Méthodes d'évaluation des politiques publiques

Introduction générale

Hélène Erkel-Rousse

Dans Économie & prévision 2014/1-2 n° 204-205 , pages I à XII
Éditions Direction générale du Trésor

ISSN 0249-4744

DOI 10.3917/ecop.204.0001

Date de mise en ligne : 14/01/2016

Article disponible en ligne à l'adresse

https://shs.cairn.info/revue-economie-et-prevision-2014-1-page-I?lang=fr

CAIRN.INFO

Découvrir le sommaire de ce numéro, suivre la revue par email, s'abonner...
Scannez ce QR Code pour accéder à la page de ce numéro sur Cairn.info.

Distribution électronique Cairn.info pour Direction générale du Trésor.
Vous avez l'autorisation de reproduire cet article dans les limites des conditions d'utilisation de Cairn.info ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Détails et conditions sur cairn.info/copyright.
Sauf dispositions légales contraires, les usages numériques à des fins pédagogiques des présentes ressources sont soumises à l'autorisation de l'Éditeur ou, le cas échéant, de l'organisme de gestion collective habilité à cet effet. Il en est ainsi notamment en France avec le CFC qui est l'organisme agréé en la matière.

Méthodes d'évaluation des politiques publiques

Introduction générale(*)

L'origine de l'évaluation des politiques publiques (EPP) en France remonte au dix-neuvième siècle. Elle est alors portée par des ingénieurs économistes comme Jules Dupuit, dont une contribution de 1844 est généralement considérée comme fondatrice du calcul économique public[1]. Cependant, en tant qu'outil d'aide à la décision publique, l'EPP s'est d'abord développée aux États-Unis, dans les années 1960. Elle s'est ensuite implantée au Royaume-Uni, dans les pays scandinaves puis dans le reste de l'Europe. Elle est encouragée depuis plus de vingt ans par les grands organismes internationaux (Organisation des Nations Unies –ONU–, Organisation pour la Coopération et le Développement Économique –OCDE–, Commission européenne). La France accuse encore un retard par rapport aux pays anglo-saxons dans ce domaine. Toutefois, elle dispose d'un vaste potentiel en matière de compétences scientifiques[2]. Depuis le milieu des années 2000, les projets de recherche en évaluation des politiques publiques s'y multiplient, ainsi que les publications d'articles scientifiques sur ces questions[3].

L'EPP recouvre un champ extrêmement large, allant bien au-delà de l'économie. Sa définition même prête à débats, même si des clarifications utiles ont été apportées à cet égard[4]. Corrélativement, les méthodes d'évaluation des politiques publiques sont multiples[5]. Si l'on se limite aux seules méthodes quantitatives de l'économie, celles-ci relèvent d'approches très diverses. Beaucoup reposent sur des cadres de raisonnements microéconomiques, parmi lesquelles le calcul économique, qui est notamment à la base des bilans coûts-avantages[6], ainsi que diverses modélisations microéconométriques, structurelles ou athéoriques[7]. D'autres se fondent sur des modélisations macroéconomiques[8]. Ces diverses approches, très complémentaires, permettent de croiser des angles de vue différents, avec l'enrichissement que cela implique. En outre, elles se nourrissent mutuellement, certaines pouvant servir d'inputs à d'autres. Enfin, elles se complètent par le fait qu'elles sont adaptées à des questions, contextes et temporalités d'évaluation différents, comme nous allons le voir sur l'exemple des méthodes microéconométriques de l'EPP, sur une partie desquelles se focalise le premier dossier de ce numéro.

(*) Je remercie Michel Houdebine, Pierre Morin, Martine Perbet, Nicolas Riedinger, William Roos et Étienne Wasmer pour leurs commentaires et suggestions très utiles sur une version antérieure de cette introduction générale. Je reste néanmoins seule responsable des erreurs et omissions de ce texte.

