

**UNIVERSITÀ DEGLI STUDI DI PADOVA**

**FACOLTÀ DI SCIENZE STATISTICHE**

Corso di laurea Specialistica in Scienze Statistiche Economiche,  
Finanziarie e Aziendali



TESI DI LAUREA

**L'EFFETTO DEL PENSIONAMENTO**

**SUL CONSUMO:**

**STUDIO EMPIRICO IN PRESENZA DI**

**ERRORE DI MISURA**

Relatore: Ch.ma Prof.ssa Alessandra R. Brazzale

Correlatore: Ch.mo Prof. Erich Battistin

Laureanda: Federica Villani

Anno Accademico 2010/2011

# Ringraziamenti

Se sono riuscita a raggiungere questo traguardo, fortemente voluto, è merito anche del sostegno di alcune persone.

Prima di tutto quello della mia relatrice, la Prof. Alessandra R. Brazzale, che in questi mesi ha saputo guidarmi in questo lungo percorso che mi ha portata fino a qui. La desidero ringraziare vivamente per la sua disponibilità, per i suoi preziosi consigli e insegnamenti che sicuramente nella vita professionale verranno utili.

Un ringraziamento profondo anche al mio correlatore, il Prof. Erich Battistin, per gli interessanti spunti, i validi suggerimenti e il suo aiuto in questi mesi.

Ringrazio inoltre il Prof. Helmut Küchenhoff per lo scambio di opinione e la Prof. Adriana Brogini per il suo appropriato consiglio quando ne ho avuto bisogno.

Inoltre desidero ringraziare il personale tecnico del Dipartimento di Scienze Statistiche per la disponibilità dimostrata.

Un ringraziamento profondo alla mia amica e collega Elena, che in questi anni si è sempre dimostrata leale, onesta e di supporto nei miei confronti. Ricorderò con lei i pomeriggi a lezione, le corse per fotocopiare appunti, i pranzi in mensa e le discussioni superficiali ma anche altrettanto profonde. Vorrei anche ricordare Roberta. Insieme abbiamo trascorso una bellissima esperienza, quella del Mimpredo, che merita essere menzionata.

Un ringraziamento sincero alla mia amica Alessandra per la sua costante presenza, il suo supporto in ogni momento della mia vita e la sua capacità a

comprendermi in qualunque occasione.

Desidero ricordare anche Anna, Martina e Gioia per i bei momenti passati insieme. E, ovviamente, Maddalena con la quale ho sempre avuto un bellissimo rapporto di amicizia, fin dalle superiori e con la quale ho sempre trascorso bellissimi momenti di festa.

Mi sento di ringraziare profondamente alcune persone che in questi ultimi due anni ho conosciuto e alle quali tengo molto. In particolare, Davide Merlin, Margherita Morpurgo, il Prof. Paolo Gubitta e Bernardo Moronato.

Il ringraziamento più sentito va ai miei genitori. Senza di loro, il loro amore e supporto, che non è mai venuto meno, non sarei mai diventata quella che ora sono oggi. Non mi hanno mai lasciata sola un momento, mi hanno sempre incoraggiata, anche nei momenti di sconforto, quando hai l'illusione di aver raggiunto un obiettivo e invece ti ritrovi a mani vuote. Quello che mi hanno insegnato è non arrendersi, rialzarsi ed affrontare con grinta la vita. Vi voglio bene.

A mia sorella Giulia, per la sua allegria e tenacia in ogni cosa che fa: questo ho imparato da lei e la ringrazio per il suo supporto in questi anni di università ma anche in quest'ultimo delicato momento, nonostante stesse preparando la maturità.

Un profondo grazie ad Alberto, per la sua costante presenza e il suo amore. Lo ringrazio, in quanto ha sempre avuto la parola giusta in ogni momento. E' riuscito a starmi vicino, con premura ma senza pressione, nei momenti più delicati della mia vita universitaria. Lo ringrazio per i bellissimi momenti trascorsi insieme.

Ringrazio la mia nonna, per le sue preghiere prima degli esami e il suo costante interesse per me e mia sorella.

Ringrazio i miei cugini, Luca, Sarah, Marco e Gianmarco e i miei zii, che mi hanno accompagnato in questo percorso.

# Indice

<b>1</b>	<b>Introduzione</b>	<b>9</b>
1.1	Motivazione del lavoro . . . . .	9
1.2	Obiettivo dello studio . . . . .	11
1.3	Organizzazione del testo . . . . .	17
<b>2</b>	<b>L'errore di misura</b>	<b>19</b>
2.1	Introduzione . . . . .	19
2.2	Modelli, conseguenze e rimedi . . . . .	20
2.2.1	Modelli per l'errore di misura . . . . .	22
2.2.2	Conseguenze nella regressione lineare semplice . . . . .	24
2.2.3	Conseguenze nella regressione lineare multipla e nei modelli non lineari . . . . .	28
2.2.4	Metodo dei momenti . . . . .	29
2.3	Metodi di correzione . . . . .	31
2.3.1	La regressione di calibrazione . . . . .	33
2.3.2	Variabili strumentali . . . . .	35
2.3.3	SIMEX . . . . .	36
2.3.4	Funzioni di punteggio corrette o condizionate . . . . .	38
2.3.5	Verosimiglianza . . . . .	40
2.3.6	Metodi semi parametrici . . . . .	44
2.3.7	Approccio bayesiano . . . . .	46
<b>3</b>	<b>Studio di simulazione</b>	<b>49</b>
3.1	Descrizione dello studio . . . . .	49

3.1.1	Variabili economiche . . . . .	51
3.1.2	Errore di misura . . . . .	52
3.1.3	Parametri della simulazione . . . . .	55
3.1.4	Scenari considerati . . . . .	59
3.1.5	Indici riassuntivi . . . . .	60
3.2	Risultati della simulazione . . . . .	62
3.2.1	Confronto fra risultati noti ed evidenza empirica . . . . .	62
3.2.2	Commenti al caso S0.M0. e S1.M0. . . . .	65
3.2.3	Tabelle per i casi S0.M1. e S1.M1. . . . .	70
3.3	Interpretazione ed implicazioni economiche . . . . .	75
3.3.1	Caso S0.M1. . . . .	75
3.3.2	Caso S1.M1. . . . .	78
<b>4</b>	<b>Un passo in avanti: correzione tramite SIMEX</b>	<b>81</b>
4.1	Il metodo SIMEX applicato al caso S0.M0. . . . .	81
4.2	Il metodo SIMEX applicato al caso S1.M1. . . . .	84
4.3	Una proposta alternativa per il caso S1.M1. . . . .	87
<b>5</b>	<b>Conclusioni e sviluppi futuri</b>	<b>91</b>
5.1	Conclusioni tratte dallo studio di simulazione . . . . .	91
5.2	Sviluppi futuri . . . . .	92
<b>A</b>	<b>Tabelle complete dello studio di simulazione</b>	<b>93</b>
<b>B</b>	<b>Stima della varianza</b>	<b>106</b>
B.1	Il metodo jackknife . . . . .	106
B.2	Il metodo bootstrap . . . . .	107
B.3	Il metodo sandwich . . . . .	107
<b>C</b>	<b>Istruzioni di R per lo studio di simulazione</b>	<b>109</b>
<b>D</b>	<b>Istruzioni di R per il metodo SIMEX</b>	<b>117</b>
	<b>Riferimenti Bibliografici</b>	<b>121</b>

# Elenco delle figure

2.1	Errore di misura classico vs berkson . . . . .	27
2.2	Metodo SIMEX . . . . .	39
3.1	Consumo per individui con età fra 61–65 anni . . . . .	56
3.2	Andamento del consumo rispetto all’eleggibilità . . . . .	57



# Elenco delle tabelle

2.1	Tabella riassuntiva per diversi modelli per l'errore di misura . . .	28
3.1	Valori teorici di $\lambda$ , del coefficiente angolare e della varianza residua . . . . .	64
3.2	Media campionaria del coefficiente angolare e della varianza residua . . . . .	65
3.3	Caso S0.M0: distorsione relativa e livello di copertura reale . . .	67
3.4	Caso S1.M0: distorsione relativa e rapporto fra varianza e distorsione . . . . .	68
3.5	Caso S1.M0: livello di copertura reale per intervalli bilaterali e unilaterali . . . . .	69
3.6	Caso S0.M1: distorsione relativa e rapporto fra varianza e distorsione . . . . .	71
3.7	Caso S0.M1: livello di copertura reale e test statistici . . . . .	72
3.8	Caso S1.M1: distorsione relativa e rapporto fra varianza e distorsione . . . . .	73
3.9	Caso S1.M1: livello di copertura reale e test statistici . . . . .	74
4.1	SIMEX al caso S0.M0. . . . .	83
4.2	SIMEX al caso S1.M1. . . . .	86
4.3	SIMEX: una proposta alternativa . . . . .	89



# Capitolo 1

## Introduzione

### 1.1 Motivazione del lavoro

Il pensionamento e, in particolare, il modo in cui le persone affrontano questo cambiamento, è un fenomeno di notevole interesse per quanto riguarda l'analisi dell'andamento del livello del consumo nell'ambito dei programmi di valutazione economica.

In base al modello del ciclo di vita di Modigliani (vedi [1] per maggiori dettagli) i consumatori risparmiano parte del reddito per far fronte alle loro esigenze di consumo al termine dell'attività lavorativa, durante il pensionamento. Questa semplice idea è alla base della teoria del ciclo di vita, formulata dal premio Nobel Modigliani: si risparmia da “giovani”, quando si dispone di un reddito, per poter spendere e finanziare le spese anche da “anziani”, quando quel reddito mancherà. È il ciclo di vita del risparmio, positivo durante la fase lavorativa, negativo durante il pensionamento. Infatti, secondo Modigliani, il pensionamento, esclusi shock di salute, è un evento pianificabile da parte degli agenti. I consumatori allocano risorse nel ciclo di vita con l'obiettivo di mantenere costante l'utilità marginale del consumo. Dunque, i consumatori arrivano a pianificare le loro scelte per perseguire l'obiettivo appena descritto. Ne segue che, intuitivamente, in corrispondenza del pensionamento, non si dovrebbe vedere un calo drastico del consumo. Questo

calo, in parte, potrebbe essere spiegato tramite una naturale diminuzione di spesa per alcune categorie di beni. Si pensi, ad esempio, alle spese inerenti la necessità di utilizzare i mezzi di trasporto, di pranzare fuori, di comprare vestiti consoni al tipo di lavoro. In pratica, ci si riferisce a quelle spese legate al lavoro.

Recenti lavori hanno però evidenziato la presenza di un crollo pesante del consumo nel momento del pensionamento, risultato che difficilmente si accosta al comportamento ideale previsto dal modello del ciclo di vita di Modigliani. Osservare una caduta drastica del consumo è assimilabile al fatto che gli agenti arrivano non preparati al pensionamento. Avere una caduta drastica del loro consumo è una conseguenza derivante dal fatto che anche il loro reddito disponibile diminuisce. Insomma, è come se il pensionamento fosse uno shock non previsto.

La presenza di questo fenomeno è stata analizzata per il Regno Unito (vedi [2] per maggiori dettagli, indicato con BBT nel seguito), per gli Stati Uniti (vedi [3] per maggiori dettagli, BSW nel seguito) e per l'Italia (vedi [4] per maggiori dettagli, MMW nel seguito) ed è noto come *the retirement consumption puzzle* (o *the retirement savings puzzle*).

BBT usa dati provenienti dall'indagine sulla spesa delle famiglie inglesi, nota come FES, <sup>1</sup> per esaminare i cambiamenti nella spesa delle famiglie nel momento del pensionamento. Gli autori trovano evidenza del fenomeno in quanto solo una parte del calo del consumo può essere spiegata e attribuita all'aumento di tempo libero che segue il pensionamento: il pensionamento dovrebbe apparire come un evento prevedibile e invece si registra un crollo del consumo tale da sembrare un comportamento provocato da uno shock inatteso. Sembra che le persone non programmino l'evento e si trovino dunque impreparate. Inoltre BBT evidenzia l'importanza di distinguere tra pensionamento volontario e involontario: infatti si osserva un crollo significativo nel consumo alimentare solamente per coloro che sono costretti ad anticipare il pensionamento a causa di problemi di salute o per la perdita del lavoro.

---

<sup>1</sup>FES: *Family Expenditure Survey*

BSW usa dati panel sulle famiglie degli Stati Uniti, tratti dal PSID<sup>2</sup>. Gli autori concludono dicendo che i risultati ottenuti non sono completamente a supporto della teoria del ciclo di vita di Modigliani ma non possono essere nemmeno esclusi da qualsiasi schema razionale: i risultati infatti possono essere spiegati dal *mental accounting*, una teoria sviluppata dall'economista statunitense Richard Thaler. Costui ipotizza che le scelte economiche delle persone sono mediate da un vero e proprio sistema di contabilità mentale, che non è neutrale e produce comportamenti non del tutto coerenti con il modello microeconomico neoclassico standard, basato sulla perfetta razionalità degli agenti.

MMW usa dati forniti dall'indagine sui consumi delle famiglie italiane, nota come ISFB<sup>3</sup> per il periodo dal 1985 al 1996. Gran parte del lavoro è dedicato ad illustrare come il consumo per diverse categorie di beni può variare con l'età degli individui e con il loro pensionamento. Il consumo di beni collegati all'attività lavorativa decresce mentre aumenta quello di beni prodotti in casa. Gli autori dunque stabiliscono che il consumo di beni collegati all'attività lavorativa (trasporti, vitto fuori casa e indumenti) diminuiscono a seguito del pensionamento mentre non ci sarebbero prove di brusche cadute per la categoria di beni non durevoli (spese per la salute, carburante e spese per la casa). I risultati sono a sostegno della tesi che in Italia non ci sarebbe un retirement consumption puzzle. Questo potrebbe essere dovuto alla presenza, in Italia, di un adeguato sostegno economico per quelle famiglie di neo pensionati che si trovano in difficoltà economica.

## 1.2 Obiettivo dello studio

La tesi prende spunto da un lavoro proposto da Battistin e coautori, collocato in letteratura nell'ambito degli studi relativi alla stima dell'effetto del pensionamento sul consumo. Il lavoro si intitola: *The Retirement Consump-*

---

<sup>2</sup>PSID: *Panel Study of Income Dynamics*

<sup>3</sup>ISFB: *Italian Survey of Family Budgets*

*tion Puzzle: Evidence from a Regression Discontinuity Approach*, scritto nel 2005 (vedi [5]). Il lavoro utilizza informazioni ricavate dalla Banca d'Italia in merito ad un'indagine sul reddito e la ricchezza delle famiglie italiane (SHIW<sup>4</sup>) per il periodo 1993–2004. I dati riguardano la spesa delle famiglie italiane circa il consumo di alimenti e beni non durevoli. Qui di seguito verranno ora presentati i passaggi rilevanti di questo lavoro.

Vengono chiamati  $Y_1$  e  $Y_0$  i consumi potenziali che si osserverebbero nel caso in cui l' $i$ -esimo individuo fosse rispettivamente in pensione oppure no. Il risultato potenziale, per ogni singola unità di osservazione, è il valore che una certa variabile, detta appunto risultato, acquista, qualora si realizza un valore specifico del trattamento. Seguendo l'approccio indicato da Donald Rubin, l'effetto causale del pensionamento sul consumo è definito come differenza in termini di consumo individuale  $\beta = Y_1 - Y_0$ , il quale però non può essere stimato. Invece, è possibile considerare e stimare l'effetto medio come

$$\beta = E\{Y_1 - Y_0\} = E\{Y_1\} - E\{Y_0\}. \quad (1.1)$$

Ovviamente è impossibile osservare per lo stesso soggetto, nello stesso istante temporale, entrambi gli stati e uno, fra i due risultati potenziali, è sempre mancante. Ma se si può ritenere che i due sottocampioni, riferiti rispettivamente alle persone pensionate e a quelle non pensionate, siano equivalenti in termini di meccanismo di assegnazione casuale, allora l'effetto causale medio del fenomeno sulla quantità oggetto di studio può essere calcolato come differenza fra i due valori attesi riferiti rispettivamente ai due sottocampioni. Per approfondimenti riguardanti il modello causale indicato da Donald Rubin si veda [6] e [7].

Nel lavoro in esame,  $R$  è una variabile binaria che rappresenta la condizione del capofamiglia, dove  $R = 1$  se il soggetto  $i$  è in pensione,  $R = 0$  se il soggetto non lo è. Quindi  $Y_1$  è il consumo che si osserverebbe nel caso di pensionamento, mentre  $Y_0$  è il consumo che si osserverebbe nel caso di non pensionamento.

---

<sup>4</sup>SHIW: *Survey on Household Income and Wealth*

Gli autori utilizzano un approccio basato sui disegni di discontinuità (*regression discontinuity approach*). In queste circostanze infatti ci si può avvalere di tale metodo in quanto la variabile  $R$  dipende da una variabile osservabile  $S^*$  e nel supporto di  $S^*$  esiste un punto noto  $\bar{s}$  dove la probabilità di essere trattato cambia in modo discontinuo. Valutando infatti un intorno di  $\bar{s}$  e utilizzando il disegno di discontinuità, si possono recuperare le proprietà di un esperimento casuale, ossia l'intorno di  $\bar{s}$  ha la stessa valenza di un confronto sperimentale. Quindi, se  $\bar{s}$  rappresenta il punto di discontinuità, una *regression discontinuity* è definita se

$$\Pr\{R = 1|\bar{s}^+\} \neq \Pr\{R = 1|\bar{s}^-\}, \quad (1.2)$$

dove  $\bar{s}^+$  e  $\bar{s}^-$  si riferiscono a quei soggetti che si trovano localmente e rispettivamente sopra e sotto rispetto a  $\bar{s}$ . Per approfondimenti sui disegni di discontinuità (RDD) si veda [8], [9], [10] e [11].

Nel lavoro di riferimento la variabile  $S^*$  rappresenta l'eleggibilità intesa come numero di anni che mancano alla pensione ( $S^*$  negativa) o che sono trascorsi dal momento in cui il soggetto è diventato idoneo alla pensione ( $S^*$  positiva); si assume che il supporto di questa variabile sia continuo. Inoltre, per la variabile  $S^*$  vale la regola deterministica

$$R = \mathbf{I}(S^* \geq \bar{s}), \quad (1.3)$$

ovvero, il valore assegnato alla variabile  $R$  è legato al fatto che il valore assunto dalla variabile  $S^*$  sia positivo. In altre parole, gli individui sono idonei al pensionamento se e solo se presentano un valore della variabile  $S^*$  non inferiore al punto di soglia  $\bar{s} = 0$ . Questa variabile dicotomica indica se l'individuo è o non è in pensione e potrà valere 1 se e solo se  $s^* \geq 0$ . Con ciò, si sottolinea il fatto che un soggetto, avente un valore dell'eleggibilità negativo, non potrà apparire in pensione in quanto dovrà versare ancora anni di contributi. D'altro canto, un soggetto con un valore positivo dell'eleggibilità, è idoneo al pensionamento e ha la possibilità di scegliere se andarci oppure continuare a lavorare.

Sia

$$Y = Y_0 + R(s)\beta \quad (1.4)$$

il consumo osservato. Per sottolineare la dipendenza fra  $R$  e  $S^*$  si scrive  $R(s^*)$ . L'espressione si semplifica a seconda del valore assunto dalla variabile  $R$ . Infatti, nel caso  $R = 1$  si ha  $Y \equiv Y_1$  mentre, se  $R = 0$  allora  $Y \equiv Y_0$ . La differenza fra i valori attesi nei due gruppi costituiti rispettivamente da coloro che riportano valori sotto e sopra la soglia di idoneità  $\bar{s}$

$$E[Y|\bar{s}^+] - E[Y|\bar{s}^-] \quad (1.5)$$

può essere scritta come

$$E[Y_0|\bar{s}^+] - E[Y_0|\bar{s}^-] + E[R(s^*)\beta|\bar{s}^+] - E[R(s^*)\beta|\bar{s}^-]. \quad (1.6)$$

Poichè vale la regola deterministica (1.3) l'espressione può essere riscritta come

$$E[Y_0|\bar{s}^+] - E[Y_0|\bar{s}^-] + E[\beta|\bar{s}^+]. \quad (1.7)$$

**Condizione 1.** Il valore atteso di  $Y_0$  condizionato ad  $S^*$  è una funzione continua di  $S^*$  in  $\bar{s}$ .

Questa condizione stabilisce che, in un mondo dove non esiste il fenomeno del pensionamento, non dovrebbe nemmeno sussistere una discontinuità attorno al punto di soglia  $\bar{s}$  ma piuttosto un andamento del consumo continuo in corrispondenza della soglia. Se vale questa condizione allora si può scrivere

$$E[\beta|\bar{s}^+] \equiv E[Y|\bar{s}^+] - E[Y|\bar{s}^-], \quad (1.8)$$

ovvero, l'effetto causale medio che il pensionamento provoca sui consumi di quei soggetti localmente idonei è definibile in termini di differenza fra i valori attesi dei consumi osservati, rispettivamente sopra e sotto la soglia  $\bar{s}$ .

Se vale la condizione 1, usando la legge dei valori attesi iterati e notando che  $E[R(s^*)\beta|R = 0, \bar{s}^+] = 0$ , allora l'espressione (1.7) produce

$$E[R(s^*)\beta|\bar{s}^+] = E[\beta|R = 1, \bar{s}^+]\Pr[R = 1|\bar{s}^+], \quad (1.9)$$

in modo che l'impatto medio sugli individui pensionati in un intorno destro di  $\bar{s}$  sia identificato da

$$E[\beta|R = 1, S^* = \bar{s}^+] = \frac{E[Y|S^* = \bar{s}^+] - E[Y|S^* = \bar{s}^-]}{E[R|S^* = \bar{s}^+]}. \quad (1.10)$$

Il risultato ottenuto fonda sul fatto che la probabilità di trovare individui in pensione, pur non essendo idonei, sia zero. Questo semplifica l'espressione (1.6), senza dover fare ulteriori assunzioni sul comportamento degli agenti. Segue che la (1.10) può essere stimata usando il metodo delle variabili strumentali, dove la variabile  $S^*$  viene usata come strumento per correggere la natura endogena della variabile  $R$  condizionata ad  $S^*$ .

Questo è ciò che dovrebbe succedere a livello teorico. Ovvero, la probabilità di andare in pensione, condizionata a diversi valori della variabile  $S^*$ , dovrà essere zero per valori di  $S^*$  appartenenti all'intorno sinistro della soglia  $\bar{s}$ . In corrispondenza di  $\bar{s}$  si dovrà evidenziare un salto a dimostrazione del fatto che una buona porzione di individui, appena raggiunta l'idoneità, decidono di andare in pensione. E' ragionevole ipotizzare che altri individui, pur essendo idonei, continuino a lavorare: la curva dunque, nell'intorno destro di  $\bar{s}$ , avrà un andamento crescente fino a che tutti siano in pensione.

Contrariamente a quello che ci si potrebbe attendere, con dati reali si osserva una porzione non trascurabile di individui che appaiono in pensione pur non essendo idonei. Ad esempio, si osserva un 2.5% di uomini che, nonostante manchi ancora un anno per andare in pensione ( $s^* = -1$ ), lo sono già. Per le donne invece si osserva un valore pari al 14%. In corrispondenza di  $s^* = -2$  invece si osserva l'1.5% di uomini e il 3.5% di donne, non ancora eleggibili ma già in pensione. Quindi, la probabilità di essere in pensione, avendo un valore negativo della variabile  $S^*$ , è diverso da zero. Questo va a contrastare la (1.3).

Gli autori cercano di spiegare tale comportamento ipotizzando una struttura per l'errore di misura contenuto in  $S^*$ . Supponendo che  $S^*$  sia una variabile misurata con errore, gli autori valutano sotto quali condizioni dell'errore di misura, si è in grado di identificare il parametro d'interesse  $\beta$  e se i risultati che si trovano sono robusti.

Si assume infatti che, nel gruppo di individui per i quali viene valutato il valore della variabile  $S^*$ , una porzione di essi riportino il valore corretto e una parte il valore errato. Si tratta di una **mistura**, che formalmente è espressa come

$$S_{\text{mistura}} = S^*Z + S_{\text{classico}}(1 - Z), \quad (1.11)$$

dove  $Z$  rappresenta una variabile binaria che vale 1 se il valore riferito è corretto, mentre è uguale a 0 altrimenti. La variabile  $S_{\text{classico}}$  invece rappresenta la variabile  $S^*$  contaminata da un errore di misura di tipo classico. Verranno forniti maggiori dettagli in seguito.

La (1.2) motiva l'ipotesi di presenza di errore di misura in  $S^*$ . Infatti, dovrebbe essere una funzione discontinua in  $\bar{s} = 0$ , in quanto una buona porzione di soggetti, appena diventano idonei, decidono di andare in pensione e per valori negativi dell'eleggibilità non si dovrebbero osservare soggetti in pensione. Con dati reali, non è così.

Lo stimatore (1.10) dipende dalla distribuzione congiunta di  $(Y, R, S^*)$ . A causa dell'errore di misura in  $S^*$  lo stimatore costruito non è consistente per il parametro d'interesse. Invece, il seguente rapporto

$$E[\beta | R = 1, S_{\text{mistura}} = \bar{s}^+] = \frac{E[Y | S_{\text{mistura}} = \bar{s}^+] - E[Y | S_{\text{mistura}} = \bar{s}^-]}{E[R | S_{\text{mistura}} = \bar{s}^+] - E[R | S_{\text{mistura}} = \bar{s}^-]} \quad (1.12)$$

identifica il parametro d'interesse ovvero l'effetto causale che il fenomeno del pensionamento provoca sul consumo, a condizione che l'errore di misura di tipo **mistura** sia compatibile con i dati a disposizione. Questa espressione può essere stimata con le variabili strumentali, utilizzando l'eleggibilità  $S$  come strumento per la scelta di pensionamento  $R$  (si veda [12] per maggiori dettagli).

In conclusione, gli autori stimano un calo del consumo relativo ai beni non durevoli del 9.8%. Tale risultato può essere comunque spiegato dal fatto che gli individui, appena entrano in pensione, cambiano stile di vita. Le spese collegate all'attività lavorativa vengono sospese. Inoltre, nel lavoro, viene mostrato che il pensionamento porta ad una riduzione del numero di adulti all'interno del nucleo familiare. Anche questa circostanza giustifica il

calo del consumo stimato. Le conclusioni tratte e i risultati ottenuti vanno a sostegno della tesi che in Italia non esisterebbe un retirement consumption puzzle, come dimostrato appunto nel lavoro di Miniaci R., Monfardini C. e Weber G., qui indicato MMW (si veda [4]).

### 1.3 Organizzazione del testo

Nel Capitolo 2 illustreremo la teoria sottostante l'errore di misura. Dopo una breve introduzione, presenteremo i modelli per l'errore di misura e i metodi di correzione proposti in letteratura.

Il Capitolo 3 è dedicato interamente allo studio di simulazione effettuato. Descriveremo le variabili economiche utilizzate, le due strutture adottate per l'errore di misura, l'assegnazione dei valori ai parametri della simulazione, gli scenari considerati ed infine gli indici calcolati. Il Capitolo 3 si conclude con l'interpretazione dei risultati ottenuti e le implicazioni economiche. I risultati trovati serviranno a rispondere ad una precisa domanda di ricerca che guida l'intero capitolo.

Nel Capitolo 4 viene considerato il metodo SIMEX per la correzione dell'errore di misura. Il metodo verrà applicato in particolare a due casi. Infine, il capitolo 4 si concluderà con una proposta alternativa del SIMEX.



# Capitolo 2

## L'errore di misura

### 2.1 Introduzione

In molte discipline scientifiche la quasi totalità delle misurazioni sono affette da errore di misura. I problemi connessi all'errata misurazione delle variabili possono essere ricondotti a cause diverse: alcuni esempi possono essere gli elevati costi di rilevazione dei dati per ottenere misure esatte, l'imprecisione degli strumenti utilizzati, analisi di laboratorio poco curate, l'intrinseca variabilità naturale o la natura soggettiva di alcune variabili oggetto di studio. Se non se ne tiene conto e non si opera una correzione, gli errori di misura sono una fonte di problemi inferenziali non irrilevanti (si veda [13] per maggiori dettagli). In letteratura si è a lungo evidenziato il fatto che la presenza di errori di misura porta a stime distorte dei parametri. Inoltre gli errori di misura possono incidere sulla varianza e sul livello di copertura degli intervalli di confidenza degli stimatori. Queste circostanze sorgono poichè, non potendo osservare la variabile d'interesse  $X$ , si devono utilizzare le informazioni riguardanti un'altra variabile  $W$  e questa non risulta perfettamente correlata con  $X$ . Gli scostamenti fra la vera variabile  $X$  e la variabile  $W$  misurata con errore sono chiamati *errori di misura* (*measurement errors*). Convenzionalmente, si usa questo termine per fare riferimento a variabili di tipo numerico, continue o discrete, mentre per variabili di tipo categoriale si

predilige un termine che fa riferimento alla natura qualitativa delle variabili e si usa la parola inglese *missclassification* per definire un'errata classificazione. Il modello statistico per dati affetti da errori di misura viene chiamato **modello per l'errore di misura**. Negli ultimi vent'anni in letteratura sono stati presentati numerosi metodi per correggere la presenza dell'errore di misura.

Nel Paragrafo 2.1 verrà fatta una breve introduzione all'errore di misura. Nel Paragrafo 2.2 verranno indicati i concetti di base e i modelli per l'errore di misura. Seguirà una presentazione sulle conseguenze alle quali si può andare incontro nel caso in cui la presenza dell'errore di misura venga ignorata. Infine, nel Paragrafo 3.2, verrà fornita una visione d'insieme sui metodi proposti in letteratura per far fronte al problema della presenza dell'errore di misura in una o più variabili esplicative, includendo la teoria sottostante.

## 2.2 Gli errori di misura: modelli, conseguenze e rimedi

Esiste una vasta letteratura che approfondisce il problema della presenza dell'errore di misura nelle variabili presenti nel modello. Per il caso della regressione lineare semplice e per i modelli lineari generalizzati si veda [14] per maggiori dettagli. Per un'analisi approfondita sui modelli non lineari si veda [15] per maggiori dettagli.

Verrà indicata con  $Y$  la variabile risposta del modello. Viene fatta una distinzione tra due tipi di variabili concomitanti: con  $Z$  verrà indicato il vettore contenente tutte quelle variabili che sono misurate senza errore mentre con  $X$  quelle che non possono essere osservate esattamente. La caratteristica sostanziale del problema dell'errore di misura nasce dal fatto che si osserva la variabile  $W$  invece di  $X$ . Il primo passo è quello di costruire un modello per l'errore di misura mentre l'obiettivo primario è quello di ottenere stime non distorte dei parametri d'interesse nella situazione di dover stimare un modello per  $Y$  in termini di  $(Z, W)$  invece di  $(Z, X)$ . Per avere successo in questa delicata operazione sono richieste attente analisi. Si può sostituire

infatti  $W$  al posto di  $X$  ma devono essere eseguite le correzioni necessarie, altrimenti si ottengono stime distorte, come verrà illustrato in seguito.

Se il meccanismo dell'errore di misura viene visto in termini di distribuzione della variabile surrogato  $W$  data la vera variabile d'interesse  $X$ , è possibile esprimere la densità congiunta delle variabili rilevanti come

$$f(W, Y, X, Z) = f(W|Y, X, Z)f(Y|X, Z)f(X|Z)f(Z) \quad (2.1)$$

Il primo termine si definisce **modello per l'errore di misura** (*measurement model*). In altre parole si tratta della densità della variabile affetta da errore di misura  $W$ , condizionata alla variabile risposta  $Y$ , alla vera variabile d'interesse  $X$  e al vettore contenente le altre variabili esplicative  $Z$ . Questo descrive come la variabile esplicativa surrogato  $W$  deriva dalla vera variabile esplicativa  $X$ , considerando il fatto che le variabili  $Y$  e  $Z$  possono influenzare questo processo.

Il secondo termine indica il **modello per la variabile d'interesse** (*disease model*) e descrive la relazione fra la variabile risposta  $Y$  e le variabili esplicative  $(X, Z)$ . Tipicamente il primo obiettivo consiste nel scoprire la forma di questa relazione.

Considerati insieme, il terzo e quarto termine costituiscono il **modello per l'esposizione** (*exposure model*) ovvero la distribuzione congiunta di  $(X, Z)$ .

Quando si lavora con gli errori di misura capita spesso di non essere in grado di stimare i parametri d'interesse servendosi solamente delle informazioni contenute nel campione formato dai valori delle variabili  $(Y, Z, W)$ . Alcuni parametri che caratterizzano la distribuzione congiunta di  $(Z, X, W)$  devono essere noti oppure stimati al fine di ottenere una stima del parametro d'interesse. Sono necessarie dunque delle **informazioni supplementari** e la selezione dipende molto dal tipo di modello scelto per l'errore di misura. In seguito si approfondirà la questione.

Infine, si precisa che questo lavoro si concentrerà sui problemi che sorgono quando si vuole stimare un modello che mette in relazione una variabile risposta  $Y$  con le vere variabili esplicative  $X$  e le variabili non affette da errore  $Z$ , date le osservazioni sulla variabile  $W$ , affette da errore, in aggiunta

a quelle su  $Y$  e  $Z$ . Anche la  $Y$  potrebbe essere affetta da errore di misura ma l'attenzione qui si restringe esclusivamente alla presenza dell'errore di misura sulle variabili esplicative.

### 2.2.1 Modelli per l'errore di misura

Per analizzare gli effetti dell'errore di misura è fondamentale costruire un modello per il processo che provoca l'errore di misura. In letteratura si distinguono due grandi gruppi:

- modelli per l'errore che plasmano la distribuzione condizionata di  $W$  dato  $(Z, X)$  e fra questi si trova il modello per l'errore di misura classico;
- modelli per l'errore che utilizzano la distribuzione condizionata di  $X$  dato  $(Z, W)$  e fra questi si trova il modello per l'errore di misura di Berkson.

Il modello più semplice per l'errore di misura è il **modello classico**

$$W = X + U \tag{2.2}$$

Questo modello risulta appropriato quando si tenta di determinare direttamente  $X$  ma si è incapaci di farlo a causa di varie fonti di errori di misura, come per esempio la taratura dello strumento, errori manuali, l'imprecisione dell'operatore. Nel modello di tipo classico è la quantità d'interesse  $x$  ad essere fissa per ciascun individuo ma il valore  $w$  che si va a misurare è perturbato da errore. Ha senso perciò ipotizzare, in queste circostanze, un modello di tipo additivo dove  $U$  ha media zero ed è indipendente da  $X$ . Infatti, dal momento che si impone che  $W$  sia una misura non distorta per  $X$  e che quindi il valore atteso di  $W$  condizionato ad  $X$  e a tutte le altre variabili  $Z$  misurate senza errore sia pari ad  $X$ , allora deve essere che  $E[U|X, Z] = 0$ . Inoltre la varianza di  $U$  può essere di tipo omoschedastico oppure eteroschedastico e viene indicata con  $\sigma_u^2$ .

Un modello alternativo è rappresentato dal **modello di Berkson**, chiamato anche **modello controllato**. In questo modello è la vera variabile  $X$  a variare

attorno a  $W$  ossia

$$X = W + U \quad (2.3)$$

dove  $U$  ha media zero ed è indipendente da  $W$ . Nel modello di Berkson si ha che  $E[X|W, Z] = W$  dal momento che  $E[U|W, Z] = 0$  e dunque  $W$  è un predittore non distorto per  $X$ . Se nel modello di tipo classico  $U$  è indipendente in media dalla vera variabile  $X$ , nel modello di Berkson  $U$  è indipendente in media dalla variabile osservata  $W$ . Tipicamente un modello di questo tipo per l'errore di misura si può trovare in studi di laboratorio o in situazioni sperimentali nel caso in cui la variabile osservata  $W$  si possa misurare con precisione perchè stabilita dal ricercatore stesso mentre la variabile d'interesse  $X$  cambia a causa di svariate fonti d'errore, intrinseche nella natura dell'oggetto stesso. Per esempio si supponga che, in un disegno sperimentale, si debba solidificare in un forno un certo materiale ad una specifica temperatura  $W$ . Nonostante l'operatore possa impostare la temperatura desiderata  $W$ , non si può conoscere con certezza la temperatura vera  $X$ , all'interno del forno, perchè questa potrebbe variare attorno al valore di  $W$ . Questo esempio fa capire bene quanto sia ragionevole l'assunzione che  $E(X|W) = W$ . In letteratura sono state sviluppate procedure di stima e di inferenza per entrambi i modelli.

