

L'évaluation des politiques publiques : enjeux, méthodes et institutions

Antoine Bozio *

16 août 2013

Résumé

Cet article présente les objectifs et les difficultés techniques et institutionnelles de l'évaluation des politiques publiques. Les enjeux de l'évaluation sont à la fois approfondissement de connaissances scientifiques, aide à la décision politique et contribution au débat démocratique. Loin de réserver les décisions politiques à un groupe d'experts restreint, l'évaluation est au contraire la condition d'appropriation des éléments du débat public par l'ensemble des citoyens. L'expertise doit se concentrer sur le choix des méthodes d'évaluation, diverses selon les données disponibles et la politique à évaluer. Une revue non technique des méthodes disponibles actuellement est présentée, contrastant leurs avantages et leurs limites. Enfin, après un panorama des institutions contribuant à l'évaluation des politiques publiques, les clés d'une réussite institutionnelle de l'évaluation sont discutées.

Mots clés : Méthodes d'évaluation ; politiques publiques

JEL : C5 ; C9 ; D04 ; H43

*Institut des politiques publiques (IPP) et Paris School of Economics (PSE). Article préparé pour le congrès des économistes belges de langue française (CEBLF) 2013.

Au vu des fortes contraintes budgétaires dans lesquelles s'inscrit aujourd'hui l'action publique, connaître l'impact des politiques mises en œuvre semble une évidence. Comment sinon améliorer l'efficacité de la dépense publique ?

Si la démarche d'évaluation a été pendant longtemps négligée, c'est avant tout parce que l'impact des différentes politiques semblait évident : les politiques d'emploi améliorent l'emploi, les politiques du logement améliorent l'accès au logement, etc. Dans un tel cadre les choix de politiques publiques consistent essentiellement à un arbitrage budgétaire entre les différentes dépenses possibles. En réalité, les modalités de l'intervention publique sont rarement évidentes : il existe souvent plusieurs politiques envisageables pour atteindre un même objectif ; les dispositifs mis en place peuvent avoir des effets indésirables non attendus, voire pas du tout l'effet attendu ; plus généralement, les mécanismes économiques sur lesquels les politiques publiques essaient d'intervenir sont complexes, souvent encore mal compris. Dès lors qu'il devient clair que l'effet des politiques publiques n'est pas chose évidente, la nécessité de leur évaluation apparaît alors beaucoup plus nettement : l'objectif est de permettre la compréhension des mécanismes sur lesquels l'action publique cherche à peser et d'identifier les interventions les plus efficaces au vu des objectifs souhaités (Bozio et Grenet 2010).

Une fois que l'intérêt de l'évaluation est admis, il reste à comprendre pourquoi évaluer une politique publique n'a rien de plus évident. L'objectif n'est pas de suivre au fil de l'eau l'évolution des bénéficiaires apparents d'une politique et de mesurer leur nombre : l'évaluation a pour but de mesurer « ce qui ne se serait pas passé en l'absence de la politique ». La difficulté que doivent surmonter toutes les méthodes d'évaluation est la construction statistique de ce contrefactuel, qui seul permet d'établir l'impact causal d'une politique.

La démarche d'évaluation repose certes sur une expertise permettant le choix et la bonne réalisation des méthodes d'évaluation, mais la réussite technique de l'évaluation ne garantit pas pour autant son utilité sociale. Le but final de l'évaluation est de pouvoir éclairer le débat public sur les principaux arbitrages en jeu et ainsi faciliter les choix démocratiques. Cette dimension de l'évaluation implique que les institutions en charge de l'évaluation soient crédibles auprès du grand public, donc indépendantes, mais aussi capables de communiquer des résultats sous une forme non technique, qui permette réellement d'alimenter le débat démocratique.

Après avoir défini ce qu'on entend par « évaluation » (partie 1), nous présenterons les différentes méthodes d'évaluation, en discutant leur pertinence selon le contexte et les données disponibles (partie 2). La partie 3 présente un cas pratique avec l'évaluation du programme *Pathways to work* au Royaume-Uni, qui a mobilisé plusieurs méthodes d'évaluation. Nous ferons ensuite un rapide panorama institutionnel de l'évaluation dans plusieurs pays européens, avant de discuter les modalités pour que l'évaluation soit une réussite démocratique et pas simplement une réussite technique (partie 4).

1 Qu'est-ce que l'évaluation des politiques publiques ?

Le mot « évaluation » est un mot-valise qui est utilisé pour décrire des approches foncièrement différentes. Avant de décrire ce qu'est – ou devrait être – l'évaluation des politiques publiques, il est utile de décrire ce qu'elle n'est pas.

1.1 Ce que n'est pas l'évaluation des politiques publiques

Le mot évaluation est souvent compris dans le langage commun comme une évaluation individuelle (évaluation d'un devoir scolaire, évaluation du travail de salariés, etc.) ou l'évaluation d'une entreprise par les consommateurs (questionnaires de satisfactions). Cette connotation du mot pour signifier un jugement de valeur sur une activité quelconque a souvent été la cause d'incompréhensions. Cela explique, par exemple, la réticence des responsables politiques ou des administrations à l'évaluation des politiques publiques, partageant une crainte de retombées négatives en cas d'évaluations négatives. Cela a aussi entraîné des critiques de la culture de l'évaluation (del Rey, A. 2013), perçue comme la domination d'une expertise bureaucratique visant à remplacer les choix démocratiques. C'est tout le contraire : l'évaluation des politiques publiques n'est pas un substitut aux choix politiques ; elle ne dit pas « ce qu'il faut faire », mais elle permet de comparer l'impact de plusieurs *instruments* de politiques publiques au vu des *objectifs* variés que les citoyens peuvent chercher à poursuivre.

Dans le domaine des politiques publiques, le mot est le plus souvent utilisé pour décrire du contrôle de gestion ou une analyse du management public : les rapports des corps d'inspection cherchent ainsi à vérifier que la dépense publique a été faite efficacement et en respectant des règles de bon fonctionnement. Il s'agit essentiellement d'une évaluation du processus de mise en place des politiques publiques : vérifier que la mise en œuvre pratique d'un dispositif correspond bien au cahier des charges défini par ses concepteurs, information des bénéficiaires potentiels, formation des agents chargés de la mettre en œuvre, etc. Pour prendre un exemple sur le cas français, on peut ainsi mesurer le nombre de fonctionnaires nécessaires pour traiter d'un certain nombre de dossiers du Revenu de solidarité active (RSA)¹ et étudier quelle organisation du travail ou quels moyens matériels permettrait d'améliorer la mise en place de cette politique. Dans le même ordre d'idée, on pourrait étudier si les bénéficiaires de la politique visée sont bien touchés (par exemple si la population qui doit bénéficier du RSA reçoit bien l'allocation), mesurer leur nombre, voire récolter leur sentiment sur leur interaction avec l'administration autour de ce dispositif.

