

L'apport de
la microéconométrie
à l'évaluation
des politiques publiques

Thierry MAGNAC

Thierry MAGNAC*

*The benefits of
microeconometric
approaches
to the evaluation
of public policies*

Key-words:

*public policies, evaluation,
selection, experimental data*

**L'apport de
la microéconométrie
à l'évaluation des
politiques publiques**

Mots-clés:

politiques publiques,
évaluations, sélection,
données expérimentales

Summary – In this paper, we survey the microeconomic literature on the quantitative evaluation of public interventions in the labour market, focussing upon the measure of programme effects for the participants to the programme (“the effect of the treatment on the treated”). We briefly summarize the main statistical problem that evaluation studies are confronted with, that is endogenous selection. We distinguish estimation methods according to the use of controlled experiment, natural experiment or non experimental data. We continue by describing an example of programme evaluation where we study the effects of public training schemes for youths, on their subsequent employment probabilities. We use non experimental data on youth employment histories at the beginning of the nineties in France and we contrast the results of the evaluation between urban and rural areas. The main result is that the effects of training schemes on employment probabilities are not significant in both areas. We continue by discussing the benefits and costs of experimental and non experimental data according to the type of public policies and their intended range and by presenting some examples of limitations, when using controlled or natural experiments, that have been described in the literature. Despite these limitations, we conclude by insisting upon the need to more closely relate the data collection process to any programme evaluation.

Résumé – Cet article décrit les recherches en microéconométrie menées sur les aspects méthodologiques de l'évaluation d'une politique publique pour l'emploi. Après un exposé rapide du problème statistique principal de l'évaluation qui est celui de la sélection endogène, on contraste les méthodes d'évaluations effectuées à partir de données de type expérimental et non expérimental. On propose un exemple d'évaluation de la politique de stages rémunérés pour les jeunes en France dans lequel on utilise des données non expérimentales sur les trajectoires professionnelles des jeunes au début des années 90 sur les marchés du travail urbains et ruraux. On montre que les effets de cette politique de stages sur l'insertion ultérieure ne sont pas significatifs. On montre ensuite l'intérêt et les limites des expériences contrôlées qui sont courantes aux Etats-Unis mais qui n'existent pas en France. L'article conclut sur la nécessité de lier les procédures de collecte de données à l'évaluation envisagée.

* INRA-LEA et CREST, ENS, 48 Boulevard Jourdan, 75014 Paris
e-mail: Thierry.Magnac@ens.fr

Cet article s'inspire beaucoup des recherches menées en collaboration avec Denis Fougère, Valérie Lechene, Sylvaine Laulom et Michael Visser. Ces recherches ont reçu l'appui financier de la Délégation interministérielle à l'insertion des jeunes. Je tiens à remercier Jean-Pierre Bompard pour les suggestions qu'il m'a faites tout au long de ce travail ainsi que les deux rapporteurs de la revue pour leurs remarques très détaillées. Mais je reste seul responsable des thèses, omissions et erreurs de cet article.

AU cours des années 80 et 90, les politiques publiques pour l'emploi se sont développées en France de manière accélérée. On rassemble d'ordinaire sous ce vocable des mesures de politique économique aussi diverses que les aides à la recherche d'emploi ou à l'insertion professionnelle, les politiques de formation, les subventions à l'emploi par l'abaissement des charges sociales ou les mesures concernant les retraites ou préretraites. On trouvera dans Marioni, Roguet et Gubian (1997), la description de ces mesures publiques et l'histoire de leur mise en œuvre. L'intérêt public pour l'évaluation de telles politiques s'est développé au même moment (Gélot et Simonin, 1997) et on en voudra pour preuve la création d'un Conseil scientifique de l'évaluation en 1990. Celui-ci a pour mission d'évaluer l'adéquation des résultats d'une politique aux objectifs que le législateur et/ou l'exécutif s'étaient fixés. En France, l'évaluation des politiques pour l'emploi est essentiellement mise en œuvre au sein des ministères, même si certaines équipes de recherche ont été associées à ces travaux.

Les recherches en micro-économétrie ont pour principal objectif de construire des outils méthodologiques d'évaluation qui puissent être appliqués à divers types de politiques publiques et non au cas par cas. Ces outils fournissent un jugement quantitatif global qui n'est qu'une partie seulement du travail d'évaluation des politiques publiques. Des jugements qualitatifs sur des mesures publiques, leur adaptation aux bénéficiaires par exemple ou l'incohérence entre diverses mesures, seront plus à même d'être donnés par les personnes qui se servent du dispositif public et par les acteurs qui mettent en œuvre ces mesures (agences pour l'emploi, missions locales, etc.). Mais l'évaluation quantitative ne se restreint pas à l'étude de l'évaluation de l'efficacité économique de ces politiques. Elle s'attache aussi à l'étude de l'évaluation des effets qu'elles ont sur l'inégalité (Heckman, Smith and Clements, 1997). En menant ces recherches, on cherche à fournir aux décideurs politiques et aux acteurs sociaux des arguments qui reposent sur des études empiriques et dont la construction a été validée par la communauté scientifique. Ces recommandations aident donc à la décision publique et elles alimentent le débat économique et social sur les choix de politique économique.

Certaines des méthodes d'évaluation ont été importées de la biostatistique puisque l'évaluation expérimentale de l'effet des traitements a été depuis longtemps développée en médecine et en biologie. Mais les recherches se sont particulièrement développées en économie du travail qui a toujours été à l'avant-garde de l'utilisation conjointe de méthodes économiques et statistiques pour fournir des recommandations de politique économique. Il va sans dire que la recherche aux Etats-Unis a été à la pointe de ces développements puisque les réflexions approfondies

sur l'évaluation de politiques publiques ont commencé dès les années 70 par l'évaluation des coûts et bénéfices des systèmes d'impôt sur le revenu (Killingsworth, 1983).

Dans cet article, on concentrera notre attention sur l'évaluation des effets de politiques publiques pour l'emploi visant à améliorer l'emploi ou les revenus des individus en laissant de côté celles qui affectent la demande de travail (subventions à des entreprises, partage d'emplois entre exploitations,...) pour lesquelles les recherches sont moins nombreuses. L'évaluation s'appuie alors sur l'observation d'histoires individuelles sur le marché du travail, par exemple, les probabilités d'emploi ou les revenus d'individus après qu'ils ont reçu une formation, une aide à la recherche d'emploi, etc. On suit ces individus pendant une période assez longue. Pour justifier les propriétés statistiques des méthodes économétriques utilisées, il est aussi nécessaire d'utiliser un nombre assez important d'observations. En utilisant ces informations et un modèle prédictif, on peut alors reconstituer les trajectoires hypothétiques des individus dans le cas où les politiques publiques auraient été différentes. On peut distinguer plusieurs effets susceptibles d'intéresser l'évaluateur (Heckman et Robb, 1985). La question la plus courante s'intéresse à l'impact de la politique publique sur ceux qui en sont les bénéficiaires. D'autres questions portent sur l'impact d'une modification ou extension de cette politique à ceux qui n'en sont pas bénéficiaires ou sur l'impact de la politique publique quand la participation varie de façon exogène (LATE, Imbens et Angrist, 1994). Comme ces mesures nécessitent d'autres hypothèses (voir Fougère *et al.*, 1997), on considérera seulement la première question de l'impact de la mesure sur les bénéficiaires pour ne pas alourdir le texte. Le lecteur intéressé pourra se référer au survey le plus récent de Heckman, LaLonde et Smith (1999).