E' utile riportare un'altra importante distinzione: la differenza fra errori di tipo differenziale e non differenziale. Si presenta un errore di misura non differenziale quando  $W$  non apporta su  $Y$  più informazioni rispetto a quelle contenute in  $X$  e  $Z$ . L'errore in  $W$  quindi è non differenziale se la distribuzione condizionata di  $Y$  dato  $(X, Z, W)$ , indicata come  $f_{Y|XZW}(y|x, z, w)$ , è la stessa di  $Y$  dato  $(X, Z)$ , indicata come  $f_{Y|XZ}(y|x, z)$ , ossia  $f_{Y|XZW}(y|x, z, w) = f_{Y|XZ}(y|x, z)$ . In questo caso,  $W$  è chiamato **surrogato** per  $X$ . Quando invece  $f_{Y|XZW}(y|x, z, w) \neq f_{Y|XZ}(y|x, z)$  l'errore è **differenziale**. In letteratura alcuni metodi statistici, atti a correggere gli errori di misura, si basano sull'assunzione che gli errori siano non differenziali e la ragione principale è che questa ipotesi semplifica molti passaggi. Ma è sempre importante domandarsi se sia ragionevole assumere un errore di questo tipo.

E' accettabile in molti casi, per esempio quando l'errore di misura è riconducibile ad errori legati ad analisi di laboratorio o agli strumenti utilizzati ma soprattutto sono tipici nelle indagini panel.

Per riflettere sulla plausibilità dell'assunzione che  $W$  possa o no essere un surrogato è utile immaginare di poter osservare direttamente  $X$  e allo stesso tempo includere nel modello di regressione anche  $W$ . Per esempio, si supponga che il vero predittore  $X$  sia un qualche valore medio temporale o spaziale di un fattore di rischio che varia con il tempo o di un'esposizione che varia nello spazio, mentre  $W$  è una misura di un singolo punto nel tempo o nello spazio. In questi casi  $W$  non contribuisce aggiungendo maggiori informazioni di quelle già contenute in  $X$  e per questo può definirsi un surrogato.

In ogni caso, questo modo di ragionare non è infallibile. Anche il particolare modello a cui si fa riferimento influenza la definizione di  $W$  come surrogato. Per esempio si supponga che  $Z$  sia formata da due componenti,  $Z = (Z_1, Z_2)$ . E' possibile avere che  $f_{Y|Z_1Z_2XW} = f_{Y|Z_1Z_2X}$  e che  $f_{Y|Z_1XW} \neq f_{Y|Z_1X}$ . Quindi  $W$  è un surrogato nel modello completo che include  $Z_1$  e  $Z_2$  ma non lo è nel modello ridotto. In altre parole, il fatto che una variabile sia o no un surrogato dipende anche dalle altre variabili incluse nel modello di riferimento.

Con gli errori di misura non differenziali è possibile stimare i parametri del modello che collega la variabile risposta  $Y$  alle vere esplicative usando il predittore  $W$  affetto da errore solamente con una quantità minima di informazioni supplementari sulla distribuzione dell'errore di misura. Inoltre, non è necessario osservare la vera variabile  $X$ . Questo non si può fare con gli errori di misura differenziali ma è necessario avere un sottocampione con dati di validazione sia del valore misurato che del vero valore.

### 2.2.2 Conseguenze nella regressione lineare semplice

Per capire gli effetti dell'errore di misura sui risultati inferenziali si prenda come riferimento il seguente modello di regressione lineare semplice

$$Y = \beta_0 + \beta_x X + \varepsilon \quad (2.4)$$

dove la variabile scalare  $X$  ha media  $\mu_x$  e varianza  $\sigma_x^2$ . L'errore casuale  $\varepsilon$  si distribuisce con media 0 e varianza  $\sigma_\varepsilon^2$ , in simboli  $\varepsilon \sim (0; \sigma_\varepsilon^2)$ . Supponiamo inoltre che la variabile d'interesse  $X$  e la variabile  $W$  siano distribuite normalmente, con  $Var(X) = \sigma_x^2$ ,  $Var(W) = \sigma_w^2$ ,  $Cov(X, W) = \sigma_{xw}$ ,  $Cov(\varepsilon, X) = 0$  e  $Cov(\varepsilon, W) = \sigma_{\varepsilon w}$ . I dati disponibili sono le osservazioni provenienti da  $(Y, W)$ . Per stimare i parametri si utilizzano, come metodo di stima, i minimi quadrati ordinari. In assenza di errore di misura, se si considera la regressione di  $Y$  su  $W$ , lo stimatore per  $\beta_x$  risulta non distorto. In presenza di errore di misura invece questo stimatore in genere è distorto e la varianza residua aumenta. La distorsione dello stimatore e il comportamento della varianza residua dipendono molto dal tipo di errore di misura adottato. Fra breve, questa affermazione verrà maggiormente chiarita.

Per cominciare, si prenda il caso dell'errore di tipo classico (2.2). La variabile  $X$  non può essere osservata e invece quello che si osserva è  $W = X + U$  dove  $U$  è indipendente da  $X$ , con media 0 e varianza  $\sigma_u^2$ . Questo è il classico modello additivo per l'errore di misura dove lo stimatore dei minimi quadrati ordinari sulla regressione di  $Y$  su  $W$  non è consistente per  $\beta_x$  ma lo è per  $\beta_{x*} = \lambda\beta_x$ , dove

$$\lambda = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2} < 1. \quad (2.5)$$

Dunque, poichè  $|\beta_x|\sigma_x^2/(\sigma_x^2 + \sigma_u^2) \leq |\beta_x|$ , si conclude affermando che lo stimatore del coefficiente angolare  $\beta_x$  è sempre distorto verso lo zero. Questo comportamento è noto in letteratura con il nome di **attenuazione** verso lo zero e la quantità  $\lambda$  viene invece chiamata **fattore di attenuazione** o **quoziente di affidabilità** (*reliability ratio*). La varianza residua della regressione di  $Y$  su  $W$  è

$$Var(Y|W) = \sigma_\varepsilon^2 + \beta_x^2 \sigma_x^2 \frac{\sigma_u^2}{\sigma_x^2 + \sigma_u^2}. \quad (2.6)$$

Questo lato del problema, spesso ignorato, è invece molto importante. La presenza dell'errore di misura non causa solo una sottostima in termini assoluti del parametro relativo al coefficiente angolare ma apporta anche un notevole aumento della varianza residua, che non è pari a  $\sigma_\varepsilon^2$ .

Ora ipotizziamo per l'errore di misura il modello di Berkson. In questo caso lo stimatore del coefficiente angolare  $\beta_x$  è non distorto per  $\beta_x$ . La varianza residua invece è più grande di quella del vero modello ed è pari a

$$\text{Var}(Y|W) = \sigma_\varepsilon^2 + \beta_x^2 \sigma_x^2 \frac{\sigma_u^2}{\sigma_x^2}. \quad (2.7)$$

L'inferenza sui coefficienti di regressione del modello lineare è robusta con un modello di Berkson per l'errore di misura.

In generale, quando  $W$  è un surrogato di  $X$ , lo stimatore del coefficiente angolare  $\beta_x$  stima in modo non distorto la quantità  $\beta_x \sigma_{xw} / \sigma_w^2$  e non  $\beta_x$ . La stima potrebbe essere più grande o più piccola di  $\beta_x$  e non è possibile definire con precisione se c'è una sovrastima oppure una sottostima come nel caso del modello classico. Per quanto riguarda la varianza residua, essa è pari a:

$$\text{Var}(Y|W) = \sigma_\varepsilon^2 + \beta_x^2 \sigma_x^2 (1 - \rho_{xw}^2). \quad (2.8)$$

Quando  $W$  è un surrogato la varianza residua non è mai inferiore a quella del vero modello ma è sempre più grande. Si può notare che questa quantità, quando  $\rho_{xw}^2 = 1$  ossia quando  $X$  e  $W$  sono perfettamente correlati, si semplifica e viene a coincidere proprio con la varianza residua del vero modello, indicata con  $\sigma_\varepsilon^2$ . Segue che il coefficiente di determinazione del modello stimato con i minimi quadrati ordinari è sempre inferiore o al massimo uguale a quello del vero modello. Per recuperare il valore di  $\beta_x$  da un'analisi dei dati osservati, occorre che  $\sigma_{xw}$  sia nota. Questa quantità si può stimare attraverso uno studio di validazione dove ci si avvale delle osservazioni per le variabili  $(X, W)$ .

Infine, nei modelli con errori differenziali, lo stimatore del coefficiente angolare  $\beta_x$  stima in modo non distorto la quantità  $(\beta_x \sigma_{xw} + \sigma_{\varepsilon w}) / \sigma_w^2$  che può essere più grande o più piccola di  $\beta_x$  a seconda del valore assunto dalle due covarianze e dalla varianza di  $W$ . Conclusioni generali sulla distorsione non sono possibili. Per quanto riguarda la varianza residua, essa è pari a

$$\text{Var}(Y|W) = \sigma_\varepsilon^2 + \beta_x^2 \sigma_x^2 - \frac{(\sigma_{xw} \beta_x + \sigma_{\varepsilon w})^2}{\sigma_w^2}. \quad (2.9)$$

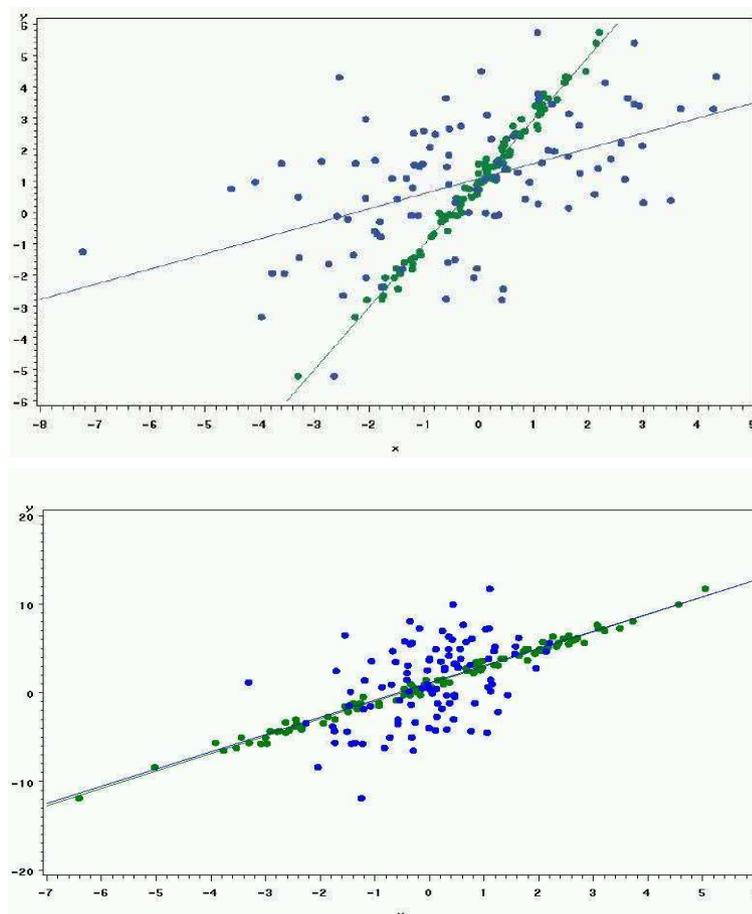


Figura 2.1: Confronto fra una regressione lineare semplice, in cui la variabile esplicativa  $X$  non è affetta da errore di misura e una regressione lineare semplice in cui la variabile esplicativa  $X$  è affetta da un errore di misura di tipo classico (grafico sopra) ed un errore di misura di tipo Berkson (grafico sotto). La variabile  $X$  affetta da errore di misura viene indicata con  $W$ . L'illustrazione si riferisce, in entrambi i casi, ad un modello additivo per l'errore di misura. I cerchi di colore verde fanno riferimento alle vere osservazioni del tipo  $(Y, X)$  e la linea passante per quei punti è la retta di regressione stimata con il metodo dei minimi quadrati ordinari. I cerchi blu e la linea passante per essi sono rispettivamente le osservazioni osservate delle variabili  $(Y, W)$  e la retta di regressione stimata con il metodo dei minimi quadrati ordinari.

In modo analogo, anche la varianza residua potrebbe essere più grande o più piccola rispetto a quella del vero modello.

Modello	Valore atteso di $\beta_x$	Varianza residua
Senza errore	$\beta_x$	$\sigma_\varepsilon^2$
Classico	$\beta_x \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2}$	$\sigma_\varepsilon^2 + \frac{\beta_x^2 \sigma_u^2 \sigma_x^2}{\sigma_x^2 + \sigma_u^2}$
Berkson	$\beta_x$	$\sigma_\varepsilon^2 + \beta_x^2 \sigma_x^2 \frac{\sigma_u^2}{\sigma_x^2}$
Surrogato	$\beta_x \frac{\sigma_{xw}}{\sigma_w^2}$	$\sigma_\varepsilon^2 + \beta_x^2 \sigma_x^2 (1 - \rho_{xw}^2)$
Differenziale	$\beta_x \frac{\sigma_{xw}}{\sigma_w^2} + \frac{\sigma_{\varepsilon w}}{\sigma_w^2}$	$\sigma_\varepsilon^2 + \beta_x^2 \sigma_x^2 - \frac{(\sigma_{xw} \beta_x + \sigma_{\varepsilon w})^2}{\sigma_w^2}$

Tabella 2.1: I valori attesi e le corrispondenti varianze residue relative al coefficiente angolare in un modello di regressione lineare del tipo (2.4), considerando diversi modelli per l'errore di misura.

### 2.2.3 Conseguenze nella regressione lineare multipla e nei modelli non lineari

I risultati riportati nella tabella e le conclusioni qualitative presentate nel caso di un modello di regressione lineare semplice possono essere generalizzate al caso di una regressione lineare multipla.

Se si ipotizza per l'errore di misura il modello di Berkson, sostituendo  $W$  al posto di  $X$ , non sussiste il problema della distorsione per gli stimatori dei parametri del modello. L'effetto principale, dovuto alla presenza dell'errore di misura, è rappresentato dall'aumento della varianza residua.

Per un modello di errore di misura classico è bene evidenziare due aspetti importanti, non presenti nel modello di regressione lineare semplice. Quando il modello include sia le variabili esplicative misurate con errore  $X$  e sia quelle misurate senza errore  $Z$ , la presenza dell'errore di misura può causare una distorsione anche nello stimatore per  $\beta_z$  oltre che sullo stimatore per  $\beta_x$ . Inoltre, l'attenuazione sul coefficiente di una variabile misurata con errore non è più una semplice funzione della varianza di questa variabile e della varianza dell'errore di misura. Quando sono presenti nel modello più variabili

esplicative misurate con errore, la distorsione negli stimatori dei coefficienti della regressione è rappresentata da una funzione non intuitiva che coinvolge la matrice di covarianza dell'errore di misura e la matrice di covarianza del vero predittore.

Gli effetti dell'errore di misura in un modello non lineare sono sostanzialmente gli stessi di un modello lineare, ovvero inducono distorsione negli stimatori dei parametri e riducono il potere di previsione del modello. Per quanto riguarda il modello per l'errore di misura di Berkson, se nei modelli di regressione lineare semplice e multipla non induce distorsione nelle stime dei parametri, nei modelli non lineari produce distorsione anche se questa è nettamente meno forte rispetto a quella causata dall'errore di misura classico.

### 2.2.4 Metodo dei momenti

Per quanto riguarda la regressione lineare semplice con un modello di tipo classico per l'errore di misura, nella (2.5) si è visto che i minimi quadrati ordinari forniscono una stima di  $\lambda\beta_x$ , dove  $\lambda$  è il fattore di attenuazione. Se  $\lambda$  fosse noto, si potrebbe ottenere una stima di  $\beta_x$  semplicemente dividendo  $\beta_{x*}$  per  $\lambda$ . Raramente però, nella pratica, il fattore di attenuazione è noto ma può essere stimato. Se  $\hat{\sigma}_u^2$  è una stima consistente della varianza dell'errore di misura e se  $\hat{\sigma}_w^2$  è la varianza del campione di  $W$ , allora una stima consistente per il fattore di attenuazione è  $\hat{\lambda} = (\hat{\sigma}_w^2 - \hat{\sigma}_u^2)/\hat{\sigma}_w^2$ . Quindi, la stima di  $\hat{\beta}_x = \beta_{x*}/\hat{\lambda}$ . L'algoritmo descritto è detto **metodo dei momenti**.

Sia  $\hat{\beta}_{x*}$  lo stimatore per il coefficiente angolare ottenuto con il metodo dei minimi quadrati della regressione di  $Y$  su  $W$ . La sua media è  $E[\hat{\beta}_{x*}] = \lambda\beta_x$  e la varianza è  $\sigma_*^2$ . Lo stimatore di  $\beta_x$  ricavato con il metodo dei momenti, se  $\lambda$  è noto, è  $\hat{\beta}_x^{MM} = \lambda^{-1}\hat{\beta}_{x*}$ , ha media  $E[\hat{\beta}_x^{MM}] = \beta_x$  e varianza  $\text{Var}[\hat{\beta}_x^{MM}] = \lambda^{-2}\sigma_*^2$ . E' chiaro che, dal momento che  $\lambda < 1$ , mentre la correzione per l'attenuazione in  $\hat{\beta}_x^{MM}$  riduce la distorsione a 0, si verifica un aumento della variabilità dovuto alla varianza dello stimatore distorto  $\hat{\beta}_{x*}$ . Se si vuole ridurre la distorsione il prezzo da pagare è un aumento della varianza. Nei casi in cui l'assenza della distorsione è un risultato di primaria importanza,

allora non si può evitare un incremento della varianza. Nei casi in cui la distorsione può essere tollerata allora è necessario prendere in considerazione l'errore quadratico medio. Gli stimatori non corretti sono dotati di una maggior distorsione ma di una varianza più piccola rispetto agli stimatori corretti. Infatti si noti che

$$EQM(\hat{\beta}_{x*}) = \sigma_*^2 + (1 - \lambda)^2 \beta_x^2 \quad (2.10)$$

$$EQM(\hat{\beta}_x^{MM}) = \lambda^{-2} \sigma_x^2. \quad (2.11)$$

Segue che

$$EQM(\hat{\beta}_x^{MM}) < EQM(\hat{\beta}_{x*}) \quad (2.12)$$

se e solo se

$$\sigma_*^2 < \frac{\lambda^2(1 - \lambda)\beta_x^2}{1 + \lambda} \quad (2.13)$$

Poiché  $\sigma_*^2$  diminuisce con l'aumentare della numerosità campionaria si può concludere che in campioni sufficientemente ampi è sempre vantaggioso, in termini di errore quadratico medio, correggere l'attenuazione causata dall'errore di misura.

Ricapitolando, le stime dei parametri risultano distorte se non si tiene conto della presenza dell'errore di misura. Volendo però risolvere questo problema, nasce un compromesso fra distorsione e varianza. Infatti, quello che succede correggendo la distorsione, è che lo stimatore corretto ottenuto è più variabile di quello distorto. E ovviamente, quando uno stimatore ha una varianza più elevata, gli intervalli di confidenza associati sono più ampi. Questa questione viene portata alla luce facendo riferimento ad una regressione lineare semplice del tipo (2.4) ed ad un modello per l'errore di misura classico, del tipo (2.2) ma è un problema molto generale. Si suppone inoltre che il fattore di attenuazione  $\lambda$  nella (2.5) sia noto. Raramente lo è nella pratica ma questa assunzione viene fatta per semplificare il ragionamento.

## 2.3 Metodi di correzione

In questo paragrafo verranno presentati vari metodi proposti in letteratura per porre rimedio al problema della presenza dell'errore di misura in una o più variabili. Per ciascun metodo verrà richiamata la teoria sottostante. L'ordine di presentazione seguirà quello proposto da Carroll, Ruppert e Stefanski in [15]:

- regressione di calibrazione;
- variabili strumentali;
- simulazione-estrapolazione;
- funzioni di punteggio corrette o condizionate;
- verosimiglianza;
- metodi semi parametrici;
- approccio bayesiano.

I metodi possono essere di tipo funzionale o strutturale. Un **metodo funzionale** non fa assunzioni sulla variabile non osservata  $X$  ma la modellazione avviene attraverso costanti fisse e non note. Spesso si è costretti a ricorrere ad una modellazione di tipo funzionale poiché ci sono poche informazioni sulla distribuzione di  $X$  nei dati. I metodi funzionali si dividono fra quelli approssimativamente consistenti dove viene rimossa una parte della distorsione degli stimatori e quelli completamente consistenti che eliminano tutta la distorsione presente. Tipicamente, i metodi completamente consistenti alla base dei modelli di regressione non lineari richiedono assunzioni sulla distribuzione dell'errore di misura. La regressione di calibrazione e il SIMEX sono esempi di metodi funzionali approssimativamente consistenti mentre i metodi delle funzioni di punteggio corrette o condizionate e alcuni metodi che usano le variabili strumentali sono completamente consistenti per un'ampia classe di modelli.

Al contrario, in un metodo strutturale la  $X$  è una variabile casuale che deve essere descritta dal modello per l'esposizione. I metodi basati sulla verosimiglianza sono strutturali.

La scelta tra un modello funzionale o strutturale dipende tipicamente dalle assunzioni fatte e in particolare dal modello per la variabile d'interesse, che mette in relazione la variabile  $Y$  con la variabile esplicativa, non nota,  $X$  e con le altre variabili esplicative  $Z$  non affette da errore.

Si noti che i termini funzionale e strutturale fanno riferimento alle assunzioni sulla  $X$  non sul modello per l'errore di misura. Il vantaggio di un metodo funzionale è quello di fornire una valida inferenza anche senza occuparsi della distribuzione della  $X$ .

Quando si lavora in presenza di errore di misura, una difficoltà che deve essere affrontata è la mancanza di informazioni supplementari al fine di evitare che non venga soddisfatta la condizione per l'identificabilità dei parametri. Il problema dell'identificazione riguarda la possibilità di individuare univocamente i parametri incogniti partendo dalle quantità note riferite all'intera popolazione. Se ciò non succede bisogna ricorrere ad informazioni aggiuntive per ottenere stime consistenti dei parametri d'interesse. Possono essere disponibili in una di queste due forme:

- interni, ossia sottoinsiemi dei dati primari;
- esterni, ossia ricavati da altri studi.

Per ciascuna di queste categorie, possiamo identificare tre tipi di dati:

- dati di validazione, se è disponibile un'osservazione diretta della variabile  $X$ ; questa osservazione viene chiamata misura *gold standard* di  $X$ ;
- misure ripetute, quando due o più osservazioni di  $W$  sono effettuate sulle medesime unità statistiche;
- variabili strumentali, quando si osserva anche un'altra variabile  $V$ , oltre a  $W$ .

Esistono due tipologie di metodi che consentono di superare il problema dell'errore di misura e migliorare l'inferenza, avendo a disposizione osservazioni sulle variabili  $(Y, Z, W)$ . Infatti, è utile fare una distinzione tra approccio statistico ed approccio econometrico. La distinzione si basa sul tipo di informazioni supplementari che vengono utilizzate.

Infatti, i metodi statistici di correzione, sfruttano poca informazione esterna per concentrarsi direttamente sull'errore e sulle sue caratteristiche. Per esempio, si deve avere un'idea del valore da assegnare a  $\sigma_u^2$ . E questa informazione la si può ricavare da studi passati oppure basandosi sulle aspettative che hanno le persone. Un metodo statistico ha bisogno di informazioni supplementari relative proprio all'errore e ricaverà queste informazioni attraverso dati di validazione oppure misure ripetute.

In econometria invece ci si chiede cosa succede a fare inferenza quando tutta l'informazione a disposizione è sulle variabili  $(Y, Z, W)$ . L'approccio al problema, utilizzato in econometria, è quello di usare le variabili strumentali. In altre parole, si sfrutta una seconda misura, affetta da errore, usata come strumento per la prima misura, affetta da errore. Ciò che comunque nei metodi econometrici non si fa è quello di intervenire direttamente sull'errore di misura.

### 2.3.1 La regressione di calibrazione

Il metodo della regressione di calibrazione (*regression calibration*), RC per abbreviare, è uno dei metodi più diffusi per correggere la presenza dell'errore di misura e viene applicato con successo a molti modelli di regressione. Questo metodo funzionale, per la sua semplicità e per la facile implementazione grazie a pacchetti già esistenti in R, è ampiamente utilizzato.

L'algoritmo alla base di questo metodo punta ad approssimare il modello che lega la  $Y$  con  $(Z, X)$  considerando la regressione lineare della vera variabile  $X$  sulla variabile surrogato  $W$  e sulle variabili non affette da errore  $Z$ . Per descrivere questo metodo è utile immaginare un approccio che attribuisce valori alla variabile non osservata  $X$ . In effetti l'idea che caratterizza questo

metodo è quella di stimare la  $X$  con  $X^*$ , ovvero con il valore previsto di  $X$  della regressione di  $X$  su  $(Z, W)$ . Per modellare e stimare la regressione di  $X$  su  $(Z, W)$  sono necessarie informazioni supplementari come dati di validazione, interni o esterni, osservazioni replicate o variabili strumentali. I parametri nel vero modello per la variabile d'interesse vengono stimati regredendo la  $Y$  su  $(Z, X^*)$ . Si noti che la  $X^*$  è la miglior stima di  $X$  usando le variabili esplicative osservate  $(Z, W)$ , ossia è la stima che meglio riesce a minimizzare l'errore quadratico medio di previsione. Questa approssimazione del vero modello per la variabile d'interesse viene poi usata per effettuare un'analisi e per ottenere le stime dei parametri. Infine gli errori standard di queste stime dovranno essere aggiustati dal momento che la  $X$  viene stimata. Solitamente si ricorre ai metodi bootstrap, jackknife o sandwich. Il metodo bootstrap richiede meno programmazione ma il calcolatore impiega molto tempo ad elaborare il risultato e questo può essere un vero problema dal momento che è realistico pensare che i ricercatori non siano propensi a volerlo utilizzare facilmente nelle loro applicazioni. Una spiegazione dei metodi bootstrap, jackknife e sandwich è riportata in Appendice.

Per riassumere, la RC si sviluppa attraverso questi due passi:

1. modellare e stimare la regressione di  $X$  su  $(Z, W)$  per ottenere  $X^*$ ;
2. regredire la  $Y$  su  $(Z, X^*)$  per ottenere le stime dei parametri della regressione.

Per alcuni modelli, come per esempio i modelli loglineari e la regressione lineare, quando la varianza di  $X$  dati  $(Z, W)$  è costante, allora l'approssimazione della RC, che lega  $X$  a  $(Z, W)$ , è esatta, fatta eccezione per un cambiamento nell'intercetta del parametro. Questo metodo dunque risulta completamente consistente. Nei modelli non lineari come la regressione logistica è approssimativamente consistente. L'approssimazione può essere svolta usando uno sviluppo in serie di Taylor, assumendo che la varianza dell'errore di misura sia piccola. In ogni caso, prendendo dei termini aggiuntivi nella serie di Taylor si possono ottenere approssimazioni migliori.

Per applicare il metodo della RC devono essere soddisfatte alcune assunzioni. In particolare si richiede che:

- il modello che lega  $X$  a  $W$  sia di tipo lineare omoschedastico;
- gli errori di misura siano non differenziali;
- la varianza dell'errore di misura sia piccola.

Quando saltano una o più fra queste assunzioni, la RC può risultare inefficiente per ridurre la distorsione, specialmente nei modelli non lineari. In particolare, l'ipotesi che l'errore di misura sia non differenziale raramente, nelle applicazioni, è soddisfatta.

Come si è già anticipato prima, il metodo della RC richiede informazioni aggiuntive per essere in grado di predire le osservazioni per la variabile non nota  $X$  dalle osservazioni di  $W$  affette da errore. L'informazione supplementare ideale è rappresentata dalle misure esatte per  $X$  (*gold standard*) in un sottoinsieme dei dati primari.

### 2.3.2 Variabili strumentali

Si prenda in considerazione il modello per l'errore di misura classico, ossia  $W = X + U$ ,  $U \sim N(0, \sigma_u^2)$ , con  $\sigma_u^2$  nota. Una **variabile strumentale**  $V$  è una misura supplementare di  $X$  che soddisfa i seguenti tre requisiti:

1. è correlata con  $X$ ;
2. non è correlata con il termine d'errore  $U = W - X$ ;
3. è un surrogato di  $X$ , nel senso che  $f_{Y|XZV}(y|x, z, v) = f_{Y|XZ}(y|x, z)$ .

Per capire meglio l'idea principale del metodo di stima che usa le variabili strumentali si prenda in esame questo semplice esempio. Si consideri un modello lineare semplice senza intercetta:  $Y = \beta_x X + \varepsilon$  e  $W = X + U$ . Non potendo osservare  $X$ , si deve lavorare con  $W$  e il modello  $Y = \beta_x W + \tilde{\varepsilon}$ . Ma  $W$  e  $\tilde{\varepsilon}$  sono correlati e questo va a violare una delle assunzioni fondamentali del

modello di regressione lineare causando così la distorsione dello stimatore dei minimi quadrati ordinari di  $\beta_x$ . L'equazione di stima  $\sum_{i=1}^n (Y_i - \beta_x W_i) W_i = 0$  è distorta perché  $W_i$  e  $Y_i - \beta_x W_i$  sono correlati. Si può costruire un'equazione non distorta sostituendo  $W_i$ , fuori dalle parentesi, con una quantità incorrelata con  $Y_i - \beta_x W_i$ . Una variabile strumentale  $V$  soddisfa questo requisito e l'equazione di stima  $\sum_{i=1}^n (Y_i - \beta_x W_i) V_i = 0$  fornisce lo stimatore consistente

$$\hat{\beta}_x = \frac{\sum_{i=1}^n Y_i V_i}{\sum_{i=1}^n W_i V_i}. \quad (2.14)$$

Per quanto riguarda i modelli lineari, la stima con le variabili strumentali è ampiamente illustrata da Fuller (si veda [14] per maggiori dettagli). Per i modelli non lineari sono stati proposti vari approcci. Tra i più comuni si può menzionare quello presentato da Carroll e Stefanski, collegato alla RC (si veda [16] per maggiori dettagli). Approcci differenti si possono trovare in Amemiya per i modelli non lineari (si veda [17], [18], [19] per maggiori dettagli) e in Buzas e Stefanski per i modelli lineari generalizzati nella forma canonica (si veda [20] per maggiori dettagli).

### 2.3.3 SIMEX

La RC è un metodo semplice per correggere la presenza degli errori di misura quando sono disponibili dati di validazione o di replicazione per modellare la relazione fra  $X$  e  $(Z, W)$ . Questo approccio però richiede che alcune assunzioni siano soddisfatte. Per contro il metodo della **simulazione-estrapolazione** (SIMEX, per abbreviare) non necessita di ipotesi sulla distribuzione delle variabili. È un metodo funzionale che si basa sulla simulazione per ridurre la distorsione dovuta all'errore di misura ed è consistente per il modello lineare classico e approssimativamente consistente per i modelli non lineari. Condivide con la RC la semplicità di applicazione e si presta bene in quei casi dove si ipotizza, per l'errore di misura, una struttura di tipo additivo. È l'unico metodo in grado di fornire una rappresentazione concreta e visiva degli effetti che l'errore di misura ha sulle stime dei parametri.

L'intuizione sottostante a questo procedimento è che l'effetto dell'errore di misura su uno stimatore possa essere determinato sperimentalmente attraverso la simulazione. In effetti la distorsione varia in modo sistematico a seconda del peso assunto dall'errore di misura. Le stime SIMEX sono ottenute mediante una procedura suddivisa in due passaggi. Nel primo passo viene aggiunto in maniera progressiva errore alle misure e vengono stimati i parametri della regressione corrispondente. Questa prima fase viene chiamata *simulazione* (*simulation step*). Successivamente viene modellata la relazione tra le stime dei parametri e la varianza degli errori di misura. In pratica si inserisce errore di misura nei dati, si cerca di coglierne l'andamento e infine si rimuove per ottenere le stime SIMEX dei parametri. Questo secondo passaggio viene chiamato *estrapolazione* (*extrapolation step*). La tecnica è stata proposta da Cook e Stefanski (si veda [21] per maggiori dettagli). Più tardi è stata sviluppata da Carroll, Küchenhoff, Lombard e Stefanski (si veda [22] per maggiori dettagli) e da Stefanski e Cook (si veda [23] per maggiori dettagli).

Le caratteristiche specifiche dell'algoritmo SIMEX verranno descritte in questa sezione prendendo come esempio la regressione lineare semplice con errore di misura classico additivo. Si suppone che  $Y = \beta_0 + \beta_x X + \varepsilon$ , con  $W = X + U$ , dove  $U$  è una variabile normale, indipendente da  $(Y, X)$ , con media zero e varianza  $\sigma_u^2$ . Nei casi in cui  $\sigma_u$  non sia nota viene stimata usando informazioni supplementari. La stima con i minimi quadrati ordinari di  $\beta_x$ , indicata con  $\hat{\beta}_x^{naive}$ , non stima in modo consistente  $\beta_x$  ma piuttosto la quantità  $\beta_x \sigma_x^2 / (\sigma_x^2 + \sigma_u^2)$  come già spiegato precedentemente. Inoltre, per questo semplice modello, l'effetto dell'errore di misura sullo stimatore dei minimi quadrati ordinari è facilmente determinato matematicamente.

Se guardiamo all'errore di misura come ad un fattore di cui si vuole conoscere l'influenza che ha sullo stimatore, si pensa subito a ricorrere alla simulazione dove si possono variare o tenere sotto controllo le quantità desiderate, come per esempio la varianza dell'errore di misura. In aggiunta ai dati originali usati per calcolare  $\hat{\beta}_x^{naive}$ , ci sono anche  $M - 1$  insiemi di

dati, generati dall'algoritmo SIMEX, ciascuno con una varianza dell'errore di misura più ampia del precedente, pari a  $(1 + \lambda_j)\sigma_u^2$ , con  $j = 1, \dots, M$ , dove  $0 = \lambda_1 < \lambda_2 < \dots < \lambda_M$ . La stima *naive* del coefficiente angolare ottenuta dal  $j$ -mo insieme di dati,  $\hat{\beta}_{x,j}$ , stima in modo consistente la quantità  $\beta_x \sigma_x^2 / [\sigma_x^2 + (1 + \lambda_j)\sigma_u^2]$ . Per ciascun  $\lambda_j \geq 0$ , si determinano

$$W_b(\lambda_j) = W + \sqrt{\lambda_j} U_b, b = 1, \dots, B, \quad (2.15)$$

dove  $[U_b]_{b=1}^B$  sono  $B$  pseudo errori generati dal calcolatore, indipendenti fra loro e con tutte le osservazioni identicamente distribuite come variabili casuali normali, ciascuna con media 0 e varianza  $\sigma_u^2$ . La regola pratica suggerisce che  $B = 100$  per un lavoro accurato. Per ciascun insieme di dati, si calcola la stima *naive*  $\hat{\beta}_x^b$  di  $\beta_x$ . Il valore medio delle  $B$  stime *naive*, indicato con  $\hat{\beta}(\lambda_j)$  è calcolato come

$$\hat{\beta}(\lambda_j) = \frac{\sum_{b=1}^B \hat{\beta}_{bx}(\lambda_j)}{B} \quad (2.16)$$

Durante la fase di estrapolazione viene modellata  $[\hat{\beta}(\lambda_j), \lambda_j]_{j=1}^M$  come una funzione di  $\lambda$ . Lo stimatore SIMEX è l'estrapolazione del modello per  $\lambda = -1$ . Lo stimatore *naive* è invece ottenuto per  $\lambda = 0$ .

