici, nei SEM la stima dei parametri risulta essere più complessa, in seguito verrà analizzata. Inoltre gli asterischi vengono riportati accanto ai termini di errori, ciò sta a indicare la varianza della variabile indipendente.

A questo punto si possiedono tutti gli elementi necessari alla realizzazione del grafo che mostrerà da un punto di vista grafico le equazioni che nel seguito dovranno essere scritte in forma di equazioni.

### 4.3.4. Identificazione

Una questione importante e allo stesso tempo difficile è quella riguardante l'identificazione. Per comprendere meglio è necessario prima capire alcuni aspetti fondamentali del processo di stima dei parametri. Questo contesto semplificato utilizzato nella rappresentazione dei modelli che prevede il solo impiego di relazioni lineari, definisce da un lato le relazioni matematiche tra i parametri del modello e dall'altro le varianze e le covarianze delle variabili osservabili. Quindi, se si è a conoscenza del valore di tutti i parametri del modello, risulta computazionalmente semplice calcolare la vera varianza e covarianza delle variabili, ma nella pratica non è così infatti i valori dei parametri non sono conosciuti, bensì da stimare. È possibile però considerare un campione di osservazioni (variabili osservabili) di cui conosciamo il comportamento e da queste ottenere i valori stimati di varianza e covarianza. In questo caso avremo il problema opposto: date le varianze e le covarianze osservate delle variabili osservabili e specificato il modello, si vorranno ricercare i valori dei parametri del modello stesso che riproducano i valori di varianza e covarianza osservati. Ciò significa che la soluzione ottenuta viene stimata per quel dato campione, allora non sarà quella esatta ma quella più vicina possibile ai valori reali, ovvero un'approssimazione con un certo grado di confidenza. Durante la computazione, la stima dei parametri avviene attraverso complesse funzioni della varianza e della covarianza, quindi per ogni parametro libero è necessario che alla fine si ottenga una di quelle funzioni. I parametri che soddisfanno questa condizione con una soluzione unica si dicono identificati; i parametri per i quali esiste più di una soluzione distinta si dicono sovra identificati. I modelli SEM con uno o più parametri sovra identificati sono di notevole interesse in quanto esiste solo un modello tra tutti in cui il problema della corrispondenza tra modello e dati risulta essere significativo e questo va trovato a seguito di un'analisi più approfondita. Modelli contenenti questa tipologia di parametri creano la possibilità di avere un modello che potrebbe essere fittato sui dati differenti in modo non soddisfacente. Invece, se sono presenti solo parametri identificati, questi sono stimati sempre perfettamente, quindi si perde la significatività per valutare la plausibilità del modello attraverso il fitting. Per i parametri liberi dati, se non è possibile esprimerli in una funzione di varianza e covarianza, si dice che sono non identificati. Un modello contenente anche solo un parametro con queste caratteristiche nella pratica risulta essere inutilizzabile, poiché la stima di questi risulta essere arbitraria e di conseguenza non interpretabile. Sfortunatamente risulta complesso verificare l'identificazione di un parametro, però è possibile esaminare due caratteristiche. Innanzitutto, dato che la scala delle variabili va definita attraverso un parametro fisso, se la condizione imposta non è rispettata significa che esistono dei parametri non definiti. Inoltre, l'effettivo valore dei parametri stabilito precedentemente non deve eccedere il valore della varianza e della covarianza del campione per le variabili osservate; in questo modo si ha una mancanza di identificazione poiché i dati per la stima dei parametri sono troppo pochi. Queste due condizioni risultano essere necessarie ma non sufficienti, l'unica vera condizione sufficiente e effettuare la verifica algebrica. I software manifestano il problema attorno ad alcuni parametri, sarà l'utilizzatore ad approfondire lo studio attorno a questi per determinare se risultano essere identificati o meno.

### 4.3.5. Disconfirmability

Una questione importante da trattare durante la specificazione di un modello appartenente a questa classe è il grado di disconformità che introduce la possibilità che il modello risulti essere inconsistente rispetto ai dati osservati. Se un modello è capace di fittare ogni set di campioni di variabili osservabili in modo perfetto, significa che non è disconforme affatto. Molti modelli hanno generalmente molti o troppi parametri, in riferimento a molte varianze e covarianze di variabili osservabili. Un modello di questo tipo non è di interesse particolarmente significativo dal punto di vista scientifico. Infatti non risulta utile al fine di spiegare in modo parsimonioso la struttura sottostante ai dati. Se un modello è disconforme significa che esistono dei gradi di libertà, ovvero che il numero effettivo di parametri stimabili è minore del numero delle varianze e delle covarianze delle variabili osservabili; il numero dei gradi di libertà è conteggiabile attraverso la differenza delle due voci sopra citate. In realtà avere qualche parametro indefinito e di conseguenza pochi gradi di libertà permette di fittare i dati abbastanza bene e di conseguenza non si può dire che il modello sia disconforme. Ovviamente non si può dire lo stesso quando i gradi di libertà risultano essere molti. Nel caso in cui si trovi un buon fit del modello, allora è possibile affermare con maggior confidenza che il modello è plausibile e rappresenta bene la struttura dei dati.

Nel processo di specificazione del modello è bene tener presente il problema della disconformità per poter concepire costrutti che non siano troppo parametrizzati. Modelli fortemente parametrizzati risultano essere rigidi, fittano molto bene il campione in esame ma non si adattano agli altri campioni della stessa popolazione.

### 4.3.6. Modelli equivalenti

Un'altra importante questione che si presenta in questa fase è quella dei modelli equivalenti. Essenzialmente due modelli risultano essere equivalenti se fittano con la stessa bontà lo stesso set di dati. Molti modelli non sono distinguibili dal punto di vista matematico, ma differiscono solo in termini di sostanziale significato e la loro conseguente interpretazione della soluzione ottenuta dopo che sono stati fittati sui dati. Dato un modello di interesse, è possibile costruirne altri ad esso alternativi ed equivalenti attraverso delle semplici regole, sviluppate da Lee e Hershberger nel 1990 sulla base del lavoro di Stelzl del 1986. Queste regole indicano delle condizioni sotto le quali è possibile replicare un percorso con un altro differente senza peggiorare la bontà del fit. Idealmente si potrebbe trovare che esistano relativamente pochi di questi modelli e che questi non siano sostanzialmente significativi. D'altra parte però, se si trova che è possibile generare un sostanziale numero di questi modelli e se sembra che questi possano essere plausibili, ci saranno dei punti da seguire nello studio dell'originale per sostenere la sua validità.

L'esistenza di modelli equivalenti solleva il problema di poter confondere il ricercatore nella fase di progettazione del modello. È possibile che si generi confusione su quale sia la vera struttura del modello se per il fenomeno osservato esistono delle alternative di spiegazione, tutte differenti e

allo stesso modo buone. Bisogna fare attenzione a non ignorare una spiegazione piuttosto che supportarne un'altra senza avere condotto prima uno studio sul tema in oggetto. Allo stesso modo non è possibile ignorare a priori l'esistenza di modelli equivalenti alternativi e assumerne a priori uno effettuando la specificazione del modello ipotizzando che questo fornisca una valida spiegazione dei dati del campione. Nella pratica il problema è forte poiché modelli equivalenti occorrono frequentemente e spesso in largo numero, per questo motivo è importante saper gestire correttamente la situazione al fine di scegliere la struttura che meglio si adatta al campione oggetto di studio. Per poter fare chiarezza su quale possa essere il modello che spiega il fenomeno nel modo migliore tra le alternative possibile trovate, è opportuno avere una conoscenza della teoria su cui si sta svolgendo lo studio; ciò risulta essere vero sia con un utilizzo del tipo confermativo, sia con quello esplorativo: la differenza che intercorre tra questi due casi è semplicemente il momento del ricorso ad uno studio teorico del caso.

### 4.3.7. Strategie

Nelle applicazioni empiriche dei SEM esistono varie strategie che possono essere impiegate per la costruzione e lo sviluppo dei modelli. Jöreskog e Sörbom individuano tre strategie: strettamente di conferma, generazione del modello e comparazione del modello. Queste vengono proposte qui di seguito.

Nella prima strategia, quella di conferma, si costruisce un modello di interesse e si valuta attraverso il fitting, ovvero alla sua adeguatezza rispetto ai dati che lo hanno generato. Se il modello oggetto di valutazione produce dei parametri interpretabili e fitta bene i dati, si conclude che possa essere plausibile per descrivere il fenomeno di studio; altrimenti il modello è da scartare. Questa procedura non è molto utilizzata nella pratica poiché lascia poca flessibilità di operazione e non dà la possibilità di indirizzare verso una soluzione positiva l'output negativo di una valutazione.

Una strategia sicuramente più utilizzata è quella di generazione del modello. La procedura prevede di iniziare con la specificazione del modello, ovvero con i passi analizzati precedentemente, e di fittare in seguito i costrutti trovati ad un set di dati appropriato. La soluzione così ottenuta viene valutata con lo scopo di apportare delle modifiche per migliorarlo dal punto di vista della parsimonia o della sua adattabilità ai dati. Risulta essere possibile anche una semplificazione dei costrutti attraverso la cancellazione di quei parametri che risultano essere poco significativi nella spiegazione del fenomeno. Molto frequentemente alcuni parametri addizionali vengono aggiunti per aumentare l'adattamento ai dati, per migliorare la bontà del modello. Un eccellente metodo per raggiungere questi obiettivi è, data una soluzione, eseguire il test di Wald per determinare i gradi con cui il fit viene deteriorato in seguito a una cancellazione di un sottogruppo di parametri liberi. Nella pratica i parametri liberi non significativi vengono convertiti in parametri fissi con valore fissato a zero. Se è possibile determinare un sottogruppo di parametri liberi da eliminare (individuati da un non rifiuto dell'ipotesi nel test di Wald), producendo solo un piccolo decremento della bontà del fit, è possibile affermare che i parametri stessi non sono di particolare utilità per

spiegare il fenomeno studiato. Il risultato di questa operazione è un modello semplificato che comunque fitta i dati in maniera molto simile e vicina rispetto al modello originario di partenza, ma sicuramente più parsimonioso. Alternativamente è possibile utilizzare il test dei moltiplicatori di Lagrange per determinare i gradi di libertà con cui il modello può migliorare a seguito della conversione di alcuni parametri da fissi a liberi, ancora con l'obiettivo di migliorare la bontà del modello, ovvero di interpretare meglio i dati. In questo modo si avrà invece un modello con un numero di parametri maggiore a quello del modello di partenza, che quindi fitterà in modo migliore i dati. Una metodologia comune per l'ultima operazione descritta (LM) è rappresentata dalla modifica degli indici (MI): alcuni programmi per la risoluzione di SEM forniscono un valore per tutti quei parametri che sono fissati a zero e che quindi non sono rappresentati nel diagramma di flusso. Questi parametri vengono convertiti al fine di ottenere un risultato migliore nel complessivo test  $\chi^2$  sulla capacità del modello di interpretare i dati, solitamente con delle performance molto buone. Nonostante la strategia di generazione dei modelli sia utilizzata ampiamente nella pratica, è opportuno approcciarsi ad essa con cautela poiché presenta molti pericoli che solitamente vengono ignorati. In particolare, è di massima importanza che ognuna delle modifiche apportate sia giustificabile e significativa. Un'altra importante questione è la necessità di validare il modello così modificato provando a fittare un set di dati nuovo, differente da quello utilizzato per la stima dei parametri; nella pratica questo pericolo è spesso ignorato. Senza tener conto di queste considerazioni, il modello così generato potrebbe avere poco significato e quindi potrebbe risultare di poca utilità poiché non spiega le relazioni davvero esistenti tra le variabili prese in considerazione. Ma ci potrebbero essere altre ragioni per utilizzare questo tipo di strategia. Si consideri un caso ideale in cui esiste un modello vero che calcola in modo esatto la varianza e la covarianza delle variabili osservabili nella popolazione di interesse, ma supponiamo che valutato sul campione non sia specificato a seguito dell'omissione di uno o più parametri. Teoricamente usando la statistica MI si lavora bene poiché riesce ad aggiungere quei parametri necessari a modificare il modello in modo tale che diventi uguale a quello vero. Nella realtà però ciò non è così vero: MacCallum nel 1986 prova a verificare questa teoria attraverso dei dati simulati, ottenendo conferma da soli 22 casi sui 160 campioni di dati analizzati. Così la teoria è stata invalidata dai risultati. Un'altra importante questione interessa il problema della capitalizzazione delle possibilità. Infatti, le modifiche che vengono apportate al modello sono individuate sulla base dei risultati ottenuti dall'analisi di un campione di dati per migliorare l'adattamento ai dati stessi. Quindi le modifiche potrebbero essere determinate in parte dalle possibili particolari caratteristiche del campione osservato, ciò implica che la modifica potrebbe essere migliorativa nel campione, ma non è possibile generalizzare il concetto estendendone la validità alla totalità dei campioni che possono essere estratti da una popolazione. Di conseguenza è possibile affermare che se la stessa strategia di generazione viene implementata partendo dallo stesso modello, ma da un campione di dati differente (oppure utilizzando la totalità della popolazione), le modifiche da apportare potrebbero essere discordanti. Da uno studio empirico condotto sempre da MacCallum, Roznowski e Necowitz (1992), si può concludere che le considerazioni fatte risultano essere veritiere nella realtà: le modifiche da implementare al modello di partenza risultavano essere di volta in volta differenti, instabili, se la dimensione del campione era minore di 400, e non completamente stabili con dimensione del campione pari a

1200. L'esperimento dimostra come le modifiche al modello risultano essere molto sensibili alle caratteristiche del campione di dati utilizzato per fare il fitting: in particolare minore sarà la dimensione del campione, maggiore sarà la diversità delle modifiche da applicare al modello per farlo adattare in modo migliore ai dati. In conclusione alla trattazione di questa strategia è necessario affermare che i ricercatori sono particolarmente interessati all'uso di questa modalità operativa, che deve essere affrontata senza dimenticare che si tratta di una tecnica esplorativa. Non sussiste nulla di sbagliato nell'applicare questa modalità, purché si riconosca che il modello ottenuto in questo modo non può essere supportato e validato se non attraverso la valutazione mediante un nuovo campione di dati. Il problema nasce quando si eseguono i passi della procedura senza nessuno sforzo di significato sostanziale di approccio alle modifiche del modello, e quando il modello risultante è trattato come se potesse essere confermato solamente in riferimento al campione di dati osservato in partenza. Concludendo è possibile affermare che questa strategia risulta essere un buon approccio solamente se utilizzata responsabilmente, senza ignorare le problematiche esposte.

La terza strategia da analizzare è quella comparativa. Si scelgono a priori dei modelli alternativi e si fittano sullo stesso campione di dati. I modelli alternativi sono la rappresentazione di posizioni teoriche differenti e alternative, oppure sono alla base di conflitti di ricerca su un oggetto di studio. Precedentemente si è discusso di modelli alternativi sviluppati all'interno di un certo contesto, risultato di una serie di modifiche date dall'evidenza statistica. In altre parole, nella prima strategia si parte dalla costruzione di un solo modello che a seguito di cambiamenti si trasforma in altri equivalenti, ma con stessa base di partenza. In questa circostanza invece si parte dalla costruzione e dallo sviluppo di modelli alla base diversi tra loro, poiché le ipotesi delle tesi che li governano sono a loro volta differenti. Una volta che i modelli sono sviluppati, questi devono essere fittati su un set di dati che sia appropriato: i risultati dovranno essere valutati e comparati tra loro. La valutazione deve rispettare alcune caratteristiche. Innanzitutto la bontà del fit, come discusso anche precedentemente, poiché bisogna tenere in mente che è importante applicare procedure di valutazione del fit che tengano conto della complessità del modello. L'attenzione però non va riposta solamente sulla fase di fitting, infatti un modello che fitta bene il campione di dati potrebbe avere altre caratteristiche non particolarmente attrattive. Infatti risulta di fondamentale importanza analizzare i parametri stimati per quanto riguarda la loro interpretabilità e la loro capacità di spiegare le relazioni esistenti tra le variabili. Per esempio un modello con buon fit ma con parametri privi di significato risulta essere di poco valore e di inutile applicazione nella realtà.

Riepilogando, relativamente alle tre strategie esaminate per la specificazione dei modelli, quella di stretta conferma è probabilmente troppo rigida e di conseguenza non viene particolarmente utilizzata. La strategia di generazione appare essere largamente impiegata anche se presenta dei difetti significativi che nella routine vengono solitamente ignorati. La strategia di comparazione riesce a evitare quei difetti in larga misura e fornisce un meccanismo in grado di rendere il modello utilizzabile in un intervallo di studio, da studio di conferma di un piccolo numero di modelli all'esplorazione di un grande numero di modelli.



## 4.4. Stima dei parametri

I modelli SEM rappresentano una serie di ipotesi in relazione a come le variabili vengono generate e create durante l'analisi del fenomeno in oggetto. I parametri del modello sono i coefficienti della regressione lineare e le varianze e covarianze delle variabili indipendenti. Ognuno di questi parametri sono di fondamentale importanza nella interpretazione del modello, delle relazioni che sono state identificate tra le variabili; questi risultano essere ignoti e di conseguenza andranno stimati. Il processo di stima è il primo passo logico da fare durante la fase di modellazione, ovvero in seguito alla fase di specificazione affrontata precedentemente. Simultaneamente alla stima, avviene il test di bontà del fit, ovvero viene applicato una statistica test che permette di comprendere se il modello risulta essere adeguato. La bontà del fit è riflessa nella similarità esistente tra la matrice di covarianze stimata ( $\Sigma_{\theta}$ ) dal modello e la matrice di covarianza reale della popolazione da cui il campione è stato preso. Sulla base di questa definizione, risulta chiaro che il test diventa un elemento fondamentale per poter interpretare correttamente i risultati ottenuti dal modello sviluppato. Nel caso in cui il modello non risultasse essere consistente se confrontato con la matrice di covarianze della popolazione, non ci sarebbero le condizioni necessarie per interpretare i parametri e di conseguenza non si potrebbe dire alcunché sulle relazioni ipotizzate tra le variabili. La statistica test è una  $\chi^2$  e dipende dalla procedura di stima scelta; infatti metodi diversi producono valori stimati e risultati di test piuttosto diversi. Per questo motivo è desiderabile avere qualche regola a riguardo in modo da poter lavorare in modo opportuno in condizioni diverse, facilitando così la messa in pratica dei SEM. Ad ogni modo è possibile affermare che i valori stimati saranno vicini a quelli veri se il modello stimato è vicino a quello vero, e che di conseguenza il test di significatività mostrerà che le conclusioni risultano essere adeguate. Sebbene esistano molte tipologie di stima e di statistiche test, qui ci si concentrerà sulle tre metodologie maggiormente utilizzate: la massima verosimiglianza (LM), i minimi quadrati generalizzati (GLS), entrambe da utilizzare sotto l'ipotesi di normalità, e la distribuzione asintotica libera (ADF). Questi metodi forniscono una stima (con la stima degli errori standard per i parametri liberi) e un test  $\chi^2$  per la bontà del modello. Lo stimatore di massima verosimiglianza è largamente utilizzato come approccio nei modelli a equazioni strutturali; a volte però può essere inadeguato in quanto opera sotto le condizioni di normalità che sono solitamente violate nella realtà. Molti ricercatori si sono interrogati sulla robustezza dello stimatore ML (ma anche di altri) per definire la grandezza dell'impatto che una violazione di questa ipotesi di normalità si rifletta in modo negativo sui risultati ottenuti. ML è risultato essere abbastanza robusto e di conseguenza risulta essere in grado di fornire soluzioni appropriate anche quando siano presenti variabili con un distribuzione diversa da quella normale. Il metodo GLS al contrario non è stato particolarmente studiato, ma si può parlare di distorsione negativa. Infine, il metodo ADF ha prodotto un risultato non consistente in quanto alcuni studiosi hanno trovato una robustezza buona, mentre altri il risultato opposto.

### 4.4.1. Concetti base nella fase di stima

L'obiettivo della stima è ottenere un valore numerico per tutti quei parametri che risultano essere sconosciuti. Per ottenere una buona stima dei parametri deve essere affrontata la questione dell'identificazione. Questa comprende lo studio delle condizioni da avere per ottenere un'unica soluzione per i parametri specificati nel modello. Una delle condizioni richieste è che il numero di parametri liberi (che verrà indicato d'ora in avanti con la lettera  $q$ ) sia minore o uguale al numero di elementi non ridondanti nella matrice di covarianza (questi elementi verranno indicati con la lettera  $p^*$ ). È possibile calcolare il numero esatto di questi elementi non ridondanti essendo a conoscenza del numero delle variabili nella matrice di covarianza ( $p$ ), nello specifico  $p^* = p * (p + 1)/2$ . Quindi  $q \leq p^*$  è solamente una condizione necessaria affinché il modello possa dirsi identificato. Ma la condizione sufficiente per operare questa dichiarazione è che tutti i parametri liberi siano identificati a loro volta. Quest'ultima condizione risulta tuttavia di non semplice comprensione e valutazione nella pratica, come invece lo è  $q \leq p^*$ , inoltre diventa veramente complesso quando i modelli sono molto grandi e di conseguenza cresce il numero dei parametri liberi. Risulta sicuramente più semplice rilevarlo attraverso l'applicazione di programmi informatici per la trattazione di SEM. Questi infatti forniscono un messaggio di errore quando esiste una relazione lineare tra dei parametri liberi, identificando così un sistema indefinito perché contenete infinite soluzioni. Altri messaggi di errori appaiono a riguardo della matrice di covarianza, quando questa non risulta essere definita positiva: se la matrice di covarianza non è definita positiva significa che non sarà possibile calcolarne l'inversa e di conseguenza il sistema risulterà essere impossibile. È possibile illustrare il concetto espresso nella condizione sufficiente attraverso un sistema di equazioni in cui sono presenti diverse incognite (parametri) da determinare. È possibile allora affermare che se il numero di incognite è maggiore delle equazioni il sistema è indefinito. Questa situazione risulta essere di particolare interesse nei modelli SEM poiché occorre frequentemente quando si lavora con variabili latenti, dove la scala dei fattori non è spesso data. Al contrario, nel caso in cui il numero delle incognite risulta essere uguale a quello delle equazioni che ne descrivono le relazioni esistenti tra loro, esiste una sola e unica soluzione che soddisfa il sistema di equazioni; allora il modello è identificato. Modelli di regressione multipla e path analisi su modelli costruiti con variabili misurabili risultano essere spesso identificati. Ciò significa che un solo valore verrà stimato per ogni parametro. La condizione di sovra identificazione si presenta quando il numero delle equazioni è superiore a quello delle incognite e di conseguenza non si ha una soluzione esatta. Sebbene non si possa conseguire una soluzione unica, si ottiene la definizione di un certo criterio che rappresenta la relazione tra le variabili.

Il concetto espresso precedentemente, esposto nel caso di più equazioni simultaneamente presenti, può essere facilmente trasferito sui modelli SEM. La condizione di identificazione o sovra identificazione deve essere verificata per ogni parametro presente nel modello, poiché solamente se la condizione è rispettata per tutti si avrà un modello in grado di fornire un'unica soluzione. Nel caso in cui tutti i parametri possano essere stimati, è possibile determinare i gradi di libertà del modello testato:  $df = p^* - q$ . Quando i gradi di libertà risultano essere positivi il modello può essere testato. Il test è solitamente basato sulla statistica test con distribuzione  $\chi^2$ ; si possono poi

avere procedure di stima differenti per la valutazione del modello selezionato, ma in ogni caso devono avere distribuzione con gli stessi gradi di libertà.

#### 4.4.2. Problemi pratici

Oltre alla questione dell'identificazione ne esistono altri che si incontrano frequentemente nell'applicazione pratica dei modelli e che possono causare problemi nella procedura di stima. Si ipotizzi di avere dei dati che soddisfino i requisiti statistici fondamentali, ovvero che siano indipendenti e identicamente distribuiti, e si ricerchino i potenziali problemi. Uno dei problemi che si incontra più frequentemente è quello di avere una matrice di covarianze non definita positiva. Questo problema si riscontra solitamente quando esiste una dipendenza lineare tra le variabili osservabili, ovvero quando una variabile è perfettamente prevedibile da un'altra. Poiché è necessario calcolare la matrice inversa di quella contenente le covarianze per poter raggiungere la stima dei parametri, nel caso in cui non si sia in grado di operare questo passaggio non si potrebbe arrivare alla soluzione finale. In altre parole, nessuna soluzione è ottenibile quando sono presenti variabili con dipendenza lineare. Allora, per evitare questo problema, le variabili da utilizzare nella specificazione del modello devono essere scelte con cura. Un altro problema comune è quello della non convergenza. La procedura di stima consiste in un processo iterativo di computazione che inizia con un dato valore per i parametri da stimare (questi valori possono essere forniti dal ricercatore o generati dal computer stesso): ad ogni ciclo il risultato ottenuto viene confrontato con quello del ciclo precedente e il processo iterativo prosegue fino a quando alcuni criteri non vengono soddisfatti. I criteri statistici necessari al fermare il ciclo iterativo verranno discussi successivamente. In pratica molti software fermano automaticamente il processo dopo un determinato numero di cicli, anche se i requisiti non sono stati raggiunti; ciò viene implementato per evitare un eccessivo consumo del time del computer, che potrebbe essere molto costoso. Il numero di cicli possibili risulta comunque essere maggiore del numero dei cicli necessari in media per raggiungere quei limiti statistici che permettono di fornire una soluzione ottimale. La non convergenza si presenta proprio quando il processo viene stoppato in quanto sono terminati i cicli a disposizione, ma non si è ottenuto nessun requisito statistico. Da notare che in questo modo la non convergenza vera e propria non può essere determinata, poiché non si ha a disposizione un numero di cicli infinito, ma solo ipotizzata con buona probabilità. Il problema della non convergenza verso una soluzione del modello in esame può derivare dalla povertà del modello stesso, ovvero dalla poca capacità che ha nello spiegare le relazioni ipotizzate essere esistenti tra le variabili. Un altro motivo di non convergenza potrebbe essere la povertà di significato nei valori di partenza; a ciò è possibile porre rimedio in quanto modificabili facilmente, in modo studiato o anche random. Bisogna comunque riportare che nel caso in cui il modello si appropriato all'interpretazione dei dati, risulta difficile che un set di valori iniziali non opportuno porti alla non convergenza il modello; ciò potrebbe avvenire solo in rare circostanze. Solo se il modello è povero, allora anche i valori iniziali avranno un ruolo importante nella capacità di determinare una soluzione per il modello.