Cette approche de la performance des administrations publiques donne lieu à la construction d'indicateurs censés évaluer l'efficacité de la dépense publique. Dans le cas français, par exemple, la loi organique relative aux lois de finances (LOLF) a mis en place un processus

1. Revenu de solidarité minimum qui a remplacé le revenu minimal d'insertion en incluant une partie variable en fonction de la rémunération d'activité.

systématique d'analyse de la performance des administrations publiques et donc la publication d'une multiplicité d'indicateurs chiffrés. Cette approche a souvent été assimilée à de « l'évaluation » et a conduit à de nombreuses critiques d'une « politique du chiffre » où les indicateurs chiffrés deviennent un objectif en soi des politiques publiques (Chauvière, M. 2012). Souvent cette approche de l'évaluation comme du contrôle de gestion a d'ailleurs pour but principal d'identifier des « économies » et donc de permettre une réduction des dépenses publiques. C'est assez clairement l'objectif assigné, en France, à la Révision générale des politiques publiques (RGPP) ou aujourd'hui la Modernisation de l'action publique (MAP).

1.2 Définir l'évaluation

Toutes ces démarches sont certainement nécessaires, mais elles sont bien différentes de ce que les universitaires appellent l'évaluation des politiques publiques, c'est-à-dire l'évaluation d'impact. L'objet est alors de mesurer l'impact d'une politique ou d'un dispositif, sur de multiples critères, au vu des objectifs qui lui ont été assignés. L'objectif de l'évaluation est avant tout un objectif scientifique : il s'agit d'approfondir nos connaissances sur la façon dont les dispositifs institutionnels existants (les instruments des politiques publiques) atteignent – ou n'atteignent pas – leurs objectifs initiaux. En sus de vouloir mesurer l'impact des politiques, il s'agit d'en comprendre les mécanismes afin d'élaborer d'autres dispositifs qui permettent plus directement de répondre à l'objectif initial.

Mesurer l'impact d'un dispositif ou d'une politique publique n'a pour autant rien d'évident. Par impact on entend un *impact causal* du dispositif sur l'ensemble des facteurs d'intérêt, ceux qui sont initialement visés par la politique bien sûr, mais aussi l'ensemble des éléments qui peuvent être malencontreusement touchés par une politique sans le vouloir. Pour mesurer l'impact causal d'une politique, il faudrait idéalement pouvoir comparer le même monde avec et sans cette politique. C'est ce que les économètres appellent généralement le problème fondamental de l'évaluation, l'impossibilité d'observer en même temps, dans le même lieu, les mêmes conditions macroéconomiques, pour les mêmes individus, ce qui se passerait avec ou sans une politique publique particulière. Pour remédier à ce problème, toutes les méthodes de l'évaluation proposent des solutions statistiques pour comparer l'impact d'un dispositif à un *contrefactuel*, un monde reconstruit en l'absence de politique. Le choix de la méthode idéale dépend alors du contexte, des données disponibles et des hypothèses nécessaires pour rendre crédible la construction de ce contrefactuel.

Les pièges de l'évaluation sont nombreux et rendent l'exercice en réalité très difficile (Conseil d'analyse économique 2013). La difficulté la plus évidente est la causalité inverse : si par exemple on réalise une corrélation entre dépenses de soins et état de santé, on obtiendra une corrélation négative, due au fait que les individus en mauvaise santé consomment plus de soins. En conclure que les soins de santé ont un impact négatif sur la santé est évidemment faux. Souvent les politiques publiques visent justement un public sélectionné sur la base de

l'indicateur que l'on cherche à améliorer : les politiques de l'emploi peuvent viser des chômeurs de longue durée, les politiques d'éducation des élèves en difficulté scolaire, etc. Ne pas prendre en compte cet *effet de sélection*, rendra caduque toute démarche d'évaluation.

Un autre piège tient à l'effet de l'incidence : les bénéficiaires d'une politique ne sont pas forcément ceux qui étaient visés à l'origine par la politique. Ce cas se matérialise le plus souvent lorsqu'un prix de marché est en jeu et peut être influencé par la politique elle-même. Par exemple, si l'on met en place des allocations logement, visant à améliorer l'accès au logement de ménages à faibles ressources, on ne peut pas faire l'hypothèse que les loyers ne vont pas réagir à la mise en place de cette politique. Si les loyers réagissent à la hausse, cela veut dire que les propriétaires des logements ont capturé une partie de la dépense publique consacrée aux allocations logement². Ce type d'effet peut avoir lieu pour les impôts, les prestations sociales ou même certaines dépenses publiques. Le mécanisme d'incidence est un cas particulier d'effets d'équilibre général, qui peuvent modifier substantiellement l'effet d'une politique publique, à court terme par rapport à moyen terme. Si la politique modifie l'offre et la demande sur un marché, l'effet en retour peut être négatif : si, par exemple, on subventionne les ménages pour acheter leur logement, et si l'offre est faiblement élastique (peu de marges pour construire de nouveaux logements), alors la subvention va simplement augmenter la demande et se traduira par une augmentation des prix, qui limitera en retour la demande.

Les méthodes de l'évaluation doivent toutes essayer d'éviter ces pièges et de proposer dans chaque cas, une estimation de l'impact causal de la politique étudiée.

2 Les méthodes de l'évaluation

Les méthodes de l'évaluation d'impact sont nombreuses et plusieurs articles ou livres de qualité sont aujourd'hui disponibles pour apprendre à les mettre en place (Angrist et Pischke 2008, Blundell et Costa Dias 2009, Givord 2010, Imbens et Wooldridge 2009). Cette partie se contente d'une présentation succincte et non technique afin d'insister sur les avantages et limites de chaque approche. Une bonne évaluation reposera avant tout sur le bon usage de la panoplie de techniques disponibles, au vu des contraintes spécifiques à chaque cas.

Au sein de l'évaluation d'impact, on distingue généralement l'évaluation *ex ante*, de l'évaluation *ex post*. L'évaluation *ex ante* est réalisée avant l'introduction d'une politique et consiste à analyser ses effets potentiels ; l'évaluation *ex post* vise à mesurer son impact réel après son entrée en vigueur. Dans les deux cas, l'efficacité de la politique étudiée est mesurée en comparant ses coûts et ses bénéfices.

