

UNIVERSITA' DEGLI STUDI DI PADOVA

FACOLTA' DI SCIENZE STATISTICHE
Corso di laurea in Scienze Statistiche
Economiche Finanziarie e Aziendali

“L’influenza della nascita di un figlio sul
benessere delle famiglie italiane”

Relatore: Prof.ssa Anna Giraldo

Laureanda: Lucia Bernardi

Anno accademico 2010/2011

INDICE

INTRODUZIONE.....	5
CAPITOLO 1: LETTERATURA.....	15
CAPITOLO 2: I DATI.....	28
2.1 L'indagine sui bilanci delle famiglie italiane della Banca D'Italia.....	28
2.2 Il campione selezionato.....	31
2.3 Analisi descrittiva.....	32
2.3.1 Caratteristiche del capofamiglia.....	34
2.3.2 Caratteristiche del coniuge.....	35
2.3.3 Caratteristiche della famiglia.....	35
2.3.4 Confronto tra il campione analizzato e i campioni sezionali degli anni presi in esame.....	37
CAPITOLO 3: STIMADEL MODELLO.....	40
3.1 Approccio empirico: stima del <i>propensity score</i> <i>matching</i>	40
3.2.1 Stima del <i>propensity score</i>	46
3.2.2 Stima dell'impatto.....	49
3.2.3 Sensibilità del <i>propensity score</i> alla specificazione del modello.....	55
CAPITOLO 4: ANALISI DI STABILITA'	56
CONCLUSIONI.....	63
APPENDICE	65
BIBLIOGRAFIA.....	67

INTRODUZIONE

L'obiettivo di questo lavoro è quello di capire se e come la nascita di un figlio possa influenzare il benessere di una famiglia.

L'andamento della nascita di un figlio è principalmente legato a tre fattori di ordine economico: l'occupazione femminile, la flessibilità del lavoro, le politiche di sostegno alla famiglia adottate nel paese in esame.

Gli studi più recenti degli organismi internazionali rilevano che i paesi caratterizzati da una minore partecipazione delle donne al mercato del lavoro sono quelli che otterrebbero dall'aumento dell'occupazione femminile un maggior vantaggio in termini di crescita (Nota Aggiuntiva "Donne, Innovazione, Crescita", Rapporto sullo stato di attuazione del PNR 2006-2008, Dipartimento Politiche Comunitarie, 2007).

Risulta che il lavoro femminile non è più un ostacolo alla natalità; anzi, si dimostra che oggi nei paesi avanzati, a differenza di quanto avveniva in passato, se le donne hanno meno opportunità di occupazione fanno meno figli. Viceversa, la fecondità è maggiore nei paesi ad elevata occupazione femminile. Gli studi sottolineano che i paesi con i tassi d'occupazione più bassi e con un tasso di natalità inferiore sono quelli che hanno una copertura di servizi più bassa, che presentano una minore disponibilità dei padri a prendere congedi parentali, dove le donne hanno un maggior carico di lavoro domestico, dove è più bassa la condivisione del lavoro di cura tra uomini e donne.

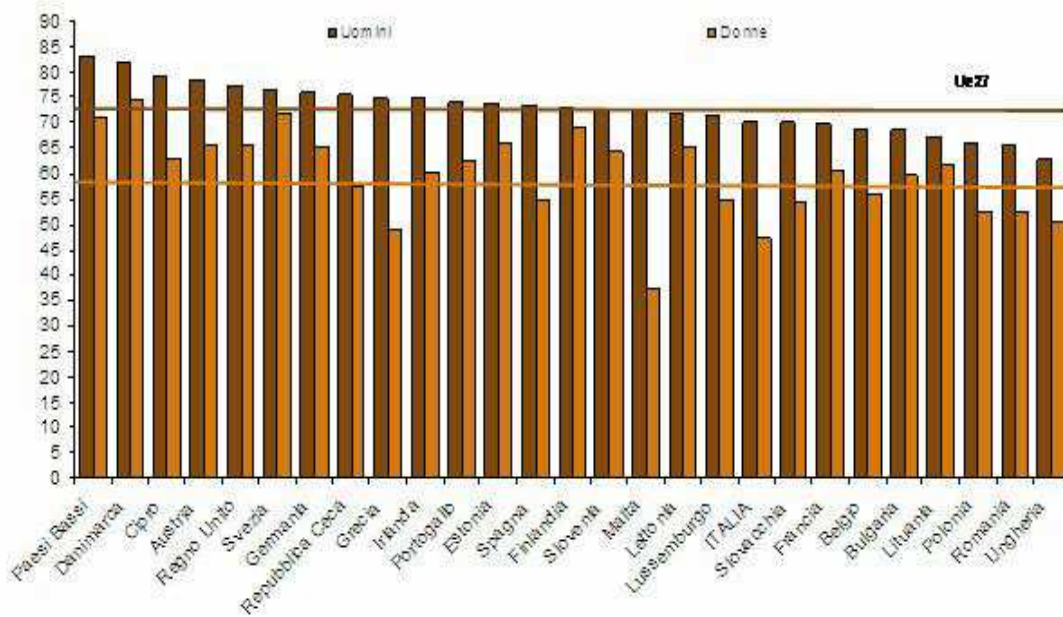
L'Unione Europea conferma tale diagnosi e pone, ormai da diversi anni, l'obiettivo dell'innalzamento dell'occupazione femminile al centro delle proprie politiche per lo sviluppo. Ricordiamo, per fare l'esempio più noto, la Strategia di Lisbona, che nel 2000 puntava a raggiungere per la media europea un tasso di occupazione femminile pari al 70 per cento entro il 2010.

Per l'Italia, la situazione attuale del mercato del lavoro femminile si presenta molto debole. La prima cosa da osservare è che il sistema economico del nostro paese è caratterizzato da un basso grado di coinvolgimento nel mercato del lavoro della popolazione in età attiva, distante da quello dei paesi dell'Unione europea comparabili al nostro per livello di sviluppo economico. Il tasso di occupazione (vale a dire il rapporto percentuale tra gli occupati di 15-64 anni e la popolazione della stessa classe di età) è l'indicatore più usato in sede nazionale e internazionale per valutare l'evoluzione del mercato del lavoro. Ebbene, secondo i dati ISTAT, in Italia il tasso di occupazione è pari, nel 2008, al 58,7 per cento, un valore inferiore di circa 7 punti a quello medio dell'Unione europea.

Il risultato dell'Italia, però, esprime e sintetizza un rilevante divario di genere: mentre per gli uomini il tasso di occupazione si colloca al 70,3 per cento, un valore prossimo a quello medio europeo, per le donne il tasso si colloca appena al 47,2 per cento, distante quasi 12 punti da quello europeo.

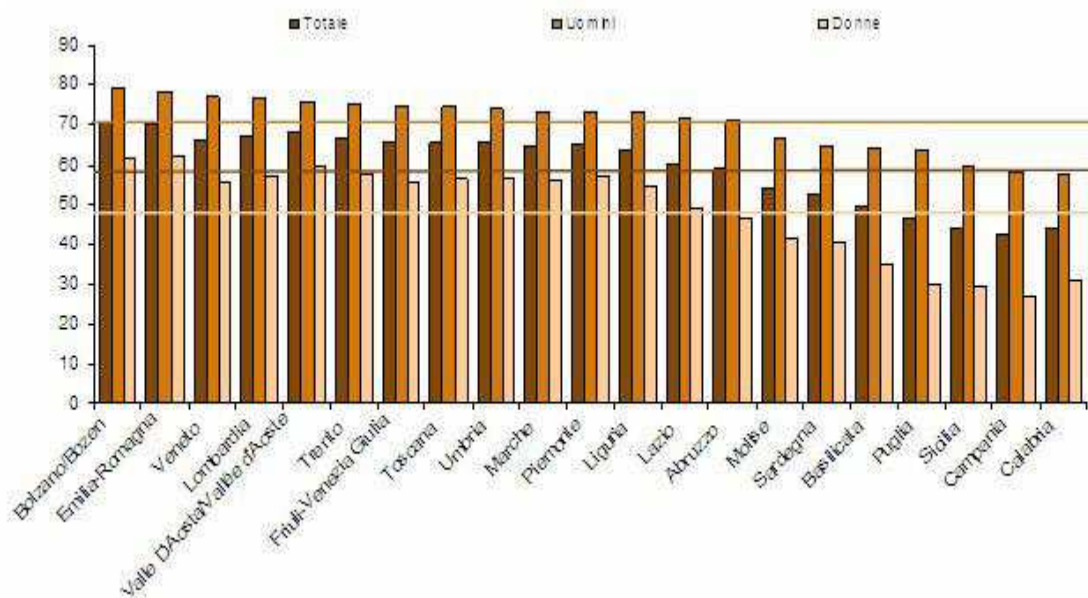
Solo Ungheria e Malta, nella lista dei 27 paesi dell'Unione europea, presentano una situazione del lavoro femminile peggiore di quella italiana.

Grafico 1: Tasso di occupazione della popolazione in età 15-64 anni per sesso nei paesi Ue - Anno 2008 (valori percentuali). Fonte: Istat 2010



La seconda cosa da osservare è che il fenomeno della bassa partecipazione delle donne al mercato del lavoro è un fenomeno concentrato soprattutto nelle regioni del Mezzogiorno, ove i tassi di occupazione femminili sono inferiori di oltre venti punti rispetto al resto del paese.

Grafico 2: Tasso di occupazione della popolazione in età 15-64 anni per sesso e regione - Anno 2008 (valori percentuali). Fonte: Istat, 2010



Un altro importante fattore che incide sullo squilibrio uomini-donne è quello generazionale. Infatti, persino nel Centro Nord, ove la situazione dell'occupazione femminile è migliore, si riscontra una forte differenziazione generazionale: le donne nella fascia d'età 25-44 hanno tassi di occupazione elevati, in media con l'Europa, mentre le donne della fascia d'età più alta mostrano una partecipazione molto più bassa.

Le donne italiane, dopo una certa età, quindi, smettono di lavorare: recenti indagini evidenziano che, nonostante gli ottimi risultati scolastici, esse hanno difficoltà a raggiungere ruoli direttivi e che, a parità di posizione professionale, percepiscono un salario inferiore a quello di un uomo: un "ambiente" che certo non offre un'adeguata motivazione.

Considerando il tasso di natalità e il tasso di occupazione femminile, entrambi bassi, in Italia ci sarebbe bisogno di più politiche per conciliare lavoro e famiglia. La letteratura economica evidenzia un'importante differenza sotto questo aspetto tra i paesi d'Europa (Del Bocca, 2002). Nei paesi del nord Europa il governo ha implementato una politica che incoraggia sia la partecipazione delle donne al lavoro sia la fertilità. Tale politica supporta, infatti, le famiglie in cui entrambi i coniugi lavorano spostando parte degli oneri economici riguardanti il bambino sullo stato: vengono messi a disposizione asili pubblici, permessi per le mamme, possibilità di part-time che permette alle donne di rimanere nel mondo del lavoro anche dopo aver avuto un figlio. Grazie a tale politica si è registrato sia un aumento del tasso di fertilità sia un aumento del tasso di occupazione femminile. Al contrario, nei paesi del sud Europa si registra una diminuzione di entrambi i tassi. In Italia in particolare, le donne risultano più in difficoltà nel conciliare lavoro e famiglia e ciò comporta spesso il dover scegliere tra avere un lavoro o avere dei figli. Il risultato di questa situazione è un basso tasso di natalità (pari secondo l'Istat nel 2010 a 1,41 figli per donna) e un basso tasso di occupazione femminile (pari al 48% contro una media Ocse del 59%). Il tasso di fecondità in Italia è fermo dagli anni '80, periodo in cui però crebbe rispetto al calo degli anni '70. I giovani italiani, anche per avere una posizione lavorativa più stabile, spesso posticipano l'età in cui avere un figlio (l'età media per donna al primo

figlio è passata da 27,7 anni dei primi anni '90 ai 30,8 anni attuali, fonte demo-istat), col rischio di non averne mai.

