

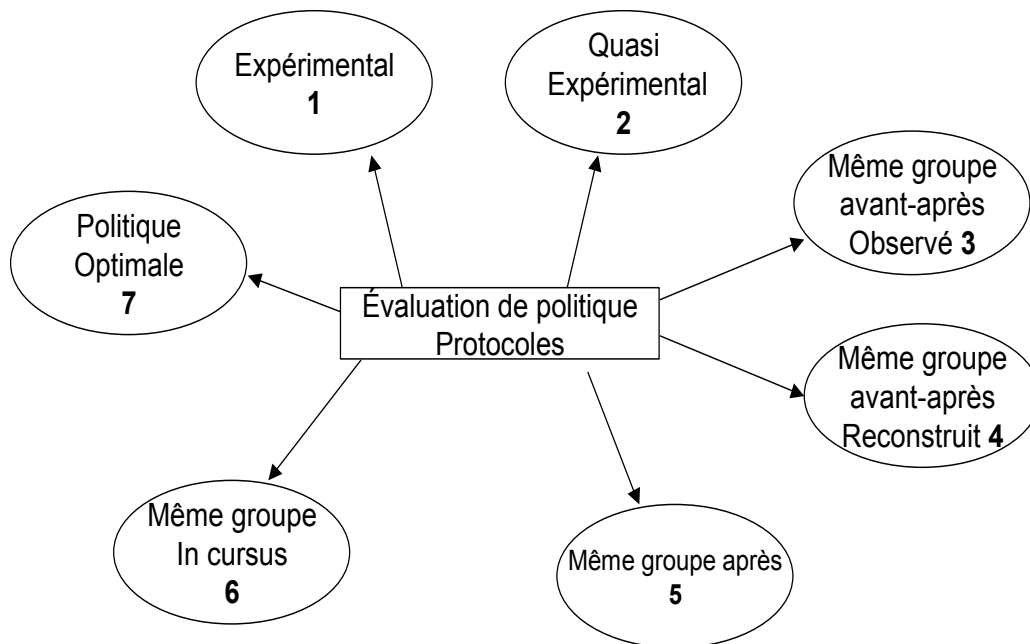
CHAPITRE 29 - PROTOCOLES ET VARIABLES D'UNE ÉVALUATION DE POLITIQUE.

L'évaluation de politique n'est pas un outil d'aide à la décision ex-ante. On l'utilise pour corriger une politique en cours de réalisation ou ex-post pour en faire l'analyse. Cela ne signifie pas pour autant qu'il convient d'en rabattre sur le plan de l'exigence méthodologique.

Une évaluation de politique demeure le résultat d'une comparaison de situations avec et sans politique. Elle exige donc un protocole comparatif utilisant une situation témoin. Dans la pratique cette exigence est souvent abandonnée au « profit » de méthodes discutables. Mais le terrain commande parfois de n'utiliser qu'une information incomplète et de qualité médiocre, ce qui rend les résultats obtenus difficilement exploitables.

7-LES PROTOCOLES METHODOLOGIQUES DE L'ÉVALUATION DES RÉSULTATS.

On peut distinguer les sept protocoles suivants.



7.1 -Le protocole d'évaluation expérimental (avec assignation aléatoire).

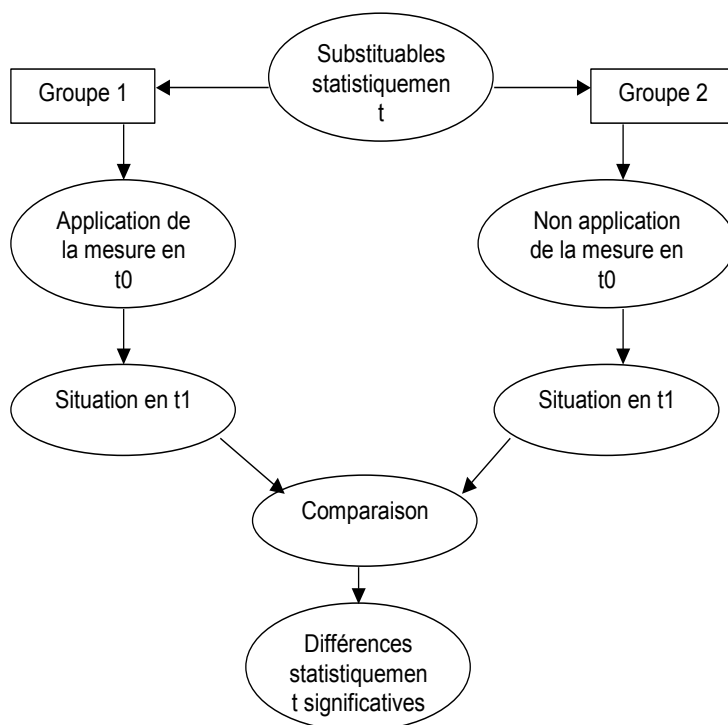
Il s'agit d'une évaluation ex post AVEC groupe de contrôle EQUIVALENT.

Ce type de technique correspond aux protocoles pratiqués dans le cadre d'expérimentations en médecine et en pharmacie. Il s'agit de la technique dite "en double aveugle".

Elle consiste en le tirage aléatoire de deux échantillons, d'individus parfaitement substituables, appartenant à une même population, auxquels on va appliquer un protocole thérapeutique.

Celui ci conduit à proposer à un groupe un traitement, dont on veut mesurer l'efficacité, et à l'autre groupe un placebo. Dans ce protocole, les patients mais aussi les praticiens, sont tenus dans l'ignorance du groupe recevant le traitement. L'évaluation résulte alors de la

comparaison des résultats obtenus au terme de l'expérimentation, et de leur signification statistique.



L'application de cette méthode aux politiques économiques et sociales présente un certain nombre de difficultés mais est garante de résultats fiables.

Le groupe 1 est nommé groupe de TRAITEMENT. Le groupe 2 groupe de CONTROLE.

- 1-L'expérimentation ne concerne que des mesures ciblées sur des individus, des entreprises, ou des territoires particuliers. Ce qui écarte les politiques globales (infrastructures, sécurité etc...
- 2-L'expérimentation peut apparaître comme contraire aux contraintes juridiques et éthiques d'égalité de traitement.
- 3-La condition de pratique "en double aveugle" peut très bien ne pas pouvoir être satisfaite (comment un groupe peut-il ignorer qu'il ne reçoit pas d'aides financières).
- 4-L'isolement du contexte expérimental, n'est pas non plus reproductible en milieu social ouvert. Les groupes peuvent avoir des contacts, et des mesures extérieures peuvent influencer les résultats.
- 5-La durée de l'expérimentation, souvent nécessaire pour lutter contre certaines pesanteurs sociales, peut affecter la mobilisation des acteurs de l'expérimentation, et la limitation de la durée peut interdire la mesure d'effets à moyen ou long terme.
- 6-La complexité des processus sociaux est souvent incompatible avec la simplification d'un protocole expérimental, destiné à mettre en évidence des relations causales supposées linéaires et univoques.
- 7-Enfin le protocole expérimental est par nécessité statique, et se trouve donc limité dans l'évaluation de processus dynamiques (changements économiques et sociaux).

7.2-Le protocole d'évaluation quasi-expérimental.