2. Cet effet a été mis en évidence par Fack (2006) dans le cas français.

2.1 L'évaluation d'impact *ex ante*

Une évaluation *ex ante* consiste à modéliser l'introduction d'une politique sur la population cible. Il s'agit d'estimer l'impact attendu au vu des connaissances actuelles. Par définition, l'évaluation *ex ante* suppose de connaître « l'impact théorique » de la mesure à évaluer.

L'intérêt premier des évaluations d'impact est de pouvoir modéliser la modification d'une législation complexe, par exemple la fiscalité ou le système de retraite. Il s'agit en premier lieu d'identifier la population touchée par la modification en question. Pour y parvenir, on utilise généralement des **modèles de micro-simulation** (Blanchet 2011). Cette forme d'évaluation repose sur une représentation plus ou moins simplifiée de l'univers économique au sein duquel évoluent les agents, qui est utilisée pour simuler, à partir d'un échantillon représentatif des agents économiques concernés (individus, ménages, entreprises, etc.), l'effet de la politique publique envisagée. Par exemple, si l'on souhaite évaluer l'impact d'une réforme des retraites consistant en une augmentation de la durée requise de cotisation (ou d'un autre paramètre du système), il est indispensable d'avoir recours à ce type de simulation. En raison de la complexité du système et des données sous-jacentes (distribution des carrières des individus), il est souvent impossible, même à des experts, de prédire quel sera l'effet d'une modification des barèmes de retraite (qui va être touché, quel effet sur l'équilibre budgétaire) sans avoir recours à une simulation détaillée de la réforme. Le modèle de micro-simulation spécifie plusieurs scénarios sans réforme (en fonction des conditions macro-économiques ou démographiques), puis simule la législation actuelle du système de retraite. Cela permet en première instance de donner des indications sur les déséquilibres budgétaires ou le niveau des pensions à plus ou moins long terme. Il est ensuite possible de simuler une réforme et de mesurer qui va être touché (au niveau individuel), et quelles en seront les conséquences budgétaires. Pour rendre l'exercice de micro-simulation plus réaliste, il est nécessaire de prendre en compte l'effet de la mesure évaluée sur les comportements des individus. Cela suppose de pouvoir « calibrer » les réactions des agents à partir d'estimations empiriques des principaux paramètres du modèle – dans notre exemple, il s'agit des comportements de départ en retraite. La modélisation des comportements dans ces modèles peut venir de règles simples ou d'estimations qui sont le plus souvent fournies par des études empiriques *ex post* réalisées antérieurement. A la limite, cette approche pourrait permettre de modéliser tous les effets en retour d'une politique, y-compris sur l'équilibre macroéconomique. Dans la pratique, la difficulté de prendre en compte des effets du second ou troisième ordre et l'incertitude croissante autour des résultats obligent à une certaine modestie et conduisent généralement à présenter plusieurs variantes de chaque simulation (Adam et Bozio 2009).

Dans l'esprit des modèles de microsimulation, on trouve les développements de **l'économétrie structurelle** (Meghir, C. 2008). Il s'agit d'une approche qui repose sur la formalisation de modèles économiques, spécifiant des paramètres « structurels » de la population et supposés invariants. L'estimation de ces modèles est réalisée sur des données variées, en coupe ou en

panel, et l'identification du modèle repose idéalement sur une réforme bien identifiée. Si le modèle est bien spécifié – chose particulièrement difficile à obtenir – cela signifie qu'il est possible de modifier à volonté la forme de l'intervention simulée pour estimer le type d'impact qu'elle est susceptible d'entraîner. Cette approche, la plus exigeante, permet théoriquement de tester de multiples variantes d'interventions mais sa limite évidente est la capacité à modéliser finement l'ensemble des comportements et interactions possibles.

2.2 L'évaluation d'impact *ex post*

Toutes les techniques de l'évaluation *ex post* consistent à construire statistiquement un *contrefactuel*, c'est-à-dire un scénario aussi proche que possible de ce qui se serait passé si la politique étudiée n'avait pas été mise en place. La mesure de l'impact d'une politique publique compare ensuite ce scénario « sans la politique », avec ce qu'il est possible d'observer quand le dispositif évalué a été effectivement mis en place.

La première technique à présenter est **l'expérimentation aléatoire**. Elle est directement inspirée des méthodes médicales pour évaluer l'effet d'un traitement : parmi les potentiels bénéficiaires d'une politique, on tire au sort qui recevra la politique (le groupe traité) et qui n'en bénéficiera pas (le groupe de contrôle). En mesurant dans les deux groupes les critères d'intérêt – qui doivent correspondre à un objectif de la politique – et en faisant la différence entre les deux groupes, on obtient une mesure causale de l'effet du dispositif ainsi expérimenté. Certains dispositifs, reposant fortement sur une auto-sélection des individus, ne peuvent être évalués de façon robuste que par cette méthode. Enfin, l'avantage de cette approche est de proposer une méthode très convaincante pour un public non spécialiste : elle ne repose pas sur des techniques économétriques élaborées et peut être comprise de façon intuitive. Cela permet d'apporter des éléments dans le débat public qui ne pourront pas être réfutés facilement et qui obligeront les parties prenantes à revoir leurs *a priori*. Cette méthode, la plus robuste en termes de protocole, a néanmoins certaines limites : on mesure précisément un effet de la politique dans le contexte et dans le pays où l'expérience a été menée. Pour généraliser la validité des résultats, il faut expérimenter à nouveau dans un autre contexte. Par ailleurs, comme pour les expérimentations médicales, cette méthode nécessite l'acceptation des participants et une procédure éthique stricte de la part des évaluateurs. De façon naturelle, cette approche est plus appropriée dans le cas de dispositifs restreints, limités localement, souvent en phase exploratoire.

