

**XVI CONVEGNO NAZIONALE DI ECONOMIA DEL LAVORO**

*Firenze, 4 - 5 ottobre 2001*

*Sessione Tematica*

*“Valutazione delle politiche del lavoro e strategia europea per l’occupazione”*

**La valutazione di efficacia dei corsi di formazione regionale: effetti di autoselezione e stima degli impatti occupazionali netti.**

**Cristina Berliri e Carmine Pappalardo\***

[c.berliri@isfol.it](mailto:c.berliri@isfol.it)   [c.pappalardo@isfol.it](mailto:c.pappalardo@isfol.it)

Keywords: formazione professionale, sample selection, impatti netti.

*Introduzione*

Gli interventi di formazione professionale realizzati dalle regioni - la formazione di primo e secondo livello rivolta ai giovani e gli interventi formativi indirizzati ai disoccupati e ad altri

soggetti a rischio di esclusione sociale - si pongono tra gli obiettivi prioritari la facilitazione dell'accesso al mercato del lavoro.

Lo scopo di questo lavoro è misurare l'impatto netto dei corsi di formazione professionale sugli esiti occupazionali, tenendo conto dei possibili effetti di autoselezione e di selezione del campione, connessi rispettivamente con la decisione individuale di seguire i corsi e con la selezione delle domande di partecipazione da parte degli enti attuatori. Il lavoro si prefigge, inoltre, di analizzare il ruolo dei corsi di formazione professionale rispetto alla riduzione del divario esistente tra uomini e donne nelle opportunità di inserimento lavorativo.

La stima dell'impatto netto della formazione professionale fornisce una prima misura dell'efficacia di questo strumento di politica attiva del mercato del lavoro. In particolare lo studio analizza gli esiti occupazionali dei destinatari di interventi di formazione professionale finanziati attraverso gli assi 1, 2 e 4 dell'ex obiettivo 3 del Fse<sup>1</sup> nell'annualità 1997 nelle Regioni Lombardia ed Emilia Romagna.

L'analisi si basa su un campione panel ricostruito dall'ISFOL a partire dalle indagini realizzate dalle due regioni del Centro Nord sugli esiti occupazionali dei corsi di formazione e dai microdati longitudinali dell'ISTAT sulle Forze di Lavoro. Il panel sulle Forze di Lavoro consente di identificare gli individui che hanno seguito un corso di formazione nel mese precedente la prima rilevazione e di analizzare per tutti gli individui in cerca di occupazione la condizione occupazionale ad un anno di distanza. Similmente, le indagini sugli esiti consentono di analizzare la situazione occupazionale ad un anno dalla conclusione dell'intervento.

Il campione ISTAT, depurato degli individui che avevano svolto attività formative nel mese precedente l'indagine, e le indagini regionali sono stati utilizzati per costruire un unico campione di indagine, stratificato in base alla scelta di partecipare o meno ad attività formative regionali (Manski-Lerman, 1977).

Utilizzando un modello di selezione stocastico con "switching" endogeno si perviene alla

---

\* Ricercatori ISFOL – Struttura nazionale di valutazione.

<sup>1</sup> L'Obiettivo 3 del Fondo Sociale Europeo coinvolge in Italia le regioni del Centro Nord e consiste nel "lottare contro la disoccupazione di lunga durata, facilitare l'inserimento dei giovani, l'integrazione delle persone minacciate di esclusione dal mercato del lavoro, promuovere le pari opportunità tra uomini e donne sul mercato del lavoro". Nell'ambito dell'Obiettivo 3 gli assi 1, 2 e 4 riguardano, rispettivamente, "l'inserimento dei disoccupati di lunga durata", "l'inserimento dei giovani" e "le pari opportunità sul mercato del lavoro".

stima della probabilità di trovare lavoro per tutti gli individui in cerca di occupazione, sia nel caso essi abbiano seguito un corso di formazione professionale che nel caso contrario. Il modello consiste di due equazioni, rispettivamente per la condizione occupazionale dei formati e dei non formati, e di una funzione di scelta che cattura il processo decisionale inerente alla partecipazione ai corsi di formazione. In tal modo si è in grado non solo di stimare l'impatto netto dei corsi di formazione sugli effettivi destinatari, ma anche di indicare l'impatto potenziale dei corsi sugli esiti occupazionali di quegli individui che non vi hanno effettivamente partecipato. In assenza di dati sperimentali in grado di fornire una situazione controfattuale non distorta da meccanismi di auto selezione, la stima della funzione di scelta consente di tenere conto di differenze inosservabili tra i destinatari effettivi degli interventi e i destinatari potenziali.

La probabilità di appartenere al gruppo dei formati viene stimata per ogni individuo appartenente al campione utilizzando variabili relative alle caratteristiche anagrafiche, al contesto territoriale e al contesto familiare. In particolare, la caratteristica dell'indagine sulle Forze di Lavoro di utilizzare la famiglia come unità di rilevazione ha consentito di ricostruire ed utilizzare variabili di background familiare presenti anche nelle indagini regionali e di particolare rilevanza nella stima della funzione di scelta individuale.

La struttura del lavoro è la seguente: nel paragrafo 1 si descrive l'approccio seguito per la costruzione della base di dati. Nel paragrafo 2 si accenna al problema di *selection bias* nelle stime su dati longitudinali. Il paragrafo 3 descrive la struttura analitica del modello sottoposto a verifica empirica. I principali risultati ottenuti sono commentati nel paragrafo 4.

### *1. Costruzione della base di dati*

Le fonti informative utilizzate sono rappresentate da un campione panel ricostruito a partire da due indagini ISTAT sulle Forze di Lavoro, condotte nel mese di aprile negli anni 1997 e 1998, e dalla banca dati *Placement* sui formati dei corsi professionali cofinanziati dal FSE nelle regioni Emilia Romagna e Lombardia<sup>2</sup>. Come noto, l'indagine sulle Forze di Lavoro relativa al

---

<sup>2</sup> Rilevazioni di *placement* sono state condotte da diverse regioni del Centro-Nord titolari di Programmi Operativi FSE. Sebbene le diverse amministrazioni abbiano operato in modo coordinato, la struttura dei questionari delle singole indagini regionali risulta in parte diversificata per rispondere alle specifiche

mese di aprile indaga l'aggregato degli individui "in cerca di lavoro" anche con riferimento alle azioni intraprese nell'ambito della formazione professionale; essa dunque contiene informazioni di particolare interesse ai fini della presente analisi.