### 4.4.3. Modelli: massima verosimiglianza

I modelli SEM vengono utilizzati con lo scopo di valutare teorie sostantive attraverso dati empirici e ipotesi di modellazione. La grande maggioranza di questi modelli è costituita da modelli di equazioni sulla covarianza, dove i dati empirici da studiare sono caratterizzati da una matrice di covarianza quadrata di ordine  $p$ , stimatore della matrice di covarianza della popolazione ( $\Sigma$ ) da cui il campione è stato estratto. Un modello di equazioni strutturali può essere specificato attraverso un vettore di  $q$  parametri ignoti e da una matrice di covarianza generata dai dati empirici ( $\Sigma_\theta$ ). L'ipotesi nulla dei modelli SEM è che sia  $\Sigma = \Sigma_\theta$ . Ogni metodo di stima, come LM, GLS o ADF, ha il proprio criterio per produrre un valore stimato per i parametri e una statistica test per valutare l'ipotesi nulla. Il metodo della massima verosimiglianza è quello utilizzato di default in molti programmi informatici per la risoluzione di modelli SEM, ma anche nelle letterature risulta essere impiegato per la descrizione statistica. Il fatto che sia il metodo maggiormente utilizzato è testimonianza della bontà del metodo stesso, di conseguenza applicazioni di metodi diversi da questo richiedono una giustificazione esplicita. Vista la vasta scala di utilizzo, si vuole studiarlo nel dettaglio nel seguito.

Il termine massima verosimiglianza descrive il principio statistico che sottende la derivazione dei parametri stimati; i valori stimati sono quelli che massimizzano la probabilità che i dati (ovvero le covarianze osservate) siano stati presi dalla popolazione di interesse. Questo è un metodo che si basa sull'ipotesi di normalità, infatti si assume che la distribuzione delle variabili endogene sia normale. Dichiarando la normalità delle variabili, allo stesso tempo si sta affermando che queste siano continue: infatti solo variabili continue hanno distribuzione normale. Perciò se le variabili non assumono valori continui, come spesso accade nella realtà, o se presentano una distribuzione molto lontana dalla normale sarà necessario implementare un metodo di stima differente che non abbia queste ipotesi alla base. Molte forme di stima con ML sono simultanee, ciò significa che le stime dei parametri vengono calcolate tutte contemporaneamente; per questo motivo il metodo si dice essere full-information. Quando tutti i requisiti sopra citati sono soddisfatti e il modello è stato correttamente specificato, lo stimatore ML è asintoticamente non distorto, efficiente e consistente. In questo senso lo stimatore è avvantaggiato sotto le condizioni ideali, rispetto a metodi partial-information che analizzano una sola equazione alla volta. Il criterio utilizzato da questo stimatore è la ricerca della minimizzazione dell'errore, ovvero si vuole ridurre al minimo la discrepanza tra la covarianza reale della popolazione e il valore predetto dal modello creato. Una differenza tra lo stimatore ML e altre tecniche statistiche standard si presenta proprio nel calcolo della stima della varianza della popolazione  $\sigma^2$ . Nelle tecniche classiche questo valore è calcolato in un singolo campione come il rapporto tra la somma dei quadrati delle distanze dalla media ( $SS$ ) e i gradi di libertà generali all'interno del gruppo [ $S^2 = SS/df$ , oppure si utilizza  $df = N - 1$ ]. Nel metodo di massima verosimiglianza invece al denominatore compare  $N$  [ $S^2 = SS/N$ ]. Ciò comporta due situazioni differenti: quando i campioni sono di piccole dimensioni  $S^2$  è uno stimatore negativamente distorto per  $\sigma^2$ ; in campioni di grandi dimensioni invece  $s^2$  sarà simile a  $S^2$ , infatti asintoticamente tenderanno allo stesso valore. Nei software per la risoluzione di SEM viene

stimata la varianza con l'operatore  $S$ , non  $s$ . I software di analisi statistica generale invece utilizzano  $s^2$ , che non è esattamente come calcolare  $S^2$  nei programmi per i SEM.

Solitamente la procedura di implementazione di ML è di tipo iterativo, ciò significa che il software produce una soluzione iniziale e in seguito cerca di migliorare questa stima attraverso sequenziali cicli di calcolo. Migliorare significa che il fit generale del modello sui dati risulta essere più vicino. Per molti modelli identificati, il fit potrebbe risultare essere perfetto. Per i modelli sovra identificati invece il fit non sarà perfetto, ma il software continuerà ad operare dei cicli fin quando l'errore di fit non scenderà sotto un valore minimo stabilito in precedenza. Quando ciò accade il processo di stima si interrompe. La pratica iterativa può convergere a una soluzione più facilmente se la procedura inizia con dei valori iniziali ragionevoli; ad esempio il software nel determinare la stima di un parametro con valore positivo, farà maggior fatica se a questo come valore iniziale si attribuisce un valore negativo. Una procedura iterativa potrebbe anche fallire, senza quindi arrivare a una soluzione stabile; solitamente quando ciò accade i programmi danno un messaggio di errore e qualsiasi set di stima è stato realizzato presenterà poca confidenza. Alcuni programmi generano di per sé i valori iniziali, ma è importante comprendere che i valori derivati in questo modo non sempre portano ad una soluzione convergente. Sebbene i valori iniziali siano spesso abbastanza buoni, a volte risulta essere necessario fornire al programma dei valori migliori, su cui si è ragionato, per garantire o facilitare la convergenza verso la soluzione, specialmente quando si trattano modelli complessi con molte equazioni simultaneamente presenti. Un'altra pratica è quella di porre un numero massimo di cicli da poter effettuare nella procedura iterativa, solitamente il valore è compreso tra 30 e 100. Ovviamente consentendo un numero maggiore di prove al computer questo avrà più possibilità di avere una soluzione convergente.

**Soluzioni inammissibili.** Anche se di solito non si riscontrano problemi quando si analizzano analisi path di modelli ricorsivi, può accadere nella stima di ML, o di altri metodi iterativi, che la soluzione a cui si converge sia inammissibile. Ciò risulta essere molto evidente quando i valori stimati dei parametri non hanno significato logico; ad esempio nel caso si trovasse un valore stimato per la varianza negativo o un valore assoluto maggiore di uno per una correlazione, è chiaro che nessuna interpretazione risulta essere plausibile. Le cause di questa situazione possono essere:

- Errore nella specificazione;
- Modello non identificato;
- Presenza di outliers nel campione che possono distorcere la soluzione;
- Una combinazione di campioni piccoli (minori di 100) e pochi indicatori per fattore;
- Cattivi valori iniziali;
- Correlazione nella popolazione estremamente bassa o alta da far risultare il modello nella pratica sotto identificato.

Un'analogia che può aiutare nel comprendere come la soluzione raggiunta possa non essere ammissibile è il seguente. Lo stimatore ML è come un religioso fanatico che crede solo nella sua specificazione del modello e quindi forza i dati sul modello piuttosto che modificare qualcosa ammettendo che la soluzione risulta essere non plausibile. Alcuni programmi sono impostati per

non permettere soluzioni inammissibili ad esempio impedendo di stimare una varianza con un valore negativo. Una volta trovata la causa dell'inammissibilità della soluzione trovata, è necessario apportare delle modifiche opportune e rimandare in esecuzione la procedura di stima. In altre parole il computer fornisce solamente un allarme, sarà l'utente stesso a validare o meno la soluzione trovata e, in caso negativo, apportare le modifiche necessarie.

**Scala.** Il metodo ML può essere di due tipologie per quanto riguarda la scala: scala libera e scala invariante. Utilizzare una scala libera significa che se la scala di una variabile viene linearmente trasformata, il parametro stimato per la variabile trasformata può essere algebricamente convertito nella metrica originale. Scala invariante invece significa che il valore della funzione di ML fittata in un particolare campione rimane sempre lo stesso senza curarsi della scala delle variabili osservate. Tuttavia lo stimatore ML può perdere queste proprietà se viene analizzata la matrice di correlazione invece che quella di covarianza. In questo caso lo stimatore assume un valore non standardizzato cosa che si riflette sul calcolo degli errori che a loro volta non saranno standardizzati. Ciò è di considerevole importanza poiché il livello di significatività della statistica di un parametro non standardizzato potrebbe non essere attribuito alla corrispondente stima standardizzata.

**Assunzioni e propagazione dell'errore.** Come già espresso, di default ML assume che le variabili siano non standardizzate. Inoltre assume che non ci siano valori mancanti quando il file di dati grezzi, iniziali, viene analizzato; anche se tuttavia esiste una forma apposita per trattare casi in cui il file sia mancante di alcuni dati. Le assunzioni statistiche per la massima verosimiglianza includono indipendenza dei dati, normalità multivariata per le variabili endogene e indipendenza delle variabili esogene e dei termini di errore. Un'ulteriore assunzione da fare quando si analizza il path diagram è che le variabili esogene siano misurate senza errore, ma allo stesso tempo questo non risulta essere un requisito necessario all'applicazione del metodo ML. Detto ciò, probabilmente la più importante assunzione che va fatta è che il modello sia specificato correttamente: di rilevante importanza a causa dell'errore di propagazione. I metodi di piena informazione, come quello di massima verosimiglianza, tendono a propagare l'errore in tutto il modello; ciò significa che un errore nella specificazione di un parametro può generare altri errori in altri parametri in qualsiasi posizione del modello. Si supponga che l'errore di misurazione della correlazione esistente tra i due indicatori di un fattore sia davvero sostanziale ma che non possa essere stimata a causa dell'identificazione. Questo errore di specificazione potrebbe propagarsi nella stima del fattore poiché caricato da questo paio di indicatori. Risulta difficile prevedere quale possa essere la direzione o la magnitudo della possibile contaminazione, ma il più grave errore di specificazione è difficile prevedere la direzione o la grandezza di questa "contaminazione", ma il più grave errore di specificazione potrebbe essere la distorsione del modello in altre parti, distanti da quella che l'ha generata. Quando avviene una mancanza nella specificazione allora i modelli a informazione parziale performano meglio di quelli a informazione totale, come quello di massima verosimiglianza. Ciò è credibile poiché i metodi a informazione parziale hanno la capacità di isolare maggiormente gli effetti di un errore nella fase di specificazione in una parte del modello invece di permettere loro di diffondere l'errore stesso in altre parti. Alcuni studi sul metodo ML e

altri stimatori per l'utilizzo di modelli contenenti variabili latenti hanno dimostrato che la distorsione risulta essere generalmente trascurabile in molti campioni quando tre fattori per la misurazione del modello stesso sono specificati correttamente. Tuttavia quando la specificazione è scorretta, altri modelli risultano essere preferibili o di integrazioni a ML.

#### **4.4.4. Interpretazione dei parametri stimati**

L'interpretazione delle stime ottenute per i parametri stimati con ML, ma anche con altri stimatori, per modelli schematizzabili attraverso path diagram risulta essere chiara e lineare. Innanzitutto, i coefficienti delle relazioni evidenziate dal diagramma sono da interpretare allo stesso modo dei coefficienti di una regressione multipla; ciò è vero sia per le soluzioni standardizzate sia per quelle non standardizzate. Per quanto riguarda invece le varianze delle stime bisogna differire i due casi a seconda della standardizzazione o meno delle soluzioni.

#### **4.4.5. Fittare il modello con la matrice di correlazione**

Si è già discusso a proposito del default di ML che assume a priori di lavorare con variabili non standardizzate. Se al contrario le variabili hanno una distribuzione standardizzata, lo stimatore risulta essere inaccurato, sia per quanto riguarda i valori dei coefficienti, sia per la stima degli errori standard. Questo accade se il modello non è a scala invariante, ovvero se il fit complessivo ai dati dipende dalla possibilità che le variabili siano o meno standardizzate. Se il modello risulta essere effettivamente a scala invariante è determinato da una combinazione piuttosto complessa di caratteristiche, inclusa la modalità in cui si attribuisce una scala ai fattori e la presenza di vincoli di uguaglianza su determinati parametri stimati. Un sintomo visibile di scala invariante si può notare dall'analisi della matrice di correlazione, infatti alcuni elementi della diagonale non risultano essere pari a zero come dovrebbero. Esiste però un metodo per correggere il fitting del modello proprio attraverso la matrice di correlazione invece che da quella di covarianza; questa prende il nome di stima vincolata o ottimizzazione vincolata. Questo metodo implica l'imposizione di vincoli non lineari per la stima di alcuni parametri al fine di garantire che il modello sia a scala invariante. Questi vincoli possono essere abbastanza complicati, di conseguenza non risulta agevole programmarli manualmente e inoltre non tutti i software per la trattazione di SEM supportano l'inserimento di vincoli non lineari. Tuttavia alcuni programmi permettono stime vincolate da eseguire automaticamente attraverso la selezione di opzioni. Questo metodo automatico accetta come input anche file di dati grezzi, non analizzati, o matrici di correlazione.

Esistono almeno tre situazioni in cui è possibile utilizzare la stima vincolata. Il primo caso si ha quando si sta conducendo un'analisi su un SEM secondaria basata su risorse dove la correlazione sia riportata ma non si possiedono i valori delle deviazioni standard; ciò potrebbe avvenire se i dati originari non sono disponibili. Esiste poi una ragione teorica per imporre vincoli di uguaglianza sui parametri standardizzati, ovvero quando gli effetti diretti di differenti predittori sulla stessa risposta si presuma poter essere uguali; quando si analizza una matrice di covarianza, vincoli di

uguaglianza sono imposti solamente nel caso di soluzioni non standardizzate. Infine si potrebbe voler riportare la statistica test di significatività correttamente per le soluzioni standardizzate, ciò significa che è necessario possedere il valore corretto degli errori standard.

#### 4.4.6. Metodi di stima alternativi

Il metodo di stima della massima verosimiglianza lavora bene nel 90% o più dei casi di modelli a equazioni strutturali presenti in letteratura. Tuttavia è opportuno conoscere altri possibili metodi che possono essere utilizzati come completamento o in sostituzione di ML. Alcune di queste alternative sono delle opzioni da utilizzare quando non tutte le assunzioni fatte per il corretto funzionamento di ML non vengono rispettate anche solo in parte. Le due violazioni classiche sono quelle di distribuzione diversa dalla Normale e la presenza di variabili non continue, anche se in numerose discipline quest'ultime vengono approssimate al continuo. I metodi che verranno presentati sono generalmente iterativi, simultanei, a piena informazione e disponibili in molti software per la risoluzione di SEM.

Due metodi che prevedono l'utilizzo di variabili endogene con distribuzione normale multivariata sono: ordinary least squares (OLS) e generalized least squares (GLS). Il primo è un tipo di modello che minimizza la somma degli errori al quadrato, ovvero la differenza tra valori veri e valori fittati dal modello. Lo stimatore così generato risulta essere non distorto, ma non è efficiente quanto ML. La convenienza del metodo OLS è nel suo requisito che prevede che tutte le variabili osservate abbiano la stessa scala. Di conseguenza il modello non è né a scala libera né a scala invariante. Un suo potenziale vantaggio può derivare dalla non necessità di avere una matrice di covarianza definita positiva, a differenza di ML che senza questa non può arrivare a soluzione; inoltre è maggiormente robusto nella determinazione dei valori iniziali. A fronte di questa ultima considerazione si può concludere che OLS potrebbe essere utilizzato per stimare i valori iniziali che poi potranno essere utilizzati da una procedura di stima attraverso ML che di conseguenza convergerà a soluzione più facilmente. GLS invece fa parte di una famiglia più ampia, ovvero quella dei Weighted least squares (WLS); questi modelli possono essere utilizzati quando l'analisi deve essere fatta su dati che non presentano una distribuzione Normale. Al contrario di OLS, lo stimatore GLS non prevede la stessa scala per le variabili, di conseguenza sarà o a scala libera o a scala invariante, infatti è possibile affermare che asintoticamente, sotto l'assunzione di normalità, il metodo è uguale a ML. Un potenziale vantaggio rispetto a ML è che richiede un minor tempo di risoluzione e un minor tempo di computazione. Tuttavia ad oggi questo vantaggio non risulta più di significativa importanza in quanto gli elaboratori hanno sviluppato capacità computazionali che non rendono di difficile implementazione la procedura ML; infatti i processori risultano essere molto veloci e le espansioni di memoria poco costose. Si può concludere che in generale lo stimatore ML risulta da preferire a entrambi i metodi alternativi appena esposti.



#### 4.4.7. Altri Modelli

Oltre ai modelli già illustrati esiste un'altra classe di modelli che si basano su altre teorie per la risoluzione dei modelli implementati. Alcuni programmi infatti utilizzano per la risoluzione il modello di Bentler e Weeks del 1980. In questo modello ogni variabile, sia osservabile che latente, viene classificata come indipendente o dipendente. Una variabile dipendente ha come caratteristica quella di avere nella rappresentazione grafica un percorso monodirezionale che entra in essa; tutte le variabili di questo tipo vengono raccolte all'interno di un vettore indicato con  $\eta$ . Le variabili rimanenti sono quelle indipendenti e vengono raccolte in un altro vettore, individuato da  $\xi$ . È così possibile rappresentare le relazioni esistenti tra le variabili indicate nel modello attraverso la seguente relazione:

$$\eta = \beta * \eta + \gamma * \xi$$

dove  $\beta$  e  $\gamma$  contengono i valori non noti dei coefficienti del diagramma path; questi sono interpretabili come i coefficienti della regressione lineare, i pesi che ogni variabile assume all'interno della relazione con un'altra variabile all'interno del modello. In aggiunta a  $\beta$  e  $\gamma$ , è possibile definire  $\varphi$ , ovvero una matrice di covarianza delle variabili contenute in  $\xi$ . Viene utilizzata l'idea di equazioni simultanee, dove ognuna è specificata individualmente per ciascuna variabile utilizzando la notazione seguente:  $V$  indica una variabile misurabile,  $F$  un fattore,  $E$  l'errore residuale,  $D$  la disturbanza o fattore residuale. Si assume inoltre un caso in cui la confirmatory factor analysis (CFA) preveda due fattori e tre indicatori per ogni fattore. Ogni variabile osservabile ( $V$ ) viene considerata come una variabile dipendente, pertanto solo due tipologie di variabili possono avere un percorso direzionale verso queste: una fattore ( $F$ ) o un errore ( $E$ ). In questo modello CFA allora esistono sei variabili  $V$  nel vettore  $\eta$ , due fattori  $F$  e sei errori  $E$  nel vettore  $\xi$ . La matrice  $\beta$  in questo caso contiene solo zeri e di conseguenza può essere ignorata; al contrario la matrice  $\gamma$  contiene i fattori basilari e i coefficienti ad essa associati con gli errori residuali; infine  $\varphi$  è la matrice di varianza e covarianza di  $F$  e  $E$  delle variabili indipendenti. Basandosi sulle equazioni simultanee rappresentate dall'equazione illustrata sopra, la matrice di covarianza del modello, ovvero delle variabili osservabili  $V$ , può essere calcolata; il calcolo è illustrato di seguito. Con tutte le variabili osservabili trattate come variabili dipendenti, la matrice di covarianza del modello è la seguente:

$$\Sigma(\theta) = G(I - \beta) - 1\gamma\varphi\gamma'(I - \beta) - 1'G'$$

dove  $G$  è una matrice contenente 0 e 1 e scelta appositamente per selezionare le variabili osservabili da tutte quelle dipendenti presenti nel vettore  $\eta$ . L'equazione esposta sopra può essere espansa in  $p^2$  equazioni, ovvero una per ogni elemento presente in  $\Sigma(\theta)$ . È possibile però considerare solamente gli elementi non ridondanti della matrice  $\Sigma(\theta)$ , infatti questa risulta essere simmetrica, e di conseguenza derivare solamente  $p^*$  equazioni indipendenti contenenti  $q$

parametri da stimare; risolvendo simultaneamente le sole equazioni necessarie si otterranno i valori  $\hat{\theta}$ .

#### 4.4.8. La funzione di fitting

Una volta che le condizioni di identificazione sono soddisfatte per le  $p^*$  equazioni menzionate in precedenza, è necessario selezionare un criterio per ottenere un'unica soluzione per i parametri. Il criterio utilizzato per la stima dei parametri è anche conosciuto come funzione di discrepanza:

$$F = f(S, \Sigma(\hat{\theta}))$$

Questa ha l'obiettivo di suggerire una linea guida per minimizzare la differenza tra la matrice di covarianza della popolazione  $\Sigma$ , ottenuta come stima della matrice di covarianza  $S$ , e la matrice di covarianza ottenuta dall'ipotesi fatta con il modello  $\Sigma(\theta)$ . La minimizzazione della funzione di discrepanza può avvenire attraverso metodi differenti, a seconda delle assunzioni che vengono fatte sulla distribuzione delle variabili. Ad esempio è possibile minimizzare attraverso la massima verosimiglianza, ottenendo:

$$F_{ML} = \log|\Sigma(\theta)| + \text{Trace}[[\Sigma(\theta)^{-1}S] - \log|S| - p$$

Il valore dei parametri stimati viene ottenuto proprio attraverso la minimizzazione della differenza tra le due matrici; infatti per ogni soluzione verrà creata una matrice di covarianza e si sceglierà quella più vicina alla matrice di covarianza stimata direttamente sul campione della popolazione di interesse.

#### 4.4.9. Errori standard

Tipicamente gli errori standard vengono derivati dalla stima dei parametri e dalla funzione di discrepanza. Questa stima degli errori standard risulta essere buona quando il campione è di grandi dimensioni, se la specificazione del modello è corretta e se sono soddisfatte le assunzioni sulla distribuzione del campione stesso. Mentre per ADF gli errori standard possono essere corretti a prescindere dalla distribuzione, ciò non è necessariamente vero per gli errori calcolati attraverso la massima verosimiglianza o con il metodo dei minimi quadrati generalizzati. Infatti per quest'ultimi si troveranno errori standard con valori lontani dalla realtà poiché la variabilità del campione viene sottostimata quando le code della distribuzione sono più grosse di quelle di una normale.

#### 4.4.10. Conclusioni

Per la trattazione è stato utilizzato il concetto di equazioni simultanee per spiegare il processo della procedura di stima nei modelli SEM. La questione dell'identificazione, della specificazione del

modello, del test di ipotesi, della funzione di discrepanza e della statistica test per valutare la bontà del fit comprendono nella procedura di stima l'utilizzo di equazioni simultanee. Si è sottolineato poi come non sia semplice scegliere un metodo di stima di default per la risoluzione dei modelli che garantisca delle buone performance sia nei test da effettuare sul modello, sia nella stima degli errori standard, sia nella valutazione della significatività dei parametri. Le performance dei vari metodi dipendono dalla natura del modello e dei dati, in particolar modo dalla loro distribuzione e dalla dimensione del campione; è possibile affermare che quando questi hanno distribuzione normale multivariata e i campioni sono abbastanza grandi, i metodi ML e GLS sono sicuramente preferibili in quanto presentano una semplicità computazionale e un'accuratezza di stima che porta all'ottenimento della correttezza per i risultati cercati, ma d'altra parte quando i dati non sono normali la situazione cambia completamente.

