



DEPARTMENT OF ECONOMICS
UNIVERSITY OF MILAN - BICOCCA

WORKING PAPER SERIES

**Investire in istruzione: meglio per lui o per lei?
Stima per genere dei rendimenti dell'istruzione in Italia**

Luca Flabbi

No. 08 - September 1997

Dipartimento di Economia Politica
Università degli Studi di Milano - Bicocca
<http://dipeco.economia.unimib.it>

INVESTIRE IN ISTRUZIONE: MEGLIO PER LUI O PER LEI? Stima per genere dei rendimenti dell'istruzione in Italia.[†]

Luca FLABBI*

(Università degli Studi di Milano - Bicocca)

20 Agosto '97

Previous estimates of Italian returns to schooling suggest that female returns are higher than male returns (Lucifora-Sestito '93). This is a common result in other industrialized countries, as well (Blau-Kahn '92 e Psacharopoulos '94).

The aim of the paper is to evaluate the stability of this hierarchy by changing specifications and estimation techniques. All the estimates are performed on the Bank of Italy *Indagine sui bilanci della famiglie nell'anno 1991*.

Results with standard techniques (i.e. ordinary least square and Heckman's correction) confirm usual hierarchy, on the contrary alternative techniques (i.e. instrumental variables) suggest an opposite hierarchy. This is a robust conclusion by varying samples, instruments, and schooling measurements.

Key words: human capital, returns to education, earnings

J.E.L. classification: J 31

[†] Questo lavoro è sviluppato partendo da un più ampio lavoro di tesi in cui sono stato seguito da Andrea Ichino, Francesco Giavazzi e Carlo Favero che colgo l'occasione per ringraziare. In particolare Andrea Ichino è stato, ed è tuttora, un costante aiuto ed incoraggiamento per il mio lavoro. Ringrazio anche i partecipanti all'XI Convegno dell'AIEL, dove una precedente versione di questo articolo è stata presentata. Infine, fondamentale per il compimento del lavoro è stata la possibilità di utilizzare le strutture dell'IGIER.

* Luca Flabbi
Università degli Studi di Milano - Bicocca
Facoltà di Economia
Piazza dell'Ateneo Nuovo, 1 - 20126 Milano

1 - Introduzione.

L'istruzione è considerata una determinante significativa dei salari in quanto fattore produttivo fondamentale nella composizione del capitale umano individuale. È allora interessante valutare il rendimento generato dall'investimento in questo particolare fattore produttivo.

L'attenzione di questo lavoro si focalizza sulle differenze di rendimento in base al genere. In particolare l'obiettivo è chiarire se l'usuale gerarchia, ovvero rendimenti femminili superiori a quelli maschili¹, possa essere considerata un risultato empirico acquisito nel mercato del lavoro italiano o non dipenda invece dalle tecniche di stima utilizzate. C'è infatti la possibilità che la distorsione delle stime con minimi quadrati ordinari, ovvero la procedura applicata dalla maggior parte dei lavori empirici che stimano rendimenti separati uomo-donna nel caso italiano², sia tale da nascondere il "vero" rapporto relativo tra i rendimenti.

Per rispondere a questa domanda sono state attuate stime con variabili strumentali, dove gli indicatori sono ispirati alla metodologia degli esperimenti naturali e sono del tutto inediti per il caso italiano. A titolo di confronto sono state applicate anche le usuali procedure OLS e con correzione di Heckman. I dati sono derivati dall'Indagine Banca d'Italia sui bilanci delle famiglie italiane nell'anno 1991.

Il paragrafo seguente contiene la specificazione e i primi risultati ottenuti con tecniche di stima standard. Il terzo discute brevemente le distorsioni che colpiscono questo tipo di stime e propone la soluzione proposta dalla metodologia degli esperimenti naturali. Infine il quarto applica quest'ultima procedura per ottenere rendimenti dell'istruzione maschili e femminili.

2 - Stime Standard.

Specificazione e dati.

Per ottenere i rendimenti dell'istruzione è stata stimata una classica funzione del salario, ovvero una regressione di salari o redditi individuali su un vettore di caratteristiche personali quali

¹ Per l'Italia una rassegna è contenuta in Lucifora-Sestito '93, per altri paesi industrializzati si veda invece Psacharopoulos '94 e Blau-Kahn '92.

² E' ad esempio questa la procedura di tutti i lavori riportati in Lucifora-Sestito '93, nonché quella del campione che ho raccolto nella tabella 2.2.

l'istruzione, l'esperienza lavorativa, le variabili che compongono il background sociale e familiare (Willis '86; Lucifora-Sestito '93).

La specificazione, come riportato nella tabella 2.1, contiene le classiche variabili di capitale umano, istruzione ed esperienza (Mincer '74, Becker '93), e alcuni controlli relativi al tipo di occupazione dell'individuo, al mercato del lavoro in cui si trova e alla sua famiglia di appartenenza. I controlli per occupazione e mercato del lavoro permettono di considerare fattori di domanda (Lucifora-Sestito '93) e di ridurre le distorsioni dei rendimenti dell'istruzione associate ad un funzionamento dei mercati non perfettamente concorrenziale³. I controlli per la famiglia di appartenenza permettono di cogliere l'influenza del lavoro di cura e riproduzione, particolarmente importante nell'offerta di lavoro femminile (Heckman-Killingsworth '86), e le dinamiche dell'*assortative mating* nella scelta del partner (Becker '91; Lam-Schoeni '93).

La banca dati da cui ho estratto il mio universo di stima è l'*Indagine dei bilanci delle famiglie italiane nell'anno 1991*, condotta per conto della Banca d'Italia. Si tratta di un campione rappresentativo della popolazione italiana che ha come oggetto soprattutto i comportamenti finanziari degli individui, ma che contiene anche i redditi e le principali caratteristiche personali. Dal campione totale disponibile ho tratto un sottocampione di 5.734 individui che contiene essenzialmente tutti i lavoratori dipendenti che abbiano lavorato a tempo pieno per tutto l'anno⁴.

Risultati.

La tabella 2.2. riporta i rendimenti dell'istruzione ottenuti applicando al campione precedente le usuali stime con minimi quadrati ordinari e correzione di Heckman. La correzione di Heckman è introdotta per risolvere la distorsione da autoselezione che coinvolge il campione di donne che offrono lavoro⁵. I risultati completi delle stime, con alcuni commenti riguardo le altre variabili contenute nella specificazione, sono riportati nell'appendice 2.

³ Sotto opportune ipotesi (Lucifora-Sestito '93), e secondo un'interpretazione strutturale del modello, il capitale umano di ogni individuo dovrebbe avere pari rendimento indipendentemente da settore, impresa ed occupazione in cui viene impiegato. Nella realtà osserviamo settori che pagano salari mediamente più alti nonché marcate differenze a seconda del mercato del lavoro in cui l'occupazione è inserita (Bettio '88 e '91, Lucifora-Reilly '90). Introdurre questi controlli può quindi assorbire le distorsioni associate a questo tipo di problemi.

⁴ I passaggi fatti per ottenere il campione, indicando le motivazioni e il costo in termini di osservazioni eliminate, sono riportati nell'appendice 1.

⁵ Per una rassegna, anche se riferita soprattutto a funzioni di offerta di lavoro, si veda Heckman-Killingsworth '86. Per lavori che invece applicano la procedura standard OLS e correzione di Heckman a funzioni del salario separate per uomini e donne, si veda ad es. Callan '91 e Wright e Ermish '90.

La stima OLS *ampia*, ovvero la stima con minimi quadrati ordinari che comprende tutte le variabili della specificazione della tabella 2.1, conferma l'usuale gerarchia dei rendimenti dell'istruzione tra generi, cioè rendimenti femminili superiori a quelli maschili. Questo è anche il risultato ottenuto dalla maggior parte dei lavori empirici applicati a campioni rappresentativi dell'economia italiana, come si può notare nella parte bassa della tabella. La stessa gerarchia è confermata da lavori analoghi applicati ad altri paesi industrializzati (Psacharopoulos '94, Blau-Kahn '92) e a paesi in via di sviluppo (Appleton-Hoddinott-Krishnan-Max '95).

Per applicare la correzione di Heckman la specificazione è stata ridotta perché alcune variabili non sono disponibili per le donne non lavoratrici oppure sono ipotizzate influenzare solo l'offerta di lavoro. Le variabili non disponibili sono i settori, le qualifiche e la tenure, mentre gli anni d'esperienza sono stati approssimati tramite l'esperienza potenziale. Le variabili che ho ipotizzato influenzino l'offerta di lavoro ma non il salario, ovvero che ho inserito nella stima probit dell'offerta di lavoro ma non nella regressione sui salari, sono principalmente gli indicatori di età e presenza figli. La specificazione ottenuta, che ho chiamato OLS *stretta*, riporta gerarchie invertite: un anno d'istruzione addizionale frutta un aumento salariale pari al 3,9% per gli uomini e al 3,1% per le donne. Questa differenza aumenta quando la stima della funzione femminile è attuata tramite correzione di Heckman, infatti in questo caso il rendimento scende al 2,8%.

Una possibile spiegazione per questa inversione può essere trovata nella differente specificazione tra stima OLS *ampia* e stima OLS *stretta*. La differenza qualificante è che nel secondo caso sono esclusi i controlli occupazionali, rappresentati invece nel primo caso con dummies per settori e qualifiche. La stima OLS *stretta* risente quindi di distorsione da omissione di variabili che agisce in modo asimmetrico sui due campioni maschili e femminili. In particolare la distorsione sembra avere un effetto maggiore sul campione maschile implicando un superiore innalzamento del loro coefficiente che diventa così superiore a quello femminile.

