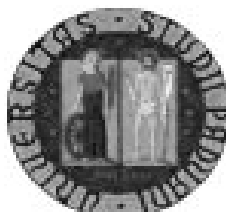


UNIVERSITÀ DEGLI STUDI DI PADOVA

FACOLTÀ DI SCIENZE STATISTICHE

CORSO DI LAUREA IN SCIENZE STATISTICHE ED ECONOMICHE



TESI DI LAUREA

**AUTONOMO O DIPENDENTE?
SCELTE LAVORATIVE E RISK SHARING
DELLE COPPIE ITALIANE**

RELATORE: CH.MO PROF. MINIACI RAFFAELE

LAUREANDO: GUIDO DALLA BONA

MATRICOLA N. 423607/SE

ANNO ACCADEMICO 2002-2003

A mio padre

Indice:

	<i>pag.</i>
INTRODUZIONE.....	1

CAPITOLO 1 L'offerta di lavoro e il risk sharing in ambito familiare

	<i>pag.</i>
<u>1 L'offerta di lavoro: introduzione al contesto intrafamiliare.....</u>	3
<u>2 Il modello unitario.....</u>	4
2.1 Analisi del modello unitario.....	4
2.2 Modello unitario con tasse e welfare.....	7
2.3 Limiti del modello unitario.....	7
<u>3 Il modello collettivo.....</u>	8
3.1 Analisi del modello collettivo.....	10
3.2 La funzione di produzione familiare.....	12
<u>4 Risk sharing: cos'è?.....</u>	14
4.1 Incertezza sui guadagni e risparmio.....	14
4.2 L'indice del grado di prudenza assoluta al rischio.....	15
4.3 Risparmio contro il rischio: conseguenze sui consumi.....	16
4.4 Risk sharing: i vantaggi sulle coppie che lavorano.....	17
<u>5 Un'analisi sulla struttura di covarianza dei guadagni intrafamiliari.....</u>	19
5.1 Presentazione dell'analisi empirica.....	20
5.2 I risultati salienti.....	21

CAPITOLO 2 L'influenza della famiglia sulle scelte individuali circa il tipo di contratto di lavoro

	<i>pag.</i>
<u>6 Offerta di lavoro: quali influenze sulle scelte individuali.....</u>	23
6.1 Il lavoro autonomo: appetibilità, difficoltà e soddisfazioni.....	24
6.2 Quando è più facile diventare lavoratori autonomi.....	26
6.3 Alcuni aspetti teorici rispetto alla scelta occupazionale.....	27
<u>7 Risk sharing nelle famiglie: analisi empiriche.....</u>	29
7.1 Test sull'ipotesi che vi sia completo risk sharing intrafamiliare e interfamiliare.....	30

8 L'analisi empirica inglese Employment Contract Matching	34
8.1 La struttura statistica dell'analisi inglese	35
8.2 I dati e i campioni utilizzati.....	37
8.3 I risultati salienti.....	40
8.4 A quali conclusioni si può arrivare	44

CAPITOLO 3 Introduzione all'analisi empirica

	<i>pag.</i>
9 Gli obiettivi.....	47
10 La fonte dei dati.....	48
10.1 Indagini sui bilanci delle famiglie nel 2000: il questionario	49
11 Come è formato il sotto-campione preso in analisi	50
11.1 La combinazione lavoratori autonomi - dipendenti tra i partner dei due sessi	51
11.2 L'ulteriore distinzione tra lavoratori dipendenti pubblici e privati	52
12 L'incidenza delle caratteristiche individuali e familiari sul tipo di lavoro	54
12.1 Le caratteristiche individuali.....	54
12.2 Le caratteristiche occupazionali	55
12.3 Alcuni aspetti demografici e geografici	56
12.4 Le caratteristiche familiari	57
13 Le variabili.....	58
13.1 La variabile dipendente	59
13.2 Le variabili indipendenti	60

CAPITOLO 4 I dati panel e i modelli probit

	<i>pag.</i>
14 I modelli con variabile dipendente discreta	63
14.1 I modelli con scelta binaria	63
14.2 La variabile latente sottostante	66
14.3 I modelli con risposte multiple.....	68
14.4 I modelli con risposte multiple ordinali	68
15 I dati longitudinali.....	71
15.1 I vantaggi dei dati panel	72
15.2 L'efficienza dei parametri stimati	73
15.3 L'identificazione dei parametri	74
15.4 Effetti fissi o casuali ?	76
15.5 I modelli con scelta binaria e dati panel.....	76
15.6 Perché sono interessato ai dati panel.....	77

<u>16</u> Il modello probit con dati panel.....	79
16.1 Il modello probit standard	79
16.2 Il modello probit standard ad effetti casuali.....	79
16.3 Gli effetti marginali.....	81
16.4 Il modello probit ordinale.....	82
16.5 Il modello probit standard bivariato	83

CAPITOLO 5 La lettura dei risultati dell'analisi empirica

	<i>pag.</i>
<u>17</u> L'impostazione dell'indagine	87
<u>18</u> La scelta binaria: risultati e commenti.....	88
18.1 Confronto fra sotto-campioni nell'univariato	88
18.2 L'output dei modelli probit standard.....	91
18.3 Alcuni risultati ottenuti nel bivariato	97
<u>19</u> La scelta tra tre opzioni: risultati e commenti.....	105
19.1 Confronto fra sotto-campioni	106
19.2 L'output dei modelli probit ordinali.....	109
<u>20</u> La "scelta" della moglie/convivente se lavorare o meno: quali influenze?	112
20.1 La "scelta" dei maschi dipende dalla condizione di lavoratrice della partner?.....	114
20.2 Quali influenze sulla condizione di lavoratrice o meno della partner.....	116
<u>21</u> Alcune considerazioni critiche sull'analisi empirica.....	121
 CONCLUSIONI	 127
 BIBLIOGRAFIA.....	 131

Introduzione

L'offerta di lavoro può essere pensata sotto un profilo individuale, ma, qualora l'individuo abbia un partner col quale forma una famiglia, assume importanza tener conto di tale contesto sociale. Infatti l'equilibrio tra i membri di una famiglia, a livello lavorativo, può essere influenzato da alcuni aspetti di carattere economico e sociale come, ad esempio, la composizione della famiglia stessa. In particolare la tesi si focalizza sul rapporto tra capofamiglia e partner rispetto alle occupazioni dei due, qualora entrambi lavorino. L'obiettivo è quello di indagare come si distribuiscono, all'interno delle coppie di coniugi/conviventi, i contratti lavorativi a seconda del grado di rischio che è loro associato.

L'incertezza dei guadagni dei capofamiglia e dei partner lavoratori può indurli a qualche forma di assicurazione contro il rischio. Se si pensa l'individuo senza tener conto del contesto familiare cui appartiene, viene logico considerare come miglior forma di assicurazione che ha a disposizione il risparmio; altrimenti l'individuo può ricorrere ad assicurazioni di tipo pubblico piuttosto che privato.

Tener conto dell'ambito familiare porta ad evidenziare come mettere assieme differenti rischi individuali, anche se tra loro correlati positivamente, riduca l'incertezza di ciascun membro della coppia, e quindi della coppia stessa; questo è, a grandi linee, quello che vedremo essere il fenomeno del risk sharing intra-coppia. Se i partner si collocano su livelli di rischio diversi si realizza una diversificazione che può assicurare ulteriormente i coniugi/conviventi. L'analisi empirica da me realizzata vuole per l'appunto indagare se in Italia vi sia una tendenza che veda i partner lavoratori appartenenti ad una stessa coppia impegnati in occupazioni con associati gradi di rischio diversi tra loro. Le categorie occupazionali cui si presume siano associati livelli di incertezza diversi sono quelle del lavoro dipendente e del lavoro autonomo; inoltre si considera l'ulteriore distinzione tra lavoro dipendente nel pubblico e nel privato.

L'elaborato è strutturato in cinque capitoli. Il primo descrive, a livello teorico, l'offerta di lavoro in ambito familiare, e si sofferma sul concetto di risk sharing; viene anche presentata un'analisi empirica sulla struttura di covarianza dei guadagni intra-familiari, riportandone i risultati salienti.

Il secondo capitolo illustra brevemente i fattori che influenzano la scelta lavorativa individuale quando si trascurano il contesto familiare di appartenenza, dando spazio anche ad alcune considerazioni sull'appeal del lavoro autonomo; inoltre vengono presentate le categorie occupazionali prese in considerazione. Il secondo capitolo è però soprattutto rivolto a descrivere alcune analisi che indagano l'eventuale presenza di risk sharing familiare; in particolare ci si sofferma sull'analisi empirica presentata in "Employment Contract Matching", di Sarah Brown, Lisa Farrel e John G. Sessions, che ha per oggetto lo studio sulla presenza di risk sharing intra-familiare e intra-coppia nel contesto inglese; è da questa lettura che prende spunto la mia analisi.

Il terzo capitolo fa da introduzione all'indagine empirica sul caso italiano. Vengono presentati gli obiettivi, la fonte dei dati, i sotto-campioni utilizzati; si misura l'incidenza di alcune caratteristiche individuali e familiari sulla scelta lavorativa individuale e si introducono le variabili dipendenti e indipendenti utilizzate.

Il quarto capitolo spiega a livello teorico come sono fatti i modelli di analisi utilizzati nell'indagine empirica, soffermandosi su alcune loro caratteristiche e sfaccettature; si spiega la scelta dei modelli probit con dati panel rispetto alle esigenze dell'analisi da me realizzata.

L'ultimo capitolo riporta i risultati ottenuti e li commenta; si traggono così delle conclusioni rispetto agli obiettivi che hanno guidato l'analisi empirica, in particolare indagando se vi sia risk sharing intra-coppia nel contesto italiano. Si mette anche in evidenza se e di quanto alcune variabili esogene diminuiscano/aumentino la probabilità che i partner facciano determinate "scelte" occupazionali, individualmente ma anche congiuntamente. Oltre al caso delle coppie in cui entrambi i partner lavorano, si illustrano e si commentano alcuni risultati ottenuti quando si prende in considerazione anche l'eventualità che l'individuo di sesso femminile non lavori; infatti si vuole vedere se, nel caso la moglie/convivente sia occupata, questa condizione influenzi o a sua volta dipenda dalla "scelta" dei partner di sesso maschile tra lavoro dipendente e autonomo. Infine si fanno alcune considerazioni critiche, sul piano teorico, rispetto all'impostazione dell'analisi empirica.

CAPITOLO 1 L'offerta di lavoro e il risk sharing in ambito familiare

1 L'offerta di lavoro: introduzione al contesto intrafamiliare

Studiare l'offerta di lavoro in ambito familiare dà rilievo e mette in luce diversi aspetti che possono essere propriamente capiti solo se collocati in tale contesto sociale. Ad esempio il sistema fiscale e i programmi di welfare che influenzano l'offerta di lavoro sono spesso pensati in prospettiva intrafamiliare. D'altra parte lo stesso equilibrio tra capofamiglia e partner, sotto il profilo dell'offerta di lavoro pensata nel contesto familiare, è a sua volta influenzato da come sono strutturati e cambiano i salari a seconda del sesso, oppure da come cambia la fertilità con l'evolversi della società.

Per studiare l'offerta di lavoro in ambito familiare occorre prima “mettersi d'accordo” su come pensare la famiglia; la si può identificare come un'unica unità decisionale, oppure, come vuole la più recente letteratura, si può vedere la famiglia come un gruppo di individui. A seconda di come si pensa la famiglia, le equazioni sull'offerta di lavoro possono rispondere a diversi modelli. Inoltre, a seconda del modello, cambia il modo in cui le politiche fiscali e di welfare incidono sull'offerta di lavoro.

Prima di approfondire tali aspetti, occorre soffermarsi brevemente sull'offerta di lavoro individuale, per poterla poi rapportare a quella calata nel contesto familiare. Uno dei modelli di offerta di lavoro è quello standard statico, in un determinato periodo di tempo t .

Si assuma che ogni individuo abbia una funzione di utilità quasi-concava

$$U(C_t, L_t, X_t)$$

con C_t , L_t e X_t rispettivamente consumo, tempo libero e caratteristiche individuali al tempo t . Dati W_t salario orario, Y_t reddito non da lavoro, e T tempo

totale a disposizione, l'utilità può essere massimizzata sotto il vincolo di bilancio individuale

$$C_t + W_t L_t = Y_t + W_t T$$

Il termine di destra include tutte le forme di reddito, e può essere indicato con M . La parte di reddito che non proviene dal lavoro può essere di due tipi, ereditario e "altro reddito non da lavoro".

Le condizioni del primo ordine prendono la seguente forma:

$$U_C(C_t, L_t, X_t) = \lambda_t, \quad U_L(C_t, L_t, X_t) \geq \lambda_t W_t$$

con λ_t utilità marginale del reddito. La disuguaglianza stretta implica che l'individuo non lavora; in tal caso il tempo libero a disposizione di un individuo viene a coincidere con quello complessivo ($L_t = T$).

Si può avere l'impressione sbagliata che le analisi sull'offerta di lavoro poggino sull'assunto che gli individui possano liberamente scegliere le loro ore di lavoro ad un dato salario con un singolo datore di lavoro; in realtà possono scegliere tra diversi pacchetti salariali, e conseguentemente il numero di ore di lavoro.

2 Il modello unitario

Il modello standard unitario dell'offerta di lavoro familiare tratta la famiglia come una singola unità decisionale. L'aspetto qualificante di questa "formula" sta nell'opportunità di mettere in evidenza il riscontro che tasse e assistenza sociale hanno sulle famiglie e non sui singoli individui; inoltre il modello dell'offerta di lavoro unitario può essere collocato in una struttura intertemporale.

2.1 Analisi del modello unitario

Si supponga una famiglia costituita da due individui in età da lavoro. Gli eventuali bambini e gli altri membri della famiglia che vi dipendano sono inclusi nel vettore delle caratteristiche della famiglia (X). I familiari sono assunti

massimizzare l'utilità combinata sul consumo, C , e il tempo libero di ciascuno dei due membri, indicato con L_1 e L_2 . Per tale famiglia, l'utilità è la seguente

$$U(C, L_1, L_2, X)$$

Il vincolo di bilancio (familiare) è ora

$$C + W_1 L_1 + W_2 L_2 = Y + W_1 T + W_2 T$$

mentre il reddito complessivo diventa

$$M = Y + W_1 T + W_2 T$$

con Y combinazione di tutti i redditi non da lavoro.

In tale contesto si considera il consumo misurato come un singolo aggregato, C . Le condizioni del primo ordine per il consumo individuale viste nel paragrafo precedente continuano a tenere, anche se ora determinano il consumo familiare; si considera la non partecipazione al mercato del lavoro da parte del secondo individuo:

$$U_{L1}(C, L_1, L_2, X) = \lambda W_1 \quad \text{e} \quad U_{L2}(C, L_1, L_2, X) \geq \lambda W_2$$

Se anche il secondo individuo lavora allora anche l'"ultima" condizione marginale è segnata dall'uguaglianza stretta.

Sostituendo l'utilità marginale del consumo $U_C = \lambda$, risulta:

$$U_{L1} - U_C W_1 = 0 \quad \text{e} \quad U_{L2} - U_C W_2 \geq 0,$$

con ciascuna utilità marginale funzione di L_1 e L_2 ; il consumo, dal vincolo di bilancio familiare, può essere scritto come $C = Y + W_1(T - L_1) + W_2(T - L_2)$.

L'offerta di lavoro ottima qui scelta soddisfa i vincoli standard di simmetria sulla domanda di consumo, la presenza di una matrice di sostituzione di Slutsky semidefinita negativa, e l'omogeneità di grado zero nei salari, nei prezzi e nel reddito (totale). L'omogeneità è soddisfatta specificando il modello dell'offerta

di lavoro in termini di salari reali e di redditi reali. La simmetria richiede l'uguaglianza tra i termini della matrice di sostituzione di Slutsky:

$$dL_i/dW_j + L^j dL_i/dM = dL_j/dW_i + L^i dL_j/dM \quad \text{per } i \neq j$$

Si deve anche testare l'eterogeneità nei termini d'errore; così le condizioni del primo ordine sulle utilità marginali danno

$$U_{L1} - U_C W_1 - \varepsilon_1 = 0 \quad \text{e} \quad U_{L2} - U_C W_2 - \varepsilon_2 \geq 0$$

dove ε_1 e ε_2 indicano i termini d'errore; $g(\varepsilon_1, \varepsilon_2)$ è la densità congiunta.

Le condizioni del primo ordine descrivono due diversi ambiti di comportamento:

(1) entrambi i partner partecipano al mercato del lavoro:

$$H_1 = T - L_1 > 0, H_2 = T - L_2 > 0 ;$$

(2) l'individuo 2 non vi partecipa:

$$H_1 = T - L_1 > 0, H_2 = T - L_2 = 0$$

con H_1 e H_2 indicanti il numero di ore di lavoro per ciascuno dei due membri adulti della famiglia.

La funzione di verosimiglianza per le famiglie che appartengono alla prima situazione, ovvero con entrambi gli "sposi" che partecipano al mercato del lavoro, è data da:

$$l_{H1>0, H2>0} = |J| g(U_{L1} - U_C W_1, U_{L2} - U_C W_2)$$

con $|J|$ termine Jacobiano, determinante della matrice incrociata delle derivate di ε_1 ed ε_2 rispetto alle ore di lavoro; ε_1 e ε_2 sono funzioni non lineari di H_1 e H_2 .

Nel caso di non partecipazione al mercato del lavoro dell'individuo 2, cioè se $U_{L2} - U_C W_2 - \varepsilon_2 > 0$, la scelta di L_1 richiede di risolvere le condizioni marginali con $L_2 = T$, ottenendo $\tilde{U}_{L1} - \tilde{U}_C W_1 - \varepsilon_1 = 0$. Conseguentemente la funzione di verosimiglianza in regime di non partecipazione dell'individuo 2 al mercato del lavoro è data da:

$$l_{H1>0,H2=0} = \int_{U_{L2}-U_C W_2}^{\infty} |k| g(\tilde{U}_{L1} - \tilde{U}_C W_1, \varepsilon_2) d\varepsilon_2$$

con $|k|$ termine Jacobiano. Le matrici J e K sono positive definite, proprietà garantita dalle condizioni di simmetria e di negatività di Slutsky.

2.2 Modello unitario con tasse e partecipazione a programmi di assistenza sociale

Fin qui si sono ignorate le tasse e le ricadute delle politiche di welfare; se invece le si include nel modello, si può supporre che le famiglie massimizzino una funzione di utilità della forma:

$$U = U(L_1, L_2, C, \varepsilon) - \eta P_B$$

con P_B indicatore di un'eventuale partecipazione a programmi di welfare o meno. L'eterogeneità non osservabile entra direttamente nell'utilità attraverso il vettore ε che corrisponde ai precedentemente visti ε_1 ed ε_2 . ηP_B "cattura" i costi del welfare.

Il vincolo di bilancio familiare è dato da:

$$C = W_1 H_1 + W_2 H_2 + Y - T(Y, W_1 H_1, W_2 H_2) + B P_B$$

dove Y indica il reddito non guadagnato, $T(\cdot)$ è una funzione delle tasse, e B sta per benefici del programma di welfare.

2.3 Limiti del modello unitario

Il modello unitario descritto nel paragrafo precedente porta con sé tre "categorie" di restrizioni.

Innanzitutto le restrizioni sulla simmetria, sulle proprietà della matrice di sostituzione di Slutsky, e sull'omogeneità di grado zero nei salari, nei prezzi e nel reddito, guadagnato col lavoro e non.

Il secondo set di restrizioni riguarda l'income pooling. Questa condizione implica che, nel contesto dell'offerta di lavoro familiare, tutte le fonti di reddito non guadagnato col lavoro vengono inserite in un'unica misura di reddito Y . Ad esempio, se entrambi i partner di una famiglia hanno un reddito non da lavoro, indicato rispettivamente con Y_1 e Y_2 , l'income pooling fa sì che

$$L_i^{Y1} = L_i^{Y2}, \text{ per } i = 1 \text{ e } 2, \text{ dove } L_i^{Y1} = dL_i/dY_1$$

Questa assunzione è alquanto controversa nel dibattito sul welfare e sulle sue riforme, perché suppone che la fonte di reddito non guadagnato col lavoro sia irrilevante, nelle decisioni sull'offerta di lavoro familiare.

Infine le condizioni sulla non partecipazione al mercato del lavoro portano all'individuo "che non partecipa" (soluzione d'angolo) un salario di riserva che proviene dal partner che invece lavora, influenzandone le scelte sull'offerta di lavoro. Come nel caso dell'assunzione sull'income pooling, questo è lontano dall'essere irrilevante, implicando come l'opzione sulla partecipazione al mercato del lavoro non influenzi l'allocazione di consumo e tempo libero all'interno della famiglia.

3 *Il modello collettivo*

La ricerca più recente si è focalizzata su come rilassare le assunzioni sulla simmetria e sull'income pooling, guardando invece alla teoria della contrattazione collettiva. I vantaggi dell'offerta di lavoro familiare analizzata sotto una prospettiva unitaria sono molteplici; vi è la possibilità di utilizzare direttamente la teoria economica del consumo, recuperando quelle che sono le preferenze osservate dal comportamento in modo netto, non ambiguo, e provvedendo a realizzare un'interpretazione dei risultati empirici; inoltre il modello unitario consente di evidenziare le ricadute delle politiche sociali e fiscali sulle famiglie, e in particolare il loro impatto sull'offerta di lavoro familiare. Va però detto che il modello standard unitario è stato sottoposto a critica dai suoi detrattori, oltre per i motivi visti nel paragrafo 2.3, proprio per la sua impostazione, che tratta tutti i membri appartenenti ad una famiglia come

un'unica unità decisionale, piuttosto che come un gruppo di individui. In ogni caso è senz'altro troppo forte sostenere che il modello unitario non sia neppure recuperabile in modo significativo, applicando la teoria del consumo alle singole utilità individuali dei membri di una famiglia; è infatti questo ciò che si ripropone di fare il modello collettivo.

Si supponga che non vi siano beni pubblici familiari (come luce e gas) e che le utilità individuali siano debolmente separate rispetto al consumo e al tempo libero privati, personali; così C_1, L_1, C_2 e L_2 indicano il "singolo" consumo e il "singolo" tempo libero dei soggetti 1 e 2 appartenenti ad una data famiglia. Separando allo stesso modo le utilità individuali, l'utilità familiare diventa

$$F(U_1(C_1, L_1, X), U_2(C_2, L_2, X))$$

dove $U_1(C_1, L_1, X)$ è la sotto-utilità del marito, e $U_2(C_2, L_2, X)$ è la sotto-utilità della moglie. Quando l'utilità familiare ha questa forma "debolmente separata" la decentralizzazione porta ad un bilancio familiare a due-stadi; il reddito totale della famiglia è distribuito tra tutti i membri della famiglia stessa. A questo punto ciascun familiare agisce come se stesse prendendo delle decisioni sull'offerta di lavoro e sul consumo condizionatamente a questo stadio di spesa iniziale.

Anche se i beni sono consumati privatamente, può accadere che essi siano misurati solo a livello familiare, cosicché i consumi individuali possano divenire non osservabili (o latenti). Tuttavia, per un singolo bene osservato, di consumo privato come ad esempio è il caso dell'offerta di lavoro, se si utilizzano le sotto-utilità individuali di marito e moglie, è spesso sufficiente identificare le singole funzioni di preferenza affinché i consumi individuali diventino osservabili.

L'approccio collettivo dà dei vantaggi, quali? Rispetto al modello unitario le cui condizioni marginali sottolineano la struttura di ottimizzazione congiunta, il modello collettivo, che fra breve sarà meglio analizzato, "rilassa" l'allocazione del reddito tra individui, mettendo in luce la singola posizione contrattuale di ciascun membro all'interno della famiglia; in tal modo l'allocazione del reddito dipende più direttamente dai singoli soggetti e dai singoli salari.

3.1 *Analisi del modello collettivo*

L'approccio voluto dalla più recente letteratura vede ogni membro della famiglia massimizzare una propria utilità "egoistica", $U_j(C_j, L_j, X)$, o al più una funzione di utilità "meno egoistica" (*ci si prende cura di*), $F_j(U_1(C_1, L_1, X), U_2(C_2, L_2, X))$, con $j = 1$ e 2 , secondo quello che è l'assunto di separabilità in sotto-utilità dell'utilità familiare visto precedentemente.

Secondo la seconda funzione di utilità descritta, quella "meno egoistica", l'unico modo attraverso cui C_1 e L_1 "entrano" nell'utilità del secondo individuo è attraverso la sotto-utilità dell'individuo 1; quindi non vi è diretto impatto sull'utilità del partner.

Le applicazioni di questo modello assumono che il processo decisionale generi soluzioni Pareto-efficienti (allocazioni di beni raggiunte le quali non è più possibile aumentare il benessere di un individuo senza ridurne quello di un altro) e che tutti i beni siano consumati privatamente, senza che vi sia produzione familiare.

Il modello specifica il problema sull'offerta di lavoro familiare come segue:

$$\max[\theta U_1 + (1-\theta)U_2], \quad \text{sotto il vincolo } M = C_1 + C_2 + W_1 L_1 + W_2 L_2$$

con θ peso dell'utilità per l'individuo 1, dato dalla funzione non negativa $\theta = f(W_1, W_2, M)$. Questo è equivalente ad una ripartizione tale che l'individuo 1 ottenga un reddito pari a $M - \varphi(W_1, W_2, X, M)$, e ne massimizzi l'utilità secondo la regola:

$$\max U_1, \quad \text{sotto il vincolo } C_1 + W_1 L_1 = M - \varphi(W_1, W_2, X, M),$$

con $\varphi(W_1, W_2, X, M)$ definito come regola di ripartizione (sharing rule).

La sharing rule può essere di tipo proporzionale; in tal caso il reddito familiare è diviso in diverse porzioni prestabilite tra tutti i membri, e inoltre non vi sono beni pubblici (comuni a tutti i componenti, come il riscaldamento). Ogni individuo della famiglia sceglie il proprio livello di consumo e massimizza la sua utilità in base a qual è il suo vincolo di bilancio individuale.

Data la Pareto-efficienza e le assunzioni neoclassiche sulle utilità individuali, le condizioni che identificano le preferenze e la regola di ripartizione richiedono un dato bene osservabile di consumo privato, che può assumersi ad esempio essere il tempo libero a disposizione di un individuo. L'intuizione è la seguente: sotto l'assunzione di bene esclusivo, vale a dire di bene consumabile in modo privato da un singolo individuo, il salario del partner ha un effetto sul soggetto in considerazione solo attraverso la sharing rule.

Sebbene il modello collettivo non poggi su restrizioni proprie di quello unitario, quali quelle sulla simmetria e sull'income pooling, ve ne sono altre che lo interessano. Se le fonti di reddito separate non sono osservabili, ed entrambi gli individui lavorano (cioè non vi sono soluzioni d'angolo), le sole restrizioni sono quelle sulla condizione di Slutsky; queste sono espresse sotto forma di derivate delle equazioni sull'offerta di lavoro rispetto alle variabili del salario e del reddito. Assumendo che le derivate rispetto al reddito siano diverse da zero, il modello collettivo restituisce le equazioni differenziali:

$$\begin{aligned}\alpha_M(L_1^{W2} / L_1^M) + \alpha(d/dM)(L_1^{W2} / L_1^M) - \alpha_{W2} &= 0, \\ \beta_M(L_2^{W1} / L_2^M) + \beta(d/dM)(L_2^{W1} / L_2^M) - \beta_{W1} &= 0,\end{aligned}$$

con α dato da

$$\alpha = -\{[(d/dM)(L_1^{W2} / L_1^M)(L_2^{W1} / L_2^M) - (d/dW_2)(L_2^{W1} / L_2^M)] / [(d/dM)(L_2^{W1} / L_2^M) / L_2^M)(L_1^{W2} / L_1^M) - (d/dW_1)(L_1^{W2} / L_1^M)]\}^{-1}$$

$\beta = 1 - \alpha$, e con derivate parziali scritte come sopra. I termini α_M , β_M , α_{wj} , β_{wj} sono le corrispondenti derivate su reddito e salario. Le equazioni differenziali sopra scritte sono analoghe alle condizioni sulla simmetria di Slutsky, mentre le disuguaglianze di Slutsky sono comparabili a:

$$\begin{aligned}(L_1^{W1} / L_1^M) + (T - L_1 - (\beta/\alpha)(L_1^{W2} / L_1^M)) &\leq 0, \\ (L_2^{W2} / L_2^M) + (T - L_2 - (\beta/\alpha)(L_2^{W1} / L_2^M)) &\leq 0\end{aligned}$$

Queste restrizioni sono sufficienti per recuperare le funzioni di preferenza e la sharing rule. Invece, le derivate della sharing rule, $\varphi(W_1, W_2, X, M)$, hanno la forma:

$$d\varphi/dM = \alpha, \quad d\varphi/Dw_2 = \alpha (L_1^{W_2} / L_1^M), \quad d\varphi/Dw_1 = (1-\alpha) (L_2^{W_1} / L_2^M)$$

Conseguentemente, avendo stimato le funzioni sull'offerta di lavoro non vincolate, in termini di salari per ciascun reddito individuale, è possibile recuperare le funzioni di preferenza individuali e la sharing rule.

3.2 La funzione di produzione familiare

Illustrare la funzione di produzione familiare subito dopo aver sommariamente spiegato il modello collettivo potrebbe sembrare alquanto azzardato. Infatti abbiamo appena visto come il modello collettivo assuma una regola di ripartizione con individui le cui preferenze sono di tipo egoistico, oppure “si prendono cura” del partner.

E' ragionevole assumere che in molte famiglie il tempo libero sia speso per produrre attivamente beni per la famiglia, come ad esempio le faccende domestiche. L'esempio appena citato si riferisce ad un'attività che può trovare nel mercato un perfetto sostituto, si pensi alla colf; ma per beni per la famiglia si intendono anche attività per le quali non è immediato trovare un perfetto sostituto, per esempio “prendersi cura dei bambini”. Quest'ultimo tipo di attività rientra in quelle che il modello unitario, di non separazione dell'offerta di lavoro familiare, trascura, dandovi poco peso, impatto, come fosse semplice tempo libero della famiglia.

Supponiamo vi sia un bene prodotto per la famiglia che richieda disponibilità di tempo da parte di entrambi i partner; se t_1 e t_2 indicano il tempo messo a disposizione da entrambi gli individui adulti per la produzione del bene, la funzione di produzione tecnologica può essere scritta come

$$G = g(t_1, t_2)$$

con g funzione concava. Il tempo non speso nel mercato del lavoro può essere utilizzato in due diversi modi, come puro tempo libero o appunto per produzione familiare. Tuttavia, dato che il tempo libero rientra nella funzione di utilità della famiglia, e dato che $g(t_1, t_2)$ è concava, le equazioni sull'utilità della famiglia, descrivendo ore di lavoro e partecipazione al mercato del lavoro, sono osservabili in modo equivalente a quelle per il modello senza produzione familiare.

Un caso particolare interviene quando l'utilità familiare è separabile rispetto al tempo non lavorativo per ciascun individuo. In questo caso, l'utilità con produzione familiare può essere scritta

$$F(U_1(C_1, L_1, G_1, X), U_2(C_2, L_2, G_2, X))$$

dove G_1 e G_2 indicano il consumo privato del bene prodotto "a casa", con $G_1 + G_2 = G$.

Se il consumo di produzione familiare non è osservato allora la presenza di G in ciascuna sub-utilità viene a sconvolgere l'assunzione di separazione. Infatti se la produzione familiare dà uguali ritorni, il costo unitario della produzione diventa una semplice funzione dei due tassi salariali

$$P^* = r(W_1, W_2)$$

Nel modello senza G , la condizione di debole separazione è sufficiente affinché ciascuna funzione di offerta di lavoro sia scritta in termini di salario proprio e allocazione di reddito. Introducendo G , allora P^* , e quindi W_1 e W_2 , entrano in ciascuna offerta di lavoro. Conseguentemente la produzione familiare è sufficiente per rompere la condizione di separazione e quindi anche la stessa restrizione di esclusione sui salari degli altri membri della famiglia nelle equazioni sull'offerta di lavoro. Il solo caso in cui ciò non avviene è quello che vede il bene prodotto "in casa", G , commerciabile, e quando la soluzione è interna e non d'angolo; in questo caso P diventa uguale al prezzo di mercato osservabile e non dipende dai salari individuali.

4 Risk sharing: cos'è?

Nei paragrafi precedenti abbiamo visto l'offerta di domanda familiare e in particolare come, a seconda che si prenda in considerazione il modello unitario piuttosto che quello collettivo, si possa realizzare all'interno della famiglia income pooling o meno. Nel paragrafo 3.2 ci si è poi soffermati anche sul consumo familiare e su come questo venga diviso all'interno della famiglia; si è accennato alla sharing rule, e all'eventuale presenza di beni pubblici, vale a dire beni di consumo che non possono essere divisi tra i familiari (si pensi a luce, gas). Vi sono poi dei beni di consumo privati, di consumo individuale. Ogni membro della famiglia, secondo quello che è l'approccio del modello collettivo, viene così ad avere una funzione di sotto-utilità che dipende da x_i e y , con x_i vettore dei beni privati consumati dall'individuo i , e y vettore dei beni di consumo pubblico della famiglia. Secondo l'approccio unitario anche i beni privati rientrano in un unico grande "contenitore" familiare. Ma al di là di ciò che viene fatto sui beni di consumo individuale, resta il fatto che i beni pubblici sono comuni a tutta la famiglia e relativamente a questo aspetto si può introdurre il concetto di assicurazione all'interno della famiglia; si mettono assieme dei beni, e quindi si mette in comune e si distribuisce il relativo rischio. Mi piace sottolineare che questo può avvenire sia se vi è income pooling (approccio unitario), sia se non c'è (approccio collettivo).

Detto questo, vado a spiegare più nel dettaglio cos'è il risk sharing, in particolare nel paragrafo 4.4; la mia tesi lo tratta soprattutto rispetto al lavoro, con le dovute conseguenze a cascata su consumi e risparmio.