Direction générale du Trésor | Téléchargé le 27/10/2026 sur https://shs.cairn.info par Courtier Théophile (IP: 81.1.6.66)

Les méthodes microéconométriques d'évaluation des politiques publiques

Permise par la disponibilité de données individuelles de plus en plus riches et favorisée par la demande croissante d'évaluations de la part des pouvoirs publics, la microéconométrie de l'EPP a connu un fort développement ces dernières années. Deux grandes approches s'y rattachent. La première, qualifiée de structurelle, se fonde sur une représentation explicite de l'économie : un cadre d'hypothèses théoriques complet et cohérent est posé sur le comportement des agents, dont découle l'évaluation. La seconde, qualifiée d'athéorique ou de non structurelle, consiste à identifier l'impact de politiques sans chercher à détailler les mécanismes à l'œuvre, en se reposant en premier lieu sur les données. On parle alors d'estimation en forme réduite.

L'approche structurelle est par nature utile pour contribuer à une évaluation de politique *ex ante* (c'est-à-dire avant la mise en œuvre de la politique). Elle peut également être utilisée dans le contexte d'une évaluation *ex post*, après la mise en œuvre de la politique. L'approche athéorique ne peut se positionner que sur l'évaluation *ex post* (c'est-à-dire après la mise en œuvre de la politique ou son expérimentation sur une sous-population), car elle doit s'appuyer sur des données permettant un recul suffisant. Toutefois, cette distinction est un peu artificielle. En effet, l'évaluation *ex post* par une méthode athéorique d'une politique appliquée antérieurement peut apporter une contribution utile à l'évaluation *ex ante* d'une mesure suffisamment similaire envisagée dans un futur proche[9].

Le premier dossier de ce numéro se focalise sur les méthodes microéconométriques athéoriques d'évaluation d'impact des politiques. Il ne s'agit pas là d'une prise de position dans les débats opposant les partisans de l'approche structurelle (par exemple Rodrik, 2008, Deaton, 2010, Keane, 2010, Langot, 2011) à ceux de l'approche athéorique (Angrist et Pischke, 2010, Imbens, 2010, parmi bien d'autres)[10]. À l'instar, notamment, de Blundell et Costa-Dias (2004), Heckman (2010), Fougère (2010) et Roux (2015), il nous semble que les deux approches ont chacune leurs vertus et leur utilité, appliquées dans un contexte auquel elles sont bien adaptées, et qu'elles sont donc plus complémentaires que concurrentes. Il s'agit en revanche de s'intéresser à un pan de la littérature très dynamique en France depuis la seconde partie des années 2000 et qui soulève des questions de méthodes suffisamment complexes et riches pour mériter qu'on y consacre un dossier.

Les méthodes microéconométriques athéoriques d'évaluation d'impact des politiques visent à identifier l'effet propre d'une politique. Comme l'illustrent les articles du dossier qui leur est consacré, elles reposent très largement sur le cadre de raisonnement proposé par Rubin (1974). Ce cadre, inspiré de l'évaluation des traitements médicaux, emprunte le vocabulaire des essais cliniques : la mesure à évaluer (politique, dispositif, programme) est désignée sous le terme de *traitement*, ses bénéficiaires sont dits *traités* et les personnes qui n'en bénéficient pas sont désignées comme *non traitées*[11]. La mesure est évaluée au regard d'une variable de résultat (par exemple le revenu). L'exercice d'évaluation d'impact cherche à comparer la situation d'une population d'individus au regard de cette variable de résultat selon que ceux-ci bénéficient ou non de la mesure, toutes choses étant inchangées par ailleurs. Cependant, on n'observe jamais simultanément la variable de résultat pour un même individu dans les cas où il est traité ou non. Selon qu'il bénéficie ou pas de la mesure, un seul de ses deux résultats potentiels est connu. Toute la difficulté consiste donc à trouver une manière d'approcher le plus correctement possible la situation qui ne s'est pas réalisée. C'est le problème du choix du "contrefactuel", celui-ci

Direction générale du Trésor | Téléchargé le 27/10/2026 sur https://shs.cairn.info par Courtier Théophile (IP: 81.1.6.66)

représentant dans l'idéal la situation en tous points identique à la situation observée, à l'application de la mesure à évaluer près.