Un altro fattore da considerare è la flessibilità degli orari di lavoro. La flessibilità degli orari di lavoro, infatti, svolge, nel nostro paese, un ruolo limitato nell'aiutare i genitori a conciliare lavoro e famiglia. Meno della metà delle imprese con più di 10 dipendenti, infatti, offre flessibilità ai propri dipendenti, mentre il 60% dei lavoratori dipendenti non è libero di variare il proprio orario di lavoro. Questa condizione, unita alla scarsa offerta di servizi pre e dopo scuola, pone ai genitori italiani molte difficoltà nello svolgere un lavoro a tempo pieno. Per questa ragione quasi un terzo delle donne (il 31%) svolge un lavoro part-time, opzione scelta però solo dal 7% degli uomini. Nel rapporto Ocse “Society at Glance” è emerso che in Italia la disparità di genere tra il lavoro non retribuito svolto da donne e uomini è tra le più ampie dei Paesi Ocse.

In Italia, inoltre, quasi la metà degli uomini (il 47,7%) e quasi un terzo delle donne (il 32,7%) nella fascia di età compresa tra i 25 e i 34 anni vive ancora con i genitori. E' questo uno dei dati principali del primo numero dell'Osservatorio Isfol (Istituto per lo sviluppo della formazione professionale dei lavoratori), che basandosi sui dati Eurostat 2008, ha messo a confronto i giovani italiani con i coetanei europei.

Nell'Eurozona gli uomini italiani si piazzano al quinto posto nella graduatoria maschile dei “mammoni”, ovvero di quelli che per un motivo o

per l'altro faticano a lasciare la casa della famiglia, mentre le donne al sesto posto in quella femminile.

Quest'ultimo fenomeno, l'allungamento del periodo di convivenza nella casa genitoriale, trova un ampio riscontro negli studi condotti da Aassve et al. del 2005: gli autori osservano che i giovani dei paesi del sud Europa (e non solo le donne) lasciano più tardi la casa d'origine rispetto a quelli del nord Europa. Gli autori si concentrano sulla relazione tra il rischio di povertà e l'abbandono della casa dei genitori. In particolare dimostrano che il rischio di cadere in povertà per un giovane aumenta quanto prima lascia la casa dei genitori.

L'obiettivo di questa tesi è dare un contributo al tema descritto analizzando l'influenza del fattore “nascita di un figlio” sul benessere delle giovani coppie.

Il capitolo 1 tratta in modo approfondito la letteratura esistente a supporto del nesso di causalità che lega il rischio di povertà al lasciare il tetto parentale.

Il capitolo 2 riguarda la costruzione dettagliata della banca dati. Essendo l'obiettivo quello di dimostrare se la nascita di un figlio possa influenzare negativamente il benessere della famiglia, è necessario il ricorso a dati di tipo longitudinale che ci garantiscano di poter confrontare le caratteristiche della stessa famiglia in due anni successivi. Uno studio che fornisce tali dati è quello condotto dalla Banca D'Italia sui Bilanci delle Famiglie

Italiane. Questa indagine è stata condotta a partire dagli anni '60; per l'analisi ho usato solo i dati delle ultime cinque rilevazioni (2000-2008). In particolare, da questi dati, ho selezionato solo quelli relativi alle famiglie con capofamiglia sotto i 41 anni. Nel capitolo 2 vengono anche riportate alcune statistiche descrittive del campione analizzato, con particolare riferimento alle caratteristiche del capofamiglia e del coniuge.

Nel capitolo 3 si analizza l'impatto della nascita di un figlio sul benessere e, conseguentemente, sullo stato di povertà della famiglia di origine. Come misure del benessere si considerano il reddito e il reddito equivalente. Per stimare correttamente l'impatto si dovrebbe confrontare il benessere di una famiglia che ha avuto un figlio con il benessere della stessa famiglia se non lo avesse avuto (situazione controfattuale). Poiché questo confronto non è attuabile, visto che i due scenari si escludono a vicenda, dobbiamo trovare un'approssimazione. Attraverso l'impiego del *propensity score matching* si è in grado di ottenere l'approssimazione cercata e condurre l'analisi.

Una seconda analisi viene condotta, sempre attraverso l'uso del *propensity score matching*, per misurare l'impatto della nascita di un figlio sullo stato di povertà della famiglia. Per condurre questa analisi ho combinato uno stimatore *difference in differences* con la procedura di *matching*. In questo modo ho potuto comparare la differenza tra la percentuale delle famiglie povere tra il tempo t+1 e il tempo t che hanno

avuto un figlio (i trattati) con la differenza tra la percentuale delle famiglie povere tra il tempo $t+1$ e t che non lo hanno avuto (i non trattati) a loro abbinati. Attraverso l'uso di questo stimatore siamo in grado di identificare l'impatto cercato: infatti, il passaggio alle differenze prime ci permette di confrontare la variazione media nei *tassi di povertà* tra i due periodi di tempo t e $t+1$ per coloro che hanno avuto un figlio con la variazione media nei tassi di povertà sullo stesso periodo per quel gruppo di famiglie che non hanno avuto un figlio e che presentano caratteristiche più simili possibili a quelli che l'hanno avuto. Ciò ci permette di neutralizzare l'effetto di eterogeneità eventualmente introdotto da possibili variabili non osservate. La stimatore *difference in differences* è quindi robusto poiché elimina eventuali distorsioni temporali.

Poiché dalle precedenti analisi sono emersi effetti contenuti ma significativi sulla media del reddito equivalente ma non sul tasso di povertà, nel capitolo 4 ho indagato ulteriormente l'effetto della nascita di un figlio sulla distribuzione del reddito equivalente andando a vedere la matrice di transizione tra classi di reddito.

Per le analisi verrà utilizzato il pacchetto *pscore* di STATA, in combinazione con il pacchetto Office di Microsoft.

Dalle analisi svolte emerge che se la nascita di un figlio ha una piccola ma significativa influenza sul reddito equivalente, non influenza invece lo stato di povertà. Per verificare la sensibilità dei dati alla particolare linea di

povertà scelta, ho ripetuto la stessa analisi nel caso in cui la linea di povertà venga posta al 40% e all'80% della mediana del reddito equivalente. In entrambi i casi viene confermato che la nascita di un figlio non fa aumentare la probabilità di diventare poveri.

Dall'ulteriore analisi svolta sulla distribuzione del reddito equivalente, attraverso l'utilizzo di matrici di transizione e la costruzione di un indice che considera solo le frequenze relative fuori dalla diagonale principale della matrice di transizione, emerge che la nascita di un figlio provoca un impoverimento delle famiglie.

CAPITOLO 1: LETTERATURA

Per avviare l'analisi sono partita da articoli che dimostrano come la probabilità di entrare in povertà sia maggiore per i giovani che lasciano la casa dei genitori, rispetto a quella dei giovani che restano a casa.

Il primo articolo "*Does leaving home make you poor? Evidence from 13 European Countries*" di Aassve, Davia, Iacovou, Mazzucco (2005) analizza la *relazione di causalità* tra l'abbandono della casa genitoriale e l'ingresso nello stato di povertà da parte dei giovani, ponendo la domanda se l'ingresso nella povertà è il risultato dell'abbandono della casa paterna o piuttosto dipendente da un insieme di fattori eterogenei o di selezione.

Molti sono, infatti, i fattori strettamente legati al rischio di povertà: trovare (o non trovare) un impiego, accudire figli, formare un'unione e così via. Le condizioni del mercato del lavoro variano considerevolmente tra paesi e queste differenze sono importanti determinanti dei tassi di povertà giovanile (Mercader-Prats, 1999). Il semplice ottenimento di un lavoro non è spesso sufficiente ad evitare l'esclusione sociale e il livello di istruzione gioca un ruolo determinante nella riduzione del rischio di povertà (Pavis e al., 2000). Il nucleo "monofamiliare" (ovvero la famiglia composta da un unico individuo) presenta un importante fattore di rischio di povertà per i giovani (Berthoud-Robson, 2003). Gli autori in questo articolo si focalizzano sul predittore considerato più significativo per l'ingresso nello

stato di povertà da parte dei giovani, l'abbandono della casa parentale (fattore che notoriamente comporta una drastica riduzione del reddito familiare dei giovani).

Utilizzando la metodologia del *propensity score matching* applicata alla banca dati dell'*European Community Household Panel (ECHP)*, gli autori dimostrano che esiste una relazione di causa-effetto tra l'uscita dalla casa parentale e l'ingresso nello stato di povertà o di privazione. L'oggetto dell'analisi è, quindi, la relazione tra l'uscita dalla casa genitoriale e l'ingresso nello stato di povertà e come questa relazione varia tra i paesi europei.

L'uscita dalla casa parentale è un fenomeno precoce nei Paesi Scandinavi e negli UK, mentre avviene molto tempo dopo nei paesi mediterranei e in Irlanda. Per effettuare lo studio della relazione tra abbandono e ingresso nella povertà, gli autori si sono avvalsi dell'*European Community Household Panel*, un'indagine longitudinale che coinvolge la maggior parte dei paesi europei ed ideata dall'Unione Europea. La prima indagine è stata condotta nel 1994 e includeva Germania, Danimarca, Olanda, Francia, Belgio, Lussemburgo, UK, Irlanda, Italia, Grecia, Spagna, Portogallo. Successivamente vennero inclusi nello studio anche: Austria nel 1995, Finlandia nel 1996, Svezia nel 1997; tutti i paesi, eccetto Svezia e Lussemburgo (la prima perchè non è un'indagine panel, il secondo a causa della piccola numerosità campionaria), sono inclusi nell'analisi. L'utilizzo

dei dati raccolti da ECHP comporta il vantaggio di poter confrontare i dati di diversi paesi; inoltre, l'ECHP è stato ideato per fornire informazioni sul reddito e sulla coesione sociale ed è, perciò, ricco di informazioni sul reddito che facilitano il calcolo dello stato di povertà dell'individuo. C'è, però, anche uno svantaggio: le informazioni retrospettive sono limitate. Di particolare rilevanza per lo studio è la mancanza di osservazioni retrospettive sul reddito della famiglia d'origine, mancanza dovuta al fatto che l'intervistato non vive più con i genitori al momento della prima raccolta dati. Per gruppi di popolazione in cui la struttura familiare è relativamente stabile, questo problema è trascurabile. Al contrario, per i giovani, per i quali la struttura del nucleo familiare è tutt'altro che stabile e fortemente dipendente dal reddito disponibile, è stato necessario applicare degli aggiustamenti e ricorrere alla misura del "reddito equivalente", ricavato come la somma dei redditi relativi all'anno t e all'anno $t+1$ per quegli individui presenti nel nucleo familiare al tempo t e usando una scala equivalente basata sul numero e l'età degli individui presenti nell'anno t .

Per verificare gli effetti che il lasciare la casa può avere sul benessere, l'analisi è stata condotta su due misure di privazione. Il primo è un indicatore monetario determinato dalla posizione del reddito individuale nella distribuzione del reddito. Il secondo è un indice non monetario che si basa su un insieme di variabili che riflettono il benessere della famiglia sugli individui che la compongono. Tutte le variabili sono dicotomiche e

prendono valore 1 se sono rilevanti per il benessere della famiglia, 0 altrimenti; l'indice di privazione costruito varia tra 0 (no privazione) e 1 (massima privazione). Essendo, inoltre, molte di queste variabili tra loro correlate ed alcune più importanti di altre, è stato necessario per la costruzione dell'indice procedere a una loro ponderazione.

Per sviluppare l'analisi vengono usati congiuntamente i dati di tutti i paesi, questo porta ad avere un campione molto ampio poi suddiviso in tre sottocampioni a seconda della fascia d'età: 20-24, 25-29, 30-34.

Per i paesi del sud Europa è stato possibile esaminare tutte e tre le fasce d'età, per Germania, Francia, Irlanda e Austria solo le prime due, per i restanti paesi solo la fascia 20-24.

Per condurre l'analisi viene usato il *propensity score matching*; al fine di ottenere stime soddisfacenti, viene imposto di avere un numero sufficiente sia di giovani che lasciano la casa sia che non la lasciano in ogni "cella" definito dalla stima del *propensity score*. Non vengono, perciò, analizzati gli intervalli che contengono meno di 500 osservazioni. La scelta di utilizzare questa tecnica è dovuta al fatto che gli autori, per analizzare il rischio di povertà, vorrebbero paragonare il rischio di povertà per gli individui che lasciano la casa parentale, con il rischio di povertà per gli stessi individui se non l'avessero lasciata (situazione controfattuale), la situazione controfattuale, però, non è osservabile. Tale problema è risolto

attraverso l'uso del *propensity score matching*, che fornisce un'approssimazione per la situazione controfattuale.