Il s'agit dans ce cas d'une évaluation ex post AVEC groupes de contrôle NON EQUIVALENTS. L'avantage principal des conceptions quasi-expérimentales est qu'elles

peuvent utiliser de sources de données existantes et sont ainsi souvent plus rapides et peu coûteuses à mettre en œuvre.

Un groupe de contrôle non équivalent, cela signifie qu'il n'y a pas satisfaction de la condition de substituabilité statistique entre les deux groupes (absence de tirage au sort). On parlera dans ce cas de protocole quasi-expérimental. Dans cette situation l'évaluateur n'intervient pas dans la constitution des groupes. La comparaison du groupe cible avec des groupes témoins imparfaits est une pratique relativement courante du fait de la difficulté de réaliser un protocole véritablement expérimental.

Bien que discutable quant aux résultats obtenus, une telle technique est bien souvent un moindre mal dans la mesure où l'expérimentation au sens strict semble interdite dans la plupart des cas dans le domaine de l'évaluation des politiques sociales notamment (politique d'attribution d'aides).

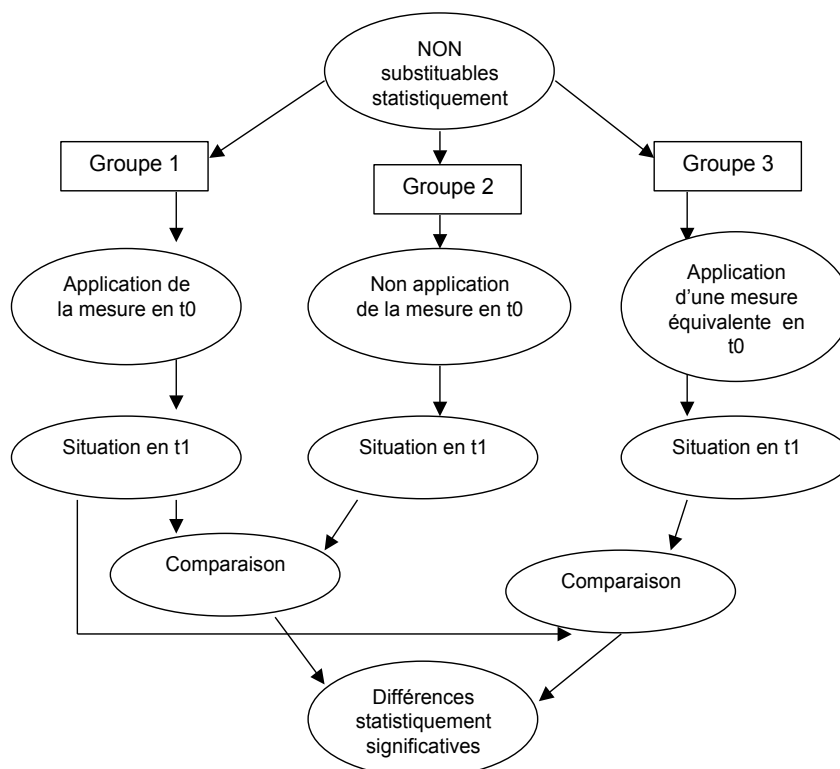
Il est possible, dans la plupart des cas, de multiplier les groupes de contrôle non équivalents, de façon à éliminer certains risques de biais, mais cela est coûteux et ne permet pas toujours d'obtenir de meilleurs résultats.

Les questions de comparaisons statistiques de séries temporelles d'observations ne présentent pas, dans ces cas, de particularités. Le recours aux méthodes de Box-Jenkins, la présence d'auto corrélations rendant les moindres carrés ordinaires inapplicables, ainsi que des tests (tests de la variance/covariance), sont ici nécessaires.

L'utilisation de la modélisation est souvent rendue indispensable pour déterminer l'influence des facteurs externes de la politique.

Dans le protocole quasi expérimental on peut distinguer différents cas, parmi lesquels :

A- Groupe de traitement au lieu X en t0 puis en t1, groupe de contrôle non aléatoire



au lieu Y en t_0 puis en t_1 .

Ex : Un programme estimé dans deux régions différentes sur une même période (Card et Krueger 1994)

B- Groupe de traitement au lieu X en t_0 , groupe de contrôle non aléatoire au lieu X en $T-1$.

Ex : Un programme estimé dans la même région mais à des périodes de temps différentes. (CEP 1992 Politique de l'eau dans le département de l'Hérault)

C- Groupe de traitement utilisant une technique R, groupe de contrôle non aléatoire utilisant une technique P.

Ex : Programme irrigation avec techniques traditionnelle et innovante.

7.3- Protocole d'évaluation sans groupe de contrôle : observation du groupe de traitement avant et après mise en oeuvre de la politique.

C'est le cas le plus fréquent. L'évaluateur ne peut que comparer la situation après la mise en oeuvre de la politique avec mise en perspective de la situation de départ. Un certain nombre de modalités est possible.

Dans le cas de cette méthode, on recourt à la comparaison d'indicateurs dans le temps. Les limites de cette pratique sont les suivantes.

- Les indicateurs utilisés après doivent avoir été saisis auparavant.
- Il sera très difficile de distinguer les changements imputables à la politique par rapport à d'autres causes influentes.
- L'observation après pose des problèmes dans la mesure où certains effets peuvent se produire à court terme, d'autres à moyen ou long terme, et parfois les mêmes effets être différents dans le temps.

7.4-Protocole d'évaluation sans groupe de contrôle : reconstruction de la situation de référence du groupe de traitement.

Il s'agit du cas d'une évaluation ex-post dont la situation de référence aurait été incomplète ou même absente. Il convient ici de reconstruire cette situation à partir des éléments disponibles au moment de l'évaluation. Ce protocole n'est pertinent qu'à la condition qu'existent des informations fiables sur la situation de départ. Dans ce cas une étude rétrospective selon les tendances passées restent parfois possible.

7.5-Protocole d'évaluation sans groupe de contrôle : observation du groupe de traitement après mise en oeuvre de la politique.

C'est naturellement une méthodologie élémentaire puisqu'il n'y a pas de terme de comparaison. Dans ces conditions, une telle pratique ne peut permettre que de constater :

- la réalité de l'application des mesures
- les effets comptables directs (nombre de bénéficiaires, coût financier...)
- les catégories de comportements.

Les techniques retenues sont en général celle de l'étude de cas (description analytique d'un événement, d'un processus, d'un organisme, d'un programme),

-d'enquêtes statistiques, présentant une meilleure validation dans la mesure où elles sont fondées sur des échantillons représentatifs.

:

7.6-Protocole d'évaluation sans groupe de contrôle : observation in cursu des effets de la politique sur le groupe de traitement.

Il s'agit de conduire l'évaluation parallèlement à la politique elle-même. Dans ce cas on utilise le recours à la mise en place d'un observatoire, ou encore l'évaluation est intégrée dans le processus d'élaboration (évaluation ex tempore).

7.7- Protocole d'évaluation par rapport à la politique idéale.

Une autre modalité de référence pour apprécier la valeur d'une politique, serait de la comparer avec la "politique optimale" possible, compte tenu des contraintes existantes. Dans ce cas, il s'agit de comparer une situation issue de l'application d'une politique avec les résultats « idéaux », possibles.

La situation à évaluer se trouvant, nécessairement entre l'absence de politique et la politique optimale.

Il est clair, cependant, que la formalisation, d'une politique optimale de référence, est en l'état actuel des possibilités techniques une totale illusion.