On ne peut approcher l'écart entre les deux résultats potentiels par l'écart de résultat entre avant et après la mise en œuvre de la mesure, d'autres événements que la mesure (notamment la conjoncture économique) influençant l'évolution du résultat dans le temps. On ne peut pas non plus l'approcher par une comparaison entre les résultats moyens des bénéficiaires et des non bénéficiaires (comparaison dite "avec / sans"), du fait que la population de bénéficiaires diffère le plus souvent de celle des non bénéficiaires avant même la mise en œuvre de la mesure, par une série de caractéristiques observables ou non. Le problème posé par la différence initiale entre les deux groupes d'individus (bénéficiaires et non bénéficiaires) est connu sous le nom de "biais de sélection". Il doit être impérativement pris en compte pour éviter que l'évaluation soit biaisée[(12)].

En pratique, les méthodes microéconométriques athéoriques d'évaluation d'impact de politiques reposent sur la comparaison entre un groupe d'individus bénéficiaires de la mesure (dit groupe de traitement ou groupe test) et un groupe d'individus non bénéficiaires (dit groupe de contrôle ou groupe témoin) judicieusement constitué pour échapper à toutes sortes de risques de biais, dont le biais de sélection, et constituer un contrefactuel valide au groupe de traitement. Les différentes méthodes se rattachant à cette approche de l'évaluation se distinguent par la manière dont est constitué ce contrefactuel. Il existe deux grandes familles de méthodes.

La première repose sur l'utilisation de bases de données individuelles constituées préalablement et indépendamment du besoin d'évaluation, couvrant une période de temps avant et après la mesure à évaluer et comportant suffisamment d'informations pertinentes tant sur des individus ayant bénéficié de la mesure que sur des individus n'en ayant pas bénéficié. Les techniques économétriques utilisées pour estimer l'impact de la mesure sont conçues pour éliminer les risques de biais, sous certaines hypothèses dont il convient de vérifier aussi soigneusement que possible le réalisme préalablement à leur utilisation. Ces techniques diffèrent par le jeu d'hypothèses assurant leur validité et par le contexte dans lequel elles peuvent s'appliquer.

La présentation et la discussion des principales grandes techniques de ce type est l'objet de l'article de **Pauline Givord**. Il s'agit d'un guide précieux pour qui est amené à utiliser ces techniques ou à apprécier la pertinence de leur utilisation dans tel ou tel contexte. L'auteure détaille les différentes étapes de l'exercice d'évaluation par ces méthodes (données nécessaires, hypothèse d'identification, interprétation des résultats) et les illustre par des exemples. Elle présente plus particulièrement la méthode de différence de différences, la technique des variables instrumentales, la régression sur discontinuité et des méthodes de sélection sur observables (dont les méthodes d'appariement). On se réfère à cette famille de méthodes en parlant d'expériences naturelles ou de méthodes quasi-expérimentales, par opposition aux expériences contrôlées.

Les expériences contrôlées, appelées aussi expérimentations randomisées, constituent la seconde famille de méthodes microéconométriques athéoriques d'évaluation d'impact. Les évaluateurs qui utilisent ce type de méthode s'appuient non pas sur des bases de données préexistantes mais sur une expérimentation *in vivo*. Inspirée des essais cliniques, celle-ci consiste à tirer au sort un groupe d'individus auquel sera appliqué le traitement (le groupe test) et un autre groupe auquel il ne sera pas appliqué (le groupe témoin). On parle alors d'assignation aléatoire. Un tirage aléatoire bien mené d'un nombre suffisant d'individus assure en principe que groupe test et groupe témoin soient comparables à l'application de la mesure près, c'est-à-dire que le biais de

Direction générale du Trésor | Téléchargé le 27/10/2026 sur https://shs.cairn.info par Courtier Théophile (IP: 81.1.6.66)

sélection soit évité. Le protocole de l'expérience devra en outre être conçu suffisamment soigneusement pour essayer de prévenir un certain nombre d'autres risques susceptibles d'induire des biais d'évaluation. Moyennant quoi, l'estimation de l'impact de la mesure pourra ensuite être menée au moyen de techniques économétriques relativement simples.