### 2.3.4 Funzioni di punteggio corrette o condizionate

I metodi RC e SIMEX sono facilmente applicabili e risultano consistenti se vengono utilizzati nel caso della regressione lineare e nei modelli log-lineari. Sono invece approssimativamente consistenti negli altri casi. Esistono altri metodi funzionali, consistenti per alcuni tipi di modelli lineari generalizzati e per alcune strutture dell'errore di misura. Si tratta dei metodi con **funzione di punteggio corretta o condizionata** (*corrected and conditional scores*).

Questi metodi sono stati sviluppati partendo dalle equazioni di stima per i parametri della regressione, nel caso di assenza di errore di misura. Un'equazione di stima è non distorta se la sua media è pari a zero. Un esempio è l'equazione di verosimiglianza basata sulla derivata prima della funzione di log verosimiglianza, fatta rispetto ai parametri. L'errore di misura provoca

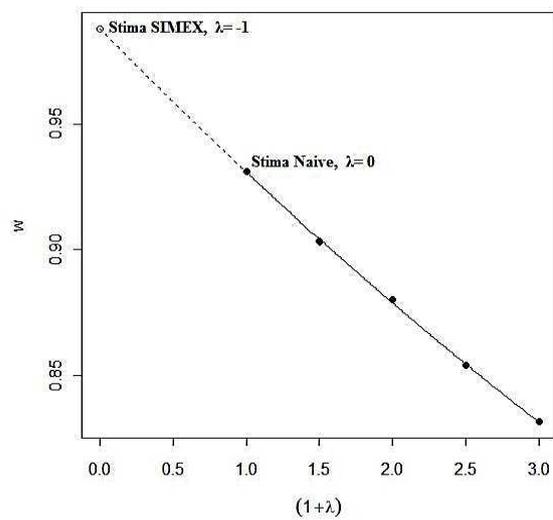


Figura 2.2: Il grafico rappresenta il principio di correzione sottostante al metodo SIMEX. Il valore di  $(1+\lambda)$  è riportato sull'asse delle ascisse, mentre il valore del coefficiente stimato è riportato sull'asse delle ordinate. La stima SIMEX è un'estrapolazione in corrispondenza di  $\lambda = -1$ . La stima naive si trova in corrispondenza di  $\lambda = 0$ .

distorsione nelle equazioni di stima che a loro volta inducono inconsistenza negli stimatori dei parametri. In effetti, il **metodo della funzione punteggio corretta** fornisce uno stimatore consistente, simile a quello che si andrebbe ad utilizzare nel caso di assenza di errore di misura. Tale metodo è stato studiato da Stefanski (si veda [24] per maggiori dettagli) e da Nakamura (si veda [25] per maggiori dettagli).

Il **metodo della funzione punteggio conditionata** è stato sviluppato da Stefanski e Carroll (si veda [26] per maggiori dettagli) e successivamente raffinato da Stefanski e Carroll, adattandolo al caso di modelli non lineari (si veda [27] per maggiori dettagli). Si tratta di un metodo funzionale e si basa sulla teoria delle statistiche sufficienti. Condizionandosi a queste, i parametri di disturbo possono essere eliminati. Si deve assumere che gli errori di misura si distribuiscano normalmente.

### 2.3.5 Verosimiglianza

I metodi che usano la **verosimiglianza** per trattare gli errori di misura sono strutturali. Rispetto ad altri metodi, come per esempio la RC, hanno ricevuto poca attenzione in letteratura a causa della loro complessità in termini computazionali e della difficoltà a controllare la validità delle assunzioni che essi richiedono. Recentemente invece alcuni risultati hanno portato alla luce i vantaggi dei metodi di verosimiglianza, più di tutto le proprietà di ottimalità degli stimatori corrispondenti. Si veda [28] e [29] per maggiori dettagli.

Qui sotto vengono riportate alcune significative differenze fra l'approccio basato sulla verosimiglianza e i metodi funzionali precedentemente descritti.

- Ci sono metodi funzionali che sono completamente consistenti e non solo approssimativamente consistenti senza dover fare assunzioni sulla distribuzione della vera variabile d'interesse  $X$ . Al contrario, i metodi basati sulla verosimiglianza, richiedono forti assunzioni sulle distribuzioni delle variabili coinvolte ma possono essere applicati a problemi più generali.

- I metodi basati sulla verosimiglianza consentono di sviluppare procedure inferenziali, per calcolare intervalli di confidenza e test, molto più attendibili rispetto a quelle solitamente adottate nei metodi funzionali, dove l'inferenza si affida a ricampionamento e ad approssimazioni normali.
- I metodi di verosimiglianza sono molto più esigenti dal punto di vista computazionale, al contrario dei metodi fin qui descritti che presuppongono l'utilizzo di pacchetti statistici standard.
- La selezione del modello e l'analisi di robustezza sono generalmente problemi più rilevanti nei metodi basati sulla verosimiglianza.
- In un contesto di regressione lineare segmentata, nelle simulazioni dove le variabili  $X$  e  $W$  sono distribuite normalmente, si evince che le stime di massima verosimiglianza sono tipicamente molto più efficienti rispetto a stime ottenute con il metodo SIMEX (si veda [30]). Assumendo inoltre che la  $X$  sia una mistura di distribuzioni normali si aggiungono informazioni considerevoli al problema. La verosimiglianza ne trae vantaggio dal momento che lo stimatore corrispondente mostra una varianza più piccola.

L'analisi basata sulla verosimiglianza richiede che tutti i componenti vengano completamente descritti da un modello parametrico.

1. Il modello che lega il risultato  $Y$  alla vera variabile  $X$ , alla variabile  $W$  affetta da errore e alle variabili  $Z$ , ossia il modello per la variabile d'interesse, deve essere specificato

$$f_{Y|XZW}(y|x, z, w; \theta_1). \quad (2.17)$$

Il modello (2.17) coincide con  $f_{Y|XZ}(y|x, z; \theta_1)$  quando si assume che l'errore di misura sia di tipo non differenziale, come si assumerà d'ora in poi. La distribuzione normale e quella logistica sono solitamente adottati per il modello (2.17).

2. Il modello che collega la variabile  $W$  affetta da errore con la vera esposizione  $X$  e le variabili  $Z$ , deve essere specificato e viene chiamato modello per l'errore di misura

$$f_{W|XZ}(w|x, z; \theta_2). \quad (2.18)$$

3. Infine, si ha bisogno di specificare il modello per la vera esposizione  $X$ , chiamato appunto modello per l'esposizione

$$f_{X|Z}(x|z; \theta_3). \quad (2.19)$$

La verosimiglianza per i dati osservati da  $(Y, W, Z)$  è ottenuta mediante integrazione fatta rispetto alle vere ma ignote variabili  $X$ . Perciò, la verosimiglianza integrata è data dalla seguente espressione

$$L(\theta; y, w, z) = \int f_{Y|XZ}(y|x, z; \theta_1) f_{W|XZ}(w|x, z; \theta_2) f_{X|Z}(x|z; \theta_3) dx, \quad (2.20)$$

dove  $\theta = (\theta_1, \theta_2, \theta_3)$ . L'integrale deve essere sostituito dalla sommatoria se  $X$  è una variabile discreta.

Se è disponibile l'informazione supplementare su  $X$ , questa può essere usata per stimare  $\theta_2$  e  $\theta_3$ . Queste stime possono poi essere sostituite al posto dei parametri corrispondenti nella (2.20) e la pseudo verosimiglianza ottenuta può così essere massimizzata rispetto ai restanti parametri  $\theta_1$ .

Finora è stato assunto che l'errore di misura fosse di tipo non differenziale. Se invece si ipotizza un errore di Berkson, allora il modello per l'errore di misura (2.18) cambia e così pure la verosimiglianza. Dopo aver specificato il modello per la variabile d'interesse (2.17), occorre determinare anche il modello che lega la  $X$  alla  $W$ ,  $f_{X|WZ}(x|w, z; \theta_4)$  e la distribuzione di  $W$ ,  $f_{W|Z}(w|z; \theta_5)$ . La funzione di verosimiglianza per  $\theta = (\theta_1, \theta_4, \theta_5)$  è pari a

$$L(\theta; y, w, z) = \int f_{Y|XZ}(y|x, z; \theta_1) f_{X|WZ}(x|w, z; \theta_4) f_{W|Z}(w|z; \theta_5) dx, \quad (2.21)$$

che può essere semplificata nella

$$L(\theta; y, w, z) = \int f_{Y|XZ}(y|x, z; \theta_1) f_{X|WZ}(x|w, z; \theta_4) dx, \quad (2.22)$$

se si considera che la  $f_{W|Z}(w|z; \theta_5)$  non apporta informazioni sul parametro d'interesse  $\theta_1$  e non dipende da  $X$ . Come prima, l'integrale verrà sostituito dalla sommatoria nel caso in cui la variabile casuale  $X$  sia discreta.

Per soddisfare la condizione di identificazione dei parametri non basta sfruttare le informazioni fornite da  $(Y, Z, W)$  ma è necessario procurare informazioni supplementari. Alcuni parametri possono essere assunti noti perchè il valore viene fornito da studi precedenti, altri possono essere stimati ricorrendo ad informazioni supplementari, come per esempio dati di validazione, sia interni che esterni.

In alcuni casi, invece, i parametri sono identificabili senza dover ricorrere ad informazioni supplementari, fatta eccezione per quelle fornite da  $(Y, Z, W)$ . Per esempio, nel caso della regressione lineare è noto che, se la variabile  $X$  non è distribuita normalmente, i parametri della regressione possono essere identificati anche senza osservazioni ripetute o dati di validazione, (si veda [14] per maggiori dettagli).

Ciò significa che le stime dei parametri saranno molto instabili se la  $X$  è vicina all'essere distribuita come una normale. In una regressione binaria, avente una  $X$  distribuita normalmente, è noto che il modello probit non è identificato senza informazioni supplementari mentre un modello logistico lo è, si veda Carroll *et al.* ([15]).

Possono sorgere alcuni problemi di tipo computazionale quando si tratta di valutare le espressioni (2.20) o il (2.21). Se la variabile vera, non nota,  $X$  è continua, sia nel caso in cui l'errore sia classico o di Berkson, il calcolo della funzione di verosimiglianza necessita la valutazione di un integrale e solitamente si utilizzano metodi numerici o approssimazioni analitiche. Ci si riferisce al metodo Monte Carlo, presentato da Carroll *et al.* (si veda [15], paragrafo 7.9.1). Inoltre, si trovano esempi di modelli di regressione lineare, logistica e probit con una distribuzione normale per  $(X, Z)$ , si veda Carroll *et al.* (si veda [15], paragrafo 7.9.2).

### 2.3.6 Metodi semi parametrici

Gli stimatori parametrici possono essere più efficienti rispetto a quelli basati su assunzioni più deboli. Non sono utilizzati spesso a causa dei problemi computazionali e il rischio di una possibile non corretta specificazione della distribuzione delle variabili affette da errore. In questo senso, l'approccio semi parametrico rappresenta una valida alternativa dal momento che non richiede che vengano specificate la distribuzione della variabile  $X$  e quella delle variabili surrogato  $W$ . Queste distribuzioni infatti vengono considerate e trattate come parametri da stimare.

In letteratura sono state proposte diverse tecniche funzionali semi parametriche allo scopo di modellare l'errore di misura. Possono essere identificati tre gruppi:

1. metodi che si focalizzano sulla stima non parametrica della distribuzione marginale o condizionata di  $X$  e  $W$ ;
2. metodi che puntano sulla stima non parametrica dei momenti per la distribuzione condizionata di  $X$  o  $W$ ;
3. metodi che utilizzano la stima non parametrica della funzione di verosimiglianza empirica.

Si riportano alcuni esempi per ciascun gruppo.

Uno tra i più famosi approcci semi parametrici appartiene al primo gruppo ed è stato proposto da Carroll *et al.* (si veda [31] per maggiori dettagli). Gli autori prendono in considerazione una regressione logistica in uno studio caso-controllo con un sottocampione di osservazioni supplementari di validazione per le variabili  $X$  e  $W$ . Non vengono fatte restrizioni sulla distribuzione degli errori di misura. Questo approccio semi parametrico si concentra su una formulazione parametrica del modello per la variabile d'interesse e del modello per l'errore di misura. La distribuzione marginale della vera variabile  $X$  viene stimata in modo non parametrico attraverso una media pesata delle distribuzioni empiriche di  $X|Y = y$  ottenute dai dati completi, ossia quando i dati per  $(Y, Z, X)$  sono tutti osservati.

Roeder *et al.* (si veda [32]) propongono un metodo molto simile a quello presentato da Carroll *et al.* ([31]) considerando sempre un modello logistico in uno studio caso-controllo. Questo metodo è collocabile all'interno del primo gruppo. La funzione di distribuzione empirica di  $X$  è usata come una prima stima della distribuzione marginale di  $X$ . All'interno del processo di stima dei parametri del modello per la variabile d'interesse la suddetta stima viene aggiornata attraverso un algoritmo EM o il metodo del gradiente. In termini di consistenza e di errori standard degli stimatori, studi di simulazione hanno dimostrato che questo metodo funziona meglio di quello proposto da Carroll *et al.* ([31]), sotto due condizioni:

1. la varianza dell'errore di misura è piccola;
2. dati di validazione, con almeno 1/8 dei dati completo.

Un esempio di approccio semi parametrico, appartenente al secondo gruppo, è quello proposto da Carroll e Stefanski (si veda [33]) e viene applicato a dati caso-controllo che sono affetti da errore di misura non differenziale con una varianza piccola. Sono richiesti dati di validazione. Gli autori si servono della media e della varianza, in particolare di  $E[Y|X]$  e di  $V[Y|X]$  per fare inferenza. Non vengono fatte assunzioni sulla distribuzione di  $X$  dato  $W$  ma solamente sui primi due momenti dell'errore di misura dato  $W$ . Gli autori propongono stimatori di tipo  $M$  per i parametri coinvolti, basati su sviluppi in serie di Taylor delle funzioni di media e varianza.

Infine, un metodo appartenente al terzo gruppo è quello proposto da Schafer (si veda [34]) che suggerisce un algoritmo EM per una stima semi parametrica della verosimiglianza in modelli di regressioni lineari, lineari generalizzati e non lineari. Si specifica un modello strutturale per la variabile d'interesse e per il modello dell'errore di misura mentre la distribuzione di  $X$ , lasciata non specificata, viene stimata da una massima verosimiglianza non parametrica. Questo metodo viene applicato in situazioni dove vengono usati diversi tipi di informazioni supplementari per definire la distribuzione dell'errore di misura.

### 2.3.7 Approccio bayesiano

La caratteristica distintiva dell'approccio bayesiano è quella di trattare tutti i parametri come variabili casuali. I metodi bayesiani per la correzione dell'errore di misura si basano sulla funzione di verosimiglianza, interpretata come densità condizionata delle osservazioni, dati i parametri. L'analisi bayesiana esige una distribuzione a priori per i parametri, che racchiuda tutta l'informazione disponibile inizialmente. Anche le variabili  $X$ , non essendo note, vengono trattate come parametri ignoti. Il prodotto fra la priori e la funzione di verosimiglianza fornisce la densità congiunta dei dati e dei parametri. Spesso vengono impiegate priori non informative.

Sia  $\pi(\theta_1, \theta_2, \theta_3)$  la densità a priori dei parametri  $\theta_1, \theta_2, \theta_3$ . Inoltre, si definisce il modello per la variabile d'interesse, il modello per l'errore di misura e quello per l'esposizione come nelle (2.17), (2.18), (2.19). Perciò, tralasciando i pedici per semplificare le scritte, si ottiene la densità congiunta dei dati per  $(Y, W)$  e per i parametri  $(\theta_1, \theta_2, \theta_3, X)$  come

$$f(y, w, x, \theta_1, \theta_2, \theta_3|z) = f(y, x, w|z; \theta_1)f(w|x, z; \theta_2)f(x|z; \theta_3)\pi(\theta_1, \theta_2, \theta_3). \quad (2.23)$$

Ottenuta la densità congiunta, si può ricavare la densità marginale dei dati, integrando rispetto ai parametri. A questo punto, dividendo la densità congiunta per la densità marginale si riesce ad ottenere la distribuzione a posteriori

$$\pi(x, \theta_1, \theta_2, \theta_3|y, w, z) = \frac{f(y, w, x, \theta_1, \theta_2, \theta_3|z)}{\int f(y, w, x, \theta_1, \theta_2, \theta_3|z) dx d\theta_1 d\theta_2 d\theta_3}. \quad (2.24)$$

Essa sintetizza tutte le informazioni disponibili sui parametri sperimentali e sulle variabili esogene e rappresenta la base per l'inferenza bayesiana.

Spesso calcolare la distribuzione a posteriori non è affatto un problema banale a causa dell'integrale al denominatore. Questo è uno dei motivi principali per cui i metodi bayesiani non sono stati ampiamente sfruttati in letteratura. Le tecniche computazionali per valutare gli integrali si basano solitamente su metodi MCMC, come fa per esempio la tecnica di ricampionamento di Gibbs (*Gibbs sampler*), un metodo che in letteratura ha ricevuto

molta attenzione. Si veda Robert e Casella ([35]) per una spiegazione approfondita sul metodo di Gibbs sampler. Si veda anche Tanner ([36]) per una visione ampia sulle metodologie per calcolare le distribuzioni a posteriori.

Per una spiegazione dettagliata sulle tecniche adottabili in ambito bayesiano per correggere l'errore di misura si rimanda a Gustafson ([37]). Inoltre Richardson e Gilks ([38]) propongono di usare il metodo Gibbs per quanto riguarda i modelli lineari generalizzati al fine di ottenere la distribuzione marginale dei parametri in un contesto Bayesiano quando sono disponibili dati di validazione per stimare il modello di esposizione. Richardson e Leblond ([39]) chiariscono che, se la distribuzione di  $X$  è mal specificata, generalmente questo crea negli stimatori una distorsione moderata ma un aumento degli errori standard a posteriori.



# Capitolo 3

## Studio di simulazione

### 3.1 Descrizione dello studio

Abbiamo condotto uno studio di simulazione per analizzare il comportamento degli stimatori per i parametri del modello principale.

Nel lavoro di Battistin *et al.* [5], gli autori vogliono stimare la relazione fra il logaritmo del consumo e scelta di pensionamento  $R$ . Il modello dal quale partono è un modello del tipo:

$$\log C = \delta_0 + \delta_1 R + \text{errore} \quad (3.1)$$

e usano la variabile eleggibilità  $S^*$  come strumento per correggere l'endogeneità della variabile  $R$ .

Nel nostro studio invece partiamo da un modello che mette in relazione il logaritmo del consumo con i valori dell'eleggibilità  $S^*$ . L'obiettivo di questo studio di simulazione è quello di cercare di capire cosa succede agli stimatori del modello principale in presenza di errore di misura, in particolar modo al parametro relativo al calo del consumo.

La domanda di ricerca, che ci ha spinto ad effettuare questo studio di simulazione, riguarda l'effetto dell'errore di misura sul parametro relativo al salto del logaritmo del consumo per quei soggetti potenzialmente idonei al pensionamento. Infatti, vogliamo capire se la presenza dell'errore di misura

1. evidenzia il salto oppure
2. elimina o attenua il salto.

Quindi, se è vera la prima opzione, sotto l'ipotesi nulla in cui il pensionamento è una decisione anticipata, non ci dovrebbe essere alcuna discontinuità nei consumi. Ci siamo chiesti se tale discontinuità possa essere osservata per effetto dell'errore di misura. Se ciò fosse vero, vorrebbe dire che, intuitivamente, l'unico meccanismo di errore che garantirebbe questa cosa, è un errore di misura con proprietà diverse subito prima e subito dopo l'aver maturato l'ammissibilità al pensionamento, tali da trasformare una relazione lineare in una relazione discontinua. Invece, lavorando sotto l'ipotesi nulla in cui gli agenti arrivano impreparati al pensionamento, la presenza dell'errore di misura dovrebbe rendere la potenza del test per il parametro relativo al salto molto elevata, così che nella maggior parte dei casi siamo portati ad accettare la presenza del salto.

Contrariamente, se è vera la seconda opzione, ci aspettiamo che, nel caso in cui, sotto l'ipotesi nulla in cui il pensionamento è una decisione anticipata e dunque non ci dovrebbe essere alcuna discontinuità nei consumi, anche se c'è la presenza di errore di misura, non si dovrebbe evidenziare una discontinuità nei consumi. Supponendo invece che ci sia una discontinuità nel profilo dei consumi, per effetto dell'errore di misura, non si dovrebbe riuscire ad identificare ed evidenziare questa discontinuità. La presenza dell'errore di misura dovrebbe rendere la potenza del test per il parametro relativo al salto del consumo molto bassa, così che, nella maggior parte dei casi, siamo portati a rifiutare la presenza del salto.

Il metodo di stima che abbiamo utilizzato è quello dei minimi quadrati ordinari. Lo studio ha permesso di analizzare l'andamento del logaritmo del consumo, tenendo sotto controllo diversi indici e considerando, nelle varie simulazioni, due diverse strutture dell'errore di misura. Abbiamo ipotizzato che l'errore di misura fosse presente in misure diverse.

### 3.1.1 Variabili economiche

Come abbiamo appena detto, usiamo un modello che mette in relazione il logaritmo del consumo con i valori dell'eleggibilità  $S^*$ , in prossimità della soglia dell'eleggibilità alla pensione.

Ipotizziamo che  $\log C|S^* \sim N(\mu_C, \sigma_C^2)$ . E' possibile ammettere la normalità della funzione di consumo. Studi empirici infatti dimostrano che, se si considerano gruppi omogenei di individui, con caratteristiche demografiche simili, quali l'età, la distribuzione del consumo è riconducibile ad una distribuzione normale. Si veda [40] per maggiori dettagli.

Descriviamo i due modelli principali dai quali simuliamo. Sotto l'ipotesi che il pensionamento è un evento pianificato da parte degli agenti e dunque non ci dovrebbe essere alcuna discontinuità nel consumo, ipotizziamo che il valore atteso della funzione di consumo segua il modello con la seguente forma lineare rispetto ai parametri:

$$E[\log C|S^*] \equiv \mu_C = \alpha + \beta S^*. \quad (3.2)$$

Sotto l'ipotesi che il pensionamento non è un evento pianificato da parte degli agenti e dunque si presenta una discontinuità nel consumo, ipotizziamo che il valore medio della funzione di consumo abbia la seguente forma lineare rispetto ai parametri:

$$E[\log C|S^*] \equiv \mu_C = \alpha + \beta S^* + \gamma I_{[S^* \geq 0]}, \quad (3.3)$$

dove con  $S^*$  viene indicata la **vera eleggibilità** ovvero quella variabile che misura, senza errore, il numero di anni che mancano alla pensione e quelli che sono trascorsi dopo aver maturato il diritto di andare in pensione. La variabile  $S^*$  assumerà valori negativi se gli individui non sono ancora idonei per andare in pensione e valori positivi se gli individui hanno già passato il punto di soglia e sarebbero idonei alla pensione. Ipotizziamo che  $S^*$  si distribuisca come una normale di media  $\mu$  e varianza  $\sigma_*^2$ . In simboli,  $S^* \sim N(\mu, \sigma_*^2)$ . L'assunzione di normalità è un'ipotesi di lavoro, che caratterizza il campione.

Indicato con  $\bar{s} \equiv 0$  il punto in cui un individuo passa dallo stato di non idoneità a quello di idoneità alla pensione, con  $\bar{s}^+$  vengono indicati gli anni già trascorsi da  $\bar{s}$  mentre con  $\bar{s}^-$  vengono indicati gli anni che mancano all'individuo per diventare idoneo a ricevere la pensione. Ciò vuol dire che gli individui risultano idonei al pensionamento se e solo se riportano un valore della variabile  $S^*$  non negativo e dunque non inferiore al punto di soglia  $\bar{s} = 0$ . Lo studio di simulazione viene effettuato in un intorno di questa soglia  $\bar{s}$ .

Lavorando sotto l'ipotesi che il pensionamento non è un evento pianificato da parte degli agenti, abbiamo ipotizzato la presenza di un salto del consumo in prossimità della soglia  $\bar{s} = 0$ . Questo salto viene indicato dal parametro  $\gamma$ , nel modello (3.3), dove compare la variabile  $I_{[S^* \geq 0]}$ . Si tratta di una variabile dicotomica che associa valore 1 se l'elemento della variabile  $S^*$  risulta positivo e 0 altrimenti. Quindi, vale 1 se l'individuo possiede un valore della vera eleggibilità positivo e quindi idoneo per andare in pensione, e zero altrimenti. Con questa variabile vogliamo individuare, nell'intero campione, il sottocampione formato da individui idonei al pensionamento. Il nostro studio assume implicitamente che tutti vadano in pensione appena possibile. Quindi  $S^*$  misura implicitamente da quanti anni si è in pensione. Dunque, il parametro  $\gamma$  rappresenta il termine differenziale che determina quanto meno consumano quegli individui che sono in pensione.

### 3.1.2 Errore di misura

Come abbiamo già detto nel Capitolo 1, nella realtà, manipolando dati reali, ciò che si osserva è una porzione non trascurabile di individui che, pur non essendo idonei al pensionamento, risultano in pensione. Infatti, contrariamente a quanto ci si potrebbe attendere, nell'intorno sinistro di  $\bar{s}$ , la probabilità di osservare individui in pensione è diversa da zero. Questo risultato inatteso induce a pensare che i valori assunti dalla variabile  $S^*$  siano affetti da errore di misura. Per questo motivo, nel nostro studio di simulazione, consideriamo la variabile  $S$  per indicare l'eleggibilità osservata, affetta da errore di misura.

Il caso in cui l'errore di misura sia insidiato proprio in questa variabile è molto interessante dal punto di vista teorico. Dal momento che vengono sfruttati i presupposti che lo assimilano ad un esperimento locale, la presenza dell'errore di misura in un intorno di  $\bar{s}$  è delicata in quanto il ricercatore potrebbe essere fuorviato dal pensare che un'unità faccia parte dell'intorno destro di  $\bar{s}$  quando in realtà appartiene all'intorno sinistro, e viceversa.

Una titubanza che potrebbe sorgere è quella di pensare che l'errore di misura sia contenuto nella variabile  $R$ , che rappresenta la condizione dell'individuo, se è in pensione oppure no. Se ciò fosse vero sarebbe una questione che si ridurrebbe ad un problema di rilevazione: in effetti è inverosimile che un individuo menta sul fatto di essere o meno in pensione. Solitamente, i questionari somministrati contengono una serie di domande di controllo per aiutare il ricercatore a capire se l'intervistato riferisce una risposta non corretta in merito alla sua condizione.

Tornando allo studio di simulazione ipotizziamo che  $S|S^* \sim N(s^*, \sigma_{s|s^*}^2)$ . Abbiamo scelto di simulare da una distribuzione gaussiana, centrata nel vero valore dell'eleggibilità, in quanto si vuole che l'eleggibilità errata, pur essendo affetta da errore di misura, riesca a cogliere, in media, quella che è la vera eleggibilità. Per quanto riguarda la varianza si è ritenuto opportuno farla variare per diversi valori a seconda delle diverse strutture adottate per l'errore di misura. In seguito verrà fornita una spiegazione dettagliata.

Le due strutture utilizzate per l'errore di misura possono essere riassunte come segue. La prima è una struttura dell'errore di tipo **classico**

$$S|S^* \equiv S_{\text{classico}} = S^* + U. \quad (3.4)$$

L'espressione (3.4) sta a significare che si è intenzionati a conoscere la vera variabile  $S^*$  ma si è impossibilitati a causa di varie fonti di errore di misura, come per esempio errori manuali di trascrizione o imprecisione dell'operatore. In questo modello addittivo la variabile  $U$  ha media zero ed è indipendente da  $S^*$ . Quindi, in simboli,  $U \sim N(0, \sigma_u^2)$ . In questo caso la varianza di  $S_{\text{classico}}$

si definisce come segue:

$$\text{Var}[S_{\text{classico}}] = \text{Var}[S^*] + \text{Var}[U] = \sigma_*^2 + \sigma_u^2. \quad (3.5)$$

Ai fini della simulazione si è ritenuto opportuno non attribuire un valore preciso a  $\text{Var}[S_{\text{classico}}]$ , ma farla variare. Identificato con  $\rho$  il rapporto fra la varianza del termine d'errore  $U$  e la varianza di  $S_{\text{classico}}$  ovvero

$$\rho = \frac{\text{Var}[U]}{\text{Var}[S_{\text{classico}}]} = \frac{\sigma_u^2}{\sigma_*^2 + \sigma_u^2}, \quad (3.6)$$

si ha

$$\sigma_u^2 = \frac{\rho}{1 - \rho} \sigma_*^2. \quad (3.7)$$

La varianza del termine d'errore  $U$  varia a seconda del valore assegnato al parametro  $\rho$ . Di conseguenza anche  $\text{Var}[S_{\text{classico}}]$  assumerà valori diversi a seconda del valore assunto da  $\sigma_u^2$ .

La seconda è una struttura dell'errore di misura più generale. Parliamo di una struttura dell'errore di tipo **mistura**. In questo caso, la mistura combina una componente erratica classica e delle vere osservazioni, non affette da errore di misura. Si assume che  $\tau$  sia il valore osservato per l'eleggibilità. Si assume che una porzione di individui, durante la rilevazione, riportino il valore vero, ovvero  $S^* = \tau$  mentre altri riferiscano un valore di  $S = \tau$ , affetto da errore di misura. Tipicamente, una mistura è una distribuzione che viene utilizzata per rappresentare delle osservazioni dove la popolazione può essere divisa in due sotto popolazioni e ciascuna sotto popolazione pesa  $Z$  e  $1 - Z$  in termini percentuali. Formalmente, viene dunque indicato

$$S|S^* \equiv S_{\text{mistura}} = S^*Z + S_{\text{classico}}(1 - Z), \quad (3.8)$$

dove  $Z$  rappresenta una variabile binaria che vale 1 se il valore riferito è corretto, mentre è uguale a 0 altrimenti. In simboli  $Z \sim \text{Bin}(1, p)$ , dove con il parametro  $p \in (0, 1)$  viene indicata la probabilità che il singolo individuo riferisca il valore corretto. In termini di popolazione il parametro  $p$  rappresenta la proporzione della popolazione che fornisce il valore corretto dell'eleggibilità. La variabile  $S_{\text{classico}}$  invece è la stessa descritta nella (3.4).

Questo modello è noto come *contaminated sampling model*, discusso anche da Horowitz e Manski. Si veda [41] per maggiori dettagli.

Si noti che la (3.8), se  $p = 0$  nella (3.8), allora l'errore di tipo mistura viene a coincidere con un errore di tipo classico.

### 3.1.3 Parametri della simulazione

In questo paragrafo spiegheremo i valori numerici assegnati ai parametri.

Nel lavoro di riferimento proposto da Battistin *et al.* [5] le informazioni relative al pensionamento e al consumo degli individui sono state ottenute dall'indagine sui bilanci delle famiglie italiane SHIW, <sup>1</sup> condotta dalla Banca d'Italia, nel periodo 1993–2004. Si tratta di un'indagine che nasce negli anni '60 con l'obiettivo di raccogliere informazioni sui redditi e i risparmi delle famiglie italiane. Negli anni lo studio si è ampliato per includere anche la ricchezza e altri aspetti inerenti i comportamenti economici e finanziari delle famiglie, come ad esempio l'uso dei mezzi di pagamento. Nelle ultime indagini il campione è formato da circa 8.000 famiglie, distribuite in circa 300 comuni italiani. Nello studio di simulazione, abbiamo voluto emulare l'indagine condotta dalla Banca d'Italia. Si è generato un campione di numerosità  $n = 8000$ . L'esperimento semplice viene ripetuto con un numero di replicazioni pari a  $B = 1000$ .

Per individuare i valori della simulazione da assegnare ai parametri  $\theta = (\alpha, \beta, \gamma, \sigma_C^2)$ , si è preso in considerazione, oltre al lavoro di riferimento proposto da Battistin *et al.* [5] anche Battistin *et al.* [40]. In quest'ultimo, gli autori basano la maggior parte delle loro analisi empiriche su dati provenienti da indagini di tipo panel sui consumi e sui redditi delle famiglie degli Stati Uniti. In particolare fanno riferimento agli studi CEX <sup>2</sup> e PSID <sup>3</sup>. Completano l'analisi utilizzando anche i dati relativi all'indagine sui consumi delle famiglie del Regno Unito. Nello specifico si fa riferimento allo studio FES <sup>4</sup>.

---

<sup>1</sup>SHIW: *Survey on Household Income and Wealth*

<sup>2</sup>CEX: *Consumer Expenditure Survey*

<sup>3</sup>PSID: *Panel Study of Income Dynamics*

<sup>4</sup>FES: *Family Expenditure Survey*

Abbiamo quindi scelto come valore plausibile per la varianza del logaritmo del consumo  $\sigma_C^2 = 0.55^2$ , come proposto dalle analisi empiriche del lavoro di Battistin *et al.* [40]. Come si può notare nella Figura 3.1, la deviazione standard del logaritmo del consumo stimata è pari a 0.55.

Il parametro  $\beta$  indica invece la pendenza della retta, che cresce se  $\beta > 0$  e decresce se  $\beta < 0$ . Per scegliere un valore di tale parametro che fosse il più possibile vicino a quello reale, ci siamo avvalsi delle informazioni derivanti dal lavoro di Battistin *et al.* [5]. Infatti, abbiamo considerato il grafico rappresentato in Figura 3.2, che vede sull'asse delle ascisse  $S^*$  ovvero gli anni che mancano o che sono trascorsi dal momento del pensionamento e sull'asse delle ordinate il logaritmo del consumo di beni non durevoli. La pendenza della retta è decrescente e per questo abbiamo associato al parametro  $\beta$  un valore negativo.

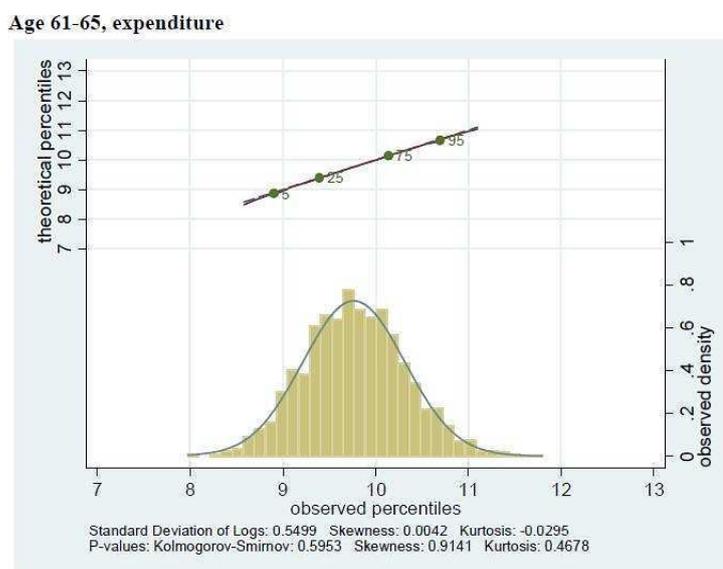


Figura 3.1: Rappresentazione empirica che mostra la distribuzione del logaritmo del consumo per individui con un'età compresa fra 61-65 anni. Il grafico proviene dal lavoro di Battistin *et al.* [40].