4.1 Incertezza sui guadagni e risparmio

La recente letteratura ha studiato gli effetti dell'incertezza sul risparmio. Questo ha contribuito a spiegare la discrepanza tra analisi teorica e risultati empirici sul modello su consumo e risparmio di un dato individuo. Ad esempio è proprio l'incertezza estesa a tutto il ciclo di vita di un individuo che può spiegare come

mai chi va in pensione decumuli la propria ricchezza in modo più lento di quanto ci si potrebbe aspettare dalla teoria economica.

L'incertezza sui guadagni fa comprendere la rilevante sensitività del consumo rispetto alle fluttuazioni del reddito; una immediata conseguenza dell'incertezza è il boom di assicurazioni private e pubbliche che si è avuto negli anni Ottanta in diversi paesi industrializzati.

Poche sono però le analisi empiriche che sono state fatte su questi aspetti, in particolare sulla correlazione tra propensione al risparmio e ricerca di assicurazione contro il rischio. Friedman ha trovato che l'attitudine al risparmio è molto più alta in quelle famiglie con occupazioni lavorative rischiose, anche se i risultati empirici emersi a tal proposito sono contraddittori (Skinner).

Carroll e Samwick hanno studiato la storia dei redditi familiari e la loro varianza, trovando come l'incertezza sui guadagni abbia un peso fondamentale sulla gestione del reddito familiare. Anche Guiso, Jappelli e Terlizzese hanno focalizzato la relazione tra varianza della distribuzione del reddito come è percepito dalla famiglia, e risparmio.

4.2 *L'indice del grado di prudenza assoluta al rischio*

L'incertezza sui guadagni può portare gli individui ad assicurare se stessi contro le fluttuazioni dei salari formando una famiglia. Questo tipo di assicurazione può così andare a sostituire le soluzioni che si possono trovare nel pubblico e anche nel privato.

L'idea che sta alla base del risk sharing, rispetto al lavoro, è che le famiglie con più redditi sono più facilmente protette contro gli shocks sul reddito rispetto a quelle che dipendono dal solo salario del capofamiglia; conseguentemente le famiglie con guadagni provenienti da più membri di una stessa famiglia risparmiano e si assicurano per via tradizionale meno di quelle con unica entrata.

Se il "rischio da reddito" non è assicurabile, l'incertezza sui guadagni aumenta il risparmio e l'accumulazione di ricchezza, se la derivata terza della funzione di utilità è positiva. Se vale questa condizione, un aumento dell'income risk implica un maggiore ricorso al risparmio. Kimball ha introdotto una misura del grado di

prudenza, il rapporto tra la derivata terza e la derivata seconda della funzione di utilità, $p(w) = -v'''(w)/v''(w)$, con v utilità di secondo periodo e w ricchezza della famiglia. Per piccoli rischi, il risparmio è proporzionale a $p(w)$.

L'indice $p(w)$ misura il grado di prudenza assoluta al rischio. Dal momento che la prudenza varia con la ricchezza, il risparmio a scopo precauzionale può essere più alto o basso a seconda che la prudenza diminuisca o aumenti con la ricchezza. Kimball e Weil sostengono che una prudenza che decresce è proprietà plausibile delle funzioni di utilità.

4.3 Risparmio contro il rischio: conseguenze sui consumi

Dato un piano di assi cartesiani (*vedere figura 1 a pag. 17*) con sulle ordinate il livello di consumo c e sulle ascisse la ricchezza w , si suppone di avere una curva crescente e concava che rappresenta la funzione di consumo (relazione tra consumi e reddito disponibile). Un aumento del rischio porta ad una caduta verso il basso della funzione di consumo (*diventa c'*), perché è necessaria più ricchezza per sostenere lo stesso livello di consumo. Alternativamente, dato un ammontare di ricchezza, il consumo corrente deve diminuire per permettere più risparmio. Se la prudenza assoluta al rischio è costante, la funzione di consumo è funzione lineare rispetto alla ricchezza e la variazione nel consumo è indipendente dal livello di ricchezza; la *linea tratteggiata* indica che ad ogni livello di ricchezza il rapporto tra consumo e ricchezza cade, mentre la propensione marginale al consumo, che misura di quanto varia il consumo quando il reddito disponibile aumenta di una unità, non cambia.

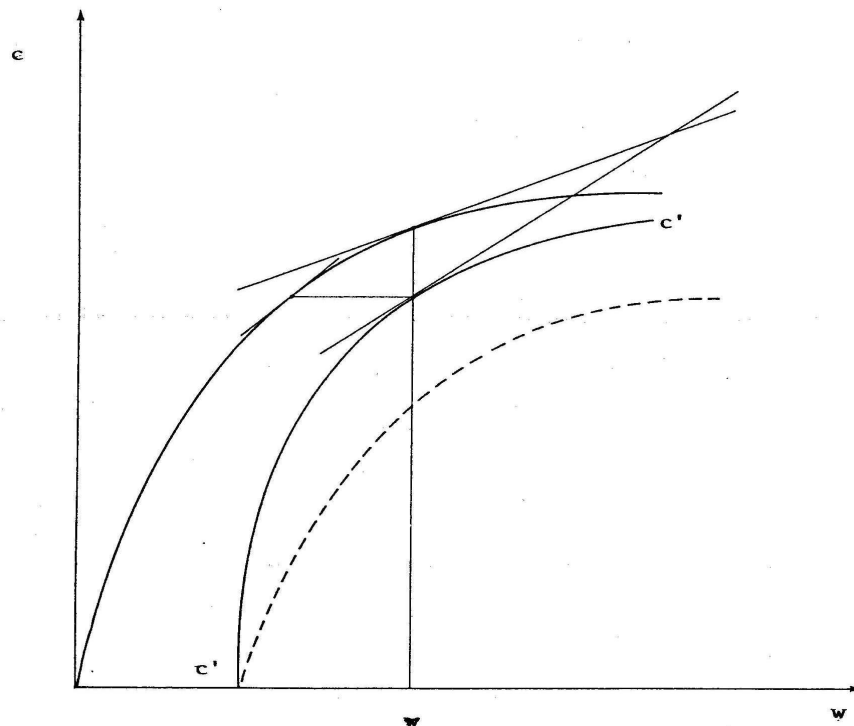


figura 1

Se tuttavia la prudenza cala con l'aumentare della ricchezza, l'ammontare di reddito addizionale che occorre per finanziare l'equivalente livello di consumo cade all'aumentare della ricchezza stessa. In questo caso la funzione di consumo è non lineare rispetto alla ricchezza, e un incremento del rischio porta la propensione media al consumo (rapporto tra consumo e reddito) a calare, ma con un calo di prudenza al rischio la propensione marginale al consumo aumenta ad ogni livello di reddito. Questo fa sì che il reddito aumenti, e inoltre fa sentire il consumatore più sicuro; entrambi gli effetti stimolano il consumo.

4.4 Risk sharing: i vantaggi sulle coppie che lavorano

L'idea che l'incertezza stimoli il risparmio è intuitiva, ma è difficile realizzare un test che misuri tale tipo di assicurazione; questo perché la soggettiva incertezza dell'individuo rispetto agli eventi futuri non è osservabile. Inoltre il risparmio a scopo assicurativo può essere la conseguenza di specifici rischi individuali, come di rischi di tipo aggregato. La stessa propensione media al consumo può essere più bassa per determinate categorie di popolazione; in particolare Friedman e

Skinner hanno testato se i lavoratori autonomi, in quanto appartenenti ad una categoria occupazionale più rischiosa di quella cui appartengono i lavoratori dipendenti, tendano a consumare meno degli altri.

Il tentativo di Friedman e Skinner ha trovato alcuni ostacoli:

- 1) non c'è evidenza che la percezione di incertezza sul reddito sia più grande per i lavoratori autonomi piuttosto che per gli altri lavoratori;
- 2) supposto che l'ordinamento degli income risks sia corretto, c'è un problema di auto-selezione: se la prudenza è correlata all'avversione al rischio, e gli individui meno avversi al rischio scelgono le occupazioni più rischiose, essi possono benissimo consumare più di quanto non facciano i lavoratori di occupazioni più sicure;
- 3) i test empirici non distinguono tra propensione media e marginale al consumo; questa distinzione è invece importante, come è evidente dal paragrafo 4.3.

Per misurare l'effetto dell'incertezza in modo empirico, è proponibile un test basato sull'idea che le famiglie con più redditi da lavoro assicurano parte del rischio associato a ciascun reddito individuale. Le famiglie che si assicurano in tal modo contro il rischio vedono attenuarsi il bisogno di ricorrere ad un risparmio a scopo precauzionale. E' facile dimostrare come mettere insieme differenti rischi individuali, anche se tra loro correlati positivamente, riduca il rischio (la varianza) di ciascun membro della famiglia, e quindi della famiglia stessa.

Date due variabili casuali u_1 e u_2 , identicamente distribuite con media zero, varianza σ_u^2 e coefficiente di correlazione ρ , la variabile casuale z , ottenuta come combinazione lineare di u_1 e u_2 con peso $1/2$, ha varianza

$$\sigma_z^2 = (1/2)(1+\rho)\sigma_u^2.$$

$$\text{Se } \rho < 1, \sigma_z^2 < \sigma_u^2.$$

Le famiglie beneficiano del risk sharing anche se i rischi non sono identicamente distribuiti. Si supponga che u_1 e u_2 abbiano un coefficiente di correlazione

positivo ρ , ma differenti varianze σ_1^2 e $k^2 \sigma_1^2$ rispettivamente, con $k > 1$; siano z_1 e z_2 due variabili casuali ottenute come combinazione lineare di u_1 e u_2

$$z_1 = \beta u_1 + (1-\beta) u_2$$

$$z_2 = (1-\beta) u_1 + \beta u_2$$

La varianza di z_1 è allora

$$E(z_1^2) = [\beta^2 + (1-\beta)^2 K^2 + 2\beta(1-\beta)\rho K] \sigma_1^2$$

Se $\rho < (1/k)$, il valore di β che massimizza $E(z_1^2)$ è

$$\beta = [k(k-\rho)]/[1+k(k-2\rho)]$$

Questo valore di β implica che $E(z_1^2) < \sigma_1^2$ e che $E(z_2^2) < k^2 \sigma_1^2$. Così, se il valore di ρ non eccede la radice quadrata del rapporto tra la più bassa e la più alta varianza, esiste una combinazione di rischi tale che il nuovo rischio individuale sia più piccolo di quello originale, precedente al risk sharing.

Si è così messo in luce come le condizioni sotto le quali il risk sharing è vantaggioso sono abbastanza generali, non ristrette al solo caso di rischi indipendenti ed identicamente distribuiti.

5 Un'analisi sulla struttura di covarianza dei guadagni intrafamiliari

A breve riporto un'indagine sui guadagni familiari che poggia sull'analisi delle strutture di covarianza e correlazione. Questo tipo di analisi è importante perché mette in evidenza diversi aspetti che si collegano a quanto detto nel paragrafo precedente. Infatti la matrice di covarianza familiare dà informazioni sia sul rischio individuale, quando viene a coincidere con la varianza, sia sulla relazione lineare che lega diversi membri appartenenti ad una stessa famiglia; la correlazione tra i guadagni dei familiari mette in evidenza il grado di pooling dei rischi all'interno della famiglia. L'analisi illustrata nei due sotto-paragrafi seguenti dà un'ulteriore informazione su come gli shock temporali incidano sulla struttura di covarianza in esame.

5.1 *Presentazione dell'analisi empirica*

Negli Stati Uniti nei primi anni Ottanta è significativamente cresciuta la disuguaglianza salariale e nei guadagni, individuali e familiari. Allo scopo di studiare questo, è stato realizzato un modello intertemporale sull'offerta di lavoro familiare che ha dato alcuni risultati interessanti circa la struttura di covarianza familiare.

Le disuguaglianze salariali in discussione sono state spesso attribuite alla crescita del commercio internazionale e allo sviluppo della tecnologia, con una sorprendente carenza di ricerca sui meccanismi che spieghino la parallela se non più rapida crescita della disuguaglianza sul reddito familiare rispetto a quello individuale. Il collegamento tra la disuguaglianza sul benessere familiare e la dispersione delle aliquote salariali ha tre aspetti di cui si deve tener conto:

- 1) l'attribuzione dei motivi della crescente disuguaglianza salariale a shock di tipo transitorio piuttosto che permanente;
- 2) il grado di correlazione più o meno forte che caratterizza i cambiamenti salariali dei lavoratori di una stessa famiglia;
- 3) come e se le famiglie cambiano i loro orari di lavoro in risposta ai mutamenti delle aliquote salariali che investono i singoli membri.

Il periodo al quale si riferisce l'analisi è quello che va dal 1979 al 1985; questo per via dei forti shock che lo attraversano e per esigenze del modello di analisi stesso che richiede un numero di anni limitato. L'indagine utilizza dati longitudinali del Panel Study of Income Dynamics, e descrive la matrice di varianza e covarianza dei guadagni di mariti e mogli. L'analisi è preceduta da una prima indagine sui trends sulla disuguaglianza incrociata e sezionale dei guadagni individuali e familiari nella prima metà degli anni '80; si prendono in considerazione diversi campioni, dalle coppie ove entrambi i partner lavorano, a quelle in cui a lavorare è il solo marito, con anche diversi tipi di dati utilizzati.

A questo punto ci si sofferma su di un'analisi che ha per oggetto le matrici di covarianza e correlazione dei salari e dei guadagni, incrociando anni (sono sette) e partner maschile e femminile.

Si studiano le autocovarianze dei salari individuali, e si focalizza l'attenzione su quale frazione di varianza dei salari è attribuibile a shock permanenti piuttosto che transitori. Poi si guarda alle covarianze incrociate dei salari di mariti e mogli, e al grado al quale queste sono attribuibili a fattori permanenti o transitori. Infine si esamina la struttura di covarianza dei guadagni dei coniugi e ci si chiede se è ragionevole assumere che gli orari di lavoro siano indipendenti dai loro salari.

5.2 *I risultati salienti*

I risultati principali di questa analisi mettono in luce come circa la metà delle variazioni dei salari siano di tipo permanente; inoltre la correlazione dei salari o guadagni delle mogli con i risultati dei mariti nei sette anni è positiva; infine la correlazione incrociata tra i salari degli "sposi" è più grande di quella relativa ai guadagni. Ne consegue che le ore non sono indipendenti dai salari, e le differenze nelle correlazioni incrociate nei salari e nei guadagni sono interamente attribuibili a fattori permanenti. E' inoltre importante sottolineare che non ci sono effetti di sostituzione incrociati intertemporali che legano le variazioni dei salari di un coniuge ai cambiamenti delle ore lavorative del partner.

Viene presentato un modello intertemporale sull'offerta di lavoro che è coerente con queste conclusioni; lo si combina con un modello con componenti negli errori per descrivere l'evoluzione dei salari orari di mariti e mogli. Il modello stima medie, varianze e covarianze dei salari e dei guadagni di mariti e mogli; il campione è quello delle coppie in cui entrambi i partner lavorano in modo continuo, sempre relativo agli anni 1979-1985. I risultati del modello sui salari denotano che la crescita della disuguaglianza incrociata e sezionale sui salari orari dei mariti è attribuibile in modo robusto proporzionalmente a fattori permanenti e transitori; la quota di shock permanenti è circa un mezzo nei primi anni Ottanta. Invece per le mogli la crescita di disuguaglianza è largamente

attribuita a fattori transitori, con la quota di shocks permanenti in calo dal 65 al 55%.

I salari degli sposi sono tra loro correlati positivamente, cosa che fa pensare che normalmente i partner siano in qualche modo legati dal punto di vista lavorativo (ad esempio potrebbero lavorare nello stesso settore industriale). Nelle coppie c'è anche correlazione positiva tra i guadagni di marito e moglie, anche se minore rispetto a quella nei salari.

Le stime dei parametri del modello sull'offerta di lavoro mettono in evidenza come l'elasticità dell'offerta di lavoro intertemporale dei mariti rispetto ai propri salari è vicina a zero, mentre per le mogli è pari a circa 0.4. La conclusione di questi risultati è che le risposte che dà il modello sull'offerta di lavoro dicono molto poco sulla crescita della disuguaglianza rispetto ai guadagni annuali per i mariti, ma spiegano circa il 50% della più modesta crescita di disuguaglianza per le donne sposate, e il 20% dell'incremento della disuguaglianza sui guadagni familiari nei primi anni Ottanta.

CAPITOLO 2 L'influenza della famiglia sulle scelte individuali circa il tipo di contratto di lavoro

6 *Offerta di lavoro: quali influenze sulle scelte individuali*

L'individuo che lavora può essere autonomo piuttosto che dipendente, ma è anche possibile distinguere tra più di due categorie di lavoro con associati determinati livelli di rischio; la principale distinzione è comunque quella tra autonomo e dipendente .

Nella scelta sul lavoro l'individuo è influenzato da tutta una serie di variabili che lo spingono in una direzione piuttosto che nell'altra, dalle disponibilità economiche al titolo di studio. La mia analisi si sofferma soprattutto sull'influenza che la famiglia e in particolare la situazione lavorativa del partner hanno sulla scelta di un data persona. Mi soffermerò nei paragrafi 2 e soprattutto 3 su tale tipo di influenza, mentre in questo paragrafo 1 analizzo come alcune variabili incidano sulla scelta sul tipo di lavoro di un individuo, senza tener conto del contesto familiare in cui è eventualmente calato.

Normalmente le persone sono maggiormente attratte dal lavoro autonomo piuttosto che da quello dipendente; nei sotto-paragrafi 6.1 e 6.2 spiegherò l'appel e le soddisfazioni che dà la prima soluzione e illustrerò quando è più facile diventare lavoratori autonomi o dipendenti. Nel sotto-paragrafo 6.3 mi soffermerò invece su alcuni aspetti teorici rispetto alla scelta occupazionale e su come questa rifletta le stesse caratteristiche individuali del lavoratore.

6.1 *Il lavoro autonomo: appetibilità, difficoltà e soddisfazione*

Al di là della scelta finale più o meno obbligata, ogni persona può soggettivamente preferire il lavoro autonomo piuttosto che quello dipendente. Va detto che le indagini a tal riguardo sono unanimi nel sottolineare come la soluzione preferita sia la prima. Ad esempio l'International Social Survey Programme del 1989 ha posto ad un campione casuale di individui di undici diversi paesi la domanda "Supponi di essere in procinto di lavorare e di poter scegliere tra diversi tipi di lavoro; tra lavoro autonomo, lavoro dipendente e altro cosa sceglieresti?"; tra gli altri, il 63% degli americani (su 1453 intervistati), il 48% dei bretoni (su 1297 intervistati) e il 49% dei tedeschi (su 1575 intervistati) ha risposto "lavoro autonomo"; nella realtà invece i lavoratori autonomi sono in percentuale inferiore.

Detto questo, le preferenze di un individuo verso un determinato tipo di lavoro dipendono dalle personali attitudini verso gli aspetti caratterizzanti delle due categorie. In particolare un'indagine inglese su dati longitudinali ha sottoposto 9mila adulti, nell'autunno 1991, ad un'intervista. Questa ha rilevato come i lavoratori autonomi enfatizzino l'uso di iniziativa e la soddisfazione propri di questo tipo di lavoro, e come i lavoratori dipendenti mettano invece l'accento sulla sicurezza e sui compensi perlopiù sicuri che dà loro il lavoro.

Un'indagine completa su appetibilità, difficoltà ed eventuale soddisfazione che dà il lavoro autonomo è stata fatta negli anni Ottanta sempre in Gran Bretagna; si tratta della British Social Attitudes Surveys.

Una prima domanda posta a tutti i lavoratori dipendenti del campione, vale a dire a 5947 persone, è: "Avete preso in seria considerazione l'eventualità di diventare lavoratori autonomi?"; negli anni 1983-84, 1986-87 e 1989 mediamente hanno risposto "Molto seriamente" o "Abbastanza seriamente" circa il 16.8% degli intervistati. A una parte di quest'ultimi è stato chiesto il motivo per cui non lo sono diventati, e circa la metà ha risposto rifacendosi all' "assenza di capitale e disponibilità economica", circa il 10% ha motivato la scelta con l'eccessivo rischio caratterizzante il lavoro autonomo, e i rimanenti hanno dato altre risposte,

ad esempio mettendo in evidenza il detrattore del “contesto economico sfavorevole”.

La National Survey of the Self-Employed dà invece voce a un campione di circa 12mila adulti intervistati in Gran Bretagna nel 1987. Agli individui diventati lavoratori autonomi, in particolare 243 imprenditori, è stato chiesto di indicare la fonte finanziaria del loro business. Il 42% si è rifatto a propri risparmi, il 15% ha usato finanziamenti familiari o di amici, e il 17% ha chiesto un prestito in banca. A conferma di questi risultati, la stessa National Survey of the Self-Employed ha intervistato 139 persone in procinto di diventare imprenditori e ha chiesto loro qual è la maggiore preoccupazione che hanno nell'intraprendere questa nuova strada; la risposta più frequente è proprio quella relativa alla difficoltà di trovare gli adeguati finanziamenti.

I lavoratori autonomi sono considerati essere persone più felici di quelli dipendenti, e questo spiega l'appetibilità del lavoro autonomo. Nel 1981 la National Child Development Study ha realizzato un test per vedere se quanto appena affermato è vero. Del resto, posto che i soldi facciano la felicità, affermazione nel cui merito non entro, sarebbe difficile fare dei ragionamenti in tal senso, dato che i guadagni denunciati dai lavoratori autonomi sono di fatto spesso lontani da quelli realizzati davvero.

L'indagine su di un campione di 9176 lavoratori 23enni ha innanzitutto messo in luce come per la maggior parte degli intervistati la soddisfazione nella vita sia altamente correlata alla soddisfazione nel lavoro. Inoltre l'indagine ha messo in evidenza come i più soddisfatti siano proprio i lavoratori autonomi. In particolare alla domanda “Quanto sei soddisfatto/insoddisfatto della tua condizione lavorativa le risposte (tra cinque proposte) riportate nella seguente *tabella 1* sono sufficientemente esaustive.

tabella 1

RISPOSTE	LAV. AUTONOMI	LAV. DIPENDENTI
“molto insoddisfatto”	1.7	2.8
“insoddisfatto”	2.9	9.6
“né l’uno né l’altro”	6.7	8.2
“soddisfatto”	42.4	50.2
“molto soddisfatto”	46.2	29.1

Va comunque detto che questi risultati sono indicativi, ma comunque opinabili; infatti può essere che i lavoratori autonomi, e in particolare gli imprenditori, rispondano più facilmente degli altri “molto soddisfatto” solo perché appartenenti ad una categoria intrinsecamente ottimistica.

6.2 Quando è più facile diventare lavoratori autonomi

Faccio ora ancora riferimento all’indagine inglese National Child Development Study, che utilizza dati longitudinali.

L’analisi mette in evidenza la correlazione positiva tra lavoro autonomo (che è la variabile dipendente) e l’aver ricevuto un’eredità o una donazione; ovviamente un’equazione che tenga conto di quest’ultima variabile acquista significatività se si prende in considerazione l’eventuale morte di un/entrambi i genitori. Questo denota una correlazione tra assetto familiare e lavoro autonomo. A tal riguardo la stessa indagine riporta come, all’aumentare dell’eredità ricevuta a livello familiare, aumenti anche la probabilità di essere lavoratori autonomi, fermo restando che il numero di lavoratori dipendenti rimane comunque preponderante. Ad esempio non hanno ricevuto alcuna eredità l’88.7% dei lavoratori dipendenti e l’83.8% dei lavoratori autonomi. Questi risultati non si discostano sensibilmente se rivolti ad un campione di lavoratori 23enni piuttosto che a dei 33enni, anche se naturalmente per quest’ultimi la probabilità di essere lavoratori autonomi aumenta, soprattutto se hanno ricevuto un’eredità/donazione.

Altri risultati che emergono sono:

- 1) in proporzione le persone occupate nel lavoro autonomo, rispetto al totale, aumentano con l'età;
- 2) gli individui di sesso femminile hanno minore probabilità di diventare lavoratori autonomi rispetto a quelli di sesso maschile;
- 3) vi è influenza diretta dell'occupazione lavorativa dei genitori su quella dei figli;
- 4) oltre all'ammontare di eredità/donazione, anche il capitale complessivamente disponibile ad una persona è positivamente correlato alla probabilità di diventare lavoratori autonomi;
- 5) entrano in gioco anche le variabili su istruzione, esperienza, e contesto economico a livello regionale.

Vediamo ora un esempio che mette assieme un po' tutti questi aspetti e che si rifà ai risultati empirici dell'indagine inglese National Child Development Study . Un maschio che viva nel sud-est dell'Inghilterra, con un tirocinio alle spalle e il cui padre è un manager di un'azienda con meno di 25 lavoratori, ha una probabilità pari al 16% di diventare lavoratore autonomo, se privo di eredità e/o donazioni; la probabilità sale al 37% se ha ricevuto un'eredità di 5000 dollari. Se il soggetto in questione è invece di sesso femminile, le probabilità di diventare lavoratore autonomo scendono rispettivamente al 7 e al 21%.

L'indagine arriva anche alla conclusione che le caratteristiche psicologiche di un individuo, in particolare quelle legate all'infanzia, hanno scarsa influenza sulla scelta del tipo di lavoro che una persona prende. Infatti i punteggi sui test psicologici sottoposti ai bambini risultano scarsamente correlati alla probabilità di diventare lavoratori autonomi piuttosto che dipendenti.

6.3 Alcuni aspetti teorici rispetto alla scelta occupazionale

Sia la teoria neoclassica che quella non-competitiva provvedono a spiegare l'allocazione di lavoratori eterogenei rispetto a lavori a loro volta eterogenei. Da un lato, i lavoratori hanno diverse attitudini e preferenze, dall'altro, i datori di

lavoro offrono occupazioni che differiscono rispetto all'esperienza richiesta, al grado di sicurezza intrinseco, e al tipo di ambiente di lavoro proposto.

L'assunto che sta alla base della teoria neoclassica è che i lavoratori cercano di massimizzare l'utilità e non semplicemente il loro ritorno monetario; questo significa che essi sono interessati sia agli aspetti pecuniari sia a quelli non legati al guadagno in senso stretto. I lavori sono differenti perché alcuni richiedono istruzione e preparazione e altri no, alcuni sono "tranquilli" e altri sono pericolosi, alcuni richiedono iniziativa personale e responsabilità al contrario di altri che sono invece monotoni e noiosi, alcuni sono sicuri rispetto ad un'eventuale interruzione del rapporto di lavoro mentre altri sono rischiosi e fortemente dipendenti dalle condizioni del mercato. Tutte queste caratteristiche dei lavori influenzano l'utilità dei lavoratori, e quindi la loro scelta.

Una ben conosciuta interpretazione dell'allocazione dei lavori e dei lavoratori, nella teoria neoclassica, è quella relativa al modello edonistico di Sherwin Rosen che spiega la co-esistenza, in equilibrio, di differenti coppie di "tipo di lavoratore – genere di lavoro" rispetto a diversi livelli salariali. Se la prima implicazione del modello, cioè che a lavori poco piacevoli corrispondano salari più alti, è stata pesantemente smontata sia dal punto di vista teorico che da quello empirico, la seconda è decisamente più incontrovertibile. Quest'ultima sostiene che i lavoratori con forti preferenze per certe caratteristiche di un determinato tipo di lavoro sceglieranno, in assenza di controindicazioni o vincoli, i lavori con queste caratteristiche.

La questione cruciale è se le grandi categorie di tipi di lavoro, quali sono lavoro autonomo e dipendente (ma anche lavoro dipendente nel pubblico e lavoro dipendente nel privato), possano essere trattate come raggruppamenti di lavori omogenei riguardo alla condizione del lavoratore, alla sicurezza intrinseca, all'istruzione e all'esperienza richieste, all'organizzazione dell'attività, e al tipo di lavoratori che ne sono attratti, con le loro caratteristiche e preferenze individuali. Non ha tanto rilievo mettere l'una contro l'altra e confrontare le diverse teorie economiche, quanto studiare in modo empirico le caratteristiche e le motivazioni che mettono insieme gruppi di persone che lavorano in stesse categorie di lavoro, vale a dire autonomo e dipendente. Va detto che la

classificazione del lavoro in autonomo, dipendente, e magari in dipendente nel pubblico e in dipendente nel privato è assimilabile ad una classificazione in settori produttivi, sulla base della natura del prodotto o del processo produttivo. Infatti vi sono una serie di aspetti che caratterizzano i tipi di lavoro come settori. In particolare, in Italia, il lavoro dipendente nel pubblico è qualificabile come lavoro sicuro, che richiede un certo livello di istruzione (anche molto elevato come nel caso della sanità e dell'insegnamento), con ore di lavoro strutturate in modo sotto-settoriale (offrendo molto tempo libero a disposizione e la possibilità di praticare attività collaterali), e che solitamente dà una certa visibilità sociale. Il lavoro dipendente nel privato rientra invece in un settore ove vigono regole di avanzamento di carriera che sono maggiormente meritocratiche e a discrezione dei propri superiori di quanto sia nel settore pubblico; si lascia così più spazio alla responsabilità personale, e vi sono più flessibilità e più mobilità. Infine il lavoro autonomo è peculiare per via dell'elevato grado di indipendenza che offre nell'organizzazione dell'attività lavorativa, e per l'alto livello di autonomia che lo caratterizza nel momento delle decisioni da prendere, anche se tutto questo è accompagnato da un certo rischio e da una qualche variabilità del reddito.

E' dunque ragionevole assumere che queste differenti caratteristiche dei tipi di lavoro e delle loro condizioni ambientali attirino individui con determinati gusti e preferenze, che è facile siano a loro volta correlate alle caratteristiche personali.

7 Influenza della famiglia nella scelta del tipo di lavoro

La scelta del settore, vale a dire se l'individuo è lavoratore autonomo, dipendente pubblico o dipendente privato, è, come appena visto, influenzata da tutta una serie di caratteristiche personali, dall'età al livello di istruzione.

La scelta dipende dalle relative opportunità di guadagno, dalle preferenze e dalle attitudini individuali rispetto alle caratteristiche dei lavori, ma anche dal rapporto con il rischio. Assumono così importanza il grado di stabilità del lavoro e la variabilità dei guadagni percepiti, oltre che il desiderio di sicurezza.

L'esposizione al rischio, oltre che dalla predisposizione personale, è influenzata anche da come si pone l'eventuale famiglia del soggetto in questione rispetto al rischio stesso. Ma oltre che nell'atteggiamento, la famiglia influenza l'esposizione al rischio in quanto porta con sé un certo livello di ricchezza da distribuire. Le variabili che assumono importanza in tal senso sono la presenza di altri redditi in famiglia, che porta ad una diversificazione del rischio, la presenza di reddito non da lavoro, e il ruolo del capofamiglia.

L'analisi sul ruolo della famiglia che vado a realizzare si concentra soprattutto sulla presenza di "altri redditi da lavoro", in particolare quello del capofamiglia e quello del partner. Infatti entrambi i soggetti possono lavorare, oppure le entrate possono venire dal solo capofamiglia. Quando vi sono più entrate vi è, *come visto da un punto di vista teorico nel capitolo 1 paragrafo 4.4*, diversificazione, e quindi meno rischio per ciascun membro, anche se in misura minore se i rischi che si mettono assieme sono correlati positivamente tra loro.

Ora analizzo quanto e in che dimensioni questo sia vero da un punto di vista empirico, soprattutto guardando a se i rischi, portati da membri di una stessa famiglia, provengano da tipi di lavoro con grado di rischio diverso o meno (alto se autonomo, basso se dipendente), per valutare se c'è risk sharing per via di una diversificazione. L'analisi empirica, che io realizzo sul caso Italiano, prende spunto da una simile fatta in Inghilterra, su cui mi soffermo nel paragrafo 8. Prima però illustro i risultati di alcuni test che verificano l'ipotesi che vi sia risk sharing intrafamiliare e/o interfamiliare.

7.1 Test sull'ipotesi che vi sia completo risk sharing intrafamiliare e interfamiliare

Il Panel Study of Income Dynamics (PSID) è la fonte dei dati utilizzati da Fumio Altonji, Joseph Hayashi e Laurence Kotlikoff per testare la presenza di completo risk sharing (completa distribuzione del rischio) all'interno e tra le famiglie americane. Studiare le dimensioni del risk sharing può essere importante per stimare la performance del mercato dell'assicurazione privata così come l'efficacia delle politiche governative in ambito sociale sull'assicurazione; è

anche importante per modellare in modo adeguato l'economia; ad avvalorare ciò vi è il fatto che molte analisi macroeconomiche diano per scontata la presenza di completo risk sharing.

I dati sulle famiglie costituiscono il terreno ideale per testare la presenza di risk sharing completo. Guardando alle preferenze e al tempo libero, i cambiamenti nel tempo del consumo delle famiglie che sono caratterizzate da completo risk sharing dipendono dai cambiamenti nelle loro risorse collettive, ma non da cambiamenti nella distribuzione delle risorse tra le famiglie stesse. Stimare la presenza di completo risk sharing attraverso un set di famiglie richiede di esaminare la correlazione incrociata e sezionale tra i cambiamenti nel consumo familiare e i cambiamenti nelle risorse a disposizione delle famiglie.

Il PSID permette di testare la presenza di completo risk sharing tra nuclei familiari legati da parentela e non. Questo perché il PSID è un campione casuale di famiglie che consiste di quelle originariamente incluse nel panel (relative all'anno 1968), ma anche di quelle che si sono formate da quest'ultime dopo un certo numero di anni rispetto a quelle prese inizialmente in considerazione. Va detto che testare il risk sharing familiare con il PSID porta a due diversi squilibri nei dati a disposizione. Innanzitutto vi è una diversa composizione delle famiglie; inoltre la quantità di anni durante i quali sono osservati i nuclei familiari non è sempre la stessa.