Questa interpretazione porta però a evidenziare solo una possibile distorsione: quella da omissione di variabili. In realtà i problemi di distorsione legati alle tecniche di stima fin qui utilizzate sono più ampi e complessi e giustificano l'utilizzo di procedure alternative.

3 - Il Problema: la Distorsione delle Stime.

A partire dal lavoro di Griliches del '77, le stime OLS dei rendimenti dell'istruzione hanno cominciato ad essere considerate distorte e inconsistenti. Il loro calcolo non è stato però abbandonato nella speranza di determinare segno ed entità della distorsione.

Su questo punto tuttavia le risposte non sono state univoche. Griliches sosteneva la possibilità sia di distorsione verso l'alto che di distorsione verso il basso, contestando l'ipotesi allora più accreditata di una sovrastima del coefficiente da parte dei minimi quadrati ordinari. La possibilità di una distorsione verso l'alto è rimasta comunque l'opinione più diffusa (Ehrenberg e Smith '91), anche se per alcuni autori questa si verifica solo nella maturità della vita lavorativa (Blackburn e Neumark '93). Le stime che utilizzano variabili strumentali e modelli a effetti fissi trovano, invece, distorsioni verso il basso⁶ che si fanno più marcate nel caso della metodologia dei cosiddetti *esperimenti naturali*⁷.

Questa indecisione dei risultati empirici riflette la teoria econometrica alla base del problema: variabili omesse, errori di misurazione, problemi di endogeneità implicano distorsioni di segno opposto, senza permettere di decidere a priori quale prevarrà. Il senso globale di questo discorso teorico può essere visto come un problema di endogeneità della variabile istruzione nella regressione dei salari. Per spiegarlo, seguendo Card '93, si può descrivere la specificazione della funzione del salario con un sistema a due equazioni:

$$[1] \quad S_i = \alpha H_i + v_i$$

$$[2] \quad \log w_i = \beta H_i + \gamma S_i + \mu_i,$$

dove H è un vettore di caratteristiche dell'individuo i^8 , S_i sono gli anni d'istruzione completati e w_i è il salario. L'endogeneità della variabile S deriva da una sua correlazione con μ_i , ovvero da $\text{Cov}(v_i, \mu_i) \neq 0$. In questo caso la stima OLS di γ , cioè il rendimento marginale dell'istruzione, è distorta e inconsistente. La correlazione che porta ad endogeneità può essere causata da (Card '93; Griliches '77):

⁶ Griliches, Hall e Hausman '78 utilizzano variabili di background familiare come strumenti per l'istruzione e stimano un coefficiente doppio rispetto alla stima OLS. Angrist e Newey '91, a fronte di un coefficiente OLS pari a 0.036, trovano un coefficiente stimato con effetti fissi pari a 0.080.

⁷ Ad esempio i cinque lavori presi in considerazione da Card '94 vedono un aumento dei rendimenti stimati con IV compreso tra il 10% e il 100%.

⁸ Si ipotizza $E(H_i v_i) = E(H_i \mu_i) = 0$, ovvero che l'unica variabile correlata coi residui nella regressione sia l'istruzione.

- variabili omesse,
- errori di misurazione,
- disomogeneità dei rendimenti all'interno della popolazione.

La distorsione da variabili omesse si riferisce essenzialmente all'omissione di variabili relative ai *background famigliari e sociali* e al concetto di *abilità* individuale. Questo tipo di variabili dovrebbero cogliere i fattori che permettono di ottenere un salario più elevato indipendentemente dalle caratteristiche di capitale umano (istruzione, esperienza, ecc.) e dai controlli (mercato del lavoro, famiglia di appartenenza, ecc.) presenti tra i regressori della funzione del salario.

Consideriamo ad esempio i talenti e le caratteristiche individuali che possiamo riassumere con il concetto di *abilità* ed indichiamoli con A . La loro omissione significa che:

$$[3] \quad \mu_i = \delta A_i + \varepsilon_i .$$

Ma l'abilità influenzerà anche gli anni di scolarità, avremo quindi:

$$[4] \quad v_i = \lambda A_i + \eta_i .$$

In conclusione $\text{Cov}(v_i, \mu_i) \neq 0$, istruzione endogena, OLS distorto. Il segno della distorsione dipende dal segno di λ e δ . δ è per definizione positivo, λ ha invece segno incerto (Cannari-D'Alessio '95). Può essere positivo se individui più abili decidono di completare un maggior numero di anni d'istruzione, ad esempio per *segnalare* la propria abilità ai datori di lavoro. Può essere negativo se i medesimi individui decidono di terminare prima i propri studi, ad esempio a causa del maggior costo opportunità da loro affrontato.

Più univoco sembra invece valutare l'effetto dei background: una famiglia benestante e di cultura elevata che vive in un quartiere non degradato ha influenza positiva sia sul livello d'istruzione che sul reddito del figlio⁹. Ciò che causa endogeneità in questo caso è la possibilità che il background possa influire sul reddito indipendentemente dalla scolarità. Questo può accadere ad esempio per trasmissione "involontaria" di capitale umano o a causa dei processi di *networking* (Montgomery '91).

La distorsione da errori di misurazione è causata da errori nel rilevare gli anni d'istruzione. Se questo accade la variabile S che noi osserviamo è uguale all'effettivo valore S^* a meno di un errore τ

⁹ Una rassegna dei lavori empirici sugli Stati Uniti è in Corcoran-Gordon-Laren-Solon '90. Lavori italiani che confermano questa intuizione sono Checchi-Ichino-Rustichini '94 e '96.

$\sim N(0, \sigma_1^2)$. Se ipotizziamo σ_1^2 positiva la stima OLS di γ sarà distorta e inconsistente. L'effetto della distorsione sarà di *attenuazione*, ovvero il coefficiente tenderà a zero al crescere delle varianze σ_1^2 . In questo caso, quindi, il segno della distorsione è determinato, ma la sua entità dipende dalla varianza dell'errore e dalla specificazione utilizzata¹⁰.

Infine anche la variabilità dei rendimenti dell'istruzione all'interno della popolazione può essere causa di distorsione. Infatti in questo caso il "vero" modello sarà:

$$[5] \quad \log w_i = \beta H_i + \gamma_i S_i + \pi_i \quad \text{con} \quad \gamma_i \neq \gamma_j \quad \text{per ogni } i \neq j.$$

Il modello stimato sarà allora:

$$[6] \quad \log w_i = \beta H_i + E(\gamma) S_i + \mu_i \quad \text{dove} \quad \mu_i = \pi_i + S_i(\gamma_i - \gamma),$$

implicando quindi una stima distorta del coefficiente.

Seguendo l'argomentazione contenuta in Card '94, due motivi possono causare la variabilità del rendimento marginale dell'istruzione. Il primo deriva dall'esistenza di caratteristiche che rendono eterogenei gli individui riguardo la redditività dell'investimento in istruzione (*distorsione da abilità*). Il secondo si basa sull'ipotesi di rendimenti marginali decrescenti dell'istruzione: nell'ambito del classico modello di school-choice di Becker '67, a cui Card fa riferimento, la quantità ottima d'istruzione S^* è tale che $\gamma_i = r_i$, dove r_i è il tasso di sconto affrontato dall'individuo per finanziare l'istruzione, ovvero il suo costo marginale. Se i mercati sono imperfetti, individui provenienti da famiglie più povere affronteranno r_i più elevati e termineranno gli studi in corrispondenza di S^* inferiori (*distorsione da tassi*). La parte di popolazione con inferiore istruzione sarà quindi composta sia da individui con bassi rendimenti (meno abili) sia da individui che affrontano tassi più elevati (più poveri). In conclusione: se prevalgono i primi avremo una correlazione positiva tra scolarità ed errori nella funzione del salario e quindi γ_{OLS} distorto verso l'alto, se prevalgono i secondi accadrà l'inverso e la distorsione sarà verso il basso.

Le prime soluzioni per risolvere i problemi di distorsione precedenti si sono focalizzate sulla distorsione dovuta ad omissione della variabile *abilità* cercando di costruire proxy della variabile e di inserirle nella regressione. Soprattutto i lavori relativi al Nord America hanno potuto cercare di

¹⁰ La rassegna presentata da Griliches '77 presenta valori compresi tra il 10 e 15 per cento a seconda delle covariate inserite nel modello.

identificare questa variabile con i test di quoziente intellettivo attuati nelle scuole dell'obbligo. Tuttavia, dubbi sono stati sollevati sul potere esplicativo di questi test¹¹ e la scarsità di dati rende difficile applicare questa correzione al caso italiano.

Altra soluzione è stata quella di costruire banche dati più precise e complete¹² in modo da ridurre errori di misura e distorsioni da omissione di variabili.

Infine l'impostazione che ha cercato di affrontare forse nel modo più completo il problema dell'endogeneità è quella che utilizza variabili strumentali per gli anni d'istruzione conseguiti. La soluzione è applicabile se esistono componenti del vettore H_i che hanno effetto sull'istruzione ma non sul salario. Chiamiamo Z queste componenti e X le rimanenti componenti del vettore H . In questo modo il sistema precedente diventa:

$$[1a] \quad S_i = \alpha X_i + \delta Z_i + v_i$$

$$[2a] \quad \log w_i = \beta X_i + \gamma S_i + \mu_i .$$

Il metodo delle variabili strumentali consiste proprio nel trovare queste componenti Z e nell'utilizzarle per ottenere una stima consistente e non distorta di γ . Per vederlo concentriamoci sulla [2a] e scriviamola come:

$$[2b] \quad \log w_i = \lambda Q_i + \mu_i \quad \text{dove: } \lambda = [\beta \ \gamma] \text{ e } Q_i = [X_i \ S_i]' .$$

In questa espressione si ha $\text{Cov}(Q, \mu) \neq 0$ che discende dall'endogeneità della sua componente istruzione, ovvero da $\text{Cov}(S, \mu) \neq 0$. Questo implica che lo stimatore OLS, pari a:

¹¹ Griliches '77 riporta il dibattito. Le critiche più recenti sottolineano che i risultati dei test, enfatizzando solo alcune capacità logiche, sono in realtà dipendenti dalla formazione familiare e sociale ed hanno quindi poco a che fare con ciò che vogliono effettivamente misurare.