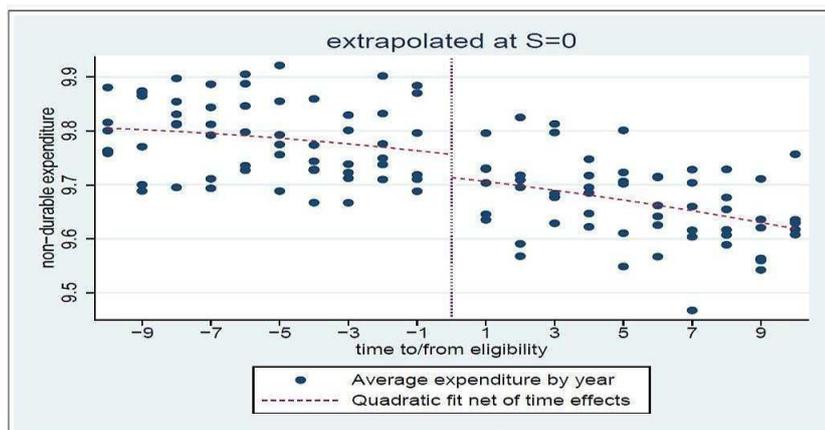


Figura 3.2: Stima dell'effetto dell'eleggibilità sul logaritmo del consumo di beni non durevoli. Il grafico è stato preso dal lavoro di Battistin *et al.* [5]

Per determinare il valore assoluto da assegnare al parametro  $\beta$  abbiamo scelto due punti del tipo  $A(x_a; y_a)$  e  $B(x_b; y_b)$ . Abbiamo applicato la classica formula che determina il valore del coefficiente angolare di una retta passante per due punti

$$\beta = \frac{y_b - y_a}{x_b - x_a}. \quad (3.9)$$

Dunque, il valore per  $\beta$ , utilizzato nelle simulazioni, è pari a  $\beta = -0.02$ . Quando la variabile dipendente è in forma logaritmica e il regressore no, il coefficiente  $\beta$  cattura la variazione relativa della variabile dipendente rispetto a una variazione unitaria della variabile esplicativa. Il valore  $\beta = -0.02$  possiamo interpretarlo come la variazione percentuale del consumo associata a una variazione di una unità dell'eleggibilità. Quindi, per esempio, passando dalla condizione in cui mancano due anni per andare in pensione ( $s^* = -2$ ) alla condizione che manca un solo anno ( $s^* = -1$ ), il consumo diminuisce del 2%.

Matematicamente parlando,  $\alpha$  è il termine noto ed è pari a  $E[\log C|S^* =$

0], quando non c'è discontinuità nel consumo. In altre parole rappresenta il valore atteso del logaritmo del consumo in corrispondenza di un valore dell'eleggibilità pari a zero, ovvero quando un individuo raggiunge esattamente l'età adeguata per andare in pensione. Abbiamo quindi scelto un punto  $A(x_a; y_a)$  e sostituito tali valori nella regressione (3.2), utilizzando  $\beta = -0.02$ . Nelle simulazioni abbiamo usato un valore di  $\alpha = 9.76$ .

Come si può notare in Figura 3.2, in prossimità del punto di soglia  $\bar{s} = 0$  il logaritmo del consumo passa da 9.76 a 9.7, registrando quindi un calo, in valore assoluto, pari a  $-0.06$ . Nelle simulazioni è stato quindi utilizzato il valore  $\gamma = -0.06$ . Vuol dire che la sotto popolazione di individui con un valore dell'eleggibilità positivo, quindi idonei al pensionamento, consumano lo 0.06 in meno rispetto a quegli individui con un valore dell'eleggibilità negativo e quindi non idonei al pensionamento.

Tornando allo studio di simulazione si ipotizza che  $S^* \sim N(\mu, \sigma_*^2)$ . Abbiamo ipotizzato che la media  $\mu = 0$ . La variabile  $S^*$  rappresenta il numero di anni che mancano al pensionamento oppure gli anni trascorsi dal momento in cui si diventa idonei ad esso. Il fatto di considerare, per questa variabile, una distribuzione normale, centrata in zero, implica che stiamo studiando una popolazione dove la probabilità di campionare un soggetto a cui mancano tre anni per andare in pensione è uguale alla probabilità di campionare un soggetto che da tre anni è idoneo al pensionamento. L'idea è quella di campionare i soggetti in maniera simmetrica attorno allo zero.

Inoltre, dal momento che si vuole studiare il fenomeno in un intorno della soglia  $\bar{s} = 0$ , si considerano solo i valori di  $S^*$  compresi fra  $-5$  e  $+5$ . Pertanto è stato scelto un valore di  $\sigma_*^2 = 4$  al fine di ottenere la gran parte delle osservazioni simulate comprese entro questi valori. Ricapitolando, in simboli,  $S^* \sim N(0, 4)$ .

Il parametro  $\rho$ , spiegato nella (3.6), è stato fatto variare secondo valori suggeriti dall'evidenza empirica. In particolare  $\rho = 10\%, 20\%, 30\%$ .

Nello studio di simulazione al parametro  $p$  sono stati attribuiti diversi valori in modo da ottenere un quadro sufficientemente esauriente di come

il consumo possa modificarsi al variare della probabilità di contaminazione nei dati. Nello specifico,  $p = 1, 0.5, 0.4, 0.3, 0.2, 0$ . Si noti che, nel caso in cui  $p = 0$ , l'errore di misura di tipo mistura coincide con il caso in cui la struttura dell'errore di misura è di tipo classico, come descritto nella (3.4).

### 3.1.4 Scenari considerati

In questo paragrafo spiegheremo gli scenari considerati nello studio di simulazione.

Nello studio di simulazione verranno presi in esame quattro diversi casi:

- analizziamo il caso in cui simuliamo sotto l'ipotesi che il pensionamento è un evento pianificato da parte degli agenti e dunque non c'è alcuna discontinuità nel consumo e stimiamo un modello che non prevede una discontinuità nel consumo. Simuliamo e stimiamo quindi da un modello del tipo (3.2), che non prevede la presenza di un salto nel consumo. Questo caso lo indicheremo con la sigla S0.M0.
- analizziamo il caso in cui simuliamo sotto l'ipotesi che il pensionamento non è un evento pianificato da parte degli agenti e dunque si presenta una discontinuità nel consumo e stimiamo un modello che prevede una discontinuità nel consumo. Simuliamo e stimiamo quindi da un modello del tipo (3.3), che prevede la presenza di un salto nel consumo. Questo caso lo indicheremo con la sigla S1.M1.
- analizziamo il caso in cui simuliamo sotto l'ipotesi che il pensionamento è un evento pianificato da parte degli agenti e dunque non c'è alcuna discontinuità nel consumo e stimiamo invece un modello che prevede una discontinuità nel consumo. Simuliamo quindi da un modello del tipo (3.2), che non prevede la presenza di un salto nell'andamento del consumo e invece stimiamo un modello del tipo (3.3), che prevede un salto del consumo. Questo caso lo indicheremo con la sigla S0.M1.

- analizziamo il caso in cui simuliamo sotto l'ipotesi che il pensionamento non è un evento pianificato da parte degli agenti e dunque si presenta una discontinuità nel consumo e invece stimiamo un modello che non prevede una discontinuità nel consumo. Simuliamo quindi da un modello del tipo (3.3), che prevede la presenza di un salto nell'andamento del consumo e invece stimiamo un modello del tipo (3.2), che non prevede un salto del consumo. Questo caso lo indicheremo con la sigla S1.M0.

Per ciascuno di questi quattro casi:

- stimiamo il modello supponendo che le variabili esplicative non siano affette da errore di misura. Questi scenari corrispondono a  $p = 1$ , ovvero quando si ipotizza che tutti gli intervistati riferiscono il valore corretto dell'eleggibilità. Inseriamo quindi nel modello stimato le variabili  $S^*$  e/o  $I_{[S^* \geq 0]}$ , non affette da errore di misura;
- stimiamo il modello con errore di misura. Quindi, inseriamo nel modello stimato la variabile eleggibilità che, affetta da errore di misura, viene indicata con  $S_{\text{classico}}$  o con  $S_{\text{mistura}}$ , a seconda della diversa struttura ipotizzata per l'errore di misura. I modelli considerano tutte le combinazioni fra valori  $\rho = 0.1, 0.2, 0.3$  e i diversi pesi dell'errore di misura, ovvero  $p = 0.5, 0.4, 0.3, 0.2, 0$ . Con  $p = 0$ , ci riferiamo al caso di errore di tipo classico in quanto l'errore di tipo mistura viene a coincidere con un errore di tipo classico.

In totale, abbiamo effettuato 64 simulazioni, ciascuna con un numero di replicazioni  $B = 1000$  e con una numerosità campionaria pari a  $n = 8000$ . Le simulazioni sono state eseguite in R.

### 3.1.5 Indici riassuntivi

In questo paragrafo presentiamo gli indici riassuntivi calcolati sullo stimatore. Il metodo di stima utilizzato è quello dei minimi quadrati ordinari. In particolar modo calcoliamo:

- la distorsione assoluta, indicata con

$$D(\hat{\theta}) = E(\hat{\theta}) - \theta. \quad (3.10)$$

- la distorsione relativa,

$$\frac{E(\hat{\theta}) - \theta}{\theta}, \quad (3.11)$$

ovvero la distorsione assoluta rapportata al valore vero del parametro, riportata in percentuale.

- l'errore quadratico medio, come misura di precisione dello stimatore, calcolato come somma fra la varianza dello stimatore e la sua distorsione assoluta al quadrato,

$$\text{EQM}(\hat{\theta}) = \text{Var}(\hat{\theta}) + D(\hat{\theta})^2. \quad (3.12)$$

- il rapporto fra la varianza dello stimatore e la sua distorsione assoluta al quadrato,

$$\frac{\text{Var}(\hat{\theta})}{D(\hat{\theta})^2}. \quad (3.13)$$

Questo rapporto serve a capire quanto pesa la varianza, in termini di variabilità, rispetto alla distorsione.

- il livello di copertura reale degli intervalli di confidenza bilaterali basati sulla statistica di Wald,

$$1 - \hat{\alpha} = \frac{1}{B} \sum_{b=1}^B I_b, \quad (3.14)$$

dove  $I_b$  è la funzione indicatrice che vale 1 se la quantità  $(\hat{\theta}_b - \theta)/\hat{s}e(\hat{\theta}_b)$  appartiene all'intervallo  $(z_{\alpha/2}, z_{1-\alpha/2})$ , con  $z_\alpha$  il quantile di livello  $\alpha$  della normale standard, e 0 altrimenti. Consideriamo il livello di confidenza nominale  $1 - \alpha = 0.95$ . Fra parentesi, indichiamo anche gli intervalli al 95% del livello di confidenza stimato, basato sull'approssimazione normale della distribuzione di  $1 - \hat{\alpha}$ . Abbiamo considerato anche livelli di

copertura reali, come calcolati sopra, per intervalli di confidenza unilaterali. Nel caso di non adeguata copertura questi valori indicano se ciò è dovuto ad un'errata centratura degli intervalli di confidenza causata dalla distorsione oppure ad una questione di efficienza dello stimatore.

Infine, abbiamo calcolato dei test di significatività sui parametri  $\alpha$ ,  $\beta$  e  $\gamma$ . Abbiamo condotto un test alla Wald del tipo

$$\frac{\hat{\theta} - 0}{se(\hat{\theta})}, \quad (3.15)$$

dove, sotto l'ipotesi nulla, si ipotizza che il parametro  $\theta$  sia zero, contro l'alternativa che sia diverso da zero. Specifichiamo che con  $R_\theta$  indichiamo la regione di rifiuto che suggerisce di rifiutare l'ipotesi nulla. Con  $A_\theta$  indichiamo la regione di accettazione che suggerisce di accettare l'ipotesi nulla. Le espressioni che presentiamo qui di seguito assumono significato diverso a seconda dello scenario considerato. In particolare, l'espressione:

- $\Pr\{R_\theta|\theta\}$  indica la potenza del test (3.15);
- $\Pr\{A_\theta|\theta\}$  indica la probabilità dell'errore di secondo tipo per il test (3.15);
- $\Pr\{R_\theta|\theta = 0\}$  indica il livello di significatività reale del test (3.15);
- $\Pr\{A_\theta|\theta = 0\}$  indica il livello di copertura reale degli intervalli di confidenza bilaterali basati sulla (3.15), calcolato come nella (3.14).

## 3.2 Risultati della simulazione

### 3.2.1 Confronto fra risultati noti ed evidenza empirica

Mostriamo che alcuni risultati ottenuti nello studio di simulazione riportano valori molto vicini a quelli previsti in letteratura [15].

Prendiamo come esempio il caso in cui simuliamo e stimiamo un modello del tipo (3.2), simuliamo sotto l'ipotesi che il pensionamento è un evento

pianificato da parte degli agenti e dunque non c'è alcuna discontinuità nel consumo e, perciò, stimiamo un modello che non prevede una discontinuità nel consumo attorno alla soglia  $\bar{s}$ . Lo scenario al quale facciamo riferimento è del tipo S0.M0. In particolare, consideriamo lo scenario in cui viene stimato un modello di regressione lineare semplice dove la variabile esplicativa è misurata con errore. Non potendo osservare la variabile  $S^*$ , si osserva la variabile  $S_{\text{classico}} = S^* + U$ . Si tratta di un modello per l'errore di misura classico. Si veda la (2.2). Coincide inoltre con il caso di errore di tipo mistura, quando  $p = 0$ . Per stimare i parametri si utilizza il metodo dei minimi quadrati ordinari.

Il modello rappresentativo del processo generatore di dati è

$$\log C = \alpha + \beta S^* + \varepsilon, \quad (3.16)$$

dove  $S^*$ , come specificato in precedenza, è la variabile che riporta i veri valori dell'idoneità al pensionamento e  $S^* \sim N(\mu, \sigma_*^2)$  dove  $\mu = 0$  e  $\sigma_*^2 = 4$ . Mentre,  $\varepsilon \sim N(0; \sigma_\varepsilon^2)$ , ovvero  $\varepsilon \sim N(0; 0.55^2)$ .

Viene invece stimato il modello

$$\log C = \alpha_e + \beta_e S_{\text{classico}} + \epsilon, \quad (3.17)$$

dove  $S_{\text{classico}} = S^* + U$  con  $U \sim N(0, \sigma_u^2)$ . E' noto in letteratura che, in un modello di questo tipo, la regressione (3.17) non stima in modo consistente il parametro d'interesse  $\beta = -0.02$  ma piuttosto  $\beta_e = \lambda\beta$ , dove

$$\lambda = \frac{\sigma_*^2}{\sigma_*^2 + \sigma_u^2} < 1. \quad (3.18)$$

Perciò, la regressione (3.17), stimata con il metodo dei minimi quadrati ordinari, produce uno stimatore del coefficiente angolare  $\beta$  che è distorto verso lo zero. La presenza dell'errore di misura causa una distorsione degli stimatori. La distorsione di cui si parla è nota in letteratura con il nome di **attenuazione verso lo zero** in quanto i parametri vengono sottostimati. Infatti, come si può notare in Tabella 3.1, lo stimatore  $\beta_e$  sottostima il parametro  $\beta = -0.02$ .

Inoltre, la varianza residua della regressione del logaritmo del consumo sulla variabile esplicativa affetta da errore  $S_{\text{classico}}$ , ovvero la  $\text{Var}(\log C|S_{\text{classico}}) = \text{Var}(\epsilon|S_{\text{classico}})$  espressa come

$$\sigma_\epsilon^2 + \frac{\beta^2 \sigma_u^2 \sigma_*^2}{\sigma_*^2 + \sigma_u^2}, \quad (3.19)$$

è maggiore rispetto a quella fornita dal vero modello che mette in relazione il logaritmo del consumo con la vera eleggibilità  $S^*$ . Infatti, come si può notare in Tabella 3.1, tale quantità è maggiore rispetto a  $\sigma_\epsilon^2 = 0.3025 = 0.55^2$  e cresce, all'aumentare di  $\rho$ .

Riproduciamo la medesima tabella usando i valori empirici di  $\beta_e$  e  $\text{Var}(\log C|S_{\text{classico}})$ , con  $\lambda$  fissato. Precisiamo che, poichè  $\lambda$  è funzione di  $\rho$ , per calcolare il valore empirico  $\beta_e$  abbiamo preso la media delle stime  $\hat{\beta}_{e,b}$ , di ogni singola replicazione  $b$ . Per calcolare il valore empirico di  $\text{Var}(\log C|S_{\text{classico}})$  abbiamo preso la media delle  $B$  stime per le varianze di  $\epsilon$ . In Tabella 3.2 sono riportati questi valori empirici e si può notare quanto questi valori siano vicini a quelli previsti dalla teoria, riportati in Tabella 3.1.

	$\lambda$	$\beta_e = \lambda\beta$	$\text{Var}(\log C S_{\text{classico}})$
<b>Errore di misura classico con <math>\rho = 0.1</math></b>	0.900	-0.0180	0.30266
<b>Errore di misura classico con <math>\rho = 0.2</math></b>	0.800	-0.0160	0.30282
<b>Errore di misura classico con <math>\rho = 0.3</math></b>	0.700	-0.0140	0.30298

Tabella 3.1: La tabella riporta i valori teorici di  $\lambda$ , calcolata come nella (3.18), i valori teorici della stima del coefficiente angolare  $\beta_e$  e il valore teorico della varianza residua della regressione del logaritmo del consumo su  $S_{\text{classico}}$ , calcolata come nella (3.19).

	$\lambda$	$\beta_e = \lambda\beta$	$\text{Var}(\log C S_{\text{classico}})$
<b>Errore di misura classico con <math>\rho = 0.1</math></b>	0.900	-0.0181	0.30269
<b>Errore di misura classico con <math>\rho = 0.2</math></b>	0.800	-0.0160	0.30292
<b>Errore di misura classico con <math>\rho = 0.3</math></b>	0.700	-0.0141	0.30297

Tabella 3.2: La tabella riporta la media campionaria del coefficiente angolare  $\beta_e$  e della varianza residua della regressione del logaritmo del consumo sulla variabile eleggibilità, affetta da errore di misura di tipo classico.

### 3.2.2 Commenti al caso S0.M0. e S1.M0.

In questo paragrafo ci limiteremo a commentare brevemente i casi S0.M0. e S1.M0. Questi non sono i casi di maggior interesse, in quanto non ci aiutano a rispondere alla nostra domanda di ricerca.

Riportiamo la Tabella 3.3. La tabella riporta i valori della distorsione relativa per i parametri  $\alpha$  e  $\beta$ . La parte inferiore della tabella riporta i valori del livello di copertura reale. Fra parentesi viene indicato l'intervallo di confidenza per il livello di copertura reale. Livello nominale considerato pari a 0.95. Le restanti tabelle, relative al caso S0.M0., contenenti gli altri valori degli indici calcolati, sono raccolte in appendice.

Stimando il modello senza errore di misura, usando come variabile esplicativa  $S^*$ , gli stimatori  $\hat{\alpha}$  e  $\hat{\beta}$  non sono distorti. Quando invece la variabile esplicativa è misurata con errore, lo stimatore  $\hat{\beta}$  produce stime distorte del parametro  $\beta$ . La distorsione si accentua al diminuire di  $p$  e all'aumentare di  $\rho$ . Questa distorsione provoca una riduzione del livello di copertura reale, in quanto gli intervalli di confidenza per il parametro  $\beta$  non sono correttamente centrati attorno al vero valore. Si vedano le Tabelle 3.3.

Invece, prestiamo attenzione al caso in cui simuliamo da un modello del tipo (3.3) e stimiamo un modello del tipo (3.2), ovvero ipotizziamo che nella

realtà il pensionamento non sia un evento pianificato e stimiamo invece un modello che non preveda una discontinuità nel consumo attorno alla soglia  $\bar{s}$ . Lo scenario al quale facciamo riferimento è del tipo S1.M0.

Anche senza considerare la presenza dell'errore di misura, la distorsione relativa del coefficiente  $\beta$  è molto forte e il rapporto fra la varianza di  $\hat{\beta}$  e la sua distorsione al quadrato è prossimo allo zero a dimostrazione del fatto che la distorsione è molto più forte della varianza di  $\hat{\beta}$ . Si vedano le Tabelle 3.4, relative al caso senza errore di misura. In effetti, stimando questo modello, si sta ommettendo un regressore rilevante. L'effetto di tale omissione influenza la componente erratica del modello. Se la variabile omessa risulta collegata con i regressori presenti nel modello si ottengono stime distorte dei parametri, traendo conclusioni errate sul peso della relazione esistente fra la variabile risposta e le variabili esplicative. Infatti viene a mancare una delle ipotesi fondamentali alla base del metodo dei minimi quadrati ordinari ovvero l'ipotesi di **stretta esogeneità** e le stime dei parametri risultano non corrette.

Questa distorsione si accentua all'aumentare della contaminazione di errore di misura nei dati, riducendo ulteriormente il rapporto fra varianza e distorsione al quadrato. I livelli di copertura reale per i parametri  $\alpha$  e  $\beta$  sono molto bassi e si discostano da quello che è il livello di copertura nominale considerato. E questo vale sia per il modello stimato senza errore di misura sia per i modelli stimati con errore di misura. L'errata centratura degli intervalli di confidenza per tali parametri sono dovuti appunto alla distorsione delle stime. Si vedano Tabelle 3.5.

VALORI DELLA DISTORSIONE RELATIVA, CALCOLATA COME $\frac{E(\hat{\theta})-\theta}{\theta}$			
VALORI OTTENUTI STIMANDO IL MODELLO SENZA ERRORE DI MISURA (P = 1): $\alpha$ : 0.0007% $\beta$ : 0.0608%			
CON ERRORE DI MISURA:	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
P = 0.5	$\alpha$ : 0% $\beta$ : 5.1%	$\alpha$ : 0.1% $\beta$ : 10.9%	$\alpha$ : 0.0007% $\beta$ : 18.03%
P = 0.4	$\alpha$ : 0.1% $\beta$ : 6.2%	$\alpha$ : 0.1% $\beta$ : 13.5%	$\alpha$ : 0.1% $\beta$ : 19.9%
P = 0.3	$\alpha$ : 0.1% $\beta$ : 7.1%	$\alpha$ : 0% $\beta$ : 14.7%	$\alpha$ : 0% $\beta$ : 23.1%
P = 0.2	$\alpha$ : 0% $\beta$ : 7.3%	$\alpha$ : 0.1% $\beta$ : 16.96%	$\alpha$ : 0.1% $\beta$ : 25.3%
P = 0	$\alpha$ : 0% $\beta$ : 9.9%	$\alpha$ : 0% $\beta$ : 20.3%	$\alpha$ : 0.1% $\beta$ : 29.8%

LIVELLO DI COPERTURA REALE PER INTERVALLI DI CONFIDENZA BILATERALI (IN AGGIUNTA: INTERVALLO DI CONFIDENZA AL 95%). LIVELLO NOMINALE CONSIDERATO: 0.95			
VALORI OTTENUTI STIMANDO IL MODELLO SENZA ERRORE DI MISURA (P = 1): $\alpha$ : 0.959 (0.944 ; 0.970) $\beta$ : 0.952 (0.936 ; 0.964)			
CON ERRORE DI MISURA:	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
P = 0.5	$\alpha$ : 0.945 (0.928 ; 0.957) $\beta$ : 0.934 (0.916 ; 0.948)	$\alpha$ : 0.940 (0.922 ; 0.953) $\beta$ : 0.886 (0.864 ; 0.904)	$\alpha$ : 0.950 (0.934 ; 0.961) $\beta$ : 0.752 (0.723 ; 0.778)
P = 0.4	$\alpha$ : 0.956 (0.940 ; 0.967) $\beta$ : 0.936 (0.918 ; 0.950)	$\alpha$ : 0.956 (0.940 ; 0.967) $\beta$ : 0.841 (0.816 ; 0.862)	$\alpha$ : 0.953 (0.937 ; 0.964) $\beta$ : 0.687 (0.657 ; 0.715)
P = 0.3	$\alpha$ : 0.963 (0.948 ; 0.973) $\beta$ : 0.916 (0.896 ; 0.932)	$\alpha$ : 0.945 (0.928 ; 0.957) $\beta$ : 0.833 (0.808 ; 0.855)	$\alpha$ : 0.957 (0.942 ; 0.968) $\beta$ : 0.602 (0.570 ; 0.632)
P = 0.2	$\alpha$ : 0.959 (0.944 ; 0.970) $\beta$ : 0.919 (0.899 ; 0.934)	$\alpha$ : 0.944 (0.927 ; 0.957) $\beta$ : 0.792 (0.765 ; 0.816)	$\alpha$ : 0.959 (0.944 ; 0.970) $\beta$ : 0.516 (0.484 ; 0.547)
P = 0	$\alpha$ : 0.951 (0.935 ; 0.963) $\beta$ : 0.886 (0.864 ; 0.905)	$\alpha$ : 0.948 (0.932 ; 0.961) $\beta$ : 0.684 (0.654 ; 0.712)	$\alpha$ : 0.957 (0.942 ; 0.968) $\beta$ : 0.334 (0.315 ; 0.374)

Tabella 3.3: La tabella in alto riporta i valori della distorsione relativa riferita agli stimatori  $\hat{\alpha}$  e  $\hat{\beta}$ . La tabella in basso riporta i valori del livello di copertura reale. Fra parentesi viene indicato l'intervallo di confidenza per il livello di copertura reale. Livello nominale considerato pari a 0.95. Le tabelle si riferiscono entrambe al **caso S0.M0**.

VALORI DELLA DISTORSIONE RELATIVA, CALCOLATA COME $\frac{E(\hat{\theta})-\theta}{\theta}$			
VALORI OTTENUTI STIMANDO IL MODELLO SENZA ERRORE DI MISURA (P = 1): $\alpha$ : 0.305% $\beta$ : 39.59%			
CON ERRORE DI MISURA:	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
P = 0.5	$\alpha$ : 0.307% $\beta$ : 51.59%	$\alpha$ : 0.307% $\beta$ : 42.1%	$\alpha$ : 0.306% $\beta$ : 31.8%
P = 0.4	$\alpha$ : 0.304% $\beta$ : 50.01%	$\alpha$ : 0.309% $\beta$ : 38.9%	$\alpha$ : 0.307% $\beta$ : 26.7%
P = 0.3	$\alpha$ : 0.307% $\beta$ : 48.6%	$\alpha$ : 0.307% $\beta$ : 35.3%	$\alpha$ : 0.309% $\beta$ : 22.7%
P = 0.2	$\alpha$ : 0.307% $\beta$ : 47.1%	$\alpha$ : 0.308% $\beta$ : 33.4%	$\alpha$ : 0.309% $\beta$ : 19.20%
P = 0	$\alpha$ : 0.305% $\beta$ : 9.217%	$\alpha$ : 0.312% $\beta$ : 20.11%	$\alpha$ : 0.305% $\beta$ : 30.68%

VALORI DEL RAPPORTO $\frac{Var(\hat{\theta})}{D(\hat{\theta})^2}$			
VALORI OTTENUTI STIMANDO IL MODELLO SENZA ERRORE DI MISURA (P = 1): $\alpha$ : 0.0385 $\beta$ : 0.0667			
CON ERRORE DI MISURA:	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
P = 0.5	$\alpha$ : 0.043 $\beta$ : 0.100	$\alpha$ : 0.040 $\beta$ : 0.123	$\alpha$ : 0.043 $\beta$ : 0.187
P = 0.4	$\alpha$ : 0.044 $\beta$ : 0.090	$\alpha$ : 0.046 $\beta$ : 0.134	$\alpha$ : 0.043 $\beta$ : 0.291
P = 0.3	$\alpha$ : 0.040 $\beta$ : 0.097	$\alpha$ : 0.040 $\beta$ : 0.172	$\alpha$ : 0.040 $\beta$ : 0.322
P = 0.2	$\alpha$ : 0.041 $\beta$ : 0.097	$\alpha$ : 0.041 $\beta$ : 0.177	$\alpha$ : 0.041 $\beta$ : 0.444
P = 0	$\alpha$ : 0.040 $\beta$ : 1.943	$\alpha$ : 0.042 $\beta$ : 0.434	$\alpha$ : 0.043 $\beta$ : 0.188

Tabella 3.4: La tabella in alto riporta i valori della distorsione relativa riferita agli stimatori  $\hat{\alpha}$  e  $\hat{\beta}$ . La tabella in basso riporta i valori del rapporto fra varianza dello stimatore e la sua distorsione al quadrato. Le tabelle si riferiscono entrambe al caso **S1.M0**.

LIVELLO DI COPERTURA REALE PER INTERVALLI DI CONFIDENZA BILATERALI (IN AGGIUNTA: INTERVALLO DI CONFIDENZA AL 95%). LIVELLO NOMINALE CONSIDERATO: 0.95			
VALORI OTTENUTI STIMANDO IL MODELLO SENZA ERRORE DI MISURA (P = 1): $\alpha$ : 0.003 (0.0007 ; 0.009) $\beta$ : 0.034 (0.024 ; 0.048)			
CON ERRORE DI MISURA:	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
P = 0.5	$\alpha$ : 0.000 (0 ; 0.005) $\beta$ : 0.091 (0.074 ; 0.111)	$\alpha$ : 0.004 (0.0013 ; 0.011) $\beta$ : 0.168 (0.146 ; 0.193)	$\alpha$ : 0.000 (0 ; 0.005) $\beta$ : 0.373 (0.343 ; 0.404)
P = 0.4	$\alpha$ : 0.001 (0 ; 0.006) $\beta$ : 0.073 (0.058 ; 0.091)	$\alpha$ : 0.005 (0.0018 ; 0.012) $\beta$ : 0.232 (0.206 ; 0.260)	$\alpha$ : 0.004 (0.0012 ; 0.0109) $\beta$ : 0.507 (0.475 ; 0.538)
P = 0.3	$\alpha$ : 0.003 (0.0007 ; 0.0095) $\beta$ : 0.103 (0.085 ; 0.124)	$\alpha$ : 0.002 (0.0003 ; 0.008) $\beta$ : 0.309 (0.280 ; 0.338)	$\alpha$ : 0.001 (0 ; 0.006) $\beta$ : 0.620 (0.589 ; 0.650)
P = 0.2	$\alpha$ : 0.001 (0 ; 0.006) $\beta$ : 0.105 (0.870 ; 0.126)	$\alpha$ : 0.002 (0.0003 ; 0.008) $\beta$ : 0.336 (0.306 ; 0.366)	$\alpha$ : 0.001 (0 ; 0.0047) $\beta$ : 0.707 (0.677 ; 0.734)
P = 0	$\alpha$ : 0.003 (0.0008 ; 0.009) $\beta$ : 0.902 (0.881 ; 0.919)	$\alpha$ : 0.001 (0 ; 0.006) $\beta$ : 0.698 (0.668 ; 0.726)	$\alpha$ : 0.001 (0 ; 0.0064) $\beta$ : 0.342 (0.313 ; 0.372)

LIVELLO DI COPERTURA REALE PER INTERVALLI DI CONFIDENZA UNILATERALI. LIVELLO NOMINALE CONSIDERATO: 0.95						
VALORI OTTENUTI STIMANDO IL MODELLO SENZA ERRORE DI MISURA (P = 1): $\alpha$ : 0.002 e 1.000 $\beta$ : 0.015 e 1.000						
CON ERRORE DI MISURA:	$\rho = 0.1$		$\rho = 0.2$		$\rho = 0.3$	
P = 0.5	$\alpha$ : 0.000 $\beta$ : 0.05	1.000 1.000	$\alpha$ : 0.002 $\beta$ : 0.092	1.000 1.000	$\alpha$ : 0.000 $\beta$ : 0.269	1.000 1.000
P = 0.4	$\alpha$ : 0.000 $\beta$ : 0.042	1.000 1.000	$\alpha$ : 0.000 $\beta$ : 0.142	1.000 1.000	$\alpha$ : 0.002 $\beta$ : 0.393	1.000 0.999
P = 0.3	$\alpha$ : 0.002 $\beta$ : 0.058	1.000 1.000	$\alpha$ : 0.001 $\beta$ : 0.21	1.000 1.000	$\alpha$ : 0.001 $\beta$ : 0.485	1.000 1.000
P = 0.2	$\alpha$ : 0.000 $\beta$ : 0.059	1.000 1.000	$\alpha$ : 0.001 $\beta$ : 0.236	1.000 1.000	$\alpha$ : 0.000 $\beta$ : 0.588	1.000 1.000
P = 0	$\alpha$ : 0.001 $\beta$ : 0.988	1.000 0.836	$\alpha$ : 0.001 $\beta$ : 1.000	1.000 0.605	$\alpha$ : 0.000 $\beta$ : 1.000	1.000 0.272

Tabella 3.5: La tabella in alto riporta i valori del livello di copertura reale. Livello nominale pari a 0.95. Fra parentesi viene indicato l'intervallo di confidenza per il livello di copertura reale. La tabella in basso riporta i valori del livello di copertura reale calcolato per intervalli unilaterali calcolati considerando ipotesi unilaterale sinistra (valore a sinistra) e destra (valore a destra). Livello nominale 0.95. Le tabelle si riferiscono entrambe al **caso S1.M0**.

### 3.2.3 Tabelle per i casi S0.M1. e S1.M1.

In questo paragrafo riporteremo le tabelle con i risultati più rilevanti dello studio di simulazione per i due casi S0.M1. e S1.M1. Nel paragrafo successivo commenteremo i risultati facendo riferimento alle tabelle qui riportate. Per le restanti tabelle, si rimanda all'Appendice. I casi S0.M1. e S1.M1 sono interessanti per la nostra domanda di ricerca in quanto riusciamo a valutare, dal punto di vista statistico e dal punto di vista dell'interpretazione economica, i valori degli indici calcolati per lo stimatore relativo al parametro  $\gamma$ , sul quale si concentra la nostra attenzione.

L'ordine delle tabelle segue quello di presentazione e analisi dei risultati utilizzato nel paragrafo successivo.

- Inizialmente consideriamo il caso in cui simuliamo un modello del tipo (3.2) e stimiamo invece un modello del tipo (3.3). Il caso al quale facciamo riferimento è del tipo S0.M1.
- Proseguiamo, considerando il caso in cui simuliamo un modello del tipo (3.3) e stimiamo sempre da un modello del tipo (3.3). Il caso al quale facciamo riferimento è del tipo S1.M1.

VALORI DELLA DISTORSIONE RELATIVA, CALCOLATA COME $\frac{E(\hat{\theta})-\theta}{\theta}$			
VALORI OTTENUTI STIMANDO IL MODELLO SENZA ERRORE DI MISURA (P = 1): $\alpha$ : 0.00046% $\beta$ : 0.802% $\gamma$ : n.d.			
CON ERRORE DI MISURA:	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
P = 0.5	$\alpha$ : 0.0014% $\beta$ : 5.94% $\gamma$ : n.d.	$\alpha$ : 0.0032% $\beta$ : 12.22% $\gamma$ : n.d.	$\alpha$ : 0.0105% $\beta$ : 20.51% $\gamma$ : n.d.
P = 0.4	$\alpha$ : 0.0016% $\beta$ : 6.42% $\gamma$ : n.d.	$\alpha$ : 0.0095% $\beta$ : 14.63% $\gamma$ : n.d.	$\alpha$ : 0.0138% $\beta$ : 22.48% $\gamma$ : n.d.
P = 0.3	$\alpha$ : 0.0025% $\beta$ : 8.10% $\gamma$ : n.d.	$\alpha$ : 0.0045% $\beta$ : 17.08% $\gamma$ : n.d.	$\alpha$ : 0.0052% $\beta$ : 24.42% $\gamma$ : n.d.
P = 0.2	$\alpha$ : 0.0082% $\beta$ : 9.43% $\gamma$ : n.d.	$\alpha$ : 0.0036% $\beta$ : 18.34% $\gamma$ : n.d.	$\alpha$ : 0.0058% $\beta$ : 26.78% $\gamma$ : n.d.
P = 0	$\alpha$ : 0% $\beta$ : 9.63% $\gamma$ : n.d.	$\alpha$ : 0.1% $\beta$ : 19.71% $\gamma$ : n.d.	$\alpha$ : 0.002% $\beta$ : 30.68% $\gamma$ : n.d.