Un deuxième groupe de techniques consiste à reproduire statistiquement une expérience aléatoire, même si elle n'a pas été mise en place avec l'introduction de la politique. On appelle ces techniques **expériences naturelles** car ce sont des expériences que les chercheurs trouvent « dans la nature ». Souvent, quand une politique est mise en place, elle se restreint à un groupe de bénéficiaires pour des raisons administratives. Ces règles de mise en place créent naturellement des groupes de contrôle, c'est-à-dire des populations dont les caractéristiques

sont très proches de la population cible mais qui ne bénéficient pas de la politique. Au sein de ce groupe de méthodes, on distingue la **double différence** (en anglais *difference in difference* ou DiD) : le principe est simple, il s'agit de comparer deux groupes d'individus, l'un dit de traitement (touché par le dispositif qu'on cherche à évaluer) et l'autre dit de contrôle qui n'est pas touché. Il n'y a pas besoin que ces deux groupes soient créés par méthode aléatoire ni qu'ils soient strictement identiques. Pour que la méthode soit valide, il suffit « qu'en l'absence de la politique les deux groupes suivent la même tendance ». La méthode consiste alors à faire la différence entre la situation après et avant la réforme dans chacun des groupes et d'en faire la différence (d'où le nom de différence en différence). Comme les groupes ne sont pas créés de façon aléatoire, la robustesse de la méthode dépend de l'hypothèse de tendance commune entre les deux groupes. C'est cette hypothèse que les évaluateurs doivent vérifier, soit en faisant des tests placebo avant la mise en place de la réforme, soit sur d'autres groupes de contrôle. Visualiser la différence en différence par un graphique représentant les deux groupes permet souvent d'offrir une présentation intuitive de l'hypothèse identifiante.

Une autre méthode existe, toujours dans le cadre des expériences naturelles, qui s'appelle **la régression autour d'une discontinuité** (en anglais *regression discontinuity design*, alias RDD). L'idée consiste à utiliser une limite légale, un plafond ou une règle administrative arbitraire, qui crée incidemment deux groupes d'individus très proches, ceux qui sont juste en-dessus et ceux qui sont juste en-dessous de la discontinuité. L'hypothèse identifiante est qu'aux alentours de la discontinuité les individus sont similaires sauf pour l'accès au dispositif étudié. Par exemple, un dispositif qui vise à favoriser l'emploi des jeunes n'est ouvert qu'aux 18-25 ans : on pourra comparer les jeunes de 26 ans moins un mois et ceux de 26 ans et un mois. Ils sont certainement très proches en termes d'accès au marché du travail, d'expérience, mais les uns auront bénéficié de la politique d'emploi quand les autres non. Pour que cette méthode fonctionne bien il faut avoir suffisamment d'observations dans la base de données pour se restreindre aux individus juste en-dessus ou en-dessous de la discontinuité : plus on s'éloigne de cette coupure arbitraire, plus les individus ont des chances d'être différents. Plus encore que la double différence, cette méthode donne un résultat local (valide autour de la discontinuité). Parfois la discontinuité n'est pas très nette et la méthode peut donner lieu à discussion. Là encore, une présentation graphique des caractéristiques des individus autour de la discontinuité est la façon la plus convaincante de mettre en évidence la stratégie retenue.

Une dernière méthode d'évaluation *ex post* mérite d'être présentée ici. Il s'agit des **techniques de matching**. Le principe est de construire un groupe de contrôle et un groupe de traitement à partir des caractéristiques observables. Pour chaque individu dans le groupe de contrôle, on va sélectionner un individu identique (pour ses caractéristiques connues) qui n'a pas été touché par le dispositif évalué : même âge, même sexe, même éducation, même salaire, etc. Plus les informations seront complètes, plus les chances de construction de groupes similaires vont être élevées. On peut aussi combiner cette technique avec la double différence

pour renforcer la validité de l'estimation. Le problème avec cette approche vient des variables inobservables (soit absentes des données, soit impossible à observer) : si des caractéristiques inobservables peuvent expliquer si un individu sera touché ou non par le dispositif, la méthode de matching risque d'être inopérante. Par exemple, si les individus doivent être volontaires pour participer à un programme, il y a de fortes chances que deux personnes apparemment identiques, mais dont l'une décide de participer et l'autre non, soient en fait différentes (par leur motivation par exemple).

2.3 L'analyse économique

Il ne suffit pas de mesurer l'impact d'une politique publique pour savoir si celle-ci est efficace ou non, au vu des objectifs assignés initialement. En effet, une politique ne peut être considérée comme efficace que si ses bénéfices, pour la société, l'emportent sur ses coûts.

L'analyse coût-bénéfice consiste à attribuer une valeur monétaire à ces deux composantes, afin de calculer la valeur nette totale de la politique considérée. Ce type d'analyse est souvent utilisé comme outil de décision *ex ante* mais peut également servir comme cadre d'analyse pour l'évaluation *ex post*. La valeur actualisée des bénéfices nets d'une politique publique est plus complexe à calculer que la valeur actualisée d'un investissement privé. En effet, alors que les décisions d'investissements peuvent être prises en considérant le coût de marché des facteurs de production et la rentabilité prévisible d'un investissement, l'évaluation des coûts et des bénéfices d'une politique publique peut rarement s'appuyer sur des prix de marché. Dans certains cas, ces prix existent mais ne reflètent pas les coûts et les bénéfices sociaux parce qu'ils ne prennent pas en compte d'éventuelles externalités, sont contaminés par des asymétries d'informations ou sont régulés par l'État. Dans d'autres cas, ces prix n'existent tout simplement pas : le coût social des dommages environnementaux causés par la pollution est par exemple difficile à évaluer en l'absence d'un marché mesurant la valeur d'un environnement préservé. Malgré ces limites, la quantification, avec une métrique similaire, des bénéfices et des gains des politiques publiques est un exercice indispensable lorsqu'on souhaite comparer les mérites respectifs de plusieurs interventions possibles. Il y a plusieurs difficultés pour passer d'une évaluation *ex post* à une analyse coût-bénéfice : d'abord il faut obtenir une mesure complète des coûts (coûts directs, coûts administratifs, coûts pour les bénéficiaires, coût du prélèvement fiscal nécessaire au financement, etc.) ; ensuite il faut avoir une idée de la durée de l'impact du dispositif – si l'impact est de courte ou longue durée, l'analyse coût-bénéfice sera foncièrement changée ; enfin il faut pouvoir estimer des coûts et bénéfices non monétaires (stigmatisation des bénéficiaires, démobilisation des familles, etc.) afin d'effectuer un bilan complet d'une intervention publique.

L'analyse coût-bénéfice permet de comparer un flux de bénéfices aux coûts d'une politique publique, mais il est souvent nécessaire d'aller plus loin pour juger de l'optimalité d'une intervention publique : prendre en compte la redistribution opérée, l'amélioration du bien-être

global, la cohérence de l'intervention avec d'autres politiques déjà en place, etc. L'analyse économique classique, souvent appelée *optimal design*, cherche à offrir des arguments explicites pour les choix sociaux au vu des choix collectifs de redistribution et de la variété de structuration des politiques publiques. Développés particulièrement dans le cadre de l'analyse des politiques fiscales et de prestations sociales, ces méthodes cherchent à maximiser le bien-être collectif, en prenant en compte un poids plus fort porté sur les individus les plus fragiles et les moins favorisés. Cette approche demande finalement beaucoup à l'analyse économique – en théorie il faudrait avoir répondu à toutes les questions théoriques et empiriques de la discipline – mais elle permet de rappeler que l'analyse des politiques publiques ne se limite pas à un calcul d'efficacité ou de performance monétaire : l'objectif ultime est l'amélioration du bien-être collectif.