Vi è una crescente letteratura sulla presenza di risk sharing completo all'interno e tra le famiglie. I primi test appartengono a Leme, che ha esaminato i tassi di crescita del consumo aggregato in una sezione incrociata di paesi. Un più recente test, basato sul Panel Study of Income Dynamics, è quello di Altug e Miller. I due, che permettono al consumo di tempo libero di essere variabile endogena, falliscono nel tentativo di rigettare l'ipotesi sulla presenza di completo risk sharing.

Hall invece ipotizza che le famiglie non abbiano accesso all'assicurazione sul reddito, e che l'unico modo che hanno per limitare l'incertezza dovuta alle fluttuazioni sul reddito sia quello di auto-assicurarsi, attraverso il ricorso al risparmio. Le famiglie che si auto-assicurano cambiano il loro livello di consumo solo quando ricevono shock impreveduti sulle risorse familiari, cosicché la

correlazione su serie temporali tra cambiamenti sul consumo e precedenti mutamenti sul salario sia nulla.

Altonji, Hayashi e Kotlikoff utilizzano anch'essi il Panel Study of Income Dynamics, e realizzano un test dinamico sull'altruismo che fa anche da test sulla presenza di risk sharing familiare. Questo test regredisce i cambiamenti familiari nel consumo di sussistenza sui cambiamenti di reddito come effetto fisso per la famiglia. Se i membri di una famiglia ripartiscono il reddito in modo completo, il cambiamento nel consumo di un dato membro, associato ai mutamenti sul reddito, sarà pienamente catturato dall'effetto fisso familiare. Perciò, visto che l'effetto dei cambiamenti sul reddito sui singoli individui è altamente significativo, come trovano gli stessi Altonji, Hayashi e Kotlikoff, ciò può essere interpretato come un rifiuto dell'ipotesi nulla di risk sharing familiare. Tuttavia questo approccio ha due lacune. Innanzitutto si assume che i beni di sussistenza siano separabili dal consumo di tempo libero; inoltre si ignora la possibile endogeneità del tempo libero stesso.

Utilizzando le fluttuazioni relative al periodo 1985-87 del PSID, si può supporre che non vi sia separabilità tra consumo di sussistenza e di tempo libero. Si può spiegare l'implicazione del completo risk sharing secondo cui i cambiamenti nel consumo nel corso di un anno sono incorrelati al tasso salariale. A questo risultato si arriva esaminando la correlazione di cambiamenti sul lungo termine così come di cambiamenti nel breve periodo del consumo e del tasso salariale. Ciò è essenziale per discriminare tra risk sharing tra le famiglie, e autoassicurazione familiare.

La metodologia seguita può essere sintetizzata come segue. Il PSID associa ad ogni famiglia un vettore che specifica il consumo, per ogni anno, disponibile alla famiglia stessa. Le ipotesi sulla presenza di risk sharing tra le famiglie e di autoassicurazione familiare implicano che l'utilità marginale abbia una particolare struttura fattoriale; per ogni struttura fattoriale può essere utilizzata una trasformazione lineare sul vettore del consumo familiare. Se l'ipotesi nulla (cioè "vi è risk sharing") è vera, la trasformazione lineare sul consumo sarà incorrelata al tasso salariale, che è un determinante della funzione di utilità marginale. Se invece è falsa, la trasformazione lineare sul consumo riguarderà ancora le utilità

marginali, e così il consumo sarà correlato al tasso salariale. Poiché la trasformazione lineare sul consumo può essere adattata ad ogni tipo di famiglia, questo test è applicabile a tutte le famiglie nel campione; non c'è bisogno di restringere il test alle famiglie osservate in uno stesso periodo temporale e con uguale composizione numerica.

I risultati dipendono da due differenti PSID panel. Uno è il panel delle famiglie originarie usato da Altug e Miller. Esso contiene tutte le famiglie originarie (costituite fin dall'anno 1968) presenti nell'indagine che va dal 1968 al 1981. Usando questo panel, si cerca di replicare il test statistico di Altug e Miller per il risk sharing familiare completo. Questo è importante perché lo studio di Altug e Miller è rilevante per il fallimento del tentativo di rigettare l'ipotesi di completo risk sharing, con p-value all'80%. Il p-value relativo alle fluttuazioni sul reddito avvenute negli anni 1985-87 è pari al 12%, molto più basso rispetto al precedente ma ancora sostanzialmente insufficiente per rifiutare l'ipotesi di risk sharing completo. Aggregando le condizioni sulla correlazione nulla rispetto al tempo, il p-value cade al 0%. Rapportando il cambiamento del consumo all'età il p-value torna a salire al 5%. Se si aggiungono i futuri e correnti cambiamenti salariali alla lista degli strumenti, il p-value cade nuovamente al 0%. Così si fallisce nel tentativo di rigettare l'ipotesi sul ricorso all'auto-assicurazione familiare al 5%, ma viene invece fortemente rigettata l'ipotesi sulla presenza di completo risk sharing tra le famiglie.

Il secondo campione è ricavato dal Panel Study of Income Dynamics relativo agli anni 1985-87, e include le famiglie formatesi da quelle originariamente incluse nel campione PSID. Si utilizza questo panel per testare la presenza di risk sharing interfamiliare e di auto-assicurazione. La presenza di risk sharing tra le famiglie è ancora una volta rifiutata; è rigettata anche l'ipotesi sulla presenza di auto-assicurazione familiare (risk sharing intrafamiliare), visto come d'altra parte la covarianza sui cambiamenti nel consumo all'interno delle famiglie sia molto piccola.

8 *L'analisi empirica inglese Employment Contract Matching*

L'analisi inglese che ora vado ad illustrare ha per oggetto lo studio sulla presenza di risk sharing intra-familiare e intra-coppia, e su come questi influenzino la scelta sui tre principali tipi di contratto di lavoro; infatti gli autori di *Employment Contract Matching*, Sarah Brown, Lisa Farrel e John G. Sessions, distinguono ancora tra lavoro autonomo e lavoro dipendente, aggiungendo però anche una modalità "intermedia" rispetto alle due più classiche, che prevede l'acquisto da parte di un datore di lavoro direttamente di una performance realizzata. Ciascuna delle tre categorie è caratterizzata da un certo livello di rischio sul reddito. Si attribuisce al lavoro autonomo il maggior grado di rischio, mentre il lavoro dipendente viene visto come attività più sicura.

Due sono i risultati che l'analisi può attendersi; infatti può accadere che le coppie costituite da capofamiglia e partner, entrambi lavoratori, vedano i due scegliere prevalentemente tipi di lavoro che rientrino nella stessa categoria (autonomo, dipendente, modalità intermedia); alternativamente può succedere che la soluzione che va per la maggiore veda i due partner impegnati in tipi di lavoro diversi, realizzando una diversificazione del portafoglio dei contratti lavorativi, e quindi favorendo un pooling dei rischi intra-familiare.

Si tratta di un'analisi empirica inglese che utilizza dati tratti dal *British Family Expenditure Surveys*, con riferimento agli anni che vanno dal 1996 al 2000.

L'indagine, che ora vado ad illustrare più nel dettaglio, assume un'importanza particolare per la realizzazione di questa tesi, visto che è proprio dalla lettura "*Employment Contract Matching: an analysis of dual earner couples and working households*" che trae spunto la mia analisi empirica, negli obiettivi e in parte anche nelle modalità di realizzazione.

8.1 La struttura statistica dell'analisi inglese

La variabile dipendente è categorica, cioè assume valore 1 se l'individuo è lavoratore dipendente, 2 se l'individuo si trova nella situazione "intermedia", 3 se l'individuo è lavoratore autonomo. Ci si aspetta che le caratteristiche individuali e familiari assumano un ruolo importante nel determinare le variazioni nelle probabilità individuali di appartenere ad una certa categoria occupazionale piuttosto che ad un'altra. L'analisi adottata utilizza un modello probit ordinale ed un modello probit ordinale ad effetti casuali (*per una spiegazione dettagliata di questi modelli rimando al capitolo 4*).

L'ordinamento adottato, realizzato supponendo che al lavoro autonomo sia associato il maggior grado di rischio e al lavoro dipendente il minore, trova supporto nei dati grezzi. Infatti la deviazione standard dei guadagni attorno alla media è 204.34 per il lavoro dipendente, 326.70 per la forma "intermedia", e 481.70 per il lavoro autonomo, con il campione formato dalle coppie con entrambi i partner che lavorano. I risultati confermano l'ordinamento scelto anche lì dove il campione sia formato da tutti i lavoratori appartenenti a famiglie con capofamiglia e partner entrambi occupati.

Il modello probit ordinale si basa su di una struttura di regressione latente con

$$y_i^* = \beta' x_i + \varepsilon_i$$

Sebbene y_i^* sia non osservabile, si osserva y_i tale che

$$y_i = 1 \text{ se } y_i^* \leq \mu_1$$

$$y_i = 2 \text{ se } \mu_1 < y_i^* \leq \mu_2$$

$$y_i = 3 \text{ se } \mu_2 < y_i^*$$

con μ e β parametri ignoti da stimare. Dato che ε_i è normalmente distribuito attraverso osservazioni con media zero e varianza pari ad uno, si ottengono le seguenti probabilità

$$P(y_i = 1) = \Phi(\mu_1 - \beta' x_i)$$

$$P(y_i = 2) = \Phi(\mu_2 - \beta' x_i) - \Phi(\mu_1 - \beta' x_i)$$

$$P(y_i = 3) = 1 - P(y_i = 1) - P(y_i = 2)$$

con $\Phi(\cdot)$ che denota una distribuzione cumulativa normale standard.

Il vettore x_i indica il set di variabili esplicative; contiene diverse variabili che rappresentano le caratteristiche individuali e familiari utili a spiegare i cambiamenti individuali osservati rispetto ai tipi di lavoro del capofamiglia e del partner.

Le caratteristiche individuali osservate sono il sesso della persona, l'età, il livello d'istruzione; vengono inoltre riportate le caratteristiche specifiche del lavoro. Le caratteristiche familiari sono quelle sul livello di reddito, sul numero di bambini in età prescolare e scolare, sull'abitazione, sull'area geografica. Si guarda anche all'età e allo status economico degli altri individui maggiorenni che vivono in famiglia, cioè se siano disoccupati, malati, in pensione, non occupati o studenti; si include anche un set di variabili dummy che indaga il tipo di lavoro delle altre persone che eventualmente vivono in famiglia. Infine, per accertare se la coppia capofamiglia-partner veda entrambi occupati in uno stesso tipo di lavoro, o se piuttosto vi sia pooling dei rischi, si utilizza l'informazione sul tipo di lavoro del partner.

Oltre al modello probit ordinale, l'analisi statistica utilizza anche un modello probit ordinale ad effetti casuali, che consente d'indagare il grado di correlazione sui tipi di lavoro (dipendente, forma "intermedia" e autonomo) nelle coppie capofamiglia-partner. Il campione del modello è costituito dalle coppie in cui entrambi i partner lavorano. Il modello è specificato come segue:

$$y_{ci}^* = \beta' x_{ic} + v_{ic} \quad , \quad v_{ic} = \alpha_c + \eta_{ic}$$

con y_{ci}^* propensione non osservabile al rischio dell'individuo i della coppia c ; y_{ic} dà il tipo di lavoro osservato dell'individuo, con $i = 1,2$ che indica ciascun membro della coppia; x_{ic} è un vettore di caratteristiche esogene che ci si aspetta influenzino la variabile non osservabile y_{ci}^* ; β è il vettore associato di coefficienti; α_c lo specifico effetto "di coppia" (cioè con i partner che tendono a scegliere lo stesso tipo di lavoro) non osservabile. Si considera una specificazione ad effetti casuali, con $\eta_{ic} \sim \text{IN}(0, \sigma_\eta^2)$, e si assume che, condizionatamente a x_{ic} , gli α_c siano $\text{IN}(0, \sigma_\alpha^2)$, e indipendenti dagli η_{ic} e dagli

x_{ic} . Questo implica che la correlazione tra i termini d'errore degli individui sposati/conviventi sia una costante data da

$$\rho = \text{corr}(v_{il}, v_{ik}) = \sigma_{\alpha}^2 / (\sigma_{\alpha}^2 + \sigma_{\eta}^2), \text{ con } l \neq k$$

ρ rappresenta la proporzione di varianza totale influenzata dalla componente effetto di coppia (cioè con i partner che tendono a scegliere lo stesso tipo di lavoro).

Infine si indaga la possibilità che vi sia pooling dei rischi, o meno, rispetto a tutti i membri lavoratori della famiglia, invece che semplicemente rispetto alla coppia. Il modello rimane lo stesso, anche se in questo caso l'indicatore c è sostituito dall'indicatore h relativo a tutta la famiglia e non più solo alla coppia; si assume che il minimo numero di componenti una famiglia sia uno, e il massimo sia sette. Il modello probit ordinario ad effetti casuali, visto nelle due varianti, differisce rispetto a quello "semplice", relativamente al vettore delle variabili esplicative, solo in quanto esclude la variabile sul tipo di lavoro del partner, visto che i dati sono riorganizzati in un panel per cui ogni membro della coppia (famiglia) rappresenta un'osservazione ripetuta per la coppia (famiglia). Inoltre il set di variabili dummy sul il tipo di lavoro degli altri membri lavoratori della famiglia è omesso quando, nel modello ad effetti casuali, i tipi di lavoro degli altri membri sono già inclusi come osservazioni direttamente nella variabile dipendente.

8.2 I dati e i campioni utilizzati

I dati sono presi dal Family Expenditure Survey (FES) per la Gran Bretagna; si tratta di una rappresentativa indagine nazionale condotta con cadenza annuale dal 1957 in poi. Ogni anno 10mila famiglie sono selezionate per prendere parte al FES, con un tasso medio di risposta pari a circa il 70%. L'obiettivo dell'indagine è quello di fornire una credibile fonte di informazione sulla spesa e sul consumo familiari, sul reddito, e su di altri aspetti finanziari della famiglia. Vengono realizzate interviste faccia a faccia a persone di età non inferiore ai 16 anni appartenenti alle famiglie prescelte. Il FES è particolarmente appropriato rispetto

agli obiettivi dell'analisi in esame, perché contiene informazioni dettagliate sul tipo di lavoro, e dà molto spazio alle caratteristiche individuali e familiari.

Si utilizzano i dati che vanno dal 1996 al 2000, e si includono gli individui che lavorano di età compresa fra i 16 e i 65 anni. Dai dati si generano due diversi campioni. Il primo è formato dalle coppie, sposate o conviventi, che lavorano, con osservazioni su entrambi i partner; il campione è di 92276 coppie “che lavorano”, che danno un totale di 18552 osservazioni. Il secondo, che porta l'analisi ad esplorare le correlazioni tra tutti i lavoratori membri di una famiglia, è di 31862 lavoratori che vivono in 19604 famiglie.

La distribuzione dei tipi di lavoro attraverso i due campioni, rispetto alle diverse caratteristiche individuali e familiari, è rappresentata dalla *tabella 2*; quest'ultima riporta i risultati relativi al campione che include i soli membri delle coppie capofamiglia-partner lavoratori (*campione 1*), e quelli relativi al campione formato da tutti i membri della famiglia che siano lavoratori (*campione 2*). Nel secondo caso è inclusa la variabile sullo status coniugale, mentre non sono più necessarie le variabili dummy sugli altri membri lavoratori appartenenti alla famiglia.

Tabella 2

Caratteristiche	Campione 1			Campione 2		
	Lav. dip.	Lav. int.	Lav. aut.	Lav. dip.	Lav. int.	Lav. aut.
SESSO						
Maschile	58.78	25.05	16.17	60.19	24.79	15.02
Femminile	75.41	17.75	6.84	76.55	17.32	6.13
STATUS MATRIMONIALE						
sposato				66.48	21.40	12.12
separato/vedovo/divorziato				71.67	18.10	10.23
single				72.48	22.06	5.46
ETA'						
16-19	92.68	4.88	2.44	84.47	13.79	1.74
29-29	66.48	27.59	5.93	69.54	24.96	5.50
30-39	66.81	22.52	10.66	66.56	23.05	10.39
40-49	68.24	19.46	12.30	67.51	19.19	12.56
oltre 50	66.05	17.97	15.98	66.83	17.70	15.47
ISTRUZIONE						
livello 1	67.47	18.42	14.12	66.35	18.24	13.41
livello 2	66.77	21.99	11.25	68.07	21.50	10.43
livello 3	66.71	23.73	9.56	67.69	23.50	8.81
livello 4	67.63	21.08	11.29	67.65	21.44	10.90
OCCUPAZIONE						
professionale	54.42	25.72	19.86	56.81	25.34	17.85
manageriale	67.92	22.79	9.29	67.23	23.29	9.47
specializzata	65.83	22.35	11.83	66.94	22.14	10.92
parzialmente specializzata	70.79	17.35	11.87	72.41	17.01	10.58
non specializzata	76.24	11.74	12.02	78.27	11.70	10.03
ABITAZIONE						
in affitto pubblico	73.05	17.76	9.19	76.01	16.77	7.22
In affitto privato	70.65	17.61	11.74	70.80	17.77	11.44
proprietario che vi abita	66.73	22.56	10.72	66.43	23.15	10.42
proprietario	63.62	17.00	19.38	77.23	17.44	15.32
REDDITO MEDIO FAMILIARE (£)	686.79	775.56	717.35	613.92	714.53	652.70
BAMBINI						
in età prescolare	0.25	0.26	0.22	0.21	0.24	0.23
5-18 anni	0.59	0.50	0.64	0.48	0.40	0.58
COMPOSIZIONE DELLA FAMIGLIA						
persona disoccupata	0.020	0.014	0.014	0.043	0.038	0.031
malata	0.004	0.002	0.003	0.039	0.030	0.020
in pensione	0.006	0.005	0.009	0.046	0.036	0.039
non occupata	0.007	0.004	0.008	0.080	0.100	0.138
studente a tempo libero	0.054	0.038	0.060	0.054	0.039	0.059
lav. dip.	0.148	0.125	0.158			
forma int.	0.033	0.039	0.030			
lav. aut.	0.006	0.006	0.015			

La *tabella 3* invece riporta la distribuzione dei tipi di lavoro nelle coppie con doppio guadagno da lavoro. E' evidente come il partner 2 abbia alta probabilità di essere lavoratore dipendente dato che questo è il tipo di lavoro più comune, al di là della situazione lavorativa del partner 1; ciò sembra favorire un pooling dei rischi intra-familiare. Va però detto che se si guardano più attentamente i dati si vede come l'incidenza di coppie con entrambi i partner lavoratori dipendenti piuttosto che autonomi sia piuttosto alta.

Tabella 3

		<i>Partner 1</i>					
		Lav. dip.		Lav. int.		Lav. aut.	
		N.	%	N.	%	N.	%
<i>Partner2</i>	Lav. dip.	8862	71.20	2338	58.88	1246	58.36
	Lav. int.	2338	18.79	1276	32.13	357	16.72
	Lav. aut.	1246	10.01	357	8.99	532	24.92

8.3 I risultati salienti

Il modello probit ordinale mette in luce una serie di risultati interessanti. Innanzitutto le donne hanno minore propensione al rischio rispetto agli uomini. L'età ha un impatto in forma concava (sebbene al livello di significatività del 10%) sul rischio. Alti livelli di istruzione sono positivamente associati ad una buona propensione al rischio. Per quanto riguarda la specializzazione richiesta, se questa è bassa, è basso anche il livello di rischio accettato. E' positivamente associato al rischio il livello di reddito, con un coefficiente piccolo in grandezza ma statisticamente significativo. Chi è proprietario di un'abitazione è propenso ad accettare alti rischi; ciò fa pensare che per chi possiede capitali e ricchezza la probabilità di essere lavoratori autonomi sia alta. E' difficile che un lavoratore con

Tabella 4

Caratteristiche	β	t-stat.
CARATTERISTICHE PERSONALI		
sexo femminile	-0.4000	-18.05
età	0.0197	2.37
età al quadrato	-0.0002	-1.81
livello 2 istruzione	0.0513	1.80
livello 3 istruzione	0.0742	2.28
livello 4 istruzione	0.1047	2.90
OCCUPAZIONE		
manageriale	-0.2243	-5.62
specializzata	-0.2065	-4.86
parzialmente specializzata	-0.1926	-4.01
non specializzata	-0.2707	-4.55
CARATTERISTICHE DELLA FAMIGLIA		
reddito familiare	0.0001	3.40
con bambini in età prescolare	0.0035	0.18
con membri in età 5-18 anni	0.0154	1.31
in affitto privato	0.0722	1.26
proprietario che vi abita	0.1412	3.52
proprietario	0.2929	5.78
COMPOSIZIONE DELLA FAMIGLIA		
persona disoccupata	-0.1912	-2.49
malata	-0.3386	-1.86
in pensione	0.0224	0.18
non occupata	-0.1184	-0.94
studente a tempo libero	-0.0017	-0.04
lav. dip.	-0.0296	-1.00
forma int.	-0.0409	-0.77
lav. aut.	0.2738	2.47
sposato/convivente con lav.int.	0.2521	10.67
sposato/convivente con la.aut.	0.5415	17.86
cut-off 1	1.0951	5.49
cut-off 2	1.9663	9.84
numero di osservazioni	18552	
log likelihood	-13944.957	
pseudo R quadro	0.1121	
statistica chi-quadro	3521.71 (53 gradi di libertà)	

associato alto rischio abbia in famiglia altre persone che siano disoccupate, cioè che non contribuiscano a massimizzare il profitto familiare. Inoltre le variabili sul tipo di lavoro degli altri membri della famiglia che lavorano suggerisce che il

fenomeno che vede i membri della famiglia occupati in lavori “simili” è più frequente se essi sono lavoratori autonomi.

Quanto al tipo di lavoro del partner, il fatto che il grado di rischio associato alla persona sia positivamente correlato al grado di rischio dello/a sposo/a o convivente va contro l’ipotesi che vi sia pooling dei rischi intrafamiliare.

I coefficienti dei cut-off sono distanti fra di loro, cosa che indica come l’ordinamento scelto sulla variabile dipendente sia adeguato.

Il modello ad effetti casuali porta a stabilire quanta parte della variazione sulle categorie di lavoro possa essere spiegata da una non osservabile correlazione interna alla coppia.

I coefficienti stimati, relativi al campione che include i soli membri delle coppie capofamiglia-partner lavoratori (*campione 1*), sono concordi a quelli riportati in *tabella 4*. Il ρ trovato è altamente significativo, e l’ampiezza suggerisce che l’11% dell’eterogeneità nell’avversione individuale al rischio (variabile latente), è spiegata dall’effetto coppia. Così, dato che l’effetto di coppia spiega l’11% della varianza non osservata, e dato che l’effetto specifico di coppia si basa sulle correlazioni attraverso la variabile dipendente nelle coppie, i risultati danno ulteriore supporto all’ipotesi che capofamiglia e partner abbiano buone probabilità di lavorare in occupazioni con stesso grado di rischio.

Per vedere se l’ipotesi di assenza di pooling dei rischi non valga anche in un più ampio contesto familiare, si prende in considerazione il campione formato da tutti i membri della famiglia che siano lavoratori (*campione 2*). Premesso che i coefficienti stimati concordano ancora una volta coi precedenti, il nuovo campione porta eterogeneità, e quindi una maggiore significatività di alcune variabili familiari.

L’ampiezza del ρ è minore, pari all’8%, ma di simile significatività; quindi rimane ancora apprezzabile uno specifico effetto, questa volta familiare e non più di coppia. L’ipotesi secondo cui capofamiglia e partner lavorano in occupazioni con stesso grado di rischio è ulteriormente rafforzata.

Tabella 5

Caratteristiche	Campione 1		Campione 2	
	β	t-stat.	β	t-stat.
CARATTERISTICHE PERSONALI				
di sesso femminile	-0.3365	-15.03	-0.3009	-17.46
età	0.0216	2.42	0.0472	9.04
età al quadrato	-0.0002	-1.83	-0.0005	-7.05
separato/divorziato/vedovo	-	-	-0.0320	-1.08
single	-	-	-0.0976	-3.80
livello 2 istruzione	0.0576	1.91	0.0660	2.87
livello 3 istruzione	0.0818	2.37	0.0947	3.62
livello 4 istruzione	0.1123	2.93	0.1439	5.01
OCCUPAZIONE				
manageriale	-0.2343	-5.55	-0.1463	-4.57
specializzata	-0.2146	-4.78	-0.1683	-4.99
parzialmente specializzata	-0.2107	-4.16	-0.1953	-5.18
non specializzata	-0.3047	-4.85	-0.3186	-6.89
CARATTERISTICHE DELLA FAMIGLIA				
reddito familiare	0.0001	3.83	0.0000	2.68
bambini in età prescolare	0.0062	0.29	0.0371	2.21
“bambini” in età 5-18 anni	0.0186	1.47	0.0121	1.26
in affitto	0.0834	1.34	0.1629	4.30
proprietario che vi abita	0.1625	3.73	0.1739	6.27
proprietario	0.3554	6.44	0.2708	7.88
COMPOSIZIONE DELLA FAMIGLIA				
persona disoccupata	-0.2447	-2.93	-0.1072	-2.68
malata	-0.4071	-2.06	-0.1696	-3.79
In pensione	0.0488	0.36	-0.1102	-2.72
non occupata	-0.1478	-1.09	0.0403	1.50
studente a tempo libero	-0.0127	-0.26	-0.0403	-1.10
lav. dip.	-0.0397	-1.23	-	-
forma int.	-0.0488	-0.84	-	-
lav. aut.	0.3466	2.87	-	-
cut-off 1	1.1278	5.28	1.6531	11.85
cut-off 2	2.0374	9.51	2.5589	18.26
ρ	0.1147	8.16	0.0768	8.11
numero di osservazioni	18552		31862	
log likelihood	-14089.011		-23844.285	
statistica chi-quadro	3158.97 (51 g. di l.)		5242.03 (50 g. l.)	

Va detto che il risultato che vede la correlazione tra i rischi dei due membri della coppia essere positiva era ampiamente prevedibile. Infatti il modello utilizzato, il probit ordinale ad effetti casuali, impone una struttura di covarianza tale che la

correlazione tra i rischi non possa che essere positiva o al più nulla. Del resto, come visto nel *paragrafo 8.1*, il modello adottato assume una specificazione ad effetti casuali con termine d'errore composito $v_{ic} = \alpha_c + \eta_{ic}$, con $\eta_{ic} \sim \text{IN}(0, \sigma_\eta^2)$, e, condizionatamente a x_{ic} , con gli α_c (specifico effetto di coppia non osservabile) che sono $\text{IN}(0, \sigma_\alpha^2)$ e indipendenti dagli η_{ic} e dagli x_{ic} . Questo implica che la struttura di covarianza degli errori è data dalla matrice:

$$V\{\alpha_c \iota_2 + \eta_c\} = \Omega = \sigma_\alpha^2 \iota_2 \iota_2' + \sigma_\eta^2 I_2,$$

con $\iota_2 = (1, 1)'$, $\eta_c = (\eta_{c1}, \eta_{c2})'$ e I_2 matrice d'identità di dimensione 2.

Quindi ρ , essendo pari a $\sigma_\alpha^2 / (\sigma_\alpha^2 + \sigma_\eta^2)$, non può essere negativo.

8.4 A quali conclusioni si può arrivare

L'obiettivo dell'analisi "Employment Contract Matching" era studiare se, all'interno della coppia (famiglia), a prevalere sia una situazione che veda entrambi i partner (familiari) lavoratori autonomi piuttosto che dipendenti, o se invece vi sia effettivamente diversificazione del portafoglio delle categorie lavorative di appartenenza.

I risultati empirici vanno chiaramente nella prima direzione, ed una spiegazione a ciò si trova in letteratura. Innanzitutto si deve ricordare l'incidenza che ha il lavoro autonomo nella scelta occupazionale del partner, vale a dire che se una persona è lavoratore autonomo è facile lo diventi anche il partner. La letteratura, nell'arrivare a questa conclusione, si sofferma però sulla distinzione lavoro autonomo – dipendente, trascurando la modalità intermedia introdotta invece dall'indagine appena vista.

Il fatto che sia probabile che le persone occupate con contratti lavorativi simili "si mettano assieme" trova due particolari spiegazioni:

- 1) come sostiene la più recente letteratura, è più facile trovare un partner, che sia della stessa opinione e quindi con cui si vada d'accordo, se quest'ultimo fa il nostro stesso tipo di lavoro; ciò può divenire preminente rispetto ai benefici da pooling dei rischi;

2) i vantaggi amplificati in termini di guadagno, che possono venire dall'unione di persone che siano entrambe lavoratori autonomi o appartenenti alla modalità intermedia, compensano il più alto livello di rischio che parallelamente si realizza.

Quale delle due spiegazioni sia preminente non è dato a sapersi, anche se in ogni caso è probabile che entrambe siano rilevanti. Data la portata delle implicazioni che derivano da una situazione che non favorisce il pooling dei rischi, quale è quella di cui si è avuto riscontro empirico, tali aspetti a livello di spiegazioni meriterebbero maggiore considerazione. Infatti se le coppie in cui i partner fanno lavori simili favoriscono una maggiore produttività familiare, parallelamente a maggiori guadagni e una più serena concordia familiare, allora questo può avere ricadute positive per l'intera economia, scoraggiando l'assenteismo e il continuo ricambio sul posto di lavoro.

CAPITOLO 3 Introduzione all'analisi empirica

9 *Gli obiettivi*

L'analisi empirica della tesi ha l'obiettivo di indagare la presenza di risk sharing intra-coppia, e come questo eventualmente influenzi la scelta dei lavoratori sposati/conviventi tra più categorie occupazionali distinte a seconda del diverso grado di rischio sul guadagno che è loro associato.

Due sono i risultati che l'analisi può attendersi; infatti può accadere che la soluzione preponderante veda i partner impegnati in tipi di lavoro diversi dal punto di vista del rischio, realizzando una diversificazione del portafoglio dei contratti lavorativi, e quindi favorendo un pooling dei rischi intra-coppia. Alternativamente può succedere che le coppie costituite da capofamiglia e partner, entrambi lavoratori, vedano i due scegliere prevalentemente tipi di lavoro che rientrino nella stessa categoria occupazionale; in questo caso vuol dire che vi è un effetto di coppia che spinge i partner ad essere occupati in attività lavorative con associato lo stesso livello di rischio del coniuge/convivente. Gli obiettivi dell'analisi empirica della tesi vengono quindi di fatto a coincidere con quelli perseguiti dall'indagine inglese "Employment Contract Matching".

Nell'analisi sul contesto italiano che a breve sarà presentata, una prima distinzione presa in considerazione è quella tra lavoro autonomo e dipendente, e ciò equivale a dire che se vi fosse risk-sharing intra-coppia i partner lavoratori tenderebbero l'uno ad essere autonomo, e l'altro ad essere lavoratore dipendente. Le categorie occupazionali possono anche essere più di due, come già visto; l'analisi empirica infatti distinguerà anche tra lavoro dipendente pubblico e lavoro dipendente privato. Ovviamente pure in questo caso la distinzione tra categorie occupazionali assume valore in quanto ad esse si presume siano associati gradi di rischio diversi.

Queste sono comunque considerazioni che ho già ampiamente esposto nei capitoli precedenti. Invece, per quanto riguarda gli strumenti utilizzati per vedere se l'ipotesi di risk sharing intrafamiliare sia vera, ovvero quali sono i modelli adottati, rimando al capitolo 4. Questo capitolo vuole invece introdurre l'analisi empirica, a partire da quale sia la fonte dei dati utilizzati e da una sua breve illustrazione. Il capitolo si sofferma poi sui sotto-campioni individuati, su come le caratteristiche individuali e familiari si collochino rispetto alle categorie occupazionali, e su di una breve descrizione delle variabili utilizzate, a partire da quelle dipendenti.

10 *La fonte dei dati*

La fonte dei dati utilizzati è l'*Indagine sui bilanci delle famiglie italiane* della Banca d'Italia. Tale indagine nasce negli anni '60 con l'obiettivo di raccogliere informazioni sui redditi e sui risparmi delle famiglie italiane. Nel corso degli anni l'oggetto della rilevazione si è andato estendendo includendo anche la ricchezza e altri aspetti inerenti i comportamenti economici e finanziari delle famiglie.

Si tratta di un'indagine campionaria; nelle ultime rilevazioni il campione è formato da circa 8mila famiglie, comprendenti 24mila individui, distribuite in circa 300 comuni italiani. I risultati dell'indagine vengono regolarmente pubblicati nei Supplementi al Bollettino Statistico della Banca. I dati raccolti presso le famiglie, in forma anonima, sono disponibili per elaborazioni e ricerche. La possibilità di avere accesso a dei dataset che contengono informazioni personali dettagliate, oltre che dati su salari e ore di lavoro, dà all'indagine della Banca d'Italia un grosso vantaggio rispetto ad altre fonti come l'Istat o l'Inps; infatti quest'ultimi sono forse sì più dettagliati nel riportare in modo accurato le informazioni su salari e ore di lavoro, ma sono carenti rispetto alle caratteristiche individuali dei soggetti in considerazione.

L'indagine avviene attraverso un questionario sottoposto alle persone e alle famiglie prescelte. La mia tesi fa riferimento al questionario del 2000, il più recente a mia disposizione.

10.1 *Indagine sui bilanci delle famiglie nel 2000: il questionario*

L'indagine ha riguardato 8001 famiglie, 22268 individui e 13814 percettori di reddito. Delle 8001 famiglie, 3873 sono famiglie panel, ovvero intervistate anche nella precedente rilevazione, e le rimanenti 4128 sono famiglie intervistate per la prima volta nell'indagine sul 2000.

Tutte le informazioni date si riferiscono alla situazione, individuale e familiare, al 31-12-2000; questo non vuol dire che di sovente vengano richieste notizie relative al corso dell'anno 2000 o anche precedenti. Il questionario è strutturato in alcune sezioni, a seconda delle informazioni richieste nel corso dell'intervista.