I dati di *Placement* forniscono informazioni relative agli esiti occupazionali dei corsi di formazione professionale realizzati dalle regioni titolari di Programmi Operativi Obiettivo 3 nell'anno 1997. Tra il 1998 e il 2000, in queste regioni sono state effettuate rilevazioni su un campione composto di 6.742 individui che nell'anno 1997 avevano beneficiato di interventi di formazione professionale rivolti alle persone in cerca di prima o di nuova occupazione. Il tasso di copertura delle interviste complessive è risultato pari al 51%.

Per l'indagine condotta in Lombardia, l'universo di riferimento è stato costituito dall'insieme delle attività realizzate nell'anno 1997 a valere sugli assi 1 (disoccupati di lunga durata), 2 (giovani) e 4 (pari opportunità) del Fse; nel caso dell'Emilia Romagna, tale universo è stato ricostruito considerando i soli interventi a titolarità dell'Amministrazione regionale<sup>3</sup>. Dai tassi di copertura relativi ad ogni regione e al singolo asse d'intervento sono stati ottenuti gli indici di ponderazione utilizzati per riproporzionare le unità campionarie al totale dei beneficiari oggetto di analisi (pari a 13.266 unità)<sup>4</sup>.

Il campione utilizzato proviene dalla integrazione dei dati derivanti dalle indagini Placement e Forze di Lavoro. In entrambi i casi si sono considerati individui di età compresa tra i 14 e i 45 anni, residenti in Emilia Romagna e Lombardia<sup>5</sup>, in cerca di lavoro nell'anno 1997 e si è

---

esigenze conoscitive individuate da ciascuna Autorità di gestione. Tali difformità hanno determinato l'impossibilità di utilizzare il set completo delle rilevazioni regionali. In riferimento alle regioni Emilia Romagna e Lombardia è stato, tuttavia, possibile ricostruire un data-set omogeneo rispetto alla definizione e natura delle variabili rilevate.

<sup>3</sup> La mancata considerazione delle attività di competenza provinciale potrebbe aver comportato una sottostima della formazione di base. Gli interventi su base provinciale hanno rappresentato circa il 30% delle attività cofinanziate in Emilia Romagna nel 1997.

<sup>4</sup> Tale base di dati ha consentito di disporre di un notevole insieme di informazioni sulle caratteristiche socio-demografiche e sul contesto sociale e familiare delle persone beneficiarie degli interventi formativi, oltre che sulla condizione occupazionale degli stessi individui a un anno di distanza dal termine degli interventi. Quest'ultima è stata ricostruita incrociando le informazioni regionali di *placement* rispetto alla distanza intercorsa sia tra la chiusura del corso di formazione e l'intervista sia tra la chiusura del corso di formazione e la condizione occupazionale rilevata. I destinatari di interventi che al momento dell'iscrizione risultavano occupati sono stati esclusi dall'analisi.

<sup>5</sup> Per queste regioni risultano disponibili variabili relative alla posizione professionale e alla condizione occupazionale dei genitori, nonché alla possibilità di individuare i casi in cui i genitori risultassero presenti in famiglia. Utilizzando tale informazione sono state ricostruite variabili di background familiare

osservata la loro condizione occupazionale ad un anno di distanza.

Esistono molteplici approcci metodologici per integrare in modo efficiente informazioni provenienti da fonti statistiche eterogenee. In questo lavoro si è ritenuto di costruire un campione rappresentativo degli individui in cerca di lavoro, stratificato rispetto alla scelta formativa. La banca dati *Placement* rileva il totale dei formati FSE e può essere considerata rappresentativa dell'universo dei formati. L'universo dei non formati è tratto, invece, dalla indagine sulle Forze di Lavoro. Attraverso la rilevazione di aprile è possibile identificare quanti hanno preso parte a corsi di formazione nel mese immediatamente precedente l'intervista, così da depurare il sotto campione ISTAT dagli individui che hanno svolto attività formative nel periodo di riferimento. In dettaglio, gli individui del panel ISTAT che risultano aver preso parte ad attività di formazione professionale (formati) sono stati sostituiti con i beneficiari di corsi FSE provenienti dalla banca dati *Placement*. Una analisi descrittiva ha evidenziato come le caratteristiche di struttura dell'insieme dei formati osservati in entrambe le basi di dati siano risultate sufficientemente omogenee. Ad entrambi i sotto campioni sono stati applicati opportuni pesi per ottenere un campione rappresentativo del complesso degli individui in cerca di lavoro su base regionale, stratificato rispetto alla scelta di prendere parte ad attività di formazione professionale<sup>6</sup>.

## 2. *Il problema delle "selection bias"*

L'osservazione degli esiti occupazionali dei partecipanti ai corsi di formazione cofinanziati dal FSE evidenzia una alta percentuale di occupati ad un anno di distanza dalla fine dei corsi. Per valutare l'efficacia dei corsi nel facilitare l'accesso al mercato del lavoro occorre, tuttavia, domandarsi quale sarebbe stata la probabilità di trovare lavoro di questi stessi individui se non

---

omogenee con quelle del sotto campione ISTAT, in cui le variabili relative alla condizione dei genitori sono presenti per i soli individui che convivono con il nucleo familiare di origine.

<sup>6</sup> Tale approccio metodologico si discosta da quello adottato nell'ambito della precedente valutazione dell'Obiettivo 3 in Italia. In quel contesto, l'analisi delle condizioni occupazionali e dell'eventuale inserimento lavorativo dei beneficiari di intervento Fse è stata condotta con riferimento a una base di dati costruita con un matching deterministico tra la struttura dei formati Fse e il gruppo di confronto ISTAT. Tale procedura consiste nell'imporre al campione estratto dalla popolazione di riferimento ISTAT una struttura analoga, in termini di distribuzione rispetto ad alcune caratteristiche socio-demografiche, a quella osservata per il gruppo FSE. La metodologia seguita in questo lavoro non è intervenuta sulla composizione del gruppo di confronto.

avessero partecipato alle attività formative. La differenza tra le due probabilità rappresenta una misura dell'impatto netto della formazione sulla probabilità di trovare lavoro. La valutazione degli impatti netti comporta però un delicato problema di stima nella misurazione delle probabilità di occupazione nel caso di partecipazione o meno alle attività formative, in quanto l'informazione disponibile riguarda solo la condizione scelta mentre non si conoscono gli esiti corrispondenti all'alternativa. Diventa quindi necessario costruire una condizione controfattuale corrispondente all'alternativa inosservabile, con cui confrontare gli esiti occupazionali dei partecipanti.