¹² Gli Stati Uniti sono il paese che ha guidato questi sviluppi soprattutto tramite due data set: PSID (*Panel Study of Income Dynamics*) e NLS (*National Longitudinal Survey*). Si tratta di due banche dati realizzate rispettivamente dalle Università del Michigan e dell'Ohio e che raccolgono dalla fine degli anni sessanta dati individuali longitudinali e che contengono numerose informazioni sulla famiglia d'origine. Molti altri paesi hanno realizzato data set con queste caratteristiche, cioè in grado di seguire gli individui per più anni consecutivi ed in modo da avere informazioni più puntuali ed aggiornate rispetto a quelle di un censimento. Un esempio è il PNAD condotto dall'*Istituto Brasileiro de Geografia e Estatística* e che comprende un campione di 100mila famiglie brasiliane. Un data set individuale dettagliato per quanto riguarda l'Italia è quello utilizzato in questo lavoro, ovvero l'*Indagine sui bilanci delle Famiglie Italiane* condotto per conto della Banca d'Italia: ha però il difetto di non avere sufficienti osservazioni longitudinali. Questa lacuna è risolta nell'*Italian Social Mobility Survey* (condotto da M. Barbagli, V. Capecchi, A. Cobalti, A. de Lillo, A. Schizzerotto) che ha però lo svantaggio di non riportare redditi e salari (per l'utilizzo congiunto di entrambi i data set italiani, si veda Checchi-Ichino-Rustichini '94).

$$[7] \quad \lambda_{OLS} = \lambda + (Q'Q)^{-1}Q'\mu ,$$

è distorto perché il secondo termine di destra è diverso da zero¹³. Il metodo delle variabili strumentali è applicabile se esiste una matrice W tale che $Cov(W,\mu) = 0$ e $Cov(W,S) \neq 0$. Se questo vale, allora è possibile ottenere il seguente stimatore:

$$[8] \quad \lambda_{IV} = \lambda + (W'Q)^{-1}W'\mu ,$$

che sarà non distorto e consistente.

Le prime proposte in quest'ottica (Griliches-Hall-Hausman '78), tutt'ora applicate (Cannari e D'Alessio '95), utilizzano come strumenti della variabile istruzione una o più variabili dedotte dal background della famiglia d'origine. Queste variabili sono considerate esogene perché il modello teorico adottato non prevede influenzino direttamente il reddito. Il modello prevede però che abbiano un effetto diretto sulla quantità e qualità dell'istruzione¹⁴. Sotto queste ipotesi possono quindi essere considerate strumenti legittimi.

Le critiche a questo approccio si sono sviluppate, sia sul piano teorico che su quello empirico, soprattutto all'interno della letteratura che studia la mobilità intergenerazionale. Il modello di Becker e Tomes '86, ad esempio, pur mantenendosi nell'ambito della teoria del capitale umano, specifica una funzione del salario che comprende anche variabili relative alla famiglia d'origine come conseguenza della massimizzazione del benessere dei figli compiuta dai genitori. Il modello di Montgomery '91 interpreta questa possibilità alla luce del *networking*, ovvero dei canali sociali privilegiati che permettono alle famiglie più abbienti di trovare ai propri figli un'occupazione proporzionata alla propria origine. I risultati empirici sembrano confermare queste impostazioni¹⁵: le variabili di background direttamente inserite nella funzione del salario sono altamente significative anche se la variabile anni d'istruzione è presente tra i regressori. In conclusione sembra criticabile ipotizzare che

¹³ Infatti per l'ipotesi di residui a media nulla, e a causa dell'endogeneità, abbiamo $E(Q'\mu) = Cov(Q,\mu) \neq 0$.

¹⁴ Per "modello" intendo le varie elaborazioni della teoria del capitale umano utilizzate in questa letteratura. Il background influisce sulla quantità di scolarità conseguita perché influisce sul suo rendimento (Becker '67 e '93).

¹⁵ Per gli Stati Uniti una rassegna dei risultati empirici è in Corcoran et al. '90. Un'applicazione che considera le differenze di genere è Neumark '88. Recenti applicazioni al caso italiano sono Checchi-Ichino-Rustichini '94 e '96.

le variabili relative alla famiglia d'origine non siano correlate direttamente con la variabile dipendente e risulta quindi illegittimo usarle come strumenti per l'istruzione.

Proprio la ricerca di variabili che possano essere legittimi strumenti per l'istruzione è l'obiettivo dell'approccio degli *esperimenti naturali* applicato allo studio di funzioni del salario. L'idea-base di questa metodologia è approssimare con dati reali esperimenti "ideali" eseguibili in un ipotetico laboratorio in cui è possibile controllare tutte le condizioni del modello e ripetere l'esperimento a parità di condizioni. L'approssimazione è necessaria perché l'economia rimane una scienza sociale che studia soggetti storici e la simulazione di laboratorio è applicabile solo a un numero limitato di casi.

Per quanto riguarda la stima di funzioni del salario un'approssimazione sufficiente può essere la seguente: considerare una popolazione di individui che sia un campione rappresentativo della popolazione reale e di cui conosciamo una serie di caratteristiche di controllo che approssimino la "parità di condizioni". Cercare poi un evento esogeno che induca un aumento nella scolarità per un sottogruppo selezionato in modo aleatorio (*gruppo di trattamento*) senza modificare quella del resto della popolazione (*gruppo di controllo*). Infine valutare la variazione indotta sui salari del gruppo di trattamento come effetto causale dell'istruzione sul salario.

L'esempio che più si avvicina a questa costruzione ideale è contenuto in Angrist e Kruger '91b: in questo caso la realizzazione aleatoria che "assegna" anni d'istruzione aggiuntivi agli individui selezionati è rappresentata dall'ordine di partenza per la guerra del Viet Nam.

Difficile è comunque il ripetersi di eventi simili. In assenza di assegnazioni "puramente" aleatorie "*one needs to identify a causal determinant of schooling that can be legitimately excluded from earnings equation*" (Card '93). La metodologia degli esperimenti naturali nella stima delle funzioni del salario si riduce quindi ad una corretta applicazione del metodo delle variabili strumentali: cerchiamo eventi su cui costruire indicatori che siano correlati con l'istruzione (*determinant of schooling*) ma non col salario (*excluded from earnings equation*).

Esempi di indicatori di questo tipo applicati agli Stati Uniti sono: il trimestre di nascita degli individui (Angrist e Kruger '91a), la composizione per sesso delle famiglie (Butcher e Case '93), la presenza o meno di un college nelle vicinanze della residenza del potenziale studente (Kane e Rouse '93 e Card '93).

In base a questa metodologia sembra quindi possibile risolvere i problemi di distorsione delle stime standard, rimuovendo i problemi di endogeneità. Una sua applicazione al caso italiano può quindi costituire la soluzione per verificare la gerarchia dei rendimenti dell'istruzione tra uomini e donne.

4 - Una Soluzione: Stime con Esperimenti Naturali.

Strumenti utilizzati.

I due strumenti utilizzati sono: l'indicatore *province*, che replica ed adatta lo strumento presentato in Card '93, e l'indicatore *riforme*, ispirato in parte ai lavori di Angrist-Krueger '91a e Checchi-Ichino-Rustichini '94.

L'intuizione di Card è la seguente: persone che vivono sufficientemente lontano da un'università affrontano un costo più elevato nel conseguire una laurea perché non possono risiedere con la propria famiglia d'origine. Il maggiore costo implicato dall'assenza di un'università nelle vicinanze può quindi essere l'evento esogeno che influisce sull'istruzione conseguita. Nell'applicare questa idea all'Italia, l'estensione territoriale considerata è la provincia. In questo modo l'area ha un'estensione territoriale tale da implicare un forte aumento dei costi, inoltre ricondurre l'analisi alla provincia permette di combinare in modo più agevole le informazioni individuali con le informazioni relative alla dislocazione territoriale dei centri universitari¹⁶.

Per verificare in modo preliminare questa prima intuizione, seguendo Card '93, è stato considerato il sottogruppo composto da individui cresciuti in province prive di università quando la loro età era pari a 19 anni. Per questo gruppo sono stati fatti regredire gli anni d'istruzione conseguiti su alcune variabili di controllo individuali¹⁷. In base a questa stima sono stati predetti gli anni d'istruzione conseguiti per chi ha vissuto in province con e senza università. Entrambi i sottocampioni sono stati poi divisi in quartili secondo gli anni di studio predetti in modo da calcolare per ogni quartile la media degli anni d'istruzione effettivamente conseguiti. I risultati ottenuti sono riportati nelle figure 4.1 e 4.2.