VALORI DEL RAPPORTO $\frac{Var(\hat{\theta})}{D(\hat{\theta})^2}$			
VALORI OTTENUTI STIMANDO IL MODELLO SENZA ERRORE DI MISURA (P = 1): $\alpha$ : 42554.826 $\beta$ : 363.355 $\gamma$ : 7374465.54			
CON ERRORE DI MISURA:	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
P = 0.5	$\alpha$ : 760583.36 $\beta$ : 13.5499 $\gamma$ : 119639.13	$\alpha$ : 1429.79 $\beta$ : 3.8055 $\gamma$ : 295.879	$\alpha$ : 124.120 $\beta$ : 1.239 $\gamma$ : 64.229
P = 0.4	$\alpha$ : 2461.563 $\beta$ : 11.669 $\gamma$ : 1147.493	$\alpha$ : 165.701 $\beta$ : 2.5889 $\gamma$ : 106.841	$\alpha$ : 83.132 $\beta$ : 1.0839 $\gamma$ : 83.094
P = 0.3	$\alpha$ : 11573.61 $\beta$ : 8.351 $\gamma$ : 1243.570	$\alpha$ : 200.963 $\beta$ : 1.865 $\gamma$ : 141.177	$\alpha$ : 575.007 $\beta$ : 0.8746 $\gamma$ : 154.684
P = 0.2	$\alpha$ : 123.839 $\beta$ : 7.34 $\gamma$ : 299.449	$\alpha$ : 1144.44 $\beta$ : 1.5434 $\gamma$ : 535.45	$\alpha$ : 449.794 $\beta$ : 0.6660 $\gamma$ : 436.756
P = 0	$\alpha$ : 355.679 $\beta$ : 2.068 $\gamma$ : 188.361	$\alpha$ : 1283.314 $\beta$ : 0.513 $\gamma$ : 723.070	$\alpha$ : 2254.956 $\beta$ : 0.1909 $\gamma$ : 1438.531

Tabella 3.6: La tabella in alto riporta i valori della distorsione relativa riferita agli stimatori  $\hat{\alpha}$ ,  $\hat{\beta}$ . Per  $\hat{\gamma}$  questo valore non è disponibile in quanto  $\gamma = 0$ . La tabella in basso riporta i valori del rapporto fra varianza dello stimatore e la sua distorsione al quadrato. Le tabelle si riferiscono entrambe al caso **S0.M1**.

LIVELLO DI COPERTURA REALE PER INTERVALLI DI CONFIDENZA BILATERALI (IN AGGIUNTA: INTERVALLO DI CONFIDENZA AL 95%). LIVELLO NOMINALE CONSIDERATO: 0.95			
VALORI OTTENUTI STIMANDO IL MODELLO SENZA ERRORE DI MISURA ( $P = 1$ ): $\alpha$ : 0.954 (0.937 ; 0.966) $\beta$ : 0.955 (0.940 ; 0.967) $\gamma$ : 0.967 (0.953 ; 0.977)			
CON ERRORE DI MISURA:	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
$P = 0.5$	$\alpha$ : 0.954 (0.938 ; 0.965) $\beta$ : 0.955 (0.939 ; 0.966) $\gamma$ : 0.955 (0.939 ; 0.966)	$\alpha$ : 0.955 (0.939 ; 0.966) $\beta$ : 0.922 (0.903 ; 0.937) $\gamma$ : 0.957 (0.942 ; 0.968)	$\alpha$ : 0.939 (0.921 ; 0.952) $\beta$ : 0.869 (0.846 ; 0.888) $\gamma$ : 0.939 (0.921 ; 0.952)
$P = 0.4$	$\alpha$ : 0.952 (0.936 ; 0.964) $\beta$ : 0.945 (0.928 ; 0.957) $\gamma$ : 0.952 (0.936 ; 0.964)	$\alpha$ : 0.949 (0.933 ; 0.961) $\beta$ : 0.916 (0.896 ; 0.932) $\gamma$ : 0.951 (0.935 ; 0.963)	$\alpha$ : 0.923 (0.904 ; 0.938) $\beta$ : 0.829 (0.803 ; 0.851) $\gamma$ : 0.935 (0.917 ; 0.949)
$P = 0.3$	$\alpha$ : 0.964 (0.949 ; 0.974) $\beta$ : 0.933 (0.915 ; 0.947) $\gamma$ : 0.961 (0.946 ; 0.971)	$\alpha$ : 0.953 (0.937 ; 0.964) $\beta$ : 0.884 (0.862 ; 0.902) $\gamma$ : 0.954 (0.938 ; 0.965)	$\alpha$ : 0.951 (0.935 ; 0.963) $\beta$ : 0.811 (0.785 ; 0.834) $\gamma$ : 0.949 (0.933 ; 0.961)
$P = 0.2$	$\alpha$ : 0.954 (0.938 ; 0.965) $\beta$ : 0.945 (0.928 ; 0.957) $\gamma$ : 0.944 (0.927 ; 0.957)	$\alpha$ : 0.953 (0.937 ; 0.964) $\beta$ : 0.895 (0.873 ; 0.912) $\gamma$ : 0.965 (0.951 ; 0.975)	$\alpha$ : 0.946 (0.929 ; 0.958) $\beta$ : 0.764 (0.736 ; 0.789) $\gamma$ : 0.948 (0.931 ; 0.960)
$P = 0$	$\alpha$ : 0.943 (0.926 ; 0.956) $\beta$ : 0.893 (0.872 ; 0.911) $\gamma$ : 0.961 (0.946 ; 0.972)	$\alpha$ : 0.973 (0.960 ; 0.982) $\beta$ : 0.706 (0.676 ; 0.734) $\gamma$ : 0.971 (0.958 ; 0.980)	$\alpha$ : 0.961 (0.946 ; 0.972) $\beta$ : 0.344 (0.315 ; 0.374) $\gamma$ : 0.949 (0.933 ; 0.961)

TEST DI SIGNIFICATIVITA' SUI PARAMETRI: PER IL PARAMETRO $\alpha$ VALUTO LA POTENZA, OVVERO $\Pr\{R_\alpha \alpha\}$ PER IL PARAMETRO $\beta$ VALUTO LA POTENZA, OVVERO $\Pr\{R_\beta \beta\}$ PER IL PARAMETRO $\gamma$ VALUTO IL LIVELLO DI SIGNIFICATIVITA', OVVERO $\Pr\{R_\gamma \gamma=0\}$			
VALORI OTTENUTI STIMANDO IL MODELLO SENZA ERRORE DI MISURA ( $P = 1$ ): $\alpha$ : 1 $\beta$ : 0.966 $\gamma$ : 0.043			
CON ERRORE DI MISURA:	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
$P = 0.5$	$\alpha$ : 1 $\beta$ : 0.977 $\gamma$ : 0.045	$\alpha$ : 1 $\beta$ : 0.955 $\gamma$ : 0.043	$\alpha$ : 1 $\beta$ : 0.941 $\gamma$ : 0.061
$P = 0.4$	$\alpha$ : 1 $\beta$ : 0.968 $\gamma$ : 0.048	$\alpha$ : 1 $\beta$ : 0.966 $\gamma$ : 0.049	$\alpha$ : 1 $\beta$ : 0.933 $\gamma$ : 0.065
$P = 0.3$	$\alpha$ : 1 $\beta$ : 0.956 $\gamma$ : 0.039	$\alpha$ : 1 $\beta$ : 0.948 $\gamma$ : 0.046	$\alpha$ : 1 $\beta$ : 0.928 $\gamma$ : 0.051
$P = 0.2$	$\alpha$ : 1 $\beta$ : 0.970 $\gamma$ : 0.056	$\alpha$ : 1 $\beta$ : 0.951 $\gamma$ : 0.035	$\alpha$ : 1 $\beta$ : 0.909 $\gamma$ : 0.052
$P = 0$	$\alpha$ : 1 $\beta$ : 1 $\gamma$ : 0.040	$\alpha$ : 1 $\beta$ : 1 $\gamma$ : 0.029	$\alpha$ : 1 $\beta$ : 1 $\gamma$ : 0.051

Tabella 3.7: La tabella in alto riporta i valori del livello di copertura reale. Il livello nominale è 0.95. Fra parentesi viene indicato l'intervallo di confidenza per il livello di copertura reale. La tabella in basso riporta i valori dei test statistici effettuati sui parametri, come indicato. Le tabelle si riferiscono entrambe al caso **S0.M1**.

VALORI DELLA DISTORSIONE RELATIVA, CALCOLATA COME $\frac{E(\hat{\theta})-\theta}{\theta}$			
VALORI OTTENUTI STIMANDO IL MODELLO SENZA ERRORE DI MISURA (P = 1): $\alpha$ : 0.0062% $\beta$ : 1.135% $\gamma$ : 2.0384%			
CON ERRORE DI MISURA:	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
P = 0.5	$\alpha$ : 0.07% $\beta$ : 7.01% $\gamma$ : 22.96%	$\alpha$ : 0.09% $\beta$ : 2.03% $\gamma$ : 29.26%	$\alpha$ : 0.099% $\beta$ : 5.764% $\gamma$ : 31.57%
P = 0.4	$\alpha$ : 0.09% $\beta$ : 7.19% $\gamma$ : 28.38%	$\alpha$ : 0.10% $\beta$ : 2.21% $\gamma$ : 33.41%	$\alpha$ : 0.114% $\beta$ : 5.971% $\gamma$ : 36.86%
P = 0.3	$\alpha$ : 0.1% $\beta$ : 9.47% $\gamma$ : 33.88%	$\alpha$ : 0.13% $\beta$ : 4.24% $\gamma$ : 41.76%	$\alpha$ : 0.149% $\beta$ : 4.15% $\gamma$ : 47.71%
P = 0.2	$\alpha$ : 0.12% $\beta$ : 10.85% $\gamma$ : 37.03%	$\alpha$ : 0.15% $\beta$ : 5.42% $\gamma$ : 49.89%	$\alpha$ : 0.165% $\beta$ : 3.97% $\gamma$ : 54.77%
P = 0	$\alpha$ : 0.15% $\beta$ : 14.52% $\gamma$ : 47.95%	$\alpha$ : 0.18% $\beta$ : 6.41% $\gamma$ : 60.76%	$\alpha$ : 0.22% $\beta$ : 2.09% $\gamma$ : 72.65%

VALORI DEL RAPPORTO $\frac{Var(\hat{\theta})}{D(\hat{\theta})^2}$			
VALORI OTTENUTI STIMANDO IL MODELLO SENZA ERRORE DI MISURA (P = 1): $\alpha$ : 413.879 $\beta$ : 513.470 $\gamma$ : 278.585			
CON ERRORE DI MISURA:	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
P = 0.5	$\alpha$ : 2.783 $\beta$ : 12.234 $\gamma$ : 2.172	$\alpha$ : 1.817 $\beta$ : 143.571 $\gamma$ : 1.334	$\alpha$ : 1.4953 $\beta$ : 16.602 $\gamma$ : 1.1824
P = 0.4	$\alpha$ : 1.912 $\beta$ : 11.946 $\gamma$ : 1.407	$\alpha$ : 1.508 $\beta$ : 120.223 $\gamma$ : 1.086	$\alpha$ : 1.0888 $\beta$ : 15.571 $\gamma$ : 0.8750
P = 0.3	$\alpha$ : 1.396 $\beta$ : 6.893 $\gamma$ : 1.028	$\alpha$ : 0.943 $\beta$ : 30.111 $\gamma$ : 0.673	$\alpha$ : 0.7210 $\beta$ : 30.005 $\gamma$ : 0.5492
P = 0.2	$\alpha$ : 1.073 $\beta$ : 4.818 $\gamma$ : 0.850	$\alpha$ : 0.653 $\beta$ : 20.424 $\gamma$ : 0.471	$\alpha$ : 0.498 $\beta$ : 29.878 $\gamma$ : 0.3758
P = 0	$\alpha$ : 0.742 $\beta$ : 2.976 $\gamma$ : 0.537	$\alpha$ : 0.433 $\beta$ : 12.333 $\gamma$ : 0.301	$\alpha$ : 0.2576 $\beta$ : 81.955 $\gamma$ : 0.1914

Tabella 3.8: La tabella in alto riporta i valori della distorsione relativa riferita agli stimatori  $\hat{\alpha}$ ,  $\hat{\beta}$  e  $\hat{\gamma}$ . La tabella in basso riporta i valori del rapporto fra varianza dello stimatore e la sua distorsione al quadrato. Le tabelle si riferiscono entrambe al caso S1.M1.

LIVELLO DI COPERTURA REALE PER INTERVALLI DI CONFIDENZA BILATERALI (IN AGGIUNTA: INTERVALLO DI CONFIDENZA AL 95%). LIVELLO NOMINALE CONSIDERATO: 0.95			
VALORI OTTENUTI STIMANDO IL MODELLO SENZA ERRORE DI MISURA (P = 1): $\alpha$ : 0.950 (0.935 ; 0.962) $\beta$ : 0.951 (0.935 ; 0.963) $\gamma$ : 0.942 (0.925 ; 0.955)			
CON ERRORE DI MISURA:	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
<b>P = 0.5</b>	$\alpha$ : 0.923 (0.904 ; 0.938) $\beta$ : 0.946 (0.930 ; 0.960) $\gamma$ : 0.904 (0.884 ; 0.921)	$\alpha$ : 0.899 (0.878 ; 0.917) $\beta$ : 0.953 (0.937 ; 0.965) $\gamma$ : 0.858 (0.834 ; 0.879)	$\alpha$ : 0.878 (0.856 ; 0.897) $\beta$ : 0.947 (0.931 ; 0.960) $\gamma$ : 0.842 (0.817 ; 0.864)
<b>P = 0.4</b>	$\alpha$ : 0.888 (0.866 ; 0.906) $\beta$ : 0.944 (0.927 ; 0.957) $\gamma$ : 0.879 (0.857 ; 0.898)	$\alpha$ : 0.848 (0.824 ; 0.869) $\beta$ : 0.949 (0.933 ; 0.961) $\gamma$ : 0.825 (0.800 ; 0.850)	$\alpha$ : 0.831 (0.806 ; 0.853) $\beta$ : 0.938 (0.921 ; 0.952) $\gamma$ : 0.767 (0.739 ; 0.793)
<b>P = 0.3</b>	$\alpha$ : 0.864 (0.841 ; 0.884) $\beta$ : 0.926 (0.907 ; 0.941) $\gamma$ : 0.837 (0.812 ; 0.859)	$\alpha$ : 0.812 (0.786 ; 0.835) $\beta$ : 0.941 (0.924 ; 0.954) $\gamma$ : 0.778 (0.751 ; 0.803)	$\alpha$ : 0.776 (0.749 ; 0.801) $\beta$ : 0.947 (0.931 ; 0.960) $\gamma$ : 0.725 (0.695 ; 0.752)
<b>P = 0.2</b>	$\alpha$ : 0.845 (0.821 ; 0.866) $\beta$ : 0.930 (0.912 ; 0.945) $\gamma$ : 0.808 (0.782 ; 0.822)	$\alpha$ : 0.741 (0.712 ; 0.768) $\beta$ : 0.931 (0.913 ; 0.945) $\gamma$ : 0.679 (0.649 ; 0.708)	$\alpha$ : 0.703 (0.673 ; 0.731) $\beta$ : 0.958 (0.943 ; 0.969) $\gamma$ : 0.615 (0.584 ; 0.645)
<b>P = 0</b>	$\alpha$ : 0.761 (0.733 ; 0.787) $\beta$ : 0.905 (0.885 ; 0.922) $\gamma$ : 0.690 (0.660 ; 0.718)	$\alpha$ : 0.670 (0.640 ; 0.699) $\beta$ : 0.940 (0.923 ; 0.953) $\gamma$ : 0.567 (0.535 ; 0.598)	$\alpha$ : 0.562 (0.530 ; 0.593) $\beta$ : 0.974 (0.961 ; 0.982) $\gamma$ : 0.446 (0.415 ; 0.477)

TEST DI SIGNIFICATIVITA' SUI PARAMETRI: PER IL PARAMETRO $\alpha$ VALUTO LA POTENZA, OVVERO $\Pr\{R_\alpha \alpha\}$ PER IL PARAMETRO $\beta$ VALUTO LA POTENZA, OVVERO $\Pr\{R_\beta \beta\}$ PER IL PARAMETRO $\gamma$ VALUTO LA POTENZA, OVVERO $\Pr\{R_\gamma \gamma\}$			
VALORI OTTENUTI STIMANDO IL MODELLO SENZA ERRORE DI MISURA (P = 1): $\alpha$ : 1 $\beta$ : 0.979 $\gamma$ : 0.831			
CON ERRORE DI MISURA:	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
<b>P = 0.5</b>	$\alpha$ : 1 $\beta$ : 0.989 $\gamma$ : 0.600	$\alpha$ : 1 $\beta$ : 0.988 $\gamma$ : 0.541	$\alpha$ : 1 $\beta$ : 0.980 $\gamma$ : 0.530
<b>P = 0.4</b>	$\alpha$ : 1 $\beta$ : 0.989 $\gamma$ : 0.550	$\alpha$ : 1 $\beta$ : 0.986 $\gamma$ : 0.496	$\alpha$ : 1 $\beta$ : 0.987 $\gamma$ : 0.442
<b>P = 0.3</b>	$\alpha$ : 1 $\beta$ : 0.994 $\gamma$ : 0.505	$\alpha$ : 1 $\beta$ : 0.993 $\gamma$ : 0.397	$\alpha$ : 1 $\beta$ : 0.989 $\gamma$ : 0.318
<b>P = 0.2</b>	$\alpha$ : 1 $\beta$ : 0.994 $\gamma$ : 0.455	$\alpha$ : 1 $\beta$ : 0.995 $\gamma$ : 0.318	$\alpha$ : 1 $\beta$ : 0.994 $\gamma$ : 0.267
<b>P = 0</b>	$\alpha$ : 1 $\beta$ : 0.998 $\gamma$ : 0.344	$\alpha$ : 1 $\beta$ : 0.998 $\gamma$ : 0.211	$\alpha$ : 1 $\beta$ : 0.996 $\gamma$ : 0.124

Tabella 3.9: La tabella in alto riporta i valori del livello di copertura reale. Il livello nominale considerato è 0.95. Fra parentesi viene indicato l'intervallo di confidenza per il livello di copertura reale. La tabella in basso riporta i valori dei test statistici effettuati sui parametri, come indicato. Le tabelle si riferiscono entrambe al **caso S1.M1**.

### 3.3 Interpretazione ed implicazioni economiche

#### 3.3.1 Caso S0.M1.

Cominciamo ad analizzare il caso in cui simuliamo sotto l'ipotesi che gli agenti siano razionali e stimiamo invece un modello per agenti irrazionali S0.M1. Simuliamo quindi da un modello del tipo (3.2), che non prevede la presenza di un salto nell'andamento del consumo e invece stimiamo un modello del tipo (3.3), che prevede un salto del consumo. In questo caso, si include nel modello stimato una variabile irrilevante, che non appartiene al vero modello previsto dal processo generatore dei dati. Questa azione non provoca la distorsione delle stime dei parametri ma, piuttosto, porta ad una perdita di efficienza in quanto, l'inclusione di un regressore irrilevante, tende ad aumentare le varianze degli stimatori rispetto a quelle ottenibili stimando il modello vero. Questo comunque vale solo se la variabile esplicativa è  $S^*$  e non la variabile eleggibilità affetta da errore di misura.

Viene fatta eccezione per il caso in cui ci sia incorrelazione fra i regressori presenti nel modello e la variabile irrilevante ma non è il nostro caso. Infatti, stimando il modello in cui la variabile esplicativa eleggibilità non è misurata con errore, ovvero usando  $S^*$  e  $I_{[S^* \geq 0]}$  come esplicative, la distorsione relativa delle stime dei parametri è molto piccola. Non possiamo dire nulla su  $\gamma$  in quanto il valore della distorsione relativa è non definito. Possiamo solo guardare il valore della distorsione assoluta (si rimanda all'Appendice). Il valore della distorsione assoluta per  $\gamma$  è molto basso. Invece, i rapporti fra varianza dello stimatore e distorsione dello stimatore al quadrato per  $\alpha$ ,  $\beta$  e  $\gamma$  sono elevati. Questi valori sono sintomo del fatto che la varianza degli stimatori  $\hat{\alpha}$ ,  $\hat{\beta}$ ,  $\hat{\gamma}$  è molto grande, rispetto alla distorsione. Si vedano le Tabelle (3.6), per lo scenario in cui si stima un modello senza errore di misura.

Riprendiamo ad analizzare i risultati nelle tabelle dove stimiamo il modello utilizzando la variabile esplicativa eleggibilità affetta da errore di misura.

Introducendo l'errore di misura, lo stimatore  $\hat{\alpha}$ , relativo all'intercetta,

non è distorto. Infatti la distorsione relativa è sempre prossima allo 0%. Tale comportamento si evince anche studiando il rapporto fra varianza dello stimatore  $\hat{\alpha}$  e distorsione di  $\hat{\alpha}$  al quadrato. Pur incrementando l'errore di misura, questo rapporto rimane elevato, a dimostrazione del fatto che la distorsione, rispetto alla varianza, è piccola. Si vedano le Tabelle 3.6. Il livello di copertura reale è prossimo al 95%. Si veda la parte superiore della Tabella 3.7.

Per quanto riguarda invece il parametro  $\beta$ , relativo al coefficiente angolare, si può notare che il valore della distorsione relativa aumenta man mano che diminuisce il valore del parametro  $p$  e aumenta il valore di  $\rho$ . Infatti, anche il rapporto fra varianza di  $\hat{\beta}$  e distorsione di  $\hat{\beta}$  al quadrato, tende a diminuire, dimostrando così il peso della distorsione dello stimatore che aumenta all'aumentare della presenza dell'errore di misura ( $p \rightarrow 0$ ). Si vedano le Tabelle 3.6. Il livello di copertura reale per tale parametro è inferiore rispetto a quello nominale, fissato al 95%. Si veda la parte superiore della Tabella 3.7. Quindi, questo problema di non adeguata copertura è da imputare principalmente alla distorsione dello stimatore  $\beta$ , indotta dalla presenza dell'errore di misura. Facciamo inoltre un'altra osservazione. Ricordiamo che, nello studio di simulazione, abbiamo assegnato al parametro  $\beta$  un valore negativo, pari a  $-0.02$ . Il valore della distorsione assoluta ottenuta per questo parametro è positiva (si vedano le tabelle in Appendice). Il parametro  $\beta$  è chiaramente sottostimato. In termini di interpretazione, ciò vuol dire che si verifica un'attenuazione della pendenza della retta che mette in relazione il logaritmo del consumo con i valori dell'eleggibilità.

Per quanto riguarda invece il parametro  $\gamma$ , relativo al salto. Non possiamo calcolare il valore della distorsione relativa in quanto il vero valore di  $\gamma$  è zero. Il rapporto è non definito. Possiamo solo basarci sui valori della distorsione assoluta (si rimanda all'Appendice). La distorsione è comunque piccola. Anche osservando il rapporto fra varianza di  $\hat{\gamma}$  e distorsione di  $\hat{\gamma}$  al quadrato si evince che la distorsione, rispetto alla varianza, è inferiore. Inoltre, non siamo in grado di notare un comportamento direttamente proporzionale fra

aumento della distorsione e incremento dell'errore di misura. Quindi non possiamo dire che la distorsione relativa dello stimatore del parametro  $\gamma$  si incrementa all'aumentare della presenza di errore di misura ( $p \rightarrow 0$ ): non è evidente questa relazione. Si vedano Tabelle 3.6. Il livello di copertura reale per il parametro  $\gamma$  è prossimo al 95%, valore imposto per il livello di copertura nominale. Si veda la parte superiore della Tabella 3.7.

Dai test di significatività riportati nella parte inferiore della Tabella 3.7, si evincono queste considerazioni.

Studiando  $\Pr\{R_\gamma | \gamma = 0\}$ , ovvero la probabilità di concludere erroneamente per la discontinuità nel consumo quando invece, nella realtà, il pensionamento è una decisione anticipata e quindi non si dovrebbe osservare una discontinuità nel consumo, i risultati ottenuti ci inducono a concludere che non c'è discontinuità nel consumo. In questi scenari della simulazione, sia la teoria economica suggerita da Modigliani che il processo generatore dei dati non prevedono la presenza del salto ed effettivamente si conclude che non c'è salto. Nonostante la presenza dell'errore di misura nei dati si conclude per l'assenza del salto, anche se viene inserito nel modello e stimato. Se nella realtà non esiste un salto del consumo, non è da imputare alla presenza dell'errore di misura, il fatto di mostrarlo. Riprendendo la domanda di ricerca dalla quale siamo partiti per questo studio di simulazione, i risultati ci dicono che siamo nel secondo caso, dove l'errore di misura non mi fa vedere un salto quando non c'è. Infatti, se dallo studio di simulazione fosse emersa la questione che la presenza dell'errore di misura riusciva a far evidenziare una discontinuità nel consumo, che nella realtà non esiste, ciò avrebbe significato che l'errore di misura possedeva proprietà diverse a destra e a sinistra del punto di discontinuità  $\bar{s}$ , tali per cui erano in grado di trasformare una relazione lineare in una relazione discontinua.

Un'ulteriore osservazione emerge se si ragiona diversamente. Infatti, se nella realtà non esiste un salto, non è possibile concludere che il pensionamento non appare come un evento pianificabile da parte degli agenti, per effetto dell'errore di misura in quanto, anche introducendo errore di misura,

si è portati a concludere che non c'è una discontinuità nel consumo.

### 3.3.2 Caso S1.M1.

Passiamo ad esaminare il caso in cui simuliamo e stimiamo un modello del tipo (3.3), ovvero ipotizziamo che nella realtà ci sia discontinuità nel consumo e per questo stimiamo un modello che preveda la presenza di un salto nel consumo attorno alla soglia  $\bar{s}$ . Lo scenario al quale facciamo riferimento è del tipo S1.M1.

Abbiamo stimato il modello in cui la variabile esplicativa eleggibilità non è misurata con errore, ovvero usiamo  $S^*$  e  $I_{[S^* \geq 0]}$  come esplicative. Non è presente una forte distorsione delle stime dei parametri. Infatti per  $\alpha$  la distorsione relativa dello stimatore  $\hat{\alpha}$  è pari allo 0.0062%, la distorsione di  $\hat{\beta}$  è pari all'1.13% e di  $\hat{\gamma}$  è pari al 2.04%. Invece, i rapporti fra varianza dello stimatore e distorsione dello stimatore al quadrato per  $\alpha$ ,  $\beta$  e  $\gamma$  sono elevati. Questi valori sono sintomo del fatto che la varianza degli stimatori  $\hat{\alpha}$ ,  $\hat{\beta}$ ,  $\hat{\gamma}$  è più grande della distorsione. Si vedano le Tabelle 3.8, per lo scenario in cui si stima un modello senza errore di misura. Il livello di copertura reale coincide con quello nominale, posto pari al 95%. Si veda la parte superiore della Tabella 3.9, dove si stima senza errore di misura.

Riprendiamo ad analizzare i risultati nelle tabelle dove stimiamo il modello utilizzando la variabile esplicativa eleggibilità affetta da errore di misura. La distorsione delle stime dei parametri  $\alpha$ ,  $\beta$  e  $\gamma$  aumenta e si discosta dalla distorsione evidenziata nello scenario sopra descritto, dove si stima il modello senza errore di misura. In particolar modo per il parametro  $\gamma$ . Per esso è evidente una forte distorsione che tende ad aumentare man mano che diminuisce il valore del parametro  $p$  e aumenta il valore di  $\rho$ . Anche il rapporto fra varianza dello stimatore  $\hat{\gamma}$  e distorsione di  $\hat{\gamma}$  al quadrato evidenzia tale comportamento. Infatti, anche guardando tale indice, si intuisce che la distorsione, in rapporto alla varianza, tende ad aumentare il suo peso all'aumentare della presenza dell'errore di misura ( $p \rightarrow 0$ ) e aumenta il valore di  $\rho$ . Si vedano le Tabelle 3.8. Infine, i valori dei livelli di copertura reali si

discostano molto rispetto a quelli individuati nello scenario di assenza di errore di misura e sono inferiori rispetto a quello nominale, posto pari al 95%. Il motivo è imputabile dunque alla distorsione che induce un errato centramento degli intervalli di confidenza. Si veda la parte superiore della Tabella 3.9.

Per quanto riguarda invece i test di significatività, riportati nella parte inferiore della Tabella 3.9, si evincono queste considerazioni.

Analizziamo la  $\Pr\{R_\gamma|\gamma\}$  ovvero la potenza del test di significatività per il parametro  $\gamma$ . La potenza del test, calcolata in presenza di errore di misura, diminuisce e si discosta molto da quella calcolata nel modello in assenza di errore di misura. In aggiunta, possiamo notare che la potenza diminuisce se aumenta la presenza dell'errore di misura ( $p \rightarrow 0$ ) e aumenta il valore di  $\rho$ . La presenza dell'errore di misura sembra ridurre la potenza del test di significatività per il parametro relativo al salto. Notiamo invece che le potenze dei test di significatività per i parametri  $\alpha$  e  $\beta$  rimangono elevate, molto vicine a 1. La presenza dell'errore di misura sembra non intaccare la potenza del test di significatività per i parametri relativi all'intercetta e al coefficiente angolare.

Dal punto di vista dell'interpretazione economica, vuol dire che la presenza dell'errore di misura porta a concludere che non c'è discontinuità nel consumo, quando in effetti c'è. Infatti, anche stimando un modello dove viene ipotizzata la presenza di un salto del consumo, la potenza è talmente bassa che, nella maggior parte dei casi, il coefficiente  $\gamma$  risulta essere non significativo.

In letteratura esiste un unico lavoro che mette in evidenza una discontinuità nel consumo. Si fa riferimento al lavoro [2]. Possiamo pensare ad un caso particolare oppure al fatto che il pensionamento non sia un evento pianificato ma che l'errore di misura sia fortemente presente nei dati a disposizione, tale da essere in grado di abbassare la potenza del test di significatività sul coefficiente  $\gamma$ , relativo al salto, facendo quindi concludere a favore del fatto che non ci sarebbe una discontinuità nel consumo. Per certi

valori della potenza riportati in tabella, saremmo spinti a dire che riusciamo, anche in presenza di un salto significativo, ad evidenziarlo in media solo una volta su dieci. Si veda, per esempio, il caso in cui  $p = 0$  e  $\rho = 0.3$ . E' come se dicessimo: consideriamo 10 lavori in ambito internazionale, ovunque c'è la presenza del salto, ma solo un lavoro riesce a dare evidenza della presenza del salto. Il motivo di ciò è che l'errore di misura agisce come confondimento e riduce la potenza del test di significatività sul coefficiente  $\gamma$ , ovvero riduce di molto la capacità di individuare l'effetto presente. Riprendendo la domanda di ricerca dalla quale siamo partiti per questo studio di simulazione, anche questi risultati ci dicono che siamo nel secondo caso, dove l'errore di misura non mi fa vedere il salto, anche quando esso è presente.

In altre parole, se gli individui non operano secondo quanto previsto dalla teoria economica, arrivando quindi impreparati al momento del pensionamento, si dovrebbe osservare un salto ma, per effetto dell'errore di misura presente nei dati, si è portati a concludere che il pensionamento è un evento pianificabile e che quindi gli agenti agiscono secondo quanto previsto dalla teoria economica.

## Capitolo 4

# Un passo in avanti: correzione tramite SIMEX

In questo capitolo, consideriamo il metodo SIMEX per la correzione dell'errore di misura. Tra i vari metodi citati in letteratura, il SIMEX, proposto da Cook e Stefanski (si veda [21] per maggiori dettagli), è diventato uno strumento valido per correggere la stima del parametro d'interesse, in presenza di errore di misura di tipo classico. E' un metodo basato sulla simulazione, in grado di ridurre la distorsione dovuta all'errore di misura.

L'idea alla base del metodo è stata trasferita anche al caso in cui le variabili affette da errore di misura siano di tipo qualitativo. Il metodo MC-SIMEX è stato proposto da Carroll, Küchenhoff, Lombard e Stefanski (si veda [22] per maggiori dettagli). Per la descrizione dettagliata del metodo si rimanda al Paragrafo 2.3.3., del Capitolo 2.

Le simulazioni, che a breve descriveremo, sono state eseguite in R, utilizzando il pacchetto SIMEX.

### 4.1 Il metodo SIMEX applicato al caso S0.M0.

Se una variabile esplicativa è affetta da errore di misura ed è continua, si può utilizzare il metodo SIMEX per correggere la presenza dell'errore di misura e

ottenere stime più attendibili. In questo paragrafo applicheremo tale metodo al caso S0.M0., quando  $\rho = 0.2$  e  $p = 0$ , ovvero in presenza di errore di misura di tipo classico. Calcoleremo i valori degli indici descritti nel Capitolo 3.

Simuliamo quindi dal modello

$$E[\log C|S^*] = \alpha + \beta S^* \quad (4.1)$$

e stimiamo, con il metodo dei minimi quadrati ordinari, il modello

$$E[\log C] = \alpha + \beta S_{\text{classico}}, \quad (4.2)$$

dove  $S_{\text{classico}} = S^* + U$  e  $U \sim N(0, \sigma_u^2)$ .

Il metodo SIMEX richiede che venga specificata la variabile affetta da errore di misura. In questo caso è  $S_{\text{classico}}$ . Inoltre viene richiesto il valore dello scarto quadratico medio della varianza dell'errore di misura  $U$ . In questo caso, poichè abbiamo definito  $\rho$  come nella (3.6), allora

$$\sigma_u = \sqrt{\rho \text{Var}(S_{\text{classico}})}. \quad (4.3)$$

In questo caso,  $\sigma_u$ , è supposto noto. Nella simulazione abbiamo utilizzato un numero di replicazioni pari a 200.

Nella Tabella 4.1 sono riportati i valori degli indici calcolati. Il metodo si adatta bene a questo caso. La distorsione della stima relativa ai parametri  $\alpha$  e  $\beta$  è piccola, la potenza del test (3.15) è pari ad 1 e il livello di copertura reale si avvicina al valore scelto per il livello di copertura nominale.

Lo svantaggio del SIMEX consiste in tempi di calcolo molto lunghi. L'applicazione del metodo non è immediata. La difficoltà principale infatti è legata all'implementazione dell'algoritmo nel calcolatore.

<b>MODELLO STIMATO CON I MINIMI QUADRATI ORDINARI:</b>	
<b><math>E[\log C] = \alpha + \beta S_{\text{CLASSICO}}</math></b>	
VARIABILE MISURATA CON ERRORE:	$S_{\text{CLASSICO}}$
SCARTO QUADRATICO MEDIO DELL'ERRORE DI MISURA:	$\sigma_u = \sqrt{\rho * \text{Var}(S_{\text{CLASSICO}})}$
<b>INDICI CALCOLATI:</b>	
$D(\hat{\theta}) = E(\hat{\theta}) - \theta$	$\alpha: 0.000066$ $\beta: 0.000426$
$\frac{E(\hat{\theta}) - \theta}{\theta} * 100$	$\alpha: 0.00067 \%$ $\beta: 2.1323\%$
$EQM(\hat{\theta}) = \text{Var}(\hat{\theta}) + D(\hat{\theta})^2$	$\alpha: 0.0000436$ $\beta: 0.0000125$
$\frac{\text{Var}(\hat{\theta})}{D(\hat{\theta})^2}$	$\alpha: 9796.4465$ $\beta: 67.9511$
Livello di copertura reale per intervalli di confidenza bilaterali. Livello nominale considerato: 0.95	$\alpha: 0.950 [0.906 ; 0.976]$ $\beta: 0.940 [0.895 ; 0.967]$
Livello di copertura reale per intervalli di confidenza unilaterali. Livello nominale considerato: 0.95	$\alpha: 0.950 \quad 0.950$ $\beta: 0.955 \quad 0.925$
Potenza del test di significatività, ovvero $\Pr\{R_\theta   \theta\}$	$\alpha: 1$ $\beta: 1$

Tabella 4.1: La tabella riporta i valori degli indici calcolati per gli stimatori  $\hat{\alpha}$  e  $\hat{\beta}$ . Risultati di simulazione basati su 200 replicazioni, con numerosità  $n=8000$ . Metodo SIMEX, dove la variabile affetta da errore di misura è  $S_{\text{classico}}$  e come scarto quadratico medio per l'errore di misura abbiamo usato  $\sigma_u$  noto.

