

L'efficacité des politiques de formation: méthodes pour l'évaluation de l'impact¹

Ragazzi, Elena

CNR-Ceris, Chercheur, via Real Collegio 30,10024 Moncalieri (To), E-mail : e.ragazzi@ceris.cnr.it, Tel : +390116824930, Fax :+390116824966

Sella Lisa

CNR-Ceris, Allocataire de Recherche, via Real Collegio 30,10024 Moncalieri (To), E-mail : l.sella@ceris.cnr.it, Tel : +390116824950, Fax :+390116824966

Résumé: Notre proposition aborde le thème de l'évaluation de l'impact net des politiques dans un contexte non expérimental. La discussion part du cas d'évaluation des politiques de formation professionnelle avec cofinancement du FSE en Piémont, une région au Nord-ouest de l'Italie.

L'évaluation d'impact est souvent négligée, surtout dans le cas des actions très vastes, à cause du coût de la récolte d'informations, de la difficulté d'apprécier précisément l'impact net, mais aussi pour des raisons de pratiques répandues dans la gouvernance des politiques mêmes. Par contre l'évaluation d'impact joue un rôle extraordinaire dans le succès d'une ligne d'intervention, en tant qu'outil fondamental de programmation (CGD 2006 ; White 2009) pour apprécier la réelle capacité des programmes à produire les effets qu'on en attend.

Dans le contexte actuel, les politiques de cohésion sociale et de providence contre les effets de la crise sont fondamentales pour réduire le manque d'emploi et les privations qui en sont la conséquence, surtout pour les groupes les plus faibles de la population et dans des contextes locaux de transition économique. Par contre, la disponibilité de ressources financières pour les politiques d'État-providence sont de plus en plus insuffisantes et il faut donc les concentrer sur les mesures les plus efficaces. Les administrations publiques, nationales et locales, doivent apprendre à penser en termes d'évaluation ex-ante, en ayant à disposition un ensemble efficace d'indicateurs, et des systèmes pour les enregistrer en temps et en heure et de façon peu coûteuse, et d'évaluation d'impact (Barca 2009 et 2012) pour rendre possible une programmation et une réorientation des actions en cours pour la relance du développement et de la cohésion, avec une requalification simultanée de l'action de gouvernance.

Cet article propose une stratégie d'évaluation quasi expérimentale sur l'insertion dans le marché du travail en aval de la formation. Nous expliquons et motivons les stratégies d'extraction d'un échantillon représentatif des qualifiés et des spécialisés dans une sélection de cours de formation achevés dans les années 2010 (Ragazzi et al, 2012) mais nous discutons également de la possibilité de mener une analyse de même efficacité en exploitant les données administratives de suivi ou surveillance, au lieu d'une enquête directe auprès des bénéficiaires.

Une partie importante de cette communication concerne aussi la possibilité d'évaluer l'impact net et d'isoler un échantillon de contrôle qui convienne ainsi à la nécessité de réduire le selection bias sur les techniques d'analyse qui peuvent être appliquées. Les avantages et les défauts des différentes stratégies sont évalués à partir de l'expérience d'évaluation des politiques formatives régionales (avec le financement du FSE) réalisé en Piémont (Italie), sur les bénéficiaires des politiques formatives et d'insertion au travail menées en 2010. La recherche a montré l'existence d'un effet net positif sur les chances d'emploi des individus, surtout sur les adultes à faible scolarité. Les politiques formatives adoptées au niveau local semblent agir sur les situations de handicap, mais n'arrivent pas à l'annuler complètement, vues les caractéristiques de fragilité des emplois trouvés.

Mots-clés : évaluation, politiques régionales, formation professionnelle.

¹ Les données présentées dans cet article ont été créées dans le cadre du service « Valutazione del POR FSE della Regione Piemonte ob. 2 "competitività regionale e occupazione" per il periodo 2007-2013 ». ("Evaluation de POR SFE de la région Piémont ob. 2 "Compétitivité régionale et emploi" pour la période 2007-2013"). Nous remercions vraiment la Région Piémont, qui reste propriétaire des données, de nous permettre de les employer à des fins de recherches et pour améliorer les méthodes d'analyse.

1. Introduction

Les analyses d'impact, bien que déjà reconnues depuis longtemps comme une sorte de «règle d'or» de l'évaluation, ont connu un succès limité dans les administrations publiques européennes et des applications assez épisodiques, surtout en ce qui concerne l'estimation des impacts nets. Les méthodes dites contrefactuelles, ont, en fait, été principalement utilisées dans les évaluations expérimentales de peu de projets pilotes, mais pas dans des contextes quasi-expérimentaux et de vaste programmation. Ceci est en grande partie dû à la complexité de leur conception et de leur mise en œuvre. De nos jours, cependant, la crise économique et la nécessité qui en découle d'utiliser les ressources publiques de plus en plus sur la base de preuves empiriques, semblent avoir apporté un nouvel éclairage sur l'utilisation des évaluations d'impact, comprise comme un outil essentiel pour démontrer les effets d'une intervention publique et fournir des éléments pour en évaluer le succès (Battesti et al. 2012). Le Rapport Barca, Un agenda pour une politique de cohésion réformée (Barca, 2009), en partie incorporé dans les propositions du Règlement de l'UE, a fourni un cadre opportun de recommandations et de lignes directrices pour une utilisation plus intense, dans la nouvelle période de programmation européenne, des analyses d'impact et, en particulier, des méthodes contrefactuelles. Dans cette optique, le présent article entend proposer les résultats d'une évaluation d'impact conduite dans la région du Piémont, en expliquant, avec une intention paradigmatique, sa méthodologie et mise en œuvre.

2. Evaluer les politiques formatives de la Région Piémont

Pour évaluer les changements observés à la suite d'une intervention publique et attribuer à cette dernière le juste effet correspondant à travers une analyse contrefactuelle crédible, il est indispensable de disposer tant d'une connaissance approfondie de la chaîne causale qui caractérise le processus (cadre logique), que d'une analyse quantitative rigoureuse, qui réponde aux problèmes causés par la distorsion systématique (selection bias) ou par les effets de contagion (Conseil d'analyse économique 2013, IEG, 2006). En conséquence, une combinaison de méthodes quantitatives et qualitatives est l'approche la plus complète pour une bonne évaluation d'impact (White, 2008, 2009).

L'intérêt démontré par les économistes pour ces questions est important, car elles permettent de combiner l'analyse économétrique aux approches basées sur la théorie de l'évaluation d'impact. En limitant le terrain à l'évaluation de l'impact des politiques de formation, de nombreuses expériences se basent sur des conceptions quasi-expérimentales, qui combinent des indicateurs de performance et les estimations de l'impact sur les programmes dédiés au développement de la force de travail (Hollenbeck, Schroeder, et King, 2005; Hollenbeck & Huang, 2006). Une question fondamentale au cœur de ces travaux, concerne l'adoption d'indicateurs d'impact brut ou net (Hollenbeck et Huang, 2007). En réalité, les indicateurs d'impact brut fournissent des mesures compréhensibles du résultat, *l'outcome*, mais ne représentent pas les références absolues, du moment qu'ils ne permettent pas d'isoler l'impact réel du programme des autres facteurs. En revanche, les impacts nets sont utilisés pour vérifier les résultats du programme au net d'une situation contrefactuelle, mais leur validité dépend de la qualité du groupe de comparaison. Par conséquent, l'approche contrefactuelle est préférable lorsqu'on peut identifier un groupe de comparaison valable (White, 2010).

Comme les données expérimentales pures (attribution aléatoire au groupe des traitées ou au groupe contrefactuel) sont rarement disponibles, du moins en Europe, de nombreuses études utilisent des approches non expérimentales ou quasi-expérimentales, proposant principalement des estimateurs de correspondances (*matching*), modèles de durée et de sélection, méthodes d'estimation OLS (Card, Kluve, et Weber, 2010). La littérature sur les politiques de formation, toutefois, ne suggère pas des impacts univoques: la plupart des travaux existants suggèrent des effets légèrement positifs à court terme (Sianesi, 2008) et des effets positifs et significatifs à moyen et long terme. Kluve (2010), cependant, cite de nombreux exemples de programmes à impact nul. Enfin, les effets négatifs sont détectés chaque fois qu'un programme aborde des cibles de population socialement défavorisée ou des secteurs économiques en déclin (Friedlander, Greenberg et Robins, 1997; Heckman et Smith, 1999).

La littérature sur les programmes de formation italiens partage les mêmes inconvénients: les quelques contributions existantes suggèrent des impacts qui vont dans le sens contraire. Par exemple, Battistin et Rettore (2002), dans une étude des programmes de formation menés dans le cadre de l'automatisation du travail de bureau, ne trouvent pas d'effets positifs sur l'emploi de personnel formé dans les 17 mois suivant la fin des cours, tandis que Bellio et Gori (2003), dans une enquête similaire, mesurent des impacts positifs sur l'employabilité des jeunes chômeurs et Berliri, Bulgarelli et Pappalardo (2002) constatent des effets particulièrement positifs pour les hommes de niveau d'études faible et moyen.

Cet article analyse les principaux enjeux et résultats de l'analyse d'impact brut et net réalisée à l'intérieur de l'évaluation d'insertion dans le marché du travail, effectuée par le groupe de travail du CNR-Ceris en 2012, sur certains cours de formation professionnelle financés par la Région du Piémont, dans le cadre du Programme Opérationnel du Fonds Social Européen 2007-2013 (POR-FSE). Les cours analysés, bien que de durée différente, ont en commun de s'être achevés en 2010, de se caractériser par un nombre important d'heures de formation et par la reconnaissance d'un diplôme final (qualification ou spécialisation). Sont exclus, par conséquent, les cours de courte durée, les cours de formation continue ainsi que les cours dirigés vers des groupes particulièrement défavorisés. Afin de rendre possible l'analyse de l'impact net, l'évaluation a été réalisée sur un échantillon représentatif de personnes sans emploi au moment

de leur inscription au cours et effectuée sur la base des données issues d'une enquête par sondage, réalisée selon la méthodologie CATI. L'échantillon a été tiré de la base de données de la surveillance du FSE et d'autres sources administratives, en accord avec les grandes lignes directrices de ce sujet. (ISFOL, 2003). L'objectif était de procéder à une évaluation des résultats d'emploi de la formation professionnelle, analysant la situation professionnelle des étudiants en Octobre 2011, c'est-à-dire, environ un an après l'achèvement de la formation. En particulier, l'analyse a visé une série d'actions de formation spécifiques et des territoires précis (provinces), en tenant compte des caractéristiques personnelles qui peuvent influencer sur les niveaux d'emploi. L'aspect le plus remarquable du travail est représenté par l'évaluation de l'impact net, qui est souvent négligé dans les applications pratiques, en raison des nombreux problèmes théoriques et méthodologiques annexes, particulièrement en ce qui concerne l'identification *ex post* d'un groupe de comparaison homogène à l'échantillon des personnes formées.