## 4.5. Non normalità

Negli ultimi 30 anni l'uso di modelli a equazioni strutturali è aumentato significativamente in seguito alle sue applicazioni in ambito di scienze sociali e comportamentali. Il riconoscimento da parte dei ricercatori sui vantaggi di questo approccio e il desiderio di implementare questa metodologia potente dal punto di vista dell'interpretazione numerica, hanno però portato, oltre alla diffusione, a un uso inappropriato di questa tecnica. La più grande causa di usi inappropriati, che di conseguenza genera il fallimento di un modello, è legata al fattore scala e alle assunzioni di normalità su cui sono basate la fase di stima e di test. Gli approcci più comunemente utilizzati per la stima dei parametri nei modelli a equazioni strutturali, come la massima verosimiglianza e i minimi quadrati generalizzati, assumono che le variabili osservabili siano continue e con una distribuzione normale multivariata. Nella pratica però spesso i dati reali violano queste ipotesi creando dei problemi nelle applicazioni dei SEM. In molte delle aree in cui si sviluppano queste teorie sostantive, le variabili risultano essere spesso non continue, in particolar modo si presentano molte variabili binarie o categoriche. In altre aree ancora le variabili presentano una distribuzione continua, che però non è assimilabile ad una normale in quanto le code di tale distribuzione possono essere più o meno spesse, oppure esiste una certa asimmetria tra parte destra e sinistra. Troppo spesso però i ricercatori ignorano queste assunzioni e le loro possibili violazioni, di conseguenza trovano dei risultati non soddisfacenti che fanno considerare il modello non buono; il problema non risiede però nella struttura del modello, che risulta essere valida, ma bensì nella cattiva applicazione da parte dei ricercatori. Per esempio nel 1990 Breckler condusse uno studio su 72 articoli apparsi su giornali riguardanti queste aree di scienza trovando che solo il 19% degli autori di questi fossero a conoscenza delle assunzioni di normalità e che meno del 10% ne tenevano esplicitamente conto. Risulta così evidente che una cattiva applicazione porta al rischio di attribuire una valutazione scadente del metodo quando in realtà il problema risiede nell'utilizzo in modo errato. Dato che spesso i dati violano le assunzioni su cui il modello si poggia, cresce in modo evidente l'interesse nel determinare la robustezza delle tecniche SEM e nello sviluppo di strategie alternative che possano fornire rimedi quando le ipotesi sono violate in modo evidente.

### 4.5.1. Panoramica sulla teoria di stima normale

Come discusso in precedenza, l'obiettivo della stima è quello di minimizzare l'entità della differenza tra ogni elemento presente in  $S$  e l'elemento corrispondente in  $\Sigma(\hat{\theta})$ . Si ricorda che  $S$  è la matrice di covarianza calcolata direttamente sui dati osservati, mentre  $\Sigma(\hat{\theta})$  è la matrice di covarianza ottenuta attraverso il set di parametri stimati  $\hat{\theta}$  in seguito alle ipotesi fatte per il modello in riferimento alle relazioni esistenti tra le variabili presenti. D'ora in avanti i parametri stimati verranno raccolti in un vettore che verrà identificato con  $\theta$ . È già stato discusso come e perché le tecniche maggiormente utilizzate siano la massima verosimiglianza e i minimi quadrati generalizzati. Entrambe le tecniche sono basate sulle stesse assunzioni, rilasciano stime molto simili e presentano proprietà statistiche molto favorevoli. La funzione di fitting di GLS può essere espressa come

$$F_{GLS} = \frac{1}{2} * tr \left[ ([S - \Sigma(\hat{\theta})]W^{-1})^2 \right]$$

In questa equazione  $W^{-1}$  rappresenta una matrice di pesi e  $tr$  è l'operatore traccia, che prende la somma degli elementi sulla diagonale della matrice risultante dalle operazioni effettuate all'interno delle due parentesi quadrate grandi. La minimizzazione di questa funzione coinvolge la minimizzazione della discrepanza tra  $S$  e  $\Sigma(\hat{\theta})$  ottenuta considerando i pesi e poi elevando la matrice così acquisita al quadrato. Come tutti i membri delle procedure di minimi quadrati pesati, GLS richiede la selezione della matrice  $W$ . La scelta più comune risulta essere quella di ricavarla a partire dalla matrice di covarianza del campione e quindi si avrà  $W = S^{-1}$ ; in questo modo la discrepanza risulta essere in accordo con la varianza e la covarianza con gli altri elementi. Questa scelta si basa su due assunzioni: innanzitutto il valore  $E(s_{ij})$ , ovvero il valore atteso della covarianza di campione tra l'elemento  $x_i$  e  $x_j$ , si assume essere  $\sigma_{ij}$ , la varianza corrispondente della popolazione; in secondo luogo la distribuzione di  $S$  quando si lavora con campioni di grandi dimensioni è assunta essere una normale multivariata. Se queste assunzioni sono soddisfatte, GLS presenta delle proprietà statistiche allettanti, quali:

- Lo stimatore è asintoticamente non distorto, ciò significa che i parametri stimati attraverso campioni con dimensioni abbastanza grandi assumono in media il valore corrispondente della popolazione; il valore quindi non è né sovrastimato, né sottostimato, ovvero  $E[\hat{\theta}] = \theta$ .
- Lo stimatore è consistente, ovvero al crescere del campione il valore dei parametri stimati è sempre più vicino a quello vero della popolazione.
- Lo stimatore è asintoticamente efficiente, quindi all'aumentare della dimensione del campione la distribuzione dello stimatore presenterà varianza sempre minore.
- Risulta possibile utilizzare una statistica test plausibile per valutare la bontà del fit del modello, ovvero quanto questo riesce a descrivere i dati reali, poiché  $(N - 1)F_{GLS}$  segue una distribuzione  $\chi^2$ ; ciò è vero sempre in riferimento ad un campione di grandi dimensioni.

Si ricorda tuttavia che queste proprietà dello stimatore GLS rimangono contingenti rispetto alle severe assunzioni discusse in precedenza. Come si è visto una di queste assunzioni riguarda la grandezza del campione da impiegare, che deve essere abbastanza grande. Se le ipotesi non sono totalmente verificate, a livello teorico non è garantita la presenza di queste proprietà e di conseguenza la robustezza dello stimatore in un contesto del genere assume una notevole importanza per gli studi empirici.

## 4.5.2. Basi teoriche per il problema

Un potenziale problema nella stima di modelli a equazioni strutturali si presenta quando la distribuzione delle variabili osservabili differisce sostanzialmente da una normale multivariata. La stima dei parametri deriva dalle informazioni contenute in  $S$ , la matrice di varianza stimata del campione, e in  $W^{-1}$ , la matrice di pesi ottima. Quando le variabili osservabili sono (a) continue ma non normali, (b) binarie o categoriche, le informazioni presenti nelle due matrici possono essere non corrette; la conseguenza di questa informazione errata è una stima inesatta. Ma analizziamo i due casi citati sopra.

**Variabili continue ma non normali.** I risultati di simulazioni effettuate su dati non normali indicano che non è bene ignorare l'assunzione di normalità multivariata assunta di default nel metodo di stima ML. Infatti, quando le variabili risultano avere una distribuzione continua ma non Normale accade che i valori stimati risultano essere comunque adeguati se si lavora su campioni di grandi dimensioni, ma che gli errori standard stimati tendono ad essere troppo piccoli, forse di circa il 25-50%: allora il modello sta dicendo che fitta bene i dati perché gli errori sono bassi, ma nella realtà questi sono più grandi e di conseguenza il modello non è buono come sembra. Questo risultato si traduce nel rifiuto dell'ipotesi nulla, ovvero la conferma che il parametro caratterizzante della popolazione sia uguale a zero, anche se spesso risulta assumere proprio quel valore, producendo un errore di prima specie. Ma questo non è il solo problema, infatti è possibile che il valore della statistica test del fit del modello tende ad essere troppo elevata e si traduce nel rifiuto dell'ipotesi nulla sulla bontà del fit, ovvero che questo sia perfettamente in grado di adattarsi ai dati della popolazione più spesso di quanto in verità accada. Quindi modelli veri tendono a essere negati troppo spesso. Il tasso di questo tipo di errore risulta essere più elevato del 50%, quando invece ci si aspetta un tasso del 5% se la distribuzione è assunta come normale, nuovamente a seconda dei dati e del modello. Il test statistico più largamente utilizzato nell'ambito dei SEM viene descritto attraverso una statistica  $\chi^2$ , che verrà discusso in seguito. A seconda dello schema e della severità della lontananza della distribuzione vera da quella Normale, il valore della statistica  $\chi^2$  potrebbe essere troppo piccolo a favore di ciò che si aspetta il ricercatore; in altre parole utilizzare test statistici che si basano sull'ipotesi di normalità della distribuzione quando questa è evidentemente violata risultano essere privi di significato, quindi avranno poca fiducia.

Un'opzione per evitare che ci sia distorsione è quella di normalizzare i dati prima di utilizzarli per l'analisi del modello attraverso un'opportuna trasformazione, in seguito sarà possibile applicare in modo corretto il metodo di massima verosimiglianza. Un'altra opzione per trattare variabili

continue ma non normali è, al contrario, quella di operare una correzione al metodo di stima. Ciò significa che è possibile lavorare direttamente sui dati originari utilizzando direttamente ML ma preferendo i robust standard errors e una correzione nella statistica test. Gli errori si definiscono robusti in quanto la loro stima, rispetto agli errori standard normali, è supposta essere maggiormente forte. Per quanto riguarda invece le correzioni, l'esempio più conosciuto è quello proposto da Santorra e Bentler nel 1994: l'idea è quella di poter aggiustare il valore della statistica test  $\chi^2$  calcolata in modo standard con ML attraverso un indicazione fornita da i gradi di curtosi. Il metodo venne applicato a numerosi modelli fornendo dei risultati in buona parte favorevoli. L'analisi dei dati grezzi è indispensabile per comprendere come interpretare e nel caso correggere i risultati ottenuti attraverso una stima che assume di default la normalità. Dei vari metodi descritti per l'analisi di modelli contenenti variabili continue ma con distribuzioni non normali, risulta essere di più semplice applicazione la correzione del metodo normale rispetto alla normalizzazione dei dati di partenza.

La variazione nelle variabili osservabili è completamente sommata attraverso la covarianza del campione solamente quando è presente una distribuzione normale multivariata. Se è violata l'ipotesi sulla distribuzione, la variazione non viene totalmente sommata, si necessita allora delle informazioni riguardanti i momenti del quarto ordine della distribuzione. In questa situazione  $S^{-1}$  non è più lo stimatore corretto per  $W^{-1}$ . La stima dei parametri può ancora essere non distorta e consistente, ma non sarà più efficiente. Questo risultato suggerisce che dal punto di vista statistico occorrono due problemi importanti quando si lavora con variabili non normali all'interno di un metodo che al contrario la richiede. La soluzione fornisce un test  $\chi^2$  sulla bontà del fit che fornisce una valutazione inaccurata del fit stesso, portando così al rifiuto di molti modelli (>5%) che in realtà sono veri; inoltre ci si aspetta che tutti le stime dei parametri siano distorte, cedendo troppi risultati significativi.

**Variabili categoriche.** Le variabili endogene non sempre presentano una distribuzione continua, infatti queste potrebbero assumere solamente valori discreti. Ciò avviene ad esempio per la variabili binarie, categoriche o ordinali. Si pensi ad esempio ad una variabile che indichi il gradimento con dei valori finiti in una scala da 0 a 2: nella scala è possibile distinguere tre livelli di gradimento, ma è difficile affermare che esistano degli intervalli regolari tra i tre numeri assegnati ai rispetti livelli. Inoltre, valori di variabili con così pochi livelli non hanno sicuramente distribuzione normale, ma non possono nemmeno essere approssimati; nonostante non esista una regola specifica per stabilire quale sia il numero di livelli di avere per poter ipotizzare una distribuzione, è possibile affermare che con scale con meno di 15 livelli sicuramente l'operazione non può essere fatta. La problematica di avere variabili non continue risiede proprio nelle assunzioni che molti stimatore utilizzano per operare: è richiesta distribuzione continua e normale. I risultati ottenuti da simulazioni al computer indicano che una stima effettuata con ML su un modello contenente variabili binarie o ordinali risulta essere inaccurata. Ovviamente queste simulazioni di studio assumono generalmente di avere una misura della popolazione vera con indicatori continui. All'interno di campione generati, gli indicatori vengono categorizzati con lo scopo di approssimare i dati derivanti da una variabile non continua. Quando c'è un solo fattore

singolo nella popolazione di interesse ma gli indicatori hanno poche categorie, il modello di misura del fattore tende ad essere rigettato troppo spesso. In altre parole la categorizzazione potrebbe suggerire in modo spurio la presenza di fattori multipli. I parametri stimati con ML e i loro rispettivi errori standard sono generalmente troppo piccoli quando i dati analizzati provengono da indicatori categorici e quando il grado di distorsione diventa elevato poiché la distribuzione è significativamente non normale. In sostanza queste due teorie sono sufficienti ad affermare che lo stimatore ML non è da utilizzare per variabili binarie o categoriche poiché non risulta appropriato.

Esistono due opzioni analitiche da poter operare nel caso in cui si debba lavorare su questo tipo di variabili. La prima opzione include il caso in cui l'outcome risulta essere ordinale o categorico ma isolato da altre variabili di questo tipo, ovvero non è fuso o combinato con variabili simili. L'approccio a questa situazione richiede di operare con uno stimatore speciale apposito per la trattazione di questa tipologia di dati, proveniente dalla famiglia dei modelli WLS. La seconda opzione invece prevede pacchi di analisi. Un pacco è la somma totale all'interno di elementi omogenei indicanti lo stesso livello all'interno della scala. L'affidabilità del punteggio tende a crescere all'aumentare degli elementi individuali. Se la distribuzione di tutte queste parcelle è normale, allora è possibile trattarle con lo stimatore ML. Le particelle sono tipicamente specificate come indicatori continui sottostanti a variabili latenti nel modello. Ad esempio, si supponga di avere un questionario con 40 elementi di una scala di gradimento. Invece di analizzare tutti e 40 gli elementi come singole variabili di risposta, si vuole parzionare in due insiemi non sovrapposti di 20 elementi ciascuno. Gli elementi all'interno di ciascun set si presume che siano omogenei, che riflettano un dominio comune. Lo score totale di ogni insieme deriva dai 20 elementi all'interno di ciascuno: si otterranno così 2 score da analizzare invece di 40 elementi. Poiché i valori così trovati possiedono una distribuzione continua e normale, è possibile optare per il metodo ML per stimare i parametri del modello, al posto di WLS che risulta essere più complesso. Questo è il ragionamento alla base presente nella teoria del parcellizzare, teoria che però presenta qualche controversia. Infatti si assume che gli elementi all'interno di ogni pacco siano noti per poter misurare i singoli costrutti presenti nel modello. Questa conoscenza può venire dalla familiarità con il dominio degli elementi o i risultati a priori delle analisi statistiche, come l'esplorazione dell'analisi fattoriale. Non è raccomandabile parcellizzare se non viene assunta l'unidimensionalità; nello specifico l'operazione potrebbe non far parte dell'analisi volta alla determinazione se il set di elementi ha un'unica dimensione. Ciò è possibile poiché potrebbe mascherare una struttura di un fattore multidimensionale in modo tale da avere delle mancanze di specificazione del modello che può però tuttavia adattarsi ragionevolmente ai dati. Esistono due differenti possibilità per operare la suddivisione in pacchetti. Una include un'assegnazione random degli elementi nei vari pacchetti, l'altra prevede di formare dei gruppi basandosi su motivi razionali; ovviamente la scelta della modalità di divisione si rifletterà sulla procedura influenzando i risultati.

### 4.5.3. Rilevamento degli scostamenti dalla normalità

**Asimmetria e curtosi, univariate e multivariate.** È disponibile un buon numero di procedure per determinare l'univariata o multivariata normalità delle variabili osservabili. Queste procedure dipendono dal calcolo dei momenti di più alto ordine; tutti i momenti di una distribuzione sono definiti come  $\left(\frac{1}{N}\right) \sum (x - \mu)^k$ , dove  $N$  è il numero delle osservazioni nel campione,  $x$  è un valore osservabile,  $\mu$  è la media vera della popolazione e  $k$  è l'ordine che si vuole calcolare (ad esempio per calcolare il momento del primo ordine si prenderà  $k = 1$ , mentre per momenti del secondo ordine si avrà  $k = 2$ , ecc.). Quando la normalità univariata è soddisfatta sono necessari solo i primi due momenti per descrivere in modo esaustivo la distribuzione della variabile osservabile, ovvero solo media e varianza rispettivamente poiché gli altri momenti sono standard e in particolare si ha il momento di ordine 3 pari a 0 (ad indicare la simmetria) e il momento di ordine 4 pari a 3 (per indicare la specifica grandezza delle code, ovvero la curtosi). Distribuzioni univariate che deviano dalla normalità possiedono una variazione nei momenti superiori al secondo, ovvero un diverso grado di simmetria e di curtosi. Trovare un valore diverso da zero per un momento del terzo ordine è segnale della presenza di asimmetria: valori negativi indicano che la distribuzione presenta un allungamento nella parte sinistra, al contrario valori positivi suggeriscono un'estensione nella coda di sinistra (relativamente a una distribuzione normale simmetrica). La curtosi è di particolare importanza nell'ambito dell'inferenza statistica, infatti identifica la misura in cui il peso della curva, ovvero la probabilità di densità, differisce da quella di una curva normale. Una curtosi maggiore di 3 è associata ad una distribuzione avente code lunghe e sottili, quindi una forma abbastanza appuntita; mentre curtosi minore di 3 indica code corte e spesse quindi una forma più schiacciata rispetto ad una normale. Molti pacchetti software, dato un insieme di osservazioni, forniscono le informazioni sui momenti del terzo e del quarto ordine. L'esame di asimmetria e di curtosi di una distribuzione univariata fornisce solamente un controllo iniziale per la normalità multivariata. Se la distribuzione di una variabile differisce significativamente da una normale univariata, sicuramente la distribuzione multivariata non potrà essere normale. Tuttavia il contrario non è vero: teoricamente la totalità delle distribuzioni univariate può essere normale, ma la distribuzione congiunta può essere sostanzialmente multivariata non normale. Di conseguenza è importante esaminare le misure di asimmetria e di curtosi per le multivariate sviluppate da Mardia. Queste funzioni costruite per il calcolo del terzo e del quarto ordine, che possiedono distribuzione normale standard approssimativa, permettono così di effettuare test sulla asimmetria e la curtosi nel caso di distribuzioni multivariate. Nell'applicazione dei modelli SEM risulta di particolare importanza la misura della curtosi.

**Outliers.** Gli outliers sono punti estremi presenti in un set di dati; questi possono creare dei difetti nella soluzione dei modelli a equazioni strutturali, anche quando la restante parte dei dati è distribuita secondo le assunzioni che richiede la maggior parte dei metodi statistici visti finora. Gli outliers tipicamente sono dovuti a errori in risposta da parte di soggetti presso i quali si stanno raccogliendo i dati, o a causa di un'errata registrazione del dato, o ancora perché potrebbero far parte di una popolazione diversa da quella da cui si sta estraendo il campione. Gli effetti provocati da questi dati sugli indici del fit del modello, sulla stima dei parametri e sul calcolo degli errori



standard possono essere molto rilevanti. Punti di outliers possono anche essere potenzialmente causa di soluzioni improprie, in cui i parametri stimati risultano essere fuori dal range di valori accettabili; si ricorda l'esempio di una varianza stimata che assume un valore minore di zero, ovviamente inaccettabile per come è definito il parametro. È possibile affrontare la questione in modi diversi per correggere la situazione verificatasi: controllo e correzione dei dati in fuori controllo, eliminazione dei dati non corretti, ridefinizione della popolazione di interesse o della specificazione del modello attraverso opportuni rimedi ottenuti studiando la fonte che genera i punti anomali stessi. In un contesto di modelli a equazioni strutturali gli approcci utilizzati per rilevare outliers sono prevalentemente due. Il primo, un approccio a modello indipendente, consiste nell'identificare i casi devianti il cui valore diverge in modo evidente dalla massa di punti: nel caso di distribuzioni univariate ciò è possibile attraverso una valutazione visiva del plot ottenuto da tutte le variabili misurabili, in sostanza si vogliono trovare quei punti che hanno una grande deviazione standard rispetto alla media della distribuzione e che sono lontani dalle altre osservazioni; per distribuzioni multivariate invece è possibile trovare i punti in fuori controllo utilizzando un calcolo statistico come la distanza di questi dalla media grazie ai principali pacchetti diagnostici di regressione che identificano questa grandezza in uno spazio multivariato. Alternativamente è proposta l'identificazione dei punti come causa della variazione dell'indice di curtosi (sempre nell'ambito multivariato). Tipicamente tutte le variabili misurate devono essere considerate insieme durante il processo di analisi. Il secondo approccio propone di identificare i punti di dati osservati che risultano essere estremi rispetto al loro valore predetto sulla base della specificazione del modello; in altre parole i dati iniziali vengono presi tutti in considerazione per stimare i parametri del modello, che una volta ottenuti permetteranno di calcolare i dati stimati da confrontare con quelli veri: i punti in cui la distanza sarà notevole saranno gli outliers del set di dati. A questo proposito Bollen e Arminger (1991) hanno proposto un modello basato su uno score dei fattori, che rappresenta il punteggio previsto di ciascun caso sul fattore ipotetico. Questi score dei fattori, a loro volta, vengono utilizzati per stimare un set di score stimati sulle variabili osservabili per ogni caso. I residui rappresentano la distanza tra score osservato e score stimato per ogni variabile misurata; questi vengono standardizzati (avranno quindi media pari a zero e deviazione standard pari a 1) e in seguito plottati in un grafico per esaminare dal punto di vista visivo i possibili punti in fuori controllo.

#### 4.5.4. Risultati di studi empirici sulla non normalità

**Variabili continue ma non normali.** Molte simulazioni sono state effettuate allo scopo di valutare le performance delle teorie basate sull'assunzione di normalità in presenza di variabili con distribuzione differente. In particolare i metodi studiati sono quello della massima verosimiglianza e dei minimi quadrati generalizzati, i cui casi affrontati per determinarne la robustezza sono avvenuti in diverse condizioni di non normalità e con dimensioni campionarie differenti. In questi studi, il valore di ogni parametro è impostato a un valore noto della popolazione. Questo valore viene comparato con la media delle diverse stime empiriche con lo scopo di studiare l'effetto che viene generato da un determinato livello di non normalità nella distribuzione dei dati. Si sono

raggiunte le conclusioni seguenti. Sia lo stimatore ML che GLS forniscono un valore della statistica test  $\chi^2$  che diventa troppo grande quando i dati cominciano a diventare non normali. Un esperimento da riportare è quello condotto da Curran, West e Finch nel 1994 dove vengono investigati attraverso un modello di analisi fattoriale tre fattori e nove indicatori, in cui ogni variabile osservabile presenta una forte non normalità (asimmetria pari a 3 e curtosi pari a 21); lo scopo era quello di comparare il valore atteso della statistica  $\chi^2$  pari a 24 con la media delle 200 simulazioni effettuata che è risultata essere pari a 37,4, ovvero una sovrastima del 50% circa. Da notare che il calcolo è avvenuto utilizzando un campione di dimensione 1000 per ogni simulazione. Infatti anche dalla comparazione dell'errore di prima specie si trovano le stesse conclusioni: invece che un errore del 5% atteso, si trovava un valore vero pari al 48% facendo la media dei valori ottenuti nelle singole simulazioni; in altre parole nel 48% dei casi un modello vero veniva rifiutato. Oltre a trarre delle conclusioni sul comportamento dei modelli nei casi di crescente non normalità, si è svolto uno studio anche per capire come la stessa statistica si modificava al diminuire delle dimensioni del campione. L'applicazione ha portato a concludere che GLS, ma in particolare ML, producono un valore per il quale cresce leggermente quando la dimensione del campione diminuisce, anche se è presente la normalità di tutte le variabili. Nello stesso studio del 1994 Curran trova che quando il campione è composto da 50 osservazioni e le variabili osservabili hanno distribuzione normale multivariata, la media della statistica di 200 osservazioni è 26,7 (con una sovrastima di circa il 10%) e il 12% dei modelli nella popolazione viene rifiutato erroneamente. Le simulazioni condotte da Anderson e Gerbing nel 1984 hanno inoltre confermato che se contemporaneamente decresce la dimensione campionaria e aumenta la non normalità, queste portano a crescere in proporzione dell'analisi i casi in cui non si converge a soluzione e quelli in cui la soluzione viene trovata ma risulta essere impropria.

Oltre alle conseguenze sulla statistica test  $\chi^2$ , è possibile capire cosa accade agli errori standard. La non normalità tende a portare una sottostima degli errori standard di diversa entità, dal moderato al notevole. Con gli stessi dati mostrati all'inizio di questa trattazione, ovvero non normalità con asimmetria pari a 3 e curtosi pari a 21, gli errori standard di correlazione tra i fattori venivano sottostimati di circa il 25%, mentre gli errori standard dei fattori caricati venivano anch'essi sottostimati ma per il 50% approssimativamente. Tale sostanziale sottostima negli errori implica che i test sulle stime dei parametri non potranno essere degne di fiducia sotto le condizioni di non normalità.