3 Un exemple pratique : l'évaluation de *Pathways to work* au Royaume-Uni

Cette partie présente de façon succincte une évaluation d'un dispositif mis en place au Royaume-Uni. La diversité des méthodes utilisées, des données mobilisées et la diversité des enseignements possibles pour d'autres évaluations en font un bon cas pratique.

3.1 Le dispositif *Pathways to work*

Le programme *Pathways to work* est un dispositif d'aide au retour à l'emploi visant la population bénéficiaire de l'allocation pour inaptitude/invalidité au Royaume-Uni (*Incapacity Benefit, IB*). Avec la baisse du chômage dans les années 1990, la population bénéficiaire de ce type d'allocation est devenue plus importante en nombre que les chômeurs. Le gouvernement du *New Labour*, élu en 1997, avait pris l'engagement de favoriser le retour sur le marché du travail d'un million de bénéficiaires de l'allocation IB avant 2010. Le programme *Pathways to work* (en français, « les chemins vers l'emploi ») était le dispositif principal, mis en place à partir de 2003, pour atteindre cet objectif ambitieux.

Pathways to work consiste en trois dispositifs séparés : i) une incitation financière, permettant aux allocataires de garder pendant 12 mois 50 % du montant de l'allocation ; ii) des entretiens obligatoires pour aider à la recherche d'emplois non pénibles ; iii) des dispositifs particuliers pour améliorer le lieu de travail ou pour mieux gérer les problèmes de santé diagnostiqués, dispositifs uniquement disponibles sur la base du volontariat (appelés *Choices*). L'ensemble représente un coût certain pour les finances publiques, en particulier les dispositifs de *Choices* qui impliquent des équipes médicales et un traitement intensif.

Le programme n'est pas ouvert à tous les allocataires. Deux groupes d'allocataires sont exclus du programme : ceux qui sont jugés médicalement trop handicapés pour espérer retourner sur le marché de l'emploi ; et ceux qui ont une incapacité au travail temporaire et

qui sont jugés aptes à retrouver un emploi sans aide particulière. Le groupe restant, jugés inaptes au travail de façon temporaire mais ayant besoin d'aide pour retourner sur le marché de l'emploi, sont enrôlés dans le programme *Pathways to work*.

3.2 L'évaluation

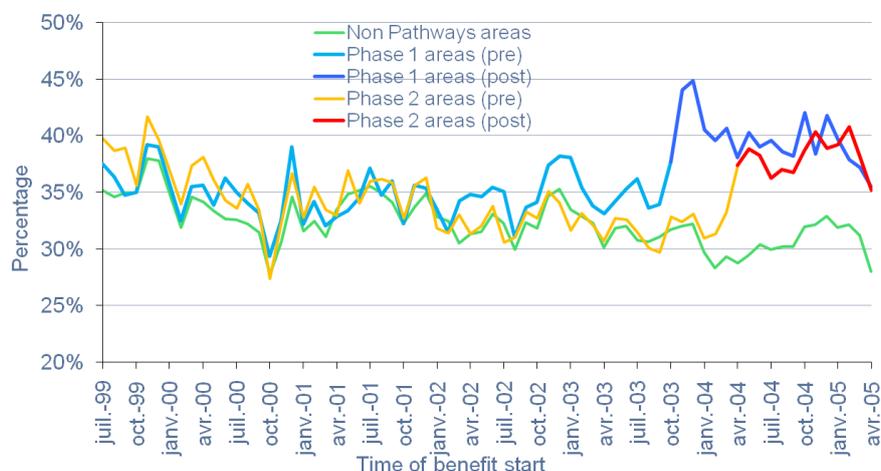
L'évaluation du programme a été réalisée par un consortium d'évaluateurs indépendants, dont l'*Institute for Fiscal Studies* (IFS), le *Policy Studies Institute* (PSI), *NatCen*, etc. L'évaluation a consisté en une multitude d'études, soit qualitative (sur le ressenti des bénéficiaires, des personnels, etc.), soit quantitative (analyse d'impact, analyse coût-bénéfice, etc.) sur une variété de critères. L'ensemble a été financé directement par le ministère du travail britannique, le *Department for Work and Pensions* (DWP).

Le principe de base de l'évaluation *ex post* repose sur une expérimentation locale, sans allocation aléatoire des bénéficiaires. Le dispositif a été mis en place progressivement dans le pays, en commençant par certains bassins de l'agence pour l'emploi (*JobCentre Plus*). Ainsi, en octobre 2003, trois localités voient la première expérimentation du dispositif, puis en avril 2004 quatre nouvelles localités sont ajoutées, et ainsi de suite jusqu'à la généralisation du dispositif en décembre 2008. Le consortium d'évaluateurs a choisi les localités témoins, c'est-à-dire des bassins de *JobCentre Plus* proches des groupes tests, similaires en termes de taux de retour à l'emploi, de taux de chômage ou d'autres caractéristiques. Des données sont collectées avant la mise en place de la politique, dans les zones de traitement et de contrôle, puis après 6 mois de mise en place, 12 mois et finalement 24 mois. Des informations sont collectées sur l'état de santé, les qualifications, l'expérience sur le marché du travail, la situation familiale, etc. Les données administratives sur le suivi des allocataires sont aussi mobilisées.

Les méthodes utilisées pour l'évaluation d'impact ont été variées. La première a consisté à effectuer une double différence en comparant les localités touchées avec les groupes de contrôle avant et après l'introduction du dispositif. Comme le programme a été étendu progressivement au pays, cela donne lieu à plusieurs occasions d'estimer une double différence, à chaque fois que des localités supplémentaires sont ajoutées. Les variables d'intérêt ont été à la fois le taux de sortie de l'allocation, le taux de retour à l'emploi, le niveau du salaire et l'état de santé déclaré. Le graphique 1 présente l'évolution du taux de sortie de l'allocation pour inaptitude dans trois zones : la zone touchée en octobre 2003, la zone touchée en avril 2004 et une zone jamais touchée par le dispositif pendant toute la période. On observe visuellement l'augmentation du taux de sortie à six mois au moment de l'introduction du dispositif. L'effet tend à baisser progressivement mais reste substantiel. Ces données descriptives ne contrôlent pour aucun élément et ne sont donc pas le coeur de l'évaluation, mais elles permettent de se rendre compte visuellement de la pertinence de la stratégie d'identification.