Pour mener à bien l'évaluation, on a besoin de deux types d'outils : un ensemble de données et un modèle statistique qui à partir des données fournit des résumés, par exemple des moyennes de l'impact de la mesure sur les bénéficiaires. Ce sont ces résumés que l'on interprète comme les résultats de l'évaluation. Les recherches menées en micro-économie depuis les années 70 ont pour objectif l'amélioration de la qualité de l'évaluation en contrôlant les deux constructions, les données et le modèle statistique. Les expérimentations biomédicales sur des êtres humains (vaccinations, essais de médicaments,...) ont montré clairement que des données expérimentales permettent l'utilisation de modèles statistiques simples qui offrent des conclusions robustes. Leur principe méthodologique est de contraster un groupe d'individus qui reçoit le traitement à évaluer (un nouveau médicament par exemple) à un groupe de contrôle qui est exclu du traitement. Les deux groupes sont constitués par des individus choisis au hasard dans la population. Les méthodes dites non expérimentales utilisent quant à elles des données d'enquêtes généralistes. La sélection dans les groupes de traitement

et de contrôle est alors généralement le résultat du libre arbitre des individus ou des agences qui gèrent la politique de l'emploi. Aux Etats-Unis, il existe des expériences contrôlées des effets des politiques publiques depuis les années 70 (voir Manski et Garfinkel, 1992; Meyer, 1995 ou LaLonde, 1995), alors qu'en Europe elles restent relativement rares (Björklund et Régner, 1996) et en France, inexistantes. Néanmoins, les expériences contrôlées en économie ne peuvent atteindre le degré de sophistication d'expériences biomédicales puisque les expériences en simple ou en double aveugle, par exemple, n'ont que peu de sens pour les politiques d'emploi (voir pourtant Fougère, 2000).

C'est pourquoi certains analystes soutiennent la thèse que l'utilisation de modèles statistiques qui prennent en compte explicitement le caractère non expérimental des données donne des diagnostics proches des méthodes expérimentales pourvu que la construction du modèle soit soumise à une batterie de tests de spécification. C'est le programme de recherche de Jim Heckman depuis 15 ans qui l'a conduit à construire une boîte à outils statistiques impressionnante (voir le survey de Heckman, LaLonde et Smith, 1999). Ses résultats les plus récents (Heckman *et al.*, 1998) montrent que les évaluations fournies à partir de données non expérimentales peuvent être robustes. Les données non expérimentales sont d'ailleurs utilisées dans la majeure partie des recherches récentes menées sur l'évaluation microéconomique en Europe (Blundell, Duncan et Meghir, 1998; Bonnal, Fougère et Sérandon, 1997; Cockx, Van der Linden et Karaa, 1998; Abbring, Van den Berg et Van Ours, 1996). Par exemple, on montre dans Magnac (1997, 2000) que l'utilisation de données non expérimentales provenant des enquêtes Emploi de INSEE sur les transitions sur le marché du travail permet d'évaluer les effets de la politique publique de stages rémunérés pour les jeunes sur la facilité qu'ils ont à s'insérer ultérieurement sur le marché du travail. On trouve que le passage par ces stages n'a pas d'effet significatif sur l'accession à un emploi stable. Cette conclusion est assez proche des effets obtenus dans d'autres pays où les politiques de formation par des stages n'ont d'efficacité que pour les femmes adultes et non pour les hommes ou les jeunes (LaLonde, 1995).

Dans cet article, on n'a pas pour objectif de résumer les recherches sur l'évaluation des politiques d'emploi en France faites sur données non expérimentales (voir Lechene et Magnac, 1996, ou Gélot et Simonin, 1997), la littérature internationale sur les expériences contrôlées (voir Fougère *et al.*, 1997; Fougère, 2000) ou les évaluations des politiques publiques pour l'emploi menées par les chercheurs de l'INRA (voir Allaire, 1997). On se propose plutôt de rappeler d'abord le problème principal de l'évaluation statistique, celui de la sélection endogène pour essayer de faire comprendre pourquoi le contrôle de la construction des données est si important. Dans la section suivante, on décrit des exemples d'évaluation de politiques publiques à partir de données non expérimentales. On détaille l'évaluation des effets des stages rémunérés sur l'insertion future

pour de jeunes adultes en contrastant les zones rurales et urbaines en France. Dans la dernière section, on revient sur les méthodes expérimentales pour en décrire les avantages, limites et conditions d'application en France. On s'intéresse aussi aux critères de choix de données non expérimentales pour mener une évaluation. On conclut par une synthèse de l'apport de ces recherches au contrôle de la qualité de la production de l'information économique.

MÉTHODES D'ÉVALUATION ET SOURCES D'INFORMATION

On précise d'abord le principe de la méthode pour quantifier l'impact de la politique publique pour ses bénéficiaires et le problème principal de cette quantification qui est celui de la sélection endogène.

Le biais de sélection

On cherche à mesurer l'impact de la politique publique sur une variable y qui représente par exemple les revenus gagnés sur le marché du travail ou la probabilité d'emploi d'un individu après son passage en stage subventionné de manière publique. Plus précisément, on note y_1 cette variable si l'individu a été bénéficiaire de la politique publique et on note y_0 cette variable s'il n'en a pas été bénéficiaire. Pour juger la mesure publique, ce que l'on voudrait apprendre à partir des données individuelles est la différence entre ces deux variables $y_1 - y_0$. Cela n'est évidemment pas possible puisque les individus ne peuvent être simultanément bénéficiaires et non bénéficiaires de la politique publique. Ce raisonnement conduit à utiliser les observations concernant des populations d'individus. Il est alors facile de calculer la différence entre les moyennes⁽¹⁾ de la variable y dans deux populations, celle qui a été bénéficiaire de la politique de l'emploi (notée $I = 1$) pour laquelle on observe y_1 et celle qui n'en a pas été bénéficiaire (notée $I = 0$) pour laquelle on observe y_0 . Cette quantification n'est pourtant pas ce que l'on cherche à mesurer à cause du problème de sélection endogène que nous exposons maintenant.

Pour des raisons qui apparaîtront plus clairement dans la suite du texte, on peut stratifier la population suivant des caractéristiques x et calculer la différence entre les moyennes de y_1 ou y_0 pour les bénéficiaires et les non-bénéficiaires dans chaque groupe de caractéristiques x . Pour être plus clair et plus précis, on utilise une notation formelle pour l'espérance conditionnelle d'une variable. Par exemple $E(Y_1 | I=1, X=x)$ est la moyenne théorique de la variable Y_1 dans le groupe des bénéficiaires $I = 1$ dotés des caractéristiques x . La différence entre les moyen-

⁽¹⁾ Ou tout autre type de résumé statistique, médiane, quartiles, etc.

nes de y_1 ou y_0 pour les bénéficiaires et les non-bénéficiaires s'écrit alors :

$$\theta(x) = E(Y_1 | I = 1, X = x) - E(Y_0 | I = 0, X = x)$$

Il devient clair que que cette différence n'est pas la mesure de l'impact de la politique sur la population des bénéficiaires puisque celle-ci s'écrit :

$$\alpha(x) = E(Y_1 | I = 1, X = x) - E(Y_0 | I = 1, X = x)$$

Il manque donc la quantité :

$$\beta(x) = E(Y_0 | I = 1, X = x) - E(Y_0 | I = 0, X = x)$$

c'est-à-dire la différence entre les moyennes de la variable d'intérêt dans les deux populations quand personne n'est bénéficiaire de la politique publique.

La quantité recherchée $\alpha(x)$ est ainsi donnée par une quantité directement mesurable à partir des données $\theta(x)$ et d'une quantité $\beta(x)$ que l'on appelle le biais de sélection endogène :

$$\alpha(x) = \theta(x) - \beta(x)$$

L'intuition de ce résultat est simple. Les différences entre les bénéficiaires et les non-bénéficiaires de la politique publique ont deux sources : l'effet réel de la politique publique et la différence initiale entre les deux groupes. Les études oubliant le biais de sélection prendront à tort la mesure de la quantité $\theta(x)$ pour la mesure de $\alpha(x)$. Dans ces études, on trouve par exemple une différence positive ($\theta(x) > 0$) des probabilités d'emploi entre utilisateurs et non-utilisateurs des stages dits de qualification, mais l'effet net peut être nul si les individus sélectionnés dans ces stages sont les plus aptes ($\beta(x) > 0$, voir par exemple Aucouturier et Gélot, 1994). De la même façon, la mesure des effets des Contrats Emploi Solidarité peut être nulle ou négative ($\theta(x) \leq 0$) sans doute pour la raison inverse ($\beta(x) < 0$). Oublier la sélection peut donc conduire à des diagnostics totalement erronés sur les politiques de l'emploi.

On expose maintenant les méthodes qui permettent de contrôler le biais de sélection suivant le type de données dont on dispose ⁽²⁾.