**Variabili categoriche.** Molti studi sono stati condotti anche per verificare le performance degli stessi stimatori trattati precedentemente in casi in cui le variabili misurabili fossero di distribuzione normale ma suddivise in categorie ordinali. Ancora una volta sono state utilizzate diverse simulazioni di analisi fattoriale e hanno fornito le seguenti conclusioni. Innanzitutto il numero di categorie ha di per se un piccolo impatto sulla statistica valutatrice della bontà del test  $\chi^2$  quando la distribuzione delle variabili categoriche può essere approssimata a una normale. Come la distribuzione delle variabili categoriche inizia a differire dalla normale, ad esempio il caso di variabili asimmetriche in sensi diversi, i valori iniziano a essere gonfiati, ingranditi. La correlazione dei fattori è solo modestamente sottostimata purché ancora una volta la

distribuzione della variabili categoriche sia approssimabile ad una normale; tuttavia la sottostima inizia a crescere evidentemente quando sono presenti poche categorie (ad esempio due o tre), quando l'entità della asimmetria cresce e quando si verifica una differenza del grado di asimmetria tra le variabili. Inoltre la stima delle varianze degli errori è particolarmente distorta rispetto agli altri parametri stimati a causa dell'influenza delle condizioni per la correlazione tra fattori; connessamente, la correlazione potrebbe essere spuria se ottenuta attraverso la varianza degli errori associata ai termini aventi gradi di asimmetria simili. Quando il numero delle categorie è ristretto, il grado di asimmetria è determinato dalla percentuale di soggetti in accordo con l'elemento. Così un set di elementi con stessa percentuale di gradimento può aumentare il fattore spurio riflettendo solamente il grado comune di asimmetria tra gli elementi.

Per quanto riguarda degli errori standard, come visto anche nel caso precedente di non normalità, la stima per ogni parametro risulta essere troppo bassa, in particolar modo quando le distribuzioni presentano asimmetrie diverse; ciò significa che il test sui parametri non può essere considerato affidabile.

#### 4.5.5. Tecniche di stima alternative

Come si è visto il problema della non normalità può crescere in due contesti differenti: presenza di variabili continue ma con una distribuzione povera, lontana dalla normale, oppure variabili continue che vengono categorizzate. La stima di base rimedia a questi due problemi in modo differente; tuttavia queste tecniche condividono lo stesso scopo, ovvero quello di ottenere una statistica test  $\chi^2$  e una stima degli errori standard che sia più vicina possibile al valore vero.

**ADF.** Browne (1984) sviluppa uno stimatore alternativo che non presenta la necessità di assumere che le variabili osservabili avessero distribuzione normale multivariata. Questa stima di "distribuzione asintoticamente libera" si basa sul calcolo di una matrice generale di pesi,  $W$ , e sullo stimatore GLS. La chiave dello stimatore ADF è l'utilizzo della matrice ottima di pesi che consente di combinare termini del secondo e del quarto ordine.  $W$  è una matrice di covarianza degli elementi di  $S$ , la quale contiene a sua volta sia le varianze che le covarianze. Quindi la matrice dei pesi in ADF ha un numero maggiore di elementi rispetto a quella che viene costruita in GLS; tuttavia questa ha la proprietà desiderabile di poter essere semplificata alla matrice di base  $W = S^{-1}$  sotto le condizioni di normalità, ovvero quando il momento del quarto ordine risulta essere nullo. A causa di questo collegamento con la funzione di fitting di GLS, ci si riferisce talvolta a questo stimatore come AGLS, ovvero arbitrary generalized least squares. ADF produce una stima asintoticamente non distorta, quindi funziona bene per campioni di grandi dimensioni; ciò avviene sia per la bontà del fit, sia per i parametri, sia per gli errori standard. Sono questi i più grandi vantaggi dello stimatore ADF in relazione ai modelli basati su teorie normali, come ML e GLS, che al contrario producono stime distorte quando la normalità viene violata. Tuttavia lo stimatore è caratterizzato da due grandi limiti nella pratica. In primo luogo è dal punto di vista computazionale impegnativo, infatti il calcolo della funzione di fitting richiede l'inversione della matrice dei pesi ottima. Basta pensare che in un caso di analisi fattoriale  $W$  è una matrice di dimensione  $p^* * p^*$ ,

dove  $p^* = \frac{1}{2} * p * (p + 1)$ ; per esempio con 15 variabili osservabili sarà necessario invertire una matrice di dimensione 120x120, contenente quindi 14400 elementi. Di conseguenza già con un numero di variabili maggiore di 20, l'implementazione del modello diventava complicata quando ancora la tecnologia non aveva fatto grandi progressi. Il secondo problema deriva dal fatto che per calcolare la matrice dei momenti del quarto ordine si necessita di un campione di grandi dimensioni per avere la sicurezza della stabilità delle stime; allora la ricerca dei dati diventa sicuramente più complessa e crea un limite all'applicazione dello stimatore.

**Statistica  $\chi^2$  riscalata e errori standard robusti.** Sebbene la statistica  $\chi^2$  utilizzata dalle teorie basate sulla normalità dei dati, non segue le aspettative quando ci sono le condizioni di non normalità, è possibile correggerla, riscalarla, per ottenere un'approssimazione che faccia riferimento alla formulazione statistica di partenza. Santorra e Bentler (1990) ha sviluppato una teoria statistica sottostante questo ridimensionamento. La teoria normale di base (ad esempio con ML o GLS) è divisa da una costante  $k$ , il cui valore è funzione della matrice dei pesi implicita del modello, dalla curtosi multivariata osservata e dai gradi di libertà del modello stesso. Come i gradi di curtosi presenti nella distribuzione iniziano ad aumentare, anche  $k$  aumenta, portando successivamente ad un evidente aggiustamento verso il basso del valore della statistica. Lo stesso ragionamento può essere applicato al calcolo degli errori standard robusti; questi errori possono essere considerati teoricamente come un aggiustamento proprio a causa di una modifica della distribuzione in particolare riferito al momento di quarto ordine, ovvero ai gradi di curtosi presenti.

**Bootstrapping.** I metodi statistici moderni di computazione forniscono un approccio completamente differente al test per la valutazione della bontà del fit e per la stima dei parametri. Anziché fare affidamento su una distribuzione teorica, come fanno i test statistici classici, è possibile immaginare di prendere ripetutamente dei campioni dalla stessa popolazione di interesse. Per ogni campione, è necessario calcolare la stima dei parametri di interesse che risultano da una distribuzione empirica del campione stesso. Nei casi in cui le assunzioni dei classici test statistici vengano violate gravemente, la distribuzione empirica che descrive quella attuale è sostanzialmente più accurata di quella teorica che si potrebbe ipotizzare. Efron e Tibshirani hanno mostrato come la distribuzione ottenuta empiricamente dai campioni può essere ragionevolmente approssimata a quella di base grazie all'utilizzo di un solo campione. Nella procedura di bootstrapping, vengono presi ripetuti campioni della stessa dimensione da un gruppo originale di dati utilizzando il riposizionamento dopo ogni estrazione. Si pensi ad una situazione in cui il gruppo di base è composto solamente da quattro possibili casi (1,2,3,4) e si ipotizzino tre possibili estrazioni dal campione originale, ovvero (1,4,1,1), (2,3,1,3) e (4,2,2,4). Si nota come gli elementi possano essere ripetuti in ogni campione bootstrap e che questi sono della stessa dimensione come il campione originale. Prendendo un numero di campioni abbastanza grande, è possibile determinare la media e la varianza della distribuzione in modo empirico. Questo approccio risulta essere di facile comprensione ma anche di facile computazione, inoltre è crescente la disponibilità di effettuare il ricampionamento da parte dei software, anche nei pacchetti di risoluzione dei modelli ad equazioni strutturali. Esistono tuttavia due complicazioni

nell'applicazione. Innanzitutto, come enfatizzano Bollen e Stine (1992), il successo della procedura di bootstrapping dipende dal comportamento nel campionamento di una statistica, essendo la stessa quando i campioni sono tratti dalla distribuzione empirica e quando vengono presi dalla stessa popolazione originaria. In secondo luogo il bootstrapping è più spesso utilizzato in modo utile per comprendere una porzione o una trasformazione della statistica di interesse. L'approccio bootstrapping utilizzato nella statistica  $\chi^2$  per la bontà del test per un modello correttamente specificato in un'analisi fattoriale si è dimostrato che spesso fornisce risultati inaccurati sotto le condizioni di normalità multivariata. Anche con un modello propriamente specificato all'interno di una popolazione, il campione originale può riflettere delle fluttuazioni, ad esempio  $s_{ij} \neq \sigma_{ij}$ . Il valore atteso della statistica  $\chi^2$  per il set dei campioni costruiti con questo metodo dal campione originale presenta tipicamente una differenza dalla stessa statistica calcolata però nel modo tradizionale. Conseguentemente, la distribuzione con questo approccio seguirà una distribuzione  $\chi^2$  non centrata, riflesso delle fluttuazioni presenti nel campione, anziché la solita distribuzione specificata dalla teoria statistica. Allora viene presentata una trasformazione che consiste in una complessa funzione dei dati originali nel campione e la loro matrice di covarianza che minimizza il problema. La valutazione da parte di Bollen e Stine (1992) mostra anche una comparazione ragionevole delle performance dell'approccio rispetto ai valori attesi dalla trattazione della teoria statistica sotto le condizioni di normalità multivariata per la statistica test  $\chi^2$  e gli errori standard degli effetti diretti e indiretti.

**Studi empirici di procedure di stima alternative.** Vari studi di simulazione sono stati condotti per determinare le performance di ADF, di  $\chi^2$  riscalata e di altri approcci ancora; da notare che invece non molte simulazioni sono state condotte per valutare il metodo di bootstrapping considerando diverse distribuzioni non normali. Per i primi due metodi si è giunti alle seguenti conclusioni:

- Tutte le simulazioni hanno riportato che ADF produce un valore della statistica  $\chi^2$  che risulta essere troppo elevato quando le dimensioni del campione sono piccole. Anche la procedura di ridimensionamento produce un valore maggiore di quello atteso, ma in modo decisamente più lieve. Infatti la sovrastima risulta essere rispettivamente all'incirca 50% e 5%, con un rifiuto di modelli veri nella popolazione più contenuto passando dal 68% di ADF al 10% del ridimensionamento. Di conseguenza è possibile affermare che per ottenere della performance adeguate è necessario lavorare con campioni di grandi dimensioni; sotto le condizioni di non normalità è necessario un campione di dimensione 1000.
- Finch, Curran e West (1994) hanno mostrato come in un campione formato da 100 osservazioni e con distribuzione fortemente non normale, ovvero con asimmetria pari a 3 e curtosi a 21, lo stimatore ADF sottostima gli errori standard in modo evidente.

#### 4.5.6. Ridefinizione di variabili

Un approccio alternativo è quello di definire in modo differente le variabili continue ma con distribuzione non normale, in modo da produrre una distribuzione che sia più vicina possibile ad

una normale. Le variabili ridefinite possono poi essere analizzate utilizzando le tecniche di stima normali (come ML o GLS) senza che queste rilascino stimatori o errori standard distorti. È necessario però definire delle metodologie per poter ridefinire queste tipologie di variabili.

**Gruppi di elementi.** Un metodo semplice e comunemente utilizzato è la costruzione di gruppi di elementi, su quali devono essere fatte delle operazioni; gli elementi possono essere sommati oppure è possibile farne la media producendo in questo modo una misura per il costrutto. Questi gruppi tipicamente presentano una distribuzione più vicina alla normale di quanto facciano gli elementi originari. Un altro vantaggio forse meno ovvio dei gruppi costituiti da elementi è che è necessario stimare un numero minore di parametri nel modello, ciò implica che le stime saranno più stabili anche quando i campioni sono di piccole dimensioni. Tuttavia bisogna notare che la costruzione dei gruppi presenta anche degli svantaggi. Un problema importante deriva dal fatto che il valore del nuovo costrutto potrebbe oscurare il fatto che più di un fattore possono essere alla base di un dato gruppo di elementi; questo problema conduce a considerare una potenziale complicazione nell'interpretazione delle relazioni e delle strutture del nuovo modello. Inoltre, l'uso di troppi pochi gruppi (diventati nel nuovo modello le variabili osservabili) come indicatori del costrutto producono test meno stringenti, rigorosi, che risultano essere così poco necessari nella fase di valutazione del modello. Anche i problemi di identificazioni possono verificarsi se vengono utilizzati troppi pochi gruppi di elementi per ogni fattore (meno di 3); in questi casi calcolando la correlazione è possibile ottenere un valore vicino allo zero, che indica indipendenza tra le variabili e porta ad avere un modello non identificato.

**Trasformazione di variabili non normali.** Una trasformazione esegue un'operazione su una variabile osservata preservando l'ordine dei valori, ma alterando la distanza presente tra valori vicini. Trasformazioni lineari, come la standardizzazione, non hanno effetto sulla distribuzione delle variabili o sui risultati dei modelli a equazioni lineari. Al contrario le trasformazioni non lineari possono potenzialmente alterare la distribuzione delle variabili osservabili come anche le relazioni esistenti tra queste, eliminando alcune forme di effetti curvilinei e le interazioni tra le variabili stesse. Nel seguito si assumerà che tutti i valori trasformati delle variabili osservabili siano maggiori di zero, una condizione che può essere ottenuta semplicemente aggiungendo una costante ad ogni osservazione. Sono disponibili due classi di approcci da cui selezionare la trasformazione appropriata. Nella prima classe di approcci, la funzione identificata da applicare alla variabile deve produrre una nuova variabile (trasformata) che abbia distribuzione più vicina alla normale della variabile di partenza. Molti studi hanno fornito delle regole da visionare per selezionare la funzione di trasformazione più potente. Tipicamente dà come risultato delle variabili trasformate ben approssimate alla normale l'applicazione di una simmetria positiva, di un logaritmo, di una radice quadrata o il calcolo del reciproco; dare una asimmetria negativa nella distribuzione spesso si tradurrà in una variabile trasformata con distribuzione più normale. Emerson e Stoto (1983) presentano una tecnica in cui funzioni di score vengono associate con determinati livelli di percentile e di seguito plottati; la pendenza delle curve nel grafico risultante permettono di identificare la trasformazione più potente. Una seconda classe di approcci è utile quando il grafico indica una possibile relazione non lineare tra delle variabili; la funzione Box-Cox

suggerisce una soluzione quando è presente nel problema una regressione non lineare: vengono ricercati i parametri che rendono vera l'equazione  $y^\lambda = b_0 + b_1 * x$ . La pendenza ( $b_1$ ) e l'intercetta ( $b_0$ ) vengono stimate simultaneamente con la funzione ottima ( $\lambda$ ) per la variabile dipendente. In pratica molte equazioni di regressione con valori di  $\lambda$  compresi in un range tra -2 e 2 vengono calcolate e da queste si sceglie quella che fornisce il massimo valore di  $R^2$ . Un altro approccio recente di esplorazione fa un passo ulteriore trovando la trasformazione di ogni variabile che produce ancora una volta il massimo valore  $R^2$  possibile tra due variabili: è l'algoritmo ACE, ovvero Alternating Conditional Expectation. Questo trova la trasformazione ottima che massimizza la relazione lineare presente tra le due variabili  $x$  e  $y$ , anche quando la trasformazione fallisce. Gli approcci Box-Cox e ACE hanno una considerevole potenza quando vengono applicati a equazioni di regressioni semplici; tuttavia gli analisti di modelli a equazioni strutturali ricordano che cercano una singola trasformazione che sia applicabile attraverso una serie di equazioni di regressione, alcune delle quali coinvolgono variabili latenti. Di conseguenza questi due approcci devono essere visti come fornitori di una guida, anziché una soluzione definitiva alla ricerca di una trasformazione singola che migliorerà la linearità del set delle relazioni coinvolte nella problematica iniziale.

Molte osservazioni possono essere fatte a riguardo delle trasformazioni. Innanzitutto, la curtosi e l'asimmetria univariate dei dati trasformati potrebbero sempre essere esaminate per apportare miglioramenti nella distribuzione di nuove variabili; questi indici sono spesso utili nella scelta tra trasformazioni in competizione. Da notare che per alcune distribuzioni di variabili osservabili non esiste sempre una funzione di trasformazione semplice che riduce in modo sostanziale l'asimmetria o la curtosi. In secondo luogo, la misura dell'asimmetria e della curtosi multivariate per le variabili originali e per quelle trasformate può essere comparata. Si ricordi che buoni comportamenti di distribuzioni univariate sono solo condizioni necessarie, ma non sufficienti per stabilire la normalità multivariata. Ancora è possibile notare che, nonostante il secondo approccio di trasformazione aumenta la linearità della relazione, non porta direttamente verso la normalità, ma linearizzando spesso si ottengono dei benefici aggiuntivi di miglioramento della distribuzione e di omoschedasticità nella misura degli errori. Inoltre le trasformazioni dei dati cambiano la variabile originale  $y$  in una nuova variabile  $y^*$ ; allora esisterà una nuova correlazione tra le variabili trasformate, diversa dalla correlazione esistente tra quelle originali. Riflettendo questo cambiamento, la statistica di fit, la stima dei parametri e degli errori standard basati sulle variabili trasformate potranno essere differenti da quelli che hanno come base le variabili originali, anche in modo sostanziale. Infine, l'applicazione dell'algoritmo ACE ad ogni variabile osservabile o di differenti trasformazioni su ogni variabile può potenzialmente portare ad una difficoltà nell'interpretazione del modello. Ancora maggiori problemi di interpretazioni possono nascere quando differenti trasformazioni vengono applicate alla stessa variabile osservabile all'interno di studi diversi. Questo è un particolare problema che si presenta durante l'utilizzo di ACE a causa della sua forte tendenza a capitalizzare possibilità di relazione che non sono attese in quanto non evidenziate durante lo studio del fenomeno in oggetto. In generale, la mancanza di una metrica da associare alle variabili trasformate è una questione da estendere, da approfondire, per poter in seguito confrontare i risultati derivanti da diversi studi. Inoltre, la metrica originale della variabile

osservata può rappresentare un'importante informazione in alcune aree delle scienze sociali; tuttavia in altre aree è più spesso assegnata una metrica arbitraria, così che diventi meno cruciale il preservamento della scala di misura.

### 4.5.7. Conclusioni

Gli effetti della non normalità nei modelli a equazioni strutturali dipendono in ugual modo sia dalla sua estensione, sia dalle sue cause. In generale però è possibile affermare che maggiore è l'estensione della non normalità nel modello, maggiore sarà il problema da considerare. La trattazione ha evidenziato quali siano gli effetti della non normalità sulla statistica  $\chi^2$  necessaria a valutare la bontà del fit, sulla stima dei parametri e degli errori standard; problemi che presentano delle rilevanti implicazioni pratiche. L'ottenimento di una statistica gonfiata per la bontà del fit porterà a ricercare una soluzione migliore per il modello operando delle modifiche anche non appropriate. La sottostima degli errori standard produce percorsi significativi e correlazione tra i fattori, nonostante queste non esistano nella popolazione originale. Questi risultati fanno pensare che si verificheranno dei fallimenti nelle repliche successive, aumentando maggiormente la confusione all'interno dell'area di ricerca.

La scelta riguardo le misure di rimedio nuovamente dipendono dall'intensità della non normalità, dalle cause che la provocano, ma anche dalla dimensione del campione. Considerando inizialmente la problematica della bontà del fit e degli errori standard per variabili continue ma con distribuzione non normale, sia lo stimatore ADF che lo scaled di Santorra-Bentler hanno mostrato di operare ottenendo delle buone performance, indipendentemente dal grado di deviazione dalla normalità se i campioni hanno dimensioni grandi e il modello è stato specificato correttamente. Bisogna però discutere di cosa significhi "campione di dimensioni grandi", su questo si effettuano molti studi e con certezza è possibile riferirsi ad un range che va dai 1000 alle 5000 osservazioni per campione; se il numero delle osservazioni risulta essere minore bisogna valutare attentamente ogni statistica singolarmente. Nel caso di piccoli campioni è raccomandabile utilizzare ugualmente gli stimatori con teoria normale, come ML e GLS, se la distribuzione ha una leggera deviazione dalla normale; mentre è preferibile Santorra-Bentler quando la non normalità si fa significativa. Sotto queste condizioni, prendere in considerazione l'utilizzo di un livello  $\alpha$  più stringente per i test sui parametri potrebbe risultare utile. In particolare per i campioni di piccole dimensioni, può essere utile considerare la ridefinizione delle variabili per ottenere variabili trasformate con distribuzione più vicina alla normale di quanto sia quella della variabili originali, sulle quali poter applicare in seguito le tecniche basate su teorie normali. Tra i metodi esposti è presente la costruzione di gruppi di elementi, metodo che sicuramente porta ad una riduzione dei dati su cui far girare il modello, ma anche dei parametri stimando, fornendo un più favorevole rapporto parametri-osservazioni, qualità indispensabile nei piccoli campioni. L'identificazione della trasformazione ottimale è sicuramente meno certa in piccoli campioni piuttosto che in quelli di dimensione maggiore; l'individuazione di una trasformazione adeguata che sia soddisfacente per la stima con teorie normali può essere raggiunta in alcuni set di dati, ma non in tutti. Ogni metodo per la ridefinizione di variabili presenta degli svantaggi: i gruppi di



elementi costruiti con i dati originali possono oscurare le strutture multifattoriali; inoltre la mancanza della metrica originale rende complicata l'interpretazione dei risultati ottenuti. Da notare che solo pochi lavori empirici sono stati specificamente affrontati per determinare gli effetti di questa tecnica sui risultati di un modello a equazioni strutturali.

## 4.6. Estensione: variabili categoriche

I modelli SEM con variabili latenti vanno rivisti per le situazioni che includono la presenza di variabili binarie, ordinali e categoriche. La metodologia disponibile fino ad oggi era da intendersi applicabile ai soli casi in cui ci fossero indicatori continui. Sviluppi riguardanti indicatori categorici sono importanti in molte applicazioni, in particolar modo nelle scienze sociali e comportamentali, dove le variabili osservabili spesso hanno un piccolo numero di categorie con delle distanze non equidistanti delle diverse classi fino ad essere binarie. Quindi, le categorie di queste variabili possono essere considerate nel seguito della trattazione come continue, ma con intervalli di scala variabili. I coefficienti di correlazione e di covarianza di Pearson sono tuttavia inadatti per queste variabili quasi continue, in modo evidente quando è presente una asimmetria nella distribuzione. Quando queste vengono forzate all'interno di un modello a equazioni strutturali tradizionale l'analisi risultante sarà distorta.

Come visto precedentemente la presenza di variabili categoriche produce stimatori distorti non solo nella statistica  $\chi^2$  e negli errori standard, ma anche nella stima dei parametri stessi. Muthén (1983) ha sviluppato uno stimatore alternativo che prende il nome di stimatore CVM, ovvero (for continuous/categorical variable methodology). Questo stimatore permette l'analisi di tutte le combinazioni di variabili binarie, categoriche e ordinali, all'interno di un modello. A differenza dei metodi basati sulla teoria normale tradizionali, lo stimatore CVM riesce a fornire una stima dei parametri non distorta, consistente ed efficiente anche quando le variabili hanno distribuzione discreta. L'approccio si basa su una assunzione forte: una variabile latente con distribuzione normale standard ( $\mu = 0$  e  $\sigma^2 = 1.0$ ) denominata  $y^*$ , è assunta come sottostante di ogni variabile misurabile  $y$ . Per le variabili binarie, la risposta "sì" può essere osservata se la posizione dell'individuo sulla sottostante  $y^*$  normalmente distribuita ha una dimensione maggiore del valore soglia stabilito; la risposta "no" al contrario la si osserva quando la dimensione assunta è minore della soglia. Generalizzando il discorso per variabili categoriche o ordinali, è possibile affermare che una volta stabilite le soglie per ogni valore, si guarderà l'entità della variabile osservabile e si otterrà la sottostante confrontando il valore con quello delle soglie stabilite. Nel caso di variabili osservabili continue,  $y$  e  $y^*$  sono assunte come equivalenti. Dato che la variabile  $y$  è considerata come solo un'approssimazione della variabile normale sottostante  $y^*$ , è necessario tracciare una distinzione tra la struttura di covarianza delle due tipologie di variabili; quando una o più variabili risultano essere categoriche, le strutture di covarianza delle variabili risultano essere molto diverse. In generale la misura di associazione tra variabili categoriche sarà attenuata relativamente alla sottostante continua  $y^*$ ; l'obiettivo dell'approccio è quello di riprodurre la stima della struttura di covarianza della variabile  $y^*$ , che si ricorda essere continua. Una soluzione in questi casi è quella di calcolare la misura dell'associazione esistente tra due variabili  $y^*$  attraverso la

correlazione tetrachoric e polychoric, mentre è possibile utilizzare la correlazione polyserial per ottenere la relazione tra due variabili  $y$ . Da notare che l'approccio è ragionevole teoricamente solamente in alcuni casi. Per esempio in molti articoli riguardanti questo argomento i ricercatori sono particolarmente interessati alla relazione esistente tra le variabili latenti sottostanti rispetto alla più semplice relazione tra le variabili osservabili nel soggetto. Per altre variabili con distribuzione continua è difficile immaginare una variabile latente sottostante normalmente distribuita. Infine, alcune variabili come la distinzione del sesso sono intrinsecamente categoriche e per queste nessuna variabile sottostante continua può esistere. L'approccio CVM di Muthén ancora una volta utilizza lo stimatore dei minimi quadrati pesati, ottenendo la seguente formula di minimizzazione per la funzione di fitting:

$$F_{WLS} = [S - \sigma(\hat{\theta})]'W^{-1}[S - \sigma(\hat{\theta})]$$

dove  $S$  è un vettore di dimensioni  $p^* \times 1$  contenente i soli elementi non ridondanti della matrice di covarianza campionaria,  $\sigma(\hat{\theta})$  è il corrispondente vettore  $p^* \times 1$  della matrice  $\Sigma(\hat{\theta})$  di covarianza implicita del modello e  $W^{-1}$  è una matrice di pesi quadrata di dimensioni  $p^* \times p^*$ . In questo caso, definito  $p$  il numero delle variabili osservabili, è possibile calcolare il valore di  $p^*$  come  $\frac{1}{2} * p * (p + 1)$ . Quando  $S$  contiene i coefficienti di correlazione di Pearson per variabili misurabili con distribuzione normale e aventi intervalli di scala, la funzione di fitting viene semplificata e corrisponde a quella dello stimatore GLS discusso precedentemente. L'approccio CVM di Muthén è molto generale e per questo può essere applicato alle variabili categoriche ordinali attraverso l'uso della correlazione polychoric e delle variabili continue che sono state censurate o troncate con la correlazione di Tobit. Questo approccio ha anche alcune limitazioni significative: l'uso richiede che venga utilizzato un campione di grandi dimensioni, dalle 500 alle 1000 osservazioni a seconda della complessità del modello. Ciò nonostante studi sulle simulazioni hanno mostrato buone performance dallo stimatore rispetto a ML, GLS e ADF. La differenza di performance è maggiormente visibile sotto le condizioni identificate nei metodi precedenti che producono delle soluzioni di poco rilievo statistico. Tra le condizioni che fanno lavorare male gli stimatori sono da ricordare un numero troppo piccolo di categorie delle variabili osservabili (due o tre) e forte asimmetria o in magnitudo ( $>1$ ) o in direzione.