Il “trattamento” è rappresentato dal gruppo di coloro che hanno abbandonato la casa parentale al tempo $t+1$; il “controllo” dal gruppo di coloro che ancora ci vivono e che sono il più simili possibile ai primi in termini di caratteristiche osservabili, misurate prima dell'evento. Ciascun individuo del primo gruppo viene confrontato con l'individuo del gruppo di controllo che presenta un *propensity score* più vicino al suo (probabilità condizionata a un vettore di variabili esplicative misurate al tempo t , prima del verificarsi dell'evento). Piuttosto che usare un singolo *nearest-neighbour*, gli autori scelgono di lavorare con la media di 3 *nearest-neighbour* al fine di ridurre la varianza delle stime. Inoltre, per evitare di ottenere stime poco significative, essi impongono nei vari confronti tra coppie il vincolo sui punteggi entro lo 0.01. Le variabili esplicative incluse nell'analisi di *matching*, misurate al tempo t , sono: l'anno dell'intervista, età, sesso, numero di fratelli e sorelle, titolo di studio dell'intervistato e dei genitori, logaritmo del reddito dei genitori e dell'intervistato, situazione lavorativa del giovane e dei genitori, se vive con entrambi i genitori, se è sposato, se ha figli, quanto grande è la casa dei genitori.

Il risultato dell'analisi mostra come la probabilità di diventare poveri è più alta per i giovani che lasciano la casa genitoriale rispetto a quella dei giovani che rimangono a vivere con i loro genitori. In questo quadro

generale esistono, però, delle interessanti differenze tra i vari paesi. Tra coloro che vivono con i genitori, i giovani scandinavi hanno minore probabilità di diventare poveri rispetto ai coetanei dei paesi mediterranei; tra coloro che decidono di lasciare la casa genitoriale, i giovani scandinavi hanno probabilità più alta di divenire poveri rispetto ai giovani dei paesi del sud Europa.

Gli autori stimano, quindi, l'effetto dell'abbandono della casa genitoriale sull'indicatore di privazione di tipo monetario, che subisce variazioni considerevoli tra i vari paesi, con un punteggio medio di 0.16 e deviazione standard dello 0.27. Tra paesi si evidenziano piccoli cambiamenti negativi per Italia e Portogallo, mentre in Francia, Germania e Paesi Scandinavi gli effetti negativi dell'abbandono della casa sul reddito sono molto più gravi.

Gli effetti sulle variabili non monetarie sono, invece, meno chiari: sebbene l'indicatore di privazione non monetario ha media e varianza simili a quello monetario, gli effetti del lasciare la casa parentale sono più contenuti con punteggi che variano dal -0.08 del Portogallo allo 0.22 della Danimarca.

Le conclusioni a cui arrivano gli autori sono inequivocabili: in quei paesi in cui si lascia presto la casa parentale il rischio di povertà associato all'evento dell'abbandono è più alto, mentre in quei paesi in cui i giovani lasciano la casa genitoriale più tardi, il rischio è più basso. Inoltre, nei paesi nordici (paesi Scandinavi, Finlandia, Olanda), dove i giovani lasciano la

casa dei genitori prima, il rischio di povertà è maggiore rispetto ai paesi come Italia, Spagna, Portogallo. Quindi l'età, il cui effetto non è però misurato direttamente in questo approccio empirico, è un *driver* importante del fenomeno oggetto di studio.

Perché allora i giovani scandinavi lasciano la casa parentale così presto se la probabilità di divenire poveri è così alta? Una possibile spiegazione è data dal fatto che mentre nei paesi del sud Europa i giovani spesso lasciano casa per costituire un nuovo nucleo familiare come mariti o mogli (potendo quindi contare su entrambi gli stipendi), nei paesi scandinavi essi tendono a vivere per un periodo prolungato da soli. L'analisi però evidenzia come anche nei casi in cui i giovani scandinavi lasciano la casa dei genitori per vivere con un'altra persona, essi comunque presentano un più alto rischio di cadere in povertà rispetto ai giovani degli altri paesi.

Il confronto dei risultati tra il gruppo dei “trattati” e la loro controparte ci rivela inoltre che i giovani scandinavi che non lasciano casa hanno minore probabilità di cadere in povertà qualora decidessero di lasciarla dei giovani scandinavi che in effetti la lasciano. Inoltre, il tasso di ingresso nella povertà è più alto per il gruppo 25-29 anni, evidenziando come la probabilità di cadere in povertà è più alta per quei giovani che lasciano prima la casa parentale.

Il secondo articolo preso in esame di Lavinia Parisi “*Leaving home and the chances of being poor: the case of young people in Southern European*

Countries” (2008); analizza la relazione tra il rischio di povertà di un giovane che ha lasciato la casa dei genitori e il reddito dei genitori stessi per i paesi del sud Europa (Italia, Spagna, Grecia, Portogallo) sempre partendo dalla banca dati dell'*European Community Household Panel* (ECHP), ma prendendo in esame solo quei giovani che lasciano la casa d'origine per formare un'unione stabile (matrimonio o convivenza), non si considerano, quindi, quelli che vanno a vivere da soli. Convenzionalmente si considerano giovani coloro che appartengono alla fascia d'età tra i 15 e i 25 anni. In questo articolo la fascia d'età considerata è leggermente diversa: i giovani hanno un'età compresa tra i 18 e i 32 anni, in ragione dell'abbandono della casa parentale al fine di formare una nuova unione.

Questo articolo basa la sua analisi sulla relazione tra la povertà giovanile dopo aver lasciato la casa parentale e il reddito dei genitori nell'anno precedente all'abbandono. Fattori come la disoccupazione o il prezzo alto delle case non vengono presi in considerazione.

Analizzando solo i giovani che lasciano la casa per andare a vivere con il partner, viene dimostrato che più la famiglia è benestante, minore è la probabilità che il ragazzo lasci la casa parentale per formare una sua famiglia. Ciò suggerisce un potenziale *sample selection bias*, perchè ci possono essere vari fattori, osservabili o meno, che possono influenzare al contempo sia la decisione di lasciare la casa da parte del giovane sia il suo status di povertà. Per risolvere questo problema viene usato un *sample*

selection model. Questo modello non tiene conto di come varia la probabilità di divenire povero dopo aver lasciato la casa dei genitori con il prolungarsi del tempo che il giovane passa nella casa parentale. Questo aspetto viene analizzato in un secondo modello di *competing risk duration model*. Questi due modelli sono complementari nell'analisi del problema.

Il primo è un modello Markoviano di primo ordine e tiene conto delle osservazioni in due anni consecutivi t e $t+1$ per tutti gli individui, dove t è l'anno in cui il ragazzo vive ancora con i genitori e $t+1$ l'anno in cui ha lasciato la casa per vivere con il partner. Un potenziale *selection bias* può essere dato dall'associazione tra il reddito della famiglia e l'inclusione nel campione; per risolvere questo problema viene usato il modello di selezione di Heckman:

$$Y_{it+1} = (y_{it}\beta + N_{it}\gamma_0 + S_{it}\gamma_1 + E_{iy+1}\gamma_2 + X_{it+1}^1\alpha + u_i > 0)$$

Questa equazione rappresenta la probabilità di diventare povero dopo aver lasciato la casa dei genitori. Poichè il principale interesse è la relazione tra lo stato di povertà del giovane dopo l'abbandono della famiglia e il reddito della famiglia d'origine, le variabili esplicative comprendono oltre alla situazione economica della famiglia d'origine, anche le caratteristiche del quartiere di residenza, le relazioni sociali al tempo t , gli anni di studio e altre caratteristiche sociali. La variabile dipendente è osservata solo per un sottoinsieme del campione: i giovani che

lasciano la casa dei genitori per vivere con un partner. In questa equazione Y_{it+1} è osservabile se e solo se una seconda variabile latente non osservabile supera una particolare soglia:

$$L_{it}^* = X_{it}^1 \xi + q_{it} \varphi + e_i \quad \text{dove } L_{it}=1 \text{ se } (L_{it}^* > 0) \quad L_{it}=0 \text{ altrimenti} \quad \text{corr}(e, u) = \rho$$

L'equazione di selezione è la probabilità di lasciare la casa dei genitori per vivere con un partner. La probabilità di lasciare la casa dipende da variabili esplicative che riflettono caratteristiche demografiche, struttura familiare e caratteristiche del quartiere. Questa equazione include anche un indice che tiene conto dell'“affollamento” della casa (*crowding index*), questo perchè giovani provenienti da famiglie numerose lasciano la casa prima, inoltre, la probabilità di vivere in una casa affollata (avere un ridotto numero di stanze e/o un alto numero di adulti), può essere negativamente associato al reddito dei genitori. Tuttavia si assume che la dimensione al tempo t (relativa al numero di adulti che condividono la casa) non sia un fattore che riguardi direttamente il reddito al tempo t+1.

I giovani possono essere intervistati al massimo sette volte (dal 1994 al 2000) e ogni individuo può contribuire più di una volta all'analisi cioè, ad esempio, si possono inserire nel *dataset* da analizzare sia i dati dell'intervistato nel biennio 1994-1995 sia quelli del biennio 1995 -1996.

Dall'analisi di indipendenza delle equazioni Y_{it} e L_{it} , condotta tramite un modello probit di Heckman per la stima di ρ , il test di Wald utilizzato rifiuta l'ipotesi nulla. La correlazione tra i termini di errore è positiva e

significativamente diversa da zero in entrambi le specificazioni. Quindi, maggiore è la probabilità di un giovane di lasciare casa , maggiore è anche la probabilità di diventare povero. Rimanere a vivere nella casa parentale è una protezione contro la povertà. Le stime inoltre suggeriscono che più la famiglia d'origine è povera al tempo t , maggiore è la probabilità del giovane di essere povero al tempo $t+1$. Questa relazione è sempre valida anche testando gli effetti del livello di istruzione, del contesto sociale e dell'introduzione del reddito del giovane espresso come percentuale del reddito familiare in t .

Per quanto riguarda le relazioni sociali, tre sono le misure che si considerano: se l'individuo esce con gli amici, se l'individuo fa parte o meno di un'associazione/club, se l'individuo frequenta i vicini. Le prime due misure sono statisticamente diverse da zero: avere, infatti, buone relazioni sociali fa diminuire il rischio di povertà; mentre far parte di un club la fa aumentare, questo risultato è spiegato da una teoria sociologica secondo la quale più sono precari i legami, più aumentano le opportunità di successo nel lavoro: infatti, le relazioni che si instaurano con gli amici impegnano meno un individuo, dandogli la possibilità di applicarsi nel lavoro garantendo, quindi, maggiore successo. Infine le caratteristiche del quartiere di residenza non influenzano la probabilità di povertà al tempo $t+1$.

Al fine di tenere conto anche dell'età in cui si lascia la casa d'origine viene stimato un secondo modello *competing risk duration model* dove con *risk* si vuole indicare che un ragazzo è potenzialmente a rischio di lasciare la casa di origine dall'età di 18 anni in poi e con *duration* il numero di anni trascorsi tra i 18 anni fino al momento in cui si decide di andare via. In questo modello ogni individuo può essere intervistato più di una volta come nel modello precedente, ma mentre nel *sample selection approach model* ciascun individuo poteva essere intervistato al massimo 6 volte, nel *competing risk duration model* può esserlo fino a 13 volte. Le variabili esplicative stimate sono le stesse del modello precedente tranne per una variabile aggiunta: il logaritmo del tempo. La stima di questa variabile suggerisce che più tardi il ragazzo lascia la casa, meno probabilità ha di diventare povero, questo può essere spiegato dal fatto che spesso i giovani non lasciano la casa d'origine per investire sullo studio, ciò dà loro la possibilità di trovare un lavoro maggiormente qualificato e remunerato. Il *competing risk duration model* mostra, inoltre, che le donne lasciano la famiglia ad un'età inferiore rispetto agli uomini.

Considerando i paesi SEC nel loro complesso, l'autrice conclude che i giovani che lasciano casa hanno maggiore probabilità l'anno successivo di diventare poveri che di uscire dalla povertà: in Portogallo un giovane che lascia la casa parentale ha una probabilità di divenire povero l'anno dopo

doppia rispetto a quella di uscirne. Italia e Grecia mostrano, invece, una tendenza opposta.