Données expérimentales

Cette présentation rend simple la démonstration de l'apport des données expérimentales. Dans ce cas, la sélection dans les groupes de

⁽²⁾ On remarquera que toutes les quantités d'intérêt peuvent s'appliquer à la population dans son entier si on ne distingue pas les caractéristiques x . Sinon, on peut aussi pour obtenir des effets moyens dans la population calculer les moyennes des effets conditionnels aux caractéristiques x , en les pondérant par la fréquence de chaque modalité x .

traitement ($I = 1$) et de contrôle ($I = 0$) s'effectue par un tirage aléatoire qui est indépendant de la variable Y_0 dont dépend le biais de sélection. Ceci veut dire qu'avant le passage en mesure publique, les distributions de la variable d'intérêt sont identiques dans les deux groupes. Cette identité, qui est conditionnelle aux variables x , implique que :

$$E(Y_0 | I = 1, X = x) = E(Y_0 | I = 0, X = x) \quad (1)$$

Le biais de sélection est donc nul :

$$\beta(x) = 0$$

Le calcul de l'impact de la politique publique est direct. Ce sont des calculs de différences de moyennes et on n'a nul besoin de modèles statistiques sophistiqués pour mener à bien l'évaluation. Cette conclusion ne reste juste que si l'hypothèse (1) est vérifiée. Ce n'est malheureusement pas toujours le cas et c'est une condition qui marque les limites des expérimentations contrôlées que l'on reverra de façon plus complète dans la dernière section.

Données non expérimentales

L'utilisation de données non expérimentales nécessite plus de travail pour construire le modèle statistique. Il existe deux grands types de méthodes dans la littérature. L'une utilise les techniques statistiques d'appariement, l'autre les méthodes à variables instrumentales. L'idée de la première technique est de supposer que si on dispose de suffisamment de caractéristiques x pour décrire le phénomène étudié, le biais de sélection $\beta(x)$ sera approximativement nul. En effet, le biais de sélection vient de l'association qui existe entre la variable d'intérêt Y_0 et la variable de sélection dans le dispositif $I = 1$ ou 0 . L'idée est donc de prendre en compte toutes les causes communes des deux variables (éducation, âge, etc.) pour que conditionnellement à ces caractéristiques les deux variables I et Y_0 ne soient plus associées. Des études récentes montrent que les résultats de cette technique peuvent être très satisfaisants (Heckman *et al.*, 1998; Lechner, 1998) mais elle est encore assez peu utilisée en économétrie alors qu'elle l'est dans la littérature statistique⁽³⁾.

L'idée de la seconde technique, dite à variables instrumentales, est d'utiliser des variables qui affectent la sélection dans le dispositif mais qui n'affectent pas directement les variables d'intérêt Y_0 et Y_1 . Un

⁽³⁾ Cette différence vient peut-être de l'interprétation différente des termes aléatoires ou de l'incertitude, que font les statisticiens et les économètres. Alors que les premiers les interprètent plutôt comme des erreurs de mesure, les seconds considèrent que les termes aléatoires sont des caractéristiques individuelles inobservables. Dans ce dernier cas, il est plus difficile de justifier de l'absence de causes communes cachées pour deux variables.

exemple célèbre est l'étude de l'impact de la formation reçue au moment du service militaire sur les revenus des individus par Angrist et Krueger, 1992. Au cours de la guerre du Vietnam et pendant la période qui a suivi, la conscription était obligatoire mais ne concernait que des personnes nées certains jours du mois, jours qui avaient été tirés au sort. En supposant raisonnablement que le jour de naissance n'affecte pas les revenus, c'est une variable instrumentale puisque le jour de naissance affecte la sélection dans le dispositif. Ce type de « construction » est aussi connu sous le nom d'expériences naturelles.

On peut montrer que $\alpha(x)$ est mesurable à partir des données mais la démonstration formelle n'est pas très éclairante (voir annexe A). Néanmoins, l'intuition de la méthode est que la variable instrumentale sert d'indicateur imparfait de la sélection dans les groupes de traitement et de contrôle. La différence entre les moyennes des deux groupes $\theta(x, z)$ – où z est la variable instrumentale – est une fonction de la quantité à mesurer $\alpha(x)$ et du biais de sélection $\beta(x, z)$. Comme la quantité $\theta(x, z)$ ne dépend de z que par le biais de sélection, on arrive à éliminer celui-ci en contrastant les sous-groupes stratifiés en fonction de z , sous-groupes des groupes de traitement et de contrôle.

Estimation en doubles différences

Un cas particulier d'évaluation quantitative par la méthode des variables instrumentales est fourni par des données non expérimentales enregistrant les revenus individuels avant et après le passage en mesure (Heckman et Robb, 1985 ou Heckman et Hotz, 1989). On suppose que, si personne n'avait été bénéficiaire de la mesure, les revenus auraient augmenté au cours du temps de la même façon pour les groupes de traitement et de contrôle par exemple d'un facteur additif $\delta(x)$. On peut alors inférer à partir des observations avant le passage en mesure ($Y_0 - \delta(x)$) le biais de sélection :

$$\beta(x) = E(Y_0 | I = 1, X = x) - E(Y_0 | I = 0, X = x)$$

La mesure de l'impact $\alpha(x)$ est appelée estimateur en doubles différences puisqu'elle est la différence entre le groupe de traitement et le groupe de contrôle des différences entre les revenus après et avant la mesure publique.

Cet estimateur a été critiqué (Ashenfelter, 1978) car l'observation complète des profils de revenus des personnes qui choisissent d'être bénéficiaires de la mesure publique montre une chute importante des revenus juste avant l'entrée en mesure (« *Ashenfelter's dip* »). C'est par exemple parce qu'elle se retrouve au chômage que la personne choisit de suivre un stage. Le biais de sélection est alors mal estimé. Néanmoins il suffit alors de remonter suffisamment loin dans le temps pour corriger cet effet.

L'ÉVALUATION SUR DONNÉES NON EXPÉRIMENTALES: DES EXEMPLES

On vient de présenter rapidement les méthodes utilisées quand on ne dispose que de données non expérimentales. On expose ici des exemples de telles évaluations: d'abord, dans un cadre d'expérience naturelle, l'évaluation de l'effet d'une politique d'incitations financières au retrait d'activité pour les femmes; ensuite, l'évaluation de l'effet d'un passage par des stages rémunérés sur les trajectoires d'emploi des jeunes – au moyen d'une procédure qui contrôle les biais de sélection – pour la population vivant en zone rurale et dans les petites villes et pour la population vivant dans les grandes villes. Dans les deux exemples, on étudie une variable discrète, le fait d'être en emploi ou non par exemple, et non une variable continue comme les revenus (voir section précédente). Il faut donc adapter les méthodes précédentes à l'étude des transitions sur le marché du travail (Ridder, 1986; Lancaster, 1990).

Le premier exemple est donné par Piketty (1998). Il étudie une mesure publique pour l'emploi introduite en 1994. Celle-ci consiste à étendre aux femmes qui travaillent et qui ont un deuxième enfant l'allocation d'aide au retrait d'activité. Cette mesure existait déjà pour les femmes à la naissance de leur premier enfant. La variable d'intérêt est le taux d'emploi féminin et la question posée est la mesure quantitative de l'incitation financière au retrait d'activité. Étudier la série temporelle du taux d'emploi des femmes qui ont deux enfants pour en inférer un impact de la mesure en 1994 ne sert à rien puisque les évolutions temporelles à partir de 1994 pourraient être attribuées à d'autres circonstances macro-économiques. Mais, comme les groupes des femmes qui ont un enfant ou plus de trois enfants n'ont pas été affectés par cette politique, ces groupes peuvent servir de groupes de contrôle. Cette « expérience naturelle » permet de construire un estimateur de l'impact de la mesure en contrastant les évolutions temporelles des taux d'emploi des femmes dans ces trois groupes. Piketty (1998) montre alors que la mesure a eu un fort impact sur le retrait d'activité des femmes. L'offre de travail féminine réagit aux incitations financières au moins dans le cas où celles-ci sont liées à la présence de jeunes enfants.