8-METHODES DE RECTIFICATION DES PROTOCOLES NON EXPERIMENTAUX ;

8-1-L'experimentation par assignation aleatoire.

Ceux qui sont favorables au modèle expérimental s'interrogent régulièrement sur la validité de modèles non expérimentaux. En effet il semble que les économistes qui font appel à des techniques quasi expérimentales n'ont pas vraiment réussi à isoler le biais de sélection.

Des auteurs comme Card (1985), Barnow (1987) ainsi que LaLonde et Maynard (1987) prétendent que les résultats de nombreuses études économétriques utilisant le modèle quasi expérimental ne permettent pas de mesurer correctement les effets des programmes évalués du fait de l'existence d'une distorsion entre les échantillons.

Les conclusions de Lalonde et Maynard (1987) notamment ont mis à mal les arguments évoqués en faveur des modèles non expérimentaux. Pour estimer l'exactitude de tels modèles, ils ont comparé les résultats d'une évaluation avec assignation aléatoire à ceux qui ont été obtenus par diverses méthodes non expérimentales.

Leurs conclusions sont qu'il ne semble pas exister de modèles non expérimentaux que les chercheurs pourraient utiliser en toute confiance pour reproduire les résultats expérimentaux.

Des travaux plus récents sont venus confirmer ces conclusions. Friedlander et Robins (1995) ont évalué deux stratégies quasi expérimentales classiques en fonction de données expérimentales tirées de quatre expériences de réforme de l'aide sociale. Ils ont comparé le groupe de traitement en un lieu à un groupe de contrôle dans un autre lieu, et comparé les résultats du groupe expérimental avec un groupe témoin antérieur au programme dans la même région.

Ils ont conclu que les estimations non expérimentales étaient habituellement très différentes des estimations expérimentales, surtout pour les échantillons de comparaison tirés de diverses régions. Ils ont également étudié deux techniques statistiques pouvant améliorer l'exactitude des estimations quasi expérimentales: l'appariement statistique pour la formation de groupes témoins étroitement appariés, et des « tests de caractérisation » pour évaluer statistiquement le modèle économétrique employé pour déterminer si les estimations qui en

décourent étaient exactes. Ni l'une ni l'autre stratégie n'ont sensiblement amélioré l'exactitude des estimations non expérimentales.

Greenberg et Wiseman (1992) ont aussi appuyé la méthode expérimentale en déclarant que, après quinze ans d'expériences, ils ont pu constater que l'assignation aléatoire est une méthode méthodologiquement supérieure à l'évaluation non expérimentale de programme et ont démontré que de telles études sont faisables.

En outre, les travaux sur les évaluations de l'*Omnibus Budget Reconciliation Act* (par Manpower Demonstration Research Corporation (MDRC), ont convaincu les décideurs de l'U.S. Department of Health and Human Services, qui financent habituellement l'évaluation des programmes de formation - que l'assignation aléatoire doit servir de fondement à l'évaluation de la réforme des programmes d'aide sociale. Les évaluations de MDRC ont été généralement jugées excellentes, ce qui a contribué à faire de l'assignation aléatoire la méthode retenue¹.

L'expérimentation présente un autre avantage, celui que les résultats sont compréhensibles et convaincants pour les décideurs (Burtless, 1995). Sans les réserves compliquées associées aux protocoles quasi expérimentaux, les analystes peuvent présenter des constatations expérimentales simples, comme celle-ci: « Le programme a augmenté le revenu annuel des participants de X \$ ». Étant donné la simplicité du résultat, il est plus probable que les décideurs utiliseront les conclusions de l'évaluation (Burtless, 1995, p. 67).

8-2-L'évaluation quasi expérimentale.

Certains, cependant, contestent le point de vue selon lequel les modèles expérimentaux sont supérieurs aux modèles non expérimentaux. Au premier plan de ce groupe se trouvent Heckman, Hotz et Dabos, 1987; Heckman et Smith, 1995). *Ils affirment que, s'il existe un ensemble de données suffisamment riches et qu'on a recours à de bonnes techniques de modélisation économétrique, il est possible de parvenir à des estimations fiables des impacts.*

Heckman et Hotz (1989) ont appliqué certains tests de caractérisation simples à des modèles fondés sur des données de l'expérience et ont constaté que les modèles les plus inexacts pouvaient être rejetés, mais qu'il restait un sous-ensemble de modèles avec lesquels ils ont pu obtenir des estimations des effets semblables à celles des résultats expérimentaux.

Heckman et Smith (1995) ont présenté des arguments convaincants qui contestent, dans une certaine mesure, les conclusions de Lalonde: la taille des échantillons était trop restreinte et les données géographiques étaient insuffisantes pour que les membres des groupes témoins soient placés dans le même marché du travail local que les participants; des données antérieures au programme n'étaient disponibles que pour une seule année, de sorte qu'il était impossible d'avoir recours à des stratégies économétriques efficaces (et que les estimations pouvaient subir l'effet du « fléchissement d'Ashenfelter »⁸); les études n'ont pas eu recours à une diversité de tests de caractérisation courants; et les méthodes non expérimentales ont connu des progrès importants depuis que ces études ont été faites.

Par exemple, Heckman et Smith (1995b) ont constaté que la dynamique de la main-d'oeuvre (c.-à-d. les déplacements entre l'emploi et le chômage, ainsi que l'entrée dans la population active et la sortie de la population active), de même que d'autres facteurs connexes comme l'âge, la scolarité, la situation matrimoniale et le revenu familial, peuvent servir à la formation

¹ Duncan J.D. Magnuson K.A (2003)

de groupes témoins qui sont « virtuellement identiques » aux groupes expérimentaux, de sorte qu'on peut tenir compte de l'effet du biais de sélection².

Heckman et Smith (1995) soulèvent également des objections sérieuses, à la fois théoriques et empiriques, relativement à la méthode expérimentale:

- L'assignation aléatoire peut modifier le bassin de personnes admissibles au programme ou changer le comportement des participants, ce qu'on appelle le « *biais de randomisation* ». Par exemple, pour former un groupe témoin, il peut arriver qu'on doive élargir le bassin de participants possibles, habituellement en élargissant certains critères d'admissibilité.
- S'il existe de proches substituts pour l'intervention expérimentale, alors peut apparaître un « *biais de substitution* ». Par exemple, certains programmes de formation consistent en l'achat de places dans des collèges communautaires; les non-participants peuvent choisir de suivre le même cours collégial s'ils paient eux-mêmes ou s'ils trouvent une autre source de financement.
- Les données expérimentales ne peuvent répondre à un grand nombre de questions qui intéressent au plus haut point les décideurs, notamment l'effet médian du programme et la proportion de participants subissant un effet positif (ou négatif) du programme. Il est impossible d'estimer des paramètres qui dépendent de la distribution combinée des résultats dans le groupe expérimental et le groupe témoin.. (Heckman et Smith, 1995, p.22)
- L'évaluation expérimentale indique aux décideurs si un programme fonctionne ou non. Elle ne précise pas généralement pourquoi un programme fonctionne ou ne fonctionne pas.
- En raison de certains facteurs organisationnels, il est difficile de procéder à une attribution aléatoire dans certains cas et de faire le meilleur travail possible dans d'autres cas. Le personnel peut saboter le processus parce qu'il est opposé à l'attribution aléatoire. En raison des coûts, on peut décider de ne pas procéder à la sélection au point optimal dans le processus de décision. La randomisation en plusieurs étapes, qui est coûteuse et qui perturbe le programme, peut être nécessaire à la production d'estimations des effets de services différents.
- Il faut souvent procéder à des analyses non expérimentales complémentaires pour remédier aux lacunes des modèles expérimentaux.