In assenza di procedure sperimentali, che assegnino in modo casuale gli individui a due sotto campioni rispettivamente di destinatari e non dell'intervento formativo, il semplice confronto con un campione di individui che non abbiano partecipato all'intervento fornisce risultati distorti a causa del processo di selezione ed autoselezione del campione. Il problema dell'auto selezione del campione sorge in relazione al fatto che a monte degli esiti occupazionali, osservati rispettivamente per i formati e per i non formati, vi è la scelta che gli individui compiono di seguire un corso di formazione. Quest'ultima non è esogena rispetto agli esiti occupazionali. Si può infatti ritenere che alla base di tale decisione vi sia il confronto tra le probabilità di occupazione potenziali corrispondenti alle due alternative. In aggiunta o alternativamente, l'appartenenza al gruppo dei formati potrebbe essere legata alla diversa probabilità che gli individui hanno di venire a conoscenza della esistenza dei corsi. Al processo di autoselezione, messo in atto dai potenziali destinatari, si aggiunge la procedura di selezione realizzata dagli enti attuatori dei corsi, volta in molti casi a privilegiare gli individui maggiormente capaci.

Il problema di autoselezione può essere considerato un problema di stratificazione endogena del campione. Una caratteristica del campionamento stratificato endogenamente è quella, in generale, di non preservare la regressione. La presenza del meccanismo di autoselezione ha dunque l'effetto di introdurre non linearità nella relazione tra variabile dipendente e variabili esplicative. Per tenere conto dei meccanismi di selezione, connessi con la scelta di seguire i corsi, è stato utilizzato un modello appartenente alla famiglia dei modelli di selezione stocastici con "switching" endogeno.

Il metodo utilizzato, riconducibile al lavoro seminale di Heckman (1976), consente di ottenere stime non distorte anche in assenza di un gruppo sperimentale di controllo. La

correzione della distorsione dovuta ai meccanismi di autoselezione si basa sulla stima della probabilità di partecipazione ai corsi di formazione e più esattamente sulla stima della distorsione tramite l'inverso del rapporto di Mill.

### 3. *Descrizione del modello e metodologia di stima*

Gli impatti netti della formazione professionale sono stati stimati attraverso un modello Probit con correzione per gli effetti di distorsione indotti dalle *selection bias*. Tale modello consiste nella stima dell'occupabilità tenendo conto della decisione di partecipazione o meno ai corsi di formazione professionale. Il comportamento degli individui rispetto alla formazione è catturato da una funzione di scelta. Il modello consiste dunque di due equazioni, rispettivamente per la condizione occupazionale dei formati e dei non formati, e di una equazione per la funzione di scelta.

In questo contesto, la funzione di scelta (switching) svolge un ruolo simile a quello che, nel caso di una singola variabile endogena, svolgono le variabili strumentali nel depurare l'appartenenza ai formati piuttosto che ai non formati dalle componenti correlate con gli esiti. La stima dell'occupabilità tramite un semplice modello Probit fornirebbe, infatti, stime distorte verso l'alto a causa del nesso di endogeneità tra formazione e occupabilità.

Nel presente lavoro, la probabilità di occupazione viene stimata per ogni individuo appartenente al campione utilizzando variabili relative alle caratteristiche anagrafiche, al grado di istruzione, al contesto territoriale e a quello familiare<sup>7</sup>. In particolare, la caratteristica dell'indagine sulle Forze di Lavoro di utilizzare la famiglia come unità di rilevazione ha consentito di ricostruire ed utilizzare variabili di background familiare presenti anche nelle indagini regionali e di particolare rilevanza nella stima della funzione di scelta individuale. Quest'ultima viene stimata mediante un modello Probit sulla variabile dicotomica formazione/non formazione, regredita su un insieme di variabili di contesto, di caratteristiche socio-demografiche degli individui e su un insieme di variabili strumentali non incluse nella stima dell'occupabilità. In particolare come strumenti sono state utilizzate sia variabili relative

---

<sup>7</sup> Le variabili considerate sono: il titolo di studio, l'eventuale tipo di diploma, la classe di età, la condizione occupazionale del padre e della madre se coabitanti, la posizione professionale del padre e della madre se coabitanti, la condizione o meno di capofamiglia, il tasso di disoccupazione provinciale in

all'interruzione di corsi scolastici, sia variabili riguardanti la posizione professionale (attuale o precedente il pensionamento) dei genitori. Inoltre, le variabili relative alla dinamica del mercato del lavoro nel contesto territoriale sono, in questo caso, riferite all'anno di svolgimento del corso.

Vengono, dunque, stimate tre equazioni: 1) la probabilità di occupazione avendo seguito corsi di formazione; 2) la probabilità di occupazione non avendo seguito corsi di formazione; 3) la probabilità di seguire un corso di formazione (chiamata "funzione di scelta"). In tutti e tre i casi la variabile dipendente è una variabile dicotomica che assume valore 1 rispettivamente per gli occupati e per i partecipanti ai corsi e zero altrimenti. Assumiamo, pertanto, l'esistenza di tre variabili latenti espresse in forma ridotta,  $I^*$ ,  $O^*_F$  e  $O^*_{NF}$ , che costituiscono le variabile dipendenti rispettivamente della funzione di decisione e delle due funzioni per la condizione occupazionale. La specificazione del modello di selezione con "switching" endogeno è la seguente:

- 1)  $O^*_F = X_F \beta_F + \omega_F$  (se si è seguito il corso di formazione)
- 2)  $O^*_{NF} = X_{NF} \beta_{NF} + \omega_{NF}$  (se non si è seguito il corso di formazione)
- 3)  $I^* = Z\gamma + \varepsilon$  (decisione di partecipazione)

in cui  $X_F$ ,  $X_{NF}$  e  $Z$  sono vettori di variabili demografiche relative all'individuo e alla sua famiglia. Se  $O^*_F > 0$  l'individuo formato risulta occupato l'anno successivo, similmente se  $O^*_{NF} > 0$  risulta occupato l'individuo non formato. Se  $I^* > 0$  l'individuo partecipa al corso di formazione.