Le second exemple est la mesure de l'effet des stages rémunérés, l'un des aspects de la politique française de lutte contre le chômage (Marrioni, Roguet et Gubian, 1997), sur l'insertion professionnelle des jeunes (voir aussi Bonnal, Fougère et Sérandon, 1997). L'entrée en stage n'est pas cette fois-ci commandée par une variable observable comme le nombre d'enfants. Le passage par un stage est le résultat de décisions prises par les jeunes et les personnes des agences chargées de la mise en œuvre de ces mesures, et les déterminants de ces décisions sont dans une large mesure inobservables. On utilise l'information dont on dispose concernant les trajectoires sur le marché du travail avant et après le passage en stage comme dans la section précédente. Mais il faut là

aussi prendre garde au biais de sélection puisque les bénéficiaires et ceux qui ne le sont pas ont des caractéristiques inobservables différentes. Le modèle est rapidement décrit dans l'annexe C. Il est appliqué à des données non expérimentales sur les trajectoires des jeunes entre 18 et 29 ans dérivées des enquêtes Emploi de l'INSEE de 1989 à 1992, pendant 38 mois.

Ces données sont décrites dans l'annexe B. La méthode repose sur des méthodes de maximum de vraisemblance conditionnel (Andersen, 1973; Chamberlain, 1984). On considère que l'hétérogénéité inobservable des individus commande seulement le nombre de passages dans chaque état du marché du travail au cours de la période, et non pas l'ordre dans lequel se font les passages. Conditionnellement à ce nombre de passages, l'évaluation de la facilité des transitions se fera en calculant la différence entre le nombre de trajectoires (stages, chômage, emploi) et le nombre de trajectoires (chômage, stages, emploi). Si cette différence est négative, cela indiquera que l'accès à l'emploi par un stage est plus facile que par le chômage. Cette procédure est développée dans Magnac (1997, 2000). On se contente ici de reporter les estimations de certains de ces effets.

On distingue six états sur le marché du travail. Les états sont les contrats à durée indéterminée, les contrats à durée déterminée, le chômage, les stages rémunérés, les études y compris les formations non rémunérées et la non-participation.

Tableau 1.
Effets des états
initiaux sur la (log-)
probabilité relative
des états de
destination
(échantillon total)

Etat de destination→ Etat initial↓	CDI	CDD	Stages	Chômage	Etudes
CDI	1.77 (0.50)	0.73 (0.54)	0.31 (0.78)	0.20 (0.47)	0.56 (0.97)
CDD	1.68 (0.51)	2.33 (0.58)	1.68 (0.77)	1.11 (0.50)	2.32 (0.14)
Stages	1.17 (0.67)	1.32 (0.63)	2.30 (0.98)	1.28 (0.62)	1.88 (1.23)
Chômage	0.99 (0.47)	1.37 (0.52)	1.91 (0.74)	1.13 (0.49)	0.43 (0.99)
Etudes	1.86 (0.52)	1.34 (0.61)	1.83 (0.87)	0.25 (0.58)	5.39 (1.08)

Notes : Le processus est de Markov du premier ordre. Les coefficients reportés sont les logarithmes des ratios de probabilité de l'équation (C.1) Les écarts types figurent entre parenthèses. Les états du marché du travail sont : l'emploi à durée indéterminée (CDI), l'emploi à durée déterminée (CDD), les stages rémunérés (Stages), le chômage et le système scolaire ou les stages non rémunérés (Etudes). La non-participation est l'état de référence.

Méthode de maximum de vraisemblance conditionnelle. Nombre d'observations $N = 915$, $\log\text{-vraisemblance}/N = -0.4635$.

Tableau 2.
Effets des états
initiaux sur la (log-)
probabilité relative
des états de
destination
(échantillon des zones
< 200 000 habitants)

Etat de destination→ Etat initial↓	CDI	CDD	Stages	Chômage	Etudes
CDI	1.70 (0.61)	0.43 (0.64)	-0.68 (0.86)	-0.15 (0.54)	0.58 (1.18)
CDD	1.65 (0.59)	2.39 (0.67)	0.90 (0.85)	1.10 (0.55)	2.28 (1.73)
Stages	1.02 (0.79)	1.20 (0.69)	1.91 (1.04)	1.18 (0.70)	0.61 (1.54)
Chômage	1.20 (0.58)	1.57 (0.58)	1.51 (0.78)	0.84 (0.55)	-0.30 (1.22)
Etudes	1.63 (0.59)	0.82 (0.69)	0.67 (1.00)	-0.20 (0.64)	5.95 (2.05)

Notes : Voir tableau 1.

Méthode de maximum de vraisemblance conditionnelle. Nombre d'observations $N = 600$, $\log\text{-vraisemblance}/N = -0.4471$.

On reporte dans les tableaux 1 et 2 les estimations des coefficients traduisant l'effet qu'ont les états initiaux sur les probabilités des états de destination en utilisant l'information annuelle de mars 1989 à mars 1992. Cette formule est donnée par l'équation (C.1) de l'annexe C. Un exemple est sans doute plus parlant. Le coefficient qui vaut 1.77 dans la cellule (CDI, CDI) dans le tableau 1 signifie que le logarithme de la probabilité d'être en contrat à durée indéterminée par rapport à la probabilité d'être non-participant (l'état de référence) à la période t augmente de 1.77 si l'état initial à la période $t - 1$ est un contrat à durée indéterminée au lieu d'être la non-participation. Ceci veut dire qu'il y a forte persistance dans l'état de contrat à durée indéterminée. A côté de ce résultat peu surprenant, plus intéressante est l'analyse de la différence entre les coefficients de la ligne Stages et Chômage dans la colonne CDI (1.17 - 0.99). Cette différence mesure l'impact de court-terme du passage par un stage rémunéré par rapport au chômage sur la probabilité d'accès à un contrat à durée indéterminée. Comme cette différence n'est pas statistiquement significative, l'effet des stages par rapport au chômage quant à l'insertion à court terme sur le marché du travail semble négligeable. On peut néanmoins penser que les erreurs de mesure dans les trajectoires sont importantes (Magnac et Visser, 1999), ou que l'agrégation de différents types de stages peut brouiller la mesure de l'effet. Pour l'améliorer, il faudrait disposer de données plus fines qui permettent de distinguer les stages et qui soient de bonne qualité. Cependant, en l'absence de telles données, la justification de cette politique publique par son effet sur l'insertion à court terme semble fragile.

On peut contraster ces estimations en stratifiant l'échantillon par diverses variables caractéristiques. Dans Magnac (1997), on utilise des stratifications suivant l'âge, le sexe ou l'éducation du père. Ici, on stratifie l'échantillon suivant la zone d'habitation. Une procédure de test montre que l'opposition principale apparaît entre zones rurales et petites villes d'un côté, les grandes villes y compris Paris de l'autre. Les estimations des mêmes coefficients pour l'échantillon, zones rurales et petites villes, sont présentées dans le tableau 2. Certaines différences entre les deux échantillons ne sont pas significatives. En particulier, l'effet d'un passage par un stage par rapport au chômage sur la probabilité d'emploi ne semble guère différent entre les deux échantillons et n'est pas non plus significatif. Les différences entre zones d'habitation sont surtout significatives quand les états de destination sont les stages et le chômage. Les transitions vers ces états sont uniformément moins probables en zone de densité d'habitation faible. Comme cette analyse se fait par rapport à la non-participation, on peut en déduire que s'il y a sortie d'emploi ou du système scolaire, les individus en zones rurales et dans les petites villes tendent à choisir plus fréquemment la non-participation que les stages et le chômage. Ceci pourrait traduire un effet de découragement plus grand des personnes jeunes dans les zones rurales et les petites villes quant à leurs perspectives de trouver un emploi.

QUELLES DONNÉES RECUEILLIR OU CONSTRUIRE POUR L'ÉVALUATION ?

Ce qui précède, mais surtout l'ensemble de cette littérature, montre que les expériences contrôlées semblent apporter un gain important dans l'évaluation des politiques publiques en termes de simplicité des méthodes statistiques et donc en termes de lisibilité et de diffusion auprès du public des résultats. Elles peuvent aussi servir à expérimenter, dans le sens traditionnel du terme, certaines politiques afin de mieux calibrer leurs paramètres au problème économique que l'on cherche à résoudre, avant de les généraliser. Certaines politiques, si elles n'entraînent que des changements mineurs, peuvent n'avoir aucun effet et l'expérimentation permet de mieux saisir à partir de quel seuil les effets ne sont plus négligeables. Il y a néanmoins de nombreuses limites à l'utilisation d'expériences contrôlées.