La sezione A indaga la struttura della famiglia, descrivendo le qualifiche dei componenti della famiglia relativamente a sesso, posizione nella famiglia, luogo e anno di nascita, stato civile, luogo di abitazione e titolo di studio. Vengono anche raccolte informazioni sui componenti usciti dalla famiglia negli anni 1999-2000 e sui nuovi entrati in famiglia; alcune domande sono riservate a capofamiglia e coniuge/convivente, e riguardano titolo di studio, condizione professionale e settore di attività.

La sezione B si focalizza su occupazione e redditi. E' interessante notare che l'indagine distingue tra soggetti non occupati, occupati dipendenti e occupati indipendenti, per poi andare più nel dettaglio delle varie professioni. Si indagano anche le opportunità di lavoro avute nel corso del 2000 e le attività lavorative svolte nella vita. Una sotto-sezione ad hoc è dedicata al lavoro non pagato e ai servizi per la famiglia. A seconda delle risposte date sulla condizione professionale, alcune domande a parte sono rivolte a lavoratori dipendenti, a liberi professionisti, imprenditori individuali, lavoratori autonomi, collaboratori coordinati e continuativi, alle imprese familiari, a soci e gestori di società, e ai pensionati, e sono raccolte nei rispettivi allegati. Un allegato sul reddito riporta le "altre entrate", quali, per esempio, borse di studio e assegni per alimenti.

La sezione C riguarda gli strumenti di pagamento e le forme di risparmio. Si occupa invece di abitazione di residenza e di altri beni immobili la sezione D, con anche gli allegati D1 (beni immobili posseduti alla fine del 2000) e D2 (beni

immobili venduti o donati nel 2000). La successiva sezione E indaga i consumi e le altre spese familiari, con una sotto-sezione che si occupa di “comportamenti di consumo”. Infine, la sezione G valuta le forme assicurative che interessano famiglie e individui, quali ad esempio le pensioni private integrative, e la sezione G riporta notizie a cura dell’intervistatore.

11 Come è formato il sotto-campione preso in analisi

Il sotto-campione su cui la tesi sviluppa l’analisi empirica è formato dalle coppie coniugi/conviventi in cui entrambi lavorano o comunque l’individuo di sesso maschile è occupato.

I coniugi/conviventi presi in considerazione sono quelli in età da lavoro. Il maschio deve avere età compresa tra i 18 e i 65 anni, mentre la persona di sesso femminile coniuge/convivente deve avere non più di 60anni. Affinché i due partner siano tali ai fini dell’analisi empirica, devono, alla domanda su quale sia la loro posizione all’interno della coppia, aver risposto “capofamiglia” (cioè principale percettore di reddito familiare) o “coniuge/convivente del capofamiglia”; ovviamente si prendono in considerazione solo le famiglie ove vi siano entrambi i ruoli. Si noti come, se in Employment Contract Matching i due partner sono distinti in base al ruolo di capofamiglia (partner 1) o meno (partner 2), nella mia tesi invece la distinzione è tra marito/convivente maschio e moglie/convivente femmina, al di là del ruolo di capofamiglia.

Relativamente allo status occupazionale, si considera la situazione prevalente nel corso dell’anno 2000. Si escludono doppioni di osservazioni già rilevate (mi è capitato di trovarne due); ciò vale anche per le osservazioni di individui che in CARCOM00 (dataset che dà le informazioni sulla condizione prevalente) dichiarano di essere, dal punto di vista dell’occupazione principale svolta nel corso del 2000, lavoratori dipendenti, per poi, in ALLB1 (dataset contenente le risposte alle domande sottoposte ai soli lavoratori dipendenti), non rispondere alla domanda che indirettamente indaga se il lavoro dipendente sia nel settore pubblico piuttosto che nel privato.

Va detto che il sotto-campione su cui si forma l'analisi empirica è quello delle coppie in cui entrambi i partner lavorano. Le coppie che rispondono alle caratteristiche fin qui elencate, e che quindi formano il campione, sono 1569, per un totale di 3138 osservazioni. Se si includono nel sotto-campione anche le coppie ove il marito/convivente di sesso maschile lavora, ma con moglie/convivente di sesso femminile non occupata, allora il numero complessivo di coppie sale a 3189 (6378 osservazioni); infatti le coppie col solo marito/convivente lavoratore sono 1620, a cui corrispondono 3240 osservazioni.

Vediamo ora come è composto il sotto-campione comprendente le sole coppie in cui entrambi i partner lavorano, rispetto alle classi occupazionali di diverso rischio, a seconda delle suddivisioni considerate.

11.1 La combinazione lavoratori autonomi - dipendenti tra i partner dei due sessi

La distinzione classica a livello occupazionale è quella tra lavoro autonomo e lavoro dipendente, con la prima che vede convenzionalmente associato un grado di rischio maggiore. La combinazione tra i due tipi di lavoro rispetto ai partner dei due sessi dà i seguenti risultati:

Tabella 6 coppie in cui entrambi i partner lavorano

FEMMINE		MASCHI					
		Lav. dip.		Lav. aut.		TOTALE	
		N.	% ↓	N.	% ↓	N.	% ↓
Lav. dip.	N.	1059	91.85	237	56.97	1296	82.60
	% →	81.71		18.29		100	
Lav. aut.	N.	94	8.15	179	43.03	273	17.40
	% →	34.43		65.57		100	
TOTALE	N.	1153	100	416	100	1569	100
	% →	73.49		26.51		100	

E' evidente come sia il partner di sesso femminile ad avere più possibilità di essere lavoratore dipendente, al di là della situazione lavorativa di quello di sesso maschile; infatti 1296 donne contro 1153 uomini sono lavoratrici dipendenti. Ciò sembra favorire un pooling dei rischi intra-familiare. Va però detto che se si guardano più attentamente i dati si vede come l'incidenza di coppie con entrambi i partner lavoratori autonomi piuttosto che dipendenti sia abbastanza alta, per lo meno a livello di tendenza e al di là della preponderanza numerica dei singoli lavoratori dipendenti. Il dato che spicca maggiormente è come il lavoratore dipendente di sesso maschile abbia altissime possibilità (91.85%) di "essere assieme" ad una lavoratrice dipendente. Inoltre il lavoratore autonomo di sesso femminile ha maggiore probabilità di convivere con una persona appartenente alla stessa categoria occupazionale di quanto non succeda a parti invertite.

11.2 *L'ulteriore distinzione tra lavoratori dipendenti pubblici e privati*

Le classi occupazionali che si presume siano associate a gradi di rischio diverso possono essere anche tre; infatti la tesi considera anche la distinzione tra lavoro dipendente nel settore pubblico e lavoro dipendente nel settore privato, oltre naturalmente alla categoria occupazionale dei lavoratori autonomi.

L'ulteriore differenziazione tra lavoratori dipendenti nel pubblico e lavoratori dipendenti nel privato è motivata dal fatto che si presume che i primi appartengano ad una categoria occupazionale con grado di rischio inferiore rispetto ai secondi. Entrambi i lavori sono comunque più sicuri rispetto a quello autonomo. In questo modo si ottengono tre categorie occupazionali con associati diversi gradi di rischio.

La combinazione tra i tre tipi di lavoro rispetto ai partner dei due sessi dà i risultati riportati in *tabella 7*.

Rispetto ai risultati evidenziati in *tabella 6*, quando si distingueva solo tra lavoro autonomo e lavoro dipendente, l'incidenza delle coppie ove i partner sono occupati in lavori appartenenti alla stessa categoria di rischio è ancora ben marcata. Infatti le probabilità, per i partner di entrambi i sessi, di "essere

assieme” a lavoratori che fanno lo stesso tipo di lavoro, sono maggiori di quelle che hanno di incrociare partner occupati in una delle altre due categorie occupazionali prese singolarmente, e spesso anche se sommate. A conferma di ciò, le probabilità di convivere con partner che fanno lo stesso tipo di lavoro sono, per entrambi i sessi, spesso superiori al 50%, anche di parecchio.

Tabella 7 coppie in cui entrambi i partner lavorano

MASCHI									
FEMMINE		Lav. dip. pubblico		Lav. dip. privato		Lav. aut.		TOTALE	
		N.	% ↓	N.	% ↓	N.	% ↓	N.	% ↓
Lav. dip. pubblico	N.	227	67.36	177	21.69	84	20.19	488	31.10
	% →	46.52		36.27		17.21		100	
Lav. dip. privato	N.	78	23.15	577	70.71	153	36.78	808	51.50
	% →	9.65		71.41		18.94		100	
Lav. aut.	N.	32	9.50	62	7.60	179	43.03	273	17.40
	% →	11.72		22.71		65.57		100	
TOTALE	N.	337	100	816	100	416	100	1569	100
	% →	21.48		52.01		26.51		100	

Già in *tabella 6* si era visto come la donna lavoratrice autonoma abbia più alte probabilità di avere partner lavoratore autonomo di quanto non succeda a parti invertite (65.57 per le donne contro 43.03 per gli uomini); in *tabella 7* si può notare come valga il discorso opposto per quanto riguarda il lavoro dipendente nel pubblico (46.52 per le donne contro 67.36 per gli uomini). Simili e intorno al

70% sono invece le probabilità per i lavoratori dipendenti nel privato di entrambi i sessi di incontrare partner appartenenti alla stessa categoria occupazionale.

Un'altro risultato che si può mettere in luce è come il partner di sesso femminile abbia maggiori probabilità di essere lavoratore dipendente, soprattutto nel settore pubblico, rispetto al partner di sesso maschile; quest'ultimo ha invece maggiori probabilità di essere lavoratore autonomo piuttosto che lavoratore dipendente nel pubblico. Quanto al lavoro dipendente nel settore privato, questo si distribuisce equamente tra i due sessi; infatti il 50% circa di mariti/conviventi ma anche di mogli/conviventi è dipendente privato.

12 *L'incidenza delle caratteristiche individuali e familiari sul tipo di lavoro*

Vediamo ora qual è la distribuzione delle diverse caratteristiche individuali e familiari, attraverso il sotto-campione formato dalle coppie con entrambi i partner lavoratori, rispetto ai tipi di lavoro con differente livello di rischio.

Si prendono in considerazione le tre categorie occupazionali (lavoro autonomo, lavoro dipendente pubblico e lavoro dipendente privato) di cui sono riportate le percentuali rispetto all'insieme dei lavoratori. Si considera anche la categoria dei lavoratori dipendenti, le cui percentuali sono naturalmente corrispondenti alla somma di quelle relative a dipendenti nel pubblico e nel privato.

12.1 *Le caratteristiche individuali*

La *tabella 8* riporta i risultati relativi al sesso dei partner, alla loro classe d'età e al loro eventuale titolo di studio. Hanno maggiori probabilità di essere lavoratrici dipendenti, soprattutto nel pubblico, le donne. Quanto alle classi d'età, si nota come all'aumentare degli anni anagrafici vi sia una tendenza ad una maggiore proporzione di lavoratori autonomi e di lavoratori dipendenti pubblici, ed una minore probabilità di incontrare lavoratori dipendenti privati; questo è evidente soprattutto se si raffrontano la prima e la quarta classe d'età costruite.

Per quanto riguarda il titolo di studio, si può osservare come al suo aumentare non corrisponda un sensibile incremento della probabilità di diventare lavoratore

autonomo, semmai vi è un leggero decremento. Si assiste invece ad una redistribuzione all'interno del lavoro dipendente; in sostanza aumentano le possibilità di lavorare nel pubblico e calano quelle di essere dipendenti privati.

Tabella 8

Caratteristiche	Lavoratore dipendente pubblico	Lavoratore dipendente privato	Lavoratore dipendente (totale)	Lavoratore autonomo
SESSO				
maschile	21.48	52.01	73.49	26.51
femminile	31.10	51.50	82.60	17.40
ETA'				
18-30	11.37	75.83	87.20	12.80
31-40	22.40	56.19	78.59	21.41
41-50	31.21	48.84	80.05	19.95
oltre 51	28.38	40.04	68.42	31.58
TITOLO DI STUDIO				
nessuno	10.00	90.00	100.00	0.00
lic. Elementare	4.49	59.18	63.67	36.38
lic. media inf.	14.65	59.71	74.36	25.64
dipl.professionale	18.63	60.84	79.47	20.53
dipl. media sup.	31.42	51.07	82.49	17.51
dipl.universitario	57.78	35.56	93.33	6.67
laurea	47.49	31.65	79.16	20.84
spec.post-laurea	30.00	30.00	60.00	40.00

12.2 Le caratteristiche occupazionali

Metto ora in evidenza alcune caratteristiche dei tipi di lavoro rispetto allo status del lavoratore e al settore di attività di riferimento.

Quanto allo status del lavoratore, i risultati relativi alla distinzione autonomo – dipendente riportati in *tabella 9* sono poco significativi; infatti per definizione operai, impiegati e dirigenti sono lavoratori dipendenti, mentre imprenditori e liberi professionisti non possono che essere lavoratori autonomi; più interessante è vedere come si distribuiscono i diversi lavori dipendenti rispetto a pubblico e privato.

Per quanto concerne il settore di attività è evidente come il lavoratore dipendente privato sia quello che si può trovare un po' in tutti i settori; ovviamente il

lavoratore autonomo è scarsamente presente nel settore dei servizi pubblici, che invece appartiene per antonomasia al lavoratore dipendente pubblico.

Tabella 9

Caratteristiche	Lavoratore dipendente pubblico	Lavoratore dipendente privato	Lavoratore dipendente (totale)	Lavoratore autonomo
STATUS DEL LAVORATORE				
operaio	10.74	89.26	100.00	0.00
impiegato	48.40	51.60	100.00	0.00
dirigente/direttivo	37.17	62.83	100.00	0.00
imprenditore/libero professionista	0.00	0.00	0.00	100.00
altro autonomo	0.00	0.00	0.00	100.00
SETTORE DI ATTIVITA'				
agricoltore	3.73	53.73	57.46	42.54
industria	0.98	79.70	80.67	19.38
servizi pubblici	73.57	23.62	97.19	2.81
altri settori (terziario, commercio, artigianato)	4.86	54.67	59.52	40.48

12.3 Alcuni aspetti demografici e geografici

La *tabella 10* mette in luce alcuni aspetti di tipo demografico e geografico.

Quanto all'ampiezza demografica del comune di residenza dell'individuo appartenente ad una coppia, il dato che emerge più chiaramente è come la proporzione di lavoratori dipendenti pubblici aumenti all'aumentare del numero di abitanti del comune. Per il resto si può notare come i lavoratori autonomi siano particolarmente presenti, in proporzione, nei comuni di dimensione medio-piccola (20.000 – 40.000 abitanti), molto meno nelle grandi città.

A livello di area geografica di appartenenza, è evidente l'alta proporzione di lavoratori dipendenti pubblici nel Sud e nelle Isole, regioni dove invece è meno probabile incontrare lavoratori dipendenti privati. Il Nord Est si distingue per la relativamente alta proporzione di lavoratori autonomi (come il Centro Italia), per la particolarmente bassa presenza di dipendenti pubblici, e per essere l'area

geografica con la maggior presenza di dipendenti privati, seppur tallonata dal Nord Ovest.

Tabella 10

Caratteristiche	Lavoratore dipendente pubblico	Lavoratore dipendente privato	Lavoratore dipendente (totale)	Lavoratore autonomo
AMPIEZZA DEMOGRAFICA DEL COMUNE DI RESIDENZA				
0-20.000	22.81	54.38	77.19	22.81
20.000-40.000	22.40	50.47	72.87	27.13
40.000-500.000	30.03	49.29	79.32	20.68
oltre 500.000	27.35	60.26	87.61	12.39
AREA GEOGRAFICA DI RESIDENZA				
Nord Ovest	23.75	57.50	81.25	18.75
Nord Est	18.50	58.33	76.84	23.16
Centro	22.95	52.73	75.68	24.32
Sud	41.49	36.81	78.30	21.70
Isole	42.50	34.58	77.08	22.92

12.4 Le caratteristiche familiari

La *tabella 11* indaga la correlazione tra alcune caratteristiche familiari e le categorie occupazionali.

All'aumentare del numero di componenti della famiglia aumenta la proporzione di lavoratori dipendenti pubblici e cala quella relativa alla presenza di dipendenti privati; il numero di componenti una famiglia non incide invece più di tanto sulla proporzione di lavoratori autonomi.

Se sono presenti bambini in età prescolare diminuiscono le possibilità di incontrare dipendenti pubblici e aumenta la presenza, in proporzione, di dipendenti privati. Vale esattamente il discorso contrario relativamente alla presenza in famiglia di componenti con età compresa tra i 6 e i 18 anni. In entrambi i casi la presenza di minorenni porta ad un lieve incremento della proporzione di lavoratori dipendenti a danno degli autonomi

Quanto al numero di percettori di reddito (esclusi i membri della coppia capofamiglia- -partner), si può notare come, all'aumentare del numero di

percettori, vi sia un sostanziale calo della proporzione di lavoratori dipendenti, soprattutto pubblici; parallelamente invece aumenta la proporzione di lavoratori autonomi.

Tabella 11

Caratteristiche	Lavoratore dipendente pubblico	Lavoratore dipendente privato	Lavoratore dipendente (totale)	Lavoratore autonomo
NUMERO DI COMPONENTI DELLA FAMIGLIA				
2	19.39	58.37	77.76	22.24
3	20.58	58.63	79.21	20.79
4	30.83	46.97	77.80	22.20
5	36.61	41.73	78.35	21.65
oltre 5	32.90	35.53	68.42	31.58
BAMBINI IN ETA' PRESCOLARE				
non presenti	27.39	50.33	77.72	22.28
presenti	22.68	56.42	79.10	20.90
COMPONENTI IN ETA' 6-18 ANNI				
non presenti	21.15	54.98	76.13	23.87
presenti	30.80	48.92	79.72	20.28
NUMERO DI PERCETTORI DI REDDITO (ESCLUSI I PARTNER)				
0	27.08	51.85	78.93	21.07
1	25.68	50.68	76.36	23.64
2	14.93	55.22	70.15	29.85
oltre 2	15.00	40.00	55.00	45.00

13 *Le variabili*

Come già detto, l'analisi empirica si avvale di alcuni modelli di regressione probit, su cui mi soffermerò nei capitoli 4 e 5. Voglio ora introdurre le variabili dipendenti ed indipendenti che verranno utilizzate, spiegando come sono state create o comunque cosa indicano. Si ricorda che il sotto-campione in cui vengono utilizzate è quello delle coppie con entrambi i partner lavoratori.

13.1 *La variabile dipendente*

Se si distingue solo tra lavoro dipendente e autonomo, si utilizza una variabile dipendente **LAVORO** che assume valori 0, se l'individuo è lavoratore dipendente, e 1, se l'individuo è lavoratore autonomo. Al lavoro dipendente si associa grado di rischio inferiore rispetto a quello autonomo.

Se invece si considera l'ulteriore distinzione tra lavoro dipendente nel pubblico e nel privato, la variabile dipendente diventa **LAV3**; si tratta di una variabile categorica che assume valore 1 se l'individuo è lavoratore dipendente pubblico, 2 se è dipendente privato, e 3 se è lavoratore autonomo. L'ordinamento adottato poggia sulla supposizione che al lavoro dipendente nel pubblico sia associato il grado di rischio inferiore e che il lavoro autonomo sia il meno sicuro dei tre. I dati grezzi confermano che tale ordinamento è adeguato. Infatti, con particolare riferimento all'associazione di un minor grado di rischio al lavoro dipendente pubblico rispetto al privato, le deviazioni standard dei guadagni (al netto di tasse e contributi) attorno alla media sono 13609,16 nel pubblico e 14007,08 nel privato. A supportare la bontà della distinzione tra dipendente pubblico – privato e lavoro autonomo sarà la distanza tra i cut-off riscontrata nei risultati dell'analisi empirica (capitolo 5).

Nell'ultimo paragrafo della tesi, quando si prenderanno in considerazione tutte le coppie con partner di sesso maschile occupato e moglie/convivente non necessariamente lavoratrice, si fa riferimento anche ad altre endogene. In particolare si utilizza **MOGLIELNL**, che assume valore 0 se la partner di sesso femminile non lavora, 1 se è invece occupata. Torna utile anche la variabile **MARITOLAVORO**, costruita esattamente come l'endogena **LAVORO**, ma che si riferisce alla situazione lavorativa dei soli mariti/conviventi; infine si utilizza anche la variabile **MARITODIP**, anch'essa riferita all'occupazione dei partner maschi, ma che assume valore 0 se il marito/convivente è lavoratore autonomo, 1 se è dipendente.

13.2 *Le variabili indipendenti*

Le variabili esplicative utilizzate naturalmente si rifanno alle caratteristiche individuali e familiari già viste, e vogliono mettere in evidenza la loro correlazione col grado di rischio. Infatti, significatività a parte, se ad esempio ad una variabile dummy corrisponde un coefficiente positivo, vuol dire che qualora un individuo abbia una caratteristica tale da corrispondere al valore 1 della dummy, questo avrà buone possibilità di incontrare alto income risk rispetto a chi ha la caratteristica opposta e compatibilmente con le altre qualità che gli appartengono. Questo ovviamente accade se la variabile dipendente è realizzata come visto nel sottoparagrafo precedente, cioè con, all'aumentare del valore, più rischio associato.

Una variabile indipendente che assume valore particolare ai fini degli obiettivi che guidano l'analisi empirica, e che vedremo essere utilizzata solo in alcuni modelli probit, è quella sulla categoria occupazionale del partner. Si tratta della variabile esplicativa **PARTNERLAVORO** che è costruita esattamente come l'endogena **LAVORO**, ma che si riferisce per l'appunto alla situazione lavorativa del partner di un dato individuo. Quando invece si considerano le tre opzioni (lavoro autonomo, dipendente pubblico e dipendente privato) non è possibile utilizzare una variabile che assuma i valori 1, 2 e 3, perché in tal caso si andrebbe ad ipotizzare una relazione lineare che non è detto vi sia. Vengono quindi costruite delle dummy; in particolare la dummy **PARTAUT**, che assume valore 1 se il partner è lavoratore autonomo e 0 altrimenti, e la dummy **PARTPUBBL**, che assume valore 1 se il partner è lavoratore dipendente pubblico e 0 altrimenti.

Le altre variabili esogene sono:

MOGLIE: variabile dummy che, dato il campione delle coppie capofamiglia-partner, assume valore 0 se l'individuo è marito/convivente e 1 se invece è di sesso femminile;

ETA: variabile che dà l'età del partner;

ETASQ: è ETA al quadrato;

NCOMP: indica il numero di componenti la famiglia di un individuo;

NPRED: indica il numero di percettori di reddito della famiglia di un dato individuo; sono esclusi dal conteggio i due partner lavoratori. Nel caso il sottocampione di riferimento includa le coppie con partner maschio lavoratore e partner femmina non necessariamente occupata, come accade nel paragrafo 20, si utilizza **NPRED12**; questa variabile esclude ancora dal conteggio i partner lavoratori, ma se per NPRED questi sono sempre due (entrambi i coniugi/conviventi), ora ve ne possono essere due oppure, nel caso la moglie non lavori, solo uno;

PRESC: variabile che indica quanti sono i bambini in età prescolare in famiglia (assume valore 0 se non ve ne sono);

SCSUP: variabile dummy che indica se l'individuo è in possesso di un diploma di media superiore o di diploma professionale della durata di tre anni (in questi casi si assume valore 1) o meno (valore 0);

UNIV: variabile dummy che indica se l'individuo è laureato (compresa la laurea breve) o anche in possesso di specializzazione post-laurea (in questi casi si assume valore 1) o meno (valore 0);

AMP40: variabile dummy che indaga se l'individuo risiede in un comune che non supera i 40.000 abitanti (valore 1) o meno (valore 0);

AMP500: variabile dummy che indaga se l'individuo risiede in un comune che non supera i 500.000 abitanti ma comunque di ampiezza demografica superiore alle 40.000 unità (valore 1) o meno (valore 0);

areaN: dummy che indica se la persona risiede nel nord Italia (valore 1) o al Centro-Sud (valore 0);

areaS: dummy che indica se l'individuo risiede nel Sud Italia (valore 1) o al Centro Nord (valore 0).

Quando le variabili ETA, ETASQ, SCSUP e UNIV sono seguite da una **M** o da una **F**, vuol dire che le esogene, che si riferiscono a caratteristiche propriamente individuali, sono costruite in modo da dare informazioni che interessano direttamente rispettivamente gli individui di sesso maschile e femminile della coppia.

CAPITOLO 4 I dati panel e i modelli probit

14 I modelli con variabile dipendente discreta

Nelle applicazioni pratiche spesso ci si trova di fronte a fenomeni che sono di natura discreta o in parte discreta e in parte continua. Ad esempio si può essere interessati ad indagare se la persona sposata/convivente sia lavoratrice o meno, oppure quante ore alla settimana lavori (nessuna o in numero positivo). Se questo tipo di variabili devono essere spiegate, un modello di regressione lineare è generalmente inappropriato. Vi sono alcuni modelli che possono essere usati per modellare variabili discrete o in parte discrete e in parte continue.

In molti casi i problemi analizzati con questo tipo di modelli sono di natura micro-economica, e così richiedono dati sugli individui, sulle famiglie o sulle aziende.

Il modello per scelta binaria e quello per scelta multipla sono i due modelli con variabile dipendente discreta che interessano ai fini della tesi.

14.1 I modelli con scelta binaria

Quando vi è scelta binaria, vale a dire quando la variabile dipendente può assumere le sole alternative discrete 0 ed 1, non è appropriato utilizzare il classico modello di regressione.

Infatti ad esempio si supponga di avere un campione di N famiglie ($i=1, \dots, N$) di cui si osserva il reddito, x_{i2} , e si voglia sapere se la casa in cui abitano è di loro proprietà o meno; ipotizziamo che venga utilizzato il classico modello di regressione. La situazione in esame può essere descritta dalla variabile dipendente binaria y_i , che viene ad assumere valori 1 (la famiglia è proprietaria) e 0 (non proprietaria). Il modello lineare può allora essere così scritto così:

$$y_i = \beta_1 + \beta_2 x_{i2} + \varepsilon_i = x_i' \beta + \varepsilon_i$$

con $x_i = (1, x_{i2})'$. A questo punto sembra ragionevole fare l'assunzione standard che $E\{\varepsilon_i | x_i\} = 0$, tale che $E\{y_i | x_i\} = x_i' \beta$. Questo implica che

$$E\{y_i | x_i\} = 1 \cdot P\{y_i = 1 | x_i\} + 0 \cdot P\{y_i = 0 | x_i\} = P\{y_i = 1 | x_i\} = x_i' \beta$$

Così, il modello lineare porta $x_i' \beta$ ad essere una probabilità che deve stare tra 0 ed 1. Questo è possibile solo se i valori degli x_i sono vincolati e se certe restrizioni su β sono soddisfatte. Normalmente ciò è difficile da mettere in pratica; inoltre il termine d'errore ε_i ha una distribuzione altamente non-normale ed è affetto da eteroschedasticità. Siccome y_i ha due soli risultati possibili (0 e 1), il termine d'errore, per un dato valore di x_i , ha anch'esso due soli risultati possibili. In particolare la distribuzione di ε_i può essere così riepilogata:

$$P\{\varepsilon_i = -x_i' \beta | x_i\} = P\{y_i = 0 | x_i\} = 1 - x_i' \beta$$

$$P\{\varepsilon_i = 1 - x_i' \beta | x_i\} = P\{y_i = 1 | x_i\} = x_i' \beta$$

Questo vuol dire che la varianza del termine d'errore non è costante ma dipende dalle variabili esplicative, con $V\{\varepsilon_i | x_i\} = x_i' \beta (1 - x_i' \beta)$; il termine d'errore viene così a dipendere dal parametro del modello, β .

Per superare questi problemi viene quindi in aiuto una classe di modelli con scelta binaria.

Questi modelli essenzialmente descrivono la probabilità che y_i assuma valore 1 direttamente, sebbene siano spesso derivati da una variabile latente sottostante. In generale si ha $P\{y_i = 1 | x_i\} = G(x_i, \beta)$ per alcune funzioni $G(\cdot)$. Questa equazione dice che la probabilità di avere $y_i = 1$ dipende dal vettore x_i contenente le caratteristiche individuali, quali ad esempio età e sesso. Chiaramente la funzione $G(\cdot)$ deve assumere valori compresi nell'intervallo $[0, 1]$. Usualmente si pone l'attenzione su funzioni della forma $G(x_i, \beta) = F(x_i' \beta)$. Anche $F(\cdot)$ deve stare tra 0 ed 1, ed è naturale scegliere F come funzione di una qualche distribuzione.

Le scelte più comuni sono la funzione di distribuzione standard normale

$$F(w) = \Phi(w) = \int_{-\infty}^w [1/(2\pi)^{1/2}] \exp\{-(1/2)t^2\} dt$$

che “restituisce” il *modello probit*, e la funzione di distribuzione standard logistica, data da

$$F(w)=L(w)=e^w / (1+ e^w)$$

che porta al *modello logit*. Una terza scelta corrisponde alla distribuzione uniforme nell'intervallo [0,1] con funzione di distribuzione

$$F(w) = 0, w < 0;$$

$$F(w) = w, 0 \leq w \leq 1;$$

$$F(w) = 1, w > 1.$$

che porta al *modello lineare di probabilità*, simile al modello di regressione visto precedentemente in questo stesso paragrafo, ma con le probabilità a cui sono assegnati valori 0 ed 1 se $x_i' \beta$ eccede rispettivamente il più alto o il più basso limite. Di fatto i modelli probit e logit sono i più utilizzati nelle applicazioni pratiche; entrambi hanno variabile casuale con valore atteso pari a 0, mentre il secondo ha varianza pari a $\pi^2/3$ invece che ad 1.

A parte i segni, i coefficienti di questi modelli con scelta binaria non sono semplici da interpretare direttamente. Un modo per interpretare i parametri e per facilitare la comparazione tra modelli differenti, è guardare alla derivata della probabilità che y_i sia uguale ad 1, rispetto a k elementi di x_i . Per i tre modelli visti si ottiene:

$$d\Phi(x_i' \beta) / dx_{ik} = \phi(x_i' \beta) \beta_k;$$

$$dL(x_i' \beta) / dx_{ik} = [e^{x_i' \beta} / (1+ e^{x_i' \beta})^2] \beta_k;$$

$$d(x_i' \beta) / dx_{ik} = \beta_k; \text{ (oppure 0),}$$

con $\phi(\cdot)$ che indica la funzione di densità normale standard. Ad eccezione dell'ultimo modello, gli effetti di un cambiamento in x_{ik} dipendono dai valori di x_i ; tuttavia in tutti i casi il segno dell'effetto di un cambiamento in x_{ik} corrisponde al segno del corrispondente coefficiente β_k .

14.2 *La variabile latente sottostante*

E' possibile, ma non necessario, derivare un modello con scelta binaria da assunzioni comportamentali sottostanti; questo porta ad una modello con variabile latente.

Si pensi all'esempio di una donna sposata che deve decidere se lavorare o meno. L'utilità differente tra avere un lavoro e non averlo dipende dal salario, ma anche da alcune caratteristiche personali, come l'età e il livello d'istruzione, e familiari. Così per la persona i si può scrivere la differente utilità di avere un lavoro o meno come una funzione di caratteristiche osservate, x_i , e di caratteristiche non osservate, ε_i . Assumendo una relazione lineare additiva si ottiene la funzione di utilità

$$y_i^* = x_i' \beta + \varepsilon_i$$

Essendo y_i^* non osservata, è definita come variabile latente. Si assume che un individuo sceglie di lavorare se l'utilità eccede un certo livello di soglia, che può essere anche zero. Conseguentemente, si osserva $y_i = 1$ (la persona sceglie di lavorare) se e solo se $y_i^* > 0$, e $y_i = 0$ (sceglie di non lavorare) altrimenti. Così si ha:

$$P\{y_i = 1\} = P\{y_i^* > 0\} = P\{x_i' \beta + \varepsilon_i > 0\} = P\{-\varepsilon_i \leq x_i' \beta\} = F(x_i' \beta),$$

con F che indica la funzione di distribuzione di $-\varepsilon_i$, oppure, nel caso di distribuzione simmetrica, la funzione di distribuzione di ε_i . In questo modo si ottiene un modello a scelta binaria, la cui forma dipende dalla distribuzione che si è assunta per ε_i . E' richiesta una normalizzazione della distribuzione di ε_i ; di solito ciò significa che la sua varianza è fissata ad un dato valore. Se è scelta una distribuzione standard normale, si ottiene un modello probit, mentre per il logit è richiesta una distribuzione logistica.

Sebbene i modelli a scelta binaria in economia possano essere spesso interpretati come derivazioni di un problema di massimizzazione dell'utilità sottostante, ciò certamente non è sempre necessario. Solitamente si definisce la variabile latente y_i^* direttamente, tale che il modello probit sia completamente descritto da

$$y_i^* = x_i' \beta + \varepsilon_i, \quad \varepsilon_i \sim \text{NID}(0,1)$$

$$y_i = 1 \text{ se } y_i^* > 0$$

$$y_i = 0 \text{ se } y_i^* \leq 0,$$

con gli ε_i indipendenti da tutti gli x_i . Per il modello logit, la distribuzione logistica prende il posto di quella normale.