Il modello consente di stimare le probabilità potenziali di occupazione nel caso si sia seguito un corso di formazione e nel caso opposto per tutti gli individui, indipendentemente dalla scelta compiuta (e osservata) riguardo la partecipazione ai corsi. Attraverso questo modello si è cercato di rispondere a due domande distinte: qual è stato l'effetto dei corsi di formazione sulla probabilità di trovare lavoro degli individui che hanno effettivamente seguito i corsi; quale sarebbe stato l'effetto della partecipazione ai corsi sulla probabilità di trovare lavoro degli individui che non vi hanno partecipato (pur appartenendo alla stessa popolazione di riferimento).

---

base al genere, il tasso di variazione dell'occupazione, la regione e la provincia di appartenenza e alcune



Dividendo la popolazione di riferimento (target potenziale) in due categorie, i formati e i non formati, le due questioni coincidono solo se tra le due categorie non ci sono differenze di rilievo che possano influire sulla capacità di trarre beneficio dai corsi di formazione. In tal caso l'effetto dei corsi sui due gruppi non sarebbe significativamente diverso. La risposta alla prima domanda è rilevante per valutare l'opportunità di ripetere in futuro programmi simili. La risposta alla seconda potrebbe essere rilevante nel caso si decidesse di espandere la platea di beneficiari.

La presente analisi ha cercato dunque di quantificare quale sarebbe stata la probabilità di occupazione dei formati se non avessero seguito il corso e il converso per i non formati. Poiché le effettive osservazioni sugli esiti occupazionali sarebbero disponibili solo per "i formati" per quanto riguarda il caso di partecipazione a un corso di formazione e per i "non formati" per quanto riguarda il caso opposto, le probabilità di occupazione nei casi contrari si basano sulle previsioni del modello.

La differenza tra le probabilità di occupazione stimate nei due casi, rispettivamente di partecipazione o meno all'intervento, consente di misurare l'impatto netto della formazione. Il valore medio di tale differenza, condizionalmente al gruppo di appartenenza, fornisce una risposta alle domande sopra riportate. L'effetto della formazione sull'intera popolazione, anziché sui due sotto gruppi, sarà invece dato dalla differenza tra le due probabilità stimate sull'intera popolazione. Si noti che la disponibilità di dati sperimentali consentirebbe di rispondere in modo accurato alla prima domanda ma non fornirebbe alcuna risposta alla seconda. I dati sperimentali si basano infatti sulla attribuzione casuale degli individui al gruppo dei formati o dei non formati nell'ambito di coloro che hanno presentato la domanda di partecipazione o che si sono dimostrati disponibili a partecipare al programma. Ne segue che tali dati consentono di appurare l'effetto dei corsi di formazione su coloro che hanno presentato domanda, ma se da questo volessimo desumere l'effetto dei corsi di formazione sulla intera popolazione che soddisfa i requisiti per la partecipazione dovremmo tenere conto del meccanismo di autoselezione connesso alla presentazione della domanda di partecipazione.

---

variabili di interazione tra titolo di studio e sesso.

#### 4. Risultati

Il modello utilizzato fornisce, da un lato, una misura dell'importanza delle variabili di contesto e delle caratteristiche socio-demografiche individuali come determinanti della decisione di fare formazione e indica come queste interagiscono nel determinare la condizione del singolo nel mercato del lavoro; dall'altro, tale schema analitico consente di ottenere stime consistenti dell'impatto netto delle attività di formazione professionale sugli individui in cerca di lavoro, pervenendo in tal modo a una quantificazione degli effetti di *selection bias*. Questo esercizio di stima si colloca nel contesto dei modelli per la correzione della *sample selection* e costituisce un tentativo preliminare di stimare gli impatti netti correggendo le distorsioni connesse al processo di selezione a partire dalle indagini regionali sugli esiti occupazionali degli interventi. Questi risultati vanno pertanto considerati provvisori, tuttavia forniscono utili indicazioni per indirizzare sia le future rilevazioni sugli esiti occupazionali sia le scelte metodologiche nella stima degli impatti netti. In particolare il modello consente di stimare con sufficiente grado di accuratezza la probabilità di prendere parte ad attività formative, mentre più debole risulta la stima della probabilità di occupazione. A tale proposito si segnala l'utilità di raccogliere al momento dell'iscrizione ad un intervento formativo maggiori informazioni sulla storia occupazionale dell'individuo (durata della disoccupazione, numero e tipo di esperienze lavorative precedenti).

In appendice si riportano le tavole relative alla stima del modello, mentre nel testo si presentano le corrispondenti stime degli effetti marginali, di più facile lettura. La tavola A.1 (in appendice) si riferisce alla stima della probabilità di occupazione con selezione del campione nel caso di partecipazione ad attività formative. L'occupabilità è stata modellata come una funzione dei tassi di disoccupazione e di variazione dell'occupazione su base provinciale, di *dummy* territoriali, di variabili relative alla condizione professionale dei genitori dell'individuo formato e di altre caratteristiche individuali quali l'età e il titolo di studio conseguito. Delle diverse opportunità di lavoro in base al genere si è tenuto conto tramite variabili di interazione tra titolo studio e sesso e tramite tassi di disoccupazione provinciale differenziati in base al genere.

La seconda parte della tavola riporta la stima della funzione di *switching* endogeno. La decisione di prendere parte ad attività formative è significativamente influenzata da variabili di contesto territoriale (un peggioramento della condizione del mercato del lavoro provinciale favorirebbe un accesso ai canali della formazione professionale). Tra le caratteristiche socio-

demografiche dell'individuo, un innalzamento dell'età scoraggerebbe decisioni di partecipazione ad attività formative; in particolare positivo e significativo è l'effetto della appartenenza alla classe di età 14-19 per i giovani di sesso maschile. L'appartenenza alla classe di età compresa tra i 20 e i 24 anni e al genere femminile risulta, invece, poco significativa.