CAPITOLO 2: I DATI

2.1 L'INDAGINE SUI BILANCI DELLE FAMIGLIE ITALIANE DELLA BANCA D'ITALIA

L'indagine sui bilanci delle famiglie italiane, svolta dalla Banca d'Italia, nasce negli anni '60 con l'obiettivo di raccogliere informazioni sui redditi e i risparmi delle famiglie italiane.

La Banca d'Italia svolge tale indagine, somministrando un questionario ad un campione di famiglie selezionato con un metodo di campionamento a due stadi. Le unità di primo stadio sono i comuni; le unità di secondo stadio sono le famiglie. Prima di procedere all'estrazione delle unità di primo stadio, queste vengono suddivise in base alla regione e alla classe di ampiezza demografica (stratificazione delle unità di primo stadio). All'interno di ogni strato i comuni nei quali effettuare le interviste sono selezionati includendo tutti quelli con popolazione superiore ai 40.000 abitanti e quelli in cui nella precedente rilevazione risiedevano le famiglie *panel* (comuni autorappresentativi) ed estraendo i comuni rimanenti con una modalità di selezione che assegna ai comuni di maggiore dimensione una probabilità più elevata di essere inclusi nel campione. In una seconda fase, per ciascun comune selezionato, le famiglie da intervistare vengono estratte casualmente dalle liste anagrafiche.

L'indagine è stata effettuata, fino al 1987, sulla base di rilevazioni indipendenti nel tempo. Dall'indagine del 1989, per favorire l'analisi dell'evoluzione dei fenomeni oggetto di indagine, è stato introdotto uno schema che prevede la presenza nel campione di una quota di unità già intervistate in occasione di precedenti indagini (famiglie *panel*).

Il questionario utilizzato nella rilevazione è predisposto seguendo una struttura modulare. Si compone di una parte di base, nella quale sono rilevati i fenomeni ai quali tutte le famiglie sono interessate, e di diversi allegati, in cui sono raccolte informazioni che riguardano soltanto specifici sottoinsiemi di famiglie.

Per ridurre l'onere di risposta per gli intervistati, si è ritenuto di sottoporre alcune sezioni del questionario solo a un sottoinsieme casuale del campione sulla base dell'anno di nascita (pari o dispari) del capofamiglia. In particolare a ogni famiglia è stato sottoposto solo uno dei due blocchi di domande, uno riguardante il benessere percepito e le condizioni della famiglia di origine, l'altro gli strumenti di pagamento e l'informazione finanziaria.

A partire dall'anno 1995 la rilevazione dei dati è stata effettuata con l'aiuto del computer (CAPI, *Computer-Assisted Personal Interviewing*), in particolare, nell'ultima rilevazione (2008), la percentuale delle interviste effettuate con il metodo CAPI è risultata del 79,5 per cento. Con questa metodologia i dati vengono rilevati presso le famiglie mediante un

questionario elettronico, che, oltre a memorizzare le informazioni, aziona una serie di controlli consentendo di risolvere, in presenza della famiglia, le eventuali incoerenze. Le restanti interviste sono state realizzate con un questionario cartaceo (PAPI, *Paper-And-pencil Personal Interviewing*) e successivamente trasferite su supporto elettronico dalla società di rilevazione, utilizzando il programma CAPI come procedura di immissione dei dati.

Il totale delle famiglie intervistate nelle indagini dal 1987 al 2008 è riportato nella tabella sottostante.

Tab. 1 Famiglie intervistate nel corso del tempo, indagini 1987-2008

Anno della 1 ^a partecipazione all'indagine	Anno di rilevazione										
	1987	1989	1991	1993	1995	1998	2000	2002	2004	2006	2008
1987	8027	1206	350	173	126	85	61	44	33	30	28
1989		7068	1837	877	701	459	343	263	197	159	146
1991			6001	2420	1752	1169	832	613	464	393	347
1993				4619	1066	583	399	270	199	157	141
1995					4490	373	245	177	117	101	84
1998						4478	1993	1224	845	636	538
2000							4128	1014	667	475	398
2002								4406	1082	672	525
2004									4408	1334	995
2006										3811	1143
2008											3632
numerosità campionaria	8027	8274	8188	8089	8135	7147	8001	8011	8012	7768	7977
quota delle famiglie panel		14,6	26,7	42,9	44,8	37,3	48,4	45	45	50,9	54,5

Le variabili rilevate dall'indagine riguardano caratteristiche socio-demografiche (tra le quali sesso, età, titolo di studio) e una serie di informazioni molto dettagliate relative alle disponibilità economiche sia a

livello individuale che familiare (reddito da lavoro, da capitale, ricchezze immobiliari, ecc.).

2.2 IL CAMPIONE SELEZIONATO

L'obiettivo è avere un campione di individui osservati due istanti temporali t e $t+1$. Il *dataset* deve garantire la possibilità di confrontare le caratteristiche di una stessa famiglia in due anni successivi, visto che la tesi si focalizza sugli effetti che la nascita di un figlio può provocare sul benessere della famiglia. Partendo dai dati della Banca D'Italia derivanti da queste indagini, l'obiettivo è quello di formare un unico *dataset* dove le famiglie selezionate sono quelle presenti almeno due volte nel campione e per le quali possiamo osservare variabili al tempo t al tempo $t+1$. Nel selezionare, per ogni anno analizzato, le famiglie con due occasioni d'indagine, ho dovuto fare attenzione che quelle selezionate per un anno, non fossero le stesse selezionate per l'anno successivo:¹se analizzo, infatti, i dati di una famiglia nel biennio 2000-2002 non la considererò per il biennio 2002-2004 affinché ogni famiglia contribuisca all'analisi solo una volta, a differenza dell'analisi svolta da Parisi (2008), dove un intervistato può contribuire con i propri dati più di una volta.

La selezione delle famiglie con le caratteristiche sopra citate, dà luogo al *dataset* finale che risulta composto da 7164 famiglie.

¹ appendice: schema di selezione delle famiglie

Dall'insieme di queste famiglie ho effettuato un'ulteriore selezione: ho selezionato solo quelle che al tempo iniziale sono composte da una coppia (con o senza figli) con capofamiglia di età fino ai 40 anni e che dopo due anni sono ancora in unione.

La scelta dei 40 anni di età è motivata dal fatto che il focus del presente lavoro è rappresentato dalle famiglie giovani. In letteratura con "giovane" si intende "colui che non è più un bambino ma che non ha ancora completato del tutto il processo di transizione all'età adulta" (Aassve, 2005). Convenzionalmente si considerano giovani coloro che appartengono alla fascia d'età tra i 15 e i 25 anni; poiché l'età media in cui si ha il primo figlio è aumentata negli anni, si è reso necessario ampliare la fascia d'età considerando giovani le famiglie con capofamiglia fino ai 40 anni.

Il *dataset* finale è composto da 759 osservazioni e riporta, oltre alle variabili rilevate direttamente dall'indagine della Banca D'Italia riguardanti caratteristiche socio-demografiche ed economiche.

2.3 ANALISI DESCRITTIVE

Il campione è composto da 759 famiglie; la tabella sotto riportata ne riassume alcune caratteristiche.

Tab. 2 Composizione del campione rispetto all'anno d'inizio, area geografica, numero di componenti, numero percettori di reddito, grandezza del comune di residenza, ricchezza, casa di proprietà

Variabile	modalità	Percentuale
Area geografica	nord-ovest	20
	nord-est	23
	centro	17
	sud	26
	isole	14
Anno d'inizio	1989	2
	1991	2
	1993	3
	1995	3
	1998	17
	2000	18
	2002	15
	2004	20
	2006	20
Numero componenti	2	21
	3	30
	4	37
	>=5	12
	comune	< 40000 abitanti
>= 40000 abitanti		49
n. percettori di reddito	1	39
	2	59
	3 o più	2
ricchezza	< 112500 euro	50
	>112500 euro	50
Casa	Affitto	43
	di proprietà	57

Elaborazione su dati della Banca D'Italia

Rispetto all'anno d'inizio le percentuali maggiori sono quelle delle famiglie intervistate per la prima volta nel 2006 (20%); rispetto all'area geografica la percentuale maggiore è quella delle famiglie del sud. Per quanto riguarda il numero di componenti la percentuale maggiore è quella delle famiglie con 4 componenti. Il 59% delle famiglie presenta due percettori di reddito e per il 57% è proprietaria della casa familiare. Il

campione risulta equamente distribuito per classi per quanto riguarda il comune di residenza (più o meno di 40.000 abitanti) e la ricchezza disponibile (più o meno di 112.500 €).

2.3.1 CARATTERISTICHE DEL CAPOFAMIGLIA

Il 58% del campione analizzato ha il capofamiglia tra i 31 e i 40 anni. La percentuale maggiore (47%) delle famiglie ha un capofamiglia con un titolo di studio non superiore alla licenza media, per quanto riguarda la qualifica lavorativa le percentuali maggiori sono operaio e impiegato (42% e 35% rispettivamente). L'8% del campione presenta il capofamiglia non occupato, di questi il 39% è disoccupato e l'12% è in cerca di prima occupazione.

Tab. 3 Caratteristiche del capofamiglia

Variabile	modalità	percentuale
classe d'età	fino 20 anni	14
	21-30	28
	31-40	58
titolo di studio	al più licenza media	47
	diploma	43
	laurea o diploma post-laurea	10
Condizione professionale	occupato	92
	disoccupato	4
	non forza lavoro	4
qualifica lavorativa	operaio	43
	impiegato	35
	libero professionista	22
non occupato	in cerca di prima occupazione	12
	disoccupato	39
	Altra condizione	49

Elaborazione su dati della Banca D'Italia

2.3.2 CARATTERISTICHE DEL CONIUGE

Per quanto riguarda le caratteristiche del coniuge, la percentuale maggiore è diplomato (44%), seguiti da un 43% di aventi titolo di studio non superiore alla licenza media, il 60% è occupato, più della metà nel settore impiegatizio (53%). La percentuale di coniugi non occupati è del 40%, di cui l'83% svolge la mansione di casalinga. La percentuale dei coniugi non occupati, se confrontata con quella del capofamiglia, risulta elevata (40% rispetto all'8%).

Tab. 4 Caratteristiche del coniuge

Variabile	modalità	percentuale
titolo di studio	al più licenza media	43
	diploma	44
	laurea o diploma post-laurea	13
stato di occupazione	occupato	60
	disoccupato	5
	non forza lavoro	34
qualifica lavorativa	operaio	32
	impiegato	53
	libero professionista	15
non occupato	in cerca di 1 occupazione	6
	disoccupato	8
	casalinga	83
	Altra condizione	3

Elaborazione su dati della Banca D'Italia

2.3.3 CARATTERISTICHE DELLA FAMIGLIA

Poichè l'analisi si prefigge di evidenziare l'impatto che ha la nascita di un figlio sul benessere della famiglia, ho focalizzato la mia attenzione sul numero di figli al tempo t. Nel campione analizzato, il numero medio di

figli è di 1,38; analizzando nel dettaglio i nuclei famigliari, si vede che la percentuale maggiore di famiglie è quella con due figli (37%).

Tab. 5 Numero di figli al tempo t

numero di figli	percentuali
0	21
1	31
2	37
3 o più	11

Elaborazione su dati della Banca D'Italia

Ho, inoltre, analizzato le differenze nel numero medio di figli rispetto all'area geografica.