Più comunemente i parametri nei modelli con scelta binaria, e in generale quelli con variabile dipendente limitata, sono stimati col metodo della massima verosimiglianza. In generale il contributo della verosimiglianza per l'osservazione i con $y_i = 1$ è dato da $P\{y_i = 1 | x_i\}$ come una funzione del vettore del parametro ignoto β , e in modo simile per $y_i = 0$. La funzione di verosimiglianza per l'intero campione è data da

$$\log L(\beta) = \prod_{i=1}^N P\{y_i = 1 | x_i; \beta\}^{y_i} P\{y_i = 0 | x_i; \beta\}^{1-y_i},$$

con β incluso nelle espressioni per le probabilità in modo da sottolineare che la funzione di verosimiglianza è una funzione di β . Normalmente si preferisce lavorare con la funzione di log verosimiglianza. Sostituendo $P\{y_i = 1 | x_i; \beta\} = F(x_i' \beta)$ si ottiene

$$\log L(\beta) = \sum_{i=1}^N y_i \log F(x_i' \beta) + \sum_{i=1}^N (1-y_i) \log (1-F(x_i' \beta))$$

Sostituendo l'appropriata forma ad F si ha un'espressione che può essere massimizzata rispetto a β ; i valori di β e la loro interpretazione dipendono da quale funzione di distribuzione è scelta. Si considerino le condizioni del primo ordine della funzione di verosimiglianza; differenziando $\log L(\beta)$ rispetto a β si ottiene:

$$d \log L(\beta) / d\beta = \sum_{i=1}^N \{ (y_i - F(x_i' \beta)) (f(x_i' \beta)) / [(F(x_i' \beta))(1 - F(x_i' \beta))] \} x_i = 0$$

con $f = F'$ derivata della funzione di distribuzione. Il termine *fra parentesi graffe* è il residuo generalizzato del modello; è uguale a $f(x_i' \beta) / F(x_i' \beta)$ per le osservazioni positive ($y_i = 1$) e a $-f(x_i' \beta) / (1-F(x_i' \beta))$ per le osservazioni nulle ($y_i = 0$). Le condizioni del primo ordine dicono così che ogni variabile esplicativa deve essere ortogonale al residuo generalizzato (sul campione intero).

14.3 *I modelli con risposte multiple*

In molte applicazioni, il numero di alternative tra cui si può scegliere è maggiore di due.

Per esempio un individuo può decidere di lavorare a tempo pieno, part-time o di non lavorare. Alcune variabili quantitative possono essere osservate giacere in certi range. Questo può accadere perché chi risponde ai questionari può essere restio a dare risposte precise, o non in grado di farlo, magari a causa di difficoltà concettuali incontrate nel rispondere; ciò può ad esempio accadere quando si risponde a questionari sul reddito o sul valore di una casa. I modelli a risposta multipla si sono sviluppati per descrivere la probabilità di ciascuna delle possibili risposte/risultati come funzione di specifiche caratteristiche spesso di tipo personale. Un importante scopo è quello di descrivere queste probabilità con un numero limitato di parametri ignoti ed in modo logicamente consistente; per esempio, le probabilità devono essere tra 0 e 1 e, per tutte le alternative assieme, ammontare ad 1.

Un'importante distinzione esiste tra modelli ordinali e non ordinali. Un modello con risposte ordinate è generalmente più parsimonioso, ma può essere applicato solo se esiste un ordinamento logico tra le alternative. La ragione di ciò è che si assume esistere una variabile latente sottostante che indirizza la scelta tra le alternative; in altre parole, i risultati saranno sensitivi all'ordinamento delle alternative, così che l'ordinamento abbia senso. I modelli non ordinali non sono sensibili al modo in cui le alternative sono numerate; in molti casi, essi possono essere basati sull'assunzione che ogni alternativa abbia un livello di utilità casuale e che gli individui scelgano l'alternativa che dà loro la più alta utilità.

14.4 *I modelli con risposte multiple ordinali*

Si supponga di poter scegliere tra M alternative, numerate da 1 ad M. Può esserci un ordinamento logico tra le alternative; ad esempio l'ordinamento può essere del tipo “non si ha la macchina”, “si ha una macchina”, “si hanno più macchine”. In questi casi si parla di modello ordinale con risposte multiple, in cui vige un qualche ordinamento; questo modello si basa su di una variabile latente

sottostante, ma con una differente corrispondenza tra la variabile latente, y_i^* , e quella osservata ($y_i = 1, 2, \dots, M$). Di solito si ha che:

$$y_i^* = x_i' \beta + \varepsilon_i$$

$$y_i = j \text{ se } \gamma_{j-1} < y_i^* \leq \gamma_j,$$

per gli ignoti γ_j con $\gamma_0 = -\infty$ e $\gamma_M = \infty$. Conseguentemente, la probabilità che l'alternativa j sia scelta è pari alla probabilità che la variabile latente sia tra i limiti γ_{j-1} e γ_j .

Se si assume che ε_i sia i.i.d. normale standard, si ottiene il modello probit ordinale. La distribuzione logistica dà il modello logit ordinale. Per $M=2$ si rientra nel sottocaso del modello a scelta binaria.

Si consideri ora un esempio sull'offerta di lavoro. Si supponga che le donne sposate rispondano alla domanda “Quanto vorresti lavorare?” con tre possibili risposte “niente”, “part-time”, “a tempo pieno”. Come vuole la teoria neoclassica, l'offerta di lavoro desiderata, misurata da queste tre risposte, dipende dalle preferenze e dal vincolo di bilancio familiare, e quindi entrano in gioco variabili individuali e sulla composizione della famiglia. Per modellare i risultati un ordinamento logico nelle risposte è quello che vede assegnare ad y_i il valore 1 se l'individuo non vuole lavorare, 2 se vuole lavorare part-time e 3 se vuole lavorare a tempo pieno. Il punto cruciale è se sia ragionevole assumere che esista un singolo indice $x_i' \beta$ tale che i più alti valori di questo indice corrispondano, in media, ai più alti valori di y_i . Se ciò avviene, si può scrivere il modello a risposta ordinale come

$$y_i^* = x_i' \beta + \varepsilon_i$$

$$y_i = 1 \text{ se } y_i^* \leq \gamma_1,$$

$$y_i = 2 \text{ se } \gamma_1 < y_i^* \leq \gamma_2,$$

$$y_i = 3 \text{ se } y_i^* > \gamma_2,$$

con y_i^* che può essere interpretato, in modo sommario, come indicatore di “buona volontà di lavorare” e di “ore di lavoro desiderate” dall'individuo.

Occorre porre il vincolo di normalizzazione sulla misura di y_i^* ; il modo più naturale è fissare la varianza di ε_i . Nel modello probit ordinale ciò significa che ε_i è NID(0,1).

Le probabilità sono allora ottenute come:

$$P\{y_i = 1 | x_i\} = P\{y_i^* \leq \gamma_1 | x_i\} = \Phi(\gamma_1 - x_i' \beta),$$

$$P\{y_i = 3 | x_i\} = P\{y_i^* > \gamma_2 | x_i\} = 1 - \Phi(\gamma_2 - x_i' \beta)$$

$$P\{y_i = 2 | x_i\} = \Phi(\gamma_2 - x_i' \beta) - \Phi(\gamma_1 - x_i' \beta),$$

con γ_1 e γ_2 parametri ignoti che sono stimati assieme a β . La stima si basa sulla massima verosimiglianza, con le probabilità appena viste che entrano nella funzione di verosimiglianza. L'interpretazione dei coefficienti β è in termini della variabile latente sottostante; nell'esempio un coefficiente β positivo significa che la corrispondente variabile contribuisce ad incrementare la "buona volontà di lavorare" della donna. L'interpretazione dei coefficienti β può anche avvenire rispetto agli effetti delle rispettive probabilità. Ad esempio si supponga che il k-esimo coefficiente β_k sia positivo; questo significa che la variabile latente y_i^* cresce se x_{ik} cresce. Di conseguenza la probabilità che $y_i = 3$ aumenta, mentre la probabilità che $y_i = 1$ decresce; l'effetto sulle situazioni intermedie è tuttavia ambiguo, visto che la probabilità che $y_i = 2$ può aumentare come diminuire.

Per illustrare le differenti *normalizzazioni sui vincoli* richieste, si considera un modello dove i vincoli non sono imposti.

$$y_i^* = \beta_1 + x_i' \beta + \varepsilon_i, \quad \varepsilon_i \sim \text{NID}(0, \sigma^2).$$

$$y_i = 1 \text{ se } y_i^* \leq \gamma_1,$$

$$y_i = 2 \text{ se } \gamma_1 < y_i^* \leq \gamma_2,$$

$$y_i = 3 \text{ se } y_i^* > \gamma_2,$$

dove la costante è portata fuori dal vettore x_i . E' possibile osservare solo se y_i sia pari ad 1, a 2 o a 3, e i soli elementi che i dati possono identificare sono le probabilità dei tre eventi, per dati valori di x_i ; queste sono esattamente le

probabilità che entrano nella funzione di verosimiglianza. Si consideri la probabilità che $y_i = 1$ (dato x_i):

$$P\{y_i = 1 | x_i\} = P\{\beta_1 + x_i' \beta + \varepsilon_i \leq \gamma_i | x_i\} = \Phi[(\gamma_i - \beta_1)/\sigma - x_i' (\beta/\sigma)]$$

Ciò evidenzia come, variando β , β_1 , σ oppure γ_i , non si hanno differenti probabilità, purché β/σ e $(\gamma_i - \beta_1)/\sigma$ rimangano gli stessi. Questo riflette un problema di identificazione, perché differenti combinazioni di parametri portano ad uguali valori di log-verosimiglianza e non c'è un unico massimo. Per risolvere questo problema, sono imposti dei vincoli di normalizzazione. Il modello standard impone che $\sigma = 1$ e $\beta_1 = 0$. L'interpretazione dei coefficienti dipende da un particolare vincolo di normalizzazione, ma le probabilità non sono sensibili ad esso. In alcune applicazioni, i limiti corrispondono ai valori osservati piuttosto che ai parametri ignoti ed è possibile stimare la varianza di ε_i .

15 I dati longitudinali

Un dataset di dati panel (o longitudinali) contiene osservazioni ripetute sulle stesse unità, che possono essere individui, famiglie o aziende; le unità si riferiscono ad un certo numero di periodi. Sebbene i dati panel siano tipicamente utilizzati ad un livello microeconomico, è diventato sempre più frequente considerare le serie temporali individuali di un certo numero di paesi o industrie ed analizzarle simultaneamente.

La disponibilità di osservazioni ripetute sulle stesse unità porta gli economisti a specificare e stimare modelli più complessi e realistici di quanto non si possa fare con una singola sezione o con una singola serie temporale.

Vi sono però degli svantaggi di natura pratica. Siccome si osservano in modo ripetitivo alcune unità, non è più appropriato assumere che differenti osservazioni siano indipendenti tra di loro. Ciò può complicare l'analisi, specie se i modelli adottati sono non lineari e dinamici. Inoltre, i dati longitudinali spesso soffrono la presenza di osservazioni mancanti, e anche quando le osservazioni perse mancano in modo casuale (random) l'analisi standard necessita di essere aggiustata.

15.1 I vantaggi dei dati panel

Un importante vantaggio dei dati panel, rispetto alla singola sezione o alla singola serie temporale, è che essi consentono di identificare certi parametri senza il bisogno di fare assunzioni restrittive.

Per esempio, i dati panel rendono possibile analizzare i cambiamenti su di un livello individuale. Si consideri il caso in cui il livello di consumo medio cresca del 2% da un anno all'altro. I dati panel possono identificare se la crescita è il risultato, ad esempio, di un incremento del 2% per tutti gli individui o se piuttosto è la conseguenza di una crescita pari al 4% per la metà della popolazione, con i rimanenti individui non interessati da cambiamenti sul consumo. Detto ciò, i dati panel non sono solo adatti a modellare o spiegare perché le unità individuali si comportano diversamente, ma anche a spiegare perché una data unità si comporta in modi diversi in differenti periodi temporali. Si indicizzino tutte le variabili con i per l'individuo ($i=1, \dots, N$) e con t per il periodo temporale ($t=1, \dots, T$). In termini generali si può specificare un modello lineare come

$$y_{it} = x'_{it} \beta_{it} + \varepsilon_{it},$$

dove β_{it} misura gli effetti parziali di x_{it} sul periodo t per l'unità i . Questo modello è però troppo generale per essere usato, e occorre strutturare i coefficienti β_{it} . L'assunzione standard, utilizzata in molti casi empirici, è che β_{it} sia costante per tutti gli i e t , eccetto, se possibile, l'intercetta. Si ottiene

$$y_{it} = \alpha_i + x'_{it} \beta + \varepsilon_{it},$$

con x_{it} vettore k -dimensionale di variabili esplicative, esclusa la costante. Questo significa che gli effetti di un cambiamento in x sono gli stessi per tutte le unità e tutti i periodi, ma il livello medio per l'unità i può essere differente da quello per l'unità j . Gli α_i catturano gli effetti di quelle variabili che sono peculiari per l' i -esimo individuo e che sono costanti sul tempo. Ne caso standard, ε_{it} è assunto essere indipendente e identicamente distribuito sugli individui e sul tempo, con

media zero e varianza σ_ε^2 . Se si trattano gli α_i come N parametri ignoti fissati, il modello si dice *standard ad effetti fissi*.

Un approccio alternativo assume che le intercette degli individui siano diverse ma che possano essere trattate come componenti casuali con distribuzione con media μ e varianza σ_α^2 . L'assunzione essenziale qui è che queste componenti siano indipendenti dalle variabili esplicative in x_{it} . Si tratta del modello *ad effetti casuali*, con gli effetti individuali α_i trattati come random. Il termine d'errore in questo modello consiste di due componenti: una componente invariante rispetto al tempo, α_i , e una componente che è incorrelata rispetto al tempo, ε_{it} . Il modello ad effetti casuali può così essere scritto:

$$y_{it} = \mu + x_{it}'\beta + \alpha_i + \varepsilon_{it},$$

con μ termine d'intercetta. Per una spiegazione più completa de modello (probit) ad effetti casuali rimando al paragrafo 16.2, mentre rimando al paragrafo 15.4 per una breve illustrazione dei vantaggi che derivano dall'utilizzo degli effetti fissi piuttosto che casuali.

Nei due sotto-paragrafi successivi vediamo ulteriori vantaggi che derivano dall'utilizzo di dati panel.

15.2 L'efficienza degli stimatori dei parametri

Siccome i dataset con dati panel sono normalmente più grandi di quelli con singola sezione trasversale o singola serie temporale, e poiché le variabili esplicative variano su due dimensioni e non una, gli stimatori basati su dati panel sono spesso più accurati di quelli basati sulle altre fonti. Perfino con campioni di stessa ampiezza, l'uso di dataset con dati panel restituisce stimatori più efficienti di quanto non avvenga con serie di sezioni trasversali indipendenti, con unità differenti campionate in ogni periodo. Per vedere questo, si consideri un modello ad effetti casuali con incluse sole dummy:

$$y_{it} = \mu_t + \alpha_i + \varepsilon_{it},$$

con ciascun μ_t parametro ignoto corrispondente alla media della popolazione al periodo t . Si supponga di essere interessati non alla media in μ_t ad un dato periodo, bensì al cambiamento di μ_t da un periodo all'altro. In generale la varianza dello stimatore efficiente per $\mu_t - \mu_s$ ($s \neq t$), $\hat{\mu}_t - \hat{\mu}_s$, è data da

$$Var(\hat{\mu}_t - \hat{\mu}_s) = Var(\hat{\mu}_t) + Var(\hat{\mu}_s) - 2Cov(\hat{\mu}_t, \hat{\mu}_s)$$

con $\hat{\mu}_t = (1/N) \sum_{i=1}^N y_{it}$ ($t=1, \dots, T$). Di solito, se un dataset con dati panel è utilizzato, la covarianza tra $\hat{\mu}_t$ e $\hat{\mu}_s$ è positiva e, se l'assunzione sul modello ad effetti casuali tiene, è pari a σ_α^2 / N . Se invece vengono utilizzati due dataset indipendenti a sezione trasversale, differenti periodi contengono differenti individui, cosicché $\hat{\mu}_t$ e $\hat{\mu}_s$ hanno covarianza nulla. Quindi se si è interessati ai cambiamenti da un periodo all'altro, i dati panel restituiscono stimatori più efficienti di quanto non facciano le singole serie temporali.

15.3 Il problema dell'identificazione dei parametri

Un secondo vantaggio che deriva dall'utilizzo di dati panel è che riduce i problemi dell'identificazione dei parametri, quali la presenza di regressori endogeni e soggetti ad errori di misura, l'errore da omissione di variabili e l'identificazione delle dinamiche individuali nel tempo.

Vi sono due alternative spiegazioni per il fenomeno che vede gli individui che hanno avuto una certa esperienza nel passato avere buone probabilità di riviverla. La prima spiegazione è che il fatto che un individuo che abbia sperimentato un certo evento vede cambiare le sue preferenze, aumentando la probabilità che riviva l'evento in futuro. La seconda spiegazione sta nel fatto che gli individui possono essere diversi nelle caratteristiche non osservate che influenzano la probabilità di sperimentare un certo evento. Heckman chiama la prima spiegazione condizione di dipendenza vera, e la seconda condizione di dipendenza spuria. L'utilizzo di dati panel facilita la distinzione tra le due, perché le storie degli individui sono osservate e possono essere incluse nel modello.

L'errore di omissione di una variabile cresce se una variabile correlata ad altre incluse è esclusa dal modello. Un classico esempio è la stima di funzioni di

produzione. Specialmente nel caso di piccole aziende, è auspicabile includere la qualità del management come input nella funzione di produzione; la qualità del management è però generalmente non osservabile. Si supponga di avere la funzione di produzione del tipo

$$y_{it} = \mu + x_{it}'\beta + m_i\beta_{k+1} + \varepsilon_{it}$$

con y_{it} logaritmo dell'output, e x_{it} vettore dimensionale degli input in forma logaritmica, entrambi per l'azienda i al tempo t , e con m_i che denota la qualità manageriale per l'azienda i , che si assume essere costante nel tempo. Ci si aspetta che la variabile non osservata m_i sia correlata negativamente con gli altri input x_{it} , dato che una alta qualità del management risulterà probabilmente in un più efficiente utilizzo degli input. Perciò, a meno che $\beta_{k+1} = 0$, l'omissione di m_i porta ad errori di stima sugli altri parametri del modello. Se sono disponibili i dati panel questo problema può essere risolto introducendo uno specifico effetto dell'azienda $\alpha_i = \mu + m_i\beta_{k+1}$ e considerandolo come un parametro ignoto fissato. Si noti che senza l'informazione addizionale non è possibile identificare β_{k+1} .

In modo simile, un effetto fisso sul tempo può essere introdotto per catturare l'effetto di tutte le variabili, osservate e non, che non variano rispetto alle unità individuali. Così i dati panel riducono gli effetti degli errori da omissione di variabili; gli stimatori in dataset con dati panel guadagnano in robustezza in caso di specificazione di un modello incompleto.

Infine, in molti casi, i dati panel forniscono strumenti interni per i regressori che sono endogeni o soggetti ad errori di misura. Infatti, le trasformazioni delle variabili originali sono spesso incorrelate con i termini d'errore del modello e correlate con le stesse variabili esplicative e quindi non c'è bisogno di variabili strumentali esterne. Per esempio, se x_{it} è correlato con α_i , si può sostenere che $x_{it} - E(x_i)$, con $E(x_i)$ valore medio rispetto al tempo per l'individuo i , è incorrelato con α_i e fornisce un valido strumento per x_{it} .

15.4 Effetti fissi o casuali ?

Se trattare gli effetti individuali α_i come fissi o casuali è questione a cui non è facile dare una risposta. L'interpretazione più comune è che l'approccio con effetti fissi è condizionale rispetto ai valori di α_i ; si considera la distribuzione di y_{it} dato α_i , con gli α_i che possono essere stimati. Ciò ha intuitivamente senso se tutti gli individui nel campione sono dello stesso genere, e non possono essere visti come se fossero stati estratti casualmente da una qualche popolazione sottostante. Questa interpretazione è probabilmente più appropriata quando gli i denotano paesi, grandi compagnie o industrie, e la predizione che si vuole fare è per un particolare paese, compagnia o industria. L'inferenza è fatta rispetto agli effetti che vi sono nel campione.

Al contrario, l'approccio con effetti casuali non è condizionale agli α_i individuali. In tal caso non si è interessati ad un particolare valore di un qualche α_i individuale; si pone l'attenzione sugli individui arbitrari che hanno certe caratteristiche. L'approccio porta a fare inferenza rispetto alle caratteristiche della popolazione.

15.5 I modelli con scelta binaria e dati panel

Il modello con scelta binaria è usualmente formulato in termini di variabile latente sottostante, cioè

$$y_{it}^* = x_{it}'\beta + \alpha_i + \varepsilon_{it},$$

e si osserva $y_{it} = 1$ se $y_{it}^* > 0$, e $y_{it} = 0$ altrimenti.

Se si trattano gli α_i come parametri fissati ignoti in sostanza si stanno includendo N variabili dummy nel modello. La funzione di verosimiglianza è data da

$$\text{Log } L(\beta, \alpha_1, \dots, \alpha_N) = \sum_{i,t} y_{it} \log F(x_{it}'\beta + \alpha_i) + \sum_{i,t} (1 - y_{it}) \log [1 - F(x_{it}'\beta + \alpha_i)]$$

Massimizzando la log verosimiglianza rispetto a β e ad α_i ($i=1, \dots, N$), risultano stimatori consistenti a condizione che il numero di periodi temporali T vada ad infinito. Per T fissato e $N \rightarrow \infty$, gli stimatori sono inconsistenti. La ragione di ciò è che per T fissato, il numero di parametri cresce con l'ampiezza del campione

N ; ciò è noto come *problema incidentale sui parametri*. Così ogni α_i può essere stimato consistentemente solo se c'è un crescente numero di osservazioni per l'individuo i , con T che tende all'infinito.

Questa situazione si verifica in tutti i tipi di modelli ad effetti fissi, anche quelli lineari; per quest'ultimi però l'inconsistenza dello stimatore di α_i per T fissato non porta necessariamente all'inconsistenza dello stimatore per β , come avviene solitamente per i modelli non lineari.

Per superare il problema incidentale sui parametri, si può utilizzare la massima verosimiglianza condizionale; si considera la funzione di verosimiglianza condizionale su di un set di statistiche t_i che siano sufficienti per α_i . Questo significa che su t_i il contributo di verosimiglianza di un individuo non dipende più da α_i , ma dipende ancora dagli altri parametri β . Infatti, a livello generale, si può scrivere la densità congiunta di y_{i1}, \dots, y_{iT} come $f(y_{i1}, \dots, y_{iT} | \alpha_i, \beta)$, che dipende dai parametri α_i e β ; se esiste una statistica sufficiente t_i , questo significa che esiste una statistica t_i tale che $f(y_{i1}, \dots, y_{iT} | t_i, \alpha_i, \beta) = f(y_{i1}, \dots, y_{iT} | t_i, \beta)$, e quindi non c'è dipendenza da α_i . Conseguentemente si può massimizzare la funzione di verosimiglianza condizionale, basata su $f(y_{i1}, \dots, y_{iT} | t_i, \beta)$, per ottenere uno stimatore consistente per β . Ad esempio per il modello lineare con errori normali una statistica sufficiente per α_i è il valore medio rispetto al tempo di y_{it} .

Nei modelli con dati panel a scelta binaria, l'esistenza di una statistica sufficiente dipende dalla forma della funzione di distribuzione, e quindi dalla distribuzione di ε_{it} . In particolare per il modello probit, che ha distribuzione normale standard, non esiste una statistica sufficiente per α_i . Questo significa che non può essere stimato in modo consistente per T fissato un modello probit ad effetti fissi.

15.6 Il motivo dell'interesse sui dati di panel rispetto all'analisi empirica

Il capitolo si sofferma a lungo sui dati panel; il motivo di ciò sta nel fatto che l'analisi empirica trattata in questa tesi vuole studiare come un dato effetto influenzi l'eterogeneità della variabile latente, e per fare ciò ci si avvale del

supporto di un modello strutturato come quello con dati panel ad effetti casuali; questo ha la seguente forma:

$$y_{it}^* = x_{it}'\beta + v_{it} \quad , \quad i = 1, \dots, N \text{ e } t = 1, \dots, T$$

$$v_{it} = \alpha_i + u_{it}$$

con un termine d'errore v_{it} a due componenti, α_i e u_{it} ; α_i denota lo specifico effetto non osservabile individuale, mentre u_{it} è un errore casuale. Inoltre, nel caso del modello probit ad effetti casuali, si assume che $u_{it} \sim \text{IN}(0, \sigma_u^2)$ e che, condizionatamente agli x_{it} , gli α_i siano $\text{IN}(0, \sigma_\alpha^2)$ e indipendenti dagli u_{it} e dagli x_{it} . Questo porta la correlazione tra due (temporalmente) successivi termini d'errore, per uno stesso individuo i -esimo, ad essere data da

$$\rho = \text{corr}(v_{it}, v_{it-1}) = \sigma_\alpha^2 / (\sigma_\alpha^2 + \sigma_u^2).$$

Fin qui si è visto come i dati panel servano ad evidenziare quanto un certo effetto influenzi la variabile latente, con α_i effetto non osservabile individuale con i fissato che sta per individuo i -esimo; si ha anche una componente temporale caratterizzante i dati di panel, con t tempo che varia da 1 a T .

L'analisi empirica rielabora quella che è la struttura appena vista, adattandola alle proprie esigenze. Infatti l'analisi assume come componente fissata la coppia, mentre a variare, questa volta senza un ordinamento come quello temporale, è l'individuo i , che può essere marito o moglie. Il modello sopra visto può quindi essere così riscritto:

$$y_{ci}^* = x_{ci}'\beta + v_{ci} \quad , \quad i = 1, \dots, N \text{ e } t = 1, \dots, T$$

$$v_{ci} = \alpha_c + u_{ci}$$

Le assunzioni su α_c e u_{ci} rimangono quelle precedentemente viste per α_i e u_{it} , ma ora α_c sta per effetto di coppia con c fissato, mentre i varia a seconda che l'individuo sia maschio piuttosto che femmina. Così la correlazione $\rho = \text{corr}(v_{c1}, v_{c2})$ viene ad indicare quanta parte dell'eterogeneità della variabile latente è spiegata dall'effetto coppia.

16 *Il modello probit con dati panel*

Nel paragrafo 14 si sono introdotti i modelli con variabile dipendente discreta, binaria e con risposte multiple ordinate o meno, e si è visto come uno dei modelli che rientrano in questa categoria sia il modello probit. Nel paragrafo 15 si parla dei modelli con dati longitudinali, soffermandosi in particolare su quali siano i vantaggi che derivano dall'utilizzo di dati panel.

Ora vediamo più nel dettaglio il modello probit, con dati panel; si illustrano il probit standard e il probit standard ad effetti casuali, ci si sofferma brevemente sul probit ordinale, e si introduce il modello probit standard bivariato.

16.1 *Il modello probit standard*

Il modello probit standard si basa su di una struttura di regressione latente con

$$y_{it}^* = x_{it}'\beta + u_{it}, \quad u_{it} \sim \text{NID}(0,1)$$

con

$$y_{it} = 1 \text{ se } y_{it}^* > 0 \quad \text{e} \quad y_{it} = 0 \text{ se } y_{it}^* \leq 0,$$

Gli u_{it} sono indipendenti da tutti gli x_{it} , con $i = 1, \dots, N$ e $t = 1, \dots, T$; si hanno y^* variabile non osservabile, y variabile osservata, x vettore osservabile delle caratteristiche esogene che influenzano y^* , e β vettore dei coefficienti associati ad x .

Si tratta di un modello con scelta binaria e distribuzione standard normale.

16.2 *Il modello probit standard ad effetti casuali*

I modelli probit ad effetti casuali sono, a livello empirico, sempre più utilizzati.

Si tratta di modelli strutturati come segue:

$$y_{it}^* = x_{it}'\beta + v_{it}, \quad i = 1, \dots, N \text{ e } t = 1, \dots, T$$

$$v_{it} = \alpha_i + u_{it}$$

con $y_{it} = 1$ se $y_{it}^* > 0$, $y_{it} = 0$ altrimenti

Oltre a y^* variabile non osservabile, y variabile osservata, x vettore osservabile delle caratteristiche strettamente esogene che influenzano y^* , e β vettore dei coefficienti associati ad x , si ha un termine d'errore v_{it} a due componenti, α_i e u_{it} ; α_i denota lo specifico effetto non osservabile individuale, mentre u_{it} è un errore casuale. Inoltre, nel caso del modello probit ad effetti casuali, si assume che $u_{it} \sim \text{IN}(0, \sigma_u^2)$. Allo scopo di marginalizzare la verosimiglianza si assume che, condizionatamente agli x_{it} , gli α_i siano $\text{IN}(0, \sigma_\alpha^2)$ e indipendenti dagli u_{it} e dagli x_{it} . Questo implica che la correlazione tra due successivi termini d'errore per lo stesso individuo è data da

$$\rho = \text{corr}(v_{it}, v_{it-1}) = \sigma_\alpha^2 / (\sigma_\alpha^2 + \sigma_u^2),$$

I parametri del modello sono facilmente stimabili, visto che la distribuzione degli y_{it}^* condizionatamente agli α_i è normale indipendente. Si ha

$$P(y_{it}=1 \mid \alpha_i, x_{it}) = P[(u_{it}/\sigma_u) > (-x_{it}'\beta - \alpha_i)/\sigma_u] = \Phi(z_{it})$$

$$\text{con } z_{it} = (-x_{it}'\beta - \alpha_i) / \sigma_u$$

e Φ funzione di distribuzione normale standard. Quindi si marginalizza l'appropriata funzione di verosimiglianza rispetto ad α , che è data da

$$\prod_{i=1}^N \left(\int_{-\infty}^{+\infty} \prod_{t=1}^T \left(1 - \Phi \left(x_{it}'\beta^* + \left(\frac{\rho}{1-\rho} \right)^{1/2} \alpha^* \right) \right)^{1-y_{it}} \times \Phi \left(x_{it}'\beta^* + \left(\frac{\rho}{1-\rho} \right)^{1/2} \alpha^* \right)^{y_{it}} \phi(\alpha^*) d\alpha^* \right)$$

con $\beta^* = \beta/\sigma_u$ e $\alpha^* = \alpha/\sigma_\alpha$. Alcuni programmi di software come Stata restituiscono le stime di β^* e ρ . Si noti come i coefficienti stimati qui siano normalizzati su σ_u .

Nel modello è possibile ottenere una stima dei parametri consistente dei coefficienti β ignorando la struttura di correlazione; così un semplice modello probit standard fornisce stime consistenti dei parametri. Però, a causa della natura

binaria dei dati, il modello probit semplice non porta a stimare il parametro di scala. Si ottengono solo le stime parametriche di β/σ_v . Quindi le stime parametriche del modello probit standard semplice sono uguali a quelle del modello probit ad effetti casuali solo quando $\sigma_\alpha^2 = 0$. Invece se $\sigma_\alpha^2 \neq 0$ le stime consistenti del probit ordinario non sono simili alle stime del probit ad effetti casuali a causa della normalizzazione.

Per convertire i coefficienti del probit ad effetti casuali da β/σ_u a β/σ_v , i coefficienti stimati devono essere moltiplicati per il fattore $\sqrt{1-\hat{\rho}}$. Ciò consente di fare validi raffronti tra le stime dei modelli probit univariato e ad effetti casuali, ma anche attraverso le stime ottenute da una differente specificazione del vettore x_{it} .

Occorre anche interpretare i vari effetti stimati in questi modelli con specifica componente non osservabile individuale. Nel probit univariato ci si aspetta che i cambiamenti che avvengano sui risultati dipendano da cambiamenti sulle caratteristiche esogene nel tempo, noti come effetti marginali.

16.3 Gli effetti marginali

Si consideri l'effetto medio di un cambiamento su x_j , con $j=1, \dots, k$, di piccola entità, sul risultato probabilistico. Nel probit standard, sotto la condizione di normalizzazione $\sigma_v = 1$, l'effetto è dato da

$$d [\text{Prob}(y_{it} = 1 | x_{it})] / d x_{jit} = d [E(y_{it})] / d x_{jit} = d [\Phi(x_{it}'\beta)] / d x_{jit} = \phi(x_{it}'\beta)\beta_j$$

Come questo effetto varia con i valori di x , lo si può valutare rispetto alle medie dei regressori oppure lo si può valutare separatamente per ciascun individuo nel campione per poi fare la media sul campione stesso.

L'associata matrice di covarianza può essere calcolata col Metodo Delta, vale a dire utilizzando le serie di Taylor di primo ordine per calcolare la matrice di covarianza nel caso di funzioni non lineari di variabili casuali. Per il modello probit standard si ottiene

$$\text{Var. Asimm.}[\phi(x'\hat{\beta})\hat{\beta}] = \text{Var. Asimm.}[\hat{y}] = [d\hat{y} / d\hat{\beta}]V[d\hat{y} / d\hat{\beta}]'$$

Con $V = \text{var. asimmm.}[\hat{\beta}]$ e gli effetti marginali $\hat{\gamma}$ valutati rispetto alle medie delle variabili; perciò $\phi(x' \hat{\beta}) = \hat{\phi}$ è uno scalare. La matrice delle derivate $[d\hat{\gamma} / d\hat{\beta}]$ è pari a $\hat{\phi}I - \phi(\hat{\beta}x')$, con I matrice d'identità k-dimensionale.

Nel caso del modello probit ad effetti casuali, invece, si prende in considerazione la normalizzazione, e ciò fa sì che, nel calcolo degli effetti marginali, i coefficienti stimati debbano essere moltiplicati per $\sqrt{1-\hat{\rho}}$. Allora si ottiene

$$d [\text{Prob} (y_{it} = 1 | x_{it})] / d x_{jit} = \phi(x'_{it} \beta) \beta_j = \phi[x'_{it} \beta^* \sqrt{1-\hat{\rho}}] [\sqrt{1-\hat{\rho}} \beta_j^*]$$

Il software restituisce le stime di β^* e ρ , e la loro matrice di covarianza Ω ; ma

$$\beta^* = \sqrt{1-\hat{\rho}} \beta^* \text{ implica che}$$

$$\text{Cov}(\hat{\beta}) = A\Omega A'$$

$$\text{con } A = \begin{bmatrix} \sqrt{1-\hat{\rho}} & 0 & \cdot & \cdot & 0 & -\hat{\beta}_1 / 2 \sqrt{1-\hat{\rho}} \\ 0 & \sqrt{1-\hat{\rho}} & 0 & \cdot & \cdot & -\hat{\beta}_2 / 2 \sqrt{1-\hat{\rho}} \\ \cdot & 0 & \cdot & \cdot & 0 & \cdot \\ 0 & \cdot & \cdot & 0 & \sqrt{1-\hat{\rho}} & -\hat{\beta}_k / 2 \sqrt{1-\hat{\rho}} \end{bmatrix}$$

matrice con k righe e k+1 colonne che contiene nella sua j-esima riga la derivata di $\hat{\beta}_j$ rispetto a tutti i $\hat{\beta}_j^*$ e a $\hat{\rho}$. Per calcolare la matrice di covarianza per gli effetti marginali propri del modello probit ad effetti casuali, si può ancora utilizzare l'equazione

$$\text{Var. Asimm.}[\phi(x' \hat{\beta}) \hat{\beta}] = \text{Var. Asimm.} [\hat{\gamma}] = [d\hat{\gamma} / d\hat{\beta}] V [d\hat{\gamma} / d\hat{\beta}]',$$

ma questa volta con V pari a $A\Omega A'$.