En premier lieu, le coût des expériences est très élevé⁽⁴⁾ même s'il ne représente qu'une faible fraction du coût total de la politique à évaluer (Heckman *et al.*, 1999). Il faudrait d'abord mettre en regard ces coûts avec la disponibilité et le coût d'usage d'enquêtes généralistes qui servent à d'autres buts mais qui donnent sans doute des évaluations de

⁽⁴⁾ L'expérience américaine sur les plans de santé, portant sur un échantillon de 20000 personnes entre 1974 et 1977, coûta 136 millions de dollars US 1984 (Manning *et al.*, 1987).

moindre qualité. Il faudrait aussi les mettre en regard avec les bénéfices attendus de l'évaluation, en termes de bien-être social, de cette politique et de la meilleure calibration que permet l'expérimentation. Exercice difficile et qui, tenté, semble très imprécis même si les auteurs montrent généralement que les bénéfices sont plus grands que les coûts (voir par exemple Manning *et al.*, 1987).

En second lieu, certains analystes de la politique de l'emploi ont souvent des réticences vis-à-vis des expériences contrôlées à cause de leur légalité. C'est pourquoi on s'intéressera dans la sous-section suivante au cadre juridique encadrant l'expérimentation biomédicale en France, et on montrera comment il pourrait être utilisé pour mener à bien des expériences contrôlées sur la politique de l'emploi.

Mais certaines politiques, comme celles du relèvement du SMIC en France, ne peuvent être évaluées expérimentalement. On montrera alors comment les évaluateurs sont revenus à l'utilisation d'enquêtes généralistes pour traiter de ces problèmes. On démontrera ensuite, brièvement, que le contrôle des expériences est souvent imparfait et que la généralisation des résultats expérimentaux est quelquefois très fragile. On finira cette section en revenant sur les critères à appliquer au choix de données non expérimentales pour mener une évaluation en l'absence de données expérimentales.

Le cadre juridique de l'expérimentation ⁽⁵⁾

Il est quelquefois avancé que les expériences contrôlées n'auraient pas de cadre légal. Pourtant, les expériences contrôlées biomédicales existent en France et les questions éthiques qu'elles posent semblent tout aussi, sinon plus, graves que dans le cas de l'emploi. En effet, même si des pré-tests sont effectués, l'expérimentation d'un nouveau médicament pour une maladie grave peut être très risquée pour la santé ou la vie même des personnes qui acceptent de participer. On expose en effet les participants du groupe de traitement à des risques liés à des effets secondaires indésirables. Mais on expose aussi les membres du groupe de contrôle à ne pas être traités par un médicament connu mais moins efficace.

Le cadre juridique pourrait s'inspirer de la loi encadrant en France l'expérimentation contrôlée biomédicale (loi du 20 décembre 1988). Trois principes la régissent : les expérimentations sont systématiquement évaluées par des comités consultatifs spécialement créés ; le consentement écrit des personnes participant à l'expérience est obligatoire ; les expérimentations sans finalité thérapeutique directe sont encadrées plus strictement. Le Conseil constitutionnel précise les conditions de recours à ces

⁽⁵⁾ Cette partie s'inspire beaucoup du rapport Fougère *et al.*, 1997.

lois expérimentales (décision 93-322 du 18 juillet 1994, p. 204): « *il est loisible au législateur de prévoir les possibilités d'expériences comportant des dérogations aux règles..., de nature à lui permettre d'adopter par la suite, au vu des résultats de celles-ci, des règles nouvelles appropriées...; toutefois il lui incombe alors de définir précisément la nature et la portée de ces expérimentations, les cas dans lesquels celles-ci peuvent être entreprises, les conditions et les procédures selon lesquelles elles doivent faire l'objet d'une évaluation conduisant à leur maintien, à leur modification, à leur généralisation ou à leur abandon* ».

On peut classer les expérimentations contrôlées en trois catégories : celles qui portent sur un projet de politique publique, celles qui portent sur des mesures existantes mais qui ne sont pas des droits pour les participants, et celles qui portent sur des mesures existantes et qui sont des droits pour les participants. Les expérimentations de la première catégorie sont les plus faciles à mettre en œuvre. Cela nécessite que l'information complète sur l'expérimentation et les résultats qui en sont attendus soit donnée aux participants. Leur consentement écrit est obligatoire. Les expérimentations dans les autres catégories semblent moins faciles à mettre en œuvre. Néanmoins, les expérimentations peuvent être construites de telle façon que le groupe de traitement reçoive des ressources supplémentaires par rapport à leurs droits, ressources qui sont de même nature que celles qu'ils obtiennent par exercice de leurs droits. Par exemple, des expérimentations portant sur les effets de l'allocation chômage pourraient accorder une prime aux chômeurs du groupe de traitement qui trouveraient un emploi avant une durée de chômage égale à trois mois. Cet exemple est celui d'une expérience menée dans l'Illinois, expérience dont on reparlera plus bas.

Evaluation de politiques macroéconomiques

Des expérimentations contrôlées sur des politiques macroéconomiques semblent impossibles. Comment évaluer, par exemple, l'impact d'un relèvement du salaire minimum, d'un abaissement des charges sociales sur les bas salaires, du passage aux 35 heures, ou d'une politique fiscale sur le niveau d'emploi ? Ce sont pourtant des questions de politique économique très importantes. En utilisant des données de séries temporelles, on pourra difficilement démêler l'influence de la politique économique de divers chocs macroéconomiques qui se seraient produits au même moment.

Pour étudier les effets du salaire minimum par exemple, certains chercheurs se sont alors mis en quête de changements contrôlables dans le niveau du salaire minimum pour en inférer l'effet sur le niveau d'emploi. Cette démarche est plus facile aux Etats-Unis puisque le niveau du salaire minimum est non seulement fixé au niveau fédéral mais aussi par les Etats. C'est ainsi que Card et Krueger (1996) ont étudié l'effet de l'augmentation du niveau du salaire minimum sur l'emploi des

jeunes en contrastant des données d'emploi recueillies dans les restaurants rapides de deux états voisins, le New Jersey et la Pensylvanie, au moment même où l'état du New Jersey donnait un coup de pouce au salaire minimum. En utilisant l'estimateur en doubles différences précédemment exposé, les auteurs concluaient à une augmentation du niveau d'emploi quand le salaire minimum augmentait. Les conclusions empiriques de cette étude restent fragiles puisque les méthodes de collecte des données par les auteurs ont été fortement remises en cause. Néanmoins, l'idée de cette collecte est exemplaire. Il faut trouver dans l'économie des contrastes entre des populations soumises à des variations contrôlées de la politique publique.

Un autre exemple est donné par l'évaluation d'une réforme fiscale par Blundell, Duncan et Meghir (1998). Dans cet article, les auteurs évaluent des réformes fiscales qui ne touchent que certains groupes dans la population, en fonction de leurs revenus. Ils montrent que les techniques de doubles différences, comme celle vue plus haut, ne s'appliquent pas directement car elles peuvent être en contradiction avec les prédictions du modèle structurel. En particulier, l'appartenance aux groupes de traitement et de contrôle semble endogène. Néanmoins, comme les techniques utilisant des expériences naturelles ne sont après tout que des techniques à variables instrumentales, il suffit d'adapter ces dernières au modèle structurel pour mener une évaluation des réformes. Le mariage entre les acquis des recherches sur les expériences naturelles et sur les modèles structurels économiques semble donc un des principaux développements à venir dans ces recherches. Besley et Case (2001) affirment ainsi que l'adoption de politiques économiques par différents états des Etats-Unis est généralement endogène pour le phénomène que l'on veut étudier. Or la technique des doubles différences, c'est-à-dire le contraste entre les états ayant ou non adopté certaines politiques, est aujourd'hui utilisée de façon routinière pour l'évaluation et elle suppose que ces changements de politiques sont exogènes. Les auteurs montrent néanmoins que l'adoption de ces politiques est influencée par des caractéristiques exogènes des gouvernements et législatures des états (proportion des femmes par exemple,...). Ils indiquent comment reformuler alors ce qu'est une expérience naturelle. C'est un retour à la discussion bien connue en économétrie du choix des variables instrumentales, mais de façon sans doute plus claire que dans les années 70.

Les limites du contrôle de l'expérimentation⁽⁶⁾

Le plan d'expérience adopté dans la plupart des expériences contrôlées, qui sont majoritairement américaines, est le suivant. Les individus éligibles pour la mesure publique sont informés de l'expérimentation.