¹⁶ Questa necessità ha però obbligato a una riduzione del campione: la procedura preliminare presentata nel seguito è applicata a quello che sarà definito campione 2, ovvero individui nati tra il 1936 e il 1962 (estremi inclusi) e che risiedono nella provincia in cui sono nati. Spiegherò la limitazione per età parlando del prossimo strumento. La seconda limitazione è invece pertinente proprio con lo strumento *province* ed è implicata da scarsità d'informazione. L'informazione necessaria sarebbe la provincia di residenza nel momento in cui l'individuo sceglie se frequentare l'università o meno, ovvero intorno ai 18 anni. Tuttavia i dati Bankitalia riportano solo la provincia di nascita e quella di residenza: si è quindi ipotizzato che individui residenti alla data dell'intervista nella medesima provincia di nascita vi abbiano abitato anche nell'intorno dei loro 18 anni. Su questo sottocampione è stato possibile costruire lo strumento. Questa limitazione implica una riduzione del 30% sul campione femminile e del 35% su quello maschile. Le limitazioni introdotte non implicano distorsioni sulla distribuzione delle caratteristiche individuali utilizzate, anche se possono causare problemi di selezione.

¹⁷ Queste comprendono: età, dummies riguardo la posizione nell'attuale famiglia di residenza (capofamiglia, coniuge, figlio), stato civile, dummies per la macroregione di residenza, dummies per la dimensione del comune di residenza.

Le figure mostrano che, sia per le donne che per gli uomini, la vicinanza di un'università implica un numero maggiore o uguale di anni di scolarità. In particolare per entrambi i campioni questa differenza è molto ampia nel primo quartile ed è pari a circa due anni d'istruzione. Si tratta di un risultato ragionevole. Il primo quartile, infatti, comprende individui con una inferiore propensione a continuare la propria formazione scolastica. Saranno soprattutto questi, quindi, gli individui che potranno reputare decisivo il maggiore costo implicato dalla lontananza dell'istituzione universitaria e sarà soprattutto per queste persone che la differenza di anni di studio conseguiti risulterà sensibile alla provincia di residenza.

In conclusione l'indicatore-strumento *province* è una variabile binaria che assume valore 1 quando la provincia in cui l'individuo viveva a 19 anni era sede di almeno un'università ed assume valore 0 in caso contrario.

L'indicatore *riforme* è costruito invece considerando come eventi esogeni le riforme del sistema scolastico italiano avvenute negli anni '60. La prima, legge n. 1859 del 31 dicembre 1962, prescrive la fusione in una scuola media unica della precedente scuola media e della scuola di avviamento professionale. La seconda, legge n. 910 dell'11 dicembre 1969, consente ai diplomati di tutte le scuole medie superiori di iscriversi a qualsiasi facoltà universitaria.

Si tratta in entrambi i casi di riforme che, contrariamente al sistema precedente, permettono sempre il passaggio a qualsiasi grado di istruzione immediatamente superiore, sono quindi eventi esogeni sufficienti a condizionare l'aumento di scolarità conseguita dalla popolazione. Lo strumento che ne deriva è una variabile binaria 0-1 che assume valore 1 se l'individuo ha avuto la possibilità potenziale di frequentare la scuola corrispondente dopo la relativa riforma.

Questo strumento ha però il difetto di utilizzare la coorte di appartenenza degli individui¹⁸ ed implica perciò il seguente problema: l'età d'uscita dalla vita lavorativa, che determina l'inclusione o meno nel campione, non è indipendente dal livello d'istruzione raggiunto. In generale i più istruiti lavorano fino a livelli d'età superiori, determinando una sovrarappresentazione di individui con elevata istruzione nel sottocampione dei lavoratori più anziani. Applicando lo strumento direttamente a questi dati non si spiegherebbe la variabilità negli anni d'istruzione di questo sottogruppo. E' stato quindi scelto di ridurre il campione ad individui con un'età compresa nell'intorno della coorte su cui hanno effetto le riforme¹⁹.

¹⁸ Avere la "possibilità potenziale di frequentare la scuola corrispondente dopo la relativa riforma" significa infatti in questo caso "persone nate dopo o nel 1951".

¹⁹ Il campione 1, a cui si riferiscono le stime probit presentate nella tab 4.1 è quello limitato all'intorno più stretto, ovvero ad individui nati nell'intervallo 1946-1962 (estremi inclusi). La limitazione superiore è introdotta per

L'analisi preliminare attuata per verificare questo indicatore è stata una serie di semplici stime probit dell'influenza delle riforme sulla probabilità di raggiungere un elevato livello d'istruzione. La tabella 4.1 mostra, come atteso, un'influenza significativa e positiva.

Risultati

I rendimenti dell'istruzione stimati sono presentati nella tabella 4.2. Gli strumenti utilizzati sono indicati nella prima colonna e corrispondono alle definizioni precedentemente date. Tutte le stime sono derivate da funzioni del salario con la specificazione presentata nella tabella 2.1. Nella tabella 4.2 sono riportati solo i rendimenti dell'istruzione ed alcune statistiche. I risultati completi sono presentati nell'appendice 2.

Le usuali statistiche implicano stime con variabili strumentali più imprecise di quelle con minimi quadrati, ma comunque a livelli di significatività accettabili e con variabilità paragonabili ai migliori studi riportati nella tabella 2.2.

Il test di Wu-Hausman rifiuta l'esogeneità della variabile istruzione sul campione maschile ma non su quello femminile. La supposta esogeneità dell'istruzione nella funzione del salario femminile però sembra, come verrà discusso nei paragrafi seguenti, più il risultato di distorsioni con effetto opposto e di un diverso effetto degli strumenti utilizzati che il segnale di una effettiva assenza di endogeneità. Risultati simili si hanno applicando il test di Sargan al campione 2, mentre per il campione 1 il test non rifiuta la validità degli strumenti al 99%.

Due aspetti meritano di essere enfatizzati riguardo i risultati trovati:

- il sistematico aumento del coefficiente stimato con variabili strumentali;
- il passaggio da un maggiore rendimento per le donne nel caso OLS ad un maggior rendimento per gli uomini nel caso con variabili strumentali.

Soprattutto il secondo risultato è cruciale perché contraddice i risultati empirici fino ad ora trovati nel mercato del lavoro italiano.

Interpretazioni dei risultati

permettere all'individuo di finire la sua carriera scolastica, la limitazione inferiore per eliminare i problemi appena discussi.

L'aumento del coefficiente stimato con variabili strumentali è un risultato in linea con i precedenti lavori che si ispirano alla metodologia degli esperimenti naturali²⁰. Questo risultato è robusto al variare dei campioni, degli strumenti, del modo di rappresentare la scolarità ed attuando le regressioni su campioni separati donna-uomo e su campioni aggregati²¹.

Riprendendo i motivi di distorsione prima presentati, questa sottostima potrebbe essere causata da errori di misurazione che coinvolgono le variabili relative all'istruzione. Analogo effetto si avrebbe con l'omissione di variabili rilevanti identificabili con l'idea di *abilità* individuale, sotto l'ipotesi che individui più *abili* terminino prima i propri studi.

Un'ultima interpretazione è possibile se ipotizziamo che i rendimenti dell'istruzione varino all'interno della popolazione e che prevalga distorsione da tassi, ovvero che, coerentemente col modello Becker-Card presentato precedentemente, individui con scolarità inferiore abbiano rendimenti marginali superiori. Gli strumenti utilizzati sono basati su indicatori discreti di interventi esogeni che agiscono su un sottocampione di individui (*gruppo di trattamento*) senza influenzare il resto del campione (*gruppo di controllo*). Se il gruppo di trattamento è selezionato in modo aleatorio, il rendimento stimato è rappresentativo della media del campione. Tuttavia se il gruppo di trattamento subisce distorsione da selezione, il coefficiente stimato è rappresentativo del rendimento marginale di questo particolare sottogruppo. In altre parole, se l'esperimento naturale induce maggiore scolarità solo su individui con particolari caratteristiche, allora la stima γ_{IV} sarà uguale al rendimento marginale dell'istruzione per il gruppo di persone con queste particolari caratteristiche (se si ipotizza non variabilità dei rendimenti all'interno di questo gruppo). Perciò se il sottocampione su cui agiscono gli strumenti ha rendimenti marginali significativamente superiori alla media, allora la stima IV che si ottiene può eccedere la semplice stima con minimi quadrati ordinari²².

Nelle stime attuate c'è la possibilità che il gruppo di controllo abbia rendimenti superiori alla media perché gli strumenti agiscono soprattutto su persone con un livello d'istruzione potenzialmente inferiore ovvero proprio quelle persone che nel modello Becker-Card hanno rendimenti più elevati.

Il primo strumento agisce soprattutto su questi individui perché corrisponde ad un evento esogeno che abbassa i costi dell'istruzione. Come è già stato messo in evidenza commentando le figura 4.1 e

²⁰ Ad esempio quelli riportati da Card '94, stimano con IV rendimenti del 10-100% più elevati di quelli conseguiti con OLS.

²¹ Le stime sui campioni aggregati e basate su altre rappresentazioni dell'istruzione, disponibili su richiesta, non sono riportate, ma confermano i risultati delle stime presentate nella tabella 4.2.

²² Una formalizzazione è presentata in Card '94, un'applicazione in Card '93. Un'esaustiva trattazione econometrica è contenuta in Heckman '97.

4.2 questo abbassamento dei costi è decisivo per individui appartenenti al quartile con istruzione potenzialmente più bassa, ovvero individui che probabilmente affrontano tassi più elevati perché provenienti da famiglie meno abbienti. Il secondo strumento agisce facilitando la mobilità tra gradi d'istruzione. Anche in questo caso è probabile che scuole che non permettevano il conseguimento di elevata scolarità fossero scelte da individui provenienti da famiglie più povere e quindi con un grado d'istruzione potenzialmente inferiore²³.