Burtless (1995) convient que les évaluations expérimentales comportent des lacunes, mais il invoque aussi le fait que les protocoles quasi expérimentaux ont parfois aussi les mêmes inconvénients. Elles proposent de plus généralement des problèmes statistiques plus graves que ceux que présentent les essais randomisés. Enfin on constate, qu'aucune règle propre à un modèle expérimental n'interdit aux chercheurs de faire appel à des méthodes non expérimentales pour l'analyse des données³.

8-3-Amélioration des méthodes quasi expérimentales

A-LA METHODE D'APPARIEMENT SUR LES SCORES DE PROPENSION.

² Duncan J.D. Magnuson K.A (2003)

³ <http://www.hrsc.gc.ca>

La plupart des évaluations recourant au processus quasi expérimental sont susceptibles de présenter des *biais de recrutement* dans leurs échantillons. La méthode d'appariement sur les scores de propension peut permettre de les réduire.

Il s'agira pour limiter le biais, de rechercher un groupe de contrôle aussi proche statistiquement que possible du groupe de traitement. Du moins en ce qui concerne les variables observables déterminantes. Pour cela on collecte le plus possible de variables explicatives pour chacun des acteurs. Le score de propension désigne la probabilité pour une personne de caractéristiques données de faire partie du groupe de traitement.

Le score de propension pour un individu i , est la probabilité de recevoir le traitement ($Z=1$) étant donné ses caractéristiques X_i : $e(x_i)=P(Z_i=1/X_i=x_i)$

« La distribution de ce score entre les groupes comparés fournit un critère de jugement de la comparabilité entre ces deux groupes : si biais de recrutement il y a, les scores auront tendance à être élevés pour les patients exposés et faibles pour les non exposés. Afin de neutraliser ces biais au maximum, un sous-échantillon de patients comparables entre les deux groupes peut être élaboré, par appariement sur les scores de propension : chaque patient exposé est apparié au patient non exposé ayant le score le plus proche, à condition que la différence entre les deux scores de soit pas trop grande. Ce sous-échantillon possède des caractéristiques proches du groupe de traitement, toutefois, il ne permet pas d'assurer la comparabilité sur les caractéristiques non observées »⁴.

Les différences peuvent être ajustées soit par régression soit par appariement de façon à constituer le groupe de contrôle le plus proche possible du groupe de traitement, en sélectionnant les personnes ayant la même probabilité d'appartenir au groupe de traitement que celles qui en font effectivement partie.

Lorsque le score de propension n'est pas connu et doit être estimé, on recourt le plus souvent à un modèle logistique.

$$q(x_i) = \log \frac{\Pr(Z_i = 1|x_{i,1}, \dots, x_{i,n})}{1 - \Pr(Z = 1|x_{i,1}, \dots, x_{i,n})} = \alpha_0 + \alpha_1 x_{i,1} + \dots + \alpha_n x_{i,n} = \log \frac{e(x_i)}{1 - e(x_i)}$$

$$\hat{e}(x_i) = [1 + \exp \{-(\hat{\alpha}_0 + \hat{\alpha}_1 x_{i,1} + \dots + \hat{\alpha}_n x_{i,n})\}]^{-1}$$

B-LA METHODE D'ESTIMATION PAR DIFFERENCES DES DIFFERENCES (D.I.D.).

Le principe de base de la DID consiste à comparer les variables pertinentes entre deux zones, l'une où le programme évalué est appliqué et une autre où il ne l'est pas. Sous conditions que d'autres programmes ou événements n'interviennent pas sur les résultats, alors on peut soustraire les valeurs des variables observées sur la zone de contrôle de celles mesurées de la zone de traitement.

⁴ www.rees-france.com

Cette méthode peut aussi être appliquée sur une même zone mais sur des périodes différentes. Il faut dans ce cas disposer de séries longues et veiller à écarter les interférences de programmes.

Card et Krueger cherche à mesurer l'effet du salaire minimum sur l'emploi. En avril 1992 le salaire minimum du New Jersey passe de 4,25 \$ à 5,05 \$. Ils recueillent des données sur l'emploi dans la restauration rapide avant et après la hausse (février et novembre). Dans cet Etat, l'emploi moyen par restaurant est en moyenne de 20,44 ETP (équivalent temps plein), avant la variation du salaire de 21,03 ETP après le changement de salaire. On pourrait penser que la variation du salaire minimum a entraîné une augmentation 0,59 ETP de l'emploi par restaurant.

Cependant un certain nombre de biais peuvent fausser cette conclusion. Les périodes d'observation retenues avant et après ne sont pas substituables du point de vue de l'activité des restaurants. L'écart de temps entre les observations peut faire que les conditions économiques générales aient changé entre février et novembre, (baisse du chômage par exemple).

Pour éliminer les biais, Card et Krueger ont besoin d'un groupe de contrôle. Ils se tournent alors vers la restauration rapide en Pennsylvanie, un Etat où les conditions économiques sont très semblables. Ils s'attendent à observer des variations d'emploi de même ampleur. En réalité c'est le phénomène inverse qu'il constate. En Pennsylvanie, l'emploi ETP a en fait diminué d'Avril 23.33 à Novembre 21.17, soit un écart de -2,16 ETP en moyenne par restaurant.

En estimant *la différence des différences* entre les deux périodes, on constate que l'emploi dans le New Jersey a augmenté de (0,59 - (-2,16)) soit 2,76 ETP et non seulement de 0,59.

C-LE MODELE DE REGRESSION PAR DISCONTINUE

La méthode de régression par discontinuité (RD) a été introduite par Thistlethwaite et Campbell (1960). On l'utilise lorsque l'assignation des individus à un programme dépend d'une règle fondée sur une variable comme le résultat d'un test ou la durée d'un épisode de chômage, etc...

Ce n'est que lorsque la valeur de la variable d'assignation excède un certain seuil que l'individu devient éligible au programme.