## 4.2 Il metodo SIMEX applicato al caso S1.M1.

Nel contesto di questo lavoro, i casi ai quali siamo maggiormente interessati sono quelli del tipo S0.M1. e S1.M1. In entrambi i casi, nel modello stimato, sono presenti due variabili affette da errore di misura, di cui una è continua ( $S_{\text{classico}}$ ) e l'altra è discreta ( $I_{[S_{\text{classico}} \geq 0]}$ ).

Il metodo SIMEX si adatta bene al caso in cui, nel modello stimato, le variabili esplicative siano continue oppure discrete. Per l'appunto:

- se la variabile esplicativa, affetta da errore di misura, è continua allora il metodo SIMEX funziona bene, come abbiamo mostrato nel paragrafo precedente.
- se la variabile esplicativa, affetta da errore di misura, è discreta allora viene trattata come una variabile di tipo qualitativo, affetta da un'errata classificazione. Occorre utilizzare il metodo MC-SIMEX, ovvero il metodo SIMEX adattato al caso di errata classificazione.

Si veda l'articolo proposto da Küchenhoff e Lederer [42] per maggiori dettagli. In questo articolo gli autori ipotizzano che il metodo SIMEX possa anche essere applicato alla situazione in cui una variabile esplicativa, misurata con errore, è discreta e un'altra variabile esplicativa ha un errore di misura classico. Ma affermano che non esiste un pacchetto disponibile in R che consenta di applicare il metodo SIMEX ad un modello di questo genere.

A questo punto, la domanda di ricerca che è sorta, è stata quella di provare a capire cosa succederebbe se dovessimo applicare il metodo SIMEX ad un modello stimato del tipo M1.

Prendiamo quindi il caso S1.M1., in cui simuliamo dal modello

$$E[\log C|S^*] = \alpha + \beta S^* + \gamma I_{[S^* \geq 0]} \quad (4.4)$$

e stimiamo, con il metodo dei minimi quadrati ordinari, il modello

$$E[\log C] = \alpha + \beta S_{\text{classico}} + \gamma I_{[S_{\text{classico}} \geq 0]}. \quad (4.5)$$

Specifichiamo che la variabile misurata con errore è una sola, per l'ap-punto  $S_{\text{classico}}$ . Separiamo due casi a seconda del valore usato per  $\sigma_U$ , ovvero per lo scarto quadratico medio della varianza dell'errore di misura  $U$ . Infatti, nel primo caso, utilizziamo il valore teorico della varianza di  $S_{\text{classico}}$ , moltiplicata per  $\rho = 0.2$ , come nella (4.3). Nel secondo caso, facciamo lo stesso, utilizzando però la varianza campionaria di  $S_{\text{classico}}$ , moltiplicata sempre per  $\rho$ .

Come possiamo evincere dalla Tabella 4.2, il metodo SIMEX, non fornisce risultati soddisfacenti, come era prevedibile. Stiamo infatti specificando che la variabile misurata con errore è solamente  $S_{\text{classico}}$ , mentre nel modello stimato c'è anche  $I_{[S_{\text{classico}} \geq 0]}$ , anche questa misurata con errore. Il fatto di ignorarla peggiora molto la situazione. La distorsione delle stime relative ai parametri  $\beta$  e  $\gamma$  è molto forte e il livello di copertura reale si discosta molto dal livello di copertura nominale, fissato a 0.95. Anche la potenza del test (3.15) per il parametro  $\gamma$  è molto bassa.

In queste due simulazioni abbiamo utilizzato un numero di replicazioni pari a 200.

Lo svantaggio del SIMEX consiste in tempi di calcolo molto lunghi. L'applicazione del metodo non è immediata. La difficoltà principale infatti è legata all'implementazione dell'algoritmo nel calcolatore.

86CAPITOLO 4. UN PASSO IN AVANTI: CORREZIONE TRAMITE SIMEX

<b>MODELLO STIMATO CON I MINIMI QUADRATI ORDINARI:</b> $E[\log C] = \alpha + \beta S_{\text{CLASSICO}} + \gamma I_{S, \text{CLASSICO}} \geq 0$		
VARIABILE MISURATA CON ERRORE:	$S_{\text{CLASSICO}}$	$S_{\text{CLASSICO}}$
SCARTO QUADRATICO MEDIO DELL'ERRORE DI MISURA:	$\sigma_u = \sqrt{\rho * \text{Var}(S_{\text{CLASSICO}})}$	$\sigma_u = \sqrt{\rho * \widehat{\text{Var}}(S_{\text{CLASSICO}})}$
<b>INDICI CALCOLATI:</b>		
$D(\hat{\theta}) = E(\hat{\theta}) - \theta$	$\alpha$ : -0.037817 $\beta$ : -0.013555 $\gamma$ : 0.077444	$\alpha$ : -0.038290 $\beta$ : -0.012421 $\gamma$ : 0.076728
$\frac{E(\hat{\theta}) - \theta}{\theta} * 100$	$\alpha$ : -0.3875% $\beta$ : +67.773% $\gamma$ : -129.073%	$\alpha$ : -0.392% $\beta$ : 62.104% $\gamma$ : -127.879%
$EQM(\hat{\theta}) = \text{Var}(\hat{\theta}) + D(\hat{\theta})^2$	$\alpha$ : 0.00162 $\beta$ : 0.00023 $\gamma$ : 0.0067	$\alpha$ : 0.00168 $\beta$ : 0.00020 $\gamma$ : 0.00663
$\frac{\text{Var}(\hat{\theta})}{D(\hat{\theta})^2}$	$\alpha$ : 0.13274 $\beta$ : 0.25450 $\gamma$ : 0.12003	$\alpha$ : 0.14508 $\beta$ : 0.31748 $\gamma$ : 0.12659
Livello di copertura reale per intervalli di confidenza bilaterali. Livello nominale considerato: 0.95	$\alpha$ : 0.24 [0.163 ; 0.338] $\beta$ : 0.45 [0.351 ; 0.553] $\gamma$ : 0.16 [0.097 ; 0.249]	$\alpha$ : 0.22 [0.146 ; 0.316] $\beta$ : 0.49 [0.389 ; 0.591] $\gamma$ : 0.18 [0.113 ; 0.272]
Livello di copertura reale per intervalli di confidenza unilaterali. Livello nominale considerato: 0.95	$\alpha$ : 0.16 1.00 $\beta$ : 0.35 1.00 $\gamma$ : 1.00 0.07	$\alpha$ : 0.15 1.00 $\beta$ : 0.37 1.00 $\gamma$ : 1.00 0.08
Potenza del test di significatività, ovvero $\Pr\{R_{\theta}   \theta\}$	$\alpha$ : 1 $\beta$ : 0.99 $\gamma$ : 0.1	$\alpha$ : 1 $\beta$ : 0.99 $\gamma$ : 0.09

Tabella 4.2: La tabella riporta i valori degli indici calcolati per gli stimatori  $\hat{\alpha}$ ,  $\hat{\beta}$  e  $\hat{\gamma}$ . Risultati di simulazione basati su 200 replicazioni, con numerosità  $n=8000$ . Metodo SIMEX, dove la variabile affetta da errore di misura è  $S_{\text{classico}}$  e come scarto quadratico medio per l'errore di misura abbiamo usato  $\sigma_u$  noto e stimato.

### 4.3 Una proposta alternativa per il caso S1.M1.

A quanto ci risulta non esiste un pacchetto in R che consenta di applicare il metodo SIMEX ad un modello con due variabili esplicative, entrambe affette da errore di misura, di cui una è continua e l'altra è discreta.

La nostra proposta è la seguente. Specifichiamo che le variabili affette da errore di misura sono due:  $S_{\text{classico}}$ , variabile continua e  $I_{[S_{\text{classico}} \geq 0]}$ , variabile discreta. Anche se la variabile  $I_{[S_{\text{classico}} \geq 0]}$  è una Bernoulli di parametro  $\pi$ , che assegna valore 1 se il valore dell'eleggibilità è positivo e 0 altrimenti, la trattiamo come una variabile continua. Per questo motivo utilizziamo il metodo SIMEX e non il metodo MC-SIMEX. Siamo consapevoli che questa è una forzatura del metodo SIMEX. Il nostro obiettivo è trovare uno strumento che ci consenta di fare un compromesso accettabile tra rigore del metodo e semplicità d'uso.

Lavoriamo in parallelo come abbiamo fatto nel paragrafo precedente, impiegando da una parte quantità teoriche e dall'altro quantità empiriche.

Il primo caso è il seguente. Per la variabile continua utilizziamo come  $\sigma_u$  la radice quadrata del prodotto fra  $\rho$  e il valore teorico della varianza di  $S_{\text{classico}}$ , come nella (4.3). Per la variabile discreta utilizziamo il valore teorico della probabilità di successo  $\pi$ . Dunque

$$\pi = \Pr(I_{[S_{\text{vero}} \geq 0]} = 1) = \Pr(S^* \geq 0) = 1/2 \quad (4.6)$$

dal momento che la variabile  $S^*$  è una Normale, centrata in zero. La varianza della variabile  $I_{[S_{\text{vero}} \geq 0]}$  è  $\pi(1 - \pi) = 1/4$ . La moltiplichiamo per  $\rho$  e ne estraiamo la radice. Poichè  $\rho$  è il valore che determina il rapporto fra la varianza del termine d'errore  $U$  e la varianza della variabile affetta da errore di misura  $S_{\text{classico}}$ , dalla varianza della variabile  $I_{[S_{\text{vero}} \geq 0]}$ , tramite  $\rho$ , recuperiamo un'approssimazione euristica per la varianza di questa variabile binaria affetta da errore.

Il secondo caso è il seguente. Per la variabile continua utilizziamo la varianza stimata di  $S_{\text{classico}}$ , moltiplicata sempre per  $\rho$ , ed estraiamo la radice per ottenere lo scarto quadratico medio. Per la variabile discreta usiamo

$\hat{\pi}$ , ovvero stimiamo il parametro  $\pi$  con la proporzione campionaria. Recuperiamo la varianza campionaria delle variabili discrete affette da errore di misura e sfruttiamo, come prima, la conoscenza su  $\rho$ . Estraiamo quindi la radice per ottenere lo scarto quadratico medio.

In Tabella 4.3 riportiamo i valori degli indici per questi due casi, riferiti agli stimatori dei parametri  $\alpha$ ,  $\beta$  e  $\gamma$ .

Come si può notare, facendo il confronto con i risultati riportati in Tabella 4.2, la situazione è nettamente migliorata. La distorsione della stima relativa ai parametri  $\beta$  e  $\gamma$  si riduce molto. Se prima il valore del livello di copertura reale era molto diverso da quello nominale, con questo metodo il livello di copertura reale si scosta meno da quello nominale, fissato al 95%. Anche il valore della potenza del test di significatività (3.15) migliora.

In queste due simulazioni abbiamo utilizzato un numero di replicazioni pari a 200.

Lo svantaggio del SIMEX consiste in tempi di calcolo molto lunghi. L'applicazione del metodo non è immediata. La difficoltà principale infatti è legata all'implementazione dell'algoritmo nel calcolatore.

<b>MODELLO STIMATO CON I MINIMI QUADRATI ORDINARI:</b> $E[\log C] = \alpha + \beta S_{\text{CLASSICO}} + \gamma I_{S_{\text{CLASSICO}} \geq 0}$		
VARIABILE MISURATA CON ERRORE:	$S_{\text{CLASSICO}}$ $I_{S_{\text{CLASSICO}} \geq 0}$	$S_{\text{CLASSICO}}$ $I_{S_{\text{CLASSICO}} \geq 0}$
SCARTO QUADRATICO MEDIO DELL'ERRORE DI MISURA:	$\sigma_u = \sqrt{\rho * VAR(S_{\text{CLASSICO}})}$ $\sigma_i = \sqrt{\rho * z * (1 - z)}$ dove $z=0.5$ valore teorico	$\sigma_u = \sqrt{\rho * \widehat{VAR}(S_{\text{CLASSICO}})}$ $\sigma_i = \sqrt{\rho * \hat{z} * (1 - \hat{z})}$ dove $\hat{z}$ valore empirico
<b>INDICI CALCOLATI:</b>		
$D(\hat{\theta}) = E(\hat{\theta}) - \theta$	$\alpha$ : -0.010525 $\beta$ : -0.004543 $\gamma$ : +0.016890	$\alpha$ : -0.011813 $\beta$ : -0.004511 $\gamma$ : +0.018873
$\frac{E(\hat{\theta}) - \theta}{\theta} * 100$	$\alpha$ : - 0.1078 % $\beta$ : 22.717% $\gamma$ : -28.150 %	$\alpha$ : -0.1210 % $\beta$ : 22.553 % $\gamma$ : -31.455 %
$EQM(\hat{\theta}) = Var(\hat{\theta}) + D(\hat{\theta})^2$	$\alpha$ : 0.00026 $\beta$ : 0.00004 $\gamma$ : 0.00100	$\alpha$ : 0.00026 $\beta$ : 0.00004 $\gamma$ : 0.00100
$\frac{Var(\hat{\theta})}{D(\hat{\theta})^2}$	$\alpha$ : 1.38209 $\beta$ : 1.11898 $\gamma$ : 0.96246	$\alpha$ : 1.38209 $\beta$ : 1.11898 $\gamma$ : 0.96246
Livello di copertura reale per intervalli di confidenza bilaterali. Livello nominale considerato: 0.95	$\alpha$ : 0.815 [0.753 ; 0.865] $\beta$ : 0.860 [0.802 ; 0.904] $\gamma$ : 0.865 [0.808 ; 0.926]	$\alpha$ : 0.865 [0.808 ; 0.908] $\beta$ : 0.880 [0.832 ; 0.933] $\gamma$ : 0.855 [0.791 ; 0.907]
Livello di copertura reale per intervalli di confidenza unilaterali. Livello nominale considerato: 0.95	$\alpha$ : 0.755 0.995 $\beta$ : 0.715 1.000 $\gamma$ : 1.000 0.710	$\alpha$ : 0.740 1.000 $\beta$ : 0.722 1.000 $\gamma$ : 1.000 0.675
Potenza del test di significatività, ovvero $Pr\{R_0   \theta\}$	$\alpha$ : 1 $\beta$ : 1 $\gamma$ : 0.7	$\alpha$ : 1 $\beta$ : 1 $\gamma$ : 0.672

Tabella 4.3: La tabella riporta i valori degli indici calcolati per gli stimatori  $\hat{\alpha}$ ,  $\hat{\beta}$  e  $\hat{\gamma}$ . Risultati di simulazione basati su 200 replicazioni, con numerosità  $n=8000$ . Metodo SIMEX, dove le variabili affette da errore di misura sono  $S_{\text{classico}}$  e  $I_{[S_{\text{classico}} \geq 0]}$ . Come scarto quadratico medio dell'errore di misura abbiamo usato  $\sigma_u$  noto e stimato e  $\sigma_i$  noto e stimato come descritto nel paragrafo.



# Capitolo 5

## Conclusioni e sviluppi futuri

### 5.1 Conclusioni tratte dallo studio di simulazione

Riprendendo la domanda di ricerca che ci ha spinto ad effettuare lo studio di simulazione, i risultati ottenuti nei casi S0.M1 e S1.M1, ci indicano che, tra le due opzioni che avevamo preventivato, quella che è risultata vera è la seconda.

Infatti, lavorando sotto l'ipotesi che il pensionamento sia un evento pianificato nella vita degli agenti e dunque non ci dovrebbe essere alcuna discontinuità nel consumo, anche se si lavora in presenza di errore di misura nelle variabili esplicative del modello, ciò non porta ad osservare una discontinuità nel consumo.

Lavorando invece sotto l'ipotesi che gli agenti arrivino impreparati al pensionamento, per effetto dell'errore di misura, non si riesce a dare evidenza della discontinuità nel consumo. Abbiamo capito che è un problema legato alla potenza. La potenza, a causa della presenza dell'errore di misura, è così bassa che, nella maggior parte dei casi, siamo portati a concludere che non c'è discontinuità nel consumo quando in effetti questa è presente. Per certi valori della potenza è come se dicessimo: consideriamo 10 lavori in ambito internazionale, ovunque c'è la presenza del salto, ma solo un lavoro riesce a

dare evidenza della discontinuità. Il motivo di ciò è che l'errore di misura agisce come confondimento e riduce la potenza del test.

## 5.2 Sviluppi futuri

Come abbiamo chiarito nel Paragrafo 4.3, il metodo suggerito è una forzatura del metodo SIMEX.

Il passo successivo sarà quello di fornire un algoritmo che tenga conto di una variabile continua e di una variabile discreta, entrambe affette da errore di misura. In altre parole, un strumento che lavori su due fronti: da un lato in grado di correggere un errore di misura di tipo classico e dall'altro un'errata classificazione, possibilmente integrando SIMEX e MC-SIMEX.

# Appendice A

## Tabelle complete dello studio di simulazione

In questa Appendice riportiamo le tabelle relative allo studio di simulazione al completo. Nel Capitolo 3 abbiamo riportato solamente quelle più interessanti dal punto di vista dell'interpretazione ma per avere un quadro completo dello studio è bene fare riferimento all'intera lista delle tabelle.

Le tabelle contengono i valori per gli indici che sono stati calcolati come descritto nel Paragrafo 3.1.5.

Le tabelle sono raggruppate per indice calcolato. L'ordine di presentazione è il seguente:

1. Distorsione assoluta;
2. Distorsione relativa
3. Errore quadratico medio
4. Rapporto fra varianza dello stimatore e distorsione dello stimatore, elevato al quadrato
5. Livello di copertura reale per intervalli di confidenza bilaterali. Fra parentesi viene riportato l'intervallo di confidenza per il medesimo.
6. Livello di copertura reale per intervalli di confidenza unilaterali.

7. Test di significatività, che assumono significato preciso a seconda dello scenario considerato. Ciascuna tabella riporta esplicitamente il test di significatività che è stato considerato.

# Appendice B

## Stima della varianza

Supponiamo che siano disponibili  $N$  osservazioni del tipo  $(y_i, x_i)$ , con  $i = 1, \dots, N$ , provenienti dalla coppia di variabili casuali  $(Y, X)$ , dove  $Y$  è la variabile risposta e  $X$  la variabile esplicativa. Sia  $\hat{\beta}$  lo stimatore del parametro  $\beta$ , ovvero il parametro del modello che lega  $Y$  a  $X$ . In questo paragrafo vengono descritti i metodi che comunemente vengono utilizzati per stimare gli errori standard dello stimatore  $\hat{\beta}$ . Si può ottenere una stima non parametrica degli errori standard avvalendosi del metodo jackknife o del metodo bootstrap. Inoltre, si può utilizzare il metodo sandwich.

### B.1 Il metodo jackknife

L'idea generale sottesa è quella di formare diversi sotto campioni di un generico campione estratto eliminando di volta in volta una singola osservazione o, nel caso di disegni che prevedono un'elevata numerosità, una piccola parte del campione originario. Il metodo jackknife comincia quindi selezionando il campione jackknife. L' $i$ -esimo campione jackknife, con  $i = 1, \dots, N$  contiene  $N - 1$  osservazioni ed è ottenuto dal campione originario, omettendo la  $i$ -esima osservazione. Ciascun sottocampione misurerà così il contributo alla variabilità totale fornita dalla piccola parte eliminata dalla stima: la procedura jackknife calcola infatti le stime dei parametri di interesse su ciascuna

replica ed utilizza la variabilità tra tali valori per ottenere una stima della variabilità campionaria totale.

Sia  $\hat{\beta}_{-i}$  lo stimatore di  $\beta$  ottenuto riducendo il campione iniziale. La varianza stimata dello stimatore è

$$\hat{V}ar(\hat{\beta}) = \frac{N-1}{N} \sum_{i=1}^N (\hat{\beta}_{-i} - \hat{\beta})^2, \quad (\text{B.1})$$

dove  $\hat{\beta} = \frac{1}{N} \sum_{i=1}^N \hat{\beta}_{-i}$ . Il metodo risulta poco efficace quando si dispone di strumenti specifici, più solidi e funzionali. Il termine è stato scelto poiché il metodo ha un'applicazione appropriata quando non è possibile utilizzare i metodi classici.

## B.2 Il metodo bootstrap

Anche questo metodo si basa sulla costruzione di campioni appropriati. Un campione bootstrap si ottiene estraendo casualmente  $n$  elementi con ripetizione dagli  $N$  dati campionari. Tale campione viene usato per ottenere la stima  $\hat{\beta}_b$  di  $\beta$ . Questa procedura viene ripetuta per un numero di volte pari a  $B$  (solitamente  $B = 50, 100$ ). La varianza stimata dello stimatore è

$$\hat{V}ar(\hat{\beta}) = \frac{1}{B-1} \sum_{b=1}^B (\hat{\beta}_b - \hat{\beta})^2, \quad (\text{B.2})$$

dove  $\hat{\beta} = \frac{1}{B} \sum_{b=1}^B \hat{\beta}_b$ .

## B.3 Il metodo sandwich

Sia  $\Psi_i(\beta)$  la funzione di stima per il parametro  $\beta$  dell' $i$ -esima osservazione,  $i = 1, \dots, N$ . Inoltre, sia  $\hat{\beta}$  lo stimatore per il parametro  $\beta$ , derivante dalla relazione  $\sum_{i=1}^N \Psi_i(\beta) = 0$ . Lo stimatore sandwich della varianza di  $\hat{\beta}$  è

$$\hat{V}ar(\hat{\beta}) = A_N^{-1}(\hat{\beta}) B_N(\hat{\beta}) (A_N^{-1})^T(\hat{\beta}), \quad (\text{B.3})$$

dove

$$A_N(\hat{\beta}) = \sum_{i=1}^N \frac{\partial \Psi_i(\beta)}{\partial \beta} \quad (\text{B.4})$$

e

$$B_N(\hat{\beta}) = \sum_{i=1}^N \Psi_i(\beta) \Psi_i(\beta)^T. \quad (\text{B.5})$$

Il metodo sandwich fornisce uno stimatore consistente della varianza. Inoltre, usa quantità empiriche per stimare la varianza di  $\hat{\beta}$ . In questo senso è robusto.

# Appendice C

## Istruzioni di R per lo studio di simulazione

Vengono riportate le istruzioni in R (versione 2.11.1 per Microsoft Windows), usate nello studio di simulazione, come descritto nel Capitolo 3. Non riportiamo le istruzioni usate per tutte le 64 simulazioni ma solamente le istruzioni per alcuni scenari, a titolo d'esempio.

### Definizione ed inizializzazione dei parametri della simulazione

```
>n = 8000
>B = 1000
>media.s.vero = 0
>sd.s.vero = 2
>var.s.vero = sd.s.vero2
>alfa = 9.76
>beta = -0.02
>gamma = -0.06
```

```
>sd.logc = 0.55
```

## Caso S0.M1. in assenza di errore di misura

```
>ris.sim.M1 = matrix(NA, nrow=R, ncol=10)
>dimnames(ris.sim.H1) = list(paste(1:1000),
c(alfa,beta,gamma,sd.alfa,sd.beta,sd.gamma,p.alfa,p.beta,p.gamma,sigma))
>set.seed(67547715)
>for(idx in 1:1000)
{
s.vero = rnorm(n, mean=media.s.vero, sd=sd.s.vero)
logc = rnorm(n, mean = alfa + beta*s.vero, sd=sd.logc)
modello.stimato = lm(logc ~ s.vero + I.s.vero)
ris.tmp = c(as.vector(summary(modello.stimato)$coef[,c(1:2,4)]),
summary(modello.stimato)$sigma)
ris.sim.M1[idx,] = ris.tmp
print(idx)
}
>risultati.S0.M1.no.errore = ris.sim.H1
>distorsione.S0.M1.no.errore = apply(risultati.S0.M1.no.errore[,1:3], 2, mean)
- c(alfa, beta,0)
>relativa.alfa.tmp = (distorsione.S0.M1.no.errore [1]/alfa)*100
>relativa.beta.tmp = (distorsione.S0.M1.no.errore [2]/beta)*100
>distorsione.relativa.S0.M1.no.errore = c(relativa.alfa.tmp,
relativa.beta.tmp, relativa.gamma.tmp)
>varianza.S0.M1.no.errore = apply(risultati.S0.M1.no.errore[,1:3], 2, var)
>eqm.S0.M1.no.errore = varianza.S0.M1.no.errore + distorsione.S0.M1.no.errore2
>varianza.S0.M1.no.errore/distorsione.S0.M1.no.errore2
>test.alfa.tmp=(risultati.S0.M1.no.errore[,1]-alfa)/risultati.S0.M1.no.errore[,4]
>p.alfa.tmp = 2 * pmin(pnorm(test.alfa.tmp), 1-pnorm(test.alfa.tmp))
```

```

>conf.alfa.tmp = ifelse(p.alfa.tmp >= 0.05, 1, 0)
>test.beta.tmp = (risultati.S0.M1.no.errore[,2]-beta)/risultati.S0.M1.no.errore[,5]
>p.beta.tmp = 2 * pmin(pnorm(test.beta.tmp), 1-pnorm(test.beta.tmp))
>conf.beta.tmp = ifelse(p.beta.tmp >= 0.05, 1, 0)
>test.gamma.tmp = (risultati.S0.M1.no.errore[,3]-0)/risultati.S0.M1.no.errore[,6]
>p.gamma.tmp = 2 * pmin(pnorm(test.gamma.tmp), 1-pnorm(test.gamma.tmp))
>conf.gamma.tmp = ifelse(p.gamma.tmp >= 0.05, 1, 0)
>prop.test(sum(conf.alfa.tmp), R, p=0.95)
>prop.test(sum(conf.beta.tmp), R, p=0.95)
>prop.test(sum(conf.gamma.tmp), R, p=0.95)
>test.alfa.tmp=(risultati.S0.M1.no.errore[,1]-alfa)/risultati.S0.M1.no.errore[,4]
>p.alfa.unilaterale.tmp = cbind(pnorm(test.alfa.tmp), 1-pnorm(test.alfa.tmp))
>conf.alfa.unilaterale.tmp = apply(p.alfa.uni.tmp, 2, function(x) ifelse(x>=0.05,
1, 0))
>apply(conf.alfa.unilaterale.tmp, 2, mean)
>test.beta.tmp=(risultati.S0.M1.no.errore[,2]-beta)/risultati.S0.M1.no.errore[,5]
>p.beta.unilaterale.tmp = cbind(pnorm(test.beta.tmp), 1-pnorm(test.beta.tmp))
>conf.beta.unilaterale.tmp = apply(p.beta.unilaterale.tmp, 2, function(x) ifelse(x>=0.05,
1, 0))
>apply(conf.beta.unilaterale.tmp, 2, mean)
>test.gamma.tmp=(risultati.S0.M1.no.errore[,3]-0)/risultati.S0.M1.no.errore[,6]
>p.gamma.unilaterale.tmp = cbind(pnorm(test.gamma.tmp), 1-pnorm(test.gamma.tmp))
>conf.gamma.unilaterale.tmp = apply(p.gamma.unilaterale.tmp, 2, function(x)
ifelse(x>=0.05, 1, 0))
>apply(conf.gamma.unilaterale.tmp, 2, mean)
>potenza.alfa = sum(risultati.S0.M1.no.errore[,7] < 0.05)/1000
>errore.secondo.tipo.alfa = sum(risultati.S0.M1.no.errore[,7] > 0.05)/1000
>potenza.beta = sum(risultati.S0.M1.no.errore[,8] < 0.05)/1000
>errore.secondo.tipo.beta = sum(risultati.S0.M1.no.errore[,8] > 0.05)/1000
>potenza.gamma = sum(risultati.S0.M1.no.errore[,9] < 0.05)/1000
>errore.secondo.tipo.gamma = sum(risultati.S0.M1.no.errore[,9] > 0.05)/1000

```

## Caso S1.M1. in presenza di errore di misura di tipo classico ( $p = 0$ ), con $\rho = 0.2$

```

>ris.sim.M1 = matrix(NA, nrow=R, ncol=10)
>dimnames(ris.sim.H1) = list(paste(1:1000),
c(alfa,beta,gamma,sd.alfa,sd.beta,sd.gamma,p.alfa,p.beta,p.gamma,sigma))
>rho = 0.2
>var.u = (var.s.vero*rho)/(1-rho)
>sd.u = sqrt(var.u)
>set.seed(126635)
>for(idx in 1:1000)
{
s.vero = rnorm(n, mean=media.s.vero, sd=sd.s.vero)
I.s.vero = as.numeric(s.vero >= 0)
s.classico = rnorm(n, s.vero, sd.u)
I.s.classico = as.numeric(s.classico >= 0)
logc = rnorm(n, mean = alfa + beta*s.vero + gamma*I.s.vero, sd=sd.logc)
modello.stimato = lm(logc ~ s.classico + I.s.classico)
ris.tmp = c(as.vector(summary(modello.stimato)$coef[,c(1:2,4)]),
summary(modello.stimato)$sigma)
ris.sim.M1[idx,] = ris.tmp
print(idx)
}
>risultati.S1.M1.classico = ris.sim.M1
>distorsione.S1.M1.classico = apply(distorsione.S1.M1.classico[,1:3], 2, mean)
- c(alfa, beta, gamma)
>relativa.alfa.tmp = (distorsione.S1.M1.classico[1]/alfa)*100
>relativa.beta.tmp = (distorsione.S1.M1.classico[2]/beta)*100
>relativa.gamma.tmp = (distorsione.S1.M1.classico[3]/gamma)*100
>distorsione.relativa.classico = c(relativa.alfa.tmp,
relativa.beta.tmp, relativa.gamma.tmp)
>varianza.S1.M1.classico = apply(risultati.S1.M1.classico[,1:3], 2, var)
>eqm.S1.M1.classico = varianza.S1.M1.classico + distorsione.S1.M1.classico2
>varianza.S1.M1.classico/distorsione.S1.M1.classico2

```

```

>test.alfa.tmp=(risultati.S1.M1.classico[,1]-alfa)/risultati.S1.M1.classico[,4]
>p.alfa.tmp = 2 * pmin(pnorm(test.alfa.tmp), 1-pnorm(test.alfa.tmp))
>conf.alfa.tmp = ifelse(p.alfa.tmp >= 0.05, 1, 0)
>test.beta.tmp = (risultati.S1.M1.classico[,2]-beta)/risultati.S1.M1.classico[,5]
>p.beta.tmp = 2 * pmin(pnorm(test.beta.tmp), 1-pnorm(test.beta.tmp))
>conf.beta.tmp = ifelse(p.beta.tmp >= 0.05, 1, 0)
>test.gamma.tmp = (risultati.S1.M1.classico[,3]-gamma)/risultati.S1.M1.classico[,6]
>p.gamma.tmp = 2 * pmin(pnorm(test.gamma.tmp), 1-pnorm(test.gamma.tmp))
>conf.gamma.tmp = ifelse(p.gamma.tmp >= 0.05, 1, 0)
>prop.test(sum(conf.alfa.tmp), R, p=0.95)
>prop.test(sum(conf.beta.tmp), R, p=0.95)
>prop.test(sum(conf.gamma.tmp), R, p=0.95)
>test.alfa.tmp=(risultati.S1.M1.classico[,1]-alfa)/risultati.S1.M1.classico[,4]
>p.alfa.unilaterale.tmp = cbind(pnorm(test.alfa.tmp), 1-pnorm(test.alfa.tmp))
>conf.alfa.unilaterale.tmp = apply(p.alfa.uni.tmp, 2, function(x) ifelse(x>=0.05,
1, 0))
>apply(conf.alfa.unilaterale.tmp, 2, mean)
>test.beta.tmp=(risultati.S1.M1.classico[,2]-beta)/risultati.S1.M1.classico[,5]
>p.beta.unilaterale.tmp = cbind(pnorm(test.beta.tmp), 1-pnorm(test.beta.tmp))
>conf.beta.unilaterale.tmp = apply(p.beta.unilaterale.tmp, 2, function(x) ifelse(x>=0.05,
1, 0))
>apply(conf.beta.unilaterale.tmp, 2, mean)
>test.gamma.tmp=(risultati.S1.M1.classico[,3]-gamma)/risultati.S1.M1.classico[,6]
>p.gamma.unilaterale.tmp = cbind(pnorm(test.gamma.tmp), 1-pnorm(test.gamma.tmp))
>conf.gamma.unilaterale.tmp = apply(p.gamma.unilaterale.tmp, 2, function(x)
ifelse(x>=0.05, 1, 0))
>apply(conf.gamma.unilaterale.tmp, 2, mean)
>potenza.alfa = sum(risultati.S1.M1.classico[,7] < 0.05)/1000
>errore.secondo.tipo.alfa = sum(risultati.S1.M1.classico[,7] > 0.05)/1000
>potenza.beta = sum(risultati.S1.M1.classico[,8] < 0.05)/1000
>errore.secondo.tipo.beta = sum(risultati.S1.M1.classico[,8] > 0.05)/1000
>potenza.gamma = sum(risultati.S1.M1.classico[,9] < 0.05)/1000
>errore.secondo.tipo.gamma = sum(risultati.S1.M1.classico[,9] > 0.05)/1000

```

## Caso S1.M1. in presenza di errore di misura di tipo di tipo mistura, $p = 0.5$ e $\rho = 0.2$

```

>ris.sim.M1 = matrix(NA, nrow=R, ncol=10)
>dimnames(ris.sim.H1) = list(paste(1:1000),
c(alfa,beta,gamma,sd.alfa,sd.beta,sd.gamma,p.alfa,p.beta,p.gamma,sigma))
>p = 0.5
>rho = 0.2
>var.u = (var.s.vero*rho)/(1-rho)
>sd.u = sqrt(var.u)
>set.seed(221513445)
>for(idx in 1:1000)
{
s.vero = rnorm(n, mean=media.s.vero, sd=sd.s.vero)
I.s.vero = as.numeric(s.vero >= 0)
s.classico = rnorm(n, s.vero, sd.u)
prob = rbinom(n, 1, p)
s.mistura = ifelse(prob <= p, s.vero, s.classico)
I.s.mistura = as.numeric(s.mistura >= 0)
logc = rnorm(n, mean = alfa + beta*s.vero + gamma*I.s.vero, sd=sd.logc)
modello.stimato = lm(logc ~ s.mistura + I.s.mistura)
ris.tmp = c(as.vector(summary(modello.stimato)$coef[,c(1:2,4)]),
summary(modello.stimato)$sigma)
ris.sim.M1[idx,] = ris.tmp
print(idx)
}
>risultati.S1.M1.mistura.p5 = ris.sim.M1
>distorsione.S1.M1.mistura.p5 = apply(distorsione.S1.M1.mistura.p5[,1:3], 2,
mean) - c(alfa, beta, gamma)
>relativa.alfa.tmp = (distorsione.S1.M1.mistura.p5/alfa)*100
>relativa.beta.tmp = (distorsione.S1.M1.mistura.p5[2]/beta)*100
>relativa.gamma.tmp = (distorsione.S1.M1.mistura.p5[3]/gamma)*100
>distorsione.relativa.mistura.p5 = c(relativa.alfa.tmp, relativa.beta.tmp,
relativa.gamma.tmp)
>varianza.S1.M1.mistura.p5 = apply(risultati.S1.M1.mistura.p5[,1:3], 2, var)
>eqm.S1.M1.mistura.p5 = varianza.S1.M1.mistura.p5 + distorsione.S1.M1.mistura.p52
>varianza.S1.M1.mistura.p5/distorsione.S1.M1.mistura.p52

```

```

>test.alfa.tmp=(risultati.S1.M1.mistura.p5[,1]-alfa)/risultati.S1.M1.mistura.p5[,4]
>p.alfa.tmp = 2 * pmin(pnorm(test.alfa.tmp), 1-pnorm(test.alfa.tmp))
>conf.alfa.tmp = ifelse(p.alfa.tmp >= 0.05, 1, 0)
>test.beta.tmp = (risultati.S1.M1.mistura.p5[,2]-beta)/risultati.S1.M1.mistura.p5[,5]
>p.beta.tmp = 2 * pmin(pnorm(test.beta.tmp), 1-pnorm(test.beta.tmp))
>conf.beta.tmp = ifelse(p.beta.tmp >= 0.05, 1, 0)
>test.gamma.tmp = (risultati.S1.M1.mistura.p5[,3]-gamma)/risultati.S1.M1.mistura.p5[,6]
>p.gamma.tmp = 2 * pmin(pnorm(test.gamma.tmp), 1-pnorm(test.gamma.tmp))
>conf.gamma.tmp = ifelse(p.gamma.tmp >= 0.05, 1, 0)
>prop.test(sum(conf.alfa.tmp), R, p=0.95)
>prop.test(sum(conf.beta.tmp), R, p=0.95)
>prop.test(sum(conf.gamma.tmp), R, p=0.95)
>test.alfa.tmp=(risultati.S1.M1.mistura.p5[,1]-alfa)/risultati.S1.M1.mistura.p5[,4]
>p.alfa.unilaterale.tmp = cbind(pnorm(test.alfa.tmp), 1-pnorm(test.alfa.tmp))
>conf.alfa.unilaterale.tmp = apply(p.alfa.uni.tmp, 2, function(x) ifelse(x>=0.05,
1, 0))
>apply(conf.alfa.unilaterale.tmp, 2, mean)
>test.beta.tmp=(risultati.S1.M1.mistura.p5[,2]-beta)/risultati.S1.M1.mistura.p5[,5]
>p.beta.unilaterale.tmp = cbind(pnorm(test.beta.tmp), 1-pnorm(test.beta.tmp))
>conf.beta.unilaterale.tmp = apply(p.beta.unilaterale.tmp, 2, function(x) ifelse(x>=0.05,
1, 0))
>apply(conf.beta.unilaterale.tmp, 2, mean)
>test.gamma.tmp=(risultati.S1.M1.mistura.p5[,3]-gamma)/risultati.S1.M1.mistura.p5[,6]
>p.gamma.unilaterale.tmp = cbind(pnorm(test.gamma.tmp), 1-pnorm(test.gamma.tmp))
>conf.gamma.unilaterale.tmp = apply(p.gamma.unilaterale.tmp, 2, function(x)
ifelse(x>=0.05, 1, 0))
>apply(conf.gamma.unilaterale.tmp, 2, mean)
>potenza.alfa = sum(risultati.S1.M1.mistura.p5[,7] < 0.05)/1000
>errore.secondo.tipo.alfa = sum(risultati.S1.M1.mistura.p5[,7] > 0.05)/1000
>potenza.beta = sum(risultati.S1.M1.mistura.p5[,8] < 0.05)/1000
>errore.secondo.tipo.beta = sum(risultati.S1.M1.mistura.p5[,8] > 0.05)/1000
>potenza.gamma = sum(risultati.S1.M1.mistura.p5[,9] < 0.05)/1000
>errore.secondo.tipo.gamma = sum(risultati.S1.M1.mistura.p5[,9] > 0.05)/1000

```



# Appendice D

## Istruzioni di R per il metodo SIMEX

### **SIMEX applicato al caso S0.M0. con errore di misura di tipo classico e $\rho = 0.2$**

Non riportiamo i comandi utilizzati per trovare i valori degli indici perchè sono i medesimi impiegati per lo studio di simulazione, adattati in modo opportuno.