Tab. 6 Numero medio di figli al tempo t per area geografica

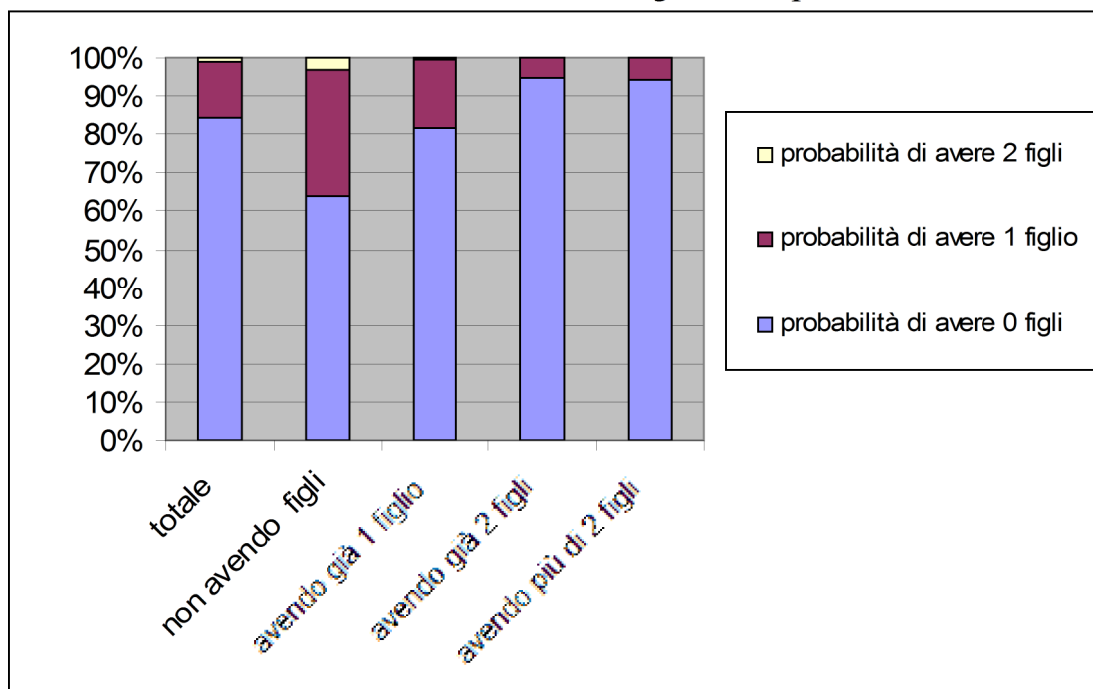
Area geografica	n medio figli
nord-ovest	1,21
nord-est	1,24
centro	1,39
sud	1,55
isole	1,55

Elaborazione su dati della Banca D'Italia

Le famiglie che risultano avere un numero medio di figli più alto sono quelle del sud e delle isole.

Ho, infine, analizzato la probabilità di avere uno o due figli in due anni (tempo trascorso tra le due occasioni d'indagine) per l'intero campione. Ho, inoltre, calcolato la stessa probabilità condizionata al numero di figli della famiglia: cioè la probabilità di avere un figlio e la probabilità di averne due se la famiglia non ha figli, se ne ha uno, se ne ha due o più.

Grafico 3 Probabilità di avere uno o due figli al tempo t+1



Il grafico evidenzia che in tutte le categorie analizzate la probabilità di avere 2 figli è molto bassa; si evidenzia, inoltre, che più il numero di figli già posseduti dalla coppia è alto, maggiore è la probabilità di non avere altri figli.

2.3.4 CONFRONTO TRA IL CAMPIONE ANALIZZATO E I CAMPIONI SEZIONALI DEGLI ANNI PRESI IN ESAME

Per completare l'analisi del campione, ho confrontato alcune sue caratteristiche con quelle dei campioni sezionali degli anni presi in esame (2000-2006). In particolare, ho analizzato la distribuzione dei campioni di tali anni rispetto all'area geografica, il numero di componenti, il numero di figli, ampiezza del comune di residenza, numero dei percettori di reddito, titolo di studio, condizione professionale, qualifica lavorativa, stato di non

occupazione². Dalla tabella sotto riportata si evidenzia che, nei quattro anni analizzati, non ci sono differenze rilevanti nelle percentuali delle famiglie rispetto ad area geografica, grandezza del comune di residenza, numero di redditi percepiti, titolo di studio, condizione professionale, qualifica lavorativa, stato di non occupazione.

Confrontando i risultati della tabella sotto riportata con quelli del campione, però, si evidenziano alcune differenze: per quanto riguarda il numero di componenti delle famiglie negli anni presi in esame, la percentuale maggiore è quella delle famiglie con due componenti. Al contrario nel campione la percentuale maggiore è quella delle famiglie con tre componenti. Anche nel numero di figli si rilevano delle differenze: negli anni presi in esame la percentuale maggiore è quella delle famiglie senza figli, mentre nel campione è quella delle famiglie con due figli. Analizzando la condizione occupazionale, si può notare che la percentuale di capofamiglia non forza lavoro nei campioni sezionali sia molto più alta di quella del campione. Confrontando, infine, il numero dei percettori di reddito si può notare che la percentuale di famiglie del campione con due percettori è molto più alta rispetto a quella del campione sezionale.

² Le caratteristiche individuali si riferiscono al capofamiglia

Tab. 7 Caratteristiche dei campioni degli anni 2000-2006

Variabile	modalità	Anno				campione costruito
		2000	2002	2004	2006	
Area geografica	nord-ovest	24	26	25	26	20
	nord-est	20	20	21	22	23
	centro	20	21	22	20	17
	sud	24	21	21	21	26
	isole	12	12	11	11	14
Numero componenti	1	18	22	24	25	/
	2	28	29	30	30	21
	3	22	22	21	21	30
	4	23	20	19	17	37
	5 o più	9	7	6	6	11
n. di figli	0	45	49	52	53	21
	1	25	24	24	24	31
	2	24	21	19	18	37
	3 o più	7	6	5	5	11
comune	< 40000 abitanti	61	60	60	61	51
	>= 40000 abitanti	40	40	40	39	49
n. percettori di reddito	1	46	48	50	48	39
	2	41	40	39	41	59
	3 o più	13	12	11	11	2
titolo di studio	al più licenza media	68	67	66	63	47
	diploma	23	24	26	27	43
	laurea o diploma post-laurea	9	8	8	10	10
Condizione professionale	occupato	59	59	60	60	92
	disoccupato	4	3	3	3	4
	non forza lavoro	38	38	37	38	4
qualifica lavorativa	operaio	21	21	22	22	43
	impiegato	23	24	25	25	35
	libero professionista	14	14	13	12	22
non occupato	in cerca di prima occupazione	1	1	1	1	11
	disoccupato	5	4	4	4	39
	Altra condizione	94	95	95	94	49

Elaborazione sui dati della Banca d'Italia

CAPITOLO 3: STIMA DEL MODELLO

3.1 APPROCCIO EMPIRICO : *PROPENSITY SCORE MATCHING*

Nella presente tesi si è interessati a valutare l'impatto che ha la nascita di un figlio sul benessere della famiglia. Come misura di benessere utilizzeremo il reddito equivalente. L'obiettivo è quello di confrontare famiglie con le stesse caratteristiche che differiscano solo per il numero di figli, in modo che, se si evidenziano differenze nel reddito equivalente, queste sono dovute esclusivamente alla presenza del figlio. Per confrontare famiglie a parità di caratteristiche usiamo il *propensity score matching*.

Per stimare l'impatto, dovremmo calcolare la differenza tra la media della variabile risultato delle famiglie dopo il trattamento che, nel mio caso, è l'aver avuto un figlio, e quella potenziale che le stesse avrebbero realizzato nel caso ipotetico in cui non fossero stati trattati (situazione controfattuale). La media della variabile risultato delle famiglie se non fossero state trattate non è osservabile; possiamo osservare solo la media delle famiglie che non hanno ricevuto il trattamento. Tuttavia, se usassimo i dati delle famiglie che non hanno ricevuto il trattamento per approssimare la distribuzione della variabile risultato controfattuale, otterremmo una stima dell'impatto distorta perchè, generalmente, la variabile risultato per i due gruppi differisce in maniera sistematica anche in assenza di trattamento. Questo è, infatti, dovuto alla presenza di un probabile

problema di autoselezione: la partecipazione al trattamento può non essere casuale ma dovuta alle particolari condizioni della famiglia. Per poter stimare l'impatto si deve trovare, allora, un'approssimazione per il controfattuale.

Per ovviare a questo problema, una possibile soluzione è quella di utilizzare un procedimento di *matching*, che consenta di abbinare a ogni individuo trattato l'individuo non trattato maggiormente somigliante in termini di determinate caratteristiche pre-trattamento. A tal fine risulta necessario il rispetto di due assunzioni. La prima è che il vettore di variabili al quale si vuole condizionare il *matching* (X) risulti indipendente dall'assegnazione al trattamento. La seconda è che anche la distribuzione della variabile risultato, condizionata all'insieme X , risulti indipendente dal trattamento. In particolare, quest'ultima assunzione è conosciuta in letteratura come CIA, ossia *conditional independence assumption* e risulta di fondamentale importanza in quanto solo se la CIA è rispettata la selezione degli individui può essere espressa in funzione delle sole caratteristiche pre-trattamento.

Tuttavia, uno dei principali problemi legati al *matching* consiste nel fatto che il processo di abbinamento risulta spesso difficoltoso nel caso di un'elevata dimensione del vettore X perchè il numero dei gruppi omogenei creati dalla stima cresce esponenzialmente: due individui possono essere identici in una dimensione e differire lungo un'altra, un abbinamento esatto

che consente di trovare un non trattato identico ad un trattato è praticamente impossibile. Per ovviare a tale inconveniente Rosenbaum e Rubin (1983) hanno suggerito l'utilizzo del *balancing scores* ($b(X)$), ossia di una funzione dell'insieme X . Infatti, hanno dimostrato che se la CIA è rispettata, non solo il vettore X è indipendente dal processo di selezione ma anche qualsiasi funzione di X possiede questa caratteristica. Tra i possibili *balancing scores*, Rosenbaum e Rubin (1983) suggeriscono l'uso del *propensity score*, un indice definibile come la probabilità di essere trattato dato il vettore X . Rosenbaum e Rubin (1983) hanno dimostrato che se la CIA è rispettata, allora la probabilità di appartenere al gruppo dei trattati o dei non trattati, condizionata al vettore X , si configura anch'essa come una variabile casuale e può essere definita come $p(X) \equiv Pr(T = 1 | x_i = x)$, dove T è una variabile dicotomica che assume valore 1 se l'individuo è trattato e 0 l'individuo non è trattato. Siccome la distribuzione del *propensity score* dipende dalla distribuzione congiunta del vettore X , il procedimento di *matching* può essere condizionato esclusivamente al *propensity score*, riducendo un problema potenzialmente multidimensionale a uno unidimensionale. Un ulteriore vantaggio legato all'utilizzo di questa tecnica risiede nel fatto che viene generata una variabile continua, il che evita, di fatto, il pericolo di trovare più individui con lo stesso valore di *propensity score*.

Per effettuare i confronti tra unità trattate e non trattate è necessario che, all'interno di ciascun blocco creato dalla procedura di stima, l'assegnazione al trattamento sia casuale rispetto alle variabili osservabili. In altre parole, all'interno di ogni singola cella, casi e controlli non devono differire statisticamente nel valore delle variabili esplicative a cui ci si condiziona. Questa proprietà è detta proprietà di bilanciamento ed è il principio cardine su cui si basa la somiglianza statistica delle unità inserite nell'esperimento. Ne segue che solo i trattati e non trattati che rispettano questa proprietà possono essere utilizzate per il calcolo del *propensity score*.

Riassumendo, le due proprietà di cui gode il *propensity score*, che ci assicurano di poter confrontare i trattati e i non trattati, sono:

1. bilanciamento delle variabili pre-trattamento dato il *propensity score*

(*balancing property*) $X \perp I | e(x)$

Se vale questa proprietà, le osservazioni con lo stesso *propensity score* hanno la stessa distribuzione rispetto alle caratteristiche osservabili X indipendentemente se sono esposti o meno al trattamento. Ciò significa che per un dato *propensity score*, l'esposizione al trattamento è casuale e quindi trattati e controlli sono, in media, identici.

2. se vale la condizione di indipendenza condizionata

$[(Y^T, Y^{NT}) | I \perp X]$ ed esiste un supporto comune $0 < P(I=1|X) < 1$ (condizioni definite da Rosenbaum e Rubin (1983) *strong ignorability* di I per le variabili risultato (Y^T, Y^{NT}) , condizionatamente alle caratteristiche X)

allora vale anche $[(Y^T, Y^{NT})|I \perp e(x)]$ cioè se vale la condizione di *strong ignorability*, condizionandosi ai valori del *propensity score*, la distribuzione della variabile risultato controfattuale degli esposti è la stessa della variabile risultato fattuale dei non esposti. Quindi:

$$E(Y^T | e(x), I=1) = E(Y^T | e(x))$$

$$E(Y^{NT} | e(x), I=0) = E(Y^{NT} | e(x))$$

Questo implica che, se si stima l'impatto mediante *matching* è sufficiente abbinare all'*i*-esimo esposto con caratteristiche X^E un non esposto con caratteristiche X^{NE} tali che $e(X_i^E) = e(X_i^{NE})$.