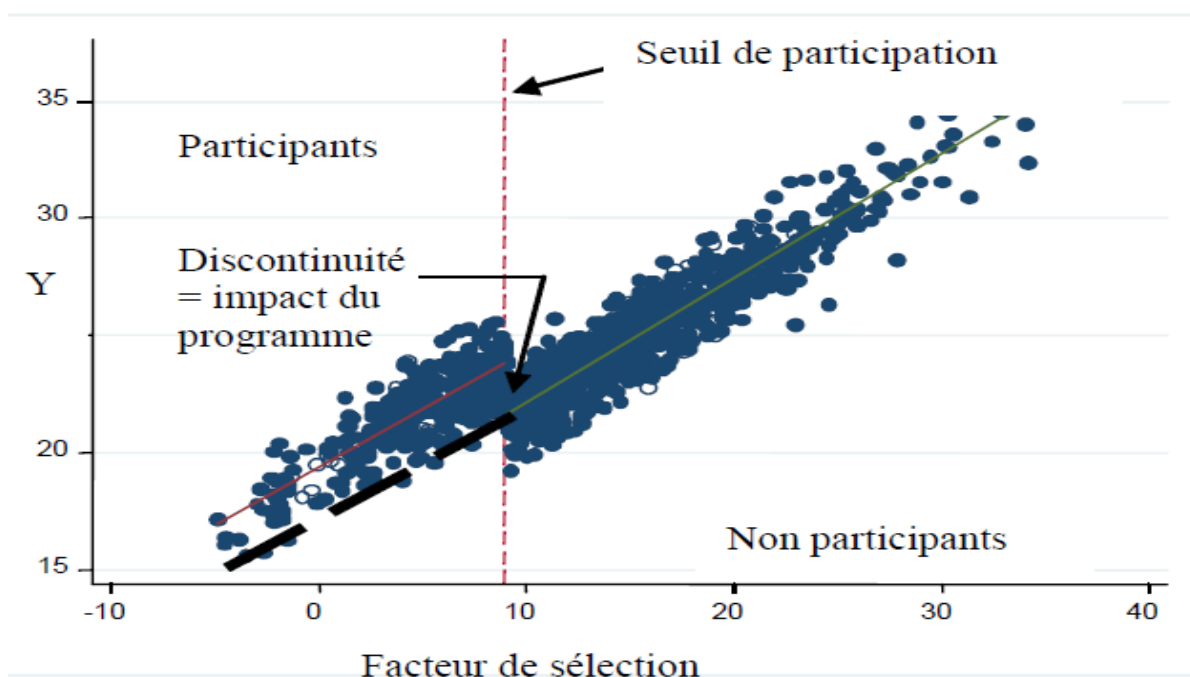
Lorsque le programme a un effet positif sur par exemple le revenu, la probabilité de trouver un emploi, ou d'autres variables pertinentes, la règle d'éligibilité crée une discontinuité dans la relation entre ces variables et la variable d'assignation.

C'est cette discontinuité qui permet d'évaluer l'effet du programme à l'aide de la méthode RD.

Le graphique suivant illustre la relation entre la variable dépendante et le facteur de sélection si le programme comporte véritablement un effet. Un saut du nuage de points est alors observé au seuil de participation du programme.

Puisque cette discontinuité survient à ce seuil très précisément, elle est alors directement associée à l'effet du programme. Cet effet peut être estimé par la différence entre le résultat moyen des participants et celui des non-participants à la « marge » du seuil de participation.

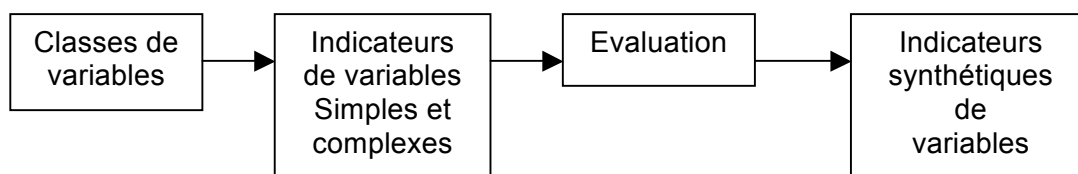
Si le nombre d'observations est suffisant, l'effet estimé du programme sera non biaisé puisqu'il s'appuie sur la comparaison des participants et des non-participants à une valeur presque identique du facteur de sélection.



(source : Santarossa, G (2008): Note d'introduction sur l'évaluation d'impact d'un programme public par la méthode de régression par discontinuité. Online at <http://mpra.ub.uni-muenchen.de/11282/> MPRA Paper No. 11282, posted 28. October 2008 / 23:34

9-LES VARIABLES DE L'ÉVALUATION GÉNÉRALE.

La mise en œuvre d'une évaluation de politiques nécessite d'identifier un certain nombre de classes de variables significatives. A partir de celles-ci, on établira des ensembles d'indicateurs simples et complexes, qui feront l'objet d'une mesure qualitative ou quantitative. Enfin des indicateurs synthétiques permettant de rendre compte des caractéristique de la politique seront élaborés.



Nous retiendrons les classes de variables suivantes :

- les OBJECTIFS,
- les BESOINS,
- les CONTRAINTES,
- les MOYENS,
- les RESULTATS
- les IMPACTS,
- la SATISFACTION

9.1 - Les objectifs.

Toute évaluation doit être conduite par rapport à des objectifs définis auparavant.

A-CATEGORIES D'OBJECTIFS.

On distinguera au cours de l'étude les catégories d'objectifs suivants:

- les principes généraux ou buts
- les objectifs stratégiques
- les objectifs opérationnels
- intermédiaires
- finaux.

La démarche doit conduire à les identifier successivement. Nous en proposons les définitions suivantes.

1-Un principe général correspond à un but de politique générale appliqué au secteur considéré. Ces principes relèvent du niveau politique le plus élevé dans la hiérarchie de la décision. Ils constituent les termes de référence du contrat politique.

2-Un objectif stratégique correspond à un niveau de la décision se situant généralement à celui du responsable politique de la conception et du contrôle des programmes. Il fixe le cadre des actions à entreprendre sans en fixer le contenu.

3-Un objectif opérationnel est un but quantifié et daté dont la responsabilité est assumée par l'instance d'exécution d'un programme. Il correspond en principe à l'intégration des contraintes identifiées dans les programmes d'action.

B-PROBLEMES D'IDENTIFICATION.

Certaines questions sont souvent évoquées concernant les objectifs d'une politique publique et leur accessibilité.

1-Il est fréquent que les objectifs ne soient pas véritablement explicités. De plus lorsqu'ils le sont, ils ne correspondent pas toujours aux objectifs réels. La question de la possibilité de l'identification des objectifs d'une politique publique est posée par certains auteurs qui suggèrent que les décideurs n'ont pas intérêt à révéler leurs objectifs, puisqu'ils seront ensuite évalués, par les électeurs sur la base de leurs résultats par rapport à ceux ci.

2-Il peut aussi exister une certaine confusion entre objectifs principaux et objectifs secondaires, ou encore entre objectifs ultimes de la politique et ses objectifs intermédiaires.

3-Les objectifs d'une organisation publique peuvent évoluer au cours du temps pour au moins trois raisons:

- Parce que certains objectifs initiaux, se révèlent impossible à atteindre, (changement de conjoncture, erreur de prévision...)
- Parce que l'on observe un changement de valeurs qui rend caduque les objectifs retenus.
- Du fait de la formation de nouvelles coalitions de pouvoir au sein de l'organisation.

4-Certains objectifs d'une même politique peuvent être en partie contradictoires. C'est par exemple le cas des politiques de développement qui se donnent des objectifs économique et de préservation de l'environnement⁵

5-Les objectifs n'ont pas tous le même horizon temporel. C'est encore le cas lorsqu'une politique se préoccupe par exemple du développement de l'emploi et de durabilité (agriculture, pêche).

6- Les objectifs n'ont pas tous le même horizon institutionnel. Dans notre étude sur le Contrat-Enfance⁶ nous avons été confrontés cette difficulté. Le fait particulier en la circonstance était que les objectifs pertinents pour l'évaluation, résultaient de l'intersection des objectifs des trois partenaires au contrat enfance (Etat, Département, commune).

C-LA VALIDATION DES OBJECTIFS.

Lorsque l'identification des objectifs est rendue difficile par l'insuffisance ou les contradictions des documents disponibles, alors leur interprétation, leur structuration et leur hiérarchisation devront être réalisées par l'évaluateur avant de les soumettre au contrôle des responsables politiques ou du comité de suivi de l'évaluation, pour validation.