Cette famille de méthodes est présentée dans l'article d'**Adrien Bouguen et Juliette Seban**. Les auteurs détaillent les conditions dans lesquelles l'assignation aléatoire permet d'évaluer sans biais l'impact d'une mesure. Ils décrivent différents protocoles d'expérimentation possibles et leurs conséquences en termes d'applicabilité, d'éthique et de précision statistique. L'article mentionne les risques qui pèsent sur les résultats d'analyse, en distinguant ceux qui sont spécifiques à l'assignation aléatoire de ceux qui ne le sont pas. En mettant bien en évidence les avantages de cette méthode, tout en en indiquant aussi les risques et les moyens de s'en prémunir, les auteurs transmettent des informations précieuses à tout économiste amené à réaliser sa première évaluation ou à utiliser de manière éclairée le travail d'évaluation d'autrui.

Les exemples concrets d'application présentés dans les autres articles de ce dossier illustrent la diversité des mesures qui peuvent être évaluées par ce type de méthode, ainsi que le grand soin qui doit être consacré en amont de l'estimation de l'impact d'une politique au choix de la méthode, selon son degré d'adéquation au problème et le niveau de réalisme des hypothèses qui sous-tendent sa validité.

Les articles de Maëlle Fontaine et Thomas Le Barbanchon, Claire Marbot et Delphine Roy, et Amélie Mauroux présentent tout d'abord des illustrations d'évaluations à partir d'expériences naturelles.

Maëlle Fontaine et Thomas Le Barbanchon s'intéressent à l'évaluation du suivi mensuel personnalisé (SMP) instauré par l'Agence nationale pour l'emploi (ANPE) en 2006. Ce dispositif consiste en un suivi précoce (dès le quatrième mois de chômage) et intensif des demandeurs d'emploi présentant un risque de chômage de longue durée faible ou modéré. Les auteurs optent pour l'utilisation de la méthode de différence de différences après avoir soigneusement vérifié que les conditions dans lesquelles elle s'applique sont respectées dans le contexte de leur étude. Ils estiment l'impact du SMP sur le retour à l'emploi et sur les radiations administratives des listes de demandeurs d'emploi en s'appuyant sur des données du Fichier Historique Statistique et du Fichier Historique Administratif de Pôle emploi, sur la période allant d'octobre 2003 à septembre 2007. Selon leurs résultats, l'intensification du suivi aurait sensiblement réduit la durée passée sur les listes de l'ANPE, de 5 à 12 %. Cette réduction proviendrait à la fois de reprises d'emploi et de radiations administratives intervenant plus rapidement après l'inscription.

Claire Marbot et Delphine Roy cherchent à évaluer l'impact de la transformation en 2007 de la réduction d'impôt pour l'emploi de salariés à domicile en crédit d'impôt. Elles estiment l'effet de cette mesure, qui concerne les ménages actifs, à partir de données fiscales exhaustives permettant un suivi longitudinal des ménages de 2006 à 2008. La technique d'estimation utilisée combine les méthodes de différence de différences et d'appariement. L'étude illustre la difficulté qu'il peut y avoir en pratique à définir ce qu'est le *traitement*. Les auteures envisagent deux définitions possibles de ce dernier. Si les deux reposent sur la variation du taux de subvention de la dépense en services à domicile liée à l'introduction du crédit d'impôt, la première s'intéresse au taux de subvention moyen, tandis que la seconde se réfère au taux de subvention marginal. Selon les estimations réalisées, la mesure aurait incité entre 86 000 et 120 000 ménages supplémentaires à utiliser des services à domicile déclarés.