Ma forse più interessante dell'interpretazione dell'aumento dei coefficienti è quella del perché questo aumento avvenga in modo asimmetrico, tanto da implicare un'inversione nella gerarchia dei rendimenti tra uomini e donne. Se infatti quasi tutti i lavori empirici sul mercato del lavoro italiano stimano rendimenti dell'istruzione femminili superiori a quelli maschili, le stime qui ottenute con variabili strumentali suggeriscono esattamente l'opposto.

La spiegazione più semplice potrebbe essere la diversa entità delle distorsioni a seconda del genere del sottocampione. Il risultato è ad esempio compatibile con una maggiore varianza negli errori di misura del sottocampione maschile. Se infatti il vero valore del coefficiente è maggiore per gli uomini, le stime OLS possono affermare il contrario se la distorsione verso il basso che coinvolge il sottocampione maschile è sufficientemente superiore a quella femminile. Analogo ragionamento si può applicare alla distorsione da abilità se individui più abili terminano prima i propri studi. Nell'ipotesi opposta otteniamo il medesimo risultato se la distorsione da abilità è combinata con quella da variabili legate al background familiare e da errori di misura. Ad esempio, se ipotizziamo analoga influenza degli errori di misura e distorsione da abilità verso l'alto maggiore per le donne, l'effetto netto è una distorsione verso il basso per entrambi i sottocampioni dove però l'abbassamento è inferiore per le donne vista la loro superiore distorsione da omissione di variabili.

Alcune argomentazioni possono essere portate a sostegno di quest'ultima ipotesi interpretativa. Dal punto di vista empirico è interessante notare come l' R^2 delle funzioni del salario femminili sia in genere inferiore a quello delle funzioni del salario maschili²⁴. Dal punto di vista teorico si può ricordare l'importanza che il fattore produttivo *intermittenza* riveste nel determinare la retribuzione globale femminile²⁵, questa importanza è invece quasi del tutto assente nel caso maschile perché la

²³ Entrambe le affermazioni sono coerenti con la letteratura sui background familiari e sociali tra cui, ad esempio, Corcoran-Gordon-Laren-Solon '90.

²⁴ Questo risultato è confermato da altri lavori empirici, ad esempio: per l'Italia, Erickson-Ichino '92; per la Gran Bretagna, Wright-Ermisch '90; per gli Stati Uniti, Blau-Kahn '92.

²⁵ Questo aspetto è enfatizzato dai modelli che si richiamano alla teoria del capitale umano (ad es. Blau e Ferber '91), ma riveste una certa importanza anche nei modelli della segmentazione ed in quelli che assumono salari d'efficienza (ad es. Bettio '90). Per altre impostazioni, invece, l'idea di intermittenza come fattore produttivo è fuorviante, ad

vita lavorativa degli uomini è generalmente più continua. Nelle funzioni del salario stimate non è stato possibile inserire variabili che tenessero conto di questo fattore: l'effetto può essere quindi una maggiore distorsione da variabili omesse per le funzioni del salario in cui questo fattore è importante, ovvero quelle femminili²⁶. Un'altra omissione può influenzare in modo asimmetrico le stime: quella relativa a variabili che considerino le determinanti dell'offerta di lavoro femminile. In genere la maggior parte della popolazione attiva maschile offre lavoro, mentre solo una porzione di quella femminile si comporta allo stesso modo. Questo implica stime delle funzioni del salario su un campione "troncato" di donne. Una possibile soluzione a questo problema consiste nell'attuare la correzione di Heckman, che ho applicato nella sezione *stime standard*, ovvero inserire tra i regressori una trasformazione non lineare (la *Mill's ratio*) delle variabili che influenzano l'offerta di lavoro femminile. L'omissione di variabili nelle funzioni del salario può quindi essere più grave per le donne perché non è inserita questa variabile che rappresenta le determinanti dell'offerta, variabile che invece ha inferiore influenza nel caso maschile.

Tutte le interpretazioni presentate finora considerano i rendimenti stimati con variabili strumentali rappresentativi della media del campione. Tuttavia se ipotizziamo variabilità dei rendimenti all'interno della popolazione e applichiamo il modello Becker-Card, l'inversione della gerarchia dei rendimenti non coinvolge l'intero campione, ma solo il sottogruppo su cui ha effetto l'evento esogeno relativo all'esperimento naturale (*gruppo di trattamento*). Gli strumenti utilizzati potrebbero in effetti agire su gruppi di trattamento con rendimenti superiori alla media perché coinvolgono soprattutto individui provenienti da famiglie meno abbienti o comunque che affrontano tassi più elevati. Ma perché il gruppo di trattamento maschile dovrebbe avere rendimenti superiori a quello femminile?

Una spiegazione può essere data ipotizzando quella che nella letteratura riguardo la discriminazione di genere viene chiamata discriminazione *pre-mercato* del lavoro (ad es. D'Amico '87 o Cain '86). Con questo concetto si fa riferimento al trattamento differenziale che coinvolge uomini e donne nell'acquisizione di caratteristiche produttive prima di entrare nel mondo del lavoro. L'esempio classico riguarda proprio l'acquisizione di scolarità.

Se vale questa ipotesi, le famiglie scelgono di dare priorità al conseguimento dell'istruzione dei figli piuttosto che a quello delle figlie. In questo caso il livello d'istruzione medio del gruppo di

esempio negli approcci marxisti ed in generale nelle impostazioni che rifiutano la divisione tra discriminazione pre-mercato e nel-mercato del lavoro (ad es. Bettio '88).

²⁶ Il lavoro di Wright e Ermisch '90 è un esempio in cui l'introduzione di variabili rappresentative dell'intermittenza porta un aumento nell' R^2 delle stime femminili, senza tuttavia colmare completamente la differenza rispetto a quelle maschili.

trattamento maschile è inferiore a quello del gruppo di trattamento femminile perché la riduzione dei costi indotta dagli strumenti è sufficiente a far proseguire gli studi del figlio maschio ma non quelli della figlia femmina. Affinché l'evento esogeno permetta anche alla o alle figlie di conseguire un maggiore livello d'istruzione, la famiglia d'origine dovrà avere un livello di reddito superiore, ovvero affrontare un tasso di sconto inferiore che sarà associato a minori rendimenti marginali.

In conclusione, secondo questa interpretazione, il maggior valore del coefficiente maschile stimato dalla procedura con esperimenti naturali non significa che il rendimento dell'istruzione sia maggiore per gli uomini, ma significa che il gruppo di trattamento maschile su cui agiscono gli strumenti ha un rendimento superiore a quello del gruppo di trattamento femminile. È da notare che secondo Heckman '97, e in parte anche Card '93, è proprio il rendimento del gruppo di trattamento il risultato più interessante al fine della valutazione di interventi di policy, come ad esempio le riforme dell'istruzione utilizzate per costruire lo strumento.

5 - Conclusioni.

L'usuale gerarchia, ovvero rendimenti dell'istruzione femminili superiori a quelli maschili, è confermata solo in parte. Come contenuto in lavori precedenti (ad es. Lucifora-Sestito '93 e Blau-Kahn '92) è confermata, anche se in modo non indipendente dalla specificazione, quando si applica la procedura OLS. E' negata invece da stime IV ispirate alla metodologia degli esperimenti naturali. In questo caso, utilizzando indicatori inediti per strumentare la variabile istruzione, i rendimenti sono mediamente più elevati e risultano superiori per gli uomini che per le donne. La distorsione delle stime con minimi quadrati ordinari sembra quindi tale da nascondere il vero rapporto relativo dei rendimenti in base al genere.

E' possibile tuttavia un'interpretazione alternativa (Card '93 e '94; Heckman '97): l'inversione non varrebbe in realtà per tutto il campione, ma solo per il sottogruppo dei meno abbienti. Il significato economico di questo risultato porterebbe però ad ipotizzare variabilità dei rendimenti all'interno della popolazione e discriminazione di genere pre-mercato del lavoro a svantaggio delle donne.

In conclusione le stime dei rendimenti dell'istruzione per uomini e donne non sembrano stabili nel loro rapporto relativo. A seconda delle specificazioni e delle tecniche di stima utilizzate, i risultati possono cambiare in modo significativo. Non sembra quindi, come i precedenti lavori facevano

supporre, che i superiori rendimenti femminili rispetto a quelli maschili siano un fatto empirico acquisito per il mercato del lavoro italiano.

APPENDICI

APPENDICE 1: L'universo di stima.

Le stime sono attuate su un sottocampione di 5.734 individui ottenuto dal campione totale dell'*Indagine Bdl*. I passaggi fatti per ottenerlo sono schematizzati nella tabella A.1.1 indicando il costo, in termini di osservazioni eliminate, implicato dalle varie limitazioni.

La prima limitazione è ovvia mentre la seconda (solo lavoratori dipendenti) è determinata dalla scarsa attendibilità dei redditi dichiarati dai lavoratori autonomi. Come riportato in Brandolini e Cannari '94, l'*Indagine Banca d'Italia* sembra sottostimare di circa il 50% il reddito dei lavoratori autonomi.

La terza limitazione, cioè considerare solo gli individui classificati come capofamiglia, partner o figli, è necessaria perché il modo in cui sono attuate le interviste non permette di risalire a variabili importanti nelle mie funzioni del salario per persone classificate in altro modo.

Il fatto che i salari siano riportati su scala annuale, e la mancanza delle informazioni necessarie per risalire da questi ai salari orari, implica la quarta e quinta limitazione (solo individui che lavorano tutto l'anno a tempo pieno). Infine alcune "pulizie" finali hanno portato all'eliminazione dello 0.6% del campione.