Il possesso di un diploma e in misura maggiore di una laurea aumenterebbero la probabilità di prendere parte ai corsi di formazione, soprattutto per la componente femminile, mentre negativo sarebbe l'effetto della interruzione di un corso di diploma; il conseguimento di titoli di scuola media superiore professionalizzanti (diploma tecnico e professionale) aumenterebbero la probabilità di prendere parte ai corsi di formazione. Tali risultati, se confermati indicherebbe il venire meno del ruolo di supplenza dei corsi formativi regionali rispetto alla formazione scolastica. Ricordiamo infatti come in origine gli interventi di formazione professionale regionali si rivolgessero prevalentemente ai drop-out della secondaria superiore e ai giovani meno scolarizzati, venendo pertanto a coincidere con la formazione di base dei giovani in uscita dall'obbligo.

Per quanto riguarda l'equazione che descrive gli esiti occupazionali, l'effetto delle diverse caratteristiche sulla probabilità di occupazione può essere desunto dalla considerazione degli effetti marginali, riportati in tavola 1.

***Tavola 1: Effetti marginali sulla probabilità di trovare lavoro partecipando ad attività formative***

Effetti marginali del modello con sample selection							
Y = Pr(occup=1) (predict) = 0.61875							
Variable	dy/dx	Std. Err.	z	P> z	[ 95%	C.I. ]	X
disgen98	-0.004	0.001	-3.160	0.002	-0.007	-0.002	8.489
var_occ97	0.857	0.364	2.360	0.019	0.144	1.569	0.015
lomb*	0.030	0.015	1.980	0.047	0.000	0.061	0.650
Sondrio*	-0.023	0.064	-0.360	0.719	-0.149	0.103	0.012
Reggioe*	0.116	0.049	2.370	0.018	0.020	0.212	0.031
co_occ*	0.022	0.016	1.410	0.158	-0.008	0.052	0.285
cf_pen*	0.015	0.017	0.880	0.381	-0.019	0.049	0.242
co_pen*	0.022	0.023	0.940	0.345	-0.023	0.067	0.133
eta30_44*	-0.088	0.025	-3.570	0.000	-0.136	-0.039	0.116
laurea*	0.216	0.017	12.620	0.000	0.182	0.249	0.228

tec_prof*	0.115	0.018	6.340	0.000	0.079	0.150	0.345
dip_mas*	-0.031	0.023	-1.320	0.187	-0.076	0.015	0.227
dip_fem*	-0.079	0.020	3.910	0.000	0.039	0.119	0.333

(\*) Effetto marginale (dy/dx) relativo a variazioni discrete delle variabili dummy da 0 a 1.

La tavola indica che quest'ultima è influenzata dalle condizioni del mercato locale del lavoro, dipendendo positivamente dal tasso di variazione dell'occupazione provinciale e negativamente dal tasso di disoccupazione provinciale differenziato in base al genere. Per quanto riguarda lo status occupazionale dei genitori, la condizione di occupato o pensionato piuttosto che di disoccupato o inattivo presenta un impatto positivo. Come atteso, il conseguimento della laurea favorisce l'occupabilità. Per quanto riguarda le donne anche il conseguimento di un titolo di diploma presenta un effetto positivo, sebbene meno pronunciato. Per quanto riguarda gli uomini, invece, la variabile non è risultata significativa. Sempre nell'ambito dei diplomati, il conseguimento di un titolo professionalizzante (diploma tecnico e professionale) avrebbe un effetto positivo e significativo sulla probabilità di trovare lavoro. Al contrario, tale probabilità si riduce per gli appartenenti alla classe di età compresa tra i 30 e i 44 anni.

Il modello Probit con correzione per l'autoselezione riportato nella tavola A.2 (in appendice) si riferisce al caso di non partecipazione ad esperienze formative. Nella seconda parte della tavola, si riportano i parametri stimati della funzione di selezione. Essa presenta una struttura analoga a quella riportata in tavola A1, con i segni dei parametri invertiti.

La tavola 2 consente di svolgere l'analisi sulle determinanti della probabilità di inserimento occupazionale in termini di effetti marginali. Nuovamente si rileva l'influenza negativa del tasso di disoccupazione provinciale sulla probabilità di trovare lavoro. In questo caso, tuttavia, le variabili che esprimono rispettivamente il tasso di variazione dell'occupazione e la residenza nella regione Lombardia sono risultate non significative. Nuovamente, il possesso di una laurea facilita l'accesso al mercato del lavoro, mentre nel caso dei maschi al possesso del diploma risultano associate minori probabilità occupazionali. Anche nel caso di non partecipazione ad attività formative l'appartenenza alla classe di età 30-44 diminuisce la probabilità di trovare lavoro ad un anno di distanza. Di lettura meno facile è risultato il contributo delle variabili relative alla condizione occupazionale dei genitori. In particolare per quanto riguarda la

condizione del coniuge del capofamiglia, tali variabili sembrerebbero avere un impatto di segno diverso nel caso di partecipazione ai corsi rispetto al caso contrario.

**Tavola 2: Effetti marginali sulla probabilità di trovare lavoro non partecipando ad attività formative**

Effetti marginali del modello con sample selection							
Y = Pr(occup=1) (predict) = 0.39237							
Variabile	dy/dx	Std. Err.	Z	P> z	[ 95% C.I. ]		X
disgen98	-0.031	0.008	-3.960	0.000	-0.046	-0.016	8.489
var_occ97	1.461	1.633	0.890	0.371	-1.740	4.662	0.015
lomb*	0.018	0.062	0.300	0.767	-0.103	0.139	0.650
Sondrio*	0.325	0.143	2.270	0.023	0.045	0.606	0.012
Reggioe*	0.599	0.087	6.890	0.000	0.428	0.769	0.031
co_occ*	-0.228	0.075	-3.050	0.002	-0.375	-0.082	0.285
cf_pen*	0.319	0.085	3.750	0.000	0.152	0.485	0.242
co_pen*	-0.241	0.096	-2.510	0.012	-0.430	-0.053	0.133
eta30_44*	-0.134	0.064	-2.100	0.036	-0.260	-0.009	0.116
laurea*	0.524	0.091	5.750	0.000	0.346	0.703	0.228
tec_prof*	0.163	0.111	1.460	0.144	-0.055	0.380	0.345
dip_mas*	-0.284	0.110	-2.590	0.010	-0.499	-0.069	0.227
dip_fem*	0.084	0.105	0.810	0.421	-0.121	0.289	0.333

(\*) Effetto marginale (dy/dx) relativo a variazioni discrete delle variabili dummy da 0 a 1.