Direction générale du Trésor | Téléchargé le 27/10/2026 sur https://shs.cairn.info par Courtier Théophile (IP: 81.1.6.66)

Amélie Mauroux étudie l'impact incitatif de la hausse en 2006 de 25 % à 40 % du taux de crédit d'impôt sur les travaux d'amélioration de l'efficacité énergétique des logements. Cette mesure est réservée aux ménages vérifiant deux conditions : être propriétaires depuis deux ans au plus d'un logement ; que celui-ci ait été achevé avant 1977. Les données utilisées proviennent des fichiers de déclarations d'impôt de 2006 à 2009, appariés aux fichiers de taxe d'habitation correspondants. Trois groupes témoins sont distingués : ceux des individus vérifiant une seule des deux conditions requises pour être éligibles à la mesure et celui des individus ne vérifiant aucune de ces conditions. L'auteure tire parti de cette configuration pour tenir compte de l'effet propre de chacun des critères d'éligibilité sur la probabilité de recours au crédit d'impôt. Elle combine un modèle d'appariement à une technique dite de triples différences. Ses résultats indiquent que, si le taux avait été maintenu à 25 %, un ménage éligible sur 15, parmi les déclarants occupant un logement achevé entre 1969 et 1976, n'aurait pas recouru au crédit d'impôt en 2006, un sur huit en 2007 et 2008.

Enfin, **Marianne Bléhaut et Roland Rathelot** présentent une comparaison troublante des résultats d'évaluations menées d'une part sur la base d'une expérience contrôlée, d'autre part en utilisant la méthode d'appariement comme s'ils ne disposaient pas du groupe de contrôle issu de l'expérimentation. Les données utilisées sont issues de l'expérimentation "jeunes diplômés" mise en œuvre entre mi-2007 et mi-2009 pour évaluer un programme d'accompagnement renforcé visant des jeunes de moins de 30 ans diplômés du supérieur et en situation de chômage de longue durée. L'évaluation de l'impact du dispositif par la méthode expérimentale (*i.e.* en utilisant l'intégralité des données rendues disponibles par l'expérience contrôlée) aboutit à l'estimation d'un effet positif important du programme sur la probabilité d'avoir retrouvé un emploi d'au moins six mois, huit mois après l'entrée dans le programme. Au contraire, l'utilisation de la méthode d'appariement sur score de propension conclut à des conséquences négatives ou nulles de l'accompagnement renforcé. Les résultats obtenus par les deux méthodes sont donc très différents. Les auteurs expliquent pourquoi ils suspectent la méthode d'appariement de conduire à des résultats erronés.

Ce dossier, il faut le souligner, a été l'occasion pour *Économie et Prévision* d'inaugurer une formule à laquelle elle n'avait encore jamais eu recours et qui a fait ses preuves dans le cadre de nombreuses autres revues : celle des commentaires. Deux articles sont commentés : celui d'Adrien Bouguen et Juliette Seban et celui de Marianne Bléhaut et Roland Rathelot. Dans un cas comme dans l'autre, la décision de publier un commentaire, non programmée à l'avance, a été prise suite à des échanges fructueux entre les auteurs et les rédacteurs de ces commentaires.

Le commentaire de l'article d'Adrien Bouguen et Juliette Seban par **Hélène Erkel-Rousse** se focalise sur un des éléments clés qui déterminent l'élaboration d'un protocole d'évaluation expérimentale (choix de la taille de l'échantillon expérimental, des proportions d'individus assignés au groupe test et au groupe témoin, etc.) : le calcul de la puissance statistique de l'évaluation. Celui-ci revêt une grande importance pratique. En effet, il conditionne la capacité à évaluer de manière satisfaisante l'effet d'une mesure de politique publique sur la base d'un protocole expérimental en influençant les choix de ses caractéristiques techniques. Le commentaire vise à documenter le débat sur le type de test d'hypothèses sur lequel fonder de préférence le calcul de puissance. Il montre que le choix d'un test unilatéral ou d'un test bilatéral permet indifféremment de répondre aux questions qui se posent à l'évaluateur, dès lors que les propriétés de ces tests sont bien maîtrisées. Il rappelle en outre l'importance pour le calcul de puissance d'une bonne formulation de l'écart-type de l'estimateur de l'effet de la politique à évaluer.