⁽⁶⁾ On trouvera dans Fougère (2000) une description beaucoup plus approfondie qu'ici des expérimentations contrôlées et de leurs limites.

Ceux qui acceptent d'y participer sont affectés aux groupes de traitement et de contrôle. Le déroulement inverse, qui consiste à tirer les groupes de traitement et de contrôle avant que la décision de participer soit prise par les individus, donne des résultats qui sont contaminés par la sélection. En effet, les refus de participer dans les groupes de contrôle et de traitement auraient *a priori* des motivations différentes. En particulier, les individus du groupe de contrôle ne gagneraient rien à participer, alors que le suivi par enquête auquel ils sont soumis pourrait leur sembler pesant. Les deux groupes ne seraient donc plus comparables. Il est aussi possible de mener l'expérimentation par sites. En effet, il semble alors plus facile d'exclure ou inclure certains sites dans le traitement. Mais il faut prendre garde au fait que les sites doivent avoir des populations similaires pour contrôler le biais de sélection.

Néanmoins, les expériences sur les politiques publiques ne peuvent atteindre le degré de contrôle des expériences biomédicales. Il peut, par exemple, exister des effets dits Hawthorne (ou placebo). Ce dernier est le nom d'un établissement industriel américain où des chercheurs ont fait, dans les années 30, diverses expériences sur les conditions de travail pour tenter d'augmenter la productivité des ateliers. Une expérience, par exemple, consistait à augmenter le degré de luminosité dans les ateliers et à comparer la productivité de la journée à la productivité moyenne passée. L'augmentation de la productivité était significative. Jusqu'au jour où on eut l'idée de baisser la luminosité dans les mêmes ateliers. La hausse de productivité était tout aussi significative. Il est donc fort possible que la sélection dans le groupe de traitement accroisse la variable d'intérêt par un effet de motivation ou de changement de routine. L'hypothèse (1) est violée et l'évaluation est invalide. C'est pourquoi, dans des expériences biomédicales, on donne tant d'insistance à des expérimentations en double aveugle – la personne qui reçoit le traitement et celle qui l'administre ne savent pas si le traitement est effectif ou placebo pour contrôler ces effets Hawthorne. Ceci ne semble pas possible en économie.

Par ailleurs, des expériences comme celles menées sur le programme de formation *Job Training and Partnership Act* aux États-Unis ont montré que si les agences, chargées de mettre en place la politique publique, sont évaluées suivant les résultats de l'expérimentation, ou si le nombre de places est limité, ces agences ont tendance à sélectionner les candidats pour lesquels le programme leur semble le plus bénéfique. De plus, beaucoup d'agences refusèrent de participer à l'expérience car elles ne voulaient pas écarter du programme de formation les personnes choisies pour faire partie du groupe de contrôle. Ceci est dû aussi, partiellement, au fait que l'expérimentation s'est déroulée au même moment ou plus tard que la mise en place du dispositif. Comme le dispositif est un droit, il est délicat d'interdire son accès aux personnes éligibles.

Le contrôle de l'expérience peut aussi être déficient du côté des bénéficiaires. L'exemple le plus célèbre est celui donné par Dubin et Rivers (1993) à propos d'une expérience réalisée auprès de chômeurs dans l'Illinois. Dans l'une des expériences, on donnait aux chômeurs un bon de 500 dollars qu'ils remettraient à leur employeur potentiel s'il les embauchait. Certaines conditions étaient imposées pour éviter la fraude. De plus cette subvention à l'embauche n'était versée que si la durée du chômage de l'individu ne dépassait pas onze semaines. Dubin et Rivers montrent que certains chômeurs n'ont pas utilisé ce bon de peur de l'impression négative que pourrait avoir l'employeur quant à leurs capacités. Ce sont d'ailleurs des chômeurs qui n'avaient pas de problèmes pour retrouver rapidement un emploi qui adoptaient le plus fréquemment cette attitude de refus. Comme l'échantillon final dans le groupe de traitement s'était autosélectionné, le résultat de l'expérience montrait de façon surprenante que les chômeurs du groupe de traitement sortaient moins vite du chômage que ceux appartenant au groupe de contrôle. Si on corrige des biais de sélection comme le font Dubin et Rivers, l'effet de la subvention est faible mais positif. La thèse avancée par ces auteurs est donc que les méthodes développées dans le cadre non expérimental sont utiles, même quand on utilise des données expérimentales.

Pour finir, on peut penser à d'autres problèmes qui affectent les expériences (voir Fougère *et al.*, 1997). Il peut exister des biais de substitution par exemple. Les membres du groupe de contrôle se voyant refuser l'accès à la formation qui est l'objet de l'expérience vont utiliser d'autres types de politiques publiques. L'évaluation devient alors la mesure de l'impact relatif de la mesure expérimentée parmi toutes les autres qui sont proposées. On peut aussi se poser la question de la généralisation de l'expérience. Il faut alors non pas mesurer l'effet du traitement sur les traités mais l'effet du traitement dans toute la population éligible. Cela réclame des hypothèses plus fortes et nécessite aussi qu'il n'y ait pas d'effets d'équilibre partiel si la politique concerne beaucoup de personnes (Heckman *et al.*, 1999). On revient alors au problème de l'évaluation des politiques macroéconomiques, évoqué plus haut.

La qualité des données non expérimentales

Les données expérimentales portant sur des politiques publiques pour l'emploi n'existent pas en France. Seules des données non expérimentales sont disponibles. Elles proviennent principalement d'enquêtes auprès des individus ou des ménages effectuées par des organismes comme l'INSEE, le CEREQ, la DARES ou d'enquêtes administratives tirées des fichiers de l'ANPE, des caisses d'assurance maladie ou de retraites. On voudrait insister sur l'importance de deux critères pouvant servir à juger de la qualité de ces données pour mener à bien des éva-

luations de politiques publiques. Le premier critère se réfère à la sélection de l'échantillon de travail. D'après les principes de la méthode décrits ci-dessus, il faut disposer de données sur les variables d'intérêt, y_0 et y_1 pour les deux groupes de bénéficiaires et de non-bénéficiaires. Il est donc par exemple exclu de ne disposer que d'un échantillon de bénéficiaires, le cas extrême de sélection d'un échantillon. Une sélection de l'échantillon qui ne pondérerait pas de la même façon des bénéficiaires et les non-bénéficiaires serait moins grave mais nécessiterait des méthodes de correction pour l'échantillonnage endogène (Coslett, 1993) qui compliquent la mesure des effets. Un échantillonnage dans la population suivant des caractéristiques exogènes, x , seulement est donc recommandé. Le deuxième critère de qualité des données porte sur la disponibilité et la pertinence des variables caractéristiques individuelles ou d'environnement dans la base de données. On vient de voir que les méthodes non expérimentales reposent, soit sur des méthodes d'appariement qui nécessitent un grand nombre de variables caractéristiques, soit sur des méthodes à variables instrumentales qui requièrent des variables qui affectent la sélection des bénéficiaires et non la variable d'intérêt. Dans ce dernier cas, disposer de plusieurs variables instrumentales permet d'effectuer des tests de validité des variables instrumentales, ce qui est un gain non négligeable dans la recherche de spécification d'un bon modèle statistique. Ainsi, l'observation d'un grand nombre de caractéristiques individuelles est essentielle. C'est en raison de ces deux critères que les données obtenues par enquêtes semblent dominer les données administratives. En effet, celles-ci ont pour base un échantillonnage qui n'est pas toujours exogène, et les renseignements individuels sont souvent réduits au minimum. Au contraire, les données d'enquêtes sont obtenues par sondages ce qui réduit les problèmes de sélection de l'échantillon. Elles sont aussi plus riches en variables caractéristiques. Notons enfin que les méthodes nécessitent un nombre d'observations très important.

CONCLUSION

Les expériences contrôlées ont un domaine d'application réduit et précis. Il ne s'agit donc pas de proposer des expériences contrôlées pour évaluer toute politique publique. Toutefois, les recherches sur ces expériences ont montré qu'il y a tout à gagner pour améliorer les procédures d'évaluation à penser la collecte des données dans le cadre des expériences contrôlées. Les recherches sur l'évaluation des politiques publiques pour l'emploi ont été d'ailleurs développées de manière plus approfondie dans les pays où l'accès aux données microéconomiques pour les chercheurs est le plus facile, et là où la communauté académique est intimement liée à cette collecte de données : les Etats-Unis, le Canada, la Grande-Bretagne et à un moindre degré les Pays-Bas. C'est en effet un autre avantage des données expérimentales que de construire un en-

semble de données qui soit destiné explicitement à l'évaluation de politiques publiques. Il est en effet beaucoup plus difficile d'extraire de l'information d'enquêtes qui n'ont pas été faites pour cela. Par exemple, si on trouve que le passage par des stages rémunérés a peu d'effets sur les probabilités d'emploi, c'est peut-être parce que les données sont de qualité médiocre. Les effets réels disparaissent au milieu d'un grand bruit de fond.