Il sottocampione finale ottenuto è costituito da 5734 individui di cui 1933 donne e 3801 uomini.

APPENDICE 2: Risultati completi delle stime.

I risultati completi delle stime con procedure standard sono presentati nella tabella A.2.1. Tutte le stime sono state attuate col software STATA®.

Per quanto riguarda le stime OLS, oltre che sulla variabile istruzione, si nota un maggior rendimento femminile anche sulle altre classiche variabili di capitale umano, come l'esperienza e la tenure, mentre una forte differenza a svantaggio delle donne si nota nei rendimenti associati alle qualifiche.

Le proxy per il lavoro di cura mostrano risultati interessanti. Per le donne la presenza di figli non ha un effetto significativo sul salario, ma solo sull'offerta di lavoro²⁷. Per gli uomini, invece, presenza e numero di figli hanno un effetto significativo e positivo sul salario²⁸.

Infine vale la pena sottolineare la significatività del reddito del partner, significatività che si estende anche alla sola presenza nel caso del sottocampione maschile. Questa influenza è di pari entità in entrambe le funzioni ma ha effetto opposto: nella regressione femminile ha segno positivo, mentre in quella maschile ha segno negativo.

Per quanto riguarda le stime con correzione di Heckman, il primo risultato interessante è la conferma che il campione di donne con reddito da lavoro non è selezionato in modo casuale: infatti quando λ è inserita tra i regressori si dimostra significativa. Il segno negativo del coefficiente indica che le donne che si offrono sul mercato del lavoro

²⁷ La stima probit dell'offerta di lavoro, non riportata ma disponibile su richiesta, stima infatti una significatività del 95% per tutte le dummy relative ad età e presenza figli.

²⁸ La significatività congiunta delle variabili relative ai figli è del 99%.

sono quelle che comandano salari inferiori²⁹. Il secondo risultato interessante è che ignorare la distorsione da autoselezione porta a sovrastimare il rendimento delle variabili di capitale umano, istruzione ed esperienza.

Le regressioni della tavola A.2.2 confrontano stime con minimi quadrati ordinari e con variabili strumentali. Nel campione 1 è stato utilizzato solo lo strumento *riforme*, mentre nel campione 2 anche lo strumento *province*.

²⁹ Questo accade perché λ ha medesimo segno di ρ , ovvero la correlazione tra i residui della funzione del salario e dell'equazione di partecipazione alla forza lavoro, ovvero la correlazione tra probabilità di partecipare e salario.

RIFERIMENTI BIBLIOGRAFICI

- Angrist, J. e Newey, W.** (1991), “Over-Identification Test in Earnings Function with Fixed Effects”, *Journal of Business and Economic Statistics*, Luglio, Pp. 317-323.
- Angrist, J. e Krueger, A.** (1991a) “Does Compulsory Schooling Affect Schooling And Earnings?”, *Quarterly Journal of Economics*, Novembre, Pp. 979-1014.
- Angrist, J. e Krueger, A.** (1991b), “Estimating the Payoff to Schooling Using the Vietnam Era Draft Lottery”, Princeton University, Industrial Relations Section, *Working Paper* No. 290.
- Appleton, S.J.; J. Hoddinott; P. Krishnan; K. Max** (1995), “Gender Differences in the Returns to Schooling in Three African Countries”, *FEEM Working Paper* No. 24.
- Becker, G.** (1967), *Human Capital and the Personal Distribution of Income*, Ann Arbor: University Of Michigan Press.
- Becker, G.** (1991), *Treatise on the Family*, Cambridge: Harvard University Press. (prima ed. 1981).
- Becker, G.** (1993), *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*, Chicago: University of Chicago Press (prima ed. 1975)
- Becker G. e Tomes, N.** (1986), “Human Capital and the Rise and Fall of Families”, *Journal Of Labor Economics*, V. 4, N. 3, Parte 2, Pp. S1-S39.
- Bettio, F.** (1988), *The Sexual Division of Labour: the Italian Case*, Oxford: Clarendon Press.
- Bettio, F.** (1990), “Segregazione e discriminazione nel mercato del lavoro: parte I (letteratura straniera)”, *Economia & Lavoro*, n.4, pp. 27-47
- Bettio, F.** (1991), “Segregazione e discriminazione nel mercato del lavoro: parte II (letteratura italiana)”, *Economia & Lavoro*, n.1, pp. 49-66.
- Blackburn, M. e Neumark, D.** (1993), “Are OLS Estimates of the Return to Schooling Biased Downward? Another Look”, *NBER Working Paper* No 4259.
- Blau, F.D. e Ferber, M.E.** (1991) “Career Plans and Expectations of Young Men and Women”, *The Journal of Human Resources*, V. 26, N. 4, Pp. 581-607.
- Blau, F.D. e Kahn, L.M.** (1992), “The Gender Earnings Gap: Some International Evidence”, *NBER Working Paper* No. 4224.
- Brandolini, A. e Cannari, L.** (1994), “The Bank of Italy’s survey of Household Income and Wealth”, in: Ando, A., Guiso, L., Visco, I. (ed.) *Saving and the Accumulation of Wealth*, Cambridge: Cambridge University Press.
- Cain, G.C.** (1986), “The Economic Analysis of Labor Market Discrimination”, in: Ashenfelter O. e Layard R. *Handbook of Labor Economics*, Amsterdam: North Holland.

- Callan, T.** (1991), “Male-female Wage Differentials in Ireland”, mimeo ESRI.
- Cannari, L. e D’Alessio, G.** (1995), “Il Rendimento dell’istruzione. Alcuni problemi di stima”, Banca d’Italia, *Tem di Discussione* N. 253.
- Card, D.** (1993), “Using Geographic Variation in College Proximity to Estimate the Return to Schooling”, NBER *Working Paper* No 4483.
- Card, D.** (1994), “Earnings, Schooling, And Ability Revisited”, Princeton University, Industrial Relations Section *Working Paper* No 331.
- Cecchi, D., Ichino, A. e Rustichini, A.** (1994), “More Equal but Less Mobile? Education, Financing, and Intergenerational Mobility in Italy and in the United States”, CEPR *Working Paper* No 1496.
- Cecchi, D., Ichino, A. e Rustichini, A.** (1996), “Immobili perchè eguali? Mobilità occupazionale e scolastica tra generazioni in Italia e Stati Uniti”, in: G. Galli: *La mobilità della società italiana*, Roma: SIPI.
- Cobalti, A. e Schizzerotto, A.** (1995), “Occupational Returns to Education in Contemporary Italy”, mimeo.
- Corcoran, M., Gordon, R., Laren, D. e Solon, G.** (1990), “Effects of Family and Community Background on Economic Status”, *American Economic Review - Papers and Proceedings*, V. 80, N. 2, Pp. 362-366.
- D’Amico, T.F.** (1987), “The Conceit of Labor Market Discrimination”, *American Economic Review - Papers and Proceedings*, V. 77, No. 2, Pp. 311-315.
- Ehrenberg, R. e Smith, R.** (1991), *Modern Labor Economics* (IV ed.), New York: Harper Collins.
- Erickson, C.L. e Ichino, A.** (1992), “Wage Differential in Italy: Market Forces, Institutions, and Inflation”, NBER *Working Paper* No. 4922.
- Griliches, Z.** (1977), “Estimating the Returns to Schooling: Some Econometric Problems”, *Econometrica*, V.45 N.1, Pp. 1-22.
- Griliches, Z., Hall, B. e Hausman, J.A.** (1978), “Missing Data and Self-Selection in Large Panels”, *Annales De L’INSEE*, V. 30, Pp. 137-176.
- Heckman, J. J.** (1997), “Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations”, *forthcoming*.
- Heckman, J.J. e Killingsworth, M.R.** (1986), “Female Labor Supply: a Survey”, in: Ashenfelter O. e Layard R. *Handbook of Labor Economics*, Amsterdam: North Holland.
- Lam, D. e Schoeni, R. F.** (1993), “Effects of Family Background on Earnings and Returns to Schooling: Evidence from Brasil”, *Journal Of Political Economy*, V. 101, N. 4, Pp 710-740.

- Lucifora C. e Reilly B.** (1990), “Wage Discrimination and Female Occupational Intensity”, *Labour*, n.4, pp. 147-168.
- Lucifora, C. e Sestito, P.** (1993), “Determinazione del salario in Italia: una rassegna della letteratura empirica”, Università Cattolica del Sacro Cuore, *Working Paper* No. 9305.
- Mincer, J.** (1974), *Schooling, Experience, and Earnings*, New York: NBER.
- Montgomery, J.** (1991), “Social Networks and Labor-Market Outcomes: Toward an Economic Analysis”, *American Economic Review*, V. 81, N. 5, Pp. 1408-1428.
- Neumark, D.** (1988), “Gender Differences in Family Effects on Human Capital and Earnings: an Empirical Study of Siblings” in: Maital, S. (ed.) *Applied Behavioural Economics* V. 1, New York: New York University Press.
- Psacharopoulos, G.** (1994), “Returns to Investment in Education: A Global Update”, *World Development*, V. 22, N. 9, Pp. 1325-1343.
- Willis, R.J.** (1986), “Wage Determinants: A Survey And Reinterpretation Of Human Capital Earnings Function”, in: Ashenfelter O. e Layard R. *Handbook of Labor Economics*, Amsterdam: North Holland.
- Wright, R.E. e Ermish, J.F.** (1990), “Male-female Wage Differentials in Great Britain”, Birkbeck College, *Working Paper* No. 9010.

TAB. 2.1: Specificazione delle funzioni del salario: motivazioni teoriche e variabili inserite.