#### 4.6.1. Il modello

Muthén considera il seguente modello per  $G$  gruppi, popolazioni, di unità di osservazioni. Il modello è presentato nel seguito. Per ogni gruppo  $g$  è osservata una variabile random dipendente, quindi endogena, le cui osservazioni vengono raccolte nel vettore  $y_g$  la cui dimensione è  $(p * 1)$  e una variabile random indipendente, quindi esogena, con la conseguente raccolta delle osservazioni nel vettore  $x_g$  di dimensioni  $(q * 1)$ ; si assumano le osservazioni dei differenti gruppi indipendenti. In quello che seguirà il pedice  $g$  potrebbe essere indicato vicino ad ogni vettore di cui si vorrà indicare l'attribuzione al gruppo  $g$ -esimo, ma per evitare di aggiungere confusione sarà omissis. Ogni variabile può essere continua o categorica, anche con ordine di categorie. Le

variabili osservate sono assunte come generate da un set di sottostanti variabili latenti continue nella seguente modalità; per ogni gruppo si assume un sistema costituito da equazioni strutturali lineari in cui appaiono  $m$  variabili latenti dipendenti ( $\eta$ ) e  $n$  variabili indipendenti ( $\xi$ ), che possono essere messe in relazione secondo la relazione

$$\eta = \alpha + B * \eta + \Gamma * \xi + \zeta$$

dove  $\alpha$  è il vettore ( $m * 1$ ) dei parametri delle intercette,  $B$  è una matrice ( $m * m$ ) dei coefficienti della regressione tra le variabili latenti dipendenti tale che gli elementi presenti sulla diagonale siano pari a zero e  $I - B$  sia una matrice non singolare,  $\Gamma$  è anch'essa una matrice ma di dimensioni ( $m * n$ ) contenente i coefficienti della regressione tra le variabili latenti dipendenti  $\eta$  e quelle indipendenti  $\xi$ , infine  $\zeta$  è un vettore random di residui contenente quindi gli errori compiuti dall'equazione. È possibile inoltre assumere una relazione lineare interna per il set di  $p$  variabili latenti di risposta  $y^*$  e il set di  $q$  variabili latenti di risposta  $x^*$ :

$$y^* = v_y + \Lambda_y * \eta + \varepsilon$$

$$x^* = v_x + \Lambda_x * \xi + \delta$$

Dove i due vettori  $v_y$  di dimensione ( $p * 1$ ) e  $v_x$  ( $q * 1$ ) contengono i parametri di intercetta,  $\Lambda_y$  ( $p * m$ ) e  $\Lambda_x$  ( $q * n$ ) sono i coefficienti della regressione tra le variabili di risposta latenti e quelle presenti nelle relazioni strutturali, infine  $\varepsilon$  ( $p * 1$ ) e  $\delta$  ( $q * 1$ ) sono i vettori contenente gli errori residui. Le variabili osservabili vengono assunte come correlate alle variabili latenti di risposta attraverso il set di  $p + q$  misure di relazioni esterne. Il modello si compone dunque di parti dove la prima esposta è quella strutturale, mentre la seconda è quella di misura. In tutta si necessita così di otto matrici: quattro sono le matrici contenenti coefficienti strutturali ( $B$ ,  $\Gamma$ ,  $\Lambda_x$ ,  $\Lambda_y$ ) e quattro sono quelle di covarianza ( $\phi$ ,  $\Psi$ ,  $\theta_\varepsilon$ ,  $\theta_\delta$ ).  $\phi$  è la matrice di covarianza delle variabili latenti esogene  $\xi$ , mentre  $\Psi$  contiene le covarianze degli errori  $\zeta$ .

Per la specificazione dei modelli di misura risultano necessarie delle assunzioni, che verranno qui esposte:

- Le variabili vengono misurate in termini di scarti dalle rispettive medie;
- Le variabili indipendenti e gli errori sono tra loro incorrelati nella stessa equazione e tra equazioni differenti;
- Gli errori della stessa equazione sono tra di loro incorrelati.

A seguito di queste assunzioni risulta possibile affermare che non sono possibili altri tipi di relazioni oltre a quelle previste dalle otto matrici sopra citate. Un'ulteriore condizione è che nessuna delle equazioni strutturali risulti essere ridondante all'interno del modello, ovvero la matrice  $B$  deve essere non singolare, quindi invertibile e definita positiva. Ciò significa che tutte le equazioni del modello che esprimono le varie variabili latenti dipendenti  $\eta$  devono essere tra loro indipendenti, in altre parole che nessuna può essere combinazione lineare di altre.

Muthén (1983) distingue tra due casi per quanto riguarda la specificazione della distribuzione delle variabili osservabili. Questa distinzione risulta essere di fondamentale importanza quando si lavora su variabili categoriche dipendenti.

**Caso A.** La specificazione della densità è per la distribuzione congiunta di  $x^*$  e  $y^*$ ; la funzione è  $f(y^*, x^*)$ .

**Caso B.** la specificazione è una distribuzione condizionale di  $y^*$  data  $x$ ; la funzione è  $f(y^*|x)$ . Questo caso è di particolare interesse per il modello quando tutte le variabili indipendenti vengono considerate come continue, con  $q = n$  e  $x = x^* = \xi$ . Allora la prima relazione in cui si è esposto il modello strutturale può essere riscritta nel modo seguente:

$$\eta = \alpha + B\eta + \Gamma x + \zeta$$

Questo caso può essere denominato a variabile  $x$  fissa, si sta allora dicendo che nessuna struttura è imposta sulla distribuzione marginale della variabile  $x$  stessa.

Specificata la funzione di densità, è possibile calcolare il vettore dei valori attesi e la matrice di covarianza di un modello SEM.

**Variabili categoriche con indicatori singoli.** Si considera il caso in cui la variabile di risposta sia categorica con indicatori singoli. Questa è la situazione in cui si presenta il caso B sopra esposto. La metodologia utilizzata in questo contesto è ben conosciuta dagli econometristi ed è una rivisitazione delle applicazioni che trattano variabili binarie e ordinali; infatti questo modello ha le sue origini dalle regressioni probit e logit. Ma al contrario della regressione multivariata, i modelli a equazioni simultanee generalmente pongono una struttura ridotta sia nei coefficienti di regressione, sia nella matrice di covarianza degli errori. Con variabili categoriche come risposta, questi modelli hanno di recente attratto l'interesse degli studiosi.

**Variabili categoriche con indicatori multipli.** Ora si considera una situazione più complessa in cui la variabile categorica in risposta ha degli indicatori multipli di variabili latenti. Si consideri il primo modello di misurazione esposto: qui la variabile latente è relativa a variabili latenti di costruzione. Con indicatori binari, anche qui vengono considerati i modelli logit, nonostante le variabili continue siano ora latenti. Il modello viene sviluppato su un numero di due categorie per poi essere esteso a quelli con più categorie, ovvero da variabili binarie a categoriche.

Il problema principale di questi modelli è quello relativo all'identificazione. Come si è già esposto, un modello, per essere correttamente identificato, deve essere costruito in modo da presentare un'unica soluzione: ciò significa che per ogni parametro si avrà un solo valore stimato. Avere più valori per lo stesso parametro significa che in realtà non sussiste una relazione vera e univoca tra le variabili considerate. Quando accade che il modello risulta essere non identificato è necessario introdurre vincoli all'interno del modello fino a quando non risulta essere corretto. Si precisa che un modello non identificato fornirà ugualmente delle stime, che per non avranno significatività. Si ricorda che la condizione perché il modello sia identificato è che i gradi di libertà siano maggiori o uguali a zero, ovvero che il numero di equazioni componenti il modello siano in numero uguale o

superiore a quello delle incognite da stimare: il modello deve avere più coefficienti di covarianza tra le variabili osservate, che parametri da stimare. Una volta risolto il problema è allora possibile procedere con la stima, che come illustrato precedentemente avviene attraverso un processo di minimizzazione della funzione di fitting.

## 4.7. Valutazione del modello

Nonostante la disponibilità di svariati metodi per la misurazione del fit del modello, si riscontra spesso una difficoltà nel determinare l'adeguatezza del modello a equazioni strutturali in quanto differenti aspetti della soluzione portano a conclusioni contrastanti a riguardo dell'estensione con cui il modello stesso si adatta ai dati osservati. Le due principali modalità per fare una valutazione sul fit coinvolgono la statistica  $\chi^2$  che infatti dà il nome al test. La valutazione può essere fatta su un modello a priori oppure empiricamente attraverso lo sviluppo di procedure di modifica; ma la trattazione seguente affronterà esclusivamente la valutazione a priori. Si considereranno inoltre le seguenti problematiche: l'utilità della statistica si basa su vari metodi di stima per la valutazione e la selezione del modello; l'elaborazione concettuale e la selezione dei criteri necessari a fittare gli indici; l'identificazione di alcuni fattori cruciali che possono influenzare l'entità del valore della statistica. Infine è bene ricordare che deve essere utilizzata una certa quota di cura quando si opera con questa statistica.

### 4.7.1. Il test $\chi^2$

La complessiva convenzione del test nell'analisi di strutture di covarianza determina l'entità della discrepanza tra le matrici di covarianza, quella campionaria e quella fittata. Prendendo  $S$  come la matrice che rappresenta lo stimatore non distorto, basato sulle  $N$  osservazioni del campione, della matrice di covarianza  $\Sigma$ , i cui elementi vengono ipotizzati essere funzione dei parametri contenuti nel vettore  $\theta$ , si può scrivere:

$$\Sigma = \Sigma(\theta)$$

I parametri sono valori stimati, di conseguenza esiste una discrepanza tra  $S = \Sigma(\hat{\theta})$  descritta dalla funzione  $F = [S, \Sigma(\theta)]$  che si vuole minimizzare. Sotto una distribuzione nota e ipotizzando un modello per la popolazione, la statistica  $T = (N - 1) * F_{min}$  ha una distribuzione asintotica (lavorando quindi con campioni di grandi dimensioni) a quella  $\chi^2$ . Questa statistica rappresenta il test da effettuare. Un grande valore di  $T$  in relazione ai gradi di libertà della distribuzione associata, indica che il modello non è una buona rappresentazione del processo che genera i dati della popolazione. In generale, si può affermare che l'ipotesi  $H_0: \Sigma = \Sigma(\theta)$  viene rifiutata quando il valore della statistica  $T$  eccede il valore  $T_\alpha$ , stabilito  $\alpha$  come livello significativo di confidenza. Nel precedente svolgimento, la statistica test  $T$  è stata derivata dal metodo di stima della massima verosimiglianza e da quello dei minimi quadrati generalizzati applicati sotto le condizioni di normalità multivariata delle variabili; esistono però anche altre statistiche disponibili basate su

altri metodi di stima per la valutazione del fit che funziona altrettanto bene. Il  $\chi^2$  test gode di una forte popolarità fin dall'inizio perché sembrava che il suo utilizzo potesse fare essere libera l'analisi fattoriale confermativa da molte decisioni soggettive che storicamente erano associate ad essa. Un test oggettivo è in grado quindi di sostituire giudizi soggettivi. Tuttavia i problemi associati a questa statistica vennero riconosciuti abbastanza presto e diedero inizio ad una serie di discussioni sul tema. Uno degli aspetti di riflessione era legato alla questione della dimensione del campione; infatti la statistica risulta essere asintotica, ovvero migliora al crescere della grandezza del campione e di conseguenza la problematica è quella di capire quale sia il numero di osservazioni necessarie affinché il test possa essere considerato affidabile. Ciò comporta il non corretto funzionamento a livello pratico. Inoltre, con l'aumentare della potenza statistica del test dato dalla crescente numero di osservazioni nel campione, emerge una banale differenza tra la matrice di covarianza stimata dal campione  $S$  e quella fittata dal modello  $\Sigma(\hat{\theta})$ , che condiziona il rifiuto o meno delle specificazione dei modelli; infatti stimando i parametri con troppe osservazioni si creerà un modello che si adatta troppo ai dati, ma che non risulta di buon utilizzo se vengono cambiate le osservazioni, causa del rifiuto di alcuni modelli che nella realtà potrebbero essere veri. Inoltre, la statistica potrebbe non essere distribuita come una  $\chi^2$  quando, tipicamente, è violata l'assunzione di normalità. Quindi questo test standard potrebbe non essere una guida abbastanza buona per valutare l'adeguatezza del modello, poiché una parte significativa del valore prodotto è il riflesso delle non specificazioni del modello stesso, della potenza del test o di violazioni di assunzioni durante la fase di stima.

#### 4.7.2. Performance del test $\chi^2$

Poiché esistono davvero molti test  $\chi^2$ , ognuno dipendente dalla scelta di  $F$  e quindi di  $T$ , alcuni hanno prestazioni migliori di altri.

Bentler (1992) esamina l'adeguatezza di sei test  $\chi^2$  per la bontà del modello sotto condizioni diverse. Come menzionato precedentemente, le statistiche ML e GLS assumono che i dati abbiano distribuzione normale; una violazione di questa ipotesi in pratica invalida il test. La teoria recente sulla robustezza però ha trovato che i metodi basati sulle assunzioni di normalità possono essere corretti e di conseguenza applicati con alcune modifiche anche in casi in cui sia presente una distribuzione non normale delle variabili. Ciò richiede che le variabili latenti di qualsiasi tipo, che solitamente sono considerate come non correlate e mutualmente indipendenti, quando infrangono la non correlazione devono avere dei parametri di varianza e covarianza stimati liberamente. A riguardo però non esistono test pratici.

Browne (1984) introduce la teoria ellittica multivariata con assunzioni di distribuzione più generali della normale. La distribuzione ellittica è simmetrica, con code che possono essere identiche a quelle di una distribuzione normale ma non necessariamente, infatti possono essere anche più o meno spesse. In queste distribuzioni il parametro relativo alla curtosi, riflettendo le assunzioni comuni delle variabili, in aggiunta agli usuali parametri, è necessario per produrre uno stimatore asintoticamente ottimo e un test  $\chi^2$  per la bontà del fit. Kano, Berkane e Bentler (1990) trovano

un aggiustamento semplice per la matrice dei pesi, utilizzando una stima della curtosi delle variabili, che così può provocare convenienze statistiche ottimali. La loro statistica della curtosi eterogenea (HK) specializza per teorie normali e ellittiche, ovvero quando le variabili rispettivamente non eccedono in curtosi o hanno curtosi omogenee. Quando sono violate sia le assunzioni delle distribuzioni normali, sia di quelle ellittiche, la statistica  $T$  può essere corretta utilizzando il fattore di Santorra e Bentler esposto precedentemente. Il test corretto denominata statistica scalata, è calcolato sulla base del modello, del metodo di stima, del momento del quarto ordine della distribuzione e funziona a prescindere dalla distribuzione delle variabili. Allo stesso modo lo stimatore ADF produce un reclamo promettente che il test statistico per il fit del modello sia non sensibile alla distribuzione delle osservazioni quando il campione ha grandi dimensioni. Hu, Bentler e Kano (1992) valutano le performance dei modelli sopra esposti quando le ipotesi distribuzionali, di indipendenza e di efficienza asintotica, vengono violate. La condizione di dipendenza è un caso in cui due o più variabili sono collegate funzionalmente, nonostante la loro combinazione lineare risulta dare un valore praticamente nullo. Si ha così un contrasto poiché, con dati relativamente normali, relazioni lineari con risultato zero implicano indipendenza. È stato preso in considerazione e studiato un modello di analisi fattoriale confermativa basato su 15 variabili osservabili con 3 fattori comuni. I risultati, ottenuti attraverso la procedura Monte Carlo, rivelano delle forti implicazioni per la pratica. In accordo con quanto precedentemente trovato, l'aspettativa teorica per la quale il metodo ADF funziona bene con qualsiasi distribuzioni è corretta solamente quando la dimensione del campione è estremamente grande (ad esempio 5000 osservazioni); infatti con campioni di piccole dimensioni lo stimatore ha delle performance povere sotto tutte le condizioni stimate e di conseguenza non è raccomandabile nelle applicazioni pratiche. Con dati generati per produrre una robustezza asintotica, i metodi basati sulla normalità (ML e GLS) hanno performance adeguate e di conseguenza sono in ogni caso da preferirsi; tuttavia con campioni di piccole dimensioni ML rifiuta modelli veri troppo frequentemente mentre GLS sembra comportarsi in modo migliore. Sfortunatamente gli esperti non possono affidarsi ai test statistici basati su teorie normali quando ci potrebbe essere una dipendenza tra le variabili latenti, poiché questi metodi essenzialmente rifiutano sempre i modelli veri sotto queste condizioni. Sia il metodo basato sulla distribuzione ellittica, sia quello sulla curtosi eterogenea tendono ad accettare modelli più frequentemente di quanto non ce ne si aspettino quando le variabili latenti hanno distribuzione indipendente. Quando le variabili invece risultano essere dipendenti il metodo ellittico rifiuta modelli più frequentemente del valore atteso, mentre il metodo HK accetta modelli troppo spesso. La statistica scalata di Santorra e Bentler, basata sulla correzione della statistica  $\chi^2$  ottenuta da ML, fornisce performance migliori in condizioni di dipendenza rispetto a tutti gli altri metodi visti; la bontà delle sue prestazioni è assimilabile a quelle ottenute con i metodi basati su teorie normali in condizioni di indipendenza. In generale è possibile affermare che è il metodo migliore, ma bisogna ricordare che ha la tendenza a rifiutare troppi modelli veri con campioni di piccole dimensioni; studi successivi però sembrano rivalutare le prestazioni di quest'ultimo modello anche con poche osservazioni. In conclusione si può affermare che con campioni di dimensioni abbastanza grandi la statistica  $\chi^2$  ottenuta da ML o da GLS come quella riscalata funzionano bene è approssimativamente allo stesso modo; inoltre se le dimensioni del campione sono estremamente grandi anche ADF propone delle performance sugli stessi standard.

### 4.7.3. Distribuzioni $\chi^2$ non centrate

Si supponga che l'ipotesi nulla  $H_0: \Sigma \neq \Sigma(\theta)$  non sia vera. In seguito anche che  $T$  non è distribuita come una  $\chi^2$ , ma che potrebbe essere distribuita come  $\chi^2$  non centrata. Un parametro di non centralità ( $\lambda$ ) e i gradi di libertà ( $df$ ) sono necessari per specificare una distribuzione  $\chi^2$  che sia non centrata e che quindi viene definita attraverso la scrittura  $\chi^2(df, \lambda)$ . Il parametro  $\lambda$  di non centralità rappresenta una misura della discrepanza tra  $\Sigma$  e  $\Sigma(\theta)$  e può essere considerato come una "statistica di cattiveria del fit". Così maggiore è il valore di  $\lambda$ , più grande sarà la discrepanza delle due misure portando la situazione reale molto lontano dall'ipotesi nulla. La distribuzione  $\chi^2$  centrata originale è un caso particolare della distribuzione non centrata in cui si presenta  $\lambda = 0$ . Cudeck e Henly (1991) classificano i diversi tipi di errori coinvolti nel fitting di un modello basandosi sulla discrepanza esistente tra le due matrici. Per esempio l'errore complessivo è definito come la misura della differenza esistente tra  $\Sigma$  e  $\Sigma(\hat{\theta})$ , mentre l'errore di approssimazione si riferisce alla discrepanza tra  $\Sigma$  e  $\Sigma(\theta)$ . Così parametri derivanti dalla non centralità sono considerati come una misura dell'errore di misurazione. Poiché questa misura di errore può essere stimata solo attraverso l'utilizzo del campione, viene costruito un intervallo di confidenza sulla popolazione per i parametri non centrati che consentirà di stabilire la non bontà del modello all'interno della popolazione e la precisione con cui questa viene determinata dal campione di dati. Come si vedrà in seguito, i parametri di non centralità possono anche essere utilizzati per generare indici di fit, nel caso in cui si presentino queste situazioni.

### 4.7.4. Indici di fit

In pratica, solamente la distribuzione  $\chi^2$  centrata è utilizzata per testare l'ipotesi nulla sopra discussa e le modifiche proposte dalla statistica scalata di Santorra e Bentler sembra funzionare bene per tutti quei casi in cui le ipotesi vengono violate, come discusso precedentemente. Tuttavia, valutando i modelli attraverso questa statistica, rimane il problema dell'eccessiva potenza nel rifiutare modelli perché su campioni di dimensioni molto grandi. ovvero, anche se la discrepanza tra modello stimato e dati è molto piccola, quasi ogni modello verrà rifiutato perché la discrepanza non è statisticamente uguale a zero. Allora la rigorosa ipotesi nulla  $H_0: \Sigma = \Sigma(\theta)$  quasi sicuramente non sarà esattamente vera perché il ricercatore non è in grado di conoscere ogni cosa che c'è da sapere sui dati. Inoltre la statistica  $\chi^2$  offre solamente una decisione dicotomica implicita nella regola statistica e che non può essere utilizzata per quantificare il grado del fit di adattarsi lungo un percorso continuo con dei contorni specificati in precedenza nel modello stesso. Per questo motivo diversi studiosi hanno ricavato degli indici di fit che possano essere in grado di spiegare quanto il fit si adatti ai dati veri. Come  $R^2$  nella regressione, questi indici si propongono di quantificare qualcosa di simile alla varianza, piuttosto che testare l'ipotesi nulla. In particolare questi indici generalmente quantificano l'estensione della varianza e della covarianza nei dati accumulata attraverso il modello. Come notato da Bentler e Bonett (1980), che introdussero diversi indici e resero popolare l'idea, questi strumenti sono costruiti per evitare alcuni dei problemi derivanti dalla dimensione del campione a alle mancanze nella specificazione



del modello a riguardo della distribuzione. Inizialmente si sperava che questi indici di fit avrebbero avuto meno ambiguità nella valutazione di adeguatezza dei modelli rispetto alla statistica  $\chi^2$ . Questa speranza ottimistica si rivelò presto non corretta sfortunatamente. La questione della dimensione del campione e della complessità del modello ancora influenzano l'entità degli indici. È necessario analizzare l'adeguatezza delle quattro tipologie di indici derivanti dai diversi metodi di stima attraverso la variazione della dimensione campionaria, ma anche a seguito della violazione di ipotesi distributive.

Nel seguito vengono espone le tipologie degli indici che possono essere utilizzati all'interno dell'ambito dei modelli SEM.

#### 4.7.5. Tipologie di indici

Una delle classificazioni maggiormente adottate per la classificazione degli indici di fit è la distinzione tra indici assoluti e indici incrementali. La prima tipologia assegna direttamente una valutazione a priori su come il modello riproduce i dati; nonostante non venga richiesta alcuna referenza del modello, una comparazione esplicita o implicita viene fatta per modelli sturi che riproducono esattamente la matrice di covarianza osservata. Come risultato questo tipo di indice è analogo a  $R^2$  nella comparazione della bontà del fit che è simile alla somma totale dei quadrati. In contrasto un indice di tipo incrementale misura il miglioramento del fit rispetto ad un modello target con maggiori restrizioni, un modello nested di base; per la modalità con cui operano questi indici vengono chiamati anche di comparazione. Il modello nullo in cui tutte le variabili osservabili sono non correlate è solitamente il modello con cui tutti gli altri vengono messi a confronto, sebbene sia possibile utilizzare altri modelli di base. Altre distinzioni possono essere fatte all'interno della tipologia degli indici incrementali, in particolare si evidenziano 3 categorie che verranno indicati come Type-1, Type-2 e Type-3. La prima tipologia usa solamente le informazioni dalla statistica ottimizzata  $T$  ottenuta dal fitting del modello di base ( $T_B$ ) e quella del modello target ( $T_T$ ). La statistica non deve assumere obbligatoriamente una particolare distribuzione, sebbene se ne necessiti per calcolare la funzione di fitting  $F$ , poiché è in ogni caso identica per entrambi i modelli a confronto. La seconda tipologia utilizza dell'informazione aggiuntiva derivante dal valore della statistica  $T_T$  ottenuta sotto l'ipotesi di una distribuzione  $\chi^2$  centrata. La terza tipologia utilizza l'informazione della prima tipologia a cui viene aggiunta il valore atteso di  $T_B$  e/o di  $T_T$  sotto una distribuzione  $\chi^2$  non centrata rilevante. Quando l'assunzione sulla distribuzione è corretta, la seconda e la terza tipologia danno le prestazioni migliori poiché utilizzano un quantitativo maggiore di informazione; tuttavia è da notare l'informazione utilizzata da questi indici potrebbe essere inappropriata poiché ogni particolare statistica  $T$  potrebbe non avere la distribuzione assunta. Per esempio si consideri un indice Type-3 potrebbe applicare una distribuzione non centrata quando quella vera risulta essere al contrario centrata. Esistono poi altre categorie di indici incrementali che però non verranno discussi nella trattazione.