On peut avancer une interprétation économique des méthodes d'évaluation microéconomique d'une politique publique. Ces méthodes fournissent de l'information économique d'une certaine qualité. On peut donc les comprendre comme des technologies de production d'un bien qui est l'information économique. Les deux facteurs de production qui déterminent la qualité de cette information, dont on a parlé dans cet article, sont le contrôle des données et les procédures statistiques utilisées. Néanmoins, même les procédures statistiques les plus sophistiquées ne peuvent identifier un faible signal au milieu d'un grand bruit si les données sont peu adaptées à l'évaluation des politiques publiques. Le contraire n'est pas vrai. Des méthodes statistiques simples peuvent être utilisées si les données proviennent d'expériences contrôlées dans la limite de leur validité bien sûr. C'est sans doute ce qui donne son attrait à ces dernières. Comme nous l'avons entrevu, les procédures d'évaluation intègrent aussi des éléments de théorie économique pour construire le modèle statistique. Cela était surtout vrai dans les années 70 et le début des années 80. La vague déferlante de l'utilisation d'expériences naturelles dans la seconde moitié des années 80 pour évaluer les politiques publiques a mis l'accent sur la qualité des données, en négligeant souvent les avancées précédentes sur les modèles structurels économiques. On assiste en ce moment à un reflux marqué de cette vague (Heckman *et al.*, 1998).

En conclusion, cette rapide description de la technologie de production de l'évaluation microéconométrique a été effectuée en prenant appui sur des exemples tirés de l'économie du travail. Il est néanmoins tout à fait facile de transposer ces méthodes à l'évaluation d'autres politiques publiques, en particulier en économie agricole. Ainsi, ces outils pourraient être utilisés pour évaluer la politique de développement rural ou si l'approche contractuelle de la régulation de l'agriculture se généralise. Mais la collecte de données est un pilier de l'évaluation, au sens de fondation, auquel il est nécessaire de penser, bien avant de mener l'évaluation.

BIBLIOGRAPHIE

- ABBRING (J. H.), VAN DEN BERG, (G.) et VAN OURS (J.), 1996 — The effect of Unemployment Insurance Sanctions on the Transition Rate from Unemployment to Employment, RM38, Amsterdam, Free University.
- ALLAIRE (G.), 1997 — Emploi rural, revenus et activités : travaux de recherche de l'INRA, INRA, Secteur Sésames.
- ANDERSEN (P. K.), 1973 — *Conditional Inference and Models for Measuring*, Copenhagen, Mentalhygienjensk Forlag.
- ANGRIST (J. D.) et KRUEGER (A. B.), 1992 — Estimating the Payoff to Schooling using the Vietnam Era Draft Lottery, NBER WP 4067.
- ASHENFELTER (O.), 1978 — Estimating the effect of training programs on earnings, *Review of Economic and Statistics*, 60, pp. 47-57.
- AUCOUTURIER (A. L.) et GÉLOT (D.), 1994 — Les dispositifs pour l'emploi et les jeunes sortant de scolarité : une utilisation massive des trajectoires diversifiées, *Economie et Statistique*, 277/8.
- BESLEY (T.) et CASE (A.), 2001 — Unnatural experiments? Estimating the incidence of endogenous policies, *The Economic Journal*, 110, pp. F672-94.
- BJÖRKLUND (A.) et RÉGNER (H.), 1996 — Experimental evaluation of European labour market policy, in: SCHMIDT (G.), O'REILLY (J.) et SCHÖNMAN (K.), *International Handbook of Labour Market Policy and Evaluation*, 89-114, Cheltenham (eds), Edward Elgar.
- BLUNDELL (R.), DUNCAN (A.) et MEGHIR (C.), 1998 — Estimating labor supply responses using tax reforms, *Econometrica*, 66, pp. 827-62.
- BONNAL (L.), FOUGÈRE (D.) et SÉRANDON (A.), 1997 — Evaluating the impact of French employment policies on individual labour market histories, *Review of Economic Studies*, 64, pp. 683-713.
- CARD (D.) et KRUEGER (A. B.), 1996 — Myth and Measurement : The New Economics of the Minimum Wage, Princeton, Princeton University Press.
- CARD (D.) et SULLIVAN (D.), 1988 — Measuring the effect of subsidized training programs on movements in and out of employment, *Econometrica*, 56(3), pp. 497-530.
- CHAMBERLAIN (G.), 1984 — Panel Data, in: GRILICHES (Z.) et INTRILIGATOR (M.) (eds), *Handbook of Econometrics*, Amsterdam, North Holland.

- COCKX (B.), VAN DER LINDEN (B.) et KARAA (A.), 1998 — Active labour market policies and job tenure, *Oxford Economic Papers*, 50, pp. 685-708.
- COSLETT (S. R.), 1993 — Estimation from endogenously stratified samples, in: MADDALA (G. S.), RAO (C. R.) et VINOD (H. D.) (eds), *Handbook of Statistics*, 11, Amsterdam, North Holland.
- DUBIN (J.) et RIVERS (D.), 1993 — Experimental estimates of the impact of wage subsidies, *Journal of Econometrics*, 56, pp. 219-242.
- FOUGÈRE (D.), 2000 — Expérimenter pour évaluer les politiques d'aide à l'emploi: les exemples anglo-saxons et nord-européens, *Revue Française des Affaires Sociales*, 54(1), pp. 111-144.
- FOUGÈRE (D.), LAULOM (S.), MAGNAC (T.) et VISSER (M.), 1997 — Etude de faisabilité d'une expérimentation contrôlée pour l'évaluation d'une mesure publique pour l'emploi, Rapport à la DIIJ, 69 pages.
- GÉLOT (D.) et SIMONIN (B.), 1997 — L'évaluation de la politique de l'emploi, in: DARES (eds), *Quarante ans de politiques de l'emploi*, Paris, La Documentation Française, pp. 277-322.
- HECKMAN (J. J.) et HOTZ (V. J.), 1989 — Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training, *Journal of the American Statistical Association*, 84(408), pp. 862-74.
- HECKMAN (J. J.) et ROBB (R.), 1985 — Alternative methods for evaluating the impact of interventions, in: HECKMAN (J. J.) et SINGER (B.) (eds), *Longitudinal Analysis of Labor Market Data*, New York, Cambridge University Press.
- HECKMAN (J. J.), LALONDE (R.) et SMITH (J.), 1999 — The economics and econometrics of active labour market programs, in: ASHENFELTER (O.) and CARD (D.), *Handbook of Labour Economics*, 3(chapter 31), Amsterdam, North Holland.
- HECKMAN (J. J.), SMITH (J.) et CLEMENTS (N.), 1997 — Making the most out of program evaluations and social experiments: Accounting for heterogeneity in program impacts, *Review of Economic Studies*, 64, pp. 487-536.
- HECKMAN (J. J.), ICHIMURA (H.), SMITH (J.) et TODD (P.), 1998 — Characterizing selection bias using experimental data, *Econometrica*, 66, pp. 1017-98.
- IMBENS (G) et ANGRIST (J. D.), 1994 — Identification and estimation of local average treatment effects, *Econometrica*, 62, pp. 467-76.
- KILLINGSWORTH (M. R.), 1983 — *Labor Supply*, Cambridge, Cambridge University Press.