En effet, on est rarement en mesure de mettre en œuvre *ex ante* une évaluation basée sur la conception expérimentale, soit pour des problèmes éthiques d'exclusion *a priori* du groupe des bénéficiaires d'individus qui sont dans le besoin du soutien des fonds publics, soit pour des exigences programmatiques. L'évaluation de l'impact, cependant, demeure une question nodale dans l'analyse de l'efficacité des politiques publiques et un effort méthodologique pour tenter une estimation fiable des impacts nets est hautement souhaitable. Juste l'identification d'un groupe de comparaison approprié a dirigé la conception de l'analyse, dans l'idée que l'évaluation de l'impact constitue un élément fondamental du processus d'élaboration des politiques, car elle permet une compréhension claire des principaux effets d'un programme et permet d'éviter l'effet de perte sèche dite *deadweight*, ou la perte de ressources causées par l'aide financière à des cibles spécifiques qui seraient toutefois atteintes, même en l'absence d'intervention publique (Sestito, 2002; Martini et Sisti, 2009).

Les sections suivantes illustrent brièvement les points principaux, y compris la description de l'objet de l'analyse et la quantification de la population cible (paragraphe 3), la description du plan d'échantillonnage (par. 4) et l'identification de l'échantillon de comparaison (par. 5), la description des indicateurs importants (par. 6) et les résultats de l'évaluation de l'impact net (par. 7). La section 8 propose, enfin, quelques indications de *policy* fondées sur les résultats de l'enquête, tandis que la section 9 décrit les enseignements méthodologiques apparus.

3. L'univers de référence: identification et quantification

L'évaluation des résultats en termes d'emploi a été menée sur les personnes formées en 2010 par certaines initiatives de formation financées par la Région du Piémont au moyen de ressources communautaires dans le champ d'application de POR-FSE 2007-2013. Les six actions analysées² se réfèrent aux directives "Activités expérimentales d'apprentissage - Obligation d'éducation" (OI), orientées vers les jeunes soumis à une formation obligatoire, et «Chômeurs - Marché du Travail» (MdT), et surtout vers la formation des travailleurs au chômage. La sélection de l'univers de référence a une importance particulière pour l'évaluation de l'impact net, surtout dans des contextes quasi-expérimentaux comme celui-ci: isoler *ex post* un groupe de *traités* (groupe principal) bien défini et pour lequel il est possible d'identifier un groupe de *non traités* (groupe comparatif) le plus homogène possible sur le principal, est un ingrédient fondamental pour une bonne évaluation de l'impact net. En outre, pour obtenir une statistique significative satisfaisante, les deux groupes doivent être constitués d'une quantité adéquate d'individus.

Pour cette raison, les initiatives de formation ont été choisies de manière à satisfaire certaines exigences particulières, qui soient essentielles à une bonne évaluation:

- un nombre de bénéficiaires suffisant pour permettre un traitement statistiquement significatif et qui ait un impact d'évaluation potentiellement intéressant;
- l'obtention d'un certificat de qualification ou de spécialisation à la fin du parcours de formation, vu que ces cours ont une charge formative importante (en termes d'heures / cours), ce qui permet une évaluation de l'impact net de la formation en comparant le placement des sujets traités à celui des sujets non traités;
- elles ne devaient pas impliquer de «poches» spécifiques de défavorisés (par exemple les ex-délinquants ou handicapés), caractérisées par un faible nombre et des particularités qui les rendent peu comparables, affectant ainsi la globalité de l'analyse;
- Elles devaient impliquer les sujets majoritairement au chômage, compte tenu de la difficulté d'évaluer l'impact net des politiques éducatives sur l'amélioration des conditions de travail (salaire net, type et durée du contrat, progression de carrière) des personnes ayant déjà un emploi avant leur inscription dans le cours.

² De façon spécifique, l'évaluation a été conduite sur des actions de qualification OI IV. H. 1199 (parcours triennal de qualification finalisée dans la réalisation de l'obligation d'instruction et de l'obligation de formation) et IV .H.11.01 (parcours annuel et biennal de qualification avec crédit d'entrée); sur les actions de qualification au Marché du Travail (MdT) III.G.06.04 (parcours de qualification pour étrangers immigrés chômeurs) et IV.I.12.01 (parcours de qualification pour les adultes à faible scolarité); sur les actions de spécialisation MdT IV.I.12.02 (cours de formation en vue d'une spécialisation, cours de formation brève et de mise à jour des compétences transversales et/ou professionnelles) et II.E.12.01 (parcours post-qualification, post diplôme et post bac qui mènent à la spécialisation). De ce fait, les axes convoqués dans l'analyse sont l'axe d'employabilité (II), l'axe d'inclusion sociale (III) et l'axe du capital humain (IV).

Par conséquent, la population cible est constituée de tous les qualifiés et spécialisés des actions sélectionnées, qui n'étaient pas employés au moment de l'inscription³. Limitant l'analyse aux seuls cours de qualification et de spécialisation, durant en moyenne plus longtemps que les autres types de cours (par exemple, licence commerciale, habilitation professionnelle, etc.), l'évaluation du placement s'est spécifiquement concentrée sur les politiques de compensation des handicaps, plutôt que sur les politiques d'accumulation du capital humain. Cela permet une évaluation plus appropriée de l'impact des politiques éducatives dans la Région du Piémont, principalement destinées aux sujets plus vulnérables, qui manifestent généralement de sérieuses difficultés d'insertion dans le monde du travail.

La population cible a été quantifiée à partir des données de surveillance sur la formation professionnelle⁴, ce qui nécessitait comme préliminaire une exploration précise de données pour éviter la présence de dossiers, *record*, en doublon⁵ et pour pallier au manque d'informations personnelles vitales pour l'enquête, comme les numéros de téléphone⁶. A la fin des opérations préliminaires, nous avons obtenu une liste d'échantillonnage de 9905 individus, à partir d'une population de référence de 11130 *dossiers*. En dépit de ce prétraitement minutieux, environ 12% des contacts ont été exclus de la liste d'échantillonnage, affectant l'efficacité du plan d'échantillonnage.

Étant donné que l'activité programmatique régionale qui concerne la formation professionnelle diffère en modalité et contenu selon la province et l'action, il a été préféré d'adopter un plan d'échantillonnage qui permettrait de distinguer ces caractéristiques, en stratifiant pour elles deux dimensions. Cependant, le nombre réduit de la population visée dans certaines provinces et actions a rendu le plan particulièrement compliqué et inefficace (voir par. 4).

Tableau 1 - Population de référence par action et par Province. Valeur %

	AL	AT	BI	CN	NO	TO	VCO	VC	Tot
% sur total FP 2010	11,7	4,3	3,6	11,6	6,5	56,8	2,6	2,9	100,0

	OI		MdT				Tot
	IV.H-11-99 OI3	IV.H-11-01 OIC	Qualification		Specialisation		
			III.G-06-04 STR	IV.I-12-01 BAS	II.E-12-01 SPD	IV.I-12-02 SPE	
% sur total FP 2010	29,0	6,8	19,8	24,1	13,6	6,7	100,0

Comme on peut le voir dans le Tableau 1, en fait, la formation professionnelle piémontaise est répartie de façon très hétérogène, au niveau territorial comme au niveau de l'action. En particulier, elle est concentrée quasiment aux deux tiers dans la seule province de Turin, tandis que le tiers restant est réparti principalement entre Alessandria et Cuneo. Tandis que, la part des personnes formées n'atteint pas les 5% à Biella, dans le Verbano-Cusio-Ossola et à Vercelli. Quant à la répartition par action, près d'un formé sur trois s'est qualifié à un cours OI triennale (OI3) et un sur quatre est issu d'un parcours de qualification de base pour adultes de faible niveau scolaire (BAS). Les parcours les moins empruntés sont la qualification OI avec des crédits d'entrées (OCE) et les cours de spécialisation sur les compétences transversales et / ou professionnelles (SPE).

³ Bien que les actions MdT soient prioritairement adressées aux individus au chômage, 16% des contacts totaux s'avéraient être employés au moment de l'inscription au cours et ont été par conséquent exclus de l'échantillon.

⁴ Le catalogue MonVISO, Surveiller et Evaluer les Interventions et Soutien à l'Emploi - Formation Professionnelle, est présente sur la banque de données "Système Décisionnel Instruction, Formation Professionnelle et Travail" (<http://www.ruparpiemonte.it>), gérée par CSI-Piemonte.

⁵ Pour obtenir un dossier unique par individu, les doublons ont été éliminés en privilégiant : a) les dossiers adaptés, dans le cas de double dossier adapté/non adapté, dans l'hypothèse où l'éligibilité obtenue signale un type de traitement auquel l'individu a participé avec succès; b) les dossiers avec admission aux examens finaux ou ceux avec abandon de la formation pour une question de travail, dans le cas de double non éligibilité; dans les autres cas, les dossiers avec un nombre majeur d'heures/cours, dans l'hypothèse qu'un cours de formation plus long signale un traitement plus approfondi et structuré, à égalité avec les autres conditions.

⁶ Les données de surveillance disponibles sur MonVISO ont été croisées avec les données administratives sur les Communications Obligatoires présentes dans le Système d'Informations du Travail dans le Piémont (SILP, <http://www.ruparpiemonte.it>) et avec les coordonnées disponibles sur l'annuaire des Pages Blanches. Globalement, 9% environ des contacts manquants ont été récupérés.

4. Le choix d'échantillonnage

Les choix d'échantillonnage pour une estimation juste ne sont pas uniques, mais dépendent des objectifs d'analyse. Dans le cas de la recherche présentée ici, cela a été particulièrement complexe pour la juxtaposition d'objectifs différents d'évaluation et de recherche :

- Estimer, de façon fiable les débouchés et résultats d'emploi des initiatives de formation financés par le FSE, avec un but évident d'intérêt public, surtout face à l'Union Européenne;
- Obtenir des données détaillées au niveau territorial, avec un but de type politique, aider la gestion et la reprogrammation⁷;
- Obtenir des données détaillées sur la base des caractéristiques individuelles de l'utilisateur, avec le but d'améliorer l'action pour les groupes cible (notamment les femmes que nous analysons ici, mais aussi les jeunes et les étrangers);
- Estimer l'impact net, dans le but d'apprécier l'efficacité des politiques, mais aussi dans le but de consolider les stratégies d'évaluation réalisables de façon systématique à l'avenir.

De toute évidence, il s'agit d'une liste considérable et difficile à concilier. En particulier, l'objectif 1, essentiel, diverge, en termes de méthodologie, du 4. En effet, si pour renforcer la qualité de l'estimation du valeur de l'échantillon en rapport avec la population il est nécessaire de soigner et élargir l'échantillon principal, pour analyser l'impact net avec une stratégie quasi-expérimentale, il est nécessaire de renforcer l'échantillon de contrôle. De la discussion avec la commission régionale est née la conviction que l'élément cognitif fondamental, en plus de la détection d'un paramètre général d'efficacité, était celui d'arriver à un niveau de détail qui permettrait une analyse particularisée des principales répartitions en lesquelles est organisée l'application des politiques de formation. D'un côté, donc, il a été demandé que soit possible une évaluation distincte pour les actions dont se composent les différentes directives, d'autre part, que soient également fournis les détails au niveau provincial. Les administrations provinciales, qui gèrent les différentes actions en termes opérationnels, ont, en fait, une certaine marge de discrétion dans la mise en œuvre des actions mêmes, afin de les adapter aux caractéristiques du système économique local et aux nécessités annexes. Le choix s'est donc porté sur un échantillon aléatoire stratifié selon les deux dimensions des actions (voir note 1), qui identifient, en quelque sorte, la typologie de politique, et des provinces.