Les résultats de l'évaluation d'impact du programme (Bewlet et al. 2007, Adam et al. 2008) conduisent à conclure que le programme a entraîné une forte hausse du taux de sortie

FIGURE 1 – Taux sortie à 6 mois de l'allocation pour inaptitude.



SOURCE : Adam et al. (2008)

à 6 mois de l'allocation pour inaptitude mais aucun effet après 12 mois, qui se traduit par une hausse, elle durable, du taux de retour à l'emploi, y-compris au bout de 24 mois. Cela veut dire que les retours en emploi viennent essentiellement de personnes qui, en l'absence du programme, serait bien sortie de l'allocation IB mais ne serait pas forcément retourné sur un emploi durable. L'effet net sur l'emploi est de l'ordre de 7 à 9 points de pourcentage pour une base de 24 %. C'est un effet très net du programme, d'autant plus si cet effet dure au-delà des premiers 24 mois observés par les enquêtes. Aucun effet n'est détecté sur l'état de santé déclaré et aucun effet non plus sur le montant des salaires obtenus.

Une deuxième approche a consisté à évaluer le dispositif particulier de *Choices*. La méthodologie utilisée a reposé, en l'absence de traitement aléatoire, sur les techniques de matching. Des données administratives et d'enquêtes très complètes ont été utilisées, permettant de retrouver un historique du marché du travail des individus remontant à trois ans et des informations détaillées sur leur situation médicale. La qualité des données permettait d'espérer que la procédure d'appariement statistique puisse permettre une identification causale du dispositif. Le problème restait qu'il fallait faire l'hypothèse que les individus, semblables en termes de caractéristiques observables, qui avaient choisi ou non de participer à ces dispositifs de *Choices*, étaient similaires. L'hypothèse était de fait très forte. Au terme de l'analyse, les évaluateurs (Adam, S., Bozio, A. et C. Emmerson 2009) ont préféré refuser de tirer une interprétation causale des résultats en soulignant qu'il était difficile de ne pas exclure une auto-sélection des individus dans les dispositifs.

Une troisième approche de l'évaluation a été la réalisation d'une analyse coût-bénéfice du programme, sur la base de l'évaluation d'impact présentée plus haut (Adam et al. 2008). Avec l'aide d'un modèle de micro-simulation, et d'une étude précise des coûts engendrés par le

programme, il a été possible de simuler l'impact du dispositif sur une population représentative de la population d'allocataires de IB, de calculer la modification des impôts payés, allocations non-versées et d'estimer les gains financiers pour l'État, les allocataires et la société dans son ensemble. La principale difficulté de l'exercice a été de faire une hypothèse sur la durée de l'effet positif sur l'emploi mesuré : plus l'effet est supposé durer au-delà de la dernière enquête, plus l'analyse coût-bénéfice est naturellement favorable.

3.3 Quels enseignements en tirer ?

Il est possible de tirer plusieurs enseignements pour l'évaluation d'une politique publique.

Le premier est que cette évaluation, de qualité, a modifié la perception dans le débat public du programme proposé. Critiqué par l'opposition, le dispositif a été de fait conservé par la nouvelle coalition au pouvoir depuis 2010 (sous un autre nom) et même généralisé. Pour autant, il est difficile de donner trop de poids à l'évaluation en soi car la généralisation de la politique a été annoncée avant que les résultats de l'évaluation soient publiés.

Le problème principal de cette grande évaluation est que les mécanismes de l'impact sont encore peu compris. Le protocole d'évaluation a permis de mesurer l'impact global du programme (aide financière, entretiens obligatoires, et aide médicale spécifique) mais les composants du dispositif n'ont pas été évalués séparément. On ne sait pas au final quelle partie de l'intervention a été la plus efficace, s'il serait possible de renforcer certains aspects ou d'en laisser d'autres de côté.

L'incapacité à évaluer de façon robuste la composante *Choices* indique qu'il est parfois indispensable de recourir à l'expérimentation aléatoire si l'on veut se donner une chance de mesurer l'impact de dispositifs où l'auto-sélection va être particulièrement forte.

Enfin, ce cas pratique renforce la conviction que le recours à des évaluateurs indépendants, soucieux avant tout de la robustesse scientifique des résultats, est une composante essentielle de la réussite d'une évaluation, ce qui invite à discuter les conditions institutionnelles, et pas seulement technique, pour réaliser des évaluations utiles à la société.

4 Les institutions de l'évaluation

Les institutions de l'évaluation sont loin d'être un sujet annexe pour la réussite du processus d'évaluation, en particulier si on juge l'évaluation à sa capacité à influencer le débat démocratique et *in fine* les choix de politiques publiques mis en œuvre.

4.1 Les grands modèles institutionnels de l'évaluation

On peut distinguer deux grandes approches institutionnelles de l'évaluation des politiques publiques.

La première conçoit l'évaluation comme une expertise au service du « Prince », destinée à aider la prise de décision publique face à la complexité du monde économique. L'évaluation est alors réalisée prioritairement par l'administration ou par des conseils d'experts directement rattachés au gouvernement. Dans ce cadre là, l'évaluation est souvent de courte durée – quelques mois – et répond à une demande pressante des décideurs. Ces analyses sont rarement publiées et ne font donc pas partie du débat démocratique. Ce modèle est assez bien représenté par la France, où l'État et les administrations publiques jouent un rôle prépondérant dans les instances d'évaluation. Les grands corps d'inspection de l'État (Inspection général des finances, Inspection générale des affaires sociales, l'Inspection générale de l'administration) ont officiellement une mission d'évaluation des politiques publiques. Ces rapports d'évaluation s'apparentent avant tout à l'évaluation de processus ou au contrôle de gestion des administrations publiques. L'objectif est d'améliorer la mise en place de l'action publique dans des délais rapides et par conséquent les évaluations d'impact y sont quasi-absents. Dans le même ordre d'idée, les analyses budgétaires réalisées en préparation du projet de loi de finances (le budget français) ne sont pas présentées au public, ni même au Parlement : le gouvernement présente les chiffres issus de ses arbitrages sans offrir la possibilité à la représentation nationale de les discuter.