Perciò gli individui nel gruppo dei trattati e quelli nel gruppo di controllo sono “abbinati” a seconda del loro *propensity score*. Ci sono differenti algoritmi attraverso cui gli individui possono essere “abbinati”: *nearest-neighbour- matching*, *radius matching*, *stratification matching*, *Kernel matching*; questi algoritmi sono descritti in Becker e Ichino(2002), Smith e Todd (2005) e Caliendo e Kopeining (2005).

Per analizzare l'impatto che ha la nascita di un figlio sul reddito userò *nearest neighbour matching*: ad ogni individuo (nel nostro caso famiglia) nel gruppo dei trattati è abbinato un individuo del gruppo di controllo che presenta il *propensity score* più vicino.

Per stimare l'impatto si può immaginare che la variabile risposta *Y* sia definita sotto due scenari, osservato o ipotetico: Y^T nel caso l'individuo

riceva il trattamento e Y^{NT} nel caso che l'individuo non riceva il trattamento (Aassve *et al*, 2005).

Tab. 8 Definizione della variabile risposta sotto lo scenario osservato e ipotetico

	Y se famiglia ha figlio	Y se famiglia non ha figlio
trattamento	$Y^T I = 1$ (oss)	$Y^{NT} I = 1$ (no oss)
controllo	$Y^{NT} I = 0$ (no oss)	$Y^T I = 0$ (oss)

Sono interessata all'effetto del trattamento sul gruppo dei trattati; per il gruppo dei trattati questo effetto è denominato ATT (*average treatment effect on the treated*) ed è la differenza tra la media della variabile risposta misurata per le famiglie che hanno avuto un figlio e la media della variabile risposta per la stesso gruppo sotto l'ipotetico scenario che queste non l'abbiano avuto.

$$ATT = E(Y_i^{NT} | I_i = 1) - E(Y_i^T | I_i = 1)$$

Poichè $E(Y_i^T | I_i = 1)$ non è osservabile, lo si può stimare con il valore osservato di $Y^T | I = 0$ sotto l'ipotesi di *strong ignorability*.

Quello che si stima è allora:

$$ATT_{PSM} = E(Y_i^{NT} | I_i = 1, i \in \theta) - E(Y_i^T | I_i = 0, i \in \theta) \quad \text{dove } \theta \text{ rappresenta}$$

il supporto comune.

Avere un supporto comune significa che i trattati e i non trattati devono assumere valori delle variabili esplicative che appartengono allo stesso supporto.

3.2.1 STIMA DEL *PROPENSITY SCORE*

Per stimare l'impatto ho usato il pacchetto *pscore* di STATA (Becker e Ichino(2002)). Il *propensity score* può essere stimato usando sia una regressione logit che probit. Nel nostro caso useremo una regressione probit.

Nella regressione stimata ho imposto la restrizione del supporto comune (*common support*), che restringe il controllo della *balancing property* a quei valori del *propensity score* che appartengono all'intersezione dei supporti del *propensity score* dei trattati e dei non trattati.

Le famiglie che fanno parte del gruppo dei trattati nella mia tesi, sono coloro che tra il tempo t e il tempo $t+1$ hanno avuto un figlio; nelle 759 famiglie analizzate ci sono 121 trattati e 638 non trattati.

Le variabili esplicative usate per stimare il *propensity score* sono: la qualifica lavorativa e il titolo di studio del capofamiglia e del coniuge, l'area geografica di provenienza, il possesso della casa, la grandezza del comune di residenza, numero dei percettori di reddito, numero di figli, ricchezza netta, reddito. Tutte le variabili si riferiscono al tempo t .

Alcune delle variabili usate sono variabili categoriali, ho dovuto, perciò, creare delle variabili *dummies* per poterle inserire nel modello come variabili esplicative. La famiglia di riferimento ha le seguenti caratteristiche:

- non ha figli

- residente al nord in un comune con meno di 40000 abitanti
- casa di proprietà
- un capofamiglia con licenza media, libero professionista³
- coniuge con licenza media e casalinga.
- reddito basso
- ricchezza bassa

Tab. 9 Stime dei coefficienti della regressione probit per stimare il

propensity score

Variabili	coefficiente	Std. Error
capofamiglia diplomato	0,29*	0,16
capofamiglia laureato	0,33	0,25
coniuge diplomato	0,00	0,17
coniuge laureato	0,08	0,25
capofamiglia impiegato	-0,10	0,18
capofamiglia operaio	-0,22	0,18
capofamiglia disoccupato	0,06	0,28
coniuge operaio	0,20	0,25
coniuge libero professionista	0,13	0,29
coniuge disoccupato	-0,03	0,28
coniuge impiegato	0,29	0,24
centro Italia	-0,09	0,19
sud Italia	0,26*	0,15
casa in affitto	-0,212	0,17
Ricchezza alta	-0,29	0,18
comune con più di 40000 ab.	-0,34***	0,13
n percettori di reddito	-0,05	0,19
coppia con un figlio	-0,58***	0,14
coppia con 2 o più figli	-1,34***	0,16
reddito medio/basso	-0,33*	0,20
reddito medio	-0,55**	0,24
reddito medio/alto	-0,52**	0,26
reddito alto	-0,52*	0,28
cost	0,209	0,38

Significatività : *p-value<0,10 ** p-value< 0,05 *** p-value<0,01

³ La scelta migliore sarebbe operaio ma il *propensity score* non risultava bilanciato, ho dovuto, perciò, cambiare base

Tab.10 blocchi creati dalla stima del *propensity score*

limiti dei blocchi	trattamento	
	0	1
0,0161	299	16
0,1	143	21
0,2	80	28
0,3	35	32
0,4	40	24

Come si può vedere dalle stime sopra riportate solo i coefficienti delle variabili coppia con un figlio, coppia con due figli, comune grande, reddito medio e reddito medio/alto sono significativi al 5%. In particolare l'aver un figlio al tempo t rispetto a non averlo fa diminuire del 58% la probabilità di averne un altro; il fatto di abitare in un comune grande rispetto ad un comune piccolo al tempo t fa diminuire la probabilità di avere un figlio del 34%; il fatto di avere un reddito medio e medio/alto rispetto ad avere un reddito basso fanno diminuire la probabilità di avere un figlio rispettivamente del 55% e del 52%. Se, invece, si aumenta il livello di significatività al 10%, anche i coefficienti delle variabili capofamiglia diplomato, sud, reddito medio/basso e reddito alto sono significativi.

Il *propensity score* viene utilizzato per costruire dei blocchi, o celle, entro cui giacciono trattati e non trattati confrontabili. I blocchi sono costituiti da intervalli del *pscore*. La tabella 10 riporta tali intervalli ed evidenzia quante famiglie hanno un *propensity score* all'interno di tali blocchi.

Con questa specificazione del modello, il *propensity score* stimato risulta bilanciato.

3.2.2 STIMA DELL'IMPATTO

Per stimare l'impatto della nascita di un figlio ho usato come variabili obiettivo il reddito familiare equivalente e il reddito familiare.

Per poter analizzare l'impatto della nascita di un figlio sul reddito della famiglia, dovendo confrontare famiglie di numerosità diversa, ho calcolato, attraverso la scala OCSE modificata, il reddito familiare equivalente. Tale scala assegna valore 1 al primo adulto, 0,5 a tutti gli altri adulti componenti la famiglia e 0,3 ai minori di 14 anni⁴.

TAb.11 Stima dell'impatto sul reddito equivalente e sul reddito

	N trattati	n controllo	ATT	Std. Error	t
reddito equivalente	121	108	-2540,431	1729,263	-1,5
reddito	121	108	534,413	2975,285	0,18

Se guardiamo al reddito familiare l'effetto dell'avere un figlio in più è nullo. Se esaminiamo il reddito equivalente, che tiene conto del diverso numero di componenti della famiglia, notiamo che l'impatto è negativo anche se la significatività della stima è bassa. Questo mi porta a concludere che la nascita di un figlio ha una piccola influenza sul reddito della famiglia.

⁴ Il reddito equivalente si ricava dalla seguente formula:

$$\text{reddito equivalente} = \frac{\text{reddito}}{1 * n.\text{primo adulto} + 0,5 * n.\text{altri adulti} + 0,3 * n.\text{minori 14 anni}}$$

Andiamo ora a calcolare l'impatto di un'altra variabile obiettivo, lo stato di povertà. Per condurre questa analisi ho combinato uno stimatore *difference in differences* con la procedura di *matching*.

A differenza di Aassve che usa come misura della povertà due indicatori uno monetario e uno non monetario (indice di privazione), per la mia analisi userò solo un indicatore monetario in quanto i dati esaminati non riportano informazioni sui beni posseduti.

La determinazione dello stato di povertà viene fatto, non all'interno del campione da noi costruito per poter condurre l'analisi, ma all'interno del campione sezionale al quale la famiglia appartiene. Stabilita una linea di povertà (fissata al 60% della mediana della distribuzione del reddito equivalente dell'intero campione sezionale) una famiglia viene considerata povera se il suo reddito equivalente è al di sotto di questa soglia.

Quindi partendo dalla distribuzione del reddito equivalente di ogni anno d'indagine preso in analisi (2000-2008), ho calcolato la mediana del reddito equivalente. Ho dovuto calcolare cinque linee di povertà distinte perchè lo stato povero/non povero va ricercato all'interno di ogni campione sezionale.

Tab. 12 Linea di povertà per gli anni 2000-2008

anno	mediana	linea di povertà
2000	13100	7860
2002	13958	8375
2004	15040	9024
2006	16492	9895
2008	17059	10235

Dopo aver calcolato le linee di povertà, ho costruito una variabile dummy che identifica se la famiglia è povera o meno: prende valore 1 se la famiglia ha un reddito equivalente al di sotto della linea di povertà, 0 se è sopra la linea di povertà. Ho, così, calcolato per trattati e non trattati lo stato di povero non povero. Per assicurarmi che l'analisi fosse a parità di caratteristiche non ho potuto, però, analizzare i trattati e i non trattati nella loro totalità, ho dovuto analizzare i trattati confrontandoli con i non trattati a loro abbinati.

Tab 13 Percentuali di poveri/non poveri al tempo t e la tempo t+1 per i trattati e i non trattati a loro abbinati⁵

		povero al tempo t	povero al tempo t+1
blocco 1	non trattati	25	23
	Trattati	13	18
blocco 2	non trattati	27	27
	Trattati	19	14
blocco 3	non trattati	26	29
	Trattati	7	11
blocco 4	non trattati	34	25
	Trattati	13	19
blocco 5	non trattati	28	22
	Trattati	29	21

Ho, infine, confrontato la situazione di ogni famiglia al tempo t e al tempo t+1, sia nel campione, sia nei trattati e nei controlli a loro abbinati.

Le tabelle sotto riportate riassumono questo confronto.

⁵ Con blocco 1= pscore compreso tra 0,016 e 0,1; blocco 2= pscore compreso tra 0,1 e 0,2; blocco 3= pscore compreso tra 0,2 e 0,3; blocco 4= pscore compreso tra 0,3 e 0,4; blocco 5= pscore maggiore di 0,4

Tab.14 Confronto tra la situazione di povertà al tempo t e al tempo t+1

	non povero al tempo t+1	
	povero al tempo t+1	povero al tempo t+1
povero al tempo t	16	9
non povero al tempo t	7	68

Tab. 15 Confronto tra la situazione di povertà al tempo t e al tempo t+1 per i trattati e i non trattati a loro abbinati

		non povero al tempo t e al tempo t+1	
		povero al tempo t+1	povero al tempo t+1
blocco 1	non trattati	16	67
	trattati	6	75
blocco 2	non trattati	20	66
	trattati	10	76
blocco 3	non trattati	23	68
	trattati	0	82
blocco 4	non trattati	17	57
	trattati	6	75
blocco 5	non trattati	17	68
	trattati	16	66

Attraverso i risultati della tabella 13 ho calcolato le seguenti variabili:

$$Y_T = \left(\frac{N.poveri_{t+1}}{N.tot} \right) - \left(\frac{N.poveri_t}{N.tot} \right)$$

$$Y_{NT} = \left(\frac{N.poveri_{t+1}}{N.tot} \right) - \left(\frac{N.poveri_t}{N.tot} \right)$$

che rappresentano rispettivamente la differenza tra la percentuale di poveri al tempo t+1 e quelli al tempo t per i trattati (Y_T) e per i non trattati (Y_{NT}). Ho, infine, usato queste variabili come variabili obiettivo per la stima dell'impatto, testando così se la differenza è pari a 0 o meno.