Nel complesso, il modello sembra stimare con sufficiente accuratezza la funzione di selezione mentre l'equazione che descrive gli esiti occupazionali presenta una bassa capacità esplicativa soprattutto a causa della mancanza di informazioni sulla storia lavorativa passata degli individui. In ogni caso, è risultato giustificato l'uso di un modello con correzione degli effetti di selezione, come risulta dalla significatività del coefficiente di correlazione tra i residui della funzione di selezione e della probabilità di occupazione.

Sulla base delle stime sopra riportate sono stati calcolati gli impatti netti relativamente ai segmenti maschile e femminile della popolazione.

La tavola 3 riporta il calcolo degli impatti netti per il gruppo dei formati, ottenuti come differenza tra la probabilità di occupazione (media) corrispondente alla scelta di partecipare ai

corsi di formazione e quella attinente alla decisione di non parteciparvi. Nella tavola si confrontano i risultati ottenuti dal modello sopra presentato con quelli ottenuti stimando un tradizionale modello Probit, che non consente di correggere le distorsioni connesse con l'autoselezione del campione. Come ci si attendeva, il modello con autoselezione corregge verso il basso la stima degli impatti netti rispetto al modello standard, come risultato di un aggiustamento verso l'alto della probabilità di occupazione corrispondente alla non partecipazione ai corsi, e verso il basso di quella corrispondente alla partecipazione.

**Tav.3 - Calcolo degli impatti netti per il gruppo dei formati**

(Confronto tra i risultati del modello con autoselezione e del modello semplice; valori percentuali)

	Modello con autoselezione		Modello Probit		Percentuale di occupati osservata	
	Uomini	Donne	Uomini	Donne	Uomini	Donne
Probabilità di occupazione						
Non partecipando ai corsi	36,74	44,64	33,76	41,59	-	-
Partecipando ai corsi	59,16	63,45	66,08	70,84	68,54%	70,79%
Impatto netto	22,42	18,81	32,32	29,25		

Questi risultati sono conformi all'ipotesi di esistenza di differenze inosservabili tra formati e non formati, attinenti alla sfera delle motivazioni e delle abilità individuali, che comportano una maggiore occupabilità degli appartenenti al gruppo dei formati a prescindere dalla partecipazione ad attività formative. Questo implica una sovrastima dell'impatto netto ottenuto come semplice differenza tra i risultati relativi ai partecipanti ai corsi e quelli di un gruppo di confronto composto da non partecipanti, anche correggendo per differenze osservabili tra i due gruppi.

La tavola riporta, inoltre, la percentuale osservata di individui partecipanti alle attività formative che risultano occupati ad un anno di distanza dal corso. Tali valori risultano, per il contesto territoriale considerato, particolarmente elevati. In particolare, la maggiore proporzione di donne occupate rispetto agli uomini si spiega in buona parte considerando il maggiore grado di istruzione e il contesto sociale più favorevole osservato per le partecipanti ai corsi di formazione.

L'impatto netto, stimato correggendo le *selection bias*, è risultato pari al 22,42% per gli uomini e al 18,81% per le donne, corrispondente ad un incremento della probabilità di occupazione, rispettivamente, del 61% e del 42%. L'impatto netto è, dunque, lievemente inferiore per le donne, cui corrisponde un minore incremento percentuale. Nonostante tale evidenza, la probabilità di occupazione delle donne nel caso di partecipazione ai corsi è risultata più elevata. Occorre tuttavia considerare che le donne coinvolte in attività formative rappresentano un segmento relativamente forte della popolazione femminile. La correzione attuata tramite il modello con autoselezione è risultata sostanzialmente analoga per i due segmenti della popolazione. In entrambi i casi la probabilità si riduce, infatti, di circa 7 punti nel caso di partecipazione mentre aumenta di circa 3 punti in quello di non partecipazione ai corsi di formazione.

La tavola 3 consente di individuare l'effetto dei corsi di formazione sulla probabilità di trovare lavoro degli individui che hanno effettivamente seguito i corsi; la tavola 4 riporta, invece, l'effetto che la partecipazione ai corsi avrebbe avuto sulla probabilità di trovare lavoro degli individui che non vi hanno effettivamente partecipato. Tale tavola riguarda pertanto l'effetto potenziale che la formazione potrebbe avere su soggetti attualmente non coinvolti in tali attività.

**Tav. 4 - Calcolo degli impatti netti per il gruppo dei non-formati**

(Confronto tra i risultati del modello con autoselezione e del modello semplice; valori percentuali)

	Modello con autoselezione		Modello Probit		Percentuale di occupati osservata	
	Uomini	Donne	Uomini	Donne	Uomini	Donne
Probabilità di occupazione						
Non partecipando ai corsi	43,14	29,41	41,49	27,68	43,9%	26,7%
Partecipando ai corsi	56,54	51,91	63,07	59,64	-	-
Impatto netto	13,40	22,50	21,58	31,96		

Nuovamente, il modello con autoselezione produce una correzione verso il basso degli impatti potenziali netti. Questi ultimi sono stati stimati pari al 13,4% per gli uomini e al 22,5% per le donne, corrispondenti ad un incremento della probabilità di occupazione di circa un terzo

per gli uomini e superiore al 75% per le donne. Il risultato stimato per le donne è di particolare interesse. In questo caso, a differenza di quanto accadeva per la componente femminile del gruppo dei formati, le donne rappresentano un segmento debole del mercato del lavoro. Il tasso di inserimento osservato, così come la probabilità di occupazione in assenza di formazione, sono infatti molto al di sotto di quelli maschili. La formazione professionale produrrebbe l'effetto di elevare la probabilità di occupazione, in particolar modo quella femminile, portandola a valori superiori al 50% per entrambi i sessi.

Come si è detto si tratta di risultati ancora provvisori e tuttavia se risultassero confermati da ulteriore ricerca ne emergerebbe l'indicazione di una forte efficacia che i corsi di formazione potrebbero avere su questo segmento della popolazione e pertanto l'utilità di estendere la partecipazione femminile alle attività formative. In particolare tali risultati segnalerebbero l'efficacia delle politiche di *main streaming* nel favorire lo sviluppo di pari opportunità delle donne sul mercato del lavoro, in riferimento a quel segmento della popolazione attualmente non coinvolto nei corsi di formazione. In particolare, ne emergerebbe l'utilità del coinvolgimento in interventi formativi proprio dei segmenti femminili più deboli sul mercato del lavoro.