Pour ce qui concerne les caractéristiques individuelles, il n'était pas faisable d'adopter des dimensions supplémentaires de stratification dans un plan qui comprenait déjà 48 strates. Par ailleurs, la théorie économique, ainsi que l'expérience, suggèrent que les différents groupes se caractérisent à la fois par des niveaux d'employabilité différents, et par des réactions différentes aux politiques actives. La possibilité de comparer les performances des politiques dans les différents groupes cibles a été assurée par la méthode de réaffectation proportionnelle, qui consiste à appliquer au sein de chaque couche d'échantillonnage, la même proportion (calculée sur la variable de référence) que dans la couche respective de la population. Par exemple, cela signifie que la proportion de femmes formées dans la directive réservée aux adultes à faible niveau d'éducation - qui a été interviewée, est égale au pourcentage de femmes qui se sont qualifiées dans cette action, et dans cette province. Cela garantit la possibilité de comparaisons entre les groupes et améliore également la précision de l'estimation, bien que de façon non quantifiable. Cette stratégie a été adoptée pour les dimensions de genre, de nationalité et d'âge, renonçant à une ou plusieurs dimensions dans les strates moins "peuplées", dans lesquelles il était impossible de respecter toutes les proportions en raison de manques d'interviews.

Sur le plan opérationnel, il a d'abord été calculé la taille de l'échantillon, dans la mesure de 1523 personnes, en appliquant la formule habituelle pour les populations finies⁸ (Cochran, 1977). L'échantillon représente 13,7% de la population visée et montre un niveau de précision satisfaisant, surtout par rapport à des travaux d'évaluation similaire (Lalla, 2003; Lalla & Fiorani, 2004; Centra & Falorsi, 2007; divers auteurs, 2010; IRPET 2011).

Par la suite, les observations de l'échantillon (individus) ont été réparties entre les 48 strates obtenues à partir du croisement des 8 provinces du Piémont avec les 6 actions objets de l'évaluation. La taille des strates était toujours calculée avec la formule de Cochran, afin de permettre au chercheur de se concentrer sur les particularités des différentes actions, en tenant compte de l'hétérogénéité territoriale. Pour améliorer la représentativité des sous-échantillons les plus critiques, les strates les plus petites ont été sur-échantillonnées, afin de réduire les erreurs d'échantillonnage associées⁹. Une fois définie la taille des strates, les individus sont extraits de manière aléatoire selon la conception de la répartition proportionnelle fondée sur les caractéristiques d'état-civil dont il a été question ci-dessus.

Cette stratégie permet un contrôle serré des erreurs d'échantillonnage, mais représente une arme à double-tranchant. Si, en effet, théoriquement, elle garantit de la représentativité des strates, dans la pratique, elle échoue en raison du nombre

⁷ Cet objectif a été partiellement abandonné, car l'excès de détails ne permettait pas d'arriver à des estimations fiables (intervalles de fiabilité trop grands) à cause de la petite taille des strates dans l'univers, et parce que les différences entre les « Province » et les actions n'étaient pas assez marquées.

⁸ La formule est, où e est l'erreur absolue dans l'estimation de la proportion P (inconnue) de la population N ; et l'abscisse du point pour lequel la distribution normale est égal à $(1-\alpha/2)$; α est le niveau de signification statistique désiré. Les valeurs sélectionnées sont: $e = 2,31$, $P = 0,5$, $\alpha = 0,1$.

⁹ Naturellement, l'erreur absolue de chaque strate est beaucoup plus élevée que celle de l'échantillon total.

insuffisant d'individus, imputables aux problématiques connexes aux listes d'échantillonnage et aux échecs dans les contacts. Dans le tableau 2, sur les résultats des contacts au cours de l'enquête, nous quantifions le fameux "*hard-core*", le «*noyau dur*» cité par Cochran (1977), c'est à dire l'ensemble de ceux qui refusent catégoriquement d'être interviewés. Bien qu'élevé (environ 20 %), il ne dépasse pas l'ensemble de ceux qui étaient impossibles à interviewer en raison d'un numéro de téléphone erroné (environ 15%) ou parce que certains sont apparus comme non admissibles à des fins d'interviews parce qu'occupant précédemment un emploi, contrairement à ce qui est indiqué dans la base de données de suivi (16%). De plus, il faut garder à l'esprit que les dossiers des sujets pour lesquels n'était pas même disponible une coordonnée de contact, avaient déjà été éliminés des listes d'échantillons. S'il est difficile, mais pas impossible, d'agir pour réduire le *noyau dur*, beaucoup peut être plutôt fait sur la qualité des listes de départ, dans le but d'influer sur la possibilité d'obtenir des données numériquement satisfaisantes dans des répartitions moins nombreuses.

Tableau 2 –Résultats des contacts durant l'enquête sur le terrain.

	Contacts	
Injoignables	667	14,5%
Numéro erroné	715	15,5%
Refus	931	20,2%
Non admissibles	739	16%
Interviewés	1559	33,8%
Total	4611	100%
Taux de report	49,4%	
Taux d'efficacité des contacts	33,8%	
Taux de non réponses	37,4%	

Enfin, il ne faut pas oublier que, pour préserver le plan d'échantillonnage, les non-répondants doivent être remplacés par d'autres individus de la même strate (Levy et Lemeshow, 2008). Cette procédure permet de maintenir intacte le niveau de représentativité, mais elle peut affecter les erreurs non dues à l'échantillonnage..

En conclusion, quelques leçons se dégagent de l'exercice d'échantillonnage effectué. Tant les résultats descriptifs, que les contrôles sur l'efficacité de l'échantillonnage (Cicchitelli, Herzl, et Montanari, p. 154), semblent indiquer qu'un plan plus simple d'échantillonnage est préférable, par exemple en ne classant les mesures qu'en trois sous-groupes (obligation d'éducation, cours de qualification de base, cours de spécialisation). Cela conduirait à des groupes plus homogènes, caractérisés par des variances mineures intra-groupe et des plus grandes différences dans les moyennes entre les groupes. Il est utile de bien s'interroger au sujet des variables qui influent réellement sur l'employabilité et sur le fait que, dans un plan d'échantillonnage efficace, les dimensions de stratification doivent conduire à des groupes très homogènes en eux et très différents entre eux. Dans le cas analysé, cela ne s'est pas vérifié de façon satisfaisante : les actions n'ont pas d'objectifs très clairs ni de cibles spécifiques, il y a des actions avec des caractéristiques très similaires entre elles (par exemple, deux actions qui conduisent à la spécialisation), tandis que de nombreuses actions contiennent des cours avec des caractéristiques (d'utilisateur, de méthode, de compétence) radicalement différentes. L'action des provinces, d'ailleurs, ne s'est pas particulièrement différenciée, sauf parfois par une présence plus importante de certaines actions. Si, dans certains cas, on rencontre des différences, elles sont dérivées de caractéristiques exogènes aux territoires, qui se rapportent difficilement au terrain des actions du FSE et par conséquent échappant à l'évaluation. Il reste à rappeler, cependant, que la stratification territoriale a produit des estimations peu fiables, en raison de la petite taille de l'échantillon dans les provinces mineures. Le plan d'échantillonnage de l'enquête suivante a montré que, même sans stratification par dimension territoriale, la fiabilité statistique des analyses par zone ne se dégrade pas. Par conséquent, il faut conclure que la quantité de la population cible par province ne serait de toute façon pas suffisante pour effectuer des analyses statistiquement significatives sur les différences territoriales.

En général, une conception si complexe, bien que très efficace pour l'estimation de l'échantillon global, finit cependant par ne pas atteindre les objectifs de détails qu'elle s'était fixée; il est en fait impossible d'arriver à des estimations ponctuelles précises dans toutes les strates, en dépit de la tentative de sur-échantillonner les strates critiques, qui demeurent malgré tout peu fiables. Ce qui cependant induit une complexité de mise en œuvre, qui se traduit par des coûts plus élevés et un allongement substantiel de la durée de réalisation.

5. Les échantillons de comparaison

Comme souligné précédemment, l'évaluation de l'impact net représente un des aspects les plus appréciables de ce travail. Pour la réaliser, une étape cruciale est l'identification d'un groupe de comparaison approprié, le plus homogène possible en regard de l'échantillon principal. En effet, dans l'analyse contrefactuelle, les "traités" et "non traités" devraient différer exclusivement par l'exposition au traitement, dans ce cas, par la fréquentation, avec des résultats positifs, d'un cours de formation professionnelle. En d'autres termes, l'analyse contrefactuelle devrait répondre à la question «Que ce serait-il passé si les politiques de formation n'étaient pas financées?».

La littérature montre qu'il n'est pas trivial de répondre à cette question (White, 2010). Plus précisément, il s'agit d'une tâche particulièrement difficile lorsque le groupe de comparaison n'a pas été conçu *ex ante*, comme dans les stratégies expérimentales d'évaluation (*randomized control trial*), mais a été identifié *a posteriori*, comme dans la conception présente (Ciravegna, Favro-Paris, Matto, et Ragazzi, 1995). En outre, l'évaluation de *placement* dans la Région du Piémont doit également estimer de manière fiable les résultats d'employabilité des politiques éducatives (cf. section 5). Par conséquent, le nombre d'entretiens consacrés à l'échantillon de contrôle est évidemment limité et la marge d'erreur est, par conséquent, assez large.

Une analyse minutieuse du contexte nous a suggéré d'extraire l'échantillon de contrôle de l'ainsi-dits *no-shows* (Bell, Orr, Blomquist, et Cain, 1995), c'est à dire de la liste des étudiants qui n'avaient pas encore un emploi au moment de l'acte d'inscription et qui n'ont pas achevé avec succès ce cours de formation (traitement). Le nombre de tels sujets n'est pas très élevé, pour cela l'échantillon de contrôle a nécessairement une dimension très réduite, par-delà des considérations précédentes sur les objectifs globaux de l'enquête. La grande homogénéité des *no-shows* relativement au groupe principal a toutefois permis d'opter pour ce choix.

Certaines stratégies ont été envisagées puis rejetées pour des problèmes d'infaisabilité, c'est à dire qu'elles étaient inadaptées: par exemple, l'absence de listes de classement pour l'admission aux cours de formation ne permet pas de comparer les résultats de l'emploi des -premiers exclus, avec ceux des -derniers admis, stratégie qui permettrait toutefois d'atténuer considérablement l'effet d'une possible distorsion systématique ou *selection bias*. Enfin, il a été décidé de ne pas retirer le groupe de contrôle des listes des chômeurs inscrits dans les centres pour l'emploi (CpI), parce que l'hétérogénéité des caractéristiques individuelles non observables (par exemple, la motivation, l'attitude proactive, les capacités individuelles, l'expérience, le milieu social de provenance) par rapport au groupe principal des personnes formées serait trop grande, en augmentant la distorsion par le *selection bias*.