Concetti teorici:	Variabili inserite:
Capitale umano generale.	<ul style="list-style-type: none"> • Scolarità: anni d'istruzione completati. • Esperienza lavorativa: anni già trascorsi nel mercato del lavoro (lineari e al quadrato), dummy=1 se precedente esperienza lavorativa.
Capitale umano specifico.	<ul style="list-style-type: none"> • Anni con l'attuale datore di lavoro (tenure).
Caratteristiche dell'occupazione.	<ul style="list-style-type: none"> • Qualifica: dummies (4). • Settori: dummies (5). • Ore totali lavorate alla settimana.
Caratteristiche del mercato del lavoro.	<ul style="list-style-type: none"> • Macroregione di residenza: dummies (3). • Dummy per residenza in una grande città.
Caratteristiche della famiglia di appartenenza.	<ul style="list-style-type: none"> • Figli: dummies indicanti età e presenza figli (3), due variabili indicanti il numero di figli per fasce d'età. • Partner: dummy per presenza partner convivente in famiglia e sua intersezione col reddito da lavoro del partner. • Posizione nella famiglia: dummy=1 se capofamiglia.

Note: fra parentesi è indicato il numero di dummies per le rispettive variabili.

TAB. 2.2: Rendimenti dell'istruzione con metodo OLS e correzione di Heckman: risultati e confronto con altri lavori - Italia.

Metodo		donne	uomini
		coeff. (<i>err. st.</i>)	coeff. (<i>err. st.</i>)
Risultati:			
	anno		
OLS (ampio) ¹	1991	0,021 (0,002)	0,018 (0,001)
OLS (stretto) ¹	1991	0,031 (0,001)	0,039 (0,001)
Heckman ¹	1991	0,028 (0,002)	
Altri lavori²:			
autori	anno		
<i>Banca dati</i> ³			
Cobalti-Schizzerotto '95 ⁴ <i>Italian Social Mobility Survey</i>	1985	0,041	0,033
Erickson-Ichino '92 ⁴ <i>Indagine Bdl</i>	1987	0,028	0,017
Blau-Kahn '92 <i>Indagine Bdl</i>	1987	0,052 (0,003)	0,040 (0,002)
Lucifora-Reilly '90 <i>Indagine ENI-IRI</i>	1985	0,040 (0,001)	0,036 (0,001)

Note:

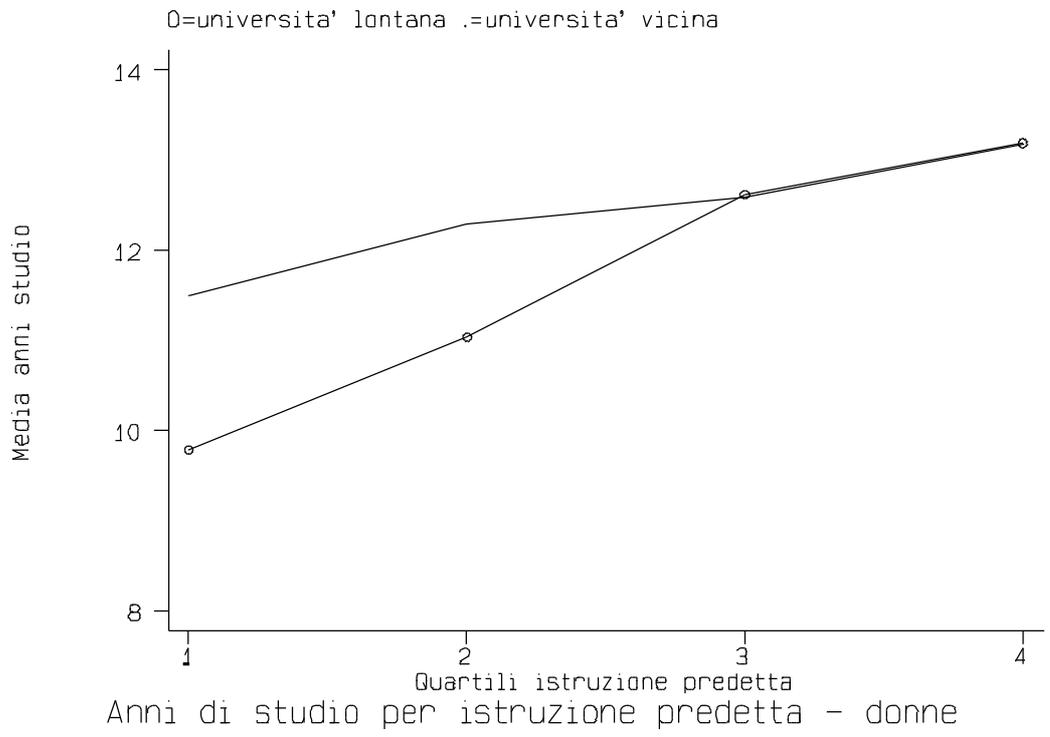
¹ I risultati completi delle stime sono riportati in appendice.

² Per attuare un confronto adeguato coi risultati trovati, sono stati selezionati lavori in cui: (i) il campione è rappresentativo dell'intera nazione (ii) la specificazione contiene almeno istruzione ed esperienza.

³ L'*Italian Social Mobility Survey* è un campione raccolto da un gruppo di sociologi (M. Barbagli, V. Capecchi, A. Cobalti, A. de Lillo, A. Schizzerotto), si differenzia fortemente dagli altri inseriti in tabella perché la variabile dipendente è un indice del prestigio sociale dell'occupazione e non una retribuzione. L'*Indagine Bdl* è quella presentata precedentemente e utilizzata nel presente lavoro: la variabile dipendente è il reddito da lavoro annuale al netto di tasse e contributi. L'*Indagine ENI-IRI* indica l'*Indagine sulle retribuzioni di fatto* attuata congiuntamente da ENI ed IRI comprendente dati di 100 imprese rappresentative dell'economia italiana: la variabile dipendente è il salario annuale lordo.

⁴ Valori calcolati come medie annue in base ai coefficienti delle dummies per titolo di studio.

FIG. 4.1



STATA™

FIG. 4.2

$$k_2^b = k_2$$

TAB. 4.1: Influenza dello strumento *riforme* sulla probabilità di raggiungere un elevato livello d'istruzione. Stime probit - Italia - 1991.

	<i>donne</i>			<i>uomini</i>		
Campione 1:	N = 1033			N = 1857		
variabile dipendente	coeff. (<i>err. st.</i>)	chi2	PseudoR ²	coeff. (<i>err. st.</i>)	chi2	PseudoR ²
Sopmed	0,403 (0,087)	21,3	0,017	0,208 (0,061)	11,4	0,004
Diploma	0,352 (0,087)	16,4	0,013	0,118 (0,061)	3,7	0,001

Note: la variabile dipendente *sopmed* = 1 quando l'individuo ha un'istruzione superiore alla media mobile calcolata su 3 coorti, la variabile dipendente *diploma* = 1 quando l'individuo ha un livello d'istruzione superiore o uguale al diploma. Campione 1 = individui nati nell'intervallo 1946-1962 (estremi inclusi).

TAB. 4.2: Rendimenti dell'istruzione per donne e uomini: stime OLS e con esperimenti naturali. - Italia - 1991.

<i>Campione e strumenti</i>		<i>donne</i>			<i>uomini</i>		
Campione 1:		N = 1033			N = 1857		
		coeff. (<i>err. st.</i>)	F	R ²	coeff. (<i>err. st.</i>)	F	R ²
	OLS	0,025 (0,002)	22,1	0,34	0,018 (0,002)	52,4	0,40
Test di Hausman		0,58			10,95		
Riforme	IV	0,036 (0,015)	17,2	0,32	0,052 (0,012)	42,1	0,29
Test di Sargan		0,00			0,00		
Campione 2:		N = 974			N = 2085		
		coeff. (<i>err. st.</i>)	F	R ²	coeff. (<i>err. st.</i>)	F	R ²
	OLS	0,022 (0,003)	20	0,33	0,017 (0,002)	64,4	0,42
Test di Hausman		1,89			17,82		
Riforme e province	GIVE	0,056 (0,027)	14,5	0,20	0,062 (0,013)	47	0,23
Test di Sargan		5,73			0,12		

Note : Variabile dipendente = logaritmo del salario annuale al netto di tasse e contributi, comprese integrazioni non monetarie. I regressori comprendono, oltre alla scolarità: esperienza, lineare e al quadrato, tenure, ore lavorate, indicatore di precedenti esperienze lavorative, 5 dummies settoriali, 4 dummies per qualifiche, una dummy per la dimensione della città di residenza, 3 dummies indicanti la macroregione di residenza, indicatore della presenza del partner in famiglia, la sua intersezione col reddito del partner, 3 dummies indicanti la presenza di figli per fasce d'età, il numero di figli per le rispettive fasce d'età, una dummy che indica se l'individuo è capofamiglia. Al campione 1 è applicato solo lo strumento *riforme*, mentre al campione 2 anche lo strumento *province*. Le regressioni complete sono presentate in appendice nella tavola A.2.2. Tutti i campioni comprendono solo i lavoratori dipendenti che abbiano lavorato a tempo pieno per tutto l'anno. I campioni si differenziano inoltre per le seguenti limitazioni: campione 1 = individui nati nell'intervallo 1946-1962 (estremi inclusi); campione 2 = individui nati nell'intervallo 1936-1962 (estremi inclusi) e che risiedono attualmente nel medesimo comune in cui sono nati.

TAB. A.1.1: Come si arriva all'universo di stima partendo dal campione iniziale dell'*Indagine sui bilanci delle famiglie italiane nell'anno 1991*.