Direction générale du Trésor | Téléchargé le 27/10/2026 sur https://shs.cairn.info par Courtier Théophile (IP: 81.1.6.66)

Le commentaire de l'article de Marianne Bléhaut et Roland Rathelot par **Sylvain Chabé-Ferret** discute la portée de ses résultats et leurs conséquences pour l'utilisation de la méthode d'appariement, fortement suspectée par les auteurs de conduire à des résultats erronés. Le commentaire explique pourquoi, en toute rigueur, les résultats de l'article ne permettent pas de répondre à la question fondamentale : les variables observées suffisent-elles à éliminer le biais de sélection ? Les résultats de l'article démontrent en revanche de manière indiscutable que les variables observées sont insuffisantes pour neutraliser simultanément sélection, attrition et effets de déplacement[13]. Ainsi, même s'ils n'en apportent pas la preuve définitive, les résultats jettent un doute sérieux sur la capacité de l'appariement à résoudre le problème de biais de sélection. Ce doute est encore renforcé à la lumière des résultats de travaux similaires (Lalonde, 1986, parmi d'autres), dont les principaux enseignements sont résumés.

Au total, ce dossier met en évidence l'intérêt des méthodes microéconométriques athéoriques d'évaluation d'impact des politiques, ainsi que les questions techniques parfois complexes que l'évaluateur doit résoudre pour présenter des estimations d'impact valides. Les applications concrètes illustrent bien le type de mesure que ces méthodes sont susceptibles d'évaluer et à quel horizon : il s'agit de dispositifs très divers mais touchant toujours une population de bénéficiaires relativement limitée et dont l'impact est estimé à relativement court terme. En effet, par nature, ces méthodes ne peuvent pas capter les phénomènes de bouclage macroéconomique (d'"équilibre général") qui se produiraient au moins à moyen long terme si la politique à évaluer touchait une proportion importante de la population. Dans ce cas, les méthodes décrites dans ce dossier ne seraient pas valides, ne serait-ce que parce que les individus du groupe de contrôle seraient alors indirectement affectés par la mesure. Une approche structurelle serait alors plus adaptée.

D'autres limites sont mentionnées et discutées dans les différents articles du dossier[14]. Concernant les méthodes d'évaluation fondées sur des expériences naturelles, le doute porte le plus souvent sur la validité interne des résultats : en dépit de toute la prudence des évaluateurs, ces méthodes permettent-elles d'aboutir réellement à une estimation d'impact non biaisée ? Pour ce qui est des expériences contrôlées, un des problèmes principalement mentionnés réside dans la difficulté à tirer de leurs résultats, souvent obtenus sur un périmètre très localisé dans l'espace et le temps, des enseignements plus généraux.

Ces considérations font en partie écho à la manière dont le champ d'application des méthodes microéconométriques d'EPP est décrit par Heckman (2000) à l'aide de sa typologie en trois points des questions caractéristiques de l'EPP : 1) l'évaluation d'une politique déjà mise en œuvre ou, du moins, testée sur une sous-population ; 2) l'évaluation d'une politique appliquée à un environnement différent de celui dans lequel elle a déjà été mise en œuvre ; 3) l'évaluation d'une politique nouvelle pas encore testée auparavant. Les méthodes microéconométriques athéoriques utilisées seules ne peuvent prétendre répondre qu'à la première de ces trois questions[15]. Les questions 2 et 3 requièrent l'utilisation de méthodes structurelles. Selon les auteurs, cette limitation du domaine d'applicabilité des méthodes athéoriques est perçue comme plus ou moins dirimante[16]. Pour sa part, Heckman (2000) considère les méthodes microéconométriques athéoriques comme très utiles dans les cas, nombreux, où la connaissance de l'ensemble des paramètres structurels de l'économie n'est pas nécessaire : ces méthodes identifient les effets de mesures plus ciblées mais sous des hypothèses plus faibles, donc plus crédibles, que celles requises pour identifier un modèle microéconomique structurel. Ainsi que le souligne Heckman (2000) et que tente de l'illustrer ce dossier, l'approche microéconométrique athéorique a donné lieu à des développements féconds et à