16.4 Il modello probit ordinale

Il modello probit ordinale è un modello con distribuzione normale standard che, a differenza di quello ordinario, ha “risposte” multiple che sono interessate da un ordinamento; ma visto che su queste caratteristiche ci si sofferma già

ampiamente nel paragrafo 14, vediamo ora un esempio in cui le “risposte” possibili e soggette a ordinamento sono tre; il modello si basa su di una struttura di regressione latente con

$$y_{it}^* = \beta' x_{it} + \varepsilon_{it}$$

Sebbene y_{it}^* sia non osservabile, si osserva y_{it} tale che

$$y_{it} = 1 \text{ se } y_{it}^* \leq \mu_1$$

$$y_{it} = 2 \text{ se } \mu_1 < y_{it}^* \leq \mu_2$$

$$y_{it} = 3 \text{ se } \mu_2 < y_{it}^*$$

con μ e β parametri ignoti da stimare. Dato che ε_i è normalmente distribuito attraverso osservazioni con media zero e varianza pari ad uno, si ottengono le seguenti probabilità

$$P(y_{it} = 1) = \Phi(\mu_1 - \beta' x_{it})$$

$$P(y_{it} = 2) = \Phi(\mu_2 - \beta' x_{it}) - \Phi(\mu_1 - \beta' x_{it})$$

$$P(y_{it} = 3) = 1 - P(y_{it} = 1) - P(y_{it} = 2)$$

con $\Phi(\cdot)$ che denota una distribuzione cumulativa normale standard.

Il vettore x_i indica il set di variabili esplicative che influenzano la variabile latente non osservabile y^* .

Anche il modello probit ordinale può essere ad effetti casuali, con la stessa struttura di equicorrelazione vista nel paragrafo 16.2.

16.5 Il modello probit standard bivariato

Si tratta di un modello probit ordinario a due equazioni; vi sono due variabili dipendenti latenti e si spiega l'influenza che hanno sulle due, prese distintamente, le variabili esogene. Il logaritmo della verosimiglianza dei modelli probit a due equazioni è

$$L = \sum_{i=1}^N \ln \Phi_2(q_{1i} \xi_i^\beta, q_{2i} \xi_i^\gamma, \rho_i^*)$$

con

$$\xi_i^\beta = x_i \beta$$

$$\xi_i^\gamma = x_i \gamma$$

$$q_{1i} = 1 \text{ se } y_{1i} \neq 0$$

$$= -1 \text{ altrimenti}$$

$$q_{2i} = 1 \text{ se } y_{2i} \neq 0$$

$$= -1 \text{ altrimenti}$$

$$\rho_i^* = q_{1i} q_{2i} \rho$$

e con $\Phi_2(\cdot)$ funzione di distribuzione normale bivariata cumulativa. La derivata assume che

$$y_{1i}^* = x_i \beta + \epsilon_{1i}$$

$$y_{2i}^* = z_i \gamma + \epsilon_{2i}$$

$$E(\epsilon_1) = E(\epsilon_2) = 0$$

$$\text{Var}(\epsilon_1) = \text{Var}(\epsilon_2) = 1$$

$$\text{Cov}(\epsilon_1, \epsilon_2) = \rho$$

con y_{1i}^* e y_{2i}^* variabili latenti non osservate; invece si osserva solo $y_{ij} = 1$ se $y_{ij}^* > 0$ e $y_{ij} = 0$ altrimenti (per $j = 1, 2$).

Nella stima di massima verosimiglianza ρ non è direttamente stimato. E' invece direttamente stimato $\text{atanh } \rho$:

$$\text{atanh } \rho = (1/2) \ln [(1+\rho) / (1-\rho)]$$

Se $\rho = 0$ allora la log verosimiglianza per i modelli probit bivariati è uguale alla somma delle log verosimiglianze dei due modelli probit univariati. Un test del rapporto di verosimiglianza può allora permettere di comparare la verosimiglianza del modello bivariato completo con la somma delle log verosimiglianze dei modelli probit univariati.

Ai fini dell'analisi empirica il modello bivariato è utile perché permette di studiare l'influenza che le variabili esogene hanno sugli individui presi in considerazione e "parallelamente" sui loro partner, a livello di scelta occupazionale e quindi di incidenza sulla propensione individuale al rischio. Infatti l'analisi empirica della tesi utilizza, nel contesto del modello bivariato,

due variabili dipendenti; una si riferisce allo status occupazionale di un certo individuo appartenente ad una coppia di coniugi/conviventi lavoratori, l'altra dà la stessa informazione ma riferita al partner della persona. Il modello restituisce un coefficiente ρ che dà una stima della correlazione tra i termini d'errore delle due variabili latenti, e che quindi fornisce il grado di correlazione interna alla coppia rispetto all'eterogeneità delle variabili latenti, cioè le avversioni individuali al rischio dei partner; tale informazione è alquanto utile se si considera che l'obiettivo dell'analisi empirica è proprio quello di indagare la correlazione intra-coppia rispetto allo status occupazionale a livello di grado di rischio associato.

CAPITOLO 5 La lettura dei risultati dell'analisi empirica

17 *L'impostazione dell'indagine*

L'analisi empirica si avvale dei modelli probit visti nel capitolo precedente. Questi restituiscono anche una stima dei coefficienti β / σ , che danno il segno dell'influenza che le singole variabili esplicative x_{ci} hanno sulla "scelta" del tipo di lavoro da parte degli individui, ammesso che i coefficienti risultino statisticamente significativi; la misura di tale influenza è invece data dagli effetti marginali. Si ricorda che il sotto-campione è formato dagli individui i appartenenti alle 1569 coppie c che vedono entrambi i partner occupati in una qualche attività lavorativa.

Dato l'obiettivo dell'analisi di studiare la presenza di una diversificazione del portafoglio dei tipi di lavoro all'interno della famiglia, l'attenzione è soprattutto rivolta ad indagare la correlazione tra lo status lavorativo di mariti e mogli (o conviventi); a tal scopo l'indagine empirica si avvale in particolare del supporto dei modelli ad effetti casuali e bivariati, che restituiscono l'informazione circa quanta parte della varianza della variabile latente sulla propensione al rischio dei singoli individui sia spiegata dall'effetto di coppia.

I risultati ed una loro lettura sono riportati nei due successivi paragrafi, a seconda che la scelta osservata sui tipi di lavoro con diverso grado di rischio da parte degli individui sia tra le modalità lavoro autonomo e dipendente, o se invece si tenga conto anche dell'ulteriore distinzione tra lavoro dipendente nel pubblico e nel privato; in tal caso si adotta, come già visto nel capitolo 3, un determinato ordinamento tra le tre opzioni a seconda del grado di rischio che si presume sia loro associato.

E' anche interessante valutare come le variabili esplicative cambino la loro influenza sulla scelta del tipo di lavoro a seconda del sotto-campione considerato; si distingue tra sotto-campione dei soli mariti/conviventi, delle sole

mogli/conviventi, e quello comprendente entrambi, sempre includendo gli individui appartenenti a coppie in cui entrambi i partner lavorano. Inoltre, nel caso di scelta binaria, si analizza tutta una serie di risultati che riguardano l'influenza che le caratteristiche individuali dei coniugi/conviventi hanno sui partner rispetto alla scelta occupazionale a livello di rischio.

Nel penultimo sotto-paragrafo, prima di trarre le conclusioni, ci si sofferma sul caso in cui la moglie/convivente non lavora, estendendo quindi il sotto-campione di riferimento; questo allo scopo di vedere se e come la "scelta" occupazionale dei partner di sesso maschile sia influenzata ed influenzata la condizione di lavoratrice o meno della donna.

Infine si ricorda che è possibile consultare il paragrafo 13 per vedere come sono costruite e cosa indicano le variabili dipendenti e indipendenti utilizzate.

18 *La scelta binaria: risultati e commenti*

Vediamo ora cosa risulta dai modelli adottati qualora la scelta osservata sul tipo di lavoro avvenga tra le due grandi categorie occupazionali del lavoro autonomo e di quello dipendente.

18.1 *Confronto fra sotto-campioni nell'univariato*

Utilizzando il modello probit standard, vediamo come cambiano gli effetti marginali a seconda del sotto-campione utilizzato. Oltre a quello comprendente entrambi i coniugi lavoratori di una coppia, si considerano i sotto-campioni che includono i soli mariti/conviventi e quindi le sole mogli/conviventi, appartenenti a coppie in cui entrambi i partner lavorano.

Sono riportati i coefficienti stimati delle variabili esogene prese in considerazione, il loro standard error, gli effetti marginali, e lo standard error degli effetti marginali. I risultati sono estesi in modo da facilitare un confronto tra i coefficienti e gli effetti marginali nei tre casi presi in considerazione. Si noti che tra le variabili esplicative non compare MOGLIE, che dà l'informazione sul

sesto del partner, essendo questa di fatto già data quando nei sotto-campioni si includono i soli maschi piuttosto che le sole femmine.

MODELLO PROBIT semplice: $y_{ci}^* = \beta'x_{ci} + \varepsilon_{ci}$, c coppia e i individuo marito/moglie (per il sotto-campione “entrambi i partner”), di sesso maschile (per il sotto-campione “mariti/conv.”) o di sesso femminile (per il sotto-campione “mogli/conv.”), con $\varepsilon_{ci} \sim NID(0,1)$ e con $y_{ci} = 1$ se $y_{ci}^* > 0$, $y_{ci} = 0$ altrimenti. y_{ci} dà il tipo di categoria occupazionale dell’individuo:

$y_{ci} = 0$ se lavoratore dipendente

$y_{ci} = 1$ se lavoratore autonomo

I RISULTATI:

• modelli probit standard

lavoro	ENTRAMBI I PARTNER		SOLI MARITI/CONV.		SOLE MOGLI/CONV.	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
eta	-.0247173	.0309989	-.0431517	.0456629	-.0414931	.0486397
etasq	.0004759	.0003556	.0005943	.0005084	.0005711	.0005824
ncomp	-.0252189	.0331913	-.0138142	.0456927	-.0018625	.0507229
npred	.0035530	.0534576	-.0175022	.0739210	.0702464	.0802709
presc	.1297595	.0585208	.1290293	.0787854	.0105928	.0927382
scsup	-.2884310	.0598361	-.1446566	.0805990	-.4292869	.0925820
univ	-.2184430	.0786578	-.0699294	.1065995	-.3302905	.1213563
partnerl.	1.1248310	.0582398	1.2825840	.0893422	1.2053150	.0836168
amp40	.0101739	.0611162	.0473455	.0830679	-.0221459	.0933696
amp500	-.2380135	.1220640	-.2803719	.1628785	-.1504070	.1894680
areaN	-.0731408	.0658244	.0166809	.0898791	-.1778281	.0999567
areaS	-.0737072	.0797215	-.0257627	.1092170	-.1253002	.1206142
_cons	.6581404	.6421672	.1162645	.9710436	.3056070	.9618873
	Eff. Marg.	S.E.	Eff. Marg.	S.E.	Eff. Marg.	S.E.
eta	-.0068050	.0085360	-.0136453	.0144406	-.0090655	.0106294
etasq	.0001310	.0000979	.0001879	.0001608	.0001248	.0001273
ncomp	-.0069430	.0091368	-.0043683	.0144481	-.0004069	.0110820
npred	.0009782	.0147174	-.0055345	.0233756	.0153476	.0175334
presc	.0357243	.0161019	.0408014	.0249052	.0023143	.0202622
scsup	-.0784558	.0160382	-.0453976	.0250891	-.0931431	.0198379
univ	-.0564332	.0189977	-.0217620	.0326376	-.0640561	.0207678
partnerl.	.3727340	.0206065	.4644075	.0314343	.3379484	.0259672
amp40	.0027973	.0167820	.0148986	.0260091	-.0048560	.0205469
amp500	-.0597374	.0276304	-.0810598	.0424790	-.0305803	.0356758
areaN	-.0201866	.0182091	.0052724	.0283947	-.0391690	.0221793
areaS	-.0199377	.0211792	-.0081072	.0342017	-.0263539	.0243920

Data la struttura della variabile dipendente che assume valori 0 ed 1, con a quest'ultimo associato un livello di rischio superiore (lavoro autonomo), è evidente come, se significativamente diversi da zero, i coefficienti, se positivi (negativi), indicano che le caratteristiche indagate influiscono positivamente (negativamente) sulla propensione al rischio degli individui; la stima degli effetti marginali dà una misura a livello probabilistico dell'influenza delle varie variabili esogene sulla scelta occupazionale

I risultati evidenziano innanzitutto come il coefficiente di PARTNERLAVORO, la variabile che dice qual'è la categoria occupazionale (con associato un certo grado di rischio) del partner, sia altamente statisticamente significativo e positivo in tutti e tre i casi; in particolar modo per i mariti/conviventi l'“essere assieme” ad una partner lavoratrice autonoma porta ad un incremento di circa 46 punti percentuali della probabilità di essere a loro volta autonomi. Invece per le persone di sesso femminile l'influenza positiva dell'essere lavoratore autonomo da parte del partner è minore, anche se pur sempre positiva; infatti la stima degli effetti marginali restituisce, per il sotto-campione delle mogli/conviventi, un coefficiente per PARTNERLAVORO pari a 0.3379484, vale a dire che, se “sono assieme” ad un lavoratore autonomo, le donne hanno una maggiore probabilità di scegliere un lavoro “poco sicuro” pari ai 33 punti circa, al contrario del caso in cui siano sposate/conviventi con lavoratori dipendenti. Per il sotto-campione che comprende entrambi i partner la probabilità di essere autonomi, se si è coniugati/conviventi con un lavoratore autonomo, aumenta di oltre i 37 punti percentuali; si può quindi dedurre come la scelta sul tipo di lavoro, in base al rischio, che riguarda un dato individuo, dipenda e sia positivamente correlata a quella che fa il coniuge/convivente.

Per il resto risultano significativi, al livello 5%, i coefficienti sul livello d'istruzione che dicono come il diploma di scuola superiore (senza laurea) e la laurea incidano negativamente sulla propensione al rischio dei singoli individui, soprattutto ed in misura maggiore se l'individuo in questione è di sesso femminile. Infatti per una donna il possesso di un diploma di scuola superiore (o professionale) o della laurea (o diploma universitario) decrementa la probabilità di scegliere il lavoro autonomo di circa 9 e 6 punti rispettivamente per i due

titoli di studio; per gli uomini l'incidenza negativa scende rispettivamente ai 5 e ai 2 punti percentuali, ma soprattutto i coefficienti non risultano statisticamente significativi. E' inoltre interessante notare come quando un individuo ha studiato oltre ciò che prevede la scuola dell'obbligo, la probabilità di scegliere un'occupazione rischiosa si cala, ma in caso di laurea o diploma universitario la misura dell'influenza negativa decresce rispetto al possesso del solo diploma di scuola superiore o professionale.

Per verificare se i vettori dei coefficienti dei due sotto-campioni "mariti/conv." e "mogli/conv." siano uguali o meno, si può utilizzare un test del rapporto di verosimiglianza:

$$2 \cdot |l_1 - l_0| = 2 \cdot |-1419.21 - (-785.10 - 591.95)| = 2 \cdot |-42.16| = 84.32 > \chi^2(13) = 22.36$$

con l_1 log-verosimiglianza per il sotto-campione "entrambi i partner", e l_0 somma delle log-verosimiglianze per i sotto-campioni che includono gli individui di sesso maschile e femminile presi singolarmente.

Il risultato vede rifiutare l'ip. nulla "uguaglianza dei vettori dei coefficienti dei due sotto-campioni".

18.2 L'output dei modelli probit standard

Si utilizzano i modelli probit standard semplice, bivariato e ad effetti casuali, e si riportano i coefficienti stimati delle variabili esogene prese in considerazione, il loro standard error, gli effetti marginali, e lo standard error degli effetti marginali.

Si noti che tra le variabili esplicative non compare, nei modelli bivariato e ad effetti casuali, l'esogena PARTNERLAVORO, che abbiamo visto dare il tipo di lavoro del partner dell'individuo. Infatti non ha senso tenerne conto in quanto entrambi i modelli riorganizzano i dati in un panel per cui ogni membro della coppia rappresenta un'osservazione ripetuta per la coppia stessa. La correlazione con il grado di rischio associato al lavoro del partner è data da ρ , strutturato come visto nel capitolo 4. Anche tra le variabili esplicative del probit standard semplice non compare l'informazione sul lavoro del partner, in modo da rendere gli effetti

marginali confrontabili con quelli degli altri due modelli, ragion per cui il set di variabili è lo stesso.

SOTTO-CAMPIONE: membri delle coppie ove entrambi i partner lavorano

I modelli, spiegati nel dettaglio nel capitolo 4 e adattati all'analisi empirica della tesi, sono sinteticamente presentati come segue:

MODELLO PROBIT semplice: $y_{ci}^* = \beta' x_{ci} + \varepsilon_{ci}$, c coppia e i individuo marito/moglie (o convivente), con $\varepsilon_{ci} \sim \text{NID}(0,1)$ e con $y_{ci} = 1$ se $y_{ci}^* > 0$, $y_{ci} = 0$ altrimenti. y_{ci} dà il tipo di categoria occupazionale dell'individuo:

$y_{ci} = 0$ se lavoratore dipendente

$y_{ci} = 1$ se lavoratore autonomo

MODELLO PROBIT bivariato: vi sono due variabili latenti e due variabili y_{ci} osservate: una dà lo status lavorativo di un dato individuo, l'altra riguarda la scelta occupazionale del partner dell'individuo stesso.

MODELLO PROBIT ad effetti casuali: $y_{ci}^* = \beta' x_{ci} + \alpha_c + u_{ci}$, con α_c effetto di coppia e con $y_{ci} = 1$ se $y_{ci}^* > 0$, $y_{ci} = 0$ altrimenti. Si assume che $u_{ci} \sim \text{IN}(0, \sigma_u^2)$ e che, condizionatamente agli x_{ci} , gli α_c siano $\text{IN}(0, \sigma_\alpha^2)$ e indipendenti dagli u_{ci} e dagli x_{ci} .

I RISULTATI:

• *probit standard semplice*

lavoro	Coeff.	Std. Err.	Eff. Marg.	S.E.
moglie	-.2539554	.0524019	-.0734545	.0150828
eta	-.0347872	.0295626	-.0100716	.0085607
etasq	.0005837	.0003380	.0001690	.0000979
ncomp	-.0189003	.0316077	-.0054720	.0091508
npred	.0455413	.0502765	.0131852	.0145549
presc	.1176692	.0562492	.0340677	.0162799
scsup	-.2719043	.0567166	-.0778683	.0160145
univ	-.2081393	.0757674	-.0569436	.0195021
amp40	.0070696	.0579185	.0020451	.0167404
amp500	-.3752218	.1190151	-.0943447	.0253071
areaN	-.1044694	.0624814	-.0303434	.0181972
areaS	-.0961669	.0757519	-.0272488	.0209906
_cons	.0014091	.6163758		

Si può mettere in evidenza come gli individui di sesso femminile abbiano una probabilità di scegliere occupazioni con guadagni poco sicuri inferiore di circa 7 punti percentuali rispetto a quella dei maschi, dato che il coefficiente di MOGLIE stimato è negativo e che il coefficiente del rispettivo effetto marginale è pari a -0.0734545 .

Se una persona ha bambini in età prescolare è generalmente ben disposta verso il rischio, con la predisposizione a scegliere il lavoro autonomo che aumenta di circa 3 punti percentuali. Questo risultato è apparentemente sorprendente, perché sarebbe più logico aspettarsi che chi ha bambini piccoli sia restio ad accettare di correre alti rischi; è però vero che chi ha bambini in età prescolare è relativamente giovane, e chi è giovane è più facile che voglia rischiare. Ecco perché il coefficiente, che evidentemente coglie questo aspetto, risulta positivo.

Altri coefficienti statisticamente significativi al livello del 5% sono quelli sul grado d'istruzione. Risulta che gli individui che hanno studiato di più, e quindi in possesso di un diploma di scuola superiore o di una laurea, hanno probabilità di essere lavoratori autonomi che calano rispettivamente di circa 8 e 6 punti per i due titoli di studio; ciò vuol dire che il grado di rischio che accettano è basso, anche se la misura dell'incidenza negativa decresce se il titolo di studio non obbligatorio è particolarmente alto, cioè se l'individuo è laureato o in possesso di

diploma universitario (l'incidenza negativa sulla probabilità di scegliere il lavoro autonomo passa per l'appunto dagli 8 ai 6 punti percentuali).

Infine risulta che chi abita nelle grandi città vede la probabilità di scegliere il lavoro autonomo diminuire di oltre 9 punti percentuali.

• *modello probit standard bivariato*

y_{ci} :	<i>lavoro (individuo)</i>		<i>partnerlavoro</i>	
	Coeff.	Std. Err.	Coeff.	Std. Err.
moglie	-.2585834	.0521968	.3802915	.0524707
eta	-.0368488	.0293338	-.0212821	.0291534
etasq	.0005951	.0003363	.0004602	.0003353
ncomp	-.0221260	.0314841	-.0323125	.0314584
npred	.0481195	.0505318	.0618989	.0505670
presc	.1182408	.0557321	.1322842	.0557921
scsup	-.2690757	.0566194	-.0728041	.0569956
univ	-.1869420	.0747178	-.0457000	.0751060
amp40	.0093598	.0577901	-.0068043	.0577455
amp500	-.3468879	.1156711	-.3800944	.1152724
areaN	-.1020627	.0621877	-.1001336	.0621476
areaS	-.0949711	.0754321	-.0848091	.0751422
_cons	.0712113	.6091654	-.7307633	.6039301
rho	.6387215	.0238625		

Il modello bivariato di cui sopra sono riportati i risultati mette in luce l'influenza che le variabili esogene che indagano le caratteristiche dei membri delle coppie hanno sulla scelta occupazionale di ciascun individuo e su quella dei partner. Quindi le caratteristiche evidenziate rispetto a sesso, età e livello d'istruzione si riferiscono ad un dato individuo e si stima come influenzino la scelta dell'individuo stesso; inoltre si stima come quelle stesse caratteristiche relative all'individuo preso in analisi influenzino la scelta del partner. Lo stesso discorso vale per le altre variabili che però, a differenza di età, sesso e titolo di studio, sono comuni alla coppia. Le due equazioni del modello bivariato sono quindi del tipo:

$$y_{individuo}^* = x_{individuo} \beta_{individuo} + \epsilon_{individuo}$$

$$y_{partner}^* = x_{individuo} \beta_{individuo}^{partner} + \epsilon_{partner}$$

Così, ad esempio, il fatto di essere di sesso femminile incide negativamente sulla propensione al rischio di un dato individuo, mentre per il partner tale influenza va nella direzione opposta. Va però detto che i coefficienti stimati, quando si riferiscono ai partner e non direttamente agli individui, risultano perlopiù non statisticamente significativi.

Il modello bivariato fornisce anche la correlazione tra individuo e partner a livello occupazionale, rispetto alle caratteristiche esogene del primo; la correlazione risulta essere piuttosto alta e questo indica che i coniugi/conviventi hanno buone probabilità di essere entrambi lavoratori autonomi piuttosto che dipendenti.

Gli effetti marginali, riportati a breve, mettono in luce quale tipo di influenza abbiano le caratteristiche degli individui sulla probabilità che siano loro stessi lavoratori autonomi, sulla probabilità che ad essere lavoratori autonomi siano i partner, e sulla probabilità che entrambi i coniugi/conviventi scelgano l'occupazione "meno sicura".

• *effetti marginali probit standard bivariato*

	<i>Pr(lavoro=1)</i>		<i>Pr(partnerlavoro=1)</i>		<i>Pr(lavoro=1,partnerl.=1)</i>	
	Eff. marg.	S.E.	Eff. marg.	S.E.	Eff. marg.	S.E.
moglie	-.0750100	.01506	.1104420	.01510	.0120669	.00896
eta	-.0197000	.00852	-.0061944	.00848	-.0059922	.00503
etasq	.0001728	.00010	.0001339	.00010	.0001087	.00006
ncomp	-.0064249	.00914	-.0094050	.00916	-.0056062	.00541
npred	.0139727	.01467	.0180165	.01472	.0113313	.00871
presc	.0343343	.01618	.0385031	.01623	.0258067	.00960
scsup	-.0772974	.01605	-.0211338	.01650	-.0349740	.00960
univ	-.0516149	.01956	-.0131445	.02135	-.0232580	.01165
amp40	.0027148	.01674	-.0019821	.01683	.0002697	.00993
amp500	-.0885313	.02538	-.0959846	.02461	-.0608647	.01267
areaN	-.0297292	.01816	-.0292343	.01819	-.0209455	.01082
areaS	-.0269993	.02098	-.0232250	.02105	-.0179445	.01216

Sono solo due gli effetti marginali stimati che risultano statisticamente significativi in tutti e tre i casi, quelli relativi alla presenza di bambini in età prescolare e all'eventuale residenza in una grande città. Se per i due membri della coppia presi singolarmente i risultati sono simili, essendo le caratteristiche prevalentemente comuni ad entrambi, è più interessante evidenziare come se vi

sono bambini di età inferiore ai sei anni aumenti di circa 2,6 punti percentuali la probabilità che entrambi i partner siano lavoratori autonomi (presumibilmente perché chi ha bambini piccoli è relativamente giovane e quindi ben disposto ad accettare alti rischi); invece se i coniugi/conviventi vivono in una grande città la probabilità scende di 6 punti. Infine i risultati danno riscontro di come se almeno uno dei partner è diplomato (scuola superiore o professionale) o laureato (o in possesso di diploma universitario) la probabilità che entrambi i partner siano lavoratori autonomi scende rispettivamente di 4 e di 2 punti percentuali circa.

• *modello probit standard ad effetti casuali*

lavoro	Coeff.	Std. Err.
moglie	-.26900595	.04202654
eta	-.03661365	.03010714
etasq	.00055464	.00034395
ncomp	-.01741792	.03676411
npred	.05365574	.05943871
presc	.10108973	.06483132
scsup	-.28207876	.05616885
univ	-.22535825	.07737099
amp40	.01742115	.06813927
amp500	-.33047655	.13483920
areaN	-.10401965	.07347768
areaS	-.09188059	.08911218
rho	.64224650	.03345590

Il modello ad effetti casuali porta a stabilire quanta parte della variazione sulle categorie di lavoro possa essere spiegata da una non osservabile correlazione interna alla coppia.

I coefficienti stimati, sempre relativi al sotto-campione che include i soli membri delle coppie capofamiglia-partner entrambi lavoratori, sono concordi con quelli restituiti dal modello probit standard semplice, anche se il coefficiente di PRESC (che indica se vi sono bambini in età prescolare presenti in famiglia) risulta ora non statisticamente significativo al livello 5%, e neppure al 10%. Essendo in ogni caso i coefficienti stimati col modello ad effetti casuali simili a quelli restituiti dal modello ordinario, lo sono anche gli effetti marginali; per tale motivo mi

astengo dal riportare quelli relativi al modello ad effetti casuali, dato anche che sono già stati commentati nell'ambito del modello ordinario.

Il ρ trovato, significativamente diverso da zero, stabilisce che il 64% circa dell'eterogeneità della variabile latente che dà il livello di avversione individuale al rischio è spiegata dall'effetto coppia. Così, dato che tale effetto spiega circa il 64% della varianza non osservata, e dato che si basa sulle correlazioni attraverso la variabile dipendente nelle coppie, i risultati vanno a confermare quanto evidenziato dal modello bivariato, con ρ alquanto vicini.

Si noti che il ρ trovato nel modello ad effetti casuali, per come è strutturato e come ho già avuto modo di evidenziare nel capitolo 2, non poteva che essere positivo o al più nullo.

18.3 Altri risultati ottenuti nel bivariato

Questo sotto-paragrafo si occupa di mettere in evidenza tutta una serie di risultati che riguardano l'influenza che le caratteristiche individuali dei partner hanno tra di loro rispetto alla scelta occupazionale a livello di rischio.

SOTTO-ANALISI 1

E' possibile confrontare i coefficienti stimati relativi agli individui e all'influenza che indirettamente hanno sui partner, restituiti dal modello probit standard bivariato, a seconda del sotto-campione preso in considerazione. Come nell'univariato, i sotto-campioni analizzati sono quello comprendente tutti i partner delle coppie con entrambi i coniugi/conviventi lavoratori, il sotto-campione con inclusi i soli soggetti di sesso maschile, e quello delle mogli/conviventi. Se ad esempio il sotto-campione è quello degli individui di sesso femminile, le due equazioni del modello bivariato sono del tipo:

$$y_{moglie}^* = x_{moglie} \beta_{moglie}^{moglie} + \epsilon_{moglie}$$

$$y_{marito}^* = x_{moglie} \beta_{moglie}^{marito} + \epsilon_{marito}$$

I risultati sono riportati in modo da facilitare il confronto fra i coefficienti.

• *modello probit standard bivariato, sotto-analisi 1*

	ENTRAMBI I PARTNER		SOLI MARITI/CONV.		SOLE MOGLI/CONV.	
lavoro	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
eta	-.0319782	.0292356	-.0479230	.0433006	-.0422935	.0447682
etasq	.0005900	.0003354	.0007188	.0004813	.0006598	.0005362
ncomp	-.0292836	.0313350	-.0232811	.0432514	-.0145070	.0466025
npred	.0336924	.0503639	.0220507	.0694353	.0723781	.0740217
presc	.1466971	.0552629	.1471769	.0748987	.0784017	.0847108
scsup	-.2758915	.0564727	-.1476767	.0763037	-.4104375	.0850149
univ	-.2040225	.0744027	-.0955173	.1013292	-.2996439	.1115879
amp40	.0066908	.0576490	.0290676	.0785433	-.0148357	.0857375
amp500	-.3472750	.1153515	-.3780863	.1549694	-.2995887	.1735695
areaN	-.1008400	.0620851	-.0471768	.0847420	-.1685405	.0919581
areaS	-.0955420	.0752061	-.0694158	.1031050	-.1296300	.1109453
_cons	-.2195779	.6051908	.2013665	.9223944	.0374641	.8851988
partner1.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
eta	-.0256911	.0290718	-.0163436	.0485172	-.0164223	.0414132
etasq	.0004320	.0003343	.0004035	.0005345	.0004006	.0004974
ncomp	-.0216820	.0312643	-.0346185	.0470584	-.0313628	.0427583
npred	.0813146	.0501473	.0989024	.0734713	.0315685	.0696463
presc	.0892643	.0550855	.0950561	.0835182	.1594208	.0758090
scsup	-.0563391	.0565227	-.0477035	.0833567	-.1104078	.0784835
univ	-.0175255	.0745317	-.0695627	.1119849	-.0419237	.1021109
amp40	-.0018904	.0574010	-.0295691	.0852877	.0116762	.0786461
amp500	-.3735981	.1149634	-.3527600	.1724479	-.3926823	.1547648
areaN	-.0972173	.0617838	-.1738870	.0912005	-.0430088	.0850091
areaS	-.0798673	.0747747	-.1232290	.1107397	-.0532001	.1023807
_cons	-.3373926	.5997629	-.7726750	1.0449920	-.4817589	.8173227
/athrho	.6914275	.0385773	.7565296	.0569338	.7647785	.0573497
rho	.5988983	.0247405	.6390283	.0336844	.6438830	.0335734

Il fatto nuovo che emerge da una prima lettura dei dati è quello relativo alla possibilità di raffrontare i coefficienti di ρ nei tre diversi casi. Il coefficiente risulta per ciascun sotto-campione altamente significativo ed elevato; ciò è sintomo di come la scelta occupazionale del partner, a livello di rischio, influenzi in modo importante la scelta degli individui. In particolare il risultato è sostanzialmente simile se si confrontano i sotto-campioni coi soli maschi e con le sole femmine, con le caratteristiche esogene che si riferiscono rispettivamente ai maschi e alle femmine e di cui si studia anche l'influenza sulla scelta del partner dell'altro sesso.

Il test del rapporto di verosimiglianza permettere di comparare la verosimiglianza del modello bivariato completo con la somma delle log verosimiglianze dei modelli probit bivariati che si riferiscono ai due sotto-campioni "mariti/conv." e

“mogli/conv.”, in modo da verificare se i vettori dei coefficienti nei due casi siano uguali o meno:

$$2 \cdot |l_1 - l_0| = 2 \cdot |-3051.46 - (-1494.59 - 1486.98)| = 2 \cdot |-69.89| = 139.78 > \chi^2(13) = 22.36$$

con l_1 log-verosimiglianza per il sotto-campione “entrambi i partner”, e l_0 somma delle log-verosimiglianze per i sotto-campioni che includono i soli individui di sesso maschile e femminile presi singolarmente.

Risulta che si rifiuta l’ipotesi nulla sull’uguaglianza dei due vettori di coefficienti.