**Indici di fit assoluti.** Il primo indice proposto tra quelli assoluti è basato sul metodo della massima verosimiglianza e definito da Jöreskog e Söbrom nel 1984; la sua formula analitica è

$$GFI_{ML} = 1 - \left[ \text{tr}(\hat{\Sigma}^{-1}S - I)^2 / \text{tr}(\hat{\Sigma}^{-1}S)^2 \right].$$

In seguito svilupparono un ulteriore indice corrispondente ma corretto, in modo tale da incorporare una funzione di penalità per l'inclusione di parametri aggiuntivi:

$$AGFI = 1 - [p * (p + 1) / 2df] * (1 - GFI_{ML})$$

Infatti, proprio come per il coefficiente  $R^2$ , all'aumentare dei parametri il modello avrà sicuramente un migliore adattamento al fit, che d'altra parte però lo rende più rigido perché troppo aderente ai dati del campione e soprattutto meno parsimonioso. Quindi al crescere dei parametri aumenterà sicuramente l'indice  $GFI_{ML}$ , ma con lui anche la penalizzazione data dal numero dei parametri; bisognerà capire caso per caso quale grandezza cresce più velocemente. L'indice  $GFI_{ML}$  è una misura della varianza e della covarianza che viene accumulata nel modello,  $\hat{\Sigma}$ , relativamente a quella già insita nella popolazione e quindi stimata nella matrice  $S$ . Quando  $\hat{\Sigma} = S$ , gli indicatori assumeranno il loro valore massimo (ovvero 1), infatti indicano che il modello è il migliore possibile poiché non aggiunge variabilità rispetto a quella già presente nei dati. Il loro valore tipicamente è maggiore di zero in modo evidente, ma è possibile che questo si avvicini molto ad esserlo. Esiste inoltre una versione di  $GFI$  e di  $AGFI$  ottenuta attraverso il metodo dei minimi quadrati generalizzati, che tuttavia risulta essere molto simile a quella fornita sopra ottenuta dalla massima verosimiglianza.

Riferendosi invece all'indice  $AIC$  è possibile individuarne uno che lo replichi per la valutazione dei modelli SEM; gli autori della correzione sono Cudeck e Browne (1983) che ne hanno sviluppata una funzione riscalata, anch'essa capace di penalizzare l'aggiunta di parametri. La formula ottenuta è:

$$CAK = [T_T / (N - 1)] + [2q / (N - 1)]$$

Ci si aspetta che questo indici lavori bene in campioni di diverse dimensioni. Ma basandosi sulla calibrazione singola campionaria, svilupparono un altro indice di cross-validation:

$$CK = [T_T / (N - 1)] + [2q / (N - p - 2)]$$

$CK$  dovrebbe essere una buona approssimazione di un indice di fit che si potrebbe ottenere con una cross-validation in un nuovo campione.  $CAK$  e  $CK$  devono essere utilizzati per scegliere uno o più modelli tra un set di opzioni plausibili, con l'obiettivo di individuare quei modelli che avranno performance migliori quando verranno utilizzati su campioni futuri. In generale, si sceglieranno quei modelli che presentano il valore minore, poiché corrispondono a miglior adattamento ai dati.

Data la disponibilità di una notevole varietà di indici, la decisione di utilizzarne uno o più tra quelli esistenti, richiede attenzione poiché è necessario considerare alcuni fattori critici che potrebbero influenzare la performance dell'indice stesso e di conseguenza la valutazione dei modelli potrebbe risultare errata.

**Type-1.** Gli indici incrementali assegnano un grado di adeguatezza del modello da valutare, detto target, in riferimento ad un modello di base, utilizzando la statistica non negativa  $T$  in entrambe le

situazioni. Un classico esempio di indice è quello costruito da Bentler e Bonett che prende il nome di *NFI*:

$$NFI = (T_B - T_T)/T_B$$

Solitamente il valore di  $T_B$  è grande; quindi se  $T_T$  ha un decremento insignificante da  $T_B$  l'indice fornisce un valore prossimo a 0, mentre se  $T_T$  risulta essere molto piccolo l'indice corrisponde al rapporto tra i valori  $T_B$  e per questo risulta essere vicino ad 1. Più genericamente è possibile affermare che l'indice *NFI* rappresenta la porzione della covarianza totale tra le variabili osservabili spiegata dal modello target quando viene utilizzato il modello nullo come riferimento. Come sviluppato da Bentler e Bonett, *NFI* è in grado di calcolare i valori anche quando le misure  $T_B$  e  $T_T$  non si presumono essere dei test statistici dato che non hanno distribuzione  $\chi^2$ . Il valore che l'indice può assumere è compreso nel range (0 – 1).

Un altro indice presente in questa categoria è quello sviluppato da Bollen nel 1986 e che per questo viene denominato *BL86*. La novità di questo indice rispetto a quello precedente è la normalizzazione dei valori attraverso i gradi di libertà di ciascuna statistica, ovvero:

$$BL86 = [(T_B/df_B) - (T_T/df_T)]/(T_B/df_B)$$

**Type-2.** Il classico indice sviluppato inizialmente da Tucker e Lewis (1973) sotto l'ipotesi di normalità e attraverso il metodo di stima ML usa informazioni dal valore atteso di  $T_T$ . Nella sua formula finale si presenta simile a *BL86*, ma con una modifica al denominatore:

$$TLI = \frac{\left[\left(\frac{T_B}{df_B}\right) - \left(\frac{T_T}{df_T}\right)\right]}{\left(\frac{T_B}{df_B}\right) - 1}$$

Lo scopo originale nello sviluppo dell'indice era quello di quantificare il grado in cui un particolare modello di esplorazione fattoriale fosse un miglioramento rispetto al modello zero quando veniva trovato attraverso la massima verosimiglianza. Generalizzando la definizione di ogni tipo di modello strutturale di covarianza sotto vari modelli di stima, così l'indice venne in seguito denominato "nonnormed", ovvero *NNFI*, poiché esso non necessita di avere il range (0 – 1), anche se  $T_B \geq T_T$ . Un altro indice appartenente a questa categoria è ancora una volta sviluppato da Bollen, che modifica il denominatore di *NFI* nel seguente modo:

$$BL89 = \frac{(T_B - T_T)}{(T_B - df_T)}$$

Come è evidente esistono delle relazioni tra questi indici, che non comportano in primo luogo la definizione della popolazione e, in secondo luogo, l'utilizzo di stimatori per questi parametri per definire il campione.

**Type-3.** Come notato precedentemente, con ipotesi nulle vere o non estremamente non specificate, l'errore di previsione è relativamente piccolo se confrontato con gli errori campionari

della matrice  $S$ , e la statistica test  $T$  può essere approssimato nei campioni di grandi dimensioni ad una distribuzione non centrata  $\chi^2(df, \lambda)$ . I parametri delle distribuzioni non centrate associati alla sequenza di modelli nested, ad esempio  $M_B, \dots, M_j, \dots, M_T, \dots, M_S$ , rappresentanti il grado della non specificazione sono ordinati come  $\lambda_B \geq \lambda_j \geq \lambda_T \geq \lambda_S = 0$ , con valore maggiore per il parametro del modello di base e valore minore per il modello saturo. Quindi la popolazione controparte di  $NFI$  è ovviamente data da

$$\delta = (\lambda_B - \lambda_T) / \lambda_B.$$

Per un dato modello base, la più piccola non specificazione di  $M_T$ , a cui corrisponde il più piccolo parametro  $\lambda_T$ , si otterrà il maggior valore dell'indice  $\delta$ . Quindi un indice non centrato assegna la riduzione di una quota indice di non adattamento del fit del modello target, in relazione a quello di base. Certamente  $\delta$  non è operativo a meno che non vengano utilizzati degli stimatori  $\lambda$  consistenti per la popolazione. L'utilizzo della statistica test  $T$  per stimare i parametri  $\lambda$  avrebbe portato all'indice  $NFI$ , ma il problema è che la statistica non è un buon stimatore per questi parametri. Per questo motivo Bentler ha sviluppato un indice per la non centralità, che stima i parametri attraverso la differenza tra la statistica  $T$  stessa e i corrispondenti gradi di libertà, ottenendo

$$BFI = \frac{[(T_B - df_B) - (T_T - df_T)]}{(T_B - df_B)}$$

Il problema di questo indicatore invece risiede nel fatto che può trovarsi al di fuori del range  $(0 - 1)$ ; allora per porre rimedio alla questione l'indice è stato modificato allo scopo di farlo rientrare nel range desiderato:

$$CFI = 1 - \max[(T_T - df_T), 0] / \max[(T_T - df_T), (T_B - df_B), 0]$$

Ogni volta che  $BFI$  è all'interno del range, risulta essere  $BFI = CFI$ .

#### 4.7.6. Problemi di valutazione degli indici di fit e criteri di selezione

Esistono dei problemi principali coinvolti nell'utilizzo degli indici di fit presi in considerazione per la valutazione della bontà dei fit: piccola distorsione del campione, effetti di stima e effetti della violazione delle ipotesi di normalità e indipendenza. Questi sono la naturale conseguenza del fatto che tipicamente sono basati sulla statistica test  $\chi^2$ . Allora, parlando in generale, quando la statistica test ha delle buona performance, ci si aspetterà una corrispondenza di buone prestazioni anche da parte dell'indice di fit. Poiché, come menzionato precedentemente,  $\chi^2$  non lavora bene quando il campione è di dimensioni ridotte, anche con gli indici di fit si riscontrerà la stessa problematica; allo stesso modo si ricordano i problemi per  $\chi^2$  derivanti dalla violazione delle ipotesi distributive che vengono fatte sulle variabili, presenti di conseguenza anche per gli indici di fit.

**Distorsione da piccoli campioni.** I metodi di stima nella modellazione di equazioni strutturali vengono sviluppati sotto assunzioni varie; una di queste è che il modello sia vero. Questa assunzione implica che ci sia uguaglianza tra le matrici di covarianza, tra quella campionaria e quella stimata dal modello, ovvero che non si possa rifiutare l'ipotesi nulla  $H_0: \Sigma = \Sigma(\theta)$ . Un'altra assunzione fatta è che stime e test sono basati su campioni di dimensioni arbitrariamente grandi, cosa che invece risulta essere quasi sempre non vera nella pratica. L'adeguatezza della statistica test è quindi probabile che venga influenzata dalla grandezza del campione, forse fornendo delle prestazioni più povere nei campioni di più piccole dimensioni e aumentando via via, probabilmente però non abbastanza da poter essere considerata asintotica. Infatti, la relazione tra grandezza del campione e adeguatezza del fit quando il modello è vero, è stata da molto tempo riconosciuta. Per esempio Bearden, Sharma e Teel (1982) trovarono che la media degli indici *NFI* è correlata positivamente alla dimensione del campione e che tende ad essere meno lontana da zero quando le osservazioni nel campione sono poche. Questo loro risultato ha sottolineato il problema principale: possibili distorsioni sistematiche negli indici di fit. Se la media degli indici di fit, calcolati attraverso differenti campioni sotto le stesse condizioni quando il modello è vero, varia sistematicamente con il valore di  $N$ , tale statistica potrebbe essere una stima distorta dei corrispondenti parametri della popolazione. Oppure, detto in altre parole, la decisione di accettare o rifiutare un particolare modello può variare in funzione della dimensione campionaria, cosa che certamente non risulta essere desiderabile.

## 4.8. Interpretazione e modifiche

### 4.8.1. Interpretazione

L'ultimo paragrafo del capitolo è dedicato alla fase che avviene una volta che sono stati stimati i parametri del modello. Come già detto, un modello sviluppato a conferma di una teoria di base si poggia su delle relazioni che si pensano esistere tra le variabili. Queste relazioni ipotizzate vengono poi tradotte in equazioni, che nel caso di modelli SEM risultano essere del tipo lineare, per le quali si vuole stimare un coefficiente. Il valore così stimato è indice della relazione stessa esistente tra le variabili; quando ad esempio il valore trovato è uguale a zero il parametro sta dicendo che non esiste relazione tra le due variabili, o che in ogni caso questa non sia significativa. Per poter concludere a riguardo della significatività o meno dei singoli coefficienti stimati è opportuno procedere con dei test di ipotesi singoli.

### 4.8.2. Modifiche

È possibile che risulti necessario apportare alcune modifiche al modello di partenza che è stato stimato. Lo scopo con cui queste modifiche vengono apportate è migliorare la vestibilità del modello stesso al campione di dati, così da avere stime più probabili delle relazioni presenti. Nella fase di modifica però bisogna fare attenzione a non irrigidire troppo il modello sul singolo

campione su cui è stato stimato il modello; si è già discusso abbondantemente infatti di come far aderire troppo un modello ai dati su cui lo si sta testando aumenta il rischio di renderlo inutilizzabile su altri campioni appartenenti alla stessa popolazione. Da un punto di vista pratico, scegliere che tipo di modifiche apportare è abbastanza semplice: molti software forniscono direttamente all'utente degli indici di modifica riferiti al miglioramento che si potrebbe ottenere a seguito dell'aggiunta di un percorso supplementare. Non è da dimenticare però che il modello riflette una teoria di base, allora le modifiche dovranno essere in accordo con questa per essere implementate; infatti, operando delle modifiche contrarie alla base teorica esposta si otterranno comunque dei risultati non accettabili da un punto di vista interpretativo.

## 5. DETERMINAZIONE DELL'ATTRATTIVITÀ DEL PF CON L'UTILIZZO DEI SEM

### 5.1. Attrattività del Project Finance

L'attrattività del Project Finance può essere valutata attraverso la capacità di raccogliere risorse e la capacità di finanziare progetti fuori bilancio ricorrendo ad un capitale di equity limitato o nullo. Esistono delle motivazioni economiche che spiegano perché la finanza di progetto risulta essere interessante e che possono essere ricondotte alle caratteristiche principali dello strumento finanziario stesso. Di seguito si entrerà nel dettaglio per mostrare quali sono queste caratteristiche sopra citate.

La finanza di progetto risulta essere da preferirsi alla corporate finance in quanto risolve due problemi finanziari quali i costi di agenzia e i costi opportunità di sottoinvestimento legati all'entità della leva. Ciò risulta possibile in quanto il progetto risulta essere fuori dal bilancio degli sponsor e in una società costruita ad hoc. Detto questo ci sono tre principali motivazioni che spingono all'utilizzo del project finance, ovvero:

- La riduzione dei costi di agenzia;
- La possibilità di utilizzare una forte quota di debito;
- La gestione del rischio.

#### 5.1.1. Costi di agenzia

Questa tipologia di costi si riferisce alle problematiche che si originano dai conflitti interni ad un'azienda, tra proprietà e controllo, tra proprietà e le altre parti connesse, e tra detentori di equity e di debito. Le motivazioni per le quali questi problemi esistono sono due: esiste una possibilità di sottoinvestimento ex-ante e di comportamenti opportunistici ex-post. Questi costi sono naturalmente legati alle caratteristiche degli asset. Asset che generano grandi margini operativi e abbondanti cash-flow possono essere condotti verso investimenti inefficienti e a produzioni di valore minore. In questo caso il costo di agenzia cresce quando i manager dell'azienda, che hanno il controllo sulle decisioni di investimento, hanno degli incentivi diversi da quelli dei detentori del capitale, ovvero dalla proprietà. La creazione di una società indipendente conferisce agli sponsor la possibilità di creare un assetto specifico aziendale tagliato su misura per il progetto da realizzare, favorendo la minimizzazione dei costi di agenzia. Nel seguito si vuole accennare alle tre tipologie di conflitti che si possono creare tra le parti.

**Conflitti tra proprietà e controllo.** Avere un equity detenuto in modo privato significa che le informazioni disponibili sono ridotte e che non esiste un mercato su cui è possibile scambiare il capitale stesso con altri investitori; queste due caratteristiche riducono l'efficienza del business. Nel

contesto della finanza di progetto questo problema può essere risolto basandosi su una proprietà concentrata con un'unica linea guida, separazione legale e un alto valore di leverage per limitare la discrezione dei manager. In altre parole, concentrare la proprietà di equity e di debito incrementa il livello di monitoraggio sulle azioni intraprese dai manager, oltre ad utilizzare la componente di debito come strumento di monitoraggio per i soggetti creditori. La facilità di monitoraggio è data anche dalla natura dei flussi di cassa, che sono originati da un solo business, quindi facilmente identificabili.

**Conflitti tra proprietà e le altre parti connesse.** Questo è una seconda tipologia di contratto che si può creare e consiste nella possibilità di assistere a dei comportamenti opportunistici che possono causare la riduzione dei ritorni finali ex post, ma anche la scelta di investimenti meno favorevoli ex ante. I fornitori di input critici sono un esempio di parte connessa che può causare un incremento dei costi di agenzia, che però attraverso la finanza di progetto possono essere ancora una volta scoraggiati. Allo scopo di evitare comportamenti opportunistici delle parti connesse gli sponsor possono utilizzare due tecniche: stipulare contratti di lungo termine, utilizzare una quota elevata di debito. La stesura di contratti di lungo termine è necessaria per legare le parti al cash-flow finale del progetto; infatti in questo caso anche le parti connesse sono interessate ad avere un ritorno finale il più possibile elevato, in quanto retribuisce anche loro. Da notare che nella tradizionale corporate finance questa tecnica risulta essere meno efficiente in quanto i flussi di cassa non possono essere attribuiti ai vari business, ma si osserva un valore conglomerato. Imporre invece un'alta leva finanziaria significa alzare i costi di default o di cambio di controllo; nessuna delle parti allora sarà incentivata a non ricercare il massimo profitto. Anche per questa tecnica, l'efficienza in corporate finance è minore perché ancora i flussi di cassa sono dati dalla sommatoria di business differenti.

**Conflitti tra detentori di debito e di equity.** Questa ultima tipologia di conflitti riguarda solitamente le decisioni di reinvestimento e di distribuzione dei cash-flow. In un contesto di finanza di progetto la distorsione dovuta all'alto indebitamento, come il trasferimento dei rischi, e al sottoinvestimento risultano essere meno rilevanti. L'opportunità del trasferimento del rischio è soffocata dai cash-flow che restringono l'ambito decisionale e dalla forte capacità di monitoraggio da parte della proprietà.

Si è quindi esposto come il project finance possa ridurre il problema dei costi di agenzia, ma anche quello del sottoinvestimento. Questi aspetti sono alla base delle altre due motivazioni precedentemente menzionate e che nel seguito verranno utilizzate.

### **5.1.2. Utilizzo di una forte quota di debito**

Ci si riferisce al fatto che il project finance riduce i sottoinvestimenti generati dal costo dell'indebitamento da parte degli sponsor. Infatti la finanza di progetto aiuta nell'attrarre opportunità di investimento attrattive senza modificare gli indici di bilancio, essendo in una società ad hoc. Come risultato si ha la capacità di preservare la situazione finanziaria a livello



corporate, in modo che non si veda aumentare il costo del debito. Il project finance appare così più efficace in quanto riduce qualsiasi ricorso agli sponsor. Ciò significa che il nuovo capitale non è sovvenzionato da reclami preesistenti dati precedenti prestiti; infatti non esistono bilanci degli anni precedenti, la SPV è come una start-up che è possibile valutare solo attraverso la valutazione dei flussi di cassa futuri che potrà generare.

### **5.1.3. Gestione del rischio**

La terza motivazione riguarda la capacità della finanza di progetto di ridurre i danni collaterali potenziali che un progetto ad alto rischio può generare. Il concetto è ancora una volta legato alla presenza di una società creata ad hoc per la realizzazione di un progetto specifico, infatti come detto più volte questo risulta essere fuori bilancio. La situazione creata è così esterna e indipendente rispetto a quella corporate; di conseguenza non ne modificherà il profilo di rischio totale, componente fondamentale per la determinazione del costo del capitale che una società deve corrispondere nel caso in cui volesse accedere a delle risorse non proprie, sul mercato.

In conclusione si può affermare come il project finance sia uno strumento molto appetibile per la realizzazione di nuovi progetti grazie alla sua capacità di ridurre tutta una serie di costi esposti fin qui. Contrariamente bisognerebbe sostenerli se si ricorresse alla classica finanza corporate, inglobando la nuova iniziativa nel bilancio aziendale. In questo caso si modificherebbe sia il profilo di rischio, sia la situazione finanziaria degli sponsor partecipanti in modo negativo, tale da generare un aumento dei costi sopra citati. Per questi motivi la finanza di progetto risulta essere preferibile dagli sponsor: si ha la possibilità di accedere a molto capitale e sostenere dei rischi anche elevati, senza modificare la situazione aziendale preesistente.

Ma ci sono altri soggetti per i quali la finanza di progetto è uno strumento attrattivo: i prestatori e gli istituti di credito in generale. Infatti risulta essere un asset class profittevole e robusto. I prestatori ricevono infatti un ritorno commisurato alla quota di rischio che si sono assunti: la società li ricompensa attraverso provvigioni e ricompense associati ai flussi generati dai prodotti e dai servizi della stessa. Altri soggetti sono i Settori Pubblici che attraverso opere di paternariato possono realizzare opere di cui altrimenti la realizzazione non sarebbe nemmeno stata presa in considerazione; il vantaggio sta nella produzione di esternalità positive che genereranno flussi positivi anche per le Amministrazioni anche se queste non partecipano direttamente ai lavori. Inoltre la realizzazione avviene eliminando quelle inefficienze tipiche del settore pubblico e i rischi possono essere trasferiti in parte ai soggetti privati che hanno in appalto l'opera.

## **5.2. Indebitamento**

Il leverage rappresenta la quota di utilizzo del capitale di debito rispetto all'equity, per finanziare gli asset del progetto. Riflette la struttura del capitale della SPV che deve far combaciare necessità di lungo termine con quelle di breve termine; nello specifico solitamente si necessita di molto capitale a lungo termine per poter finanziare la costruzione dell'opera, ma anche di breve termine

per la gestione quotidiana della cassa. Si assume che il livello di indebitamento possa essere una proxy dell'attrattività dello strumento finanziario, in quanto, come motivato precedentemente, è possibile avere impieghi di equity anche nulli senza incrementare una serie di costi: allora si cercherà di lavorare con valori di indebitamento abbastanza alti. Per questo motivo si è visto come un progetto in project finance è capace di utilizzare una leva che sarebbe inimmaginabile nella corporate finance, poiché porterebbe ad uno spostamento negativo dell'equilibrio aziendale. Sia la struttura del capitale, che di conseguenza il leverage, sono fatti su misura per il progetto. Si può affermare che il grado di leverage dipende da alcuni fattori che sono: le caratteristiche del progetto come la capacità di generare ritorni e la profittabilità, i requisiti dei prestatori, i parametri del contratto di prestito come interessi e maturità, la qualità dei contratti stipulati con le altre parti coinvolte nell'iniziativa.

Un alto livello di leverage migliora il ritorno sull'equity fondamentalmente per due ragioni. La prima motivazione riguarda il confronto tra il costo del capitale di debito e di equity: il primo è sicuramente minore del secondo, allora è conveniente che questo sia in una quota maggiore all'interno della struttura finanziaria; anche perché in qualche modo si sta trasferendo una parte di rischio su soggetti terzi. La seconda motivazione ha origine dalla componente fiscale, infatti si genera un risparmio dovuto alla deducibilità degli interessi corrisposti agli istituti di credito, al contrario dei dividendi. Come risultato si ha che il rischio totale da sopportare decresce e che la contribuzione finale per gli sponsor aumenta.

La variabile LEVA è quindi una variabile osservabile dipendente.

### **5.3. Dimensione del progetto**

La finanza di progetto è una tecnica finanziaria utilizzata per aumentare il ricorso al capitale di debito per finanziare progetti di grandi dimensioni e complessi, che vengono strutturati come una nuova e indipendente società ad hoc. Il numero di infrastrutture costruite grazie ad un finanziamento di questo tipo sono cresciuti nel tempo e sembra essere ancora in crescita. La finanza di progetto però sembra essere attrattiva per il finanziamento di iniziative di grandi dimensioni. La dimensione del progetto influenza l'entità dei rischi, specialmente nel caso in cui i progetti includano delle difficoltà di realizzazione e un accesso costoso alle risorse o alle tecnologie. In altre parole, progetti più grandi risultano essere più rischiosi, e solitamente un solo istituto non è in grado di finanziarli totalmente da solo. Inoltre, si assume che ci sia una relazione positiva e convessa tra l'incremento dei costi da sostenere e la grandezza dell'investimento totale; infine sempre una relazione positiva convessa è presente tra i costi e la grandezza del finanziamento, ovvero al crescere della somma totale si associano delle spese maggiori. Le dinamiche espresse sono tipiche del project finance e di conseguenza non si ritrovano nella corporate finance. Progetti di grandi dimensioni implicano poi una complessità contrattuale che stipuli le linee guida per il comportamento delle parti coinvolte. Questa complessità richiede che ci siano delle figure competenti e esperte dell'ambito in questione, sia dal punto di vista legale che

finanziario. Infine le società di progetto possono richiedere un indebitamento minore per tagliare l'esposizione degli sponsor ai rischi.