A l'intérieur du groupe de ceux qui ont arrêté la formation, choisis comme échantillon de contrôle, a été isolé le sous-groupe de ceux qui ont arrêté la formation pour avoir trouvé un emploi pendant la période des cours de formation (C2). Intuitivement, ce groupe représente une contrepartie «forte», avec des caractéristiques non observables qui ont plus de valeur sur le marché du travail. Par conséquent, les écarts observés par rapport à C2 représentent une sorte de sous-estimation de l'impact de la formation professionnelle, parce qu'ils estiment les effets de la formation dans le cas d'un *selection bias* négatif. En revanche, l'échantillon comparatif C1 rassemble les sujets qui ont abandonnés les cours pour des raisons diverses, à l'exclusion de l'obtention d'un emploi¹⁰.

Pour ces raisons, la stratégie choisie est bien loin de représenter un choix optimal, mais dans le contexte actuel, il semblait être la plus souhaitable¹¹. Enfin, il faut remarquer que le groupe de comparaison n'est pas présent pour les actions OI, puisque les individus de 14/16 ans sont obligatoirement inscrits à l'école, sauf les *décrocheurs*, c'est à dire ceux qui abandonnent les études qui, logiquement, ne constituent pas une référence valide et sont beaucoup plus difficiles à atteindre. Par conséquent, pour ces actions il n'est pas possible de procéder à une évaluation d'impact net.

¹⁰ Il est nécessaire de se rappeler le très haut pourcentage d'individus qui n'ont pas communiqué les motifs de leur abandon de la formation (51,2%) ou qui ne se sont jamais présentés aux cours (28,4%)/ Probablement, ces groupes cachent un bon nombre de sujets qui ont trouvé du travail, mais il n'y a pas de raison d'en déduire une *selection bias* négative, comme en C2.

¹¹ Dans tous les cas, dans la même communauté d'évaluateurs il n'existe pas encore d'accord substantiel quand aux stratégies optimum d'évaluation de l'impact (White, 2010). Une évaluation rigoureuse demanderait un contexte purement expérimental, qui définirait *ex ante* soit le groupe des traités soit le groupe de contrôle, à partir d'individus ayant des caractéristiques homogènes. Beaucoup suggèrent des stratégies optimales de pure *randomisation*, mais elles sont difficiles à mettre en œuvre.

Tableau 3 –Univers et échantillon de comparaison*

		II.E-12-01 SPD	IV.I-12-01 BAS	IV.I-12-02 SPE	Total
C1 (raisons diverses excluant l'obtention d'un emploi)	Population	670	431	212	1313
	Echantillon	189	103	28	320
	% population	20,0	27,1	32,1	24,3
	% cible	41,0	-12,0	-58,8	0,3
C2 (raison de travail)	Population	148	x	x	148
	Echantillon	70	x	x	70
	% population	47,3	x	x	47,3

* Les actions sans groupe de comparaison ne se rapportaient pas à un nombre adéquat d'individus.

Le Tableau 3 montre les dimensions des échantillons contrefactuels. L'écart des pourcentages entre l'échantillon cible et effectif (% cible) met en évidence les difficultés rencontrées en phase d'échantillonnage, à cause de la faible quantité des *dossiers* de l'univers de comparaison. En particulier, le plan d'allocation proportionnelle ne s'est maintenu dans aucun des deux échantillons, et en C2, la stratification territoriale¹² a aussi été abandonnée. Dans tous les cas, les deux échantillons contrefactuels représentent un intéressant *benchmark* (référence) pour l'évaluation de l'impact net.

6. L'impact lourd des politiques de formation sur les placement

Pour évaluer l'impact brut des formations professionnelles sur le placement des personnes formées, avec une attention particulière aux sujets fragiles (ex : femme, étrangers , jeunes), nous nous sommes concentrés sur l'analyse des conditions professionnelles principales après le cours de formation dans le moyen terme, c'est à dire environ 12 mois après sa conclusion (octobre 2011). En observant les variations des conditions d'emploi des individus au chômage au commencement de leur parcours, on évalue en fait l'impact d'employabilité lourd des formations, sans entrer dans la question de ce qui se serait vraisemblablement passé en leur absence (impact net sur l'employabilité).

Pour rendre l'analyse plus immédiate, on a aussi calculé les trois **indicateurs de résultats** liés entre eux, illustrés dans la Figure 1 (Berliri, De Vincenzi&Villante, 2003), qui évaluent respectivement le quota de personnes formées occupées, le quota - plus important - des personnes formées insérées dans le marché du travail (employés, employés en CIG, stagiaires) et enfin le taux de succès de la formation, ou plutôt le quota de ceux qui se sont insérés, unis au quota de ceux qui continuent leur propre parcours de formation ou d'instruction.

Figure 1 - Indicateurs de placement, définition

$$\text{Taux d'occupation} = \frac{\text{Formés employés}}{\text{Total Formés}}$$

$$\text{Taux d'insertion} = \frac{\text{Formés avec activité professionnelle (employé + stage + empl. en CIG)}}{\text{Total des Formés}}$$

$$\text{Taux de succès} = \frac{\text{Formés avec activité professionnelle + Etudiants}}{\text{Total formés}}$$

Tableau 4 - Indicateurs de placement* et conditions professionnelles précédentes des formés en Octobre 2011

Valeurs %

<i>Conditions professionnelles Octobre 2011</i>	%	t. occ.	t. inser.	t. succ.
---	---	---------	-----------	----------

¹² C2 est présent seulement pour l'action SPD, avec stratifications territoriales Turin / autres provinces

Employé	41,8	41,8 (±2,2)		
Employé en CIG	1,6			
Stagiaire	1,8	45,2	(±2,3)	
Etudiants	7,3	52,5	(±2,3)	
En recherche d'emplois	46,0			
Inactifs	1,5			
Total %	100,0			
V.A.	1558			

* Entre parenthèses les marges d'erreurs sur 95% de confiance

A une année environ de l'accomplissement du parcours de formation, le taux d'occupation global des formés se trouve avec une probabilité de 95% entre 39,5% et 44% (Tableau 4), pendant que le taux de succès varie entre 50,2% et 54,8%. Ce qui signifie que, parmi ceux qui ont commencé les cours en étant au chômage, au moins un formé sur deux rencontre une issue positive, soit par insertion professionnelle ou par reprise de scolarité, pendant que deux formés sur cinq ont trouvé un véritable emploi. Les *performances*, cependant, sont différentes selon l'action (tableau 5).

Les meilleurs résultats d'insertion professionnelle s'observent dans les actions de qualification de base pour adultes ayant une scolarité faible, surtout dans le secteur de la santé et d'assistance sociale. Ceci signale l'efficacité des connaissances de base transmises dans les cours de durée moyenne, également en situation persistante de crise économique, tel qu'actuellement. En revanche, les cours OI montrent de faibles taux d'emplois, autour de 30%, mais les taux importants de réussite, comparables à ceux des action MdT, signalent un important retour aux études, particulièrement appréciable lorsqu'il s'agit de sujets défavorisés ayant des difficultés d'apprentissage. Enfin, le taux élevé d'insertion qui caractérise les cours OI avec des crédits d'entrée par rapport aux cours de trois ans traditionnels signale l'importance de la formation sur le terrain, *on-the-job* (ex. stages), tandis que la précarité d'emploi de la CIG est plus importante dans l'action pour les adultes étrangers et dans les cours SPE.

Tableau 5 - Indicateurs de placement et écart correspondant avec confiance de 95%, par action.

		T. d'occupation		T. d'insertion		T. de succès	
Action / Indicateur		%	Ecart	%	Action Indicateur /	%	Ecart
OI	IV.H-11-99 – OI3	27,0	±4,80	29,8	±4,96	49,6	±5,41
	IV.H-11-01 – OIC	31,9	±5,17	37,4	±5,37	45,3	±5,53
MdL	IV.I-12-01 – BAS	53,8	±4,63	55,7	±4,61	63,4	±4,59
	III.G-06-04 – STR	42,6	±7,45	47,1	±7,53	47,1	±7,53
	II.E-12-01 – SPD	46,9	±4,45	50,0	±4,45	51,2	±4,41
	IV.I-12-02 – SPE	40,4	±9,23	45,7	±9,37	50,0	±9,41
Totale		41,8	±2,25	2,25	±2,27	2,27	±2,27

Au regard des différences de genre, les femme formées montrent des résultats plutôt meilleurs par rapport aux hommes, soit en terme de taux d'emploi supérieur, soit en taux de chômage¹³ inférieur (Tableau 6). Le résultat, cependant, est conditionné aux effets de composition: dans l'échantillon, en fait, se trouve un grand nombre d'hommes jeunes qui souffrent du taux élevé de chômage chez les jeunes. Pour épurer les résultats de tels effets, il faut procéder avec l'analyse multivariée (paragraphe 7.b).

Dans l'ensemble, cependant, l'étude de l'impact sur l'emploi brut de la formation professionnelle montre des résultats positifs, compte tenu surtout de l'atténuation du désavantage des groupes généralement fragiles, tels que les femmes et les étrangers. Leurs taux d'emploi après le cours, en fait, sont au moins aussi bons que ceux de leurs homologues, sinon

¹³ Il est notable que le haut taux de chômage ici rapporté diffère de la définition Eurostat/ISTAT. Cela, en fait, ne tient pas seulement compte de la déclaration d'un individu d'être à la recherche d'un travail, mais requiert plutôt que l'individu soit en recherche active et immédiatement disponible à le commencer.

meilleurs, et montrent également un degré plus élevé de cohérence entre les contenus de la formation et l'emploi qui en découle (résultats non montrés).

Tableau 6 - Conditions professionnelles principales à Octobre 2011 par genre, nationalité, âge.

<i>Conditions professionnelles octobre 2011</i>	Tot	<i>Genre</i>		<i>Nationalité</i>		<i>Age</i>	
		F	M	ITA	STR	A (26+)	G
Employés	41,8	43,7	39,7	41,9	41,3	51,1	35,0
Employés en CIG	1,6	1,5	1,7	1,2	2,9	2,1	1,2
Stagiaire	1,8	2,0	1,6	2,2	0,3	0,8	2,6
Etudiants	7,3	6,4	8,3	8,3	3,8	1,4	11,7
En recherche d'emploi	46,0	44,2	48,0	45,0	49,9	42,9	48,3
Inactifs	1,5	2,2	0,7	1,4	1,8	1,8	1,2
Total	100,0	100,0	100,0	100,0	100,0	100,0	100,0
V.A.	1558	814	744	1219	339	660	898

Toutefois, le handicap des sujets fragiles, persiste clairement si quelques indicateurs de qualité de travail sont analysés, tels les salaires nets, le type et la durée du contrat (résultats non montrés). Les femmes et les étrangers, en particulier, ont un accès correct à des postes de travail pour une durée indéterminée, surtout dans le champ des soins aux personnes âgées, mais il s'agit d'emplois peu stables, vu le taux élevé de résiliation de contrat. En examinant les contrats à durée déterminée, au contraire, les mêmes sujets montrent en moyenne des conditions de travail moins stables, à l'exception des jeunes de moins de 26 ans, qui grâce à l'institution du contrat d'apprentissage accèdent fréquemment à des contrats à durée déterminée de longue durée, mais montrent globalement un taux d'emploi très réduit par rapport aux adultes (Tableau 6).

Au final, le désavantage des sujets fragiles persiste quand on considère le salaire net (résultats non montrés). Les femmes, les étrangers et les jeunes, en fait, se situent principalement dans une gamme de salaire en dessous de 1000 euros et aucun étranger n'arrive à gagner plus de 1500 euros.