Le istruzioni si riferiscono al caso descritto nel Paragrafo 4.1

```
>classico.S0.M0.simex = matrix(NA, nrow=200, ncol= 6)
>dimnames(classico.S0.M0.simex) = list(paste(1:200),c(alfa,beta,sd.alfa,sd.beta,-
p.alfa,p.beta))
>rho = 0.2
>var.u= ro*var.s.vero
>sd.u = sqrt(var.u)
>set.seed(22311)
>for(idx in 1:200)
{
s.vero = rnorm(n, mean=media.s.vero, sd=sd.s.vero)
s.classico = rnorm(n, s.vero, sd.u)
```

```

logc = rnorm(n, mean = alfa + beta*s.vero, sd=sd.logc)
mod.tmp = lm(logc ~ s.classico, x = TRUE)
simex.model=simex(model=mod.tmp, SIMEXvariable=s.classico, measurement.error=sd.u)
ris.simex.tmp = c(as.vector(summary(simex.model)$coef$jackknife[,c(1:2,4)],
mode=any))
classico.S0.M0.simex [idx,] = ris.simex.tmp
print(idx)
}

```

## SIMEX applicato al caso S1.M1. usando il valore teorico della varianza di $S_{\text{classico}}$

Le istruzioni si riferiscono al caso descritto nel Paragrafo 4.2

```

>classico.S1.M1.simex = matrix(NA, nrow=200, ncol= 9)
>dimnames(classico.S1.M1.simex) = list(paste(1:200),c(alfa,beta,gamma, sd.alfa,-
sd.beta,sd.gamma,p.alfa,p.beta, p.gamma))
>rho = 0.2
>var.u= ro*var.s.vero
>sd.u = sqrt(var.u)
>set.seed(2232321)
>for(idx in 1:200)
{
s.vero = rnorm(n, mean=media.s.vero, sd=sd.s.vero)
I.s.vero = as.numeric(s.vero >= 0)
s.classico = rnorm(n, s.vero, sd.u)
I.s.classico = as.numeric(s.classico >= 0) logc = rnorm(n, mean = alfa + beta*s.vero
+ gamma*I.s.vero, sd=sd.logc)
mod.tmp = lm(logc ~ s.classico + I.s.classico, x = TRUE)
simex.model=simex(model=mod.tmp, SIMEXvariable=s.classico, measurement.error=sd.u)
ris.simex.tmp = c(as.vector(summary(simex.model)$coef$jackknife[,c(1:2,4)],

```

```

mode=any))
classico.S1.M1.simex[idx,] = ris.simex.tmp
print(idx)
}

```

## SIMEX applicato al caso S1.M1. usando il valore empirico della varianza di $S_{\text{classico}}$

Le istruzioni si riferiscono al caso descritto nel Paragrafo 4.2

```

>classico.S1.M1.simex = matrix(NA, nrow=200, ncol= 9)
>dimnames(classico.S1.M1.simex) = list(paste(1:200),c(alfa,beta,gamma, sd.alfa,-
sd.beta,sd.gamma,p.alfa,p.beta, p.gamma))
>for(idx in 1:200)
{
s.vero = rnorm(n, mean=media.s.vero, sd=sd.s.vero)
I.s.vero = as.numeric(s.vero >= 0)
s.classico = rnorm(n, s.vero, sd.u)
I.s.classico = as.numeric(s.classico >= 0)
var.stimata = var(s.classico)
var.m.e = rho * var.stimata
sd.m.e = sqrt(var.m.e)
logc = rnorm(n, mean = alfa + beta*s.vero + gamma*I.s.vero, sd=sd.logc)
mod.tmp = lm(logc ~ s.classico + I.s.classico, x = TRUE)
simex.model=simex(model=mod.tmp, SIMEXvariable=s.classico, measurement.error=sd.m.e)
ris.simex.tmp = c(as.vector(summary(simex.model)$coef$jackknife[,c(1:2,4)],
mode=any))
classico.S1.M1.simex[idx,] = ris.simex.tmp
print(idx)
}

```

## SIMEX applicato al caso S1.M1. usando le quantità teoriche

Le istruzioni si riferiscono al caso descritto nel Paragrafo 4.3

```
>var.u = (var.s.vero*rho)/(1-rho)
>var.u
>sd.u = sqrt(var.u)
>sd.u
>z=0.5
>var.I.teorica = ro*z*(1-z)
>sd.teorica = sqrt(var.I.teorica)
>classico.S1.M1 = matrix(NA, nrow=R.200, ncol=9)
>dimnames(classico.S1.M1) = list(paste(1:200), c(alfa,beta,gamma,sd.alfa,sd.beta,-
sd.gamma,p.alfa,p.beta,p.gamma))
>for(idx in 1:200)
{
s.vero = rnorm(n, mean=media.s.vero, sd=sd.s.vero)
I.s.vero = as.numeric(s.vero >= 0)
s.classico = rnorm(n, s.vero, sd.u)
I.s.classico = as.numeric(s.classico >= 0)
logc = rnorm(n, mean = alfa + beta*s.vero + gamma*I.s.vero, sd=sd.logc)
mod.tmp = lm(logc ~ s.classico + I.s.classico, x = TRUE)
simex.model=simex(model=mod.tmp, SIMEXvariable=c(s.classico,I.s.classico),
measurement.error=c(sd.u , sd.teorica))
ris.simex.tmp = c(as.vector(summary(simex.model)$coef$jackknife[,c(1:2,4)],
mode=any))
classico.S1.M1 [idx,] = ris.simex.tmp
print(idx)
}
```

## SIMEX applicato al caso S1.M1. usando le quantità empiriche

Le istruzioni si riferiscono al caso descritto nel Paragrafo 4.3

```
>classico.S1.M1 = matrix(NA, nrow= R.200, ncol=9)
>dimnames(classico.S1.M1) = list(paste(1:200), c(alfa,beta,gamma,sd.alfa,sd.beta,-
sd.gamma,p.alfa,p.beta,p.gamma))
>for(idx in 1:200)
{
s.vero = rnorm(n, mean=media.s.vero, sd=sd.s.vero)
I.s.vero = as.numeric(s.vero >= 0)
s.classico = rnorm(n, s.vero, sd.u)
I.s.classico = as.numeric(s.classico >= 0)
var.stimata = var(s.classico)
var.m.e = rho * var.stimata
sd.m.e = sqrt(var.m.e)
z= sum(I.s.classico)/n
var.I.2 = rho*z*(1-z)
sd.I.2 = sqrt(var.I.2)
logc = rnorm(n, mean = alfa + beta*s.vero + gamma*I.s.vero, sd=sd.logc)
mod.tmp = lm(logc ~ s.classico + I.s.classico, x = TRUE)
simex.model=simex(model=mod.tmp, SIMEXvariable=c(s.err,I.s.err),
measurement.error=c(sd.m.e,sd.I.2))
ris.simex.tmp = c(as.vector(summary(simex.model)$coef$jackknife[,c(1:2,4)],
mode=any))
classico.S1.M1 [idx,] = ris.simex.tmp
print(idx)
}
```



# Bibliografia

- [1] F. Modigliani. *Il risparmio nel ciclo vitale*. CLUEB, Bologna, 1995.
- [2] J. Banks, R. Blundell, and S. Tanner. *Is There a Retirement-Savings Puzzle?*, volume 88, pages 769–788. *American Economic Review*, 1998.
- [3] B. D. Bernheim, J. Skinner, and S. Weinberg. *What Accounts for the Variation in Retirement Wealth Among U.S. Households?*, volume 91, pages 832–857. *American Economic Review*, 2001.
- [4] R. Miniaci, C. Monfardini, and G. Weber. *Is There a Retirement Consumption Puzzle in Italy?* London IFS, 2003.
- [5] E. Battistin, A. Brugiavini, E. Rettore, and G. Weber. *The Retirement Consumption Puzzle: Evidence from a Regression Discontinuity Approach*, volume 99, pages 2209–2226. *American Economic Review*, 2009.
- [6] G. W. Imbens and Donald B. Rubin. *Rubin causal model*. Palgrave Macmillan, Basingstoke, 2008.
- [7] Donald B. Rubin. *Causal Inference Using Potential Outcomes: Design, Modeling, Decision*, volume 100, pages 322–331. *The Journal of the American Statistical Association*, 2004.
- [8] D. S. Lee and T. Lemieux. *Regression Discontinuity Designs in Economics*. National Bureau of Economic Research, Cambridge, 2009.

- [9] D. S. Lee and D. Card. *Regression Discontinuity Inference with Specification Error*. Journal of Econometrics, 2007.
- [10] E. Battistin and E. Rettore. *Ineligibles and eligible non-participants as a double comparison group in regression-discontinuity designs*. Journal of Econometrics, 2007.
- [11] D. L. Thistlethwaite and D.T. Campbell. *Regression discontinuity analysis: an alternative to the ex post facto experiment*, volume 51, pages 309–317. Journal of Educational Psychology, 1960.
- [12] G. W. Imbens and J. D. Angrist. *Identification and estimation of local average treatment effects*, volume 62, pages 467–476. Econometrica, 1994.
- [13] B. Armstrong. *Exposure measurement error: consequences and design issue*. In Exposure Assessment in Occupational and Environmental Epidemiology, Oxford University Press, Oxford, 2003.
- [14] W. A. Fuller. *Measurement Error Models*. John Wiley & Sons, New York, 1987.
- [15] R. J. Carroll, D. Ruppert, and L. A. Stefanski. *Measurement Error in Nonlinear Models*. Chapman & Hall, London, 1995.
- [16] R. J. Carroll and L. A. Stefanski. *Measurement error, instrumental variables and corrections for attenuation with applications to meta analyses*, volume 13, pages 1265–1282. Statistics in Medicine, 1994.
- [17] Y. Amemiya. *Instrumental variable estimation for the nonlinear errors-in-variables model*, volume 28, pages 273–289. Journal of Econometrics, 1985.
- [18] Y. Amemiya. *Instrumental variables estimation of the nonlinear measurement error model*. In Statistical Analysis of Measurement Error

- Models and Application, American Mathematics Society, Providence, 1990.
- [19] Y. Amemiya. *Two-stages instrumental variable estimators for the non-linear errors-in-variables model*, volume 44, pages 311–332. *Journal of Econometrics*, 1990.
- [20] J. S. Buzas and L.A. Stefanski. *Instrumental variable estimation in generalized linear measurement error models*, volume 91, pages 999–1006. *Journal of the American Statistical Association*, 1996.
- [21] J. Cook and L.A. Stefanski. *A simulation extrapolation method for parametric measurement error models*, volume 89, pages 1314–1328. *Journal of the American Statistical Association*, 1994.
- [22] R. J. Carroll, H. Kuchenhoff, F. Lombard, and L.A. Stefanski. *Asymptotics for the SIMEX estimator in structural measurement error models*, volume 91, pages 242–250. *Journal of the American Statistical Association*, 1996.
- [23] L.A. Stefanski and J. Cook. *Simulation extrapolation: the measurement error jackknife*, volume 90, pages 1247–1256. *Journal of the American Statistical Association*, 1995.
- [24] L.A. Stefanski. *Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models*, volume 18, pages 4335–4358. *Communications in Statistics, Series A*, 1989.
- [25] T. Nakamura. *Corrected score functions for errors-in-variables models: methodology and application to generalized linear models*, volume 77, pages 127–137. *Biometrika*, 1990.
- [26] L.A. Stefanski and R. J. Carroll. *Covariate measurement error in logistic regression*, volume 13, pages 1335–1351. *The Annals of Statistics*, 1985.

- [27] L.A. Stefanski and R. J. Carroll. *Conditional scores and optimal scores in generalized linear measurement error models*, volume 74, pages 703–716. *Biometrika*, 1987.
- [28] D.W. Schafer and K. G. Purdy. *Likelihood analysis for errors-in-variables regression with replicate measurements*, volume 83, pages 813–824. *Biometrika*, 1996.
- [29] H. Küchenhoff and R.J. Carroll. *Segmented regression with errors in predictors: semi-parametric and parametric methods*, volume 16, pages 169–188. *Statistics in Medicine*, 1997.
- [30] H. Küchenhoff and R.J. Carroll. Biases in segmented regression with errors in predictors, 1995.
- [31] R. J. Carroll, M. H. Gail, and J. H. Lubin. *Case-control studies with errors in covariates*, volume 88, pages 185–199. *Journal of the American Statistical Association*, 1993.
- [32] K. Roeder, R. J. Carroll, and B. G. Lindsay. *A semiparametric mixture approach to case-control studies with errors in covariables*, volume 91, pages 722–732. *Journal of the American Statistical Association*, 1996.
- [33] R. J. Carroll and L. A. Stefanski. *Approximate quasilielihood estimation in models with surrogate predictors*, volume 85, pages 652–663. *Journal of the American Statistical Association*, 1990.
- [34] D. W. Schafer. *Semiparametric maximum likelihood for measurement error model regression*, volume 57, pages 53–61. *Biometrics*, 2001.
- [35] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, New York, 2004.
- [36] M. A. Tanner. *Tools for Statistical Inference: methods for the exploration of posterior distributions and likelihood functions*. Springer-Verlag, New York, second edition, 1993.

- [37] P. Gustafson. *Measurement Error and Misclassification in Statistics and Epidemiology*. Chapman & Hall/CRC Press, 2004.
- [38] S. Richardson and W. Gilks. *Conditional independence models for epidemiological studies with covariate measurement error*, volume 12, pages 1703–1722. *Statistics in Medicine*, 1993.
- [39] S. Richardson and L. Leblond. *Some comments on misspecification of priors in Bayesian modeling of measurement error problems*, volume 16, pages 203–213. *Statistics in Medicine*, 1997.
- [40] E. Battistin, R. Blundell, and A. Lewbel. *Why is Consumption More Log Normal Than Income? Gibrat Law Revisited*, volume 117, pages 1140–1154. *Journal of Political Economy*, 1999.
- [41] J. L. Horowitz and C. F. Manski. *Identification and robustness with contaminated and corrupted data*, volume 62, pages 281–302. *Econometrica*, 1995.
- [42] W. Lederer and H. Kuchenhoff. *A short Introduction to the SIMEX and MCSIMEX*, volume 6, pages 26 – 31. 2006.

Tabella 1. Distorsione assoluta, calcolata come  $D(\hat{\theta}) = E(\hat{\theta}) - \theta$

Simulo i dati sotto l'ipotesi che non ci sia discontinuità e stimo il modello senza discontinuità <b>(S0.M0)</b>			
P = 1 $\alpha$ : 0.00006 $\beta$ : 0.00001 $\gamma$ : -			
	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
P = 0.5	$\alpha$ : -0.0001 $\beta$ : 0.0010	$\alpha$ : 0.0001 $\beta$ : 0.0021	$\alpha$ : 0.00006 $\beta$ : 0.0034
P = 0.4	$\alpha$ : 0.0001 $\beta$ : 0.0012	$\alpha$ : 0.00008 $\beta$ : 0.0026	$\alpha$ : 0.0001 $\beta$ : 0.0039
P = 0.3	$\alpha$ : 0.0001 $\beta$ : 0.0014	$\alpha$ : -0.0001 $\beta$ : 0.0029	$\alpha$ : -0.00007 $\beta$ : 0.0046
P = 0.2	$\alpha$ : -0.0001 $\beta$ : 0.0014	$\alpha$ : 0.00006 $\beta$ : 0.0033	$\alpha$ : 0.00009 $\beta$ : 0.0050
P = 0	$\alpha$ : -0.0002 $\beta$ : 0.0019	$\alpha$ : -0.0004 $\beta$ : 0.0040	$\alpha$ : 0.00006 $\beta$ : 0.0059

Simulo i dati sotto l'ipotesi che ci sia discontinuità e stimo il modello senza discontinuità <b>(S1.M0)</b>			
P = 1 $\alpha$ : -0.0297 $\beta$ : -0.01208 $\gamma$ : -			
	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
P = 0.5	$\alpha$ : -0.0299 $\beta$ : -0.0103	$\alpha$ : -0.0300 $\beta$ : -0.0084	$\alpha$ : -0.0299 $\beta$ : -0.0063
P = 0.4	$\alpha$ : -0.0296 $\beta$ : -0.0099	$\alpha$ : -0.0302 $\beta$ : -0.0078	$\alpha$ : -0.0299 $\beta$ : -0.0053
P = 0.3	$\alpha$ : -0.0300 $\beta$ : -0.0097	$\alpha$ : -0.0299 $\beta$ : -0.0071	$\alpha$ : -0.0301 $\beta$ : -0.0045
P = 0.2	$\alpha$ : -0.0299 $\beta$ : -0.0094	$\alpha$ : -0.0300 $\beta$ : -0.0067	$\alpha$ : -0.0301 $\beta$ : -0.0038
P = 0	$\alpha$ : -0.0297 $\beta$ : 0.00184	$\alpha$ : -0.0305 $\beta$ : 0.0040	$\alpha$ : -0.0298 $\beta$ : 0.0061

Simulo i dati sotto l'ipotesi che non ci sia discontinuità e stimo il modello con discontinuità <b>(S0.M1)</b>			
P = 1 $\alpha$ : 0.000044 $\beta$ : 0.00016 $\gamma$ : -0.000004			
	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
P = 0.5	$\alpha$ : 0.00013 $\beta$ : 0.0011 $\gamma$ : -0.00026	$\alpha$ : 0.0003 $\beta$ : 0.0024 $\gamma$ : -0.0014	$\alpha$ : 0.00102 $\beta$ : 0.0041 $\gamma$ : -0.0025
P = 0.4	$\alpha$ : 0.00015 $\beta$ : 0.0012 $\gamma$ : -0.00023	$\alpha$ : 0.0009 $\beta$ : 0.0029 $\gamma$ : -0.0019	$\alpha$ : 0.0013 $\beta$ : 0.0045 $\gamma$ : -0.0023
P = 0.3	$\alpha$ : 0.0002 $\beta$ : 0.0016 $\gamma$ : -0.00025	$\alpha$ : 0.00043 $\beta$ : 0.0034 $\gamma$ : -0.0017	$\alpha$ : 0.0005 $\beta$ : 0.0048 $\gamma$ : -0.0017
P = 0.2	$\alpha$ : 0.0011 $\beta$ : 0.0018 $\gamma$ : -0.0012	$\alpha$ : 0.00035 $\beta$ : 0.0036 $\gamma$ : -0.0010	$\alpha$ : 0.00056 $\beta$ : 0.0053 $\gamma$ : -0.0016
P = 0	$\alpha$ : 0.0005 $\beta$ : 0.002 $\gamma$ : -0.00025	$\alpha$ : 0.00023 $\beta$ : 0.0039 $\gamma$ : -0.0004	$\alpha$ : 0.00019 $\beta$ : 0.00586 $\gamma$ : -0.0003

Simulo i dati sotto l'ipotesi che ci sia discontinuità e stimo il modello con discontinuità <b>(S1.M1)</b>			
P = 1 $\alpha$ : -0.000599 $\beta$ : 0.00022 $\gamma$ : 0.00122			
	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
P = 0.5	$\alpha$ : -0.0069 $\beta$ : -0.0014 $\gamma$ : 0.0138	$\alpha$ : -0.0086 $\beta$ : -0.0004 $\gamma$ : 0.0176	$\alpha$ : -0.0097 $\beta$ : 0.0012 $\gamma$ : 0.0189
P = 0.4	$\alpha$ : -0.0087 $\beta$ : -0.0014 $\gamma$ : 0.0170	$\alpha$ : -0.0100 $\beta$ : -0.0004 $\gamma$ : 0.0200	$\alpha$ : -0.0112 $\beta$ : 0.0012 $\gamma$ : 0.0221
P = 0.3	$\alpha$ : -0.0102 $\beta$ : -0.0019 $\gamma$ : 0.0203	$\alpha$ : -0.0123 $\beta$ : -0.0008 $\gamma$ : 0.0251	$\alpha$ : -0.0146 $\beta$ : 0.00083 $\gamma$ : 0.0286
P = 0.2	$\alpha$ : -0.0114 $\beta$ : -0.0022 $\gamma$ : 0.0222	$\alpha$ : -0.0149 $\beta$ : -0.0011 $\gamma$ : 0.0299	$\alpha$ : -0.0168 $\beta$ : 0.00079 $\gamma$ : 0.03286
P = 0	$\alpha$ : -0.0143 $\beta$ : -0.0029 $\gamma$ : 0.0288	$\alpha$ : -0.0182 $\beta$ : -0.0013 $\gamma$ : 0.0364	$\alpha$ : -0.0217 $\beta$ : 0.0004 $\gamma$ : 0.04359

Tabella 2. Distorsione relativa (%), calcolata come  $\frac{E(\hat{\theta})-\theta}{\theta} * 100$

Simulo i dati sotto l'ipotesi che non ci sia discontinuità e stimo il modello senza discontinuità <b>(S0.M0)</b>			
P = 1 α: 0.0007% β: 0.0608% γ: -			
	ρ = 0.1	ρ = 0.2	ρ = 0.3
P = 0.5	α: 0% β: 5.1%	α: 0.1% β: 10.9%	α: 0.0007% β: 18.03%
P = 0.4	α: 0.1% β: 6.2%	α: 0.1% β: 13.5%	α: 0.1% β: 19.9%
P = 0.3	α: 0.1% β: 7.1%	α: 0% β: 14.7%	α: 0% β: 23.1%
P = 0.2	α: 0% β: 7.3%	α: 0.1% β: 16.6%	α: 0.1% β: 25.3%
P = 0	α: 0% β: 9.9%	α: 0% β: 20.3%	α: 0.1% β: 29.8%

Simulo i dati sotto l'ipotesi che ci sia discontinuità e stimo il modello senza discontinuità <b>(S1.M0)</b>			
P = 1 α: 0.305% β: 39.59% γ: -			
	ρ = 0.1	ρ = 0.2	ρ = 0.3
P = 0.5	α: 0.307% β: 51.59%	α: 0.307% β: 42.1%	α: 0.306% β: 31.8%
P = 0.4	α: 0.304% β: 50.01%	α: 0.309% β: 38.9%	α: 0.307% β: 26.7%
P = 0.3	α: 0.307% β: 48.6%	α: 0.307% β: 35.3%	α: 0.309% β: 22.7%
P = 0.2	α: 0.307% β: 47.1%	α: 0.308% β: 33.4%	α: 0.309% β: 19.20%
P = 0	α: 0.305% β: 9.217%	α: 0.312% β: 20.11%	α: 0.305% β: 30.68%

Simulo i dati sotto l'ipotesi che non ci sia discontinuità e stimo il modello con discontinuità <b>(S0.M1)</b>			
P = 1 α: 0.00046% β: 0.802% γ: n.d.			
	ρ = 0.1	ρ = 0.2	ρ = 0.3
P = 0.5	α: 0.0014% β: 5.94% γ: n.d.	α: 0.0032% β: 12.22% γ: n.d.	α: 0.0105% β: 20.51% γ: n.d.
P = 0.4	α: 0.0016% β: 6.42% γ: n.d.	α: 0.0095% β: 14.63% γ: n.d.	α: 0.0138% β: 22.48% γ: n.d.
P = 0.3	α: 0.0025% β: 8.10% γ: n.d.	α: 0.0045% β: 17.08% γ: n.d.	α: 0.0052% β: 24.42% γ: n.d.
P = 0.2	α: 0.0082% β: 9.43% γ: n.d.	α: 0.0036% β: 18.34% γ: n.d.	α: 0.0058% β: 26.78% γ: n.d.
P = 0	α: 0% β: 9.63% γ: n.d.	α: 0.1% β: 19.71% γ: n.d.	α: 0.002% β: 30.68% γ: n.d.

Simulo i dati sotto l'ipotesi che ci sia discontinuità e stimo il modello con discontinuità <b>(S1.M1)</b>			
P = 1 α: 0.0062% β: 1.135% γ: 2.0384%			
	ρ = 0.1	ρ = 0.2	ρ = 0.3
P = 0.5	α: 0.07% β: 7.01% γ: 22.96%	α: 0.09% β: 2.03% γ: 29.26%	α: 0.099% β: 5.764% γ: 31.57%
P = 0.4	α: 0.09% β: 7.19% γ: 28.38%	α: 0.10% β: 2.21% γ: 33.41%	α: 0.114% β: 5.971% γ: 36.86%
P = 0.3	α: 0.1% β: 9.47% γ: 33.88%	α: 0.13% β: 4.24% γ: 41.76%	α: 0.149% β: 4.15% γ: 47.71%
P = 0.2	α: 0.12% β: 10.85% γ: 37.03%	α: 0.15% β: 5.42% γ: 49.89%	α: 0.165% β: 3.97% γ: 54.77%
P = 0	α: 0.15% β: 14.52% γ: 47.95%	α: 0.18% β: 6.41% γ: 60.76%	α: 0.22% β: 2.09% γ: 72.65%

Tabella 3. Errore quadratico medio, calcolato come  $EQM(\hat{\theta}) = V(\hat{\theta}) + D(\hat{\theta})^2$

Simulo i dati sotto l'ipotesi che non ci sia discontinuità e stimo il modello senza discontinuità <b>(S0.M0)</b> .			
P = 1 $\alpha$ : 0.00004 $\beta$ : 0.000008 $\gamma$ : -			
	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
P = 0.5	$\alpha$ : 0.00004 $\beta$ : 0.00001	$\alpha$ : 0.00004 $\beta$ : 0.00002	$\alpha$ : 0.00004 $\beta$ : 0.00002
P = 0.4	$\alpha$ : 0.00004 $\beta$ : 0.00001	$\alpha$ : 0.00004 $\beta$ : 0.00002	$\alpha$ : 0.00004 $\beta$ : 0.00003
P = 0.3	$\alpha$ : 0.00004 $\beta$ : 0.00001	$\alpha$ : 0.00004 $\beta$ : 0.00002	$\alpha$ : 0.00004 $\beta$ : 0.00003
P = 0.2	$\alpha$ : 0.00004 $\beta$ : 0.00001	$\alpha$ : 0.00004 $\beta$ : 0.00002	$\alpha$ : 0.00004 $\beta$ : 0.00003
P = 0	$\alpha$ : 0.00004 $\beta$ : 0.00001	$\alpha$ : 0.00004 $\beta$ : 0.00002	$\alpha$ : 0.00004 $\beta$ : 0.00004

Simulo i dati sotto l'ipotesi che ci sia discontinuità e stimo il modello senza discontinuità <b>(S1.M0)</b> .			
P = 1 $\alpha$ : 0.00094 $\beta$ : 0.00015 $\gamma$ : -			
	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
P = 0.5	$\alpha$ : 0.00095 $\beta$ : 0.00011	$\alpha$ : 0.00093 $\beta$ : 0.00008	$\alpha$ : 0.00093 $\beta$ : 0.00005
P = 0.4	$\alpha$ : 0.00092 $\beta$ : 0.00011	$\alpha$ : 0.00095 $\beta$ : 0.00007	$\alpha$ : 0.00094 $\beta$ : 0.00004
P = 0.3	$\alpha$ : 0.00093 $\beta$ : 0.00010	$\alpha$ : 0.00094 $\beta$ : 0.00006	$\alpha$ : 0.00094 $\beta$ : 0.00003
P = 0.2	$\alpha$ : 0.00093 $\beta$ : 0.00010	$\alpha$ : 0.00094 $\beta$ : 0.00005	$\alpha$ : 0.00094 $\beta$ : 0.00002
P = 0	$\alpha$ : 0.00092 $\beta$ : 0.00001	$\alpha$ : 0.00091 $\beta$ : 0.00002	$\alpha$ : 0.00094 $\beta$ : 0.00004

Simulo i dati sotto l'ipotesi che non ci sia discontinuità e stimo il modello con discontinuità <b>(S0.M1)</b> .			
P = 1 $\alpha$ : 0.00013 $\beta$ : 0.00003 $\gamma$ : 0.0004			
	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
P = 0.5	$\alpha$ : 0.00014 $\beta$ : 0.00002 $\gamma$ : 0.00040	$\alpha$ : 0.00014 $\beta$ : 0.00003 $\gamma$ : 0.00040	$\alpha$ : 0.00015 $\beta$ : 0.00004 $\gamma$ : 0.00044
P = 0.4	$\alpha$ : 0.00014 $\beta$ : 0.00003 $\gamma$ : 0.00042	$\alpha$ : 0.00015 $\beta$ : 0.00003 $\gamma$ : 0.00043	$\alpha$ : 0.00016 $\beta$ : 0.00004 $\gamma$ : 0.00044
P = 0.3	$\alpha$ : 0.00013 $\beta$ : 0.00003 $\gamma$ : 0.00040	$\alpha$ : 0.00013 $\beta$ : 0.00003 $\gamma$ : 0.00040	$\alpha$ : 0.00014 $\beta$ : 0.00004 $\gamma$ : 0.00041
P = 0.2	$\alpha$ : 0.00014 $\beta$ : 0.00002 $\gamma$ : 0.00042	$\alpha$ : 0.00014 $\beta$ : 0.00003 $\gamma$ : 0.00038	$\alpha$ : 0.00014 $\beta$ : 0.00005 $\gamma$ : 0.00041
P = 0	$\alpha$ : 0.00008 $\beta$ : 0.00001 $\gamma$ : 0.00015	$\alpha$ : 0.00007 $\beta$ : 0.00002 $\gamma$ : 0.00014	$\alpha$ : 0.00007 $\beta$ : 0.00004 $\gamma$ : 0.00015

Simulo i dati sotto l'ipotesi che ci sia discontinuità e stimo il modello con discontinuità <b>(S1.M1)</b> .			
P = 1 $\alpha$ : 0.00014 $\beta$ : 0.00003 $\gamma$ : 0.00044			
	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
P = 0.5	$\alpha$ : 0.00018 $\beta$ : 0.00003 $\gamma$ : 0.00060	$\alpha$ : 0.00021 $\beta$ : 0.00002 $\gamma$ : 0.00072	$\alpha$ : 0.00023 $\beta$ : 0.00002 $\gamma$ : 0.00076
P = 0.4	$\alpha$ : 0.00022 $\beta$ : 0.00003 $\gamma$ : 0.00070	$\alpha$ : 0.00025 $\beta$ : 0.00002 $\gamma$ : 0.00084	$\alpha$ : 0.00029 $\beta$ : 0.00002 $\gamma$ : 0.00099
P = 0.3	$\alpha$ : 0.00025 $\beta$ : 0.00003 $\gamma$ : 0.00084	$\alpha$ : 0.00029 $\beta$ : 0.00002 $\gamma$ : 0.00105	$\alpha$ : 0.00035 $\beta$ : 0.00002 $\gamma$ : 0.00125
P = 0.2	$\alpha$ : 0.00027 $\beta$ : 0.00003 $\gamma$ : 0.00091	$\alpha$ : 0.00037 $\beta$ : 0.00002 $\gamma$ : 0.00132	$\alpha$ : 0.00042 $\beta$ : 0.00002 $\gamma$ : 0.00155
P = 0	$\alpha$ : 0.00036 $\beta$ : 0.00003 $\gamma$ : 0.00127	$\alpha$ : 0.00047 $\beta$ : 0.00002 $\gamma$ : 0.00173	$\alpha$ : 0.00065 $\beta$ : 0.00002 $\gamma$ : 0.00238

Tabella 4. Rapporto fra varianza dello stimatore e distorsione dello stimatore al quadrato, calcolata come  $\frac{Var(\hat{\theta})}{D(\hat{\theta})^2}$

Simulo i dati sotto l'ipotesi che non ci sia discontinuità e stimo il modello senza discontinuità <b>(S0.M0)</b> .			
P = 1 α: 10256.16 β: 62436.65 γ: -			
	ρ = 0.1	ρ = 0.2	ρ = 0.3
P = 0.5	α: 3325.952 β: 8.628	α: 2637.865 β: 1.756	α: 8885.07 β: 0.5853
P = 0.4	α: 1952.312 β: 5.360	α: 5406.074 β: 1.132	α: 1982.594 β: 0.453
P = 0.3	α: 2089.24 β: 4.545	α: 3406.320 β: 0.932	α: 6541.654 β: 0.322
P = 0.2	α: 3751.661 β: 3.892	α: 9181.741 β: 0.682	α: 4008.25 β: 0.274
P = 0	α: 786.574 β: 2.246	α: 218.724 β: 0.434	α: 7875.532 β: 0.179

Simulo i dati sotto l'ipotesi che ci sia discontinuità e stimo il modello senza discontinuità <b>(S1.M0)</b> .			
P = 1 α: 0.0385 β: 0.0667 γ: -			
	ρ = 0.1	ρ = 0.2	ρ = 0.3
P = 0.5	α: 0.043 β: 0.100	α: 0.040 β: 0.123	α: 0.043 β: 0.187
P = 0.4	α: 0.044 β: 0.090	α: 0.046 β: 0.134	α: 0.043 β: 0.291
P = 0.3	α: 0.040 β: 0.097	α: 0.040 β: 0.172	α: 0.040 β: 0.322
P = 0.2	α: 0.041 β: 0.097	α: 0.041 β: 0.177	α: 0.041 β: 0.444
P = 0	α: 0.040 β: 1.943	α: 0.042 β: 0.434	α: 0.043 β: 0.188

Simulo i dati sotto l'ipotesi che non ci sia discontinuità e stimo il modello con discontinuità <b>(S0.M1)</b> .			
P = 1 α: 42554.826 β: 363.355 γ: 7374465.54			
	ρ = 0.1	ρ = 0.2	ρ = 0.3
P = 0.5	α: 760583.36 β: 13.5499 γ: 119639.13	α: 1429.79 β: 3.8055 γ: 295.879	α: 124.120 β: 1.239 γ: 64.229
P = 0.4	α: 2461.563 β: 11.669 γ: 1147.493	α: 165.701 β: 2.5889 γ: 106.841	α: 83.132 β: 1.0839 γ: 83.094
P = 0.3	α: 11573.61 β: 8.351 γ: 1243.570	α: 200.963 β: 1.865 γ: 141.177	α: 575.007 β: 0.8746 γ: 154.684
P = 0.2	α: 123.839 β: 7.34 γ: 299.449	α: 1144.44 β: 1.5434 γ: 535.45	α: 449.794 β: 0.6660 γ: 436.756
P = 0	α: 355.679 β: 2.068 γ: 188.361	α: 1283.314 β: 0.513 γ: 723.070	α: 2254.956 β: 0.1909 γ: 1438.531

Simulo i dati sotto l'ipotesi che ci sia discontinuità e stimo il modello con discontinuità <b>(S1.M1)</b> .			
P = 1 α: 413.879 β: 513.470 γ: 278.585			
	ρ = 0.1	ρ = 0.2	ρ = 0.3
P = 0.5	α: 2.783 β: 12.234 γ: 2.172	α: 1.817 β: 143.571 γ: 1.334	α: 1.4953 β: 16.602 γ: 1.1824
P = 0.4	α: 1.912 β: 11.946 γ: 1.407	α: 1.508 β: 120.223 γ: 1.086	α: 1.0888 β: 15.571 γ: 0.8750
P = 0.3	α: 1.396 β: 6.893 γ: 1.028	α: 0.943 β: 30.111 γ: 0.673	α: 0.7210 β: 30.005 γ: 0.5492
P = 0.2	α: 1.073 β: 4.818 γ: 0.850	α: 0.653 β: 20.424 γ: 0.471	α: 0.498 β: 29.878 γ: 0.3758
P = 0	α: 0.742 β: 2.976 γ: 0.537	α: 0.433 β: 12.333 γ: 0.301	α: 0.2576 β: 81.955 γ: 0.1914

Tabella 5. Livello di copertura reale per intervalli di confidenza bilaterali (tra parentesi: intervallo di confidenza al 95% ). Livello nominale considerato 0.95.