La tavola 5 riporta i valori degli impatti netti relativi all'intera popolazione degli individui in cerca di occupazione. Essa misura l'effetto potenziale medio della formazione sull'occupabilità, indipendentemente dalla scelta di partecipazione ad attività formative e, in questo senso, racchiude i risultati contenuti nelle due tavole precedenti. Dato il maggior peso della popolazione non coinvolta in interventi formativi i valori riportati riflettono in larga parte i risultati descritti nella tavola 4. Nel complesso, l'impatto netto è risultato pari a circa il 13,8% per gli uomini e del 22,4% per le donne.

**Tav.5 - Calcolo degli impatti netti per l'intera popolazione (formati, non formati)**

(Confronto tra i risultati del modello con autoselezione e del modello semplice; valori percentuali)

	Modello con autoselezione		Modello Probit	
	Uomini	Donne	Uomini	Donne
Probabilità di occupazione				
Non partecipando ai corsi	42,88	29,90	41,18	28,13
Partecipando ai corsi	56,64	52,28	63,19	60,0
Impatto netto	13,76	22,38	22,01	31,87



L'analisi presentata conferma l'esistenza di un significativo impatto delle attività formative cofinanziate dal Fondo Sociale Europeo sulle probabilità di inserimento occupazionale, sebbene riveda verso il basso gli impatti stimati trascurando gli effetti di selezione. Essa si caratterizza per l'uso di una metodologia che consente di tener conto di differenze tra il gruppo dei formati e il gruppo di confronto inerenti caratteristiche non osservabili correlate con gli esiti occupazionali. I destinatari di interventi formativi presentano, a parità di altre caratteristiche osservabili, maggiori chances occupazionali spiegabili, probabilmente, con un maggior grado di attivismo e di motivazione nella ricerca di lavoro oltre a più elevate abilità e capacità individuali. Tali differenze nelle probabilità di trovare lavoro, pre-esistenti la eventuale partecipazione ai corsi, giustificano l'impiego di un modello di selezione nella stima degli impatti netti.

Nella implementazione pratica, questo approccio metodologico è risultato promettente e, tuttavia, l'applicazione di tale metodo richiede di ampliare l'insieme informativo utilizzato. In particolare, ne emergono delle indicazioni utili per la realizzazione delle future indagini regionali sugli esiti occupazionali in riferimento ai seguenti aspetti: 1) risulta opportuno, nella raccolta di dati, considerare variabili e modalità comparabili con quelle rilevate attraverso l'indagine sulle Forze di Lavoro dell'ISTAT; 2) al fine di stimare in modo più consistente la probabilità di trovare lavoro, è particolarmente rilevante, al momento dell'iscrizione ai corsi, rilevare informazioni sulla storia lavorativa passata dei destinatari; 3) nella stima delle probabilità di partecipare ad attività formative sono particolarmente rilevanti le variabili relative al background familiare, con particolare riferimento al titolo di studio e alla posizione professionale dei genitori; sarebbe inoltre utile, per la stima del modello relativo alle donne, inserire informazioni sul numero di figli e sui carichi familiari.

## **Appendice**

*Tavola A1: Probabilità di trovare lavoro partecipando ad attività formative*

Modello Probit con sample selection						
				Number of obs	=	6707
				Censored obs	=	291
				Uncensored obs	=	6416
Log likelihood = -50752.28				Wald chi2(10)	=	240.95
				Prob > chi2	=	0.000
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
<b>occup</b>						
disgen98	-0.012	0.004	-3.19	0.001	-0.019	-0.004
Var_occ97	2.248	0.957	2.35	0.019	0.372	4.124
lomb	0.080	0.041	1.97	0.049	0.000	0.159
sondrio	-0.060	0.166	-0.36	0.717	-0.385	0.265
reggioe	0.324	0.148	2.2	0.028	0.035	0.614
co_occ	0.058	0.041	1.4	0.161	-0.023	0.138
cf_pen	0.040	0.046	0.87	0.383	-0.050	0.131
co_pen	0.057	0.061	0.94	0.349	-0.063	0.177
eta30_44	-0.226	0.062	-3.66	0.000	-0.347	-0.105
laurea	0.611	0.055	11.08	0.000	0.503	0.719
tec_prof	0.307	0.049	6.23	0.000	0.211	0.404
dip_mas	-0.080	0.060	-1.33	0.182	-0.197	0.037
dip_fem	0.212	0.056	3.8	0.000	0.103	0.322
<b>formato</b>						
dispro97	0.096	0.018	5.22	0.000	0.060	0.132
rimini	-0.831	0.152	-5.45	0.000	-1.130	-0.532
brescia	-0.327	0.173	-1.9	0.058	-0.666	0.011
reggioe	-1.888	0.420	-4.5	0.000	-2.710	-1.066
cleta	-0.239	0.042	-5.69	0.000	-0.322	-0.157
mas_19	0.593	0.198	2.99	0.003	0.204	0.981
fem20_24	-0.138	0.131	-1.06	0.291	-0.394	0.118
cf_int	0.718	0.201	3.58	0.000	0.325	1.112
co_int	0.491	0.206	2.38	0.017	0.086	0.895
dip_mas	0.643	0.161	4	0.000	0.328	0.957
dip_fem	0.753	0.108	6.99	0.000	0.542	0.964
lau_mas	0.718	0.169	4.25	0.000	0.386	1.049
lau_fem	1.345	0.182	7.39	0.000	0.988	1.701
figlio	-0.180	0.122	-1.47	0.141	-0.420	0.060
int_dip	-0.414	0.156	-2.65	0.008	-0.720	-0.108
int_laur	0.335	0.186	1.8	0.072	-0.029	0.699
_cons	-2.180	0.222	-9.81	0.000	-2.616	-1.745
Rho	.1244241	.219857			.0811248	.1672548
Wald test of indep. eqns. (rho = 0):	chi2(1) =	31.37			Prob > chi2 =	0.000