Sotto questa prospettiva la dimensione del progetto può essere considerata una misura del rischio di progetto, quindi si ipotizza che abbia un impatto negativo sul livello di leverage: ad esempio per progetti di piccola entità ci si possono aspettare indebitamenti anche del 100% e viceversa.

La variabile DIMENSIONE è una variabile osservabile indipendente.

## **5.4. Settore**

Nel primo capitolo si è esposto quali siano i settori in cui il project finance venga maggiormente applicato. La tipologia di industria risulta essere rilevante per la determinazione delle caratteristiche principali del contratto di project finance da applicare. Questa infatti risulta essere collegata ad alcuni aspetti cruciali come l'elasticità della domanda, la specificità degli asset necessari e la definizione dei diritti di proprietà. Il settore di appartenenza influenza l'investimento infrastrutturale poiché quest'ultimo è solitamente strettamente collegato alla specificità delle risorse necessarie alla realizzazione dell'opera. Alcuni settori poi sono ritenuti maggiormente proficui di altri, poiché contengono asset collaterali. Gli asset collaterali sono quegli asset non specifici e tangibili, in grado di sopportare alti livelli di indebitamento. Questa caratteristica infatti spiega l'impatto positivo di alcuni settori nell'attrarre iniziative in project finance. Le forze competitive all'interno di ogni settore sono fondamentali perché determinano il contesto in cui si opera, i rischi specifici a cui il progetto andrà incontro. Questi rischi sono associati alla possibile fluttuazione di alcuni fattori che determinano l'andamento economico dell'opera stessa. Un elemento competitivo caratteristico di ogni settore è la presenza di prodotti sostitutivi, il loro prezzo e l'elasticità incrociata che presentano con i prodotti e servizi che saranno messi sul mercato. I settori possono avere caratteristiche più o meno diverse anche a seconda del Paese in cui l'opera viene realizzato, poiché ciò è in stretto contatto con l'economia e le politiche locali. A seconda del settore di appartenenza si avrà l'utilizzo di diverse tecnologie, più o meno conosciute: tecnologie nuove possono portare a dei livelli di complessità maggiore, che aumentano la percezione del rischio. Di fatto, si può notare come la finanza di progetto venga utilizzata principalmente in settori tradizionali, con tecnologie conosciute. Ragionevolmente si pensa che la percezione del rischio diminuisca se questi sono da sopportare in un contesto noto, se la situazione risulta essere chiara, comprensibile. In accordo quindi con quanto detto finora, un'iniziativa di project finance basata su una tecnologia conosciuta avrà un rischio tecnologico minore, nonché si avrà su di essa una padronanza maggiore in quanto le informazioni a disposizione risultano essere maggiori.

Si può così concludere affermando che l'impatto del settore di appartenenza sull'attrattività del project finance può essere sia negativo che positivo a seconda delle caratteristiche del settore stesso, in funzione dei fattori sopra descritti.

La variabile SETTORE è una variabile osservabile categorica indipendente; nel modello verrà quindi convertita in tante variabile dummy quante sono le categorie identificate nei dati. Le variabili qualitative possono essere inserite come esplicative nei modelli statistici attraverso la codifica di tipo dummy: questa è configurata come una variabile binaria che assume valore 1 se l'unità statistica possiede una data caratteristica e 0 altrimenti.

## **5.5. Regione**

Per il fattore localizzazione è possibile effettuare un discorso simile a quello svolto per il settore di appartenenza dell'iniziativa. Infatti all'interno di uno stesso Paese è possibile evidenziare zone in cui sia più facile organizzare un'idea imprenditoriale, piuttosto che in un'altra. Un fattore che potrebbe spingere alla realizzazione di un'opera in una determinata area è la facilità con cui è possibile approvvigionarsi delle materie di input; si pensi quindi alla presenza di infrastrutture sul territori che possano facilitare le relazioni commerciali. Un altro esempio si trova pensando alle differenze che intercorrono per dover costruire un impianto industriale su un territorio pianeggiante piuttosto che su uno montuoso; quindi le caratteristiche geologiche sono un altro fattore determinante nella scelta delle localizzazione. Le caratteristiche geologiche devono però essere messe in relazione con i progetti stessi: se una strada è più facilmente costruibile in pianura, un impianto che converta la forza di un fiume in energia elettrica necessita di un terreno che presenti forte dislivello. Un terzo e ultimo fattore è determinato dalle utenze che usufruiranno dei servizi o dei prodotti erogati del progetto stesso; è bene che la scelta della localizzazione tenga in conto anche dove sono localizzate le persone che saranno la fonte delle entrate di cassa.

Concludendo si può dire che, come per la variabile settore, ci sia aspettano impatti sia positivi che negativi riguarda all'attrattività che questi possono suscitare per il project finance; ciò avviene in funzione dei fattori sopra descritti.

La variabile AREA è una variabile osservabile categorica indipendente; anche per questa variabile verranno create le variabili dummy.

## **5.6. Impegno degli investitori**

L'impegno degli investitori si riferisce alla presenza di detentori di capitale di equity e all'adozione di strumenti di trasferimento del rischio. I maggiori indicatori di questo impegno sono: il numero degli investitori, la presenza di fondi privati, la presenza del settore pubblico. Il capitale di equity è una delle possibili risorse finanziarie per un'iniziativa di business. Questo può essere raccolto attraverso diverse tipologie di attori, quali gli sponsor, le istituzioni finanziarie e i fondi privati. Gli investitori di equity ricevono una distribuzione degli utili solo dopo che sono stati ripagati i fornitori di debito, se è rimasta una componente positiva dei cash-flow generati; quindi la loro remunerazione risulta essere particolarmente legata alla profittabilità del progetto. È logico quindi supporre che gli investitori richiedano un ritorno favorevole per essere coinvolti all'interno di un'iniziativa. Gli sponsor sono in grado di condividere i rischi tra di loro, con gli altri partecipanti

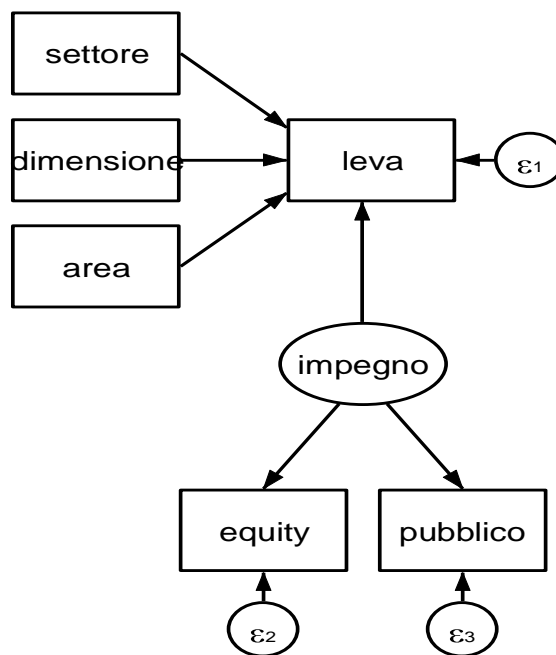
alla raccolta di equity, ma anche con i detentori di debito. Sotto l'assunzione di relazione positiva tra costi e dimensione dell'investimento, si può specificare che questi siano più elevati se associati a una singola firma su un dato investimento, piuttosto che la stessa somma ma ottenuta attraverso l'utilizzo di contratti multipli per lo stesso investimento. Ci si aspetta che il numero degli investitori sia collegato in modo positivo con l'attrattività del project finance, sia che siano detentori di equity o di debito, poiché ciò implica un minore rischio per l'equity e una maggiore garanzia per il debito. In secondo luogo bisogna considerare la presenza di fondi privati, che rappresentano un altro aspetto rilevante per determinare l'impegno degli investitori. Negli anni recenti i progetti infrastrutturali hanno attratto particolarmente questa categoria di investitori e, come notato precedentemente il project finance ha un largo utilizzo nella realizzazione di questa tipologia di opere. I fondi privati risultano come dei finanziatori puri e in questo differiscono dagli sponsor. Gli sponsor hanno la possibilità di essere coinvolti nello sviluppo del progetto, nelle decisioni a riguardo delle operations e potenzialmente hanno accesso alla cassa per compensare eventuali conti della società di progetto. I fondi privati invece sottoscrivono solo una componente residuale dei cash-flow, infatti, vengono pagati dopo tutte le altre parti (operatori, venditori, detentori di debito). Ovviamente il contratto viene stipulato allo scopo di ottenere dei ritorni interessanti. La sottoscrizione si accompagna di due aspetti positivi: il primo è la rappresentazione di una risorsa alternativa in un contesto in cui il settore privato presenta delle mancanze; il secondo è quello di dare valore alle applicazioni strategiche rendendole più veloci, meno costose e migliori. È anche possibile mettere in relazione la presenza di fondi privati con le dimensioni del progetto: ci si aspetta una maggiore presenza nei progetti di grandi dimensioni e viceversa. Per quanto esposto ci si aspetta che la presenza di investitori di fondi dia un'attrattività maggiore allo strumento finanziario PF, non solo a seguito del capitale versato, ma anche delle competenze fornite all'iniziativa. Il terzo e ultimo aspetto da affrontare per determinare l'impegno degli investitori è la presenza del settore pubblico, poiché anche questa variabile potrebbe influenzare l'attrattività di progetto. Il coinvolgimento del settore pubblico nella realizzazione di opere infrastrutturali è uno degli scopi principali in molti Paesi. I governi locali girano sul settore privato la realizzazione di molte opere di utilità pubblica, come strade, cimiteri, ospedali e così via. Il trend di austerità e crisi che caratterizza questo ultimo periodo, ha portato il settore pubblico a dare concessioni ai privati per le opere infrastrutturali, che solitamente vengono realizzate attraverso il project finance. Per attrarre investitori privati, i governi locali sono intervenuti con varie supporti che variano, da Paese in Paese e di progetto in progetto, sia nell'entità sia nella modalità. Le principali modalità di supporto risultano essere: finanziamenti diretti sotto forma di equity, incentivi statali che costituiscono una specie di scudo fiscale, garanzie per mitigare i rischi, sussidi.

Ricapitolando si può affermare che entità dell'equity, presenza di fondi privati e presenza del settore pubblico, determinano quello che viene definito l'impegno degli investitori. Per tutte e tre le categorie al crescere della loro presenza, cresce l'impegno totale. Si vuole poter misurare l'impegno in quanto incide in modo positivo sull'attrattività del project finance.

La variabile IMPEGNO è una variabile latente misurata attraverso tre indicatori, ovvero tre variabili misurabili. Le variabili indicatori sono osservabili e indipendenti e sono: EQUITY, FONDI\_PRIVATI e PUBBLICO.

## 5.7. Modello

Finora si è cercato di esplicitare una teoria che potesse essere la base di un modello statistico capace di determinare l'attrattività dello strumento finanziario project finance. Per quanto esposto in precedenza è possibile considerare il livello di leverage come una proxy dell'attrattività dello strumento finanziario studiato. Sono state individuate delle variabili che possono influenzare l'oggetto di studio in questione e attraverso relazioni teoriche ipotizzate tra queste, precedentemente enunciate, si è ora in grado di definire un modello. Si procederà inizialmente con un'interpretazione grafica del modello, per poi svolgere la conversione delle relazioni raffigurate in equazioni lineari, si procederà in seguito con l'analisi dei dati utilizzati, la stima dei parametri incogniti del modello e l'interpretazione dei risultati per capire se la teoria enunciata ha una conferma quantitativa nel modello.



A questo punto è necessario convertire il diagramma in equazioni, che il modello utilizzerà per stimare i parametri.

$$\begin{aligned}
 &LEVA = \\
 &= \text{coeff} * DIMENSIONE + \sum_{n-1} (\text{coeff}_i * SETTORE_i) + \sum_{m-1} (\text{coeff}_i * AREA_i) + \text{coeff} * IMPEGNO \\
 &\quad + ERRORE
 \end{aligned}$$

Il modello di misura per la variabile latente IMPEGNO è descritto nel seguito.

$$\text{PUBBLICO} = \text{funzione di IMPEGNO} + \text{ERRORE}$$

$$\text{EQUITY} = \text{funzione di IMPEGNO} + \text{ERRORE}$$

## 5.8. Analisi dei dati

### 5.8.1. Database iniziale

I dati sui quali si farà girare il modello sono presi da una raccolta contenente svariati progetti. La raccolta si intitola "Guida agli operatori del Project Finance" ed è redatta annualmente da Finlombardia. Lo scopo con il quale la guida è stata concepita è quello di diffondere la cultura di questo nuovo strumento finanziario e di favorire l'incontro tra pubblico e privato. L'edizione presa in considerazione è la nona, relativa all'anno 2011. Si vuole ora offrire una panoramica sui dati. Il numero di progetti a disposizione è di 1498.

**Leva.** In alcuni casi è dichiarato in modo diretto dal compilatore del questionario ma in altri casi si è provveduto alla sua determinazione attraverso i valori di finanziamento e equity. Per quanto riguarda i valori dichiarati sono presenti 824 missing. Sono 1600 i casi in cui è possibile calcolarlo, ovvero sono quei casi in cui sono presenti tutte le altre informazioni necessarie. Ovviamente sono stati riscontrati dei casi in cui i due valori differiscono. Non potendo indagare su ogni progetto quali siano le ragioni di questa discrepanza si è deciso di utilizzare il valore dichiarato dove presente e quello calcolato negli altri casi. In questo modo i dati mancanti risultano essere 472.

**Dimensione.** La variabile dimensione vuole quantificare la grandezza economica del progetto. Si ritiene che una buona approssimazione del valore economico di un business sia dato dal valore dell'investimento totale che bisogna sostenere per realizzarlo. La grandezza del progetto allora è assimilabile alla totalità del capitale necessario per realizzarlo. Per questa variabile i valori mancanti sono 168.

**Settore.** La variabile in questione indica l'ambito in cui è svolta la realizzazione dell'opera. In precedenza si è discusso del perché il project finance viene utilizzato particolarmente nella realizzazione di opere infrastrutturali. Queste però riguardo un ambito abbastanza ampio di settori, anche molto differenti tra loro. Di seguito ne è riportata una divisione.

SETTORE	NUMERO PROGETTI
Edilizia cimiteriale	30
Edilizia scolastica	7
Energia	940
Gestione e smaltimento rifiuti	44
Infrastrutture viarie	78
Parcheggi	48
Risorse idriche	36

Sanità	49
Sport e spettacolo	24
Sviluppo immobiliare	65
Telecomunicazioni	41
Trasporto ferroviario	8
Trasporto pubblico locale	18
Turismo e cultura	42
Altro	68
Totale	1498

Tabella 1. Settori di realizzazione dei progetti

**Area.** Anche questa variabile indica un ambito in cui viene intrapresa l'iniziativa, ovvero quello geografico. Come descritto prima la regione geografica è significativa in quanto delinea il contesto esterno, infatti la situazione socio-economica in Italia varia da regione a regione. Di seguito è riportata la ripartizione dei progetti nelle varie regioni italiane, dove specificato nella descrizione del progetto.

REGIONE	NUMERO PROGETTI
Abruzzo	19
Basilicata	15
Calabria	45
Campania	85
Emilia Romagna	69
Friuli Venezia Giulia	16
Lazio	78
Liguria	20
Lombardia	173
Marche	35
Molise	13
Piemonte	44
Puglia	235
Sardegna	75
Sicilia	139
Toscana	63
Trentino Alto Adige	10
Umbria	8
Valle d'Aosta	9
Veneto	52
Non disponibile	295
Totale	1498

Tabella 2. Regione di realizzazione dei progetti

**Pubblico.** Questa variabile indica la presenza del settore pubblico all'interno del progetto. Le caratteristiche della collaborazione pubblico-privato e le relative conseguenze sono già state esposte nella trattazione della teoria di base per il modello SEM; si ricorda che la presenza del settore pubblico è di fondamentale importanza in quanto modifica i flussi finanziari. La variabile PUBBLICO è di tipo binario: si assume un valore pari a 1 quando il pubblico è presente e il valore 0 quando il progetto è totalmente privato. Per questo dato non ci sono valori mancanti.



PROPIETÀ	NUMERO PROGETTI
Privata	1094
Pubblica	404
Totale	1498

Tabella 3. Partecipazione del pubblico

**Numero di sponsor e presenza di fondi.** Nella trattazione teorica del modello sono stati esposti come fattori della variabile latente IMPEGNO, il numero di sponsor che partecipano con una quota di equity e la possibile presenza di fondi privati. Si ricorda che il numero di sponsor favorisce la condivisione del rischio fornendo una visione caratterizzata da un rischio totale inferiore; la presenza di fondi privati invece garantisce prestazioni del progetto più efficienti, in quanto soggetto specializzato. Non esistono però informazioni a riguardo per i progetti censiti, o meglio: mentre per la prima l'informazione non è nemmeno stata richiesta, per la seconda è riportato un numero di progetti molto piccolo per il quale si riporta la presenza di fondi, ma per tutti gli altri progetti l'informazione non è riportata. In realtà è possibile ovviare alla mancanza di una delle due variabili, ovvero a quella del numero di sponsor: la soluzione è quella di utilizzare una variabile che possa sostituirla con una che abbia lo stesso significato all'interno del modello. L'impegno dei finanziatori può essere misurato anche attraverso il valore dell'equity all'interno della struttura finanziaria, infatti più gli sponsor immettono equity nel progetto, più si prendono in carico il rischio, meno sarà la parte di rischio rimasta da condividere.

## 5.8.2. Preparazione del dataset

Al dataset iniziale sono state apportate delle variazioni per renderlo utilizzabile all'interno della stima. In particolare sono state operate due modifiche: eliminazione dei missing e riduzione delle categorie all'interno delle variabili presenti nel modello.

**Missing.** Come si è detto la prima modifica riguarda l'eliminazione dei dati mancanti. In precedenza, per ogni variabile citata facente parte del modello, è stato riportato il numero dei progetti per il quale il valore risultava mancante. La decisione che è stata presa è quella di non considerare nella stima dei parametri tutti quei progetti che avessero anche solo un valore non riportato. Si è quindi passati da un numero di osservazioni (progetti) iniziali di 1498 a quello di 846. Ora per ogni progetto sono presenti tutte le informazioni.

**Riduzione del numero di categorie.** Nel modello sono presenti tre variabili discrete; di queste una risulta essere binaria (PUBBLICO), mentre le altre due sono categoriche (SETTORE e AREA). La riduzione quindi va operata solamente su quest'ultime due. Per quanto riguarda il settore di appartenenza del progetto, data la divisione troppo fine, si è deciso di costruire dei macro settori: si sono raccolte le 15 categorie iniziali in sole 5 macrocategorie. Il raggruppamento segue perfettamente la classificazione fornita nel primo capitolo al paragrafo "Settori di Impiego". La classificazione prevedeva cinque categorie: energia e ambiente in cui sono presenti i progetti energetici e di gestione e smaltimento dei rifiuti; risorse idriche; trasporti in cui sono raccolti le infrastrutture viarie, il trasporto ferroviario e il trasporto pubblico locale; sport, spettacolo, cultura

e turismo; opere pubbliche varie in cui sono raccolte tutti gli altri settori sopra elencati. Si può così ricostruire la tabella riassuntiva riportando la suddivisione finale.

SETTORI	CATEGORIA	NUMERO PROGETTI
Energia e gestione dei rifiuti	1	606
Risorse idriche	2	25
Trasporti	3	42
Sport, spettacolo, turismo e cultura	4	36
Opere pubbliche varie	5	137

Tabella 4. Categorie dei settori

Anche per questa categoria, invece che utilizzare una classificazione troppo fine, si è voluto convogliare le varie regioni italiane in macroaree, caratterizzate quindi da una stessa situazione dal punto di vista economico e infrastrutturale (per i dettagli si rimanda al paragrafo dedicato in precedenza). Si hanno così ancora una volta cinque macroaree, ovvero: nord-est, nord-ovest, centro, sud, isole. Le regioni sono state così raggruppate:

- Del nord-est fanno parte Emilia Romagna, Friuli Venezia Giulia, Trentino Alto Adige e Veneto;
- A nord-ovest si trovano Liguria, Lombardia, Piemonte e Valle d'Aosta;
- Del settore centrale fanno parte Lazio, Marche, Toscana e Umbria;
- Al sud si trovano Abruzzo, Basilicata, Calabria, Campania, Molise e Puglia;
- Infine il settore insulare con Sardegna e Sicilia.

Riassumendo i progetti in una nuova tabella si ottiene la seguente divisione.

AREA	CATEGORIA	NUMERO PROGETTI
Nord-Est	1	106
Nord-Ovest	2	168
Centro	3	119
Sud	4	286
Isole	5	167

Tabella 5. Categorie delle aree geografiche

All'interno del software le dummies vanno generate attraverso degli appositi comandi che generano tante dummies quante sono le categorie presenti. Quando poi queste verranno utilizzate all'interno del modello bisognerà inserire  $N - 1$  dummies. Ciò è necessario in quanto inserendole tutte, si avrebbe una colonna come combinazione lineare delle altre e di conseguenza la matrice risulterebbe non invertibile. Come esposto precedentemente, una matrice non invertibile implica la non possibilità di ottenere delle stime. Si riporta di seguito la generazione delle dummies per le tre variabili discrete sopra citate all'interno del software Stata, con il quale in seguito si procederà alla stima dei parametri.

2 . tabulate settore, generate (settore)			
settore	Freq.	Percent	Cum.
1	606	71.63	71.63
2	25	2.96	74.59
3	42	4.96	79.55
4	36	4.26	83.81
5	137	16.19	100.00
Total	846	100.00	

3 . tabulate pubblico, generate (pubblico)			
pubblico	Freq.	Percent	Cum.
1	669	79.08	79.08
2	177	20.92	100.00
Total	846	100.00	

4 . tabulate area, generate (area)			
area	Freq.	Percent	Cum.
1	106	12.53	12.53
2	168	19.86	32.39
3	119	14.07	46.45
4	286	33.81	80.26
5	167	19.74	100.00
Total	846	100.00	

Tabella 6. Generazione delle variabili dummy

## 5.9. Stima

Ora il dataset può ritenersi pronto per l'utilizzo all'interno del modello per la stima dei coefficienti. Prima di procedere con il processo di stima, è necessario premettere che il dataset sul quale si sta lavorando risulta essere poco corposo, con poche informazioni. Per questo motivo il software scelto (Stata 12.0) non è in grado di stimare i coefficienti lanciando direttamente un modello SEM. infatti, lanciando l'intero modello si ottiene una soluzione non convergente; la stima si ferma solo al raggiungimento delle iterazioni massime. Si è allora proceduto con la scomposizione del modello con lo scopo di trovare la problematica. Così si inizia lanciando la sola analisi fattoriale.

```

8 . sem (impegno -> equity) (impegno -> pubblico1), iterate(160)

Endogenous variables
Measurement:  equity pubblico1

Exogenous variables
Latent:      impegno
Fitting target model:

Iteration 0:  log likelihood = -6596.6058 (not concave)
Iteration 1:  log likelihood = -6596.6058 (not concave)
Iteration 2:  log likelihood = -6596.6058 (not concave)
Iteration 3:  log likelihood = -6596.6058 (not concave)
Iteration 4:  log likelihood = -6596.6058 (not concave)
Iteration 5:  log likelihood = -6596.6058 (not concave)
Iteration 6:  log likelihood = -6596.6058 (not concave)
Iteration 7:  log likelihood = -6596.6058 (not concave)
Iteration 8:  log likelihood = -6596.6058 (not concave)
Iteration 9:  log likelihood = -6596.6058 (not concave)
Iteration 10: log likelihood = -6596.6058 (not concave)
Iteration 11: log likelihood = -6596.6058 (not concave)
Iteration 12: log likelihood = -6596.6058 (not concave)
Iteration 13: log likelihood = -6596.6058 (not concave)
Iteration 14: log likelihood = -6596.6058 (not concave)
Iteration 15: log likelihood = -6596.6058 (not concave)
Iteration 16: log likelihood = -6596.6058 (not concave)
Iteration 17: log likelihood = -6596.6058 (not concave)
Iteration 18: log likelihood = -6596.6058 (not concave)
Iteration 19: log likelihood = -6596.6058 (not concave)
Iteration 20: log likelihood = -6596.6058 (not concave)
Iteration 21: log likelihood = -6596.6058 (not concave)
Iteration 22: log likelihood = -6596.6058 (not concave)
Iteration 23: log likelihood = -6596.6058 (not concave)
Iteration 24: log likelihood = -6596.6058 (not concave)
Iteration 25: log likelihood = -6596.6058 (not concave)
Iteration 26: log likelihood = -6596.6058 (not concave)
Iteration 27: log likelihood = -6596.6058 (not concave)
Iteration 28: log likelihood = -6596.6058 (not concave)
Iteration 29: log likelihood = -6596.6058 (not concave)
Iteration 30: log likelihood = -6596.6058 (not concave)
    
```

Tabella 7. Non convergenza del modello fattoriale

Come si evince dal risultato, la parte di analisi fattoriale non riesce a giungere la convergenza poiché nel software il metodo utilizzato è quello della massima verosimiglianza, che risulta avere delle difficoltà quando le variabili di misura sono troppo poche. In questo caso sono presenti solo due variabili sottostanti alla variabile impegno, mentre si necessita di almeno tre misure. Per questo motivo si è proceduto con l'analisi delle componenti principali. Questa metodologia appartiene all'analisi fattoriale ed è utilizzata nell'ambito della statistica multivariata allo scopo di ridurre un numero elevato di variabili in un numero minore di variabili latenti. In questo caso si passerà da due variabili indipendenti e osservabili (EQUITY e PUBBLICO) ad un solo fattore, l'IMPEGNO. Di seguito vengono riportati i risultati dell'analisi fattoriale.

```

15 . pca equity pubblico1

Principal components/correlation          Number of obs   =    846
                                         Number of comp. =     2
                                         Trace           =     2
                                         Kmo             =    1.0000

Notation: (unrotated = principal)

```

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	1.15127	.302531	0.5756	0.5756
Comp2	.848734	.	0.4244	1.0000

```

Principal components (eigenvectors)

```

Variable	Comp1	Comp2	Unexplained
equity	-0.7071	0.7071	0
pubblico1	0.7071	0.7071	0

Tabella 8. Analisi delle componenti principali

Quello che è stato ora illustrato è la stima del modello di misura. Precedentemente si è discusso a proposito della composizione di un modello ad equazioni strutturali, ovvero è diviso in due parti: il modello di struttura e quello di misura. Come si può notare le variabili spiegano solo il 58% del fattore. Ciò potrebbe essere dato dal fatto che le due variabili di misura non presentano una correlazione tra loro abbastanza elevata da essere sottostanti ad un fattore comune. Per questo motivo il modello di misura non risulta essere particolarmente significativo. Avere un modello di analisi fattoriale non significativo può significare che non ci sia particolare differenza a fronte o meno del suo utilizzo. In altre parole si sta affermando che probabilmente introdurre una variabile latente di questo tipo all'interno di una regressione è molto simile ad una regressione pura che regredisce direttamente sulla variabile, piuttosto che sul fattore. Si procederà comunque alla determinazione della variabile latente IMPEGNO, per poi inserirla all'interno del modello di regressione lineare dove la variabile dipendente è la LEVA, come esposto precedentemente. In seguito si effettuerà una regressione semplice verso tutte le variabili indipendenti per verificare se effettivamente i modelli differiscono in modo non significativo.