Limitazioni	Numero individui
Campione iniziale	24.998
<ul style="list-style-type: none"> • lavoratori • dipendenti • capofamiglia, partner o figli 	<ul style="list-style-type: none"> - 16.456 - 2.100 - 127
totale parziale	6.315
che lavorano:	
<ul style="list-style-type: none"> • tutto l'anno • a tempo pieno 	<ul style="list-style-type: none"> - 353 - 192
totale parziale	5.770
elimino osservazioni che riportano:	
<ul style="list-style-type: none"> • più di un partner convivente • età non coerente con esperienza lavorativa • valori missing per le variabili inserite nella funzione del salario 	<ul style="list-style-type: none"> - 11 - 4 - 21
universo di stima	5.734
di cui:	
donne	1.933
uomini	3.801

Nota: si riporta la procedura utilizzata per ottenere il campione su cui sono state attuate le stime. Accanto alla causa della limitazione del campione è indicato, preceduto dal segno meno, il numero di osservazioni eliminate.

TAB. A.2.1: Stime delle funzioni del salario con OLS e correzione di Heckman. Risultati completi.

	Donne			Uomini						
	OLS ampio	OLS stretto	Heckman	OLS ampio	OLS stretto					
N	1933	1933	1933	3801	3801					
F	48,9	85,5	78,3	148,6	213,9					
R ²	0,371	0,308	0,310	0,475	0,361					
	Coeff.	Err. st.	Coeff.	Err. st.	Coeff.	Err. st.	Coeff.	Err. st.	Coeff.	Err. st.
annistud	0,021	0,002	0,031	0,001	0,028	0,002	0,018	0,001	0,039	0,001
esp. (anni/10)	0,125	0,022	0,185	0,019	0,165	0,021	0,105	0,016	0,182	0,016
esp. (anni2/1000)	-0,270	0,049	-0,312	0,049	-0,269	0,052	-0,209	0,032	-0,281	0,032
tenure	0,005	0,001			0,003	0,001			0,003	0,001
ore sett. lav.	0,002	0,001			0,004	0,001			0,004	0,001
già esp. lav.	0,005	0,013	-0,029	0,012	0,014	0,021	0,033	0,010	-0,002	0,010
industria	0,045	0,031			0,055	0,014			0,055	0,014
commercio	0,064	0,032			0,054	0,015			0,054	0,015
credito ass.	0,070	0,035			0,137	0,020			0,137	0,020
pub. amm.	0,085	0,030			0,046	0,014			0,046	0,014
impiegati	0,080	0,015			0,112	0,011			0,112	0,011
quadri	0,203	0,023			0,250	0,015			0,250	0,015
dirigenti	0,368	0,049			0,552	0,025			0,552	0,025
città>500mila ab.	0,011	0,011	0,014	0,011	0,012	0,011	0,016	0,008	0,005	0,008
nord	0,029	0,013	0,030	0,014	0,027	0,014	0,048	0,010	0,058	0,010
sud	-0,006	0,015	0,009	0,015	0,029	0,017	-0,040	0,011	-0,021	0,011
pres. partner	0,010	0,020	0,023	0,019	0,028	0,018	0,051	0,020	0,097	0,020
partn.redd*pres	0,005	0,002	0,004	0,002	0,003	0,002	-0,004	0,001	-0,006	0,001
fig. min. età<14	-0,010	0,018			0,019	0,015			0,019	0,015
fig. min. età>14	0,022	0,018			0,037	0,016			0,037	0,016
num. fig. età<14	0,023	0,015			0,022	0,009			0,022	0,009
num. fig. età>14	0,005	0,011			0,017	0,007			0,017	0,007
capofamiglia	0,095	0,018	0,112	0,018	0,076	0,023	0,103	0,019	0,107	0,019
λ					-0,054	0,023				
costante	9,092	0,054	9,204	0,025	9,269	0,037	9,171	0,040	9,202	0,040

Note: Variabile dipendente = logaritmo del reddito da lavoro annuale al netto di tasse e contributi, comprese integrazioni non m studio o licenza elementare, nessuna esperienza lavorativa, operai, agricoltura e costruzioni, centro, città =<500.000 ab., nessun

TAB. A.2.2: Stime delle funzioni del salario con OLS, IV e GIVE. Risultati completi (campione1).

Campione 1	Donne				Uomini			
	OLS		IV		OLS		IV	
N	1033		1033		1857		1857	
F	22,1		17,2		52,4		42,1	
R ²	0,34		0,32		0,40		0,29	
	Coefficienti	Err. st.	Coefficienti	Err. st.	Coefficienti	Err. st.	Coefficienti	
annistud	0,025	0,002	0,036	0,015	0,018	0,002	0,052	
esp. (anni/10)	0,094	0,045	0,090	0,045	0,041	0,037	0,103	
esp. (anni2/1000)	-0,225	0,130	-0,160	0,157	-0,030	0,099	-0,043	
tenure	0,007	0,002	0,007	0,002	0,004	0,001	0,004	
ore sett. lav.	0,004	0,001	0,006	0,002	0,004	0,001	0,006	
già esp. lav.	0,004	0,017	0,005	0,017	0,034	0,013	0,032	
industria	0,088	0,048	0,084	0,049	0,059	0,019	0,029	
commercio	0,098	0,050	0,095	0,051	0,067	0,021	0,046	
credito ass.	0,119	0,053	0,105	0,057	0,134	0,028	0,068	
pub. amm.	0,124	0,047	0,113	0,050	0,052	0,020	0,015	
impiegati	0,052	0,022	0,007	0,064	0,106	0,015	-0,007	
quadri	0,155	0,030	0,101	0,078	0,230	0,020	0,074	
dirigenti	0,357	0,058	0,275	0,123	0,582	0,039	0,314	
città>500mila ab.	0,009	0,014	0,008	0,015	0,021	0,011	0,026	
nord	0,053	0,017	0,051	0,018	0,058	0,014	0,061	
sud	0,020	0,020	0,021	0,020	-0,026	0,015	-0,012	
pres. partner	-0,014	0,028	-0,017	0,028	0,029	0,027	0,054	
partn:red*pres	0,006	0,002	0,006	0,002	-0,005	0,001	-0,008	
fig. min. età<14	-0,012	0,022	-0,006	0,024	0,014	0,019	0,019	
fig. min. età>14	0,024	0,026	0,028	0,027	0,044	0,027	0,051	
num. fig. età<14	0,023	0,015	0,024	0,015	0,021	0,010	0,029	
num. fig. età>14	0,009	0,015	0,011	0,015	0,002	0,011	0,002	
capofamiglia	0,108	0,024	0,108	0,024	0,098	0,025	0,080	
costante	8,974	0,082	8,821	0,219	9,220	0,060	8,763	

TAB. A.2.2: (continua, campione 2)

Campione 2	Donne				Uomini			
	OLS		IV		OLS		IV	
N	974		974		2085		2085	
F	20,0		14,5		64,4		46,9	
R ²	0,33		0,2		0,42		0,23	
	Coefficienti	Err. st.						
annistud	0,022	0,003	0,056	0,027	0,017	0,002	0,062	0,002
esp. (anni/10)	0,086	0,040	0,084	0,044	0,104	0,028	0,178	0,028
esp. (anni2/1000)	-0,160	0,095	-0,050	0,135	-0,208	0,058	-0,214	0,058
tenure	0,005	0,001	0,005	0,002	0,003	0,001	0,002	0,001
ore sett. lav.	0,002	0,001	0,007	0,003	0,004	0,001	0,006	0,001
già esp. lav.	-0,017	0,019	0,000	0,024	0,043	0,013	0,038	0,013
industria	0,121	0,048	0,088	0,058	0,064	0,017	0,030	0,017
commercio	0,150	0,050	0,114	0,062	0,061	0,019	0,053	0,019
credito ass.	0,165	0,055	0,098	0,079	0,136	0,026	0,051	0,026
pub. amm.	0,161	0,047	0,097	0,072	0,039	0,018	-0,008	0,018
impiegati	0,057	0,024	-0,083	0,113	0,118	0,014	-0,044	0,014
quadri	0,184	0,033	0,016	0,137	0,254	0,019	0,043	0,019
dirigenti	0,303	0,069	0,011	0,242	0,575	0,034	0,214	0,034
città>500mila ab.	0,008	0,015	0,010	0,017	0,017	0,010	0,017	0,010
nord	0,061	0,019	0,052	0,022	0,058	0,014	0,061	0,014
sud	0,023	0,020	0,031	0,022	-0,027	0,014	-0,004	0,014
pres. partner	0,006	0,030	0,006	0,033	0,046	0,026	0,058	0,026
partn:redd*pres	0,005	0,002	0,003	0,002	-0,005	0,001	-0,009	0,001
fig. min. età<14	-0,009	0,025	0,016	0,033	0,026	0,020	0,044	0,020
fig. min. età>14	0,000	0,025	0,024	0,033	0,045	0,021	0,050	0,021
num. fig. età<14	0,027	0,017	0,026	0,019	0,023	0,010	0,027	0,010
num. fig. età>14	0,012	0,013	0,011	0,014	0,009	0,008	0,005	0,008
capofamiglia	0,133	0,027	0,151	0,033	0,057	0,026	0,036	0,026
costante	9,037	0,084	8,603	0,355	9,220	0,056	8,642	0,056

Note: Variabile dipendente = logaritmo del reddito da lavoro annuale al netto di tasse e contributi, comprese integrazioni titolo di studio o licenza elementare, nessuna esperienza lavorativa, operai, agricoltura e costruzioni, centro, città =<500.000 ab