La seconde approche conçoit l'évaluation comme une composante essentielle du processus démocratique : l'expertise se destine alors d'abord aux citoyens ou à leurs représentants, les parlementaires, et vise à faciliter le débat public en clarifiant les principaux arbitrages en jeu. Ainsi conçue, l'évaluation peut s'étaler sur un délais long – plusieurs années – qui dépasse largement l'horizon de l'agenda politique. Les évaluations ne deviennent utiles que si elles sont publiées et si un processus institutionnel (indépendance, discussions, etc.) lui a donné une certaine crédibilité. Les États-Unis se rattachent en partie à ce modèle, dans la mesure où une grande partie de l'évaluation des politiques publiques est réalisée par des institutions liées au Congrès : le *Congressional Budget Office* (CBO) a par exemple un accès direct aux informations de l'administration et publie des analyses déterminantes pour le vote du budget ; le *Government Accountability Office* (GAO), au départ une chambre d'audit gouvernemental, s'est mis à commissioner de plus en plus d'évaluations rigoureuses avec l'aide d'universitaires spécialisés. Le poids politique du Congrès américain donne à ces institutions une assise financière et institutionnelle très forte. Le cas britannique offre aussi un exemple de ce second modèle avec une place importante de l'expertise issue du monde universitaire. Des instituts de recherche indépendants, comme par exemple l'*Institute for Fiscal Studies* (IFS)³, jouent ainsi un rôle primordial au Royaume-Uni, non seulement dans l'évaluation des politiques publiques, dans les débats budgétaires, mais aussi dans la communication au grand public de l'état des connaissances sur les différentes options de politiques publiques. Dans le même esprit, le *Centraal Planbureau* aux Pays-Bas, une institution publique, joue un rôle

3. On peut aussi citer le *National Institute of Economic and Social Research* (NIESR), le *King's Fund*, le *Policy Studies Institute* (PSI) ou le *Centre for Economic Performance* (CEP).

très important dans le débat public en analysant les propositions des plates-formes politiques avant les élections et en proposant une analyse des choix budgétaires réalisés par le gouvernement. La France a aussi des institutions qui tiennent de ce second modèle, bien qu'aujourd'hui encore peu développées. Si le Parlement est la première institution en charge de l'évaluation des politiques publiques d'après l'article 24 de la Constitution, son rôle a été jusqu'à présent très limité. La Cour des comptes, elle aussi en charge officiellement de l'évaluation des politiques publiques (article 47-2 de la Constitution), joue un rôle nettement plus important. A l'origine, cette juridiction de l'ordre administratif avait pour mission principale de contrôler les comptes des administrations publiques et de certifier les comptes de l'État. L'évaluation a été récemment ajoutée à ses missions premières et la Cour des comptes est aujourd'hui en pointe dans la publication de rapports d'évaluation des différentes politiques publiques. La force de la Cour est son indépendance et sa capacité à bousculer le pouvoir politique et les administrations publiques avec des publications qui animent le débat public. Sa faiblesse est qu'elle privilégie souvent une approche juridique ou comptable de l'évaluation finalement assez éloignée des méthodes d'évaluation d'impact qu'on a présentées plus haut. Enfin, la statistique publique française, qui rassemble l'Insee et les services statistiques ministériels (SSM), joue un rôle clé dans l'évaluation à deux égards. D'abord ces différents services sont chargés de collecter et de mettre en forme les données statistiques documentant les différentes politiques publiques. Ces données d'enquête ou de sources administratives sont cruciales pour décrire les politiques effectivement mises en place mais aussi pour permettre des évaluations, dont la qualité dépendra en grande partie de la disponibilité des données. Ensuite, les services de recherche de ces administrations, notamment à l'Insee, contribuent directement à la réalisation d'évaluations d'impact et à la constitution d'un corpus de connaissance sur les politiques publiques.

Ces deux approches institutionnelles ne sont évidemment pas des alternatives strictes – il s'agit plus d'un idéal-type à la Max Weber –, mais elles représentent deux options assez différentes du rôle que peut jouer l'évaluation des politiques publiques dans le fonctionnement démocratique. Le choix des institutions de l'évaluation dépendra en effet assez fortement de l'idée que l'on se fait du fonctionnement du système politique et de l'indépendance de ses institutions, que ce soit la branche exécutive, législative ou judiciaire.

4.2 Comment favoriser une démocratie de l'évaluation ?

Aussi réussie soit-elle sur le plan technique, l'évaluation des politiques publiques ne peut orienter utilement l'action publique que si elle s'inscrit dans un cadre institutionnel qui facilite la communication des résultats de l'évaluation aux décideurs publics et aux citoyens. Le modèle d'une évaluation au service de la démocratie est – à notre sens – la condition *sine qua non* pour l'efficacité de la démarche d'évaluation. Informer les décideurs de l'impact des politiques est certes une chose nécessaire, mais ce n'est pas suffisant pour entraîner l'adhésion

large des citoyens à des politiques dont ils ne comprendraient pas la motivation. L'évaluation des politiques publiques doit donc être le moyen de favoriser l'appropriation des éléments du débat public par l'ensemble des citoyens.

Au-delà de la diversité des approches institutionnelles, l'évaluation des politiques publiques doit s'efforcer de conjuguer trois exigences : l'indépendance des instances d'évaluation, la multiplication des contacts avec le monde universitaire et le souci de la pédagogie. L'indépendance est la clé qui permet de garantir la crédibilité de l'évaluation : il est indispensable que les résultats de l'évaluation ne puissent être contrôlés par ceux qui mettent en place la politique. Les contacts avec le monde universitaire, en particulier avec la démarche d'évaluation des publications par la communauté scientifique, réduit le risque d'évaluation de complaisance, mais aussi assure que les techniques les plus récentes et les méthodologies les plus robustes seront employées. Enfin, le souci de la pédagogie et la communication au grand public des résultats de l'évaluation sont essentiels pour la réussite de cette démarche.

Le choix des institutions de l'évaluation doit arbitrer entre deux impératifs en partie contradictoires. D'un côté, il est nécessaire de maintenir une indépendance de fait des évaluateurs de politiques publiques. Ceci pousse à les placer hors de la dépendance du pouvoir exécutif, soit sous l'autorité du pouvoir législatif, soit du pouvoir judiciaire (voir par exemple les Hautes Cours, ou la Cour des comptes française), ou encore dans l'indépendance du monde universitaire. D'un autre côté, une bonne évaluation repose sur un processus auquel doivent participer d'assez près les administrations en charge de mettre en place la politique en question. Les informations sur le fonctionnement du dispositif sont souvent collectés en premier lieu par ces administrations et il est souvent idéal de pouvoir mettre en place le protocole d'évaluation en même temps que la mise en place de la politique. Il est donc important que le processus d'évaluation ne les mette pas de côté. Le Conseil d'analyse économique français a ainsi proposé de séparer la fonction de commanditaire de l'évaluation (hors pouvoir exécutif, parlement ou cour des comptes), de la fonction d'évaluateur (plutôt des universitaires) et du rôle des administrations publiques (Conseil d'analyse économique 2013). Ce « tryptique de l'évaluation » aurait pour but de combiner les différentes expertises nécessaires à la réalisation d'évaluation de qualité, tout en maintenant dans une certaine distance institutionnelle les différents intervenants. Quelle institution remplit le rôle du commanditaire importe finalement peu, si ce n'est qu'il est important qu'elle puisse avoir suffisamment de poids institutionnel pour garantir l'indépendance et la crédibilité des évaluations.