Si è affermato che, grazie all'analisi fattoriale, si è in grado di produrre una variabile (prima latente) che avrà stessa dimensione delle stesse variabili. Essenzialmente, per tutti i 486 progetti per i quali viene indicato sia l'equity, che la presenza del settore pubblico, viene creato un valore

che indica l'impegno profuso dagli sponsor nel progetto stesso. A questo punto allora è possibile procedere con la regressione lineare. Questa avrà come variabile dipendente la LEVA, che regredirà su alcune variabili osservabili indipendenti (SETTORE, AREA e DIMENSIONE) e su una variabile latente determinata attraverso l'analisi fattoriale (IMPEGNO). Di seguito si riporta la regressione corrispondente e i relativi risultati.

7 . regress leva dimensione settore2 settore3 settore4 settore5 area2 area3 area4 area5 impegno						
Source	SS	df	MS			
Model	11.9324706	10	1.19324706	Number of obs =	846	
Residual	20.6886564	835	.024776834	F( 10, 835) =	48.16	
Total	32.621127	845	.038604884	Prob > F =	0.0000	
				R-squared =	0.3658	
				Adj R-squared =	0.3582	
				Root MSE =	.15741	

leva	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
dimensione	.000028	.0000158	1.78	0.076	-2.92e-06	.0000589
settore2	-.1129668	.0362708	-3.11	0.002	-.1841595	-.0417741
settore3	-.2333585	.0319059	-7.31	0.000	-.2959837	-.1707334
settore4	-.1172189	.0289876	-4.04	0.000	-.1741161	-.0603217
settore5	-.1449403	.0176915	-8.19	0.000	-.1796653	-.1102154
area2	.019412	.0198387	0.98	0.328	-.0195276	.0583515
area3	-.0102439	.0212577	-0.48	0.630	-.0519687	.0314809
area4	-.0494675	.0187376	-2.64	0.008	-.0862459	-.0126891
area5	-.0894237	.0203566	-4.39	0.000	-.1293798	-.0494675
impegno	.0735051	.0096437	7.62	0.000	.0545763	.0924339
_cons	.8459078	.0168619	50.17	0.000	.8128111	.8790045

Tabella 9. Regressione lineare di LEVA

## 5.10. Interpretazione

### 5.10.1. Coefficienti

La *Tabella 9* riporta i risultati della regressione lineare multipla della variabile dipendente LEVA, contro le variabili indipendenti SETTORE, AREA e DIMENSIONE, oltre alla variabile latente IMPEGNO precedentemente determinata attraverso l'analisi fattoriale. I regressori risultano essere nove, allora i parametri stimati risultano essere dieci; infatti comprendono i coefficienti dei regressori più l'intercetta. Stimare i parametri significa scegliere una retta di regressione basata sull'evidenza empirica. La stima dei parametri si basa sul criterio dei minimi quadrati, dove si cerca di minimizzare la sommatoria degli errori elevati al quadrato; gli errori sono la differenza tra i valori osservati e quelli stimati:

$$\sum(y - \hat{y})^2$$

Il coefficiente rappresenta il senso della relazione tra una variabile indipendente e quella dipendente; in altre parole è l'espressione di quanto varia la LEVA in funzione di uno scostamento unitario della variabile indipendente corrispondente. Il modello ci dice che all'aumentare della dimensione del progetto di un'unità, la leva del progetto aumenta sensibilmente, infatti il valore del coefficiente risulta essere molto piccolo. L'ipotesi fatta nello sviluppo della teoria finanziaria è comunque rispettata: si era ipotizzato che al cresce della dimensione del progetto ci potesse

essere maggiore interesse a utilizzare la finanza di progetto rispetto alla corporate finance tradizionale. Lo stesso discorso si può applicare per la variabile che descrive l'impegno dei finanziatori, infatti il coefficiente è anche in questo caso positivo ma di impatto decisamente maggiore; si sta affermando che all'aumentare unitario della variabile impegno, aumenta quasi in modo unitario anche la leva utilizzata nel progetto. Ancora l'ipotesi teorica è confermata: si attendeva un impatto positivo e proprio questo risultato è stato ottenuto. Per ora sono stati commentati i coefficienti di variabili continue, ma ora è necessario affrontare la valutazione dei coefficienti delle variabili categoriche. Innanzitutto è necessario ricordare che all'interno del modello vanno inserite tante dummies quante sono le categorie meno una; quest'ultima dummy che viene esclusa dal modello è la variabile di riferimento. Ogni variabile dummy rappresenta una modalità di una variabile indipendente non metrica e ciascuna variabile indipendente non metrica, con  $k$  categorie, può essere rappresentata attraverso  $k - 1$  variabili dummy. Esistono due forme per la rappresentazione di variabili dummy, ma la più comune è quella in cui ogni categoria è codificata con 1 o con 0. I coefficienti di regressione delle variabili dummy rappresentano le differenze tra le medie, calcolate sulla variabile dipendente, di ciascun gruppo di rispondenti individuati dalle diverse variabili dummy e la categoria di riferimento (vale a dire il gruppo omissivo). Le differenze tra i gruppi possono essere valutate direttamente poiché i coefficienti di regressione sono espressi nella stessa unità di misura della variabile dipendente. È importante capire che le variabili dummy non modificano la natura della relazione, ma attribuiscono ai diversi gruppi soltanto differenti intercette. Questa forma di codifica è particolarmente appropriata quando esiste un gruppo di riferimento logico (che nel nostro caso è rappresentato dall'energia per la variabile settore e al Nord-Est per l'area). Ogni volta che si ricorre alla codifica tramite variabili dummy, bisogna essere consapevoli dell'esistenza del gruppo di confronto e ricordare che i coefficienti di regressione rappresentano le differenze tra le medie dei gruppi rispetto a quella del gruppo di riferimento. È possibile notare come la media dei valori di leva dei progetti realizzati nel settore dell'energia sia maggiore rispetto a quelle di tutti gli altri valori. Il settore con la leva media minore risulta essere quello dei trasporti. Vengono riportati in tabella il valore delle medie nei rispettivi settori.

SETTORE	LEVA MEDIA
Energia	0,83
Risorse idriche	0,62
Trasporti	0,63
Sport, cultura, turismo	0,67
Altre opere pubbliche	0,78

Tabella 10. Leve medie in ogni settore

Lo stesso discorso può essere effettuato sull'area geografica di sviluppo e realizzazione dei progetti. Si ricorda che l'area di riferimento è il Nord-Est. Nella Tabella 9 si legge che l'unica area ad avere una leva media maggiore è il Nord-Ovest, ma soprattutto che più si scende lungo la penisola più il valore della leva media diminuisce. Si riportano ancora i valori in una tabella.

AREA	LEVA MEDIA
Nord-Est	0,7781
Nord-Ovest	0,7784
Centro	0,7784
Sud	0,7765
Isole	0,7772

Tabella 11. Leve medie in ogni area

### 5.10.2. Significatività

Una volta interpretato il valore del coefficiente è necessario determinare la sua significatività. In altre parole si vuole implementare un test che verifichi se il coefficiente sia ragionevolmente diverso da zero e se, di conseguenza, abbia realmente un ruolo all'interno del modello. L'ipotesi da testare è la seguente:

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0$$

La statistica è rappresentata da una  $t$  di Student, il cui valore è calcolato direttamente dal software e riportato nella *Tabella 9*, con il corrispondente p-value. Il livello di confidenza utilizzato nel modello è del 95% pari a  $(1 - \alpha)$ , di conseguenza è possibile calcolare il valore di  $\alpha$  da confrontare con quello del p-value.

$$\alpha = 5\%$$

$$\frac{\alpha}{2} = 2,5\%$$

Nel caso in cui il valore del p-value sia minore di  $\alpha/2$  è possibile rifiutare l'ipotesi  $H_0$ , quindi affermare che si può rifiutare l'ipotesi nulla. Tornando alla *Tabella 9* è possibile verificare la significatività di ogni coefficiente. Il p-value di dimensione è molto vicino al limite, quindi per determinare l'effettiva significatività del coefficiente è necessario svolgere un'analisi più approfondita, che si rimanda; per il momento si è deciso di lasciare la variabile all'interno del modello. Per tutte le altre variabili c'è significatività statistica in quanto il p-value risulta essere uguale a zero. Un discorso a parte va svolto per la variabile area che probabilmente risulta avere poca significatività, infatti alcune dummies hanno valori di p-value molto alti. Un dubbio poteva sorgere anche confrontando il valore dei coefficienti con le medie riportate in tabella.

### 5.11. Bontà del modello

La bontà del modello di adattarsi è determinata attraverso degli indici, che sono stati abbondantemente trattati nella parte teorica. L'indice più utilizzato è il coefficiente  $R^2$ ; questo misura la quantità di variabilità spiegata dalla regressione in termini percentuali.  $R^2$  è quindi un indice che può assumere dei valori compresi tra 0 e 1; il modello sarà migliore quanto più l'indice

sarà vicino ad 1. Nel caso studiato si ottiene un indice abbastanza basso, pari a 0,3658. Il valore basso non deve sorprendere più di tanto, infatti all'inizio della trattazione si è affermato che il dataset non è molto corposo. Un altro indice che è interessante osservare è la versione corretta dell'indice appena citato. La correzione consiste nel diminuire l'indice al crescere del numero dei parametri inseriti. È ovvio come si possa aumentare l'adattamento del modello inserendo sempre più variabili di regressione, d'altra parte però si sta fittando molto bene il campione ma si avranno dei risultati scarsi quando lo si applicherà ad altri campioni appartenenti alla stessa popolazione. Il valore corretto risulta essere 0,3592; quindi non troppo lontano dal precedente valore.

## **5.12. Conclusioni**

Al termine di questa trattazione applicativa è importante valutare il lavoro svolto con spirito critico, per avere la possibilità, in futuro, di arricchire la ricerca. La problematica principale del modello sviluppato risiede nella natura e nella qualità dei dati e delle informazioni a disposizione. Si ricorda che questi dati utilizzati sono stati estratti da una guida che raccoglie le risposte a questionari volontari che i promotori hanno sottoposto alle banche. Pertanto, appare necessario uno studio più dettagliato della tematica finanziaria con l'obiettivo di raccogliere un numero maggiore di informazioni per ogni progetto, o di ottenere informazioni maggiormente significative rispetto a quelle al momento presenti. Una possibile giustificazione a questa carenza informativa si può ricercare nella relativa giovinezza dello strumento finanziario, per il quale non sono ancora stati creati database ricchi e consistenti, che possano essere efficaci all'interno di uno studio di inferenza statistica. Inoltre, sarebbe opportuno avere la possibilità di inserire nel dataset anche progetti appartenenti alle stesse categorie infrastrutturali di realizzazione, per i quali però è stato scelto uno strumento di finanziamento differente. In questo modo si avrebbe l'opportunità di confrontare lo strumento finanziario in sé, con le differenti alternative di scelta presenti sul mercato.



## 6. BIBLIOGRAFIA

Per la trattazione sul Project Finance e i Project Bonds.

BASSANINI, F.; DEL BUFALO, G.; REVIGLIO, E. (2011). *Financing Infrastructure in Europe Project Bonds, Solvency II and the "Connecting Europe" Facility*. Astrid Rassegna.

BODNAR, G.M.; COMER, B. (1996). *Project Finance Teaching Notes*. The Wharton School.

BREALEY, R.A.; COOPER, I.A.; HABIB, M.A.(1996). *Using Project Finance To Fund Infrastructure Investments*. Journal of Applied Corporate Finance.

DAILAMI, M.; HAUSWALD, R. (2003). *The Emerging Project Bond Market: Covenant Provisions and Credit Spreads*. World Bank Policy Research Working Paper 3095, July.

DEMIRAG, I; KHADAROO, I; STAPLETON, P; STEVESON, C.(2011). *Public Private Partnership Financiers' Perceptions Of Risks*. The Institute of Chartered Accountants of Scotland, Edinburgh.

EPEC (2010). *Capital Markets In PPP Financing*. EIB, Luxembourg

ESTY, B. J. (2003). *The Economic Motivations for Using Project Finance*. Harvard Business School. Boston, United States.

GARDNER, D; WRIGHT, J. (2010). *Project Finance*. Encyclopedia Of Debt Finance, Euromoney Institutional Investor.

GATTI, S. (2008). *Project Finance in Theory and Practice: Designing, Structuring, and Financing Private and Public Projects*. Academic Press Advanced Series, Elsevier Inc.

GROOBEY, C.; PIERCE, J.; FABER, M.; BROOME, G. (2010). *Project Finance Primer for Renewable Energy and Clean Tech Projects*. Wilson Sonsini Goodrich & Rosati.

LAM, P.T.I.; CHIANG, Y.H.; CHAN, S.H. (2011). *Critical Success Factors for Bond Financing of Construction Projects in Asia*. Journal of Management Engineering.

LATHAM AND WATKINS, (2009). *Why Project Bonds Now?* Client Alert, number.

MERTON, R.C. (1974). *On The Pricing of Corporate Debt: The Risk Structure of Interest Rates*. Journal of Finance.

MOODY'S (2010). *Generic Project Finance Methodology*. Global Project Finance, Rating Methodology.

REGAN,M.; SMITH, J.; LOVE, P. (2011). *Infrastructure procurement: learning from private and public partnership experiences "down under"*. Environment and Planning C: Government and Policy.

SCANNELLA, E. (2012). *Project Finance In The Energy Industry: New Debt-Based Financing Models*. International Business Research.

STANDARD & POOR'S (2011). Industry Report Card: Global Project Finance Rating Activity is picking up, New York.

TOFT, M.A. (2008). *Project Finance And Political Risk - An Empirical Study Of The Relationship Between Project Finance And Political Risk*. Århus School of Business.

Per la trattazione sui SEM.

ANDERSON, J. C.; GERBING, D. W. (1984). *The Effects Of Sampling Error On Convergence, Improper Solutions, And Goodness-Of-Fit Indices For Maximum Likelihood Confirmatory Factor Analysis*. Psychometrika, 49, 155-173.

BEARDEN, W. D.; SHARMA, S.; TEEL, J. E. (1982). *Sample Size Effects On Chi-Square And Other Statistics Used In Evaluating Causal Models*. Journal of Marketing Research, 19, 425-430.

BENTLER, P. M. (1992). *On The Fit Of Models To Covariances And Methodology To The Bulletin*. Psychological Bulletin, 112, 400-404.

BENTLER, P. M.; BONETT, D. G. (1980). *Significance Tests And Goodness-Of-Fit In The Analysis Of Covariance Structures*. Psychological Bulletin, 88, 588-606.

BENTLER, P. M.; WEEKS, D. G. (1980). *Linear Structural Equations With Latent Variables*. Psychometrika, 45, 289-307.

BOLLEN, K. A. (1986). *Sample Size And Bentler And Bonett's Nonnormed Fit Index*. Psychometrika, 51, 375-377.

BOLLEN, K. A.; ARMINGER, G. (1991). *Observational Residual In Factor Analysis And Structural Equation Models*. In P.V. Marsden (Ed.), Sociological methodology 1991 (Vol. 21, pp. 235-262). Oxford: Blackwell.

BOLLEN, K. A.; STINE, R. A. (1992). *Bootstrapping Goodness-Of-Fit Measures In Structural Equation Models*. Sociological Methods and Research, 21, 205-229.

BRECKLER, S. J. (1990). *Applications Of Covariance Structure Modeling In Psychology: Cause For Concern?* Psychological Bulletin, 107, 260-273.

BROWN, T. A. (2006). *Confirmatory Factor Analysis For Applied Research*.

BROWNE, M. W. (1984). *Asymptotically Distribution-Free Methods For The Analysis Of Covariance Structures*. British Journal of Mathematics and statistical psychology, 37, 62-83.

- CUDECK, R.; BROWNE, M. W. (1983). *Cross-Validation Of Covariance Structures*. *Multivariate Behavioral Research*, 18, 147-167.
- CUDECK, R.; HENLY, S. J. (1991). *Model Selection In Covariance Structures Analysis And The "Problem" Of Sample Size: A Clarification*. *Psychological Bulletin*, 109, 512-519.
- CURRAN, P. J.; WEST, S. G.; FINCH, J. F. (1994). *The Robustness Of Test Statistics And Goodness-Of-Fit Indices In Confirmatory Factor Analysis*. Manuscript submitted for publication.
- DAVEY A., SALVA J. (2010). *Statistical Power Analysis with Missing Data. A Structural Equation Modeling Approach*. Routledge, Taylor & Francis Group.
- EFRON, B.; TIBSHIRANI, R. (1986). *Bootstrap Methods For Standard Errors, Confidence Intervals And Other Measures Of Statistical Accuracy*. *Statistical Science*, 1, 54-74.
- EMERSON, J. D.; STOTO, M. A. (1983). *Understanding Robust And Exploratory Data Analysis* (pp. 97-127). New York: Wiley.
- HANCOCK, G. R.; MUELLER, R. O. (2006). *Structural Equation Modeling. A second Course*. Università del Maryland e the George Washington University.
- HOYLE, R. H. (1995). *Structural Equation Modeling. Concepts, Issues and Applications*. SAGE Publications, International Educational and Professional Publisher.
- HU, L. T.; BENTLER, P. M.; KANO, Y. (1992). *Can Test Statistics In Covariance Structure Analysis Be Trusted?* *Psychological Bulletin*, 112, 351-362.
- JORESKOG, K. G.; SORBOM, D. (1984). *LISREL VI user's guide (3rd ed.)*. Moorsville, IN: Scientific Software.
- JORESKOG, K. G.; SORBOM, D. (1993). *LISREL 8: Structural Equation Modeling With The SIMPLIS Command Language*. Hillsdale.
- KLINE, R. B. (2010). *Principles and Practice of Structural Equation Modeling. Terza edizione*. Series Editor's Note by Todd D. Little.
- LEE, S. Y.; HERSHBERGER, S. (1990). *A Simple Rule For Generating Equivalent Models In Covariance Structure Modeling*. *Biometrika*, 57, 239-251.
- MARCOULIDES, G. A., SCHUMACKER, R. E. (2009). *New Developments And Techniques In Structural Equation Modeling*. Lawrence Erlbaum Associates, Publishers Mahwah, New Jersey.
- MCCALLUM, R. C. (1986). *Specification Searcher In Covariance Structure Modeling*. *Psychology*, 2, 13-20.
- MCCALLUM, R. C.; ROZNOWSKI, M.; NECOWITZ, L. B. (1992). *Model Modifications In Covariance Structure Analysis: The Problem Of Capitalization On Chance*. *Psychological Bulletin*, 111, 490-504.

- MUTHÉN, B. (1993). *Latent Variable Structural Equation Modeling With Categorical Data*. Journal of Econometrics 22, pag 43-65. North-Holland Publishing Company.
- RAYKOV, T.; MARCOULIDES, G. A. (2006). *A First Course In Structural Equation Model. Seconda Edizione*. Michigan State University e California State University, Fullerton.
- SATORRA, A.; BENTLER, P. M. (1990). *Model Conditions For Asymptotic Robustness In The Analysis Of Linear Relations*. Computational Statistics and Data Analysis, 10, 235-249.
- SATORRA, A.; BENTLER, P. M. (1994). *Corrections To Test Statistics And Standard Errors On Covariance Structure Analysis*. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis* (pp. 399–419). Thousand Oaks, CA: Sage.
- SCHUMACKER, R. E.; LOMAX, R.G. (2004). *A Beginner's Guide to structural equation modeling, seconda edizione*. Routledge Academic.
- STELZL, I. (1986). *Changing Causal Relationship Without Changing The Fit: Some Rules For Generating Equivalent LISREL-Models*. Multivariate Behavioral Research, 21, 309-331.
- THOMPSON, B. (2004). *Exploratory And Confirmatory Factor Analysis. Understanding Concepts And Applications*. American Psychological Association, Washington DC.
- TUCKER, L. R.; LEWIS, C. (1973). *A Reliability Coefficient For Maximum Likelihood Factor Analysis*. Psychometrika, 38, 1-10.



## **RINGRAZIAMENTI**

Per cominciare vorrei ringraziare il Professor Rocco Roberto Mosconi per avermi dato la possibilità di svolgere il lavoro conclusivo dei miei studi in una materia che ha saputo suscitare la mia attenzione sin dalle prime lezioni, ma anche per aver avuto la disponibilità di venire incontro alle mie esigenze.

Un ringraziamento particolare va alla mia famiglia, ad ogni suo componente. In particolare i miei genitori che hanno contribuito al raggiungimento di questo importante obiettivo, supportandomi moralmente ed economicamente in tutti questi lunghi anni. Sono sicura che, nonostante alcune volte ci possano essere delle piccole incomprensioni, saranno sempre al mio fianco felici di potermi aiutare in ogni difficoltà. Per questo spero che possano essere fieri di me e dei miei risultati.

Non posso ora non pensare agli amici, a tutte quelle persone con cui trascorro il quotidiano, le serate e il tempo libero. Con loro ho la fortuna di poter condividere momenti difficili, tristi, ma soprattutto felici e divertenti. Li vorrei ringraziare per come sono, per quello che fanno per me e vorrei che sapessero che gli voglio bene. Una riferimento speciale va alle amiche del liceo, a quei legami che non si sa in quale modo creano dei rapporti che non potranno mai finire: ci sarà sempre un filo che ci legherà e niente e nessuno potrà romperlo; nemmeno un oceano. Possiamo stare lontane o non sentirci, ma in ogni occasione in cui ci vediamo o sentiamo è come tornare ai tempi in cui su quei banchi di scuola condividevamo davvero tutto, nulla è cambiato.

Infine, ma non sicuramente per importanza, devo ringraziare la persona più speciale del mondo, in grado di ascoltarmi, di capirmi, di farmi ridere, di supportarmi e di sopportarmi, di farmi sentire unica, di risolvere i miei problemi e di rendere anche la più piccola cosa un ricordo fantastico, ma anche di un altro milione di cose; potrei andare avanti per ore e ore. Con te ogni momento è unico e indimenticabile: lo sai che mi ricordo tutto! Stiamo condividendo un percorso importante, in cui siamo cresciuti e diventati quelli che siamo oggi; io sono molto orgogliosa dell'uomo che sei diventato e spero che tu lo possa essere di me almeno la metà di quanto lo sono io. Non riesco ad immaginare come potrei essere più felice. Ti amo.

Quindi grazie a tutti voi, per essere andati oltre alle apparenze, per aver avuto la voglia e la pazienza di conoscere la persona che sono davvero. Grazie del vostro sostegno.