Per vedere quanto il risultato sia sensibile alla particolare linea di povertà prescelta, ho, inoltre, provato a spostare la linea di povertà

aumentandola del 20% e diminuendola del 20%. Per entrambe le situazioni ho calcolato Y_T e Y_{NT} .

Le tabelle sotto riportate descrivono la distribuzione dei trattati e dei non trattati, a loro abbinati, rispetto alle diverse linee di povertà. Si riportano, inoltre, i risultati delle stime.

Tab.16 Distribuzione dei trattati e dei non trattati abbinati ponendo la linea di povertà al 40%

		povero al tempo t	povero al tempo t+1
blocco 1	non trattati	8	9
	Trattati	0	0
blocco 2	non trattati	10	9
	Trattati	0	5
blocco 3	non trattati	15	9
	Trattati	0	0
blocco 4	non trattati	9	6
	Trattati	3	6
blocco 5	non trattati	10	18
	trattati	20	8

Tab.17 Distribuzione dei trattati e dei non trattati ponendo la linea di povertà all' 80%

		povero al tempo t	povero al tempo t+1
blocco 1	non trattati	38	36
	Trattati	38	38
blocco 2	non trattati	43	43
	Trattati	29	29
blocco 3	non trattati	39	41
	Trattati	25	18
blocco 4	non trattati	51	40
	Trattati	34	44
blocco 5	non trattati	45	45
	trattati	42	42

Tab.18 Stima dell'impatto ponendo la linea di povertà al 60%

	N trattati	n controllo	ATT	Std. Error	t
blocco 1	16	29	0,028	0,003	8,312
blocco 2	21	18	0,027	0,003	8,246
blocco 3	28	21	0,026	0,003	10,128
blocco 4	32	8	0,025	0,004	5,915
blocco 5	24	2	0,031	0,003	10,262

Tab.19 Stima dell'impatto ponendo la linea di povertà al 40%

	N trattati	n controllo	ATT	Std. Error	t
blocco 1	16	29	0	0,002	0,157
blocco 2	21	18	0,009	0,002	4,366
blocco 3	28	21	0,007	0,002	3,139
blocco 4	32	8	0,004	0,003	1,523
blocco 5	24	2	0,006	0,002	3,314

Tab.20 Stima dell'impatto ponendo la linea di povertà al 80%

	N trattati	n controllo	ATT	Std. Error	t
blocco 1	16	29	0,003	0,004	0,781
blocco 2	21	18	-0,003	0,002	-1,85
blocco 3	28	21	0,002	0,003	0,943
blocco 4	32	8	0,004	0,003	1,338
blocco 5	24	2	0,003	0,004	0,906

Analizzando i risultati della tabella 18 si può notare che l'impatto è significativamente diverso da 0 per tutti e tre i blocchi, ma molto piccolo, ciò significa che la nascita di un figlio non fa aumentare la probabilità di diventare poveri al tempo $t+1$. Lo stesso risultato lo si riscontra nella stima dell'impatto ponendo la linea di povertà al 40%, fatta eccezione per il blocco 1 dove l'impatto non è significativo. Per quanto riguarda l'ATT stimato ponendo la linea di povertà all'80%, è significativamente diverso da 0 solo per il blocco 2, sebbene la significatività sia bassa, e conferma i risultati sopra citati.

3.2.3 SENSIBILITA' DEL PROPENSITY SCORE ALLA SPECIFICAZIONE DEL MODELLO

Una potenziale limitazione del metodo del *propensity score matching* è la necessità di stimare il *propensity score*, poichè può essere sensibile alla specificazione con la quale viene stimato (Rajeev R. Dehejia and Sadek Wahba). Per testare la sensibilità ho eliminato dal modello iniziale alcune variabili.

Nella tabella sottostante vengono riportati i risultati della stima dell'impatto sul reddito equivalente per diverse specificazioni del modello.

Tab. 21 Effetto del trattamento per diverse specificazioni del modello

	n	effetto del	standard	
	osservazioni	trattamento	Error ⁶	valore di t
modello iniziale	718	-2240,431	1668,204	-1,52
tolta n percettori	689	-2150,428	1728,297	-1,244
tolta n percettori dummy studio del coniuge	688	-2447,718	1823,143	-1,343
tolta n percettori dummy studio del coniuge e ricchezza	696	-2465,145	1727,287	-1,427

L'effetto del trattamento varia tra $-2150,428$ e $-2465,145$, i dati, perciò, non sono sensibili alla specificazione del modello.

Poichè l'interesse è avere i due campioni, quello dei trattati e quello dei non trattati, più simili possibile la specificazione migliore del modello è quella del modello iniziale avendo più variabili esplicative .

⁶ Lo standard error è stato calcolato usando un bootstrap con 500 replicazioni

CAPITOLO 4: ANALISI DI STABILITA'

In questa sezione ho elaborato un'ulteriore analisi sugli effetti della nascita di un figlio sul reddito equivalente di una famiglia. Per determinare questo effetto analizzerò la posizione del reddito equivalente di ciascuna famiglia rispetto alla distribuzione del reddito equivalente. Poichè la distribuzione del reddito equivalente varia ogni anno, ho determinato tale distribuzione all'interno del campione sezionale al quale la famiglia appartiene. Determinata la distribuzione ho calcolato i quintili, utili per identificare la classe di appartenenza (posizione) di ciascuna famiglia.

Tab.22 Quintili calcolati per anno

anno	Quintili
2000	1° = < 7868,67
	2° = compreso tra 7868,67 e 11356,88
	3° = compreso tra 11356,88 e 14970,36
	4° = compreso tra 14970,36 e 20441,37
	5° = > 20441,37
2002	1° = < 8635,21
	2° = compreso tra 8635,21 e 12087,6
	3° = compreso tra 12087,6 e 15790,53
	4° = compreso tra 15790,53 e 21323,39
	5° = > 21323,39
2004	1° = < 9466,66
	2° = compreso tra 9466,66 e 13205,53
	3° = compreso tra 13205,53 e 17055,78
	4° = compreso tra 17055,78 e 22873,88
	5° = > 22873,88
2006	1° = < 10480
	2° = compreso tra 10480 e 14534,22
	3° = compreso tra 14534,22 e 18792,24
	4° = compreso tra 18792,24 e 24866,82
	5° = > 24866,82
2008	1° = < 10613,56
	2° = compreso tra 10613,56 e 14927,80
	3° = compreso tra 14927,80 e 19426,75
	4° = compreso tra 19426,75 e 25800
	5° = > 25800

Ho potuto, così, creare una variabile che identifica in quale quintile cade il reddito equivalente di ogni famiglia e confrontare il quintile dell'anno t con quello dell'anno t+1 della stessa famiglia.

Per confrontare la posizione di ogni famiglia nelle due occasioni d'indagine, ho costruito delle matrici di transizione; per far sì che le famiglie fossero confrontate a parità di variabili esplicative, ho confrontato le matrici di transizione dei trattati con quelle dei non trattati abbinati.

Tab.23 Matrice di transizione dei quintili

		quintile al tempo t+1				
		1	2	3	4	5
quintile al tempo t	1	0,18	0,05	0,03	0,01	0,001
	2	0,06	0,07	0,03	0,01	0,01
	3	0,02	0,04	0,09	0,04	0,02
	4	0,004	0,02	0,04	0,08	0,04
	5	0	0,01	0,01	0,05	0,11

Tab. 24 Matrice transizione dei quintili per i non trattati blocco 1

		quintile al tempo t+1				
		1	2	3	4	5
quintile al tempo t	1	0,25	0,06	0,03	0,00	0,00
	2	0,06	0,07	0,04	0,00	0,00
	3	0,02	0,04	0,09	0,04	0,02
	4	0,00	0,01	0,02	0,07	0,04
	5	0,00	0,01	0,01	0,04	0,06

Tab.25 Matrice transizione dei quintili per i trattati blocco 1

		quintile al tempo t+1				
		1	2	3	4	5
quintile al tempo t	1	0,38	0,00	0,00	0,00	0,00
	2	0,06	0,06	0,00	0,00	0,00
	3	0,06	0,00	0,06	0,00	0,00
	4	0,06	0,06	0,06	0,06	0,00
	5	0,00	0,00	0,00	0,13	0,00

Tab. 26 Matrice transizione dei quintili per i non trattati blocco 2

		quintile al tempo t+1				
		1	2	3	4	5
quintile al tempo t	1	0,16	0,05	0,04	0,01	0,01
	2	0,06	0,06	0,03	0,03	0,01
	3	0,00	0,01	0,10	0,05	0,01
	4	0,01	0,03	0,06	0,08	0,02
	5	0,00	0,00	0,01	0,04	0,12

Tab. 27 Matrice transizione dei quintili per i trattati blocco 2

		quintile al tempo t+1				
		1	2	3	4	5
quintile al tempo t	1	0,24	0,00	0,00	0,00	0,00
	2	0,00	0,05	0,00	0,00	0,05
	3	0,00	0,10	0,05	0,00	0,05
	4	0,00	0,05	0,05	0,10	0,10
	5	0,00	0,00	0,00	0,10	0,10

Tab. 28 Matrice transizione dei quintili per i non trattati blocco 3

		quintile al tempo t+1				
		1	2	3	4	5
quintile al tempo t	1	0,10	0,01	0,04	0,00	0,00
	2	0,06	0,06	0,01	0,01	0,01
	3	0,01	0,03	0,09	0,08	0,00
	4	0,00	0,01	0,04	0,16	0,06
	5	0,00	0,03	0,00	0,01	0,18

Tab. 29 Matrice transizione dei quintili per i trattati blocco 3

		quintile al tempo t+1				
		1	2	3	4	5
quintile al tempo t	1	0,18	0,04	0,04	0,00	0,00
	2	0,07	0,04	0,00	0,00	0,00
	3	0,04	0,04	0,00	0,04	0,00
	4	0,04	0,00	0,07	0,04	0,00
	5	0,00	0,00	0,04	0,21	0,14

Tab. 30 Matrice transizione dei quintili per i non trattati blocco 4

		quintile al tempo t+1				
		1	2	3	4	5
quintile al tempo t	1	0,14	0,09	0,00	0,00	0,00
	2	0,00	0,03	0,03	0,00	0,00
	3	0,00	0,00	0,09	0,03	0,00
	4	0,00	0,03	0,06	0,09	0,14
	5	0,00	0,03	0,03	0,03	0,20

Tab. 31 Matrice transizione dei quintili per i trattati blocco 4

		quintile al tempo t+1				
		1	2	3	4	5
quintile al tempo t	1	0,09	0,03	0,03	0,00	0,00
	2	0,06	0,06	0,00	0,00	0,00
	3	0,03	0,06	0,09	0,00	0,09
	4	0,00	0,00	0,13	0,06	0,03
	5	0,00	0,00	0,00	0,13	0,09

Tab. 32 Matrice transizione dei quintili per i non trattati blocco 5

		quintile al tempo t+1				
		1	2	3	4	5
quintile al tempo t	1	0,13	0,03	0,05	0,03	0,00
	2	0,00	0,08	0,03	0,00	0,03
	3	0,08	0,05	0,03	0,00	0,08
	4	0,00	0,00	0,00	0,13	0,03
	5	0,00	0,03	0,00	0,03	0,23

Tab. 33 Matrice transizione dei quintili per i trattati blocco 5

		quintile al tempo t+1				
		1	2	3	4	5
quintile al tempo t	1	0,25	0,04	0,00	0,04	0,00
	2	0,08	0,04	0,04	0,08	0,00
	3	0,00	0,04	0,00	0,00	0,00
	4	0,00	0,08	0,04	0,04	0,00
	5	0,00	0,00	0,04	0,08	0,08

Per valutare l'accordo dei valori dei quintili nelle due occasioni d'indagine, analizziamo la stabilità. A tal fine vengono usati due indici: Cohen's K e Weighted K.

L'indice Cohen's K è un indicatore statistico della proporzione di accordo tra variabili qualitative:

$$K = \frac{\sum \sum w_{ij} \pi_{ij} - \sum \sum w_{ij} \pi_{i.} \pi_{.j}}{1 - \sum \sum \pi_{i.} \pi_{.j}}$$

dove π_{ij} denota la probabilità corrispondente alla classificazione i -esima categoria nel tempo t e j -esima categoria del tempo $t+1$; quindi π_{ii} è la probabilità che il valore dei due quintili sia la stessa nelle due occasioni d'indagine. Poichè le due osservazioni sono indipendenti, si può scrivere π_{ii} come $\pi_{ii} = \pi_{i.} \pi_{.i}$ e la probabilità di uguale accordo come $\Pi_e = \sum \pi_{i.} \pi_{.i}$.