Direction générale du Trésor | Téléchargé le 27/10/2026 sur https://shs.cairn.info par Courtier Théophile (IP: 81.1.6.66)

de nombreuses évaluations d'impact éclairantes. Le tout est de bien mesurer ce qu'elle peut dire ou ne pas dire, selon le contexte, et de ne pas se limiter à cette seule approche. En d'autres termes, le recours à plusieurs méthodes complémentaires, fondées sur des paradigmes et modes de pensées différents, est sans doute plus fécond que des débats opposant certaines méthodes à d'autres. Ceci nous mène naturellement au second dossier.

Regards sur l'évaluation des politiques publiques

Dans le second dossier de ce numéro spécial, deux auteurs apportent leur regard sur l'EPP.

Étienne Wasmer discute les avantages et inconvénients de l'interdisciplinarité en matière d'EPP. Sa réflexion s'appuie sur les enseignements de Tetlock (2005) concernant la capacité prédictive des experts. Celui-ci suggère que des experts de science politique et d'économie auraient de moins bonnes performances prédictives lorsque leurs modes de raisonnement sont plus étroits ou centrés sur un seul paradigme. Ce profil d'experts est désigné sous le terme de *hérisson*, par opposition au profil de *renard*, correspondant à des experts qui accumulent des informations et raisonnements variés et ne croient pas aux grands schémas explicatifs. Tout en reconnaissant que prévision et évaluation diffèrent, Étienne Wasmer envisage l'extrapolation du résultat de Tetlock (2005) à l'EPP. Il passe ensuite en revue les coûts de transaction générés par les échanges interdisciplinaires, puis il évoque les principes d'organisation de la science et l'existence d'une tension entre le besoin de spécialisation pour la gestion des carrières des chercheurs et les gains de l'ouverture interdisciplinaire. En conclusion, il suggère que le but ultime de l'EPP devrait être de pouvoir quantifier, au sein de modèles de prévision *ex ante*, la subtilité du contexte et les enjeux de la mise en œuvre des mesures de politiques publiques, telles qu'elles sont analysées par les différentes sciences sociales.

Stéphane Gregoir constate que, si le discours politique est favorable à l'EPP, l'organisation de l'administration et des institutions françaises, de même que leur histoire récente, ne traduit pas clairement ce souhait. En témoigne une multiplicité d'acteurs sans institutions fortement identifiées comme chargées de la mission d'EPP. L'auteur considère qu'une véritable institutionnalisation de cette mission serait pourtant utile. Elle permettrait l'accumulation des expériences en matière de méthodologies pertinentes pour la conduite des EPP et de modalités et consignes de mise en œuvre des mesures de politiques nouvelles. Elle favorisait la professionnalisation et la diffusion des bonnes pratiques. Elle faciliterait la conduite régulière de programmes d'évaluation et les rendrait plus visibles. L'auteur envisage quels pourraient être les acteurs de cette institutionnalisation et quelles actions seraient susceptibles d'y contribuer. Selon lui, certains organismes existants comme la Cour des Comptes ou l'Insee pourraient être en mesure de devenir les institutions coordinatrices et porteuses de l'évaluation *ex ante* et *ex post* à condition d'être renforcés en moyens et en compétences.

Les deux articles expriment des nuances concernant les rôles respectifs des différents types d'acteurs de l'EPP. Étienne Wasmer insiste sur l'importance des chercheurs pour l'EPP, notamment sur le plan de l'innovation méthodologique et en raison de leur indépendance statutaire. Celle-ci leur permet de s'atteler à toutes sortes de questions mais nécessite une grande vigilance quant à la transparence des liens d'intérêt et quant à la neutralité. Stéphane Gregoir privilégie une approche plus