7. L'évaluation des impacts nets

a) Les écarts d'emplois nets

Un premier pas pour évaluer l'impact net de la formation professionnelle sur l'emploi, ou plutôt, l'impact sur le net de ce qui aurait pu advenir sans les formations, consiste à confronter la distribution des conditions professionnelles depuis octobre 2011 dans l'échantillon des formés, avec celle des échantillons de comparaison C1 et C2. Précisément, l'échantillon C1 est disponible pour SPD, BAS et SPE et regroupe les individus qui n'ont pas terminé leur formation pour des raisons diverses, tandis que l'échantillon C2 est disponible seulement pour SPD et regroupe le sous-ensemble d'individus qui ont abandonné le cours parce qu'ils ont trouvé un travail.

La seule action avec un évident effet net positif est BAS, c'est à dire, l'action qui finance les cours de qualification de base pour adultes de faible niveau scolaire (Tableau 7). Dans BAS, en fait, le taux d'emploi des formés est le double par rapport au groupe de ceux qui ont abandonné le cours¹⁴. En revanche, dans SPE, on constate la situation opposée, même si avec des différences inférieures et des estimations bien moins fiables¹⁵: deux formés sur cinq ont été employés en octobre 2011, pendant que le ratio observé pour l'échantillon de ceux qui ont abandonné la formation est de un sur deux.

Au final, l'action de spécialisation pour des individus déjà en possession d'un diplôme (principalement jeunes) ne présente pas de différences significatives entre formés et ceux qui ont abandonné la formation, à moins de ne pas prendre en compte les individus ayant abandonné parce qu'ils ont trouvé du travail. Dans un tel cas, le quota des employés parmi les non formés est supérieur de plus de 10 pour cent. L'impact de cet écart important, toutefois, se redimensionne considérablement, jusqu'à quasiment s'annuler, si nous considérons les résultats dans une perspective

¹⁴ L'écart est hautement significatif, si on considère que l'erreur standard du taux d'occupation dans BAS est estimé avec une fiabilité de 95% à 2,36 dans l'échantillon principal et 3,84 pour le C1.

¹⁵ L'erreur standard aux 95% estimé dans SPE pour le taux de chômage est de 4,71 dans l'échantillon principal et 8,91 pour le C1.

dynamique : environ un tiers des non formés, en fait, a perdu sa position avantageuse sur le marché du travail, se retrouvant de nouveau au chômage après environ une année, pendant que quasiment la moitié des formés a trouvé un emploi. Ces estimations, toutefois, sont liées aux éventuels effets de composition: si un échantillon est constitué d'individus intrinsèquement plus forts sur le marché du travail, c'est à dire avec des caractéristiques observables et/ou non observables qui les rendent plus intéressants au-delà de la formation, l'effet correspondant sur l'employabilité ne pourra pas être distinct de l'effet de la formation. Pour identifier avec clarté les effets imputables au traitement (formation), il est nécessaire d'utiliser des techniques d'analyse multivariée.

Tableau 7 - Conditions professionnelles principales à octobre 2011 par action, dans l'échantillon principal (P) et écarts par rapport aux échantillons de comparaison C1 et C2. Valeurs%

	Actions						
	II.E.12.01 SPD			IV.I.12.01 BAS		IV.I.12.02 SPE	
	P	ΔC1	ΔC2	P	ΔC1	P	ΔC1
<i>Activité professionnelle octobre 2011</i>							
Employé	46,9	-0,3	11,4	53,8	-26,6	40,4	9,6
Employé en CIG	1,0	-1	1,1	1,3	-1,3	3,2	-3,2
Stagiaire	1,9	-1,9	2,3	0,8	0,2	2,1	-2,1
Etudiants	7,0	2	-0,7	1,3	0,6	4,3	-0,7
En recherche d'emploi	41,8	-1,6	-12,6	40,9	20,2	47,8	-4,9
Inactif	1,2	3,0	-1,2	1,9	6,8	2,2	1,4
Total	100,0	0,0	0,0	100,0	0,0	100,0	0,0

b. L'analyse multivariée

Le deuxième pas dans l'évaluation de l'impact net des politiques de formation consiste en **l'analyse probit à variables multiples** des probabilités de trouver un emploi, qui mesure en termes de pourcentage l'impact net de la formation professionnelle sur la probabilité d'être employé une année plus tard, tout en faisant en même temps l'examen des effets que les caractéristiques individuelles ont sur de telles probabilités.

Le modèle de régression présenté dans le Tableau 8 mesure un impact positif et significatif du nombre d'années d'instruction sur la probabilité d'emploi, tandis que l'impact de la nationalité, si extra-communautaire, est négatif. L'âge, en revanche, a un impact positif mais décroissant: *ceteris paribus*, la probabilité d'être employé croît avec l'âge, mais à vitesse décroissante. Au final, le genre et les variables de contexte (diplôme et métier des parents) n'ont pas d'effets significatifs.

Les politiques de formation, en revanche, montrent un effet positif et significatif : un formé avec des caractéristiques individuelles moyennes (31 ans, citoyen UE, 12 ans de scolarité) a une probabilité 10% supérieure de trouver un emploi par rapport à un non-formé avec les mêmes caractéristiques. En outre, quelques estimations sur les classes d'âge signalent un effet positif marginal qui croît jusqu'à 30-35 ans, mais qui décline ensuite (5,7% à 20 ans, 10% à 35, 3,7% à 50). Il faut toutefois garder à l'esprit que la majeure partie des individus présents dans l'univers sont jeunes ou de jeunes adultes: moins de 20% des individus à plus de 35 ans et les fréquences se réduisent considérablement pour les âges plus avancés, rendant moins sûres les estimations relatives aux âges plus élevés.

Tableau 8 - Modèle probit sur l'échantillon principal et sur l'échantillon de comparaison C1 (SPD, BAS, SPE)

Probit regression				
		Number of obs	=	1177
		LR chi2(9)	=	37.74
		Prob > chi2	=	0.0000

occupati	Coef.	P> z	[95% Conf. Interval]	
Age	.0868817	0.001	.0356085	.138155
Age2	-.0012975	0.001	-.0020391	-.0005559
noeu	-.3447926	0.020	-.6353264	-.0542588
deg_instr	.0280054	0.016	.0051179	.0508929
père_univ	.0297011	0.840	-.258698	.3181001
mère_univ	-.206108	0.213	-.5302953	.1180794
père_auto	.0883131	0.495	-.1655696	.3421958
mère_auto	-.3371343	0.205	-.8586388	.1843703
formation	.2449484	0.005	.0756024	.4142945
_cons	-1.883176	0.000	-2.711122	-1.05523

Expression : Pr(occupati), predict()				
at	age	=	31.29312	(mean)
	age2	=	1077.978	(mean)
	noeu	=	.0705183	(mean)
	degré d'instruction	=	12.13934	(mean)
	père_univ	=	.0934579	(mean)
	mère_univ	=	.0739167	(mean)
	père_auto	=	.104503	(mean)
	mère_auto	=	.0229397	(mean)
	0.formation	=	.259983	(mean)
	1.formation	=	.740017	(mean)

Delta-method				
	Margin	P> z	[95% Conf. Interval]	
Formation				
0	.3980683	0.000	.3419558	.4541808
1	.4946536	0.000	.4610224	.5282848

Au final, il est possible de fournir une estimation plus fiable de l'effet net de la formation en se servant des ainsi-nommés 'effets marginaux moyens' (AME, *Average Marginal Effect*). En pratique, on calcule la moyenne de l'effet marginal de la formation sur tous les individus de l'échantillon, y compris les non-traités, en estimant la différence de leur employabilité s'ils sont formés ou non formés, à égalité avec toutes les autres variables observées, au contraire de l'effet marginal plus traditionnel calculé sur l'individu moyen, ou sur un individu unique qui présente des caractéristiques moyennes pour toutes les autres variables¹⁶.

En accord avec les résultats précédents, l'AME estime un impact net significatif des formations sur l'employabilité des individus de l'échantillon de +9,4 pour-cent (résultat non montré), ou les formés ont une probabilité de travail dans la période moyenne (environ à une année de la formation) supérieure environ de 10%.

En ajoutant au modèle une variable relative au genre et une interaction entre genre et formation, qui capture la différence d'effet net de la formation sur les hommes et les femmes, apparaît un résultat important sur la récupération du désavantage pour des cibles spécifiques (non démontré): la formation compense quasi totalement le handicap d'employabilité attribué au genre. En particulier, l'effet négatif d'être une femme (-0,36) est quasi totalement compensé pour les femmes formées, qui jouissent d'une augmentation ultérieure d'employabilité (+0,34), outre que celle attribuable à la formation (+0,39). Le Tableau 9 montre la différence de genre dans les effets marginaux moyens de la

¹⁶ Concrètement, il s'agit de calculer pour chaque individu de l'échantillon quelle probabilité d'être employé est estimé par le modèle, en partant des conditions individuelles, en supposant (indépendamment de ce qui se passe dans la réalité) qu'il a préalablement participé à la formation, et puis au contraire ensuite, qu'il n'a pas achevé son parcours. La différence entre les deux probabilités est l'effet marginal individuelle, la moyenne de tels effets marginaux est l'AME

formation. Le coefficient relatif aux hommes n'est pas significatif, tandis que dans le cas des femmes, la formation augmente leur employabilité de presque 15 pour-cent.

Tableau 9 - Effets marginaux moyens de la formation sur le genre.

Average marginal effects		Number of obs = 1217				
Model VCE : OIM						
		Delta-method				
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
1.formatzion						
_at						
Femme	.1479593	.0412309	3.59	0.000	.0671483	.2287703
Homme	.0204777	.0495784	0.41	0.680	-.0766941	.1176495

Note: dy/dx for factor levels is the discrete change from the base level.

De ce fait, résulte de façon évidente l'efficacité des actions formatives dans la compensation du désavantage des sujets particulièrement fragiles sur le marché du travail, cela à l'exemple des femmes et des extra-communautaires (résultat non démontré)

c) *Propensity score matching et selection bias*

Les mêmes résultats sont confirmés par l'analyse du *propensity score* (Rosenbaum & Rubin, 1983), largement appliquée dans le champ d'évaluation pour l'estimation des effets moyens du traitement (ATE). Une telle technique est particulièrement appréciée dans des contextes non expérimentaux, dans lesquels il n'est pas possible de sélectionner aléatoirement les sujets à traiter et ceux qui sont exclus du traitement. De tels contextes, en fait, sont sujets *selection bias* qui est l'erreur d'estimation qui arrive si on attribue des différences d'employabilité des sujets traités et non traités au traitement même dans le cas où les deux groupes diffèrent par contre aussi systématiquement par quelques caractéristiques non observables.

Le *propensity score matching* mesure une augmentation de l'employabilité des personnes formées respectivement aux non formés de 5,5% (Tableau 10), oscillant entre 5 et 8% en fonction des spécificités (résultat non démontré). Ce résultat est en accord avec quelques études récentes sur des données italiennes concernant l'impact des politiques de formation sur l'employabilité des formés (Berliri & Pappalardo, 2001; Berliri, Bulgarelli, & Pappalardo, 2002; IRer, 2006; IRPET, 2011).