• *effetti marginali probit standard bivariato, sotto-analisi 1*

ENTRAMBI I PARTNER			SOLI MARITI/CONV.		SOLE MOGLI/CONV.	
	Eff. Marg.	S.E.	Eff. Marg.	S.E.	Eff. Marg.	S.E.
<i>Pr (lavoro=0,partnerl.=0)</i>			<i>P (lavoro=0,partnerl.=0)</i>		<i>P (lavoro=0,partnerl.=0)</i>	
eta	.0110268	.00923	.0140268	.01440	.0094714	.01365
etasq	-.0001954	.00011	-.0002309	.00016	-.0001838	.00016
ncomp	.0097422	.00992	.0102756	.01428	.0096783	.01411
npred	-.0220714	.01595	-.0183101	.02283	-.0170744	.02289
presc	-.0450746	.01751	-.0488984	.02492	-.0497875	.02518
scsup	.0626325	.01767	.0425818	.02500	.0793552	.02572
univ	.0393498	.02296	.0322553	.03268	.0437849	.03280
amp40	-.0009048	.01822	-.0032938	.02590	-.0010164	.02594
amp500	.1255504	.03026	.1270540	.04143	.1237395	.04181
areaN	.0379566	.01970	.0344795	.02802	.0323274	.02809
areaS	.0331101	.02315	.0326547	.03324	.0290998	.03311
<i>Pr (lavoro=1,partnerl.=1)</i>			<i>P (lavoro=1,partnerl.=1)</i>		<i>P (lavoro=1,partnerl.=1)</i>	
eta	-.0058671	.00491	-.0057251	.00790	-.0064671	.00738
etasq	.0001040	.00006	.0001054	.00009	.0001119	.00009
ncomp	-.0051861	.00527	-.0060648	.00775	-.0041768	.00765
npred	.0116703	.00848	.0138813	.01226	.0113310	.01228
presc	.0240220	.00931	.0231041	.01363	.0218061	.01379
scsup	-.0336217	.00935	-.0172217	.01349	-.0585201	.01395
univ	-.0222160	.01137	-.0154452	.01711	-.0378634	.01556
amp40	.0004929	.00968	-.0013528	.01407	-.0009349	.01408
amp500	-.0593985	.01287	-.0590249	.01790	-.0553335	.01842
areaN	-.0202518	.01057	-.0251665	.01530	-.0241623	.01534
areaS	-.0172893	.01184	-.0198414	.01707	-.0194105	.01704

Non sono riportati gli effetti marginali relativi all’influenza delle caratteristiche esogene degli individui sulla loro scelta occupazionale e su quella dei partner,

essendo i coefficienti stimati per PARTNERLAVORO (una delle due variabili dipendenti) non statisticamente significativi né al 5% né al 10%; questo risultato era ampiamente prevedibile in quanto, per saggiare la scelta occupazionale del partner, non si prendono in considerazione le sue caratteristiche, bensì quelle del coniuge/convivente. L'eccezione è rappresentata dall'esplicativa amp500, che però, non essendo possibile che due partner appartenenti ad una famiglia non risiedano nella stessa città, è di scarso interesse rispetto alla peculiarità di questa sotto-analisi.

Vediamo invece qual è l'influenza che le caratteristiche esogene degli individui hanno sulla probabilità che la coppia sia formata da partner che scelgono lo stesso tipo di lavoro a livello di rischio associato.

Risulta che se un individuo è diplomato alle scuole superiori o è laureato, la probabilità che la coppia sia formata da partner entrambi lavoratori autonomi diminuisce rispettivamente di poco più di 3 e di 2 punti percentuali; se l'individuo che ha queste caratteristiche a livello scolastico è di sesso femminile, l'incidenza negativa aumenta rispettivamente a quasi i 6 e i 4 punti. Ancora una volta l'influenza negativa sulla propensione al rischio, pur permanendo, è di minore entità se il titolo di studio "non dell'obbligo" aumenta di livello.

Infine, se un individuo vive in una grande città, la probabilità che la coppia veda entrambi i partner accettare un alto grado di incertezza diminuisce di circa 6 punti, a livello percentuale; l'incidenza non cambia sostanzialmente se l'individuo che vive nella grande città è uomo piuttosto che donna.

Quanto alla situazione che vede entrambi i partner essere lavoratori dipendenti, gli effetti marginali stimati mettono in luce l'incidenza negativa, simile per tutti e tre i sotto-campioni, della presenza in famiglia di bambini in età prescolare, di poco inferiore ai 5 punti percentuali (presumibilmente perché si coglie il fatto che chi ha bambini piccoli è relativamente giovane e quindi, in quanto tale, ben disposto ad accettare alti rischi), e quella positiva relativa all'eventuale residenza in una grande città, superiore ai 12. Se almeno un individuo della coppia è in possesso di un diploma di scuola superiore (ma non è laureato), la probabilità che entrambi i coniugi/conviventi siano lavoratori dipendenti aumenta di oltre i 6

punti percentuali; se ad essere diplomata (scuola superiore o professionale) è la donna, la probabilità aumenta ulteriormente di quasi altri 2 punti.

SOTTO-ANALISI 2

Un altro modello probit bivariato, di cui sono riportati i risultati a breve, studia la correlazione tra le scelte occupazionali rispetto al rischio tra mariti e mogli, o conviventi, sempre appartenenti alle coppie che vedono entrambi i partner occupati in una qualche attività lavorativa. A differenza del modello precedentemente visto, ora si studia l'influenza delle caratteristiche esogene di ciascun membro della coppia sull'individuo stesso; ciò vuol dire che per i maschi si prendono in considerazione le sole caratteristiche su età e livello di istruzione dei mariti/conviventi, indicate con M, mentre per le femmine si guarda ad età e livello d'istruzione delle mogli/conviventi, segnalate con F; ovviamente per altre variabili esogene, come ad esempio NCOMP (numero di componenti della famiglia cui appartiene la coppia), non ha senso specificare se sono riferite agli uomini piuttosto che alle donne, in quanto comuni ad entrambi. Le due equazioni del modello bivariato sono quindi del tipo:

$$y_{moglie}^* = x_{moglie} \beta_{moglie}^{moglie} + \varepsilon_{moglie}$$

$$y_{marito}^* = x_{marito} \beta_{marito}^{marito} + \varepsilon_{marito}$$

• *modello probit standard bivariato, sotto-analisi 2*

y_{ci} :	<i>lavoro mariti/conv.</i>		<i>lavoro mogli/conv.</i>	
	Coeff.	Std. Err.	Coeff.	Std. Err.
etaM	-.0540181	.0412179	-	-
etaF	-	-	-.0516581	.0424485
etasqM	.0007481	.0004580	-	-
etasqF	-	-	.0007441	.0005083
ncomp	-.0157721	.0430270	-.0081776	.0463832
npred	.0262381	.0691562	.0751941	.0738003
presc	.1267722	.0744942	.0615999	.0841244
scsupM	-.1824328	.0709452	-	-
scsupF	-	-	-.3952021	.0788869
univM	-.1346905	.0952745	-	-
univF	-	-	-.3283488	.1046169
amp40	.0361116	.0784835	-.0086050	.0856790
amp500	-.3613333	.1546747	-.2853040	.1733455
areaN	-.0513665	.0846729	-.1685990	.0919466
areaS	-.0707876	.1029892	-.1263763	.1109078
_cons	.4083703	.8798540	.2484677	.8415819
rho	.6459321	.0336690		

I risultati non sono sostanzialmente nuovi, anche se, rispetto ai modelli univariati visti nel sotto-paragrafo 18.1, non sono incluse tra le esogene sia la variabile MOGLIE, perché l'informazione sul sesso è già data, sia PARTNERLAVORO, essendo il modello trattato nel bivariato; l'informazione sulla correlazione tra le scelte occupazionali di mariti/conviventi e mogli/conviventi è alta e altamente statisticamente significativa, ed è pari a circa 0.65.

Il modello bivariato qui proposto restituisce una stima degli effetti marginali, relativamente alla probabilità che gli individui di sesso maschile e femminile presi singolarmente siano lavoratori autonomi, che tiene conto anche dell'influenza delle caratteristiche esogene degli uni sugli altri per via della correlazione.

• *effetti marginali probit standard bivariato, sotto-analisi 2*

	<i>Pr(maschi autonomi)</i>		<i>Pr(femmine autonome)</i>	
	Eff. marg.	S.E.	Eff. marg.	S.E.
etaM	-.0176106	.01344	-	-
etaF	-	-	-.0129297	.01063
etasqM	.0002439	.00015	-	-
etasqF	-	-	.0001862	.00013
ncomp	-.0051419	.01403	-.0020468	.01161
npred	.0085540	.02255	.0188206	.01847
presc	.0413294	.02428	.0154181	.02105
scsupM	-.0589307	.02269	-	-
scsupF	-	-	-.0981945	.01933
univM	-.0426366	.02925	-	-
univF	-	-	-.0739194	.02104
amp40	.0117321	.02541	-.0021564	.02150
amp500	-.1053650	.03947	-.0630846	.03330
areaN	-.0167671	.02767	-.0424732	.02329
areaS	-.0227883	.03273	-.0305738	.02590

Rispetto agli effetti marginali che risultano statisticamente significativi al 5%, risulta che la probabilità che i mariti/conviventi siano lavoratori autonomi cala di quasi 6 punti percentuali se diplomati, di oltre 10 se vivono in una grande città; per le femmine la probabilità di essere lavoratrici autonome diminuisce rispettivamente di circa 10 e 7 punti percentuali se diplomate o laureate.

Dato che emerge come sia probabile che i partner facciano la stessa scelta occupazionale a livello di rischio associato, è utile guardare anche agli effetti marginali che mettono in luce come incidano le diverse variabili esogene, a livello probabilistico, rispetto alla situazione che veda entrambi i partner essere lavoratori autonomi piuttosto che dipendenti.

	<i>Pr(entrambi autonomi)</i>		<i>Pr(entambi dipendenti)</i>	
	Eff. marg.	S.E.	Eff. marg.	S.E.
etaM	-.0041129	.00315	.0134976	.01030
etaF	-.0063857	.00526	.0065440	.00539
etasqM	.0000570	.00004	-.0001869	.00011
etasqF	.0000920	.00006	-.0000943	.00006
ncomp	-.0022117	.00761	.0049769	.01409
npred	.0112928	.01223	-.0160818	.02269
presc	.0172671	.01358	-.0394804	.02468
scsupM	-.0139490	.00550	.0449816	.01728
scsupF	-.0483202	.00968	.0498742	.01918
univM	-.0104456	.00753	.0321910	.02175
univF	-.0388742	.01186	.0350452	.00953
amp40	.0017043	.01399	-.0078714	.02578
amp500	-.0625741	.01894	.1158755	.04245
areaN	-.0248680	.01535	.0343723	.02799
areaS	-.0202729	.01701	.0330891	.03311

Rispetto agli effetti marginali che risultano statisticamente significativi al 5%, l'informazione data è ancora una volta quella relativa al grado d'istruzione dei partner, e su come questo incida sulla possibilità che la coppia sia formata da due lavoratori entrambi autonomi o dipendenti.

I risultati confermano quanto visto nella sotto-analisi precedente (1), con l'unica differenza che, solo rispetto alla probabilità che si formino coppie di lavoratori entrambi dipendenti, la differenziazione rispetto al sesso, relativa al conseguimento di un titolo di studio, è meno marcata.

SOTTO-ANALISI 3

Un altro modello probit bivariato è stato invece realizzato per studiare la correlazione tra le scelte occupazionali di coniugi e conviventi quando tra le variabili esogene sono incluse, sia per i maschi che per le femmine, non solo le informazioni comuni ai partner e quelle proprie di ciascun individuo, ma anche quelle su età e livello d'istruzione del partner. Quindi si studia l'influenza che hanno le caratteristiche comuni, l'età e il livello d'istruzione dei mariti/conviventi, e l'età e il grado d'istruzione delle mogli/conviventi, sia sui maschi che sulle femmine.

Le due equazioni del modello bivariato sono quindi del tipo:

$$y_{moglie}^* = x_{moglie} \beta_{moglie}^{moglie} + x_{marito} \beta_{marito}^{moglie} + \varepsilon_{moglie}$$

$$y_{marito}^* = x_{marito} \beta_{marito}^{marito} + x_{moglie} \beta_{moglie}^{marito} + \varepsilon_{marito}$$

• *modello probit standard bivariato, sotto-analisi 3*

<i>y_{ci}</i> :	<i>lavoro mariti/conv.</i>		<i>lavoro mogli/conv.</i>		<i>Pr(entrambi autonomi)</i>	
	<i>Coeff.</i>	<i>Std. Err.</i>	<i>Coeff.</i>	<i>Std. Err.</i>	<i>Eff. marg.</i>	<i>S.E.</i>
etaM	-.0771986	.0589955	.0435627	.0656410	-.0004240	.01071
etaF	.0352675	.0563245	-.0728413	.0606403	-.0063649	.01003
etasqM	.0009505	.0006477	-.0001897	.0007176	.0000482	.00012
etasqF	-.0003008	.0006697	.0007463	.0007205	.0000698	.00012
ncomp	-.0245343	.0435079	-.0271093	.0475093	-.0052084	.00781
npred	.0155911	.0700101	.0739487	.0748555	.0103369	.01241
presc	.1575861	.0764936	.0978198	.0856737	.0240008	.01396
scsupM	-.1483839	.0838140	.1303514	.0926296	.0045406	.01510
scsupF	-.0454551	.0858814	-.4601130	.0937707	-.0596594	.01529
univM	-.1244495	.1205751	.1675459	.1350507	.0095541	.02243
univF	.0543883	.1224371	-.3859395	.1345977	-.0426919	.01834
amp40	.0182874	.0791574	-.0009729	.0866607	.0012609	.01417
amp500	-.3943646	.1558422	-.3042638	.1750867	-.0557209	.01979
areaN	-.0484093	.0851506	-.1522798	.0925547	-.0226148	.01543
areaS	-.0742817	.1032001	-.1164668	.1118364	-.0193419	.01715
_cons	.1346750	.9563299	-.4342961	1.073946	-	-
rho	.6510184	.0334348				

Il coefficiente di correlazione è analogo a quello precedentemente trovato, mentre diversi coefficienti stimati non sono statisticamente significativi, risultato che ci si poteva aspettare visto che le equazioni del modello bivariato adottato prevedono che anche le caratteristiche del partner influenzino la scelta occupazionale degli individui. Per questo motivo non riporto i risultati sugli effetti marginali relativi alle probabilità che maschi e femmine presi singolarmente scelgano il lavoro autonomo quando si misura anche l'influenza subita rispetto alle caratteristiche dei partner.

Quanto alla probabilità che entrambi i partner siano lavoratori autonomi, i risultati sugli effetti marginali, quando sono statisticamente significativi, confermano quanto già emerso dalle due sotto-analisi precedenti.

19 La scelta fra tre opzioni: risultati e commenti

Vediamo ora cosa risulta dai modelli adottati qualora la scelta osservata sul tipo di lavoro avvenga fra tre grandi categorie occupazionali, quella del lavoro autonomo e quelle che tengono conto della distinzione tra lavoro dipendente nel pubblico e nel privato.

19.1 Confronto fra sotto-campioni

Oltre a quello comprendente entrambi i coniugi lavoratori di una coppia, si considerano i sotto-campioni che includono i soli mariti/conviventi e quindi le sole mogli/conviventi, appartenenti a coppie in cui entrambi i partner lavorano. Si confrontano, e a tal scopo i risultati sono riportati come segue, i coefficienti e gli effetti marginali stimati con un modello probit ordinale. La variabile dipendente y_{ic} è categorica, cioè assume valore 1 se l'individuo è lavoratore dipendente nel pubblico, 2 se l'individuo è dipendente nel privato, e 3 se l'individuo è lavoratore autonomo. Si ha

$$y_{ci}^* = \beta' x_{ci} + \varepsilon_{ci}$$

con y_{ci}^* non osservabile, mentre si osserva y_{ic} tale che

$$y_{ci} = 1 \text{ se } y_{ci}^* \leq \mu_1$$

$$y_{ci} = 2 \text{ se } \mu_1 < y_{ci}^* \leq \mu_2$$

$$y_{ci} = 3 \text{ se } \mu_2 \leq y_{ci}^*$$

con μ (cut off) e β parametri ignoti da stimare. Ovviamente si ha $\varepsilon_{ci} \sim \text{NID}(0,1)$.

I risultati empirici confermano la bontà dell'ordinamento scelto, che vede associare al lavoro dipendente nel pubblico il grado di rischio inferiore e al lavoro autonomo quello maggiore; infatti, per tutti e tre i sotto-campioni, i due cut off risultano ben distanti tra loro, indizio questo dell'adeguata separabilità, in termini di grado di rischio associato, delle tre categorie occupazionali prese in considerazione.

• *modelli probit ordinali*

	ENTRAMBI I PARTNER		SOLI MARITI/CONV.		SOLE MOGLI/CONV.	
lav3	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
eta	-.0443088	.0239212	-.0335123	.0369697	-.0649616	.0357865
etasq	.0005771	.0002775	.0004032	.0004137	.0006778	.0004337
ncomp	-.0765287	.0257971	-.0481904	.0368776	-.0786344	.0372876
npred	.0578300	.0424803	.0145770	.0604353	.1495396	.0616732
presc	.1270332	.0450144	.1777616	.0639492	-.0208717	.0658037
scsup	-.3866913	.0466434	-.2047163	.0653694	-.5335788	.0682383
univ	-.6445029	.0633474	-.3454114	.0892158	-.9067338	.0928190
partpubbl	-.5313428	.0524230	-.5605372	.0702818	-.6656421	.0826251
partaut	.7111089	.0536925	.8820888	.0857496	.7311249	.0714971
amp40	.0280173	.0472199	.0515023	.0668739	.0180982	.0678847
amp500	-.0143465	.0863432	-.0625494	.1214682	.0636540	.1248964
areaN	-.0328828	.0508259	.0491459	.0720389	-.1182735	.0730109
areaS	-.2473658	.0629378	-.1524554	.0891572	-.3365254	.0907158
_cut1	-2.070225	.4900871	-1.826182	.7831412	-2.796885	.7016285
_cut2	-.4755490	.4889094	-.2279357	.7818663	-1.102978	.6986774
	Eff. Marg.	S.E.	Eff. Marg.	S.E.	Eff. Marg.	S.E.
eta	-.0120293	.00650	-.0104351	.01151	-.0138642	.00765
etasq	.0001567	.00008	.0001255	.00013	.0001446	.00009
ncomp	-.0207766	.00701	-.0150056	.01148	-.0167822	.00798
npred	.0157001	.01153	.0045390	.01882	.0319149	.01307
presc	.0344879	.01223	.0553516	.01992	-.0044545	.01404
scsup	-.1032481	.01228	-.0630295	.01990	-.1129943	.01474
univ	-.1430026	.01134	-.0985132	.02312	-.1385780	.01144
partpubbl	-.1281367	.01135	-.1603627	.01838	-.1142644	.01207
partaut	.2228904	.01874	.3142284	.03259	.1858394	.02118
amp40	.0075784	.01271	.0159492	.02060	.0038509	.01440
amp500	-.0038740	.02319	-.0191070	.03638	.0139982	.02828
areaN	-.0089376	.01383	.0152809	.02237	-.0253809	.01577
areaS	-.0631207	.01507	-.0460374	.02607	-.0646556	.01574

I soli coefficienti che risultano essere significativi al livello del 5% per tutti e tre i sotto-campioni, e che quindi sono confrontabili, sono quelli che danno l'informazione sull'influenza che hanno titolo di studio e lavoro del partner sulla scelta occupazionale a livello di rischio accettato.

Risulta che il possesso di un diploma di scuola superiore (o diploma professionale), e ancor più la laurea (anche laurea breve), incidono negativamente sulla propensione al rischio degli individui; infatti, se in possesso di diploma di scuola superiore o di laurea, gli individui vedono diminuire la probabilità di "saltare" di categoria occupazionale (da dip. pubbl. a dip. priv. o da dip. priv. a lavoro autonomo), accettando così più alti rischi, rispettivamente di

circa 10 e 14 punti percentuali. Quindi, se si considerano le tre categorie di rischio, all'aumentare del titolo di studio non obbligatorio aumenta anche l'incidenza negativa sulla propensione al rischio. Questi risultati sono più "forti" per le femmine (11 e 14 punti in meno) che per i maschi (6 e 9 punti in meno).

Per quanto riguarda lo status occupazionale del partner i coefficienti risultano, in tutti e tre i casi, altamente significativi. I risultati, pur concordi a livello di segno, si discostano abbastanza a seconda che il sotto-campione considerato sia quello dei soli mariti/conviventi piuttosto che quello dei soli individui di sesso femminile. Infatti se il partner è lavoratore pubblico (basso rischio) le donne vedono diminuire la probabilità di lavorare in occupazioni con associato più alto rischio di circa 11 punti percentuali; per gli uomini questa probabilità cala invece di ben 16 punti. Se invece il partner è lavoratore autonomo (alto rischio) la probabilità di lavorare in occupazioni "meno sicure" aumenta, per le donne, di quasi 19 punti percentuali, e, per i maschi, di oltre 31. In ogni caso, seppur in modo più accentuato per i mariti/conviventi, l'occupazione del partner incide, con correlazione positiva, sulla scelta lavorativa a livello di rischio accettato dagli individui.

Per verificare se i vettori dei coefficienti dei due sotto-campioni "mariti/conv." e "mogli/conv." siano uguali o meno, si utilizza il test del rapporto di verosimiglianza:

$$2 \cdot |l_1 - l_0| = 2 \cdot |-2886.34 - (-1445.43 - 1360.67)| = 2 \cdot |-80.24| = 160.48 > \chi^2(14) = 23.69$$

con l_1 log-verosimiglianza per il sotto-campione "entrambi i partner", e l_0 somma delle log-verosimiglianze per i sotto-campioni che includono gli individui di sesso maschile e femminile presi singolarmente. Il risultato vede rifiutare l'ipotesi nulla sull'uguaglianza dei due vettori di coefficienti.

19.2 L'output dei modelli probit ordinali

Vediamo ora come influiscono le diverse caratteristiche individuali sulla scelta occupazionale a livello di rischio associato quando si valuta come si pongono gli individui rispetto alle tre opzioni; inoltre è interessante guardare se e di quanto, utilizzando il modello ad effetti casuali, cambia la porzione di eterogeneità non osservata sull'avversione individuale al rischio spiegata dall'effetto coppia rispetto al caso di scelta binaria.

SOTTO-CAMPIONE: membri delle coppie ove entrambi i partner lavorano

I modelli, spiegati nel dettaglio nel capitolo 4 e adattati all'analisi empirica della tesi, sono sinteticamente presentati come segue:

MODELLO PROBIT ORDINALE semplice: la variabile dipendente y_{ci} è categorica, cioè assume valore 1 se l'individuo è lavoratore dipendente nel pubblico, 2 se l'individuo è dipendente nel privato, e 3 se l'individuo è lavoratore autonomo. Si ha

$$y_{ci}^* = \beta' x_{ci} + \varepsilon_{ci}, \varepsilon_{ci} \sim \text{NID}(0,1)$$

con y_{ci}^* non osservabile, mentre si osserva y_{ci} tale che $y_{ci} = 1$ se $y_{ci}^* \leq \mu_1$, $y_{ci} = 2$ se $\mu_1 < y_{ci}^* \leq \mu_2$, e $y_{ci} = 3$ se $\mu_2 < y_{ci}^*$, con μ (cut off) e β parametri ignoti da stimare.

MODELLO PROBIT ORDINALE ad effetti casuali: $y_{ci}^* = \beta' x_{ci} + \alpha_c + u_{ci}$,

con α_c effetto di coppia e con $y_{ci} = 1$ se $y_{ci}^* \leq \mu_1$, $y_{ci} = 2$ se $\mu_1 < y_{ci}^* \leq \mu_2$, e $y_{ci} = 3$ se

$\mu_2 < y_{ci}^*$, con μ (cut off) e β parametri ignoti da stimare. Si

assume che $u_{ci} \sim \text{IN}(0, \sigma_u^2)$ e che, condizionatamente agli x_{ci} , gli α_c siano $\text{IN}(0, \sigma_\alpha^2)$ e indipendenti dagli u_{ci} e dagli x_{ci} .

I RISULTATI:

- *modello probit ordinale*

	Coeff.	Std. Err.	Eff. Marg.	S.E.
lav3				
moglie	-.2937631	.0416688	-.0832674	.01181
eta	-.0666954	.0234438	-.0189266	.00666
etasq	.0007640	.0002716	.0002168	.00008
ncomp	-.0858593	.0253059	-.0243649	.00719
npred	.1189076	.0416054	.0337432	.01181
presc	.1096959	.0446308	.0311291	.01267
scsup	-.4134967	.0456154	-.1152969	.01254
univ	-.6849169	.0618193	-.1582134	.01153
amp40	.0285409	.0464266	.0080707	.01308
amp500	-.0622737	.0850090	-.0172827	.02306
areaN	-.0470380	.0499984	-.0133689	.01423
areaS	-.3190933	.0614746	-.0838718	.01491
_cut1	-2.8567550	.4830882		
_cut2	-1.3616510	.4814753		

Innanzitutto va detto che il modello con le tre opzioni restituisce un numero maggiore di coefficienti significativamente diversi da zero al 5% rispetto a quelli ottenuti con i modelli applicati alla scelta binaria, cioè senza tener conto della evidentemente utile distinzione tra lavoro dipendente nel pubblico e nel privato.

Le donne hanno minore propensione al rischio rispetto agli uomini, con la probabilità di essere lavoratrici dipendenti private piuttosto che pubbliche, o autonome invece che dipendenti private, che scende di oltre gli 8 punti percentuali. L'età, altra variabile con coefficiente significativo al 5%, ha un impatto negativo sull'accettazione del rischio pari quasi ai 2 punti percentuali.

I soggetti che appartengono a famiglie numerose vedono diminuire la possibilità di scegliere occupazioni con associato un più alto livello di rischio di 2.4 punti; va però detto che al crescere del numero di percettori di reddito all'interno della famiglia aumenta anche la propensione al rischio degli individui, con un incremento a livello percentuale superiore ai 3 punti. La presenza in famiglia di bambini in età prescolare è correlata positivamente al grado di accettazione del rischio da parte dei coniugi/conviventi lavoratori; infatti, dato che l'effetto marginale stimato per PRESC è pari a 0.0311291, se vi sono bambini con età inferiore ai sei anni presenti in famiglia, la probabilità individuale di accettare un

più alto rischio aumenta di poco più di 3 punti percentuali; potrebbe sembrare strano che chi ha bambini piccoli sia ben disposto verso l'incertezza, ma è presumibile che il risultato colga il fatto che chi ha figli in età prescolare è relativamente giovane e quindi, in quanto tale, ben disposto ad accettare alti rischi.

Come già visto, il conseguimento di un diploma di scuola superiore (senza laurea) e ancor più la laurea hanno un impatto negativo sulla probabilità che l'individuo sia lavoratore con associato un più alto grado di rischio, rispettivamente pari ai 12 e ai 16 punti, a livello percentuale; quindi, rispetto alle tre categorie occupazionali con diversa incertezza intrinseca, all'aumentare del titolo di studio non obbligatorio diminuisce la propensione al rischio.

Chi risiede nel Sud Italia o nelle Isole ha bassa disponibilità ad accettare il rischio, con la probabilità di scegliere occupazioni meno sicure che scende di oltre 8 punti percentuali.

La distanza tra i cut off avvalora ancora una volta la validità della separazione delle categorie lavorative in tre aree con diversi gradi di rischio associato e la bontà dell'ordinamento scelto.

• *modello probit ordinale ad effetti casuali*

lav3	Coeff.	Std. Err.
moglie	-.30044588	.03516528
eta	-.06219384	.02422201
etasq	.00068812	.00027980
ncomp	-.08581855	.02950934
npred	.13199266	.04889849
prescol	.10065996	.05178857
scsup	-.39016217	.04645878
univ	-.67343978	.06663809
amp40	.02889722	.05449262
amp500	-.06453211	.09945259
areaN	-.04783432	.05876391
areaS	-.31823882	.07269750
_cut1	-2.7882536	.50946530
_cut2	-1.3113802	.50332085
rho	.47876240	.02613550

I coefficienti stimati, relativi al sotto-campione che include i soli membri delle coppie capofamiglia-partner lavoratori, sono concordi con quelli ottenuti dal modello ordinale semplice, sia nel segno che a livello di significatività; fa eccezione il solo coefficiente di PRESC (numero di bambini in età prescolare presenti in famiglia) che risulta significativamente diverso da zero solo al livello del 10%. Conseguentemente anche gli effetti marginali ottenuti col modello ad effetti casuali saranno simili a quelli relativi al modello ordinale semplice, ragion per cui non sono riportati né tanto meno commentati.

Il ρ trovato è altamente significativo, e l'ampiezza suggerisce che il 48% circa dell'eterogeneità nell'avversione individuale al rischio (variabile latente) è spiegata dall'effetto coppia. Così l'ampiezza del ρ , seppur minore rispetto a quella ottenuta col modello ad effetti casuali con scelta binaria, ma di simile significatività, suggerisce come uno specifico effetto di coppia sia ancora apprezzabile. Quest'ultimo spiega circa il 48% della varianza non osservata, e dato che si basa sulle correlazioni attraverso la variabile dipendente nelle coppie, i risultati danno ulteriore supporto all'ipotesi che capofamiglia e partner abbiano buone probabilità di lavorare in occupazioni con stesso grado di rischio.

20 La “scelta” della moglie/conv. se lavorare o meno: quali influenze?

L'analisi empirica fin qui vista ha considerato l'eventuale presenza di risk sharing su di un sotto-campione formato dalle coppie in cui entrambi i partner lavorano. E' altresì interessante vedere se e come cambiano i risultati qualora si consideri anche il caso in cui la moglie/convivente non lavori, e si condizioni la scelta occupazionale dell'uomo tra lavoro dipendente e autonomo alla presenza di una partner che lavora. In tal modo si può vedere se la “scelta” del marito tra le due categorie occupazionali con associati diversi gradi di rischio è indipendente dalla condizione di lavoratrice (o meno) da parte della partner di sesso femminile, o se invece vi dipenda.

Inoltre altro risultato interessante è quello relativo alla misura dell'eventuale condizionamento della “scelta” tra lavoro autonomo e dipendente dei

mariti/conviventi sull'essere lavoratrice o meno da parte della partner femmina; infatti in tal modo si può vedere come la "scelta" del maschio incida sulla possibilità che la donna lavori e quindi limiti il proprio rischio, visto e considerato che anche la condizione di lavoratore può essere vista come un'opportunità di auto-assicurazione da parte dell'individuo. In particolare si può valutare se e come alcune caratteristiche esogene influenzino la condizione delle partner femmine di lavoratrici o meno.

Considero solo il caso che a non lavorare, all'interno della coppia, sia il partner di sesso femminile, visto che i casi di coppie con uomo non lavoratore sono troppo pochi per ottenere risultati empirici di rilievo statistico.

Se si includono nel sotto-campione anche le coppie ove il marito/convivente di sesso maschile lavora, ma con il partner di sesso femminile non occupato, allora il numero complessivo di coppie sale a 3189 (6378 osservazioni); infatti le coppie col solo marito/convivente lavoratore sono 1620, a cui corrispondono 3240 osservazioni.

Vediamo ora come è composto il sotto-campione comprendente i membri delle coppie in cui il marito/convivente lavora mentre tale condizione non è necessaria per la donna; si considerano le classi occupazionali del lavoro dipendente e autonomo.

Tabella 12 coppie in cui almeno il partner di sesso maschile lavora

		MASCHI					
FEMMINE		Lav. dip.		Lav. aut.		TOTALE	
		N.	% ↓	N.	% ↓	N.	% ↓
Non occupate	N.	1159	50.13	461	52.57	1620	50.80
	% →	71.54		28.46		100	
Lav. dip.	N.	1059	45.80	237	27.02	1296	40.64
	% →	81.71		18.29		100	
Lav. aut.	N.	94	4.07	179	20.41	273	8.56
	% →	34.43		65.57		100	
TOTALE	N.	2312	100	877	100	3189	100
	% →	72.50		27.50		100	

20.1 La “scelta” dei maschi dipende dalla condizione di lavoratrice della partner?

Si stimano i coefficienti e gli effetti marginali del modello probit con selezione del sotto-campione. Si considera un'equazione costruita in modo da vedere sì come cambia la probabilità che i mariti/conviventi siano lavoratori autonomi (e quindi non dipendenti) a seconda delle caratteristiche esogene dei mariti stessi, ma condizionatamente al caso in cui le partner femmine lavorino; la variabile dipendente MOGLIELNL dice per l'appunto se la moglie/convivente lavora (assume valore 1) o non lavora (valore 0), e se ne misura l'influenza che hanno su tale “scelta” le caratteristiche esogene di entrambi i partner. Alcune caratteristiche sono in comune alla coppia, altre sono invece proprie di mariti e mogli (o conviventi); per questo le esplicative che danno l'informazione su età e livello d'istruzione sono contrassegnate con una M se si riferiscono ai mariti/conviventi, e con una F se riguardano i partner di sesso femminile. L'equazione del modello probit con selezione del sotto-campione è quindi del tipo:

$$y_{marito}^* = x_{marito} \beta_{marito}^{marito} + \varepsilon_{marito} \mid L_{moglie} = 1$$

con $L_{moglie}^* = x_{marito} \beta_{marito}^{moglie} + x_{moglie} \beta_{moglie}^{moglie} + \varepsilon_{moglie}$

L_{moglie} è MOGLIELNL ed assume valore 0 se la moglie/convivente non lavora, 1 se è invece occupata.