Simulo i dati sotto l'ipotesi che non ci sia discontinuità e stimo il modello senza discontinuità <b>(S0.M0)</b> .			
P = 1 $\alpha$ : 0.959 (0.944 ; 0.970) $\beta$ : 0.952 (0.936 ; 0.964) $\gamma$ : -			
	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
P = 0.5	$\alpha$ : 0.945 (0.928 ; 0.957) $\beta$ : 0.934 (0.916 ; 0.948)	$\alpha$ : 0.940 (0.922 ; 0.953) $\beta$ : 0.886 (0.864 ; 0.904)	$\alpha$ : 0.950 (0.934 ; 0.961) $\beta$ : 0.752 (0.723 ; 0.778)
P = 0.4	$\alpha$ : 0.956 (0.940 ; 0.967) $\beta$ : 0.936 (0.918 ; 0.950)	$\alpha$ : 0.956 (0.940 ; 0.967) $\beta$ : 0.841 (0.816 ; 0.862)	$\alpha$ : 0.953 (0.937 ; 0.964) $\beta$ : 0.687 (0.657 ; 0.715)
P = 0.3	$\alpha$ : 0.963 (0.948 ; 0.973) $\beta$ : 0.916 (0.896 ; 0.932)	$\alpha$ : 0.945 (0.928 ; 0.957) $\beta$ : 0.833 (0.808 ; 0.855)	$\alpha$ : 0.957 (0.942 ; 0.968) $\beta$ : 0.602 (0.570 ; 0.632)
P = 0.2	$\alpha$ : 0.959 (0.944 ; 0.970) $\beta$ : 0.919 (0.899 ; 0.934)	$\alpha$ : 0.944 (0.927 ; 0.957) $\beta$ : 0.792 (0.765 ; 0.816)	$\alpha$ : 0.959 (0.944 ; 0.970) $\beta$ : 0.516 (0.484 ; 0.547)
P = 0	$\alpha$ : 0.951 (0.935 ; 0.963) $\beta$ : 0.886 (0.864 ; 0.905)	$\alpha$ : 0.948 (0.932 ; 0.961) $\beta$ : 0.684 (0.654 ; 0.712)	$\alpha$ : 0.957 (0.942 ; 0.968) $\beta$ : 0.334 (0.315 ; 0.374)

Simulo i dati sotto l'ipotesi che ci sia discontinuità e stimo il modello senza discontinuità <b>(S1.M0)</b> .			
P = 1 $\alpha$ : 0.003 (0.0007 ; 0.009) $\beta$ : 0.034 (0.024 ; 0.048) $\gamma$ : -			
	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
P = 0.5	$\alpha$ : 0.000 (0 ; 0.005) $\beta$ : 0.091 (0.074 ; 0.111)	$\alpha$ : 0.004 (0.0013 ; 0.011) $\beta$ : 0.168 (0.146 ; 0.193)	$\alpha$ : 0.000 (0 ; 0.005) $\beta$ : 0.373 (0.343 ; 0.404)
P = 0.4	$\alpha$ : 0.001 (0 ; 0.006) $\beta$ : 0.073 (0.058 ; 0.091)	$\alpha$ : 0.005 (0.0018 ; 0.012) $\beta$ : 0.232 (0.206 ; 0.260)	$\alpha$ : 0.004 (0.0012 ; 0.0109) $\beta$ : 0.507 (0.475 ; 0.538)
P = 0.3	$\alpha$ : 0.003 (0.0007 ; 0.0095) $\beta$ : 0.103 (0.085 ; 0.124)	$\alpha$ : 0.002 (0.0003 ; 0.008) $\beta$ : 0.309 (0.280 ; 0.338)	$\alpha$ : 0.001 (0 ; 0.006) $\beta$ : 0.620 (0.589 ; 0.650)
P = 0.2	$\alpha$ : 0.001 (0 ; 0.006) $\beta$ : 0.105 (0.870 ; 0.126)	$\alpha$ : 0.002 (0.0003 ; 0.008) $\beta$ : 0.336 (0.306 ; 0.366)	$\alpha$ : 0.001 (0 ; 0.0047) $\beta$ : 0.707 (0.677 ; 0.734)
P = 0	$\alpha$ : 0.003 (0.0008 ; 0.009) $\beta$ : 0.902 (0.881 ; 0.919)	$\alpha$ : 0.001 (0 ; 0.006) $\beta$ : 0.698 (0.668 ; 0.726)	$\alpha$ : 0.001 (0 ; 0.0064) $\beta$ : 0.342 (0.313 ; 0.372)

Simulo i dati sotto l'ipotesi che non ci sia discontinuità e stimo il modello con discontinuità <b>(S0.M1)</b> .			
P = 1 $\alpha$ : 0.954 (0.937 ; 0.966) $\beta$ : 0.955 (0.940 ; 0.967) $\gamma$ : 0.967 (0.953 ; 0.977)			
	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
P = 0.5	$\alpha$ : 0.954 (0.938 ; 0.965) $\beta$ : 0.955 (0.939 ; 0.966) $\gamma$ : 0.955 (0.939 ; 0.966)	$\alpha$ : 0.955 (0.939 ; 0.966) $\beta$ : 0.922 (0.903 ; 0.937) $\gamma$ : 0.957 (0.942 ; 0.968)	$\alpha$ : 0.939 (0.921 ; 0.952) $\beta$ : 0.869 (0.846 ; 0.888) $\gamma$ : 0.939 (0.921 ; 0.952)
P = 0.4	$\alpha$ : 0.952 (0.936 ; 0.964) $\beta$ : 0.945 (0.928 ; 0.957) $\gamma$ : 0.952 (0.936 ; 0.964)	$\alpha$ : 0.949 (0.933 ; 0.961) $\beta$ : 0.916 (0.896 ; 0.932) $\gamma$ : 0.951 (0.935 ; 0.963)	$\alpha$ : 0.923 (0.904 ; 0.938) $\beta$ : 0.829 (0.803 ; 0.851) $\gamma$ : 0.935 (0.917 ; 0.949)
P = 0.3	$\alpha$ : 0.964 (0.949 ; 0.974) $\beta$ : 0.933 (0.915 ; 0.947) $\gamma$ : 0.961 (0.946 ; 0.971)	$\alpha$ : 0.953 (0.937 ; 0.964) $\beta$ : 0.884 (0.862 ; 0.902) $\gamma$ : 0.954 (0.938 ; 0.965)	$\alpha$ : 0.951 (0.935 ; 0.963) $\beta$ : 0.811 (0.785 ; 0.834) $\gamma$ : 0.949 (0.933 ; 0.961)
P = 0.2	$\alpha$ : 0.954 (0.938 ; 0.965) $\beta$ : 0.945 (0.928 ; 0.957) $\gamma$ : 0.944 (0.927 ; 0.957)	$\alpha$ : 0.953 (0.937 ; 0.964) $\beta$ : 0.895 (0.873 ; 0.912) $\gamma$ : 0.965 (0.951 ; 0.975)	$\alpha$ : 0.946 (0.929 ; 0.958) $\beta$ : 0.764 (0.736 ; 0.789) $\gamma$ : 0.948 (0.931 ; 0.960)
P = 0	$\alpha$ : 0.943 (0.926 ; 0.956) $\beta$ : 0.893 (0.872 ; 0.911) $\gamma$ : 0.961 (0.946 ; 0.972)	$\alpha$ : 0.973 (0.960 ; 0.982) $\beta$ : 0.706 (0.676 ; 0.734) $\gamma$ : 0.971 (0.958 ; 0.980)	$\alpha$ : 0.961 (0.946 ; 0.972) $\beta$ : 0.344 (0.315 ; 0.374) $\gamma$ : 0.949 (0.933 ; 0.961)

Simulo i dati sotto l'ipotesi che ci sia discontinuità e stimo il modello con discontinuità <b>(S1.M1)</b> .			
P = 1 $\alpha$ : 0.950 (0.935 ; 0.962) $\beta$ : 0.951 (0.935 ; 0.963) $\gamma$ : 0.942 (0.925 ; 0.955)			
	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
P = 0.5	$\alpha$ : 0.923 (0.904 ; 0.938) $\beta$ : 0.946 (0.930 ; 0.960) $\gamma$ : 0.904 (0.884 ; 0.921)	$\alpha$ : 0.899 (0.878 ; 0.917) $\beta$ : 0.953 (0.937 ; 0.965) $\gamma$ : 0.858 (0.834 ; 0.879)	$\alpha$ : 0.878 (0.856 ; 0.897) $\beta$ : 0.947 (0.931 ; 0.960) $\gamma$ : 0.842 (0.817 ; 0.864)
P = 0.4	$\alpha$ : 0.888 (0.866 ; 0.906) $\beta$ : 0.944 (0.927 ; 0.957) $\gamma$ : 0.879 (0.857 ; 0.898)	$\alpha$ : 0.848 (0.824 ; 0.869) $\beta$ : 0.949 (0.933 ; 0.961) $\gamma$ : 0.825 (0.800 ; 0.850)	$\alpha$ : 0.831 (0.806 ; 0.853) $\beta$ : 0.938 (0.921 ; 0.952) $\gamma$ : 0.767 (0.739 ; 0.793)
P = 0.3	$\alpha$ : 0.864 (0.841 ; 0.884) $\beta$ : 0.926 (0.907 ; 0.941) $\gamma$ : 0.837 (0.812 ; 0.859)	$\alpha$ : 0.812 (0.786 ; 0.835) $\beta$ : 0.941 (0.924 ; 0.954) $\gamma$ : 0.778 (0.751 ; 0.803)	$\alpha$ : 0.776 (0.749 ; 0.801) $\beta$ : 0.947 (0.931 ; 0.960) $\gamma$ : 0.725 (0.695 ; 0.752)
P = 0.2	$\alpha$ : 0.845 (0.821 ; 0.866) $\beta$ : 0.930 (0.912 ; 0.945) $\gamma$ : 0.808 (0.782 ; 0.822)	$\alpha$ : 0.741 (0.712 ; 0.768) $\beta$ : 0.931 (0.913 ; 0.945) $\gamma$ : 0.679 (0.649 ; 0.708)	$\alpha$ : 0.703 (0.673 ; 0.731) $\beta$ : 0.958 (0.943 ; 0.969) $\gamma$ : 0.615 (0.584 ; 0.645)
P = 0	$\alpha$ : 0.761 (0.733 ; 0.787) $\beta$ : 0.905 (0.885 ; 0.922) $\gamma$ : 0.690 (0.660 ; 0.718)	$\alpha$ : 0.670 (0.640 ; 0.699) $\beta$ : 0.940 (0.923 ; 0.953) $\gamma$ : 0.567 (0.535 ; 0.598)	$\alpha$ : 0.562 (0.530 ; 0.593) $\beta$ : 0.974 (0.961 ; 0.982) $\gamma$ : 0.446 (0.415 ; 0.477)

Tabella 6. Livello di copertura reale per intervalli di confidenza unilaterali, calcolati considerando ipotesi unilaterale sinistra (valore a sx) e destra (valore a dx). Livello nominale 0.95.

Simulo i dati sotto l'ipotesi che non ci sia discontinuità e stimo il modello senza discontinuità <b>(S0.M0)</b> .						
P = 1 $\alpha$ : 0.951 e 0.954 $\beta$ : 0.957 e 0.952 $\gamma$ : -						
	$\rho = 0.1$		$\rho = 0.2$		$\rho = 0.3$	
P = 0.5	$\alpha$ : 0.941 $\beta$ : 0.978	0.957 0.904	$\alpha$ : 0.945 $\beta$ : 0.989	0.949 0.810	$\alpha$ : 0.951 $\beta$ : 0.997	0.948 0.657
P = 0.4	$\alpha$ : 0.952 $\beta$ : 0.987	0.951 0.885	$\alpha$ : 0.954 $\beta$ : 0.997	0.961 0.763	$\alpha$ : 0.963 $\beta$ : 1.000	0.949 0.575
P = 0.3	$\alpha$ : 0.962 $\beta$ : 0.979	0.946 0.874	$\alpha$ : 0.944 $\beta$ : 1.000	0.949 0.746	$\alpha$ : 0.951 $\beta$ : 1.000	0.963 0.469
P = 0.2	$\alpha$ : 0.953 $\beta$ : 0.984	0.946 0.883	$\alpha$ : 0.947 $\beta$ : 0.999	0.944 0.681	$\alpha$ : 0.954 $\beta$ : 1.000	0.957 0.395
P = 0	$\alpha$ : 0.947 $\beta$ : 0.989	0.950 0.828	$\alpha$ : 0.935 $\beta$ : 0.999	0.954 0.559	$\alpha$ : 0.957 $\beta$ : 1.000	0.954 0.251

Simulo i dati sotto l'ipotesi che ci sia discontinuità e stimo il modello senza discontinuità <b>(S1.M0)</b> .						
P = 1 $\alpha$ : 0.002 e 1.000 $\beta$ : 0.015 e 1.000 $\gamma$ : -						
	$\rho = 0.1$		$\rho = 0.2$		$\rho = 0.3$	
P = 0.5	$\alpha$ : 0.000 $\beta$ : 0.05	1.000 1.000	$\alpha$ : 0.002 $\beta$ : 0.092	1.000 1.000	$\alpha$ : 0.000 $\beta$ : 0.269	1.000 1.000
P = 0.4	$\alpha$ : 0.000 $\beta$ : 0.042	1.000 1.000	$\alpha$ : 0.000 $\beta$ : 0.142	1.000 1.000	$\alpha$ : 0.002 $\beta$ : 0.393	1.000 0.999
P = 0.3	$\alpha$ : 0.002 $\beta$ : 0.058	1.000 1.000	$\alpha$ : 0.001 $\beta$ : 0.21	1.000 1.000	$\alpha$ : 0.001 $\beta$ : 0.485	1.000 1.000
P = 0.2	$\alpha$ : 0.000 $\beta$ : 0.059	1.000 1.000	$\alpha$ : 0.001 $\beta$ : 0.236	1.000 1.000	$\alpha$ : 0.000 $\beta$ : 0.588	1.000 1.000
P = 0	$\alpha$ : 0.001 $\beta$ : 0.988	1.000 0.836	$\alpha$ : 0.001 $\beta$ : 1.000	1.000 0.605	$\alpha$ : 0.000 $\beta$ : 1.000	1.000 0.272

Simulo i dati sotto l'ipotesi che non ci sia discontinuità e stimo il modello con discontinuità <b>(S0.M1)</b> .						
P = 1 $\alpha$ : 0.962 e 0.951 $\beta$ : 0.958 e 0.945 $\gamma$ : 0.949 e 0.964						
	$\rho = 0.1$		$\rho = 0.2$		$\rho = 0.3$	
P = 0.5	$\alpha$ : 0.949 $\beta$ : 0.968 $\gamma$ : 0.956	0.952 0.936 0.947	$\alpha$ : 0.960 $\beta$ : 0.991 $\gamma$ : 0.950	0.937 0.870 0.958	$\alpha$ : 0.952 $\beta$ : 0.994 $\gamma$ : 0.931	0.926 0.790 0.950
P = 0.4	$\alpha$ : 0.956 $\beta$ : 0.974 $\gamma$ : 0.950	0.954 0.923 0.952	$\alpha$ : 0.947 $\beta$ : 0.982 $\gamma$ : 0.945	0.945 0.872 0.949	$\alpha$ : 0.950 $\beta$ : 0.991 $\gamma$ : 0.932	0.922 0.747 0.958
P = 0.3	$\alpha$ : 0.966 $\beta$ : 0.976 $\gamma$ : 0.949	0.955 0.897 0.962	$\alpha$ : 0.959 $\beta$ : 0.990 $\gamma$ : 0.948	0.956 0.819 0.961	$\alpha$ : 0.965 $\beta$ : 0.997 $\gamma$ : 0.944	0.941 0.709 0.962
P = 0.2	$\alpha$ : 0.964 $\beta$ : 0.984 $\gamma$ : 0.950	0.944 0.913 0.962	$\alpha$ : 0.957 $\beta$ : 0.993 $\gamma$ : 0.952	0.952 0.825 0.961	$\alpha$ : 0.965 $\beta$ : 0.999 $\gamma$ : 0.939	0.939 0.649 0.963
P = 0	$\alpha$ : 0.936 $\beta$ : 0.988 $\gamma$ : 0.953	0.946 0.821 0.947	$\alpha$ : 0.951 $\beta$ : 1.000 $\gamma$ : 0.936	0.944 0.602 0.947	$\alpha$ : 0.948 $\beta$ : 1.000 $\gamma$ : 0.949	0.949 0.232 0.957

Simulo i dati sotto l'ipotesi che ci sia discontinuità e stimo il modello con discontinuità <b>(S1.M1)</b> .						
P = 1 $\alpha$ : 0.944 e 0.958 $\beta$ : 0.950 e 0.951 $\gamma$ : 0.947 e 0.937						
	$\rho = 0.1$		$\rho = 0.2$		$\rho = 0.3$	
P = 0.5	$\alpha$ : 0.861 $\beta$ : 0.914 $\gamma$ : 0.986	0.983 0.975 0.844	$\alpha$ : 0.822 $\beta$ : 0.940 $\gamma$ : 0.991	0.994 0.960 0.778	$\alpha$ : 0.799 $\beta$ : 0.977 $\gamma$ : 0.996	0.989 0.918 0.752
P = 0.4	$\alpha$ : 0.823 $\beta$ : 0.902 $\gamma$ : 0.994	0.995 0.975 0.780	$\alpha$ : 0.767 $\beta$ : 0.938 $\gamma$ : 0.994	0.993 0.957 0.741	$\alpha$ : 0.738 $\beta$ : 0.966 $\gamma$ : 0.998	0.992 0.921 0.669
P = 0.3	$\alpha$ : 0.783 $\beta$ : 0.900 $\gamma$ : 0.999	0.997 0.984 0.746	$\alpha$ : 0.756 $\beta$ : 0.929 $\gamma$ : 0.997	0.991 0.967 0.664	$\alpha$ : 0.672 $\beta$ : 0.965 $\gamma$ : 1.000	0.999 0.930 0.589
P = 0.2	$\alpha$ : 0.762 $\beta$ : 0.896 $\gamma$ : 0.998	0.995 0.987 0.703	$\alpha$ : 0.657 $\beta$ : 0.905 $\gamma$ : 0.999	0.997 0.968 0.573	$\alpha$ : 0.596 $\beta$ : 0.973 $\gamma$ : 1.000	0.999 0.939 0.512
P = 0	$\alpha$ : 0.660 $\beta$ : 0.844 $\gamma$ : 0.999	0.997 0.984 0.573	$\alpha$ : 0.552 $\beta$ : 0.917 $\gamma$ : 1.000	1.000 0.983 0.430	$\alpha$ : 0.425 $\beta$ : 0.970 $\gamma$ : 1.000	1.000 0.966 0.312

Tabella 7. Test di significatività.

Simulo i dati sotto l'ipotesi che non ci sia discontinuità e stimo il modello senza discontinuità <b>(S0.M0)</b> . Per il parametro $\alpha$ valuto la potenza, ovvero la $\Pr\{R_\alpha \alpha\}$ Per il parametro $\beta$ valuto la potenza, ovvero la $\Pr\{R_\beta \beta\}$			
P = 1 $\alpha$ : 1 $\beta$ : 1 $\gamma$ : -			
	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
P = 0.5	$\alpha$ : 1 $\beta$ : 1	$\alpha$ : 1 $\beta$ : 1	$\alpha$ : 1 $\beta$ : 1
P = 0.4	$\alpha$ : 1 $\beta$ : 1	$\alpha$ : 1 $\beta$ : 1	$\alpha$ : 1 $\beta$ : 1
P = 0.3	$\alpha$ : 1 $\beta$ : 1	$\alpha$ : 1 $\beta$ : 1	$\alpha$ : 1 $\beta$ : 0.999
P = 0.2	$\alpha$ : 1 $\beta$ : 1	$\alpha$ : 1 $\beta$ : 1	$\alpha$ : 1 $\beta$ : 1
P = 0	$\alpha$ : 1 $\beta$ : 1	$\alpha$ : 1 $\beta$ : 1	$\alpha$ : 1 $\beta$ : 1

Simulo i dati sotto l'ipotesi che non ci sia discontinuità e stimo il modello senza discontinuità <b>(S0.M0)</b> . Per il parametro $\alpha$ valuto l'errore di secondo tipo, ovvero la $\Pr\{A_\alpha \alpha\}$ Per il parametro $\beta$ valuto l'errore di secondo tipo, ovvero la $\Pr\{A_\beta \beta\}$			
P = 1 $\alpha$ : 0 $\beta$ : 0 $\gamma$ : -			
	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
P = 0.5	$\alpha$ : 0 $\beta$ : 0	$\alpha$ : 0 $\beta$ : 0	$\alpha$ : 0 $\beta$ : 0
P = 0.4	$\alpha$ : 0 $\beta$ : 0	$\alpha$ : 0 $\beta$ : 0	$\alpha$ : 0 $\beta$ : 0
P = 0.3	$\alpha$ : 0 $\beta$ : 0	$\alpha$ : 0 $\beta$ : 0	$\alpha$ : 0 $\beta$ : 0.001
P = 0.2	$\alpha$ : 0 $\beta$ : 0	$\alpha$ : 0 $\beta$ : 0	$\alpha$ : 0 $\beta$ : 0
P = 0	$\alpha$ : 0 $\beta$ : 0	$\alpha$ : 0 $\beta$ : 0	$\alpha$ : 0 $\beta$ : 0

Tabella 8. Test di significatività

Simulo i dati sotto l'ipotesi che non ci sia discontinuità e stimo il modello con discontinuità <b>(S0.M1)</b> . Per il parametro $\alpha$ valuto la potenza, ovvero la $\Pr\{R_\alpha \alpha\}$ Per il parametro $\beta$ valuto la potenza, ovvero la $\Pr\{R_\beta \beta\}$ Per il parametro $\gamma$ valuto l'errore di primo tipo, ovvero $\Pr\{R_\gamma \gamma=0\}$			
P = 1 $\alpha$ : 1 $\beta$ : 0.966 $\gamma$ : 0.043			
	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
P = 0.5	$\alpha$ : 1 $\beta$ : 0.977 $\gamma$ : 0.045	$\alpha$ : 1 $\beta$ : 0.955 $\gamma$ : 0.043	$\alpha$ : 1 $\beta$ : 0.941 $\gamma$ : 0.061
P = 0.4	$\alpha$ : 1 $\beta$ : 0.968 $\gamma$ : 0.048	$\alpha$ : 1 $\beta$ : 0.966 $\gamma$ : 0.049	$\alpha$ : 1 $\beta$ : 0.933 $\gamma$ : 0.065
P = 0.3	$\alpha$ : 1 $\beta$ : 0.956 $\gamma$ : 0.039	$\alpha$ : 1 $\beta$ : 0.948 $\gamma$ : 0.046	$\alpha$ : 1 $\beta$ : 0.928 $\gamma$ : 0.051
P = 0.2	$\alpha$ : 1 $\beta$ : 0.970 $\gamma$ : 0.056	$\alpha$ : 1 $\beta$ : 0.951 $\gamma$ : 0.035	$\alpha$ : 1 $\beta$ : 0.909 $\gamma$ : 0.052
P = 0	$\alpha$ : 1 $\beta$ : 1 $\gamma$ : 0.040	$\alpha$ : 1 $\beta$ : 1 $\gamma$ : 0.029	$\alpha$ : 1 $\beta$ : 1 $\gamma$ : 0.051

Simulo i dati sotto l'ipotesi che non ci sia discontinuità e stimo il modello con discontinuità <b>(S0.M1)</b> . Per il parametro $\alpha$ valuto l'errore di secondo tipo, ovvero la $\Pr\{A_\alpha \alpha\}$ Per il parametro $\beta$ valuto l'errore di secondo tipo, ovvero la $\Pr\{A_\beta \beta\}$ Per il parametro $\gamma$ valuto il livello di confidenza, ovvero $\Pr\{A_\gamma \gamma=0\}$			
P = 1 $\alpha$ : 0 $\beta$ : 0.034 $\gamma$ : 0.957			
	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
P = 0.5	$\alpha$ : 0 $\beta$ : 0.023 $\gamma$ : 0.955	$\alpha$ : 0 $\beta$ : 0.045 $\gamma$ : 0.957	$\alpha$ : 0 $\beta$ : 0.059 $\gamma$ : 0.939
P = 0.4	$\alpha$ : 0 $\beta$ : 0.032 $\gamma$ : 0.952	$\alpha$ : 0 $\beta$ : 0.034 $\gamma$ : 0.951	$\alpha$ : 0 $\beta$ : 0.067 $\gamma$ : 0.935
P = 0.3	$\alpha$ : 0 $\beta$ : 0.044 $\gamma$ : 0.961	$\alpha$ : 0 $\beta$ : 0.052 $\gamma$ : 0.954	$\alpha$ : 0 $\beta$ : 0.072 $\gamma$ : 0.949
P = 0.2	$\alpha$ : 0 $\beta$ : 0.043 $\gamma$ : 0.944	$\alpha$ : 0 $\beta$ : 0.049 $\gamma$ : 0.965	$\alpha$ : 0 $\beta$ : 0.091 $\gamma$ : 0.948
P = 0	$\alpha$ : 0 $\beta$ : 0 $\gamma$ : 0.961	$\alpha$ : 0 $\beta$ : 0 $\gamma$ : 0.971	$\alpha$ : 0 $\beta$ : 0 $\gamma$ : 0.949

Tabella 9. Test di significatività

Simulo i dati sotto l'ipotesi che ci sia discontinuità e stimo il modello senza discontinuità <b>(S1.M0)</b> . Per il parametro $\alpha$ valuto la potenza, ovvero la $Pr\{R_\alpha \alpha\}$ Per il parametro $\beta$ valuto la potenza, ovvero la $Pr\{R_\beta \beta\}$			
P = 1 $\alpha$ : 1 $\beta$ : 1 $\gamma$ : -			
	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
P = 0.5	$\alpha$ : 1 $\beta$ : 1	$\alpha$ : 1 $\beta$ : 1	$\alpha$ : 1 $\beta$ : 1
P = 0.4	$\alpha$ : 1 $\beta$ : 1	$\alpha$ : 1 $\beta$ : 1	$\alpha$ : 1 $\beta$ : 1
P = 0.3	$\alpha$ : 1 $\beta$ : 1	$\alpha$ : 1 $\beta$ : 1	$\alpha$ : 1 $\beta$ : 1
P = 0.2	$\alpha$ : 1 $\beta$ : 1	$\alpha$ : 1 $\beta$ : 1	$\alpha$ : 1 $\beta$ : 1
P = 0	$\alpha$ : 1 $\beta$ : 1	$\alpha$ : 1 $\beta$ : 1	$\alpha$ : 1 $\beta$ : 1

Simulo i dati sotto l'ipotesi che ci sia discontinuità e stimo il modello senza discontinuità <b>(S1.M0)</b> . Per il parametro $\alpha$ valuto l'errore di secondo tipo, ovvero la $Pr\{A_\alpha \alpha\}$ Per il parametro $\beta$ valuto l'errore di secondo tipo, ovvero la $Pr\{A_\beta \beta\}$			
P = 1 $\alpha$ : 0 $\beta$ : 0 $\gamma$ : -			
	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
P = 0.5	$\alpha$ : 0 $\beta$ : 0	$\alpha$ : 0 $\beta$ : 0	$\alpha$ : 0 $\beta$ : 0
P = 0.4	$\alpha$ : 0 $\beta$ : 0	$\alpha$ : 0 $\beta$ : 0	$\alpha$ : 0 $\beta$ : 0
P = 0.3	$\alpha$ : 0 $\beta$ : 0	$\alpha$ : 0 $\beta$ : 0	$\alpha$ : 0 $\beta$ : 0
P = 0.2	$\alpha$ : 0 $\beta$ : 0	$\alpha$ : 0 $\beta$ : 0	$\alpha$ : 0 $\beta$ : 0
P = 0	$\alpha$ : 0 $\beta$ : 0	$\alpha$ : 0 $\beta$ : 0	$\alpha$ : 0 $\beta$ : 0

Tabella 10. Test di significatività

Simulo i dati sotto l'ipotesi che ci sia discontinuità e stimo il modello con discontinuità <b>(S1.M1)</b> Per il parametro $\alpha$ valuto la potenza, ovvero la $\Pr\{R_\alpha \alpha\}$ Per il parametro $\beta$ valuto la potenza, ovvero la $\Pr\{R_\beta \beta\}$ Per il parametro $\gamma$ valuto la potenza, ovvero la $\Pr\{R_\gamma \gamma\}$			
P = 1 $\alpha$ : 1 $\beta$ : 0.979 $\gamma$ : 0.831			
	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
P = 0.5	$\alpha$ : 1 $\beta$ : 0.989 $\gamma$ : 0.600	$\alpha$ : 1 $\beta$ : 0.988 $\gamma$ : 0.541	$\alpha$ : 1 $\beta$ : 0.980 $\gamma$ : 0.530
P = 0.4	$\alpha$ : 1 $\beta$ : 0.989 $\gamma$ : 0.550	$\alpha$ : 1 $\beta$ : 0.986 $\gamma$ : 0.496	$\alpha$ : 1 $\beta$ : 0.987 $\gamma$ : 0.442
P = 0.3	$\alpha$ : 1 $\beta$ : 0.994 $\gamma$ : 0.505	$\alpha$ : 1 $\beta$ : 0.993 $\gamma$ : 0.397	$\alpha$ : 1 $\beta$ : 0.989 $\gamma$ : 0.318
P = 0.2	$\alpha$ : 1 $\beta$ : 0.994 $\gamma$ : 0.455	$\alpha$ : 1 $\beta$ : 0.995 $\gamma$ : 0.318	$\alpha$ : 1 $\beta$ : 0.994 $\gamma$ : 0.267
P = 0	$\alpha$ : 1 $\beta$ : 0.998 $\gamma$ : 0.344	$\alpha$ : 1 $\beta$ : 0.998 $\gamma$ : 0.211	$\alpha$ : 1 $\beta$ : 0.996 $\gamma$ : 0.124

Simulo i dati sotto l'ipotesi che ci sia discontinuità e stimo il modello con discontinuità <b>(S1.M1)</b> Per il parametro $\alpha$ valuto l'errore di secondo tipo, ovvero la $\Pr\{A_\alpha \alpha\}$ Per il parametro $\beta$ valuto l'errore di secondo tipo, ovvero la $\Pr\{A_\beta \beta\}$ Per il parametro $\gamma$ valuto l'errore di secondo tipo, ovvero la $\Pr\{A_\gamma \gamma\}$			
P = 1 $\alpha$ : 0 $\beta$ : 0.021 $\gamma$ : 0.169			
	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$
P = 0.5	$\alpha$ : 0 $\beta$ : 0.011 $\gamma$ : 0.400	$\alpha$ : 0 $\beta$ : 0.012 $\gamma$ : 0.459	$\alpha$ : 0 $\beta$ : 0.020 $\gamma$ : 0.470
P = 0.4	$\alpha$ : 0 $\beta$ : 0.011 $\gamma$ : 0.450	$\alpha$ : 0 $\beta$ : 0.014 $\gamma$ : 0.504	$\alpha$ : 0 $\beta$ : 0.013 $\gamma$ : 0.558
P = 0.3	$\alpha$ : 0 $\beta$ : 0.006 $\gamma$ : 0.495	$\alpha$ : 0 $\beta$ : 0.007 $\gamma$ : 0.603	$\alpha$ : 0 $\beta$ : 0.011 $\gamma$ : 0.682
P = 0.2	$\alpha$ : 0 $\beta$ : 0.006 $\gamma$ : 0.545	$\alpha$ : 0 $\beta$ : 0.005 $\gamma$ : 0.682	$\alpha$ : 0 $\beta$ : 0.006 $\gamma$ : 0.733
P = 0	$\alpha$ : 0 $\beta$ : 0.002 $\gamma$ : 0.656	$\alpha$ : 0 $\beta$ : 0.002 $\gamma$ : 0.789	$\alpha$ : 0 $\beta$ : 0.004 $\gamma$ : 0.876

Grafici 1. Curve di potenza nel caso S1.M1.