L'indice Cohen's K può variare tra 0 e 1: è 1 quando c'è perfetto accordo, 0 quando l'accordo è uguale a quello atteso per categorie.

L'indice Weighted K è appropriato quando le categorie sono ordinate perchè, in questo caso, il disaccordo dipende dalla differenza tra i punteggi. Esso permette a ogni cella (i, j) di essere "pesata" a seconda del grado di disaccordo tra la i -esima e la j -esima categoria.

$$K = \frac{\sum \pi_{ii} - \sum \pi_{i.} \pi_{.i}}{1 - \sum \pi_{i.} \pi_{.i}} = \frac{\Pi_0 - \Pi_e}{1 - \Pi_e}$$

I pesi w_{ij} usati sono calcolati come segue:

$$w_{ij} = 1 - |S_i - S_j| / |SR - S_1|$$

dove S_i è il punteggio della colonna i e R è il numero delle categorie.

Tab. 34 Indici di stabilità

		Cohen's k	Weighted k
blocco 1	trattati	0,38	0,56
	non trattati	0,41	0,63
blocco 2	trattati	0,4	0,61
	non trattati	0,41	0,59
blocco 3	trattati	0,24	0,57
	non trattati	0,48	0,64
blocco 4	trattati	0,25	0,51
	non trattati	0,42	0,65
Blocco 5	trattati	0,25	0,47
	non trattati	0,46	0,56

L'analisi di questi indici, per i primi tre blocchi, ci suggerisce che i dati si discostano dalla diagonale principale. Tali risultati sono utili per affermare che i dati non sono stabili.

L'informazione fornita da questi indici non permette di trarre conclusioni sull'influenza che la nascita di un figlio può avere sulla diminuzione del valore dei quintili tra la prima e la seconda occasione d'indagine.

Dobbiamo, allora, introdurre un nuovo indice che ci fornisca tale informazione. L'indice che andiamo a introdurre misura la direzione del cambiamento tra la prima occasione d'indagine e la seconda e tiene conto solo delle frequenze relative fuori dalla diagonale principale della matrice di transizione. Denominando "a" la somma delle frequenze relative al di sopra della diagonale principale e "b" quelle al di sotto, un indice con queste caratteristiche (Trivellato- Giraldo, 2006) è dato da:

$$H = \frac{a - b}{a + b}$$

H può variare tra -1 (tutte le frequenze rilevati sono sotto la diagonale principale) e 1 (tutte la frequenze rilevanti sono sopra la diagonale principale), prende valore 0 nel caso di perfetto accordo.

La tabella sotto riportata riassume il valore di questo indice applicato alle matrici di transizione costruite in precedenza.

Tab.35 Indice H per i trattati e i non trattati

		H
	trattati	-1,00
Blocco 1	non trattati	0,05
	trattati	-0,2
Blocco 2	non trattati	0,07
	trattati	-0,65
Blocco 3	non trattati	0,09
	trattati	-0,37
Blocco 4	non trattati	0,25
	trattati	-0,29
Blocco 5	non trattati	0,17

Il risultati sopra riportati evidenziano che nel caso dei trattati, l'indice assume un valore negativo in tutti i gruppi. Ciò significa che le frequenze al di sotto della diagonale principale sono più rilevanti rispetto a quelle al di sopra. Per i non trattati l'indice è molto basso, perciò, trascurabile. Confrontando l'indice per i trattati e i non trattati di ogni blocco, si può concludere, quindi, che la nascita di un figlio provoca una riduzione del valore dei quintili e, quindi, un impoverimento della famiglia.

CONCLUSIONI

Nella tesi ho esaminato se la nascita di un figlio possa avere influenza sul benessere della famiglia. I dati usati per elaborare la tesi sono quelli relativi all'indagine sui bilanci condotta dalla Banca D'Italia, in particolare ho selezionato quelli relativi agli anni 2000-2008.

Per sviluppare l'analisi ho usato la tecnica del *matching*, servendomi del *nearest-neighbour-matching* come algoritmo per "abbinare" i trattati e i non trattati. L'impatto è stato calcolato su diverse variabili, tutte monetarie.

Le prime variabili analizzate sono il reddito equivalente e il reddito. Per quanto riguarda il reddito la nascita di un figlio non ha alcun impatto su questa variabile. L'impatto, invece, sul reddito equivalente risulta negativo, anche se la significatività è bassa. Entrambi i risultati ottenuti sono in linea con le mie aspettative.

Una terza variabile analizzata è lo stato di povertà della famiglia: ho costruito una variabile che identifica se la famiglia è povera o meno nelle due occasioni d'indagine. Attraverso la costruzione di una linea di povertà per ogni anno d'indagine, posta al 60% della mediana, ho determinato una variabile che prende valore 1 se il reddito della famiglia è al di sotto di questa soglia, 0 se non lo è. Attraverso l'uso di uno stimatore *difference in differences* combinato con la procedura di *matching* ho analizzato l'impatto sullo stato di povertà. L'analisi evidenzia che tale variabile non è

influenzata dalla nascita di un figlio, essendo l'impatto stimato molto basso. Per verificare la sensibilità dei dati alla particolare linea di povertà scelta, ho ripetuto la stessa analisi nel caso in cui la linea di povertà venga posta al 40% e all'80% della mediana del reddito equivalente. In entrambi i casi viene confermato che la nascita di un figlio non fa aumentare la probabilità di diventare poveri. Il risultato ottenuto mi ha sorpreso in quanto all'inizio dell'analisi immaginavo che la nascita di un figlio andasse ad influenzare lo stato di povertà di una famiglia. Una spiegazione può essere data dal fatto che le famiglie che hanno avuto un figlio, hanno fatto questa scelta considerando la loro possibilità economica.

Ho condotto, infine, un'ulteriore analisi: partendo dalla distribuzione del reddito equivalente, ho calcolato i quintili e creato una variabile che identifica la posizione del reddito equivalente di ogni famiglia, rispetto a questi. Per verificare se il reddito equivalente delle famiglie rimane nello stesso quintile o meno, ho costruito una matrice di transizione per ogni blocco creato dalla stima del *propensity score*. Attraverso la costruzione di un indice che considera solo le frequenze relative fuori dalla diagonale principale della matrice di transizione, ho potuto verificare che la nascita di un figlio influisce sulla diminuzione del valore del quintile dove cade il reddito equivalente. Questo risultato è in linea con le mie aspettative, inoltre, va a confermare quanto già evidenziato dalle analisi precedenti.

APPENDICE: Tabella di selezione delle variabili

Nella tabella vengono evidenziate le famiglie prese in esame per l'analisi nei bienni 2000-2002, 2002-2004, 2004-2006, 2006-2008.

	2000	2002	2004	2006	2008	PANEL
1 occasione						solo 89
2 occasioni						89-91
3 occasioni						89-91-93
4 occasioni						89-91-93-95
5 occasioni						89-91-93-95-98
6 occasioni	X					89-91-93-95-98-00
7 occasioni	X	X				89-91-93-95-98-00-02
8 occasioni	X	X	X			89-91-93-95-98-00-02-04
9 occasioni	X	X	X	X		89-06
10 occasione	X	X	X	X	X	89-08
1 occasione						solo 91
2 occasioni						91-93
3 occasioni						91-93-95
4 occasioni						91-93-95-98
5 occasioni	X					91-93-95-98-00
6 occasioni	X	X				91-93-95-98-00-02
7 occasioni	X	X	X			91-93-95-98-00-02-04
8 occasioni	X	X	X	X		91-93-95-98-00-02-04-06
9 occasione	X	X	X	X	X	91-08
1 occasione						solo 93
2 occasioni						93-95
3 occasioni						93-95-98
4 occasioni	X					93-95-98-00
5 occasioni	X	X				93-95-98-00-02
6 occasioni	X	X	X			93-95-98-00-02-04
7 occasioni	X	X	X	X		93-95-98-00-02-04-06
8 occasione	X	X	X	X	X	93-95-98-00-02-04-06-08
1 occasione						solo 95
2 occasioni						95-98
3 occasioni	X					95-98-00
4 occasioni	X	X				95-98-00-02
5 occasioni	X	X	X			95-98-00-02-04
6 occasioni	X	X	X	X		95-98-00-02-04-06
7 occasione	X	X	X	X	X	95-98-00-02-04-06-08
1 occasione						solo 98
2 occasioni	X					98-00
3 occasioni	X	X				98-00-02
4 occasioni	X	X	X			98-00-02-04
5 occasioni	X	X	X	X		98-00-02-04-06
6 occasione	X	X	X	X	X	98-00-02-04-06-08
1 occasione	X					solo 2000
2 occasioni	X	X				00-02
3 occasioni	X	X	X			00-02-04
4 occasioni	X	X	X	X		00-02-04-06
5 occasione	X	X	X	X	X	00-02-04-06-08

1 occasione	X				solo 2002
2 occasioni	X	X			02-04
3 occasioni	X	X	X		02-04-06
4 occasione	X	X	X	X	02-04-06-08
1 occasione		X			solo 2004
2 occasioni		X	X		04-06
3 occasione		X	X	X	04-06-08
1 occasione			X		solo 2006
2 occasione			X	X	06-08
1 occasione				x	SOLO 08

BIBLIOGRAFIA

- Aassve A., Davia M., Iacovou M., Mazzucco S. (2005). *Does leaving home make you poor? Evidence from 13 European Countries*. *European Journal of Population* 23(4): 315–338.
- Agresti A. (1990), *Categorical Data Analysis*. New York: Wiley
- Baldini M., Toso S., 2009, *Disuguaglianza, povertà e politiche pubbliche*. Il Mulino, Bologna
- Becker S., Ichino A.(2002). *Estimation of average treatment effects based on propensity score*. *The STATA journal*, 2(4),358-377
- Berthoud R., Robson K. (2003). *Early motherhood and disadvantage: A comparison between ethnic groups*. ISER working paper no. 2003-29. Colchester: University of Essex.
- Del Boca D. (2002) *The Effect of Child Care and Part-time Employment on Labor Supply and Fertility*, *Journal of Population Economics* 14 (Spring): 3.
- Dipartimento Politiche Comunitarie (2007). *Rapporto sullo stato di attuazione del PNR 2006-2008. Nota Aggiuntiva “Donne, Innovazione, Crescita”*.
- Guerini e Associati. *Osservatorio Isfol (2011). Vol. 1*
- Leuven E., Sianesi B. (2003). *Psmatch2 STATA module to perform full Mahalanobis and propensity score matching. Common support graphing, and covariate imbalance testing, version 3.0.0. 13 April 2006*.
- Martini A., Sisti M. (2005). *Valutare il successo delle politiche pubbliche*. Il Mulino
- OECD (2011), *Society at a Glance 2011 - OECD Social Indicators*

- Parisi L.(2008). *Leaving Home and the Chances of Being Poor: The Case of Young People in Southern European Countries*. Labour 22 (Special Issue) 89–114, June 2008
- Pavis S., Platt S., Hubbard G. (2000). *Young people in Scotland: Pathways to social inclusion and exclusion*, Joseph Rowntree Foundation.
- Rajeev H., Dehejia, Wahba S., *Propensity score matching methods for nonexperimental causal studies*. The Review of Economics and Statistics, February 2002, 84(1): 151-161
- Rosenbaum P., Rubin D. (1983). *The central role of propensity score in observational studies for casual effects*. Biometrika, 70(1), 41-55.
- Supplemento al Bollettino Statistico indagini campionarie. I bilanci delle famiglie Italiane nell'anno 2008 (www.bancaditalia.it)
- Trivellato U., Giraldo A. (2006), *Assessing the “Choosiness” of Job-seekers. An Exploratory Approach and Evidence for Italy*. Labour, 20 (1). pp. 1-36.

RINGRAZIAMENTI

Un ringraziamento alla professoressa Anna Giraldo per la sua disponibilità e pazienza nel seguirmi nella stesura della tesi.

Un ringraziamento alla mia famiglia che durante questi anni mi ha sempre sostenuto.

Un ringraziamento particolare a Elena.