Tableau 10 - Estimation du propensity score.

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
occupées	Unmatched	.505076142	.408496732	.09657941	.03354997	2.88
	ATT	.464052288	.408496732	.055555556	.040095579	1.39

Au final, quelques estimations ultérieures ont été effectuées afin d'éliminer l'erreur de sélection (Heckman, 1976). Prenant en compte un modèle à deux équations, qui décrit respectivement, la probabilité de participer et d'obtenir un certificat final pour des individus inscrits à un cours de formation, et puis la probabilité successive d'obtenir un emploi si ils ont achevés les cours avec succès, les résultats (non démontrés) ne révèlent pas une présence significative de *selection bias*: quelques estimations révèlent une *selection bias* faiblement significative et négative, mais qui ne se révèle pas résister à des spécifications légèrement différentes. Pour autant, il est possible de conclure qu'il n'existe pas de caractéristiques non observables des individus qui influeraient de façon significative, tout d'abord sur leur décision de participer à un parcours de formation et ensuite sur leur probabilité d'obtenir un travail, indépendamment de l'effet imputable à la formation reçue. Il est plausible que ce résultat soit reproductible au processus précis de sélection des échantillons de contrôle, qui s'avère être très homogène à l'échantillon principal. Le résultat des estimations de la *selection bias* suggère, en fait, que la méthodologie probit standard soit efficace pour évaluer l'impact net de la formation professionnelle dans l'échantillon sélectionné, sans besoin de corrections ultérieures dues aux distorsions de la sélection. Si l'échantillon de contrôle avait été sélectionné parmi les chômeurs inscrits au Centres pour l'Emploi, les écarts dans les caractéristiques inobservables par rapport aux personnes formées auraient probablement été tels qu'ils

aurait rendu nécessaire une correction des estimations, en appliquant la méthode suggérée par Heckman. Au-delà de ces considérations, la multiplicité des techniques adoptées suggèrent une efficacité particulière des politiques de formation dans la compensation du handicap des individus particulièrement fragiles sur le marché du travail, comme d'autres résultats documentés nous le signalent (Antonelli & Nosvelli, 2002).

8. Synthèse des principaux résultats

Ce travail, à travers une pluralité de dispositifs statistiques et économétriques, confirme quelques résultats déjà connus relativement à l'impact des formations dans la Région du Piémont. La donnée qui émerge, confirmée par diverses méthodologies, est celle d'un impact net positif mais modeste, principalement sur le renforcement de l'employabilité des sujets désavantagés (Lauro & Ragazzi, 2011).

En ce qui concerne le placement, la meilleure performance s'observe dans les cours de qualification de base, souvent générée par quelques types professionnels spécifiques (ex : un opérateur socio-sanitaire). Néanmoins, les bons résultats des ces cours persistent également quand ces typologies professionnelles sont exclues de l'analyse. Les cours de spécialisation produisent des performances inférieures sur l'employabilité au regard de celles des cours de qualification, probablement parce que l'acquisition de compétences supplémentaires a un effet mineur par rapport à l'acquisition d'un bagage solide et cohérent, orienté par une qualification.

Il reste à considérer que les cours de qualification de base sont les plus fréquentés par les sujets désavantagés: les bonnes performances sur l'employabilité observées dans de tels cours, surtout confrontées au fléchissement général de l'économie, confirment la capacité des formations professionnelles piémontaises à intervenir sur la faiblesse d'employabilité des cibles fragiles, respectant en cela la nature de sa *mission* historique. Ceci a lieu non tant parce que les cours de base ont une meilleure offre formative comparée aux cours de spécialisation, mais parce qu'ils interviennent proprement sur des sujets pour lesquels la formation professionnelle est majoritairement en mesure de donner des résultats.

La différence des sujets fragiles (femmes, étrangers), face à la possibilité de trouver un emploi, est quoiqu'il en soit compensé seulement partiellement: même quand les formés trouvent un travail, les caractéristiques de cet emploi sont souvent "faibles" en termes qualitatifs (type de contrat, durée, revenu net). Ce point n'émerge pourtant pas de l'analyse multivariée, parce qu'elle a été conduite à partir d'un indicateur dichotomique de l'employabilité, qui différencie simplement les travailleurs des chômeurs, laissant à part toutes les autres considérations relatives à la qualité de l'emploi. En outre, le handicap des jeunes persiste, en particulier dans les actions de la Directive OI. Parmi les personnes formées de cette Directive, le nombre élevé de retour à la scolarité témoigne de l'efficacité des interventions des formations dans la construction d'un solide capital humain.

L'effet positif mais modéré de la formation suggère fortement l'exigence d'intégrer celle-ci, au niveau global, à d'autres politiques actives du travail, qui prêtent une attention particulière et agissent sur les mécanismes d'entrée dans le MdT (Marché du Travail). En particulier, l'utilisation d'instruments (suivi direct et personnalisé) qui aident les personnes à choisir les actions de formation les plus adaptées, demeure souhaitable, tout comme le recours aux services pour l'emploi additionnels, qui supportent les personnes dans la gestion de leur capital humain et les font entrer dans les réseaux relationnels qui conduisent vers l'emploi. Au final, pour ce qui concerne la cohérence des emplois, en moyenne, les individus qualifiés se trouvent avoir reçu des offres de travail en rapport avec le cours fréquenté.

D'autres détails sont inclus dans le rapport intégral sur l'emploi, dont le téléchargement est disponible sur le site www.ceris.cnr.it.

9. Leçons de méthode et de procédure pour l'évaluation de l'efficacité des politiques de formation

De concert avec l'administration adjudicatrice, le travail sur le Premier Rapport de Placement sur les politiques de formation de la Région Piémont a été conçu principalement dans un rôle d'expérimentation et de consolidation méthodologique. En fait, si l'évaluation d'impact est une pratique déjà peu diffusée dans notre pays, l'application des méthodes pour l'évaluation de l'impact net à des programmes entiers (ou à des parties de ceux-ci) est une expérience quasi inédite. Pour cela, à l'intérieur du Rapport, une grande place a été consacrée à une description extrêmement détaillée de tous les choix méthodologiques effectués et, dans quelques cas, il a été choisi d'adopter une pluralité de méthodes (comme dans l'estimation de la distorsion de la sélection), qui permettent soit de valider les résultats, soit de vérifier l'applicabilité et la portée informationnelle de chacun d'eux.

Le travail de prospection méthodologique est actuellement en train de se développer avec la seconde édition du Rapport de Placement, à paraître bientôt, mais déjà, de cette première expérience sont apparus des indications méthodologiques et procédurales significatives, qui peuvent se traduire en leçon générale pour l'exercice de l'évaluation de l'efficacité. Par la suite nous avons organisé les enseignements principaux par macro-thèmes.

L'utilisation des données administratives et de suivi

Dans le domaine de l'évaluation de l'efficacité, les données administratives et de suivi sont une source incontournable, riche et précieuse, également par la nécessité de micro-données nominatives. Elles contiennent des données extrêmement riches sur les individus, qui permettent de réduire le nombre d'informations à acquérir à travers la recherche sur le terrain (questionnaire). Mais il est clairement apparu qu'il n'est pas possible de se fier à l'aveuglette aux données extraites des banques de données, parce que seule une connaissance approfondie des procédures d'élaboration des données, de leur enregistrement et de l'organisation du système informatique permettent de construire des listes fiables et complètes et d'utiliser correctement les données individuelles qu'elles contiennent.

Les difficultés rencontrées dans la phase de planification de l'enquête ne concernent pas seulement le manque de fiabilité des informations ou des données manquantes, qui sur un mode ou un autre peuvent être gérées, mais concernent plutôt et avant tout la pleine compréhension de la réelle portée informative des données enregistrées. De telles difficultés semblent, entre autre, indiquer que le chemin de l'utilisation des données de suivis et d'administration aux fins d'évaluation de l'efficacité soit complexe et, au stade actuelle des systèmes, non applicables dans des délais brefs. Les systèmes d'informations, créés essentiellement pour des finalités de gestion, font qu'aujourd'hui les données sont pratiquement impossibles à utiliser sur un mode systématique, rapide et fiable, puisqu'il reste difficile d'intégrer les différentes sources, de comprendre la nature exacte d'une donnée et de la traduire en indicateurs clairement interprétables. De ce fait, bien que présentes, les données demeurent souvent inutilisables à moins d'un considérable investissement préventif pour les rendre accessibles, compatibles, "certifiées".

En outre, contrairement à ce que l'on pourrait attendre, les systèmes informatifs en question ne se résument pas à un ensemble d'informations codifiées, mais leur utilisation correcte demande une série de connaissances tacites, parfois secrètes, mais surtout dispersées (dans plusieurs organisations, privées ou publiques, par le phénomène de décentralisation de l'administration et celui de l'externalisation). Pour comprendre la nature d'une donnée et ses règles de construction, il est souvent nécessaire de trouver les personnes qui, à l'intérieur des organisations, détiennent de tels savoirs. Ceci se vérifie parce que les systèmes de suivi que les autorités voudraient utiliser aujourd'hui pour l'analyse et l'évaluation, ont été tirés des *bases de données* administratives, souvent avec des procédures de type additif plus que vraiment et proprement par intégration. De ce fait, à différentes étapes, ils ont subi des modifications incrémentales et inopinées et non de vraies et réelles réformes, raison pour laquelle il est vraiment complexe aujourd'hui de reconstruire une systématique¹⁷.

En synthèse, dans le processus d'évaluation, nous rencontrons des informations et connaissances dispersées ou inutilisables pour des limites non seulement méthodologiques, mais également procédurales. L'administration n'est pas toujours consciente de ce problème. Une part du devoir des évaluateurs devient alors celui de faire surgir de telles limites.

L'échantillonnage

L'analyse du *placement*, comme d'autres analyses qui alimentent l'activité d'évaluation, se base généralement sur des sondages. Vouant adopter un échantillon aléatoire représentatif, il est nécessaire de mieux en définir auparavant les dimensions de stratification. Ceci se décline en deux questions différentes.

1. *Existe-t-il une variabilité effective dans les dimensions présélectionnées?* Ou, les groupes qui découlent de la stratification sont-ils assez différents? Si non, la stratification devient un fardeau inutile qui comporte une perte nette d'efficacité par rapport à un simple échantillonnage aléatoire.
2. *Quelle est la dimension des strates dans la polulation et dans l'échantillon?* Ou, combien d'observations retombent dans chaque strate? En fait, il faut éviter de créer des classifications trop détaillées, qui mènent à une taille de l'échantillon insuffisante à garantir des estimations précises des données de détail. Des fourchettes de 15, 20, 30 points de pourcentage rendent inutiles les estimations destinées à la programmation et appesantissent beaucoup les phases de recherches sur le terrain, avec une augmentation des coûts, un allongement des temps, et un accroissement des probabilités d'erreurs et décalages.