Il reste que la validation des objectifs stratégiques ou opérationnels auprès des responsables d'une organisation présente le risque d'une appropriation de formulation et parfois même de contenu d'objectifs.

D-OBJECTIFS ET INDICATEURS .

Les objectifs identifiés doivent être formulés de façon à ce que leur traduction en descripteurs de résultats puisse être facilement réalisée. Cela ne concernent que les objectifs opérationnels intermédiaires ou finaux.

9.2 - Les besoins.

A-DEFINITION.

On peut définir un besoin comme une exigence ressentie par un ou plusieurs acteurs. Celle-ci peut avoir pour origine des raisons physiologiques ou socio culturelles. Une telle nécessité peut être perçue de façon consciente ou non.

B-IDENTIFICATION.

L'identification du besoin résulte de l'identification d'un écart entre deux situations.

Cet écart peut résulter de la comparaison entre une situation constatée et, soit une situation de référence explicite (indicateurs d'objectif, normes internationales...), soit une situation de référence implicite, la période précédente. L'identification d'un besoin semble nécessiter la consultation de l'individu concerné. On dit assez que la participation est indispensable à une

⁵ B.Perret op cit p20.

⁶ C.E.P-Conseil Général de l'Hérault 1996.

identification correcte des besoins. Pourtant, beaucoup d'exemples attestent que l'individu est:

- incapable de connaître ses besoins en certaines circonstances,
- incapable de distinguer le besoin d'intérêt général,
- incapable d'estimer son besoin.

Ce qui justifie d'une différence considérable entre préférences et besoins, et accrédite le rôle tutélaire des acteurs publics.

Dans le cas du Contrat-Enfance, les acteurs sont fondés à reconnaître leurs besoins concernant certains aspects de la politique d'accueil et pas d'autres. Ainsi en ce qui concerne la garde proprement dite, c'est naturellement le cas, pas forcément en ce qui concerne le développement, l'éveil, la prévention ou la socialisation. Ce qui pose le problème de l'identification de la méthodologie des besoins d'accueil.

9.3-Les contraintes.

On peut identifier deux types de contraintes qui vont peser sur la mise en oeuvre d'une politique, des contraintes extérieures et des contraintes intérieures.

A-LES CONTRAINTES EXTERIEURES.

Les contraintes extérieures que l'on constate généralement sont les suivantes:

-La présence de logiques d'acteurs spécifiques qui constituent un frein dans la mesure où ceux ci peuvent poursuivre des objectifs spécifiques pour partie, dans le cadre de politiques auxquelles la politique évaluée sera articulée.

-La méconnaissance du système que constitue la politique considérée. Ainsi dans le domaine de la politique de l'accueil, le niveau de l'information initial joue un rôle limitant dans l'accès aux structures.

-Les mentalités existantes, ainsi que l'image des actions dans l'opinion peuvent également jouer un rôle. Par exemple dans notre étude, la faveur ou l'opposition aux structures collectives d'accueil et d'éveil, peuvent être un handicap à l'efficacité d'une politique.

-Le niveau des revenus des acteurs est aussi une contrainte dans toutes les politiques où des modalités tarifaires sont mises en oeuvre.

-La disposition d'espace enfin, d'une toute autre nature, mais combien importante dans le développement de politiques urbaine doit être également soulignée.

B- L'EXISTENCE DE CONTRAINTES INTERIEURES.

Il faut également préciser l'existence de contraintes intérieures ou propres aux acteurs responsables de la politique. Il peut s'agir de:

-l'insuffisance budgétaire ou de l'endettement d'une collectivité locale par exemple, qui peut lui interdire de développer les actions souhaitées,

-l'incompétence technique et le retard spécifique pris dans un domaine, peut aussi empêcher une collectivité de retenir des ambitions capables de satisfaire les besoins.

9.4 - Les moyens ou ressources de la politique.

Les différentes catégories de moyens que nous étudierons seront respectivement les moyens techniques, humains, financiers et institutionnels.

-Les **moyens techniques** seront constitués par les structures elles-mêmes et ce qui les compose.

-Les **moyens humains**, par le personnel direct et indirect mis en oeuvre.

-Les **moyens financiers** recouvriront l'ensemble des dépenses d'investissement et de fonctionnement nécessaires.

-Enfin, **les moyens institutionnels** doivent également être mentionnés. Ils constituent le cadre juridique des possibilités d'action de l'institution.

9.5 -Les resultats

A-DEFINITION.

Un résultat correspondra à une variation d'utilité des acteurs, mesurée à partir du produit technique direct des actions de la politique mise en oeuvre. Le danger est le décalage parfois entre résultat technique et variation d'utilité (ha irrigués en pure perte).

Distinguer les résultats des impacts proprement dits, permet d'éviter des confusions. Cette distinction est d'ailleurs faite par les auteurs Anglo-Saxons sous le terme de OUTPUTS et OUTCOME⁷.

Nous conviendrons de distinguer trois types de résultats principaux.

-Les résultats techniques, qui dans le meilleur des cas correspondront ex-post aux objectifs opérationnels ex-ante. Il s'agira de places de crèche par exemple, ou de ménages connectés à un réseau.

-La variation d'utilité globale mesurée à partir des situations ex-ante et ex-post des bénéficiaires. Il s'agira ici de variation de surplus monétaire.

-La satisfaction des usagers (différente de l'utilité), peut être considérée comme un résultat particulier de la mise en oeuvre d'une politique. On peut la définir comme l'appréciation subjective d'une situation. Le problème est que la mesure de celle-ci ne peut pas permettre de savoir si le besoin est couvert. Cette satisfaction se mesurera par enquête, la plupart du temps à partir de carte de satisfaction sur laquelle l'enquêté choisira un niveau correspondant à son sentiment.

-La satisfaction des préférences des usagers pose le problème de la rationalité de ces préférences. En présence du risque d'irrationalité, comment peut-on considérer la satisfaction comme un indicateur constitutif de l'évaluation de l'efficacité. De plus si l'on admet l'hypothèse de l'existence de préférences indéterminées⁸, que signifie alors la mesure de leur satisfaction?

-Par ailleurs existe t-il des indicateurs permettant de mesurer en toute circonstance la satisfaction?

⁷ HALL R.(1991):Organisations; structures, processes and outcomes. Englewood Cliffs; N.J. Prentice Hall.

PERROW C.(1961). The analysis of goals in complex organizations American Sociological Review 26,December (pp 688-699).

⁸ THALER R. (1980): Toward a positive theory of consumer choice, Journal of Economic Behavior

-La difficulté de mesurer la satisfaction des usagers d'une politique peut aussi tenir au fait qu'il peut exister de nombreux outputs, pour de nombreux usagers, et que ceux ci peuvent en avoir une appréciation partielle. Cas d'une politique d'éducation.

-De plus un certain nombre d'éléments peuvent intervenir dans la satisfaction. Un service peut comporter un élément principal et un élément secondaire, et l'appréciation ou la dépréciation de l'un peut masquer l'effet réel de la politique. Cas de la politique d'accueil de la petite enfance, pour laquelle l'enquête de satisfaction mesure davantage la qualité des locaux ou les horaires de l'accueil que la qualité des prestations sur l'éveil cognitif, l'éveil personnel ou l'éveil social.