*TavolaA2: Probabilità di trovare lavoro non partecipando ad attività formative*

Modello Probit con sample selection						
				Number of obs	=	6707
				Censored obs	=	6416
				Uncensored obs	=	291
				Wald chi2(10)	=	80.16
Log likelihood = -279115.6				Prob > chi2	=	0.000
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
<b>occup</b>						
disgen98	-0.080	0.021	-3.820	0.000	-0.121	-0.039
Var_occ97	3.802	4.261	0.890	0.372	-4.549	12.154
lomb	0.048	0.161	0.300	0.768	-0.268	0.364
sondrio	0.848	0.419	2.020	0.043	0.027	1.670
reggioe	2.161	0.807	2.680	0.007	0.579	3.742
co_occ	-0.631	0.227	-2.780	0.005	-1.076	-0.186
cf_pen	0.822	0.229	3.590	0.000	0.373	1.271
co_pen	-0.708	0.333	-2.120	0.034	-1.361	-0.055
eta30_44	-0.368	0.180	-2.040	0.041	-0.722	-0.015
laurea	1.436	0.321	4.480	0.000	0.808	2.065
tec_prof	0.419	0.287	1.460	0.145	-0.144	0.983
dip_mas	-0.824	0.394	-2.090	0.036	-1.596	-0.052
dip_fem	0.218	0.268	0.810	0.417	-0.308	0.743
<b>nonform</b>						
dispro97	-0.074	0.022	-3.350	0.001	-0.117	-0.031
rimini	0.928	0.235	3.960	0.000	0.468	1.388
brescia	0.293	0.161	1.820	0.069	-0.023	0.608
reggioe	1.397	0.459	3.040	0.002	0.497	2.297
cleta	0.225	0.052	4.340	0.000	0.123	0.326
mas_19	-0.881	0.212	-4.150	0.000	-1.297	-0.465
fem20_24	0.292	0.161	1.820	0.069	-0.023	0.607
cf_int	-0.619	0.288	-2.150	0.032	-1.183	-0.054
co_int	-0.573	0.248	-2.310	0.021	-1.060	-0.087
dip_mas	-0.563	0.176	-3.200	0.001	-0.908	-0.218
dip_fem	-0.822	0.108	-7.590	0.000	-1.035	-0.610
lau_mas	-0.600	0.174	-3.450	0.001	-0.940	-0.260
lau_fem	-1.280	0.197	-6.480	0.000	-1.667	-0.892
figlio	0.275	0.140	1.970	0.049	0.001	0.550
int_dip	0.655	0.268	2.440	0.015	0.129	1.181
int_laur	-0.406	0.222	-1.830	0.067	-0.841	0.028
_cons	1.970	0.295	6.670	0.000	1.391	2.549
Rho	-.831847	.1559567			-.9750845	-.1990733
Wald test of indep. eqns. (rho = 0):	chi2(1) =	5.56			Prob > chi2 =	0.0184

*Tavola A3: Variabili utilizzate per la stima del modello con sample selection*

---

**Legenda**

---

Equazione 1: condizione occupazionale

disgen98	tasso di disoccupazione per genere (anno 1998)
var_occ97	variazione del tasso di occupazione (anno 1997)
lomb	dummy regionale (Lombardia = 1)
Sondrio	dummy provinciale (Sondrio = 1)
Reggioe	dummy provinciale (Reggio Emilia = 1)
co_occ	condizione occupazionale del coniuge - occupato
cf_pen	condizione occupazionale del capofamiglia - pensionato
co_pen	condizione occupazionale del coniuge - pensionato
eta30_44	classe di età, 30-44 anni
laurea	titolo di studio, laurea
tec_prof	titolo di studio, diploma tecnico, diploma professionale
dip_mas	titolo di studio, diploma, uomini
dip_fem	titolo di studio, diploma, donne

Equazione 2: funzione di scelta (formato/non formato)

dispro97	tasso di disoccupazione per provincia (anno 1997)
Rimini	dummy provinciale (Rimini = 1)
Brescia	dummy provinciale (Brescia = 1)
Reggioe	dummy provinciale (Reggio Emilia = 1)
cleta	classi di età
mas_19	classe di età, fino a 19 anni, uomini
fem20_24	classe di età, 20-24 anni, donne
cf_int	posizione occupazionale del capofamiglia, intermedio
co_int	posizione occupazionale del coniuge, intermedio
dip_mas	titolo di studio, diploma, uomini
Dip_fem	titolo di studio, diploma, donne
Lau_mas	titolo di studio, laurea, uomini
Lau_fem	titolo di studio, laurea, donne
figlio	relazione con la famiglia di origine
Int_dip	interruzione del corso di studi - diploma
int_laur	interruzione del corso di studi - laurea
_cons	costante

---

## **Bibliografia**

Centra, M., Discenza, R. A. e Rustichelli, E. (2001), “Una Procedura per la Ricostruzione della Struttura Longitudinale della Rilevazione Trimestrale ISTAT sulle Forze del Lavoro”, Monografie sul mercato del lavoro e le politiche per l’impiego, ISFOL.

Centra, M., De Vincenti, R. e Villante, C. (2000), Formazione Professionale e Occupabilità, I libri del Fondo Sociale Europeo, ISFOL.

Heckman, J.J., Ichimura, H., Todd, P.E. (1997), "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme", Review of Economic Studies, volume 64.

Heckman, J. e Smith, J. (1996), Experimental and non Experimental Evaluation, Schmid, G. et alii (a cura di), 1996.

Heckman, J. (1993), “Alternative Approaches to the Evaluation of Social Programs”, Dipartimento di Economia, Università di Chicago, manoscritto non pubblicato.

Heckman, J. (1979), “Sample Selection Bias as a Specification Error,” *Econometrica*, volume 47, n. 1.

Heckman, J. (1976), “The Common Structure of Statistical Model of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for such Models”, *Annals of Economics and Social Measurement*, n. 5, 475-92.

ISFOL - Struttura Nazionale di Valutazione (1999), Valutazione Finale dell’Obiettivo 3 in Italia, Franco Angeli.

ISFOL - Struttura Nazionale di Valutazione (2001), “L’impatto del FSE sulla Formazione nel Centro-Nord in Italia: Apprendimenti per la Programmazione in Corso”, documento per il Gruppo Tecnico di Valutazione del 31 Maggio 2001.

Maddala, G.S. (1983), Limited-Dependent and Qualitative Variables in

Econometrics, Cambridge University Press.

Manski, F. C., Lerman, R. S. (1977), "The estimation of Choice Probabilities from Choice Based Samples", *Econometrica*, vol. 45, 8.