Le choix d'un plan d'échantillonnage plus simple, dans l'exercice actuellement en cours pour le second Rapport de Placement, a montré l'efficacité d'une telle simplification. La capacité prédictive de l'échantillon dans chaque cellule du

¹⁷ Pour clarifier le concept, nous prenons comme exemple le champ de la situation d'emploi des sujets qui sont passés par le système de formation professionnelle. Il existe un légende qui liste les différentes modalités admises, mais nous n'avons trouvé aucun document qui permettrait de comprendre à quelle période cet état sur l'emploi se référerait (par exemple au moment de l'acte d'inscription- ou au début du cours ou bien durant le cours), ni quand la donnée a été enregistrée, ou qui l'a validé (il s'agit d'une auto déclaration de l'inscrit au cours ou bien a-t-il été vérifié sur les bons documents?) Pour résoudre ces doutes il a été d'abord nécessaire de comprendre quand tel champ a été inséré au système (c'est à dire par rapport à quelle activité administrative ou de suivi), et ensuite trouver et contacter ceux qui normalement l'alimentent.

nouveau schéma d'échantillonnage est clairement supérieure, sans occasionner une aggravation de la fiabilité des analyses sur des sous-ensembles pertinents pour la programmation. Les proportions de l'univers pertinentes pour l'évaluation peuvent en fait être reproduites dans les échantillons à travers les affectations proportionnelles et être maintenues également par l'effet de l'échantillonnage aléatoire simple inter-strate. Dans ce mode, l'erreur standard relative aux agrégats de grand nombre ne s'aggrave pas, en préservant la capacité prédictive, alors que les agrégats plus petits ne seraient de toutes les façons analysables avec fiabilité non plus qu'avec un échantillonnage stratifié *ad hoc*.

Impact net et distorsion de sélection

Avec toutes les limitations mises en évidence et les améliorations encore possibles, l'évaluation présentée ici démontre clairement qu'estimer l'impact net est possible, également pour les politiques qui ne sont pas des interventions expérimentales ou des projets pilotes, mais sont consolidées et d'application de grande ampleur.

Dans des actions de grande ambition comme celles évaluée ici, financées par des fonds FSE, une stratégie d'évaluation expérimentale pure, avec une affectation aléatoire à l'échantillon principal et à celui de contrôle, est impraticable. Le choix de méthodes quasi expérimentales est une route possible, mais qui n'élimine pas totalement les risques de distorsion non observables en amont, ni en aval, à travers les méthodes économétriques expérimentées. La possibilité d'examiner les résultats d'impact net en termes relatifs est beaucoup plus fiable: pour quelle cible cela a-t-il une plus grande efficacité nette? Dans quels contextes? Avec quels instruments? Les réponses à de telles questions peuvent être employées d'un côté pour atteindre une gestion rationnelle des ressources financières disponibles, de l'autre pour améliorer l'efficacité de l'intervention. En outre, les immenses différences observées entre les indicateurs d'impact net et lourd montrent combien il est essentiel de se confronter avec un groupe contrefactuel, tout aussi imparfait qu'il puisse être.

Ces observations, évidemment, n'oublie pas la difficulté inhérente à l'opération, liée surtout à la très forte *selection bias* qui caractérise les politiques évaluées, du fait que celui qui fréquente les cours de formation ne présente pas les caractéristiques moyennes de la population, ni de la population au chômage. En particulier dans le Piémont, nous observons une empreinte notable du système de formation professionnelle comme action pour la compensation des handicaps, comme dans la majeure partie de la réalité italienne, mais différemment de la réalité étrangère. Ceci ouvre la discussion sur le *benchmark* à adopter comme échantillon de comparaison pour l'analyse contrefactuelle et sur les méthodes pour l'analyse correcte des données. L'importance de la *selection bias* rend difficile de construire un plan d'évaluation de l'impact net avec des méthodes quasi expérimentales, parce que le handicap, surtout le plus grave, qui dérive des caractéristiques sociales, est pratiquement impossible à mesurer et dans tous les cas difficile à observer, pour des raisons objectives de *privacy*. Par conséquent il est irréaliste de vouloir construire un échantillon de comparaison issu de listes génériques des chômeurs, puisque les variables qui déterminent la décision de participer à la formation et le niveau de employabilité individuel sont généralement inobservables.

Ce raisonnement explique aussi le choix (fort, survenu en aval de nombreuses discussions, mais duquel nous sommes convaincus) d'extraire l'échantillon de comparaison d'entre ceux qui se sont inscrits à la formation mais n'ont pas fréquenté le cours ou qui, de toute façon, n'ont pas obtenu le diplôme final. Avec cette stratégie, nous avons obtenu un échantillon de comparaison d'une faible quantité, mais qui réduit au maximum la *selection bias*, ou les différences non observables entre traités et non traités. Par contre, en utilisant les listes génériques des chômeurs comme groupe de comparaison, la distorsion de la sélection serait probablement négative: celui qui fréquente les cours des formations n'est pas plus fort, en tant qu'il se caractérise par une attitude proactive qui le rend plus apte et séduisant pour le marché du travail, comme il apparaît dans la littérature internationale, mais il est plus fragile, parce qu'il arrive à la formation après avoir expérimenté des échecs et/ou de longues périodes de chômage. Ceci n'est pas en soi un élément critiquable des paramètres de *policy*, au contraire, remarquer que les politiques de formation agissent non seulement pour l'amélioration de l'employabilité, mais aussi dans la compensation des handicaps, représente un résultat important, qui doit être pris en considération dans l'élaboration du schéma d'évaluation.

Caractéristiques individuelles et employabilité.

Bien que les résultats du Rapport de Placement sur les déterminants de l'employabilité soient en phase avec ce que rapporte la littérature, le thème cependant doit être traité en évitant sa banalisation. Si, par exemple, se confirme encore une fois que l'âge augmente la probabilité d'emploi, des différences notables apparaissent également avec évidence pour ce qui touche aux aspects plus qualitatifs de l'emploi (durée et type de contrat, rémunération), pour les femmes mais aussi pour les autres catégories de sujets fragiles. L'évaluation ne peut donc se limiter à des indicateurs quantitatifs d'insertion dans le monde du travail, mais doit s'enrichir d'analyses plus approfondies pour ne pas représenter une image déformée des résultats.

Une autre raison qui rend complexe à évaluer les déterminants de l'emploi est l'effet des interactions. Par exemple, les étrangers ont des difficultés majeures à trouver un emploi, mais un tel effet n'est pas univoque, parce qu'il varie si nous considérons des sous-groupes mieux détaillés (étrangers communautaires ou extra-communautaires, par exemple). En

outre, cela s'atténue si nous introduisons une variable qui représente le chômage de longue durée, les chômeurs de longue durée étant, en général, très difficiles à placer sur le marché du travail. Si nous considérons une telle variable en interaction avec la nationalité, cette dernière devient non-significative, montrant qu'en réalité ce sont les étrangers en chômage de longue durée qui sont fragiles, et non les étrangers *tout-court*. En généralisant de tels exemples, cela nous rappelle l'importance de l'analyse multidimensionnelle, laquelle doit toutefois vérifier de nombreuses hypothèses différentes, puisqu'un résultat apparemment général (dans l'exemple, pour les étrangers) peut être entraîné par un effet de composition (dans l'exemple, par la présence de nombreux chômeurs de longue durée parmi les étrangers extra-communautaires).

Qualité de l'évaluation et définition des objectifs de policy

Egalement d'un point de vue méthodologique, l'activité d'évaluation est compliquée par la faible clarté et définition des objectifs des politiques. Un problème détecté par notre enquête concerne l'hétérogénéité de l'objet d'analyse: à l'intérieur de chaque action les cours financés apparaissent extrêmement divers en durée, finalité et destinataires. Afin de ne pas proposer de généralités erronées, il apparaît donc utile de distinguer clairement les objectifs et les instruments qui caractérisent chaque action de formation spécifique, en proposant des évaluations plus limitées sur quelques aspects importants.

Il est compréhensible que la structure des actions découle d'une longue stratification de changements advenus dans le temps, également pour tenir compte d'exigences complexes et non directement connexes à l'objectif immédiat de l'employabilité des destinataires. Il faut pourtant reconnaître avec clarté que la situation actuelle est bien éloignée de l'idée d'une programmation "*evidence based*", dans laquelle soit clairement identifiable le flux logique "analyse des exigences" / "définitions des objectifs" / "choix des instruments" / "évaluation des résultats" / "reprogrammation".

Il est prévisible qu'un tel changement dans la *forma mentis* demande beaucoup de temps pour se produire et puisse advenir seulement à la suite d'une prise de conscience des autorités de gestion et par la suite d'une série de changements sur des parties de la programmation, en évitant des révolutions traumatiques pour l'administration elle-même et pour les systèmes qui dépendent des politiques. Aujourd'hui l'évaluateur se trouve à devoir observer l'efficacité des actions dont les objectifs sont hétérogènes et souvent peu homogènes au regard des instruments choisis. Si l'objectif est vague et la cible générique, l'évaluation est peu incisive parce que l'efficacité est déterminée par le croisement d'un trop grand nombre de variables. L'évaluateur aujourd'hui, retrouve également un rôle maïeutique, celui de faire émerger, où il est possible, les raisons réelles qui ont amenées à la création d'une action précise et de ses caractéristiques. Une évaluation réussie, néanmoins, peut n'être exécutée qu'à travers sa pleine intégration à l'intérieur du système de suivi. Dans ce sens, l'évaluation n'est pas un simple instrument pour le contrôle administratif, mais devient plutôt un support pour la *policy-making*. L'intégration permettrait entre autre d'insérer le schéma d'évaluation au cours de la phase de programmation même de la politique, permettant de faire le choix d'une approche optimale, même quand il n'est pas possible ou souhaitable d'adopter une approche expérimentale, qui représente techniquement la solution optimale.

10. Futurs développements de la recherche

L'enquête conduite pour ce premier Rapport de Placement présente de nombreuses marges d'extension, qui se verront développées dans une future ligne de recherches, de façon à rendre toujours plus complète la définition et la quantification de l'impact net de la formation professionnelle en objet.

Tout d'abord, les enquêtes économétriques développées jusqu'ici se sont concentrées sur un indicateur dichotomique de l'état de l'emploi observé à environ une année de la fin des cours (employés vs chômeurs), définissant sur un mode un peu simplifié l'employabilité des individus comme probabilité d'être employé en octobre 2011. Les modèles probit proposés ont exploré l'incidence des divers facteurs sur la probabilité d'emploi, étudiant la signification statistique des impacts correspondants sur des cibles spécifiques. Ces analyses seront étendues dans le futur, en utilisant un indicateur différent de la situation de l'emploi, capable de saisir non pas tant les emplois ponctuels, qui fournissent des indications partielles et limitées sur l'employabilité, mais plutôt avant tout, la durée globale des épisodes de travail dans le cours de l'année. Analysant les déterminants du nombre de mois travaillés, nous pouvons comprendre aussi les aspects liés à la capacité des individus de garder un emploi, qu'une analyse de placement basée sur l'état de l'emploi dans un instant bien défini n'est pas en mesure de développer.

En outre, il serait intéressant de se focaliser sur l'employabilité également d'un point de vue dynamique: si dans l'échantillon principal les personnes formées passent d'un état de chômage avant-cours à un éventuel état d'emploi post-cours, ou bien perdurent dans le chômage, l'échantillon de comparaison est composé d'un bon pourcentage d'individus qui n'ont pas terminé la formation parce qu'employés entretemps. Ces individus, si au chômage en octobre 2011, ont expérimenté au moins une transition du chômage à l'emploi, qu'une analyse statique ne réussit pas à saisir. Avec l'aide des données administratives sur les Communications Obligatoires, il est possible d'étendre l'analyse aux périodes de travail, en étudiant l'employabilité comme probabilité d'effectuer une certaine transition, plutôt que comme la probabilité d'être employé à un moment donné.