- LALONDE (R.), 1995 — The promise of public sector-sponsored training programs, *Journal of Economic Perspectives*, 9(2), pp. 149-68.
- LANCASTER (T.), 1990 — *The Econometric Analysis of Transition Data*, Cambridge, Cambridge University Press.
- LECHENE (V.) et MAGNAC (T.), 1996 — L'évaluation des politiques publiques d'insertion des jeunes sur le marché du travail. Les jeunes et l'emploi, Cahiers Travail et Emploi, Paris, La Documentation Française.
- LECHNER (M.), 1998 — *Training the East German Labour Force: Microeconomic Evaluations of Continuous Vocational Training after Unification*, Heidelberg, Physica.
- MAGNAC (T.), 1997 — Les stages et l'insertion professionnelle des jeunes: une évaluation statistique, *Economie et Statistique*, 304-305, pp. 75-94.
- MAGNAC (T.), 2000 — Subsidized training and youth employment, *The Economic Journal*, 110, pp. 805-37.
- MAGNAC (T.) et VISSER (M.), 1999 — Transition models with measurement errors, *Review of Economic and Statistics*, 81, pp. 466-74.
- MANNING (W. G.), NEWHOUSE (J. P.), DUAN (N.), KEELER (E. B.), LEIBOWITZ (A.) et MARQUIS (M. S.), 1987 — Health insurance and the demand for medical care: evidence from a randomized experiment, *American Economic Review*, 77, pp. 251-277.
- MANSKI (C. F.) et GARFINKEL (I.), 1992 — *Evaluating Welfare and Training Programs*, Cambridge, Harvard University Press.
- MARIONI (P.), ROGUET (B.) et GUBIAN (A.), 1997 — La politique de l'emploi, in: DARES (eds), *Quarante ans de politiques de l'emploi*, Paris, La Documentation Française, pp. 45-138.
- MCFADDEN (D.), 1974 — Conditional logit analysis of qualitative choice behaviour, in: ZAREMBKA (P.), *Frontiers in Econometrics*, New York, Academic Press, pp. 105-42.
- MEYER (B.), 1995 — Lessons from the US unemployment insurance experiments, *Journal of Economic Literature*, 33, pp. 91-131.
- PIKETTY (T.), 1998 — L'impact des incitations financières au travail sur les comportements individuels, *Economie et Prévision*, 132-133, pp. 1-35.
- RIDDER (G.), 1986 — An event history approach to the evaluation of training, recruitment and employment programmes, *Journal of Applied Econometrics*, 1, pp. 109-126.

ANNEXES

A. Estimateur à variables instrumentales

On suppose pour simplifier que la variable Z prend deux valeurs (0 et 1). Cette variable affecte la probabilité de participer et donc :

$$P(I = 1 \mid X = x, Z = 1) \neq P(I = 1 \mid X = x, Z = 0) \quad (\text{A.1})$$

mais elle n'affecte pas directement la variable Y_0 et donc :

$$E(Y_0 \mid X = x, Z = 1) = E(Y_0 \mid X = x, Z = 0) = E(Y \mid X = x) \quad (\text{A.2})$$

ni la variable Y_1 et donc la différence entre Y_1 et Y_0 . Donc :

$$\alpha(x, z = 1) = \alpha(x, z = 0) = \alpha(x) \quad (\text{A.3})$$

On part de l'expression de $\beta(x, z)$:

$$\beta(x, z) = E(Y_0 \mid I = 1, X = x, Z = z) - E(Y_0 \mid I = 0, X = x, Z = z)$$

Comme (A.2) et la loi des espérances itérées impliquent :

$$\begin{aligned} E(Y_0 \mid X = x) &= E(Y_0 \mid X = x, Z = z) = \\ &= E(Y_0 \mid I = 1, X = x, Z = z)P(I = 1 \mid X = x, Z = z) \\ &+ E(Y_0 \mid I = 0, X = x, Z = z)P(I = 0 \mid X = x, Z = z) \end{aligned}$$

on obtient :

$$P(I = 1 \mid X = x, Z = z)\beta(x, z) = E(Y_0 \mid X = x) - E(Y_0 \mid I = 0, X = x, Z = z)$$

En éliminant la quantité inconnue $E(Y_0 \mid X = x)$, on a :

$$\begin{aligned} P(I = 1 \mid X = x, Z = 1)\beta(x, 1) - P(I = 1 \mid X = x, Z = 0)\beta(x, 0) = \\ E(Y_0 \mid I = 0, X = x, Z = 0) - E(Y_0 \mid I = 0, X = x, Z = 1) \end{aligned}$$

On note :

$$\begin{aligned} p_1(x) &= P(I = 1 \mid X = x, Z = 1), \quad p_0(x) = P(I = 1 \mid X = x, Z = 0) \\ d(x) &= E(Y_0 \mid I = 0, X = x, Z = 0) - E(Y_0 \mid I = 0, X = x, Z = 1) \end{aligned}$$

quantités qui sont mesurables à partir des données.

En utilisant aussi (A.3), on a donc les trois équations :

$$\begin{aligned} p_1(x)\beta(x, 1) - p_0(x)\beta(x, 0) &= d(x) \\ \alpha(x) &= \theta(x, 1) - \beta(x, 1) \\ \alpha(x) &= \theta(x, 0) - \beta(x, 0) \end{aligned}$$

d'où on tire en éliminant les quantités inconnues $\beta(x, 1)$ et $\beta(x, 0)$:

$$\alpha(x) = \frac{\theta(x, 1)p_1(x) - \theta(x, 0)p_0(x) - d(x)}{p_1(x) - p_0(x)}$$

qui est bien une fonction de quantités mesurables et par (A.1), $p_1(x) \neq p_0(x)$.

B. Les données sur les trajectoires sur le marché du travail

Les données sont extraites de l'enquête Emploi de l'INSEE longitudinalisée (1990-1992). L'information utilisée est celle qui porte sur les calendriers d'activité reportés mois par mois pendant trois ans. Les individus dans l'échantillon ont entre 18 et 29 ans en 1992 et doivent être restés dans l'échantillon pendant les trois ans. La sélection de l'échantillon est décrite de façon plus approfondie dans Magnac (1997). Les six états sur le marché du travail qui sont distingués dans l'estimation sont ceux qui sont déclarés dans l'enquête à quelques nuances près. Le service militaire est assimilé à la non-participation et les indépendants sont assimilés aux CDI.

C. Le modèle estimé

On utilise le modèle à variables latentes de McFadden (1974). A chaque période t , la variable latente y_{ijt}^* décrit la propension de l'individu i à être dans l'état j . Cette propension est déterminée par les variables d'état retardées, y_{it-1} , et par des composantes d'hétérogénéité, ϵ_{ijt} :

$$y_{ijt}^* = \sum_{k=0}^K \delta_{kj} \mathbf{1}\{y_{it-1} = k\} + \epsilon_{ijt}$$

où $\mathbf{1}\{\cdot\}$ est la fonction indicatrice et l'ensemble des états est $\{0, 1, \dots, K\}$. Les composantes d'hétérogénéité inobservable, ϵ_{ijt} , sont décomposées en effets individuels spécifiques à chaque état, α_{ij} , qui sont censés décrire l'hétérogénéité inobservable entre individus qui génère la sélection, et des composantes résiduelles :

$$\epsilon_{ijt} = \alpha_{ij} + u_{ijt}$$

Le lien entre variables latentes et variables observables est donné par le fait que l'état observé a une propension maximale :

$$y_{it} = j \quad \text{if} \quad y_{ijt}^* = \text{Max}_k (y_{ikt}^*)$$

On suppose de plus que la distribution de u_{ijt} , conditionnellement aux effets fixes α_{ij} , est à valeurs extrêmes et est indépendante entre états, périodes et individus. On peut alors montrer que :

$$\frac{\frac{P(y_{it} = j \mid y_{it-1} = k; \delta, \alpha)}{P(y_{it} = 0 \mid y_{it-1} = k; \delta, \alpha)}}{\frac{P(y_{it} = j \mid y_{it-1} = 0; \delta, \alpha)}{P(y_{it} = 0 \mid y_{it-1} = 0; \delta, \alpha)}} = e^{\delta_{kj}} \quad (\text{C.1})$$

où 0 est l'état de référence (la non-participation), et où les normalisations $\delta_{0j} = \delta_{j0} = 0$ sont implicites. L'interprétation de δ_{kj} est facile. Si δ_{kj} est positif, les chances d'être dans l'état j par rapport à l'état 0 quand l'état initial est k sont plus grandes que quand l'état initial est 0. Les effets des états initiaux k et k' (par exemple les stages et le chômage) sur les chances d'être en CDI (l'état j) par rapport à la non-participation peuvent être mesurés par la différence entre δ_{kj} et $\delta_{k'j}$.