I RISULTATI:

- *modello probit con selezione del sotto-campione*

Maritolavoro	Coeff.	Std. Err.
etaM	-.0420694	.0457420
etasqM	.0006562	.0005111
ncomp	-.0223273	.0433454
npred12	.0092031	.0775776
presc	.1436421	.0759006
scsupM	-.1441811	.0823430
univM	-.0831004	.1281005
amp40	.0261906	.0787518
amp500	-.4063996	.1595367
areaN	-.0444966	.0852174
areaS	-.0899953	.1253892
_cons	.0363793	1.007194

Moglielnl	Coeff.	Std. Err.
etaM	.0309343	.0392094
etaF	.1177871	.0370283
etasqM	-.0004485	.0004289
etasqF	-.0013778	.0004365
ncomp	-.0386478	.0295711
npred12	-.2735349	.0436036
presc	-.1660665	.0539951
scsupM	-.0515748	.0568315
scsupF	.6240434	.0574800
univM	.1566627	.0974451
univF	1.270401	.1044351
amp40	-.1075061	.0564749
amp500	-.3883195	.1021028
areaN	.1013248	.0638773
areaS	-.6563265	.0685199
_cons	-2.686222	.6293296

rho	.0523970	.1790977
-----	----------	----------

Il coefficiente di rho dice se la “scelta” del marito/convivente tra lavoro autonomo e dipendente in qualche modo dipende dalla condizione della partner di lavoratrice o meno; la correlazione risulta pari a 0.05 circa ed è quindi bassa, risultato questo che mette in evidenza come la “scelta” dei mariti/conviventi di essere lavoratori autonomi non dipenda dalla eventuale condizione di lavoratrice da parte della partner.

I risultati sugli effetti marginali permettono di confrontare la probabilità condizionata a quella in assenza di selezione. Il confronto avviene tra la stima degli effetti marginali restituita dal probit con selezione del sotto-campione e

quella del probit standard qualora il sotto-campione (solo per il probit standard) includa solo i mariti che hanno moglie che lavora.

- *confronto tra effetti marginali probit con selezione del sotto-campione e senza selezione*

	<i>(con selezione)</i> <i>Pr(maritolavoro=1 moglielnl=1)</i>		<i>(senza selezione)</i> <i>Pr(lavoro=1)</i>	
	Eff. Marg.	S.E.	Eff. Marg.	S.E.
etaM	-.0142878	.01482	-.0150982	.01408
etaF	-.0013054	.00446	-	-
etasqM	.0002225	.00016	.0002302	.00016
etasqF	.0000153	.00005	-	-
ncomp	-.0069726	.01430	-.0068606	.01405
npred12	.0060821	.02290	.0063979	.02243
presc	.0494542	.02513	.0477698	.02443
scsupM	-.0467212	.02742	-.0494767	.02442
scsupF	-.0067194	.02287	-	-
univM	-.0286240	.03753	-.0336539	.03171
univF	-.0111743	.03762	-	-
amp40	.0098355	.02592	.0091355	.02541
amp500	-.1158184	.04059	-.1147938	.03898
areaN	-.0158359	.02804	-.0152247	.02769
areaS	-.0222518	.03372	-.0221610	.03274

I risultati, in quanto simili, confermano come, relativamente al sotto-campione considerato, non abbia senso tener conto del condizionamento dell'informazione che dice se la moglie/convivente lavora o meno sulla scelta occupazionale dei partner di sesso maschile tra lavoro autonomo e dipendente.

20.2 Quali influenze sulla condizione di lavoratrice o meno della partner

Se un individuo non lavora rischia sicuramente più di chi, in quanto in possesso di un'occupazione lavorativa, può, in parte, assicurarsi contro il rischio; vi sono poi, come abbiamo visto, occupazioni più o meno rischiose.

All'interno della coppie, lì dove, come quasi sempre capita, il partner di sesso maschile lavora, una buona metà delle mogli/conviventi non è occupata (il 50.80% del sotto-campione considerato). Le coppie con la partner femmina non occupata sono interessate da diversificazione del portafoglio dei rischi presenti, vi è quindi risk sharing. E' a mio parere interessante valutare se ed

eventualmente come la condizione di lavoratrice o meno da parte della moglie/convivente dipende dalla “scelta” dei maschi di essere lavoratori dipendenti, con associato più basso rischio, piuttosto che lavoratori autonomi, con associato grado di incertezza superiore.

Anche in questo caso si stimano i coefficienti e gli effetti marginali del modello probit con selezione del sotto-campione. Si considera un’equazione costruita in modo da vedere come cambia la probabilità che le partner femmine lavorino o meno a seconda delle caratteristiche esogene proprie delle mogli/conviventi oltre che quelle in comune alla coppia, condizionatamente al caso in cui il marito/convivente sia lavoratore dipendente; la variabile MARITODIP dice se l’uomo è lavoratore dipendente (assume valore 1) o autonomo (valore 0), e se ne misura l’influenza che hanno su tale “scelta” le caratteristiche esogene dei mariti/conviventi stessi, ma anche quelle delle partner di sesso femminile. L’equazione del modello probit con selezione del sotto-campione è quindi, in questo caso, del tipo:

$$L_{moglie}^* = x_{moglie} \beta_{moglie}^{moglie} + \varepsilon_{moglie} \mid y_{marito} = 1$$

con $y_{marito}^* = x_{marito} \beta_{marito}^{marito} + x_{moglie} \beta_{moglie}^{marito} + \varepsilon_{marito}$

Anche in questo caso L_{moglie} è MOGLIELNL ed assume valore 0 se la moglie/convivente non lavora, 1 se è invece occupata.

• *modello probit con selezione del sotto-campione*

Moglielnl	Coeff.	Std. Err.
etaF	.1335887	.0342552
etasqF	-.0016490	.0004273
ncomp	-.0105447	.0347381
npred12	-.3126783	.0536757
presc	-.1432436	.0735986
scsupF	.7306760	.0622346
univF	1.426411	.1134613
amp40	-.0910717	.0672543
amp500	-.2995908	.1217268
areaN	.2077629	.0760704
areaS	-.6936878	.0807046
_cons	-2.525254	.7126434

Maritodip	Coeff.	Std. Err.
etaF	-.0294067	.0370426
etaM	.1061234	.0390241
etasqF	.0003477	.0004327
etasqM	-.0014012	.0004238
ncomp	-.0166894	.0293329
npred12	.0037938	.0425410
presc	-.1736899	.0526527
scsupF	-.0683118	.0585209
scsupM	.0704988	.0574637
univF	-.0668537	.0986142
univM	.0566914	.0987729
amp40	-.0135390	.0570071
amp500	.2154084	.1020899
areaN	-.0646773	.0640862
areaS	.1477074	.0687726
_cons	-.5587294	.6188663

rho -.1313338 .4065036

Il risultato saliente è quello relativo al coefficiente di correlazione, che mette in evidenza come la condizione della donna, lavoratrice o meno, non dipenda più di tanto dal fatto di avere partner lavoratore dipendente invece che autonomo; rho è abbastanza basso (in termini di valore assoluto) ed è negativo. Questo vuol dire che la probabilità che la moglie/convivente lavori dipende debolmente, e in modo negativo, dall'essere lavoratore dipendente da parte del partner.

Riporto ora le stime degli effetti marginali che danno la probabilità condizionata e quella marginale.

• *effetti marginali probit con selezione del sotto-campione*

	<i>Pr(moglielnl=1 maritodip=1)</i>		<i>Pr(moglielnl=1)</i>	
	Eff. Marg.	S.E.	Eff. Marg.	S.E.
etaF	.0527693	.01405	.0531994	.01385
etaM	.0027021	.00850	-	-
etasqF	-.0006518	.00018	-.0006567	.00017
etasqM	-.0000357	.00011	-	-
ncomp	-.0046493	.01388	-.0041992	.01384
npred12	-.1251680	.02111	-.1245188	.02170
presc	-.0618084	.02560	-.0570443	.02964
scsupF	.2837323	.02279	.2823950	.02514
scsupM	.0017865	.00565	-	-
univF	.4598778	.02246	.4461312	.05373
univM	.0014250	.00567	-	-
amp40	-.0368113	.02685	-.0362185	.02672
amp500	-.1135579	.04574	-.1188147	.04789
areaN	.0814335	.03024	.0825090	.03002
areaS	-.2677852	.03000	-.2711523	.03014

Rispetto ai risultati statisticamente significativi al 5%, si può mettere in evidenza come la probabilità che la partner di sesso femminile lavori, condizionatamente all'essere lavoratore dipendente da parte del coniuge/convivente, aumenta, di 5 punti percentuali, all'aumentare dell'età anagrafica della donna stessa. Invece al crescere del numero di percettori di reddito o in caso vi siano bambini in età prescolare presenti all'interno della famiglia, la probabilità condizionata cala rispettivamente di oltre 12 e 6 punti. Se la moglie/convivente ha come titolo di studio il diploma di scuola superiore o la laurea, la probabilità condizionata che ha di lavorare aumenta rispettivamente di 28 e di ben 46 punti percentuali. Altri tre risultati statisticamente significativi sono di carattere demografico e geografico. Se la coppia vive in una grande città la probabilità che la donna lavori, condizionatamente al fatto di avere marito/convivente lavoratore dipendente, diminuisce di circa 11 punti percentuali; il calo è invece quantificabile intorno ai 27 punti se la coppia vive nel Sud Italia (o Isole), mentre se l'area geografica di residenza è quella dell'Italia settentrionale la probabilità condizionata di lavorare che ha la donna aumenta di 8 punti percentuali.

Come era per certi versi facile aspettarsi, la probabilità (non condizionata) che la donna lavori cala, di quasi 6 punti percentuali, nel caso vi sia presenza di

bambini in età prescolare in famiglia. Come nel caso degli altri coefficienti, questo risultato è simile a quello restituito dalle probabilità condizionate; era facile aspettarsi questo in quanto il rho basso trovato già aveva messo in evidenza come non abbia più di tanto senso tener conto del condizionamento dell'informazione che dice se il marito/convivente è lavoratore dipendente piuttosto che autonomo sulla condizione della partner di sesso femminile di lavoratrice o meno.

Altri risultati ottenuti dalla stima degli effetti marginali danno la seguente informazione:

$$Pr(\text{moglielnl}=1, \text{maritodip}=1) \quad Pr(\text{moglielnl}=0, \text{maritodip}=1)$$

	Eff. Marg.	S.E.	Eff. Marg.	S.E.
etaF	.0335872	.01209	-.0433315	.01184
etaM	.0195528	.00947	.0156126	.00846
etasqF	-.0004174	.00015	.0005326	.00015
etasqM	-.0002582	.00011	-.0002061	.00010
ncomp	-.0061538	.01123	.0006235	.01121
npred12	-.0905970	.01695	.0918541	.01694
presc	-.0738260	.02059	.0162716	.02057
scsupF	.1939611	.02021	-.2167004	.01865
scsupM	.0129328	.01096	.0103095	.00961
univF	.3160475	.03356	-.3385466	.01720
univM	.0103252	.01901	.0082068	.01349
amp40	-.0291238	.02188	.0246436	.02165
amp500	-.0555298	.03790	.1227345	.03997
areaN	.0483505	.02453	-.0698394	.02427
areaS	-.1742499	.02450	.2225760	.02550

La probabilità che la coppia sia formata da partner di sesso maschile lavoratore dipendente e moglie/convivente che lavora cresce di oltre 3 punti percentuali all'aumentare dell'età della donna, e di quasi 2 al crescere dell'età dei mariti/conviventi (gli effetti marginali di etaF e etaM sono statisticamente significativi al 5%). La stessa probabilità congiunta cala invece di circa 9 punti percentuali all'aumentare del numero di percettori di reddito presenti all'interno della famiglia, di oltre 7 punti se vi sono bambini in età prescolare. Se il partner di sesso femminile è diplomato (scuola superiore o professionale) la probabilità congiunta aumenta di circa 19 punti percentuali, di quasi 32 se invece è laureato. La probabilità che la coppia “maschio lavoratore dipendente – donna che lavora”

si formi aumenta di quasi 5 punti percentuali se vive nel Nord Italia, mentre se la coppia risiede nel Sud o nelle Isole la probabilità congiunta cala di oltre 17 punti. Rispetto agli effetti marginali che risultano statisticamente significativi al 5%, si può affermare che la probabilità che la coppia veda il marito/convivente lavorare come dipendente e la partner di sesso femminile non occupata cala di poco più di 4 punti percentuali all'aumentare dell'età della moglie/convivente; la probabilità congiunta invece aumenta di oltre 9 punti al crescere del numero di percettori di reddito presenti in famiglia. Se il partner di sesso femminile è diplomato o laureato la probabilità che si formi la coppia “maschio lavoratore dipendente – donna non occupata” cala rispettivamente di quasi 22 e 34 punti percentuali. Invece la probabilità che questo tipo di coppia si formi aumenta di oltre 12 e 22 punti percentuali se rispettivamente vive in una grande città o nel Sud Italia (o Isole), mentre se la coppia risiede nel Nord la probabilità congiunta cala di quasi 7 punti.

21 Alcune considerazioni critiche sull'analisi empirica

L'analisi empirica da me realizzata ha studiato se vi sia correlazione tra la scelta dei partner tra lavoro autonomo e dipendente, allo scopo di vedere se le coppie realizzano diversificazione del portafoglio delle due categorie lavorative di appartenenza, cui sono associati diversi gradi di rischio. Così facendo non si è tenuto conto della dipendenza dalla “condizione” di lavoratore da parte dei membri delle coppie, dando per acquisito il fatto che entrambi sono lavoratori.

In particolare si è guardato alle probabilità congiunte che entrambi i partner siano lavoratori autonomi piuttosto che dipendenti, o che il maschio sia dipendente e la moglie autonoma, o viceversa, senza considerare il condizionamento o comunque l'informazione sulla condizione di occupati o meno da parte dei due coniugi/conviventi. Sarebbe stato invece più opportuno non trascurare, in seno a queste probabilità congiunte, la dipendenza dalla condizione di lavoratrice da parte della donna, aspetto più rilevante rispetto alla condizione di lavoratori dei coniugi/conviventi di sesso maschile; infatti, come già detto, le coppie con

partner maschio non occupato sono numericamente troppo poche perché la cosa possa assumere rilievo statistico.

Di fatto si è guardato ad una probabilità congiunta del tipo, ad esempio, Pr (*maschio lavoratore autonomo, femmina lavoratrice autonoma*), e non, come sarebbe stato più corretto, alla probabilità condizionata Pr (*maschio lavoratore autonomo, femmina lavoratrice autonoma | la partner femmina lavora*).

Un modo per tener conto dell'informazione che mi dice che la partner di sesso femminile lavora, è quello di "passare" da un modello probit bivariato con struttura di regressione latente

$$y_F^* = x_{moglie} \beta_{moglie}^{moglie} + x_{marito} \beta_{marito}^{moglie} + \varepsilon_F$$

$$y_M^* = x_{marito} \beta_{marito}^{marito} + x_{moglie} \beta_{moglie}^{marito} + \varepsilon_M$$

ed ove si osservano y_F e y_M tali che

$$y_F = 1 \text{ se } y_F^* > 0, y_F = 0 \text{ altrimenti}$$

$$y_M = 1 \text{ se } y_M^* > 0, y_M = 0 \text{ altrimenti}$$

(y_M e y_F assumono valore 1 quando i coniugi/conviventi sono lavoratori autonomi, 0 se dipendenti) ad un modello probit trivariato che vede aggiungere alla precedente struttura di regressione latente l'equazione

$$L_F^* = x_{moglie} \beta_{moglie}^{moglie} + x_{marito} \beta_{marito}^{moglie} + \varepsilon_L$$

e si osserva $L_F = 1$ se $L_F^* > 0$, $L_F = 0$ altrimenti; L_F assume valore 1 quando la moglie/convivente lavora, 0 se non è occupata.

Un modello trivariato di questo tipo presenta però un problema di osservabilità, dato che, qualora la moglie non lavori, y_M e y_F non sono per l'appunto osservabili.

I tre termini d'errore, presi congiuntamente nel caso di probit trivariato, hanno distribuzione normale standard:

$$\begin{pmatrix} \varepsilon_L \\ \varepsilon_M \\ \varepsilon_F \end{pmatrix} \sim N_3(\underline{0}, \Omega)$$

dove Ω è la matrice di correlazione

$$\begin{bmatrix} 1 & \rho_{LM} & \rho_{LF} \\ \rho_{LM} & 1 & \rho_{MF} \\ \rho_{LF} & \rho_{MF} & 1 \end{bmatrix}$$

Non potendo però realizzare un modello probit trivariato, come detto precedentemente, scelgo di stimare la probabilità condizionata $\Pr(y_M, y_F | L_F)$. Quanto ai termini d'errore, occorre quindi vedere come si distribuisce la funzione $(\varepsilon_M, \varepsilon_F) | \varepsilon_L > 0$. Essa si distribuisce come una *skew normal* bivariata; per capire questo vediamo quando una variabile casuale si distribuisce come una *skew normal*.

Se una variabile casuale Z ha funzione di densità

$$f_2(z; \alpha) = 2 \phi_2(z; \Omega) \Phi(\alpha^T z), \quad (-\infty < z < \infty)$$

con ϕ_2 funzione di densità della normale bivariata standardizzata, e Φ funzione di ripartizione normale standardizzata, allora si può dire che Z è una variabile casuale normale asimmetrica con parametro α che regola il grado di asimmetria in $(-\infty, \infty)$. In generale, come sostenuto da Azzalini e Dalla Valle, dato il vettore

$$\begin{pmatrix} X_0 \\ X \end{pmatrix} \sim N_{k+1}(\underline{0}, \Omega), \quad \text{con matrice di correlazione } \Omega = \begin{pmatrix} 1 & \delta^T \\ \delta & \bar{\Omega} \end{pmatrix} \quad \text{e } X_0$$

componente scalare, allora la variabile casuale $X | X_0 > 0$ si distribuisce come una *skew normal* k -dimensionale $SN_k(\bar{\Omega}, \alpha)$, con α vettore dei parametri di shape

$$\alpha = \frac{\Omega^{-1} \delta}{(1 - \delta^T \Omega^{-1} \delta)^{1/2}}$$

I parametri α dipendono quindi anche dal vettore (di correlazioni) δ .

La matrice di correlazione, nel caso empirico d'interesse, è

$$\Omega = \begin{bmatrix} 1 & \rho_{LM} & \rho_{LF} \\ \rho_{LM} & 1 & \rho_{MF} \\ \rho_{LF} & \rho_{MF} & 1 \end{bmatrix}, \quad \text{e può essere scritta come } \Omega = \begin{pmatrix} 1 & \delta^T \\ \delta & \bar{\Omega} \end{pmatrix}, \quad \text{con il vettore}$$

$$\delta = \begin{bmatrix} \rho_{LM} \\ \rho_{LF} \end{bmatrix} \quad \text{e con la matrice } \bar{\Omega} = \begin{bmatrix} 1 & \rho_{MF} \\ \rho_{MF} & 1 \end{bmatrix}. \quad \text{In tal modo si ottiene}$$

$Z = \begin{pmatrix} \varepsilon_M \\ \varepsilon_F \end{pmatrix} | \varepsilon_L > 0 \sim SN_2(\underline{0}, \bar{\Omega}, \alpha)$; la variabile casuale Z ha funzione di

densità $f_2(z; \alpha) = 2 \phi_2(z; \bar{\Omega}) \Phi(\alpha^T \omega^{-1} z)$, con $\omega = \text{diag}(\sigma_{MM}, \sigma_{FF})$, vettore dei parametri di shape $\alpha = \begin{bmatrix} \alpha_M \\ \alpha_F \end{bmatrix}$, $\phi_2(z; \bar{\Omega})$ funzione di densità della normale

bivariata standardizzata con matrice di correlazione $\bar{\Omega}$, e $\Phi(\cdot)$ funzione di

ripartizione $N(0,1)$. Dopo gli opportuni sviluppi algebrici la funzione di densità può essere scritta nel seguente modo:

$$f_2(z; \alpha) = \frac{1}{2\Pi\sqrt{1-\rho_{MF}^2}} \exp\left(-\frac{1}{2(1-\rho_{MF}^2)} [z_M^2 - 2\rho_{MF}z_Mz_F + z_F^2]\right) * \Phi\left(\frac{\alpha_M z_M}{\sigma_{MM}} + \frac{\alpha_F z_F}{\sigma_{FF}}\right).$$

La funzione di log-verosimiglianza, adattata al caso discreto, è:

$$\ln L = \sum y_M y_F \ln \Pr(y_M=1, y_F=1) + y_M(1-y_F) \ln \Pr(y_M=1, y_F=0) + \\ + (1-y_M)y_F \ln \Pr(y_M=0, y_F=1) + (1-y_M)(1-y_F) \ln \Pr(y_M=0, y_F=0),$$

con, dato che $y_{M,F} = 1$ se $\varepsilon_{M,F} > -x_{M,F} \cdot \beta_{M,F}$ e $y_{M,F} = 0$ altrimenti, funzioni di ripartizione :

$$\Pr(y_M=0, y_F=0) = \frac{1}{2\Pi\sqrt{1-\rho_{MF}^2}} \Phi\left(\frac{\alpha_M x_M \beta_M}{\sigma_{MM}} + \frac{\alpha_F x_F \beta_F}{\sigma_{FF}}\right) *$$

$$* \int_{-\infty}^{-x_M \beta_M} \int_{-\infty}^{-x_F \beta_F} \exp\left(-\frac{1}{2(1-\rho_{MF}^2)} [t_M^2 - 2\rho_{MF}t_M t_F + t_F^2]\right) dt_F dt_M$$

$$\Pr(y_M=0, y_F=1) = \frac{1}{2\Pi\sqrt{1-\rho_{MF}^2}} \Phi\left(\frac{\alpha_M x_M \beta_M}{\sigma_{MM}} + \frac{\alpha_F x_F \beta_F}{\sigma_{FF}}\right) *$$

$$* \int_{-\infty}^{-x_M \beta_M} \int_{-x_F \beta_F}^{+\infty} \exp\left(-\frac{1}{2(1-\rho_{MF}^2)} [t_M^2 - 2\rho_{MF}t_M t_F + t_F^2]\right) dt_F dt_M$$

$$\Pr(y_M=1, y_F=0) = \frac{1}{2\Pi\sqrt{1-\rho_{MF}^2}} \Phi\left(\frac{\alpha_M x_M \beta_M}{\sigma_{MM}} + \frac{\alpha_F x_F \beta_F}{\sigma_{FF}}\right) *$$

$$* \int_{-x_M \beta_M}^{+\infty} \int_{-\infty}^{-x_F \beta_F} \exp\left(-\frac{1}{2(1-\rho_{MF}^2)} [t_M^2 - 2\rho_{MF}t_M t_F + t_F^2]\right) dt_F dt_M$$

$$\Pr(y_M=1, y_F=1) = \frac{1}{2\Pi\sqrt{1-\rho_{MF}^2}} \Phi\left(\frac{\alpha_M x_M \beta_M}{\sigma_{MM}} + \frac{\alpha_F x_F \beta_F}{\sigma_{FF}}\right) *$$

$$* \int_{-x_M \beta_M}^{+\infty} \int_{-x_F \beta_F}^{+\infty} \exp\left(-\frac{1}{2(1-\rho_{MF}^2)} [t_M^2 - 2\rho_{MF}t_M t_F + t_F^2]\right) dt_F dt_M$$

Detto ciò il passo successivo sarebbe quello di calcolare la funzione di log-verosimiglianza, e massimizzarla; in questo modo si possono stimare i parametri β ed è possibile ottenere gli effetti marginali sulle probabilità congiunte in cui si tiene conto della dipendenza dalla condizione di lavoratrici da parte delle donne.

Conclusioni

A livello di risultati empirici, i vari modelli probit utilizzati, sia quando si distingue tra lavoro dipendente e autonomo, sia quando si fa l'ulteriore differenziazione tra lavoro dipendente pubblico e privato, concordano su come alcune caratteristiche influenzino la "scelta" degli individui, appartenenti a coppie di partner entrambi lavoratori, su quale grado di rischio associato alle categorie occupazionali accettare.

In generale è più probabile che un individuo sia lavoratore autonomo e non dipendente, e quindi accetti più rischio, se ha bambini in età prescolare, con la probabilità che aumenta di circa 3 punti percentuali; probabilmente ciò è spiegabile in quanto si viene a cogliere un aspetto anagrafico. Infatti chi ha bambini piccoli è presumibilmente relativamente giovane e quindi più propenso ad accettare alti rischi. La probabilità invece cala di circa 7 punti se la persona è di sesso femminile, di 6 e 8 punti se ha un titolo di studio non dell'obbligo (rispettivamente diploma e laurea), di 9 punti se vive in una città con oltre 500mila abitanti. L'incidenza negativa del possesso di un titolo di studio superiore alla scuola dell'obbligo è maggiore per le femmine piuttosto che per i maschi. Un risultato particolarmente rilevante è quello che mette in evidenza come se il partner di un individuo è lavoratore autonomo aumenta sensibilmente (di circa 37 punti percentuali) la probabilità che anch'esso "scelga" la categoria occupazionale meno sicura; ciò vale in misura ancor maggiore se l'individuo in questione è di sesso maschile, quando la probabilità di essere autonomo, se la partner è anch'essa lavoratrice autonoma, aumenta di ben 46 punti percentuali.

Un'ulteriore distinzione realizzata è quella tra lavoro dipendente pubblico e privato, oltre alla categoria occupazionale più rischiosa, quella del lavoro autonomo. La scelta delle tre opzioni e dell'ordinamento crescente a livello di rischio associato dip. pubbl. – dip. priv. – lav. autonomo, si è rivelata opportuna; infatti le distanze tra i cut-off ne confermano l'adeguatezza, ed inoltre diversi coefficienti stimati guadagnano in significatività quando si considerano le tre opzioni.

A livello riassuntivo risulta che la probabilità che un individuo sia lavoratore dipendente privato e non pubblico, o che sia lavoratore autonomo invece che dipendente privato, cresce di circa 3 punti percentuali all'aumentare del numero di percettori di reddito presenti in famiglia o se vi siano bambini in età prescolare (probabilmente perché in tal caso i partner sono relativamente giovani); si è invece meno propensi ad accettare rischi più elevati se si è di sesso femminile (la probabilità cala di circa 8 punti percentuali), all'aumentare dell'età anagrafica (2 punti), se si appartiene a famiglie numerose (2 punti), o se si vive nel Sud Italia o nelle Isole (8 punti). Un risultato in parte nuovo che emerge rispetto a quelli ottenuti con scelta binaria è quello relativo all'incidenza del possesso di un titolo di studio non dell'obbligo; infatti l'incidenza sulla probabilità di "scegliere" occupazioni meno sicure è sì ancora negativa, ma se nel caso di scelta binaria la probabilità calava di più punti percentuali in caso di diploma piuttosto che se laureati, con le tre opzioni all'aumentare del titolo di studio la misura dell'incidenza negativa cresce, passando all'incirca dai 12 ai 16 punti percentuali.

L'obiettivo principale dell'analisi empirica era studiare se, all'interno delle coppie con tutti e due i partner occupati, a prevalere sia una situazione che veda entrambi essere lavoratori autonomi piuttosto che dipendenti pubblici o privati, o se invece vi sia effettivamente diversificazione del portafoglio delle categorie lavorative, più o meno rischiose, di appartenenza.

Nel caso di scelta binaria, i modelli bivariati, costruiti in modo da mettere in evidenza tutta una serie di risultati che riguardano l'influenza che le caratteristiche individuali dei coniugi/conviventi hanno sulla scelta occupazionale a livello di rischio dei partner, restituiscono coefficienti di correlazione sempre intorno a 0.6. Questo vuol dire che vi è alta correlazione tra la "scelta" occupazionale, a livello di rischio accettato, di mariti e mogli, o conviventi. Sempre nel caso di scelta binaria, il modello ad effetti casuali stabilisce che il 64% circa dell'eterogeneità della variabile latente che dà il livello di avversione individuale al rischio è spiegata dall'effetto coppia. Così, dato che tale effetto spiega circa il 64% della varianza non osservata, e dato che si basa sulle correlazioni attraverso la variabile dipendente nelle coppie, i risultati vanno

a confermare quanto evidenziato dal modello bivariato. Quando si studia la “scelta” tra le tre opzioni del lavoro dipendente pubblico, privato, e del lavoro autonomo, il ρ ottenuto col modello ad effetti casuali è altamente significativo, e l’ampiezza suggerisce che il 48% circa dell’eterogeneità nell’avversione individuale al rischio è spiegata dall’effetto coppia. Così l’ampiezza del ρ , seppur minore rispetto a quella ottenuta col modello ad effetti casuali con scelta binaria, ma di simile significatività, suggerisce come uno specifico effetto di coppia sia ancora apprezzabile.

Si può quindi concludere che sia il coefficiente della variabile indipendente che indica il tipo di lavoro del partner (utilizzato nei modelli standard e ordinali semplici quando si raffrontano diversi sotto-campioni), che i coefficienti di correlazione restituiti dai modelli ad effetti casuali e nel bivariato, non danno supporto alla presenza di diversificazione delle categorie occupazionali, con associati diversi rischi, intra-coppia nel contesto italiano. Del resto già una prima lettura dei dati dell’analisi, nel capitolo 3, aveva evidenziato come l’incidenza delle coppie ove i partner siano occupati in lavori appartenenti alla stessa categoria di rischio fosse ben marcata. Ciò fa pensare che i coniugi/conviventi diano maggior peso ad altri aspetti piuttosto che ai vantaggi che derivano da una diversificazione del portafoglio delle categorie di appartenenza con associati determinati livelli di rischio. Come sostengono Sarah Brown, Lisa Farrell e John G. Sessions in *Employment Contract Matching*, può essere che sia più facile trovare un partner con cui si sia della stessa opinione e si vada d’accordo se quest’ultimo fa il nostro stesso tipo di lavoro; può anche essere che l’opportunità di amplificare i guadagni da lavoro autonomo, o comunque l’esigenza di coinvolgere il partner in questo tipo di attività, prevalga sui benefici da pooling dei rischi.

L’ultimo sotto-paragrafo del capitolo amplia il sotto-campione di riferimento, con l’obiettivo di vedere se e come la condizione delle donne di lavoratrici o non occupate influenzi e sia influenzata dalla scelta dei partner tra lavoro autonomo e dipendente. Un primo risultato che emerge è come l’eventuale “scelta” dei partner maschi di essere lavoratori autonomi non dipenda dal fatto che la moglie/convivente lavora, visto che vi è bassa correlazione tra la potenziale

condizione di lavoratrice della donna e la possibilità che il marito/convivente sia lavoratore autonomo piuttosto che dipendente. Inoltre si valuta quali caratteristiche incidono, e in che misura, sulla scelta della donna se lavorare o meno condizionatamente all'essere lavoratore dipendente da parte del coniuge/convivente; tra le altre cose, emerge come l'età e il titolo di studio dei partner maschi non siano statisticamente significativi rispetto alla probabilità condizionata che le mogli/conviventi lavorino. Del resto l'eventualità che vi sia "forte" risk sharing intra-coppia, con la donna non occupata e quindi maggiormente esposta al rischio rispetto al partner, non dipende neanche dalla "scelta" occupazionale dei coniugi/conviventi maschi.

Bibliografia

Arulampalam W. (1999) 'A Note on Estimated Effects in Random Effect Probit models', in *Oxford Bulletin of Economics and Statistics*, volume 61, pp. 597-602.

Azzalini A. e A. Capitanio (1999) 'Statistical Applications of the Multivariate Skew Normal Distribution', in *Journal of the Royal Statistical Society. Series B*, vol 61 (parte 3), pp.579-602.

Azzalini A. e A. Dalla Valle (1996) 'The Multivariate Skew-normal Distribution', in *Biometrika*, vol. 83(4), pp. 715-726.

Bardasi E. e C. Monfardini (1997) 'The Choise of the Working Sector in Italy: a Trivariate Probit Analysis'.

Behrman J. R. (1997) 'Intrahousehold Distribution and the Family', in *Handbook of Population and Family Economics (a cura di Mark R. Rosenweig e Oded Stark)*, vol. 1A, pp. 125-187.

Bergstrom T. C. (1997) , 'A Survey of Theories of the Family', in *Handbook of Population and Family Economics*, vol. 1A, pp. 21-79.

Blanchflower D. G. e A. J. Oswald (1998) 'What Makes an Entrepreneur?', in *Journal of Labour Economics*, vol. 16 (1), pp. 26-60.

Blundell R. e M. Thomas (2000) 'Labour Supply: a Review of Alternative Approaches', in *Handbook of Labour Economics (a cura di O. Ashenfelter e R. Layard)*, vol. 2, pp. 1559-1695.

Brown S., L. Farrell e J. G. Sessions (2002) 'Employment Contract Matching: An Analysis of Dual Earner Couples and Working Households'.

Guiso L. e T. Jappelli (1994) 'Risk Sharing and Precautionary Saving'.

Hayashi F., J. Altonji e L. Kotlikoff (1997) 'Risk-sharing Between and Within Families', in *Understanding Saving: Evidence from the United States and Japan* (a cura di F. Hayashi), pp. 261-294.

Hyslop D. R. (2001) 'Rising U.S. Earnings Inequality and Family Labour Supply: The Covariance Structure of Intrafamily Earnings', in *The American Economic review* (trimestrale a cura dell'American Economic Association), vol. 91(4), pp.755-777.

Taylor M. P. (1996) 'Earnings, Indipendance or Unemployment: Why Become Self-employed?', in *Oxford Bulletin of Economics and Statistics*, vol. 30, pp. 194-204.

Verbeek M. (2000) 'Models with Limited Dependent Variables' e 'Models Based on Panel Data', in *A Guide to Modern Econometrics*, pp. 177-223 e pp. 809-350.

Ringraziamenti

Ringrazio di tutto cuore mio padre che tanto ha contribuito al raggiungimento di questo mio traguardo e che avrebbe tanto voluto vedermi laureato; in qualche modo so che mi è vicino anche in questo momento. Un grazie immenso va a mia madre e a mio fratello che, nonostante il periodo difficile che sta attraversando la nostra famiglia, mi sono stati tanto d'aiuto e di supporto, con pazienza (fin troppa!) e affetto.

Ringrazio i miei amici dell'università, mia cugina Marisa per le preziose lezioni "informatiche", e Marianna per aver supportato mio fratello nei momenti in cui diventavo insopportabile.

Un grazie particolare va al professor Miniaci, che in questi mesi mi ha seguito con costanza e pazienza, sempre disponibile a darmi consigli utili e "dritte" efficaci.