Il n'en reste pas moins vrai que l'employabilité est, pour ainsi dire, une variable latente, ou une variable multidimensionnelle non directement observable, mais qu'elle peut être décrite en utilisant une série d'indicateurs qui en saisissent les divers aspects d'*output* - par exemple la durée de l'emploi, la forme contractuelle, le salaire - et d'*input*, comme les compétences, la sociabilité, les caractéristiques individuelles. Pour développer ce type d'analyse, il est nécessaire d'utiliser des techniques économétriques particulières, parmi lesquelles l'ainsi-nommée 'modèle à équations structurelles' (Kelm, 2000), en mesure d'estimer les relations causales qui incluent des variables latentes parmi les variables indépendantes, et parmi aussi les variables à expliquer. L'adoption de telles techniques permettra alors de décrire l'employabilité sur un mode plus complet, en en étudiant les déterminants, y compris latents.

Au final, une troisième ligne de recherches que nous avons l'intention de développer concerne la définition de l'échantillon de contrôle. Des analyses multivariées apparaissent qu'il existe deux ordres de différences entre l'échantillon des personnes formées et l'échantillon des personnes qui ont abandonné la formation, qui a été choisi comme comparaison. D'un côté en effet, le test sur les différences des effets marginaux moyens (AME) de la formation, suggèrent que les personnes formées se définissent en moyenne par des caractéristiques observables plus séduisantes pour le marché du travail, ou les formés semblent les plus forts de l'échantillon de contrôle, pour ce qui peut être observé d'eux.

Simultanément, les estimations à la Heckman ne révèlent pas une *selection bias* significative et stable, mais suggèrent que l'échantillon des personnes formées se trouve en légère position de désavantage au regard de l'échantillon de contrôle sur la base de variables non observables, qui s'intègrent donc à l'estimation de l'erreur. Ce résultat, apparemment contradictoire, signale que la procédure élaborée pour le choix de l'échantillon de comparaison a sélectionné des individus très homogènes par rapport aux personnes formées. Ils représentent néanmoins un sous-ensemble très particulier de chômeurs non formés: celui qui s'inscrit à des cours de formation a des caractéristiques différentes du chômeur moyen, soit qu'il soit allé au bout des cours, soit qu'il les ait abandonnés. Pourtant, nous sommes portés à supposer (et nous voudrions le vérifier) que, choisissant l'échantillon de contrôle dans un sous-ensemble différent de chômeurs, on observe une *selection bias* très nettement négative au désavantage des personnes formées, conséquence du fait que la cible de la formation professionnelle dans l'action sélectionnée soit caractérisée par des individus très fragiles, de qui la formation compense au moins partiellement le handicap face au marché du travail. Poursuivre la recherche dans de telles directions permettrait, ainsi, d'approfondir l'analyse des caractéristiques non observables des personnes formées et de leur incidence comparé aux autres sous-ensemble de chômeurs.

Bibliographie :

- ✓ Différents auteurs (2010). Attività di valutazione in itinere del POR Marche, FSE ob2 2007/2013. Tiré de http://www.istruzioneformazioneilavoro.marche.it/fse/Rapporto%20di%20valutazione%202010_POR%20Marche%20FSE%202007_2013.pdf
- ✓ Antonelli, G., & Nosvelli, M. (2002). Monitoraggio e valutazione delle politiche del lavoro per una "nuova economia". Bologna: Il Mulino.
- ✓ Barca, F. (2009). An agenda for a reformed cohesion policy. A place-based approach to meeting European Union challenges and expectations. Independent report prepared at the request of Danuta Hübner, Commissioner for Regional Policy.
- ✓ Battesti, JP, Bondaz, M, Marigeaud, M. (et alt.) (2012). Cadrage méthodologique de l'évaluation des politiques publiques partenariales :guide / Inspection générale de l'administration, Inspection générale des finances, Inspection générale des affaires sociales. Paris : IGA : IGF : IGAS.
- ✓ Battistin, E., & Rettore, E. (2002). Testing for programme effects in a regression discontinuity design with. *Journal of the Royal Statistical Society. Series A* , 165 (1), 39-57.
- ✓ Bell, S., Orr, L., Blomquist, J., & Cain, G. (1995). Program Applicants as a Comparison Group in Evaluating Training Programs: Theory and a Test. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.
- ✓ Bellio, R., & Gori, B. (2003). Impact evaluation of job training programs: selection bias in multilevel models. *Journal of Applied Statistics* , 30 (8), 893-907.
- ✓ Berliri, C., & Pappalardo, C. (2001). La misurazione dell'impatto occupazionale dei programmi formativi. Tratto da http://www.lavoro.gov.it/NR/rdonlyres/49AEFE8C-46E6-46CB-920A-A340C39ECEAE/0/012_misurazione_impatto.pdf
- ✓ Berliri, C., Bulgarelli, A., & Pappalardo, C. (2002). Valutazione della qualite dei programmi formativi attraverso la stima di occupabilità. *Lavoro e Relazioni Industriali* , 1, 25-54.
- ✓ Berliri, C., De Vincenzi, R., & Villante, C. (2003). Linee guida per la valutazione degli effetti occupazionali del Fondo sociale europeo 2000-2006. Metodologie per la valutazione di programma. ISFOL.
- ✓ Burtless, G., & Greenberg, D. (2004). Evaluating Workforce Program Using Experimental Methods. 2004 National Workforce Investment Colloquium. Arlington, VA.
- ✓ Card, D., Kluve, J., & Weber, A. (2010). Active Labour Market Policy Evaluations: a Meta-Analysis. *The Economic Journal* , 120, F452-77.

- ✓ Centra, M., & Falorsi, D. (2007). Strategie di campionamento per il monitoraggio e la valutazione delle politiche. Roma: ISFOL.
- ✓ CGD. (2006). When Will We Ever Learn? Washington, DC: Center for Global Development.
- ✓ Cicchitelli, G., Herzel, A., & Montanari, G. E. (1992). Il campionamento statistico. Bologna: Il Mulino.
- ✓ Ciravegna, D., Favro-Paris, M., Matto, M., & Ragazzi, E. (1995). La valutazione delle politiche attive del lavoro: esperienze a confronto. Torino: Edizioni Centro Congressi Torino Incontra.
- ✓ Cochran, W. (1977). Sampling Techniques, III ed. New York: John Wiley & Sons.
- ✓ Conseil d'analyse économique (2013), L'évaluation des politiques publiques, note du février 2013
- ✓ Friedlander, D., Greenberg, D. H., & Robins, P. K. (1997). Evaluating government training programs for the economically disadvantaged. *Journal of Economic Literature* , 4, 1809-55.
- ✓ Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection, and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement* , 5 (4).
- ✓ Heckman, J. L., & Smith, J. (1999). The economics and econometrics of active labour market programs. In O. Ashenfelter, & D. Card, *The Handbook of Labour Economics*, vol. 3. Amsterdam: North-Holland.
- ✓ Hollenbeck, K. (2005). On the Use of Administrative Data for Workforce Development Program Evaluation. ETA Occasional Paper (2005-09).
- ✓ Hollenbeck, K., & Huang, W.-J. (2003). Net Impact and Benefit-Cost Estimates of the Workforce Development System in Washington State. Upjohn Institute Technical Report (TR03-018).
- ✓ Hollenbeck, K., & Huang, W.-J. (2006). Net Impact and Benefit-Cost Estimates of the Workforce Development System in Washington State. Upjohn Institute Technical Report (06-020).
- ✓ Hollenbeck, K., & Huang, W.-J. (2007). Workforce Program Performance Indicators for The Commonwealth of Virginia. Upjohn Institute Technical Report (08-024).
- ✓ Hollenbeck, K., Schroeder, D., King, C. T., & W.-J., H. (2005). Net Impact Estimates for Services Provided through the Workforce Investment Act. ETA Occasional Paper 2005-2006 .
- ✓ IEG. (2006). Impact Evaluation Experience of the Independent Evaluation Group of the World Bank. Washington, DC: World Bank.
- ✓ IReR. (2006). Il ruolo del FSE nelle politiche formative regionali. Scenari per il 2006. Tratto da <http://www.consiglio.regione.lombardia.it/documents/38960/9400/fondo+soc+europeo+sint+consiglio.pdf>
- ✓ IRPET. (2011). Valutazione di impatto degli interventi formativi per i disoccupati e gli inoccupati. Il caso della Toscana. Tratto da http://www.irpet.it/storage/pubblicazioneallegato/341_StudixConsiglio%207_11%20completo.pdf
- ✓ ISFOL. (2003). Linee guida per la valutazione degli effetti occupazionali del Fondo sociale europeo 2000-2006. ISFOL, Struttura nazionale di valutazione Fse.
- ✓ Kluge, J. (2010). The effectiveness of European active labor market programs. *Labor Economics* , 17, 904-918.
- ✓ Lalla, M. (2003). Il disegno della seconda indagine sulle condizioni economiche e sociali delle famiglie nella Provincia di Modena. Modena: Università degli Studi di Modena, Materiali di discussione.
- ✓ Lalla, M., & Fiorani, G. (2004). Strategia di campionamento per l'oni economiche e sociali delle famiglie nella Provincia di Modena. <http://www.capp.unimo.it/pubbl/altrepubl/archivio/LallaFioraniCampionamentoIsfol.pdf>
- ✓ Lauro, C., & Ragazzi, E. (2011). Sussidiarietà e campionamento per l'oni economiche e sociali delle famiglie nella Provincia di Modena. Milano: Mondadori Education S.p.A.
- ✓ Levy, P., & Lemeshow, S. (2008). Sampling of populations. Methods and applications, Fourth edition. Hoboken: Wiley.
- ✓ Martini, A., & Garibaldi, P. (1993). L'informazione statistica per il monitoraggio e la valutazione degli interventi di politica del lavoro. *Economia e Lavoro* , 3.
- ✓ Martini, A., & Sisti, M. (2009). Valutare il successo delle politiche pubbliche. Metodi e casi. Il Mulino.
- ✓ Rosenbaum, P. R., & Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* , 70, 41-55.
- ✓ Sestito, P. (2002). Il mercato del lavoro in Italia: come of the Propensity Score in Observational Studies fo
- ✓ Sianesi, B. (2008). Differential effects of active labour market programs for the unemployed. *Labour Economics* , 15, 370-99.
- ✓ White, H. (2010). A Contribution to Current Debates in Impact Evaluation. *Evaluation* , 16 (2), 153-164.
- ✓ White, H. (2011). Impact evaluation - The experience of the Independent Evaluation Group of the World Bank. Independent Evaluation Group, The World Bank.
- ✓ White, H. (2008). Of Probits and Participation: The Use of Mixed Methods in Quantitative Impact Evaluation. NONIE Working Paper (7).
- ✓ White, H. (2009). Theory-based Impact Evaluation: Principles and Practice. 3ie Working Paper (3).