-Enfin la satisfaction peut être difficilement mesurable, et les protocoles d'enquête peuvent parfois introduire des biais dans sa mesure (biais de complaisance vis à vis de l'enquêteur). Ces difficultés justifient donc les précautions retenues dans l'identification et l'évaluation de la satisfaction.

B-DIFFICULTES DE LA MESURE DES RESULTATS.

La mesure des résultats pose un certain nombre de problèmes.

1-L'identification d'indicateurs. Un indicateur doit présenter certaines caractéristiques pour être utilisable. Il doit nécessairement être fiable, pertinent, sensible, facile à mobiliser, clair, comparable et peu coûteux.

2-L'existence d'informations sur la réalisation des opérations programmées, et donc postule l'existence d'une instance de suivi de ces opérations. Dans la pratique de l'évaluation, il n'est pas rare de rencontrer des manques d'information.

3-Les résultats peuvent être hétérogènes, qualitatifs ou quantitatifs, monétaires ou non....

4-Les résultats peuvent appartenir à des horizons temporels très différents.

5-La responsabilité des résultats peut être partagée entre plusieurs acteurs ou structures, ce qui gêne la mesure de l'efficacité d'une structure par rapport à une autre.

C-LES INDICATEURS DE RESULTATS.

On doit distinguer deux catégories d'indicateurs de résultats :

-les **indicateurs simples** (techniques, économique, sociaux ,...) sont relatifs aux actions ou aux projets :

EX : la production, la VA, l'emploi, le revenu, le taux de chômage....

-les **indicateurs complexes** relèvent des sous programmes ou des programmes. Par exemple :

-indicateurs de développement endogène

1-EFFETS DE SYNERGIE

Articulation technique projet-autres projets

Articulation économique projet-autres projets

Articulation spatiale projet-autres projets

2-EFFETS D'INTÉGRATION

Articulation projet-filière locale
 Articulation projet-politique locale

3-EFFETS DE VALORISATION

amélioration de ressources locales mal utilisées
 utilisation de ressources locales non utilisées
 confortement d'activités existantes
 flexibilité (mobilité culturelle, commerciale..)
 augmentation des exportations

4-EFFETS D'AUTONOMIE

développement équipements collectifs
 indépendance (réduction des importations)
 attraction d'investissements
 développement de la formation
 capacité d'autofinancement

-indicateurs de durabilité (BM).

Cet indicateur mesure la probabilité que les effets attendus d'un programme d'investissement puissent être observés durant la totalité de la durée de vie de cet investissement. Cette question prend tout son sens lorsque l'on observe que, dans un grand nombre de projets ou de programmes, la suppression de l'aide signe la suppression des effets.

En fait la durabilité peuvent s'apprécier sur les plans :

- financier** (capacité autonome de refinancement d'un équipement)
- environnementaux** (capacité de préserver ou de compenser les consommations de ressources naturelles)
- économiques** (capacité d'assurer la fourniture de facteurs et d'intrants)
- sociaux** (capacité à reproduire le capital humain)

-indicateurs de développement institutionnel (BM).

Cet indicateur mesure la capacité d'une action à permettre l'amélioration du fonctionnement des institutions. Il s'agit donc de mettre en indicateur un effet cognitif attendu.

9.6 -Les impacts (outcomes).

Il existe de nombreuses catégories d'impacts ; cela peut aller de la valeur ajoutée à l'apprentissage cognitif. Les impacts sont constitués de tous les types d'effets positifs ou négatifs, souhaités ou non, qui sont générés par un ou plusieurs programmes, et qui ne correspondent pas aux différents objectifs tels qu'ils ont été identifiés. Ils concernent plus spécifiquement les acteurs non directement ciblés par la politique (tiers absents), et se situent dans le moyen-long terme. On peut identifier de nombreuses catégories d'impacts :

1-Impacts économiques: Il s'agit des principaux éléments de la valeur ajoutée tels que les effets primaires directs, indirects, secondaires, ainsi que les effets sur l'emploi.

2-Impacts sociaux.

3-Impacts environnementaux

4-Impacts institutionnels

5-Impacts cognitifs et comportementaux etc...

L'efficacité d'une organisation peut aussi se mesurer par sa capacité à intégrer les résultats de ses expériences, en particulier, mais pas seulement, de ses échecs, dans sa démarche

future. L'analyse de défaillance constitue donc bien une étape dans l'étude du processus d'apprentissage collectif. Cependant, il ne faut pas réduire l'apprentissage à l'intégration des expériences négatives dans la modification des conduites. Les réussites sont aussi porteuses de contenus cognitifs. Surtout si le succès est suspecté⁹.

« Le succès de l'évaluation dans le secteur public peut être mesuré par l'institutionnalisation du feed back aussi bien dans la phase de formulation que dans celle de la planification ou de la mise en œuvre d'une politique publique.... En d'autres termes, l'ampleur des changements affectant le comportement des acteurs peut être considérée comme la mesure du succès de l'évaluation »¹⁰.

Pour la Banque Mondiale (IED) : « L'évaluation d'impact est une identification systématique des effets positifs ou négatifs, intentionnels ou non, sur des ménages, les institutions, et l'environnement, générés par une activité de développement, en l'occurrence un programme ou un projet ». (source IED).

10- TECHNIQUES DE L'EVALUATION DES INDICATEURS DE RESULTAT ET D'IMPACT.

En principe rien ne s'oppose à ce que l'ensemble des indicateurs soit renseigné quantitativement. Cependant dans de nombreux cas les informations disponibles seront insuffisantes ou simplement qualitatives.

Les méthodes utilisées par l'évaluation quantitative des indicateurs sont nombreuses et varient en fonction de la nature des résultats.

10.1-Les méthodes quantitatives.

Ces méthodes nécessitent l'utilisation d'une situation de référence. Elles sont donc réservées en principe aux protocoles expérimental ou quasi-expérimental. Il s'agit principalement des méthodes présentées ci-après.

A-METHODES MICROECONOMIQUES VUT DE VALORISATION DES BIENS ET SERVICES MARCHANDS ET NON MARCHANDS.

L'ensemble des outils techniques d'une évaluation de politique nécessaires pour en mesurer les résultats, correspond à celui utilisé dans l'évaluation économique des projets. C'est d'ailleurs bien naturel, puisque l'opérateur de transformation d'une politique est bien in fine un projet. Dans le cas des méthodes microéconomiques, on pourra utiliser les méthodes coûts-avantages, coûts-efficacité ou coûts-utilité avec ou sans prix de référence, et pour les effets non monétaires, on pourra recourir aux méthodes de valorisation. Il s'agira donc d'utiliser des méthodes micro, méso et macroéconomiques. On utilisera naturellement également des méthodes mono ou multi-critères.

1-La méthode coûts-avantages.

Elle consiste à identifier et à mesurer le surplus net à partir des coûts et des avantages d'une opération d'investissement, définis respectivement comme des consommations et des créations de ressources du point de vue de la collectivité. La démarche consiste à élaborer

⁹ GARRABÉ M (1992): Méthode d'évaluation de la pertinence *Revue Tiers monde* 129 Janvier-Mars (pp209-228).

¹⁰ B DENTE (1998) : La politique d'évaluation des politiques publiques : in *Evaluation des politiques publiques* . Paris L'harmattan.

successivement un compte économique (comprenant des éléments « cachés ») puis un compte coûts-avantages à partir du compte financier. Dans le cadre de cette méthode il est nécessaire de rectifier les prix des biens et services appartenant aux secteurs abrité et administré.