La question du financement des travaux d'évaluation n'est pas anodine non plus. Il est difficile d'imaginer autre chose qu'un financement public de l'évaluation, mais pour garantir l'indépendance des évaluateurs, il est nécessaire que ces financements ne soient pas attachés à la production de résultats attendus par les commanditaires. Une règle générale devrait être la garantie *ex ante* de la publication des résultats de l'évaluation – quels que soient ces résultats ; mais plus fondamentalement, le financement des évaluations devrait être suffisamment éloigné

des commanditaires directs, par exemple lié à une activité de recherche et à la publication de ces évaluations dans des revues à comité de lecture. Lors de la mise en place d'une politique publique, prévoir l'évaluation – et un budget pour réaliser celle-ci – dès sa conception est probablement la meilleure façon de garantir que le processus d'évaluation ne va pas dépendre d'une volonté particulière et intéressée.

Enfin, pour réussir le processus d'évaluation, il vaut mieux privilégier peu d'évaluations, mais de qualité, à une profusion d'évaluations qui apportent peu en termes de connaissance. A ce titre, la multiplication – parfois la demande systématique – d'évaluations, réalisées en peu de temps, est un gaspillage de ressources. Consacrer des budgets importants sur des projets qui se donnent le temps de réaliser des travaux substantiels – travaux sur lesquels il sera possible ensuite de s'appuyer – est nettement préférable à la dispersion des moyens et à la précipitation.

5 Conclusion

Cet article a proposé une définition ambitieuse de l'évaluation des politiques publiques comme une partie intégrante du processus démocratique moderne. L'accès des citoyens à des évaluations rigoureuses et honnêtes des politiques publiques mis en œuvre est aujourd'hui la condition *sine qua non* d'un débat public de qualité. Contrairement à certaines idées reçues, l'évaluation n'est pas un processus visant à remplacer la décision démocratique par un gouvernement d'experts. C'est au contraire une démarche visant à faire vivre la démocratie dans un monde complexe, où les citoyens doivent pouvoir bénéficier d'informations, les plus justes possibles, sur les politiques publiques sur lesquels ils doivent se prononcer.

L'évaluation d'une politique publique repose sur l'identification de l'impact causal des dispositifs mis en œuvre. De multiples méthodes statistiques ont été mises au point ces dernières années pour permettre d'approcher rigoureusement la mesure des effets des politiques publiques. Le choix des méthodes et la qualité de leur réalisation dépendent de l'attention des évaluateurs à prendre en compte tous les facteurs concomitants et à bien mettre en avant un contrefactuel crédible.

Pour que l'évaluation soit une réussite démocratique, il n'est pourtant pas suffisant pour les évaluateurs de maîtriser les techniques d'évaluation, il faut aussi qu'ils sachent défendre leur indépendance, garantir leur crédibilité, et communiquer au grand public les résultats de leurs analyses. Pour ce faire, il est nécessaire que le processus d'évaluation soit effectué avec une séparation des rôles entre commanditaire, administration en charge de la politique et évaluateurs, qui respecte les prérogatives de chacun.

En guise de conclusion à ce rapide tour d'horizon de l'évaluation, il est peut-être nécessaire de rappeler qu'une bonne évaluation ne fait pas forcément une bonne politique publique. L'évaluation permet en premier lieu d'apprendre à mieux comprendre l'impact que les inter-

ventions publiques peuvent avoir, de partager la connaissance sur celles-ci, mais au final, la responsabilité des politiques publiques reste dans les mains des citoyens, et à travers le débat démocratique, des décideurs publics.

Références

- Adam, S., Bozio, A., Emmerson, C., Greenberg, D. et G. Knight (2008), ‘A cost-benefit analysis of Pathways to Work for new and repeat incapacity benefits claimants’, *Department for Work and Pensions Research Report* (498).
- Adam, S., Bozio, A. et C. Emmerson (2009), ‘Can we evaluate the impact of the Choices package in Pathways to Work?’, *Department for Work and Pensions Working Papers* (60).
- Adam, S. et A. Bozio (2009), ‘Dynamic scoring’, *OECD Journal on Budgeting* **9**(2).
- Angrist, J. et S. Pischke (2008), *Mostly Harmless Econometrics : An Empiricist’s Companion*, Princeton University Press.
- Bewley, H., Dorsett, R. et G. Haile (2007), ‘The impact of Pathways to Work’, *Department for Work and Pensions Research Report* (435).
- Blanchet, D. (2011), ‘Microsimuler l’avenir des retraites en France : l’exemple du modèle Destinie’, *Cahiers québécois de démographie* **40**(2), 209–238.
- Blundell, R. et M. Costa Dias (2009), ‘Alternative Approaches to Evaluation in Empirical Microeconomics’, *Journal of Human Resources* **44**(3), 565–640.
- Bozio, A. et J. Grenet (2010), *Économie des politiques publiques*, Repères, La Découverte.
- Chauvière, M. (2012), ‘L’obligation de résultats contre les innovations sociales et l’expérimentation’, *Informations sociales* (174).
- Conseil d’analyse économique (2013), ‘Evaluation des politiques publiques’, *Note du CAE* (1).
- del Rey, A. (2013), *La tyrannie de l’évaluation*, Cahiers libres, La Découverte.
- Fack, G. (2006), ‘Are housing benefit an effective way to redistribute income? Evidence from a natural experiment in France’, *Labour Economics* **13**(6), 747–771.
- Givord, P. (2010), ‘Méthodes économétriques pour l’évaluation de politiques publiques’, *Documents de Travail de la DESE* (G2010–08).
- Imbens, G. et J. Wooldridge (2009), ‘Recent Developments in the Econometrics of Program Evaluation’, *Journal of Economic Literature* **47**(1), 5–86.
- Meghir, C. (2008), ‘Dynamic Models for Policy Evaluation’, *IFS working papers* (WP06-08).