2-La méthode coûts-efficacité.

Cette méthode reprend les principes de la méthode précédente pour la mesure des coûts. Par contre il n'est pas nécessaire de mesurer les avantages, ceux-ci étant considérés comme donnés.

3-Les méthodes de valorisation.

Ces méthodes sont nombreuses. Leur principe repose sur le fait que les biens ou services consommés ou produits par un projet ne relèvent pas systématiquement d'une valorisation à partir du marché naturel. Il conviendra donc de recourir à d'autres types de marché. Les marchés de substitution, les marchés implicites, les marchés générés ou les marchés théoriques.

4-La modélisation microéconomique.

La plupart du temps on recourt à des modèles linéaires qui présentent de bonnes conditions d'estimation. Ils sont généralement de la forme: $Y = a^1X^1 + a^2X^2 + a^3X^3 + \dots$, Avec Y la variable expliquée, X^1, X^2, \dots les variables explicatives et a^1, a^2, \dots les coefficients à estimer.

Exemple :L'évaluation des effets de la politique concernant les zones d'action prioritaires (ZEP) par la direction de la prospective au ministère de l'éducation a utilisé cette méthode.

Deux échantillons de 2000 élèves entrant en CE2 et en sixième, ont été constitués, pour lesquels on a renseigné les variables suivantes: variables socio-démographiques pertinentes, niveau en français et en mathématiques, ainsi que l'appartenance à une zone ZEP. On a ensuite estimé un ensemble de relations entre les niveaux dans les disciplines et les différentes variables explicatives, chaque coefficient estimé représentant l'influence d'une variable, en supposant les autres variables fixées à leur niveau de référence¹¹.

Le résultat montre que l'appartenance à une ZEP influe de façon significative sur les niveaux en CE2, mais disparaît à l'entrée en sixième. De telle sorte qu'on peut en conclure que la politique conduite dans les ZEP est efficace.

La validité de la méthode économétrique, dépend naturellement de la capacité du modèle à rendre compte de la réalité. Plus la réalité, ou plus la politique mise en oeuvre seront complexes et moins la modélisation sera efficace. Le choix des variables explicatives et leur indépendance vis à vis de la variable expliquée sont surtout aisés, dans des cas relativement simple.

B-METHODES MESOECONOMIQUES VAT DE VALORISATION DES BIENS ET SERVICES MARCHANDS.

1-La méthode des effets.

Par rapport aux méthodes classiques microéconomiques de type coûts-avantages, la méthode des effets est une méthode méso-économique. Elle consiste en l'identification et la

¹¹ MONNIER 1986

mesure des différents types d'effets résultant de l'implantation et du fonctionnement d'un projet sur l'ensemble des branches de l'activité économique.

Tout accroissement de la demande à l'économie provoquée par la mise en oeuvre d'un projet, peut être simulé et permettre de mesurer les principaux effets que l'on peut en attendre. Ces effets pourront être de trois sortes; effets productifs (primaire directs et indirects), effets distributifs (secondaires) et effets d'importation.

L'accroissement de demande initiale se décompose en augmentation de productions locales et en accroissement d'importations. Celui-ci aura pour effet la diminution des ressources en devises et l'augmentation des taxes perçues. L'augmentation de la production locale se traduira par une création de valeurs ajoutées et un accroissement des consommations intermédiaires, qui pourront être produites localement ou importées.

2-La méthodes des multiplicateurs.

La méthode des multiplicateurs consiste à utiliser un système d'indices construit pour une zone déterminée, de taille variable, qui se substitue à un tableau inputs-outputs, mais l'information disponible autorise rarement un niveau infra régional. De plus en toute rigueur la substitution n'est pas totale. Il ne faut pas confondre les multiplicateurs régionaux issus par exemple de la théorie de la base économique avec les taux inclus de la méthode des effets.

3-L'analyse structurelle résiduelle¹².

Il s'agit d'une méthode dite « top-down » au sens où l'on projette une tendance observable au niveau d'un pays à une région de celui-ci. L'effet attendu d'une politique est déterminée en comparant la situation réelle de la région avec une construction de ce que la situation aurait pu être si elle avait subi uniquement les tendances nationales. Elle nécessite deux étapes successives : la construction de la situation contrefactuelle et la comparaison. La modélisation contrefactuelle est relativement simpliste, elle consiste par exemple concernant l'emploi, à appliquer à la structure des emplois régionaux les tendances de l'emploi national.

C-METHODES MACROECONOMIQUES.

Il s'agit ici des modèles macroéconomiques construits pour simuler les conséquences d'une opération d'investissement de grande envergure. Leur réalisme dépend de leur taille, mais il rendent rarement de bons résultats dans le cas des projets à effets régionalisés.

D-LES METHODES MULTICRITERES.

Il s'agit d'outils relativement complexes permettant de prendre en compte plusieurs critères simultanément pour mesurer les effets ou l'impacts d'une opération. Parmi les plus connues y a les méthodes de pondération, les méthodes Electre ou la méthode Qualiflex.

10.2-LES METHODES QUALITATIVES.

Les informations pourront aussi être qualitatives, et pourront résulter d'enquêtes auprès des décideurs, pour les objectifs par exemple. En certaines circonstances, une méthode DELPHI (enquête d'experts) pourra être utilisée pour obtenir une information non mobilisable. L'information qualitative pose la question de son organisation en niveaux, et donc la question des seuils de classification. Il existe un grand nombre de méthodes qualitatives.

¹² (Means3 p93)

1-La cartographie conceptuelle d'impacts¹³

La cartographie conceptuelle d'impacts est une technique d'aide à la « clarification » des impacts anticipés d'une action publique, obtenus à partir des représentations que se font les acteurs de cette action. Ces représentations sont formalisées par des descripteurs proposés aux acteurs. Il s'agit d'une modalité de la technique du « concept mapping » déjà utilisé dans certain cas. Des logiciels d'analyses de données permettent de donner une image graphique de ces représentations.

1-Le système d'information géographique (SIG)¹⁴

Il s'agit d'un outil d'information graphique qui permet de représenter une réalité complexe à partir par exemple de représentations cartographiques superposées associant des faits matériels et relationnels. La comparaison avant et après une opération d'investissement de systèmes cartographiques permettant de rendre compte des effets de cette opération. Le danger est que la situation avant n'est pas, nous le savons, une situation de référence.

10.3-LES METHODES DE NOTATION DES RESULTATS

Une évaluation quantitative, mais surtout qualitative peut être exprimée sous la forme d'un indice ou d'une note. L'élaboration de celle-ci nécessite le recours à des enquêtes décideurs et des enquêtes acteurs.

Le principe de la notation consiste à attribuer une valeur quantitative à une observation qualitative à partir d'une échelle choisie. La question est celle de l'arbitraire du système retenu. Lorsqu'il n'est question que de comparaison de notes à l'intérieur d'un même système, la notation peut permettre une meilleure lisibilité des résultats. Lorsque, par contre, la note est sortie de son référentiel, alors elle perd toute signification.

La matrice score présentée ici, est un exemple d'un système de notation. Il en existe d'autres¹⁵

1 -la méthode MAREVA.

¹³ (Means 3 p49)

¹⁴ (Means 3 p87)

¹⁵ Evaluation de la politique d'accueil de la petite enfance/CEP-Conseil général de l'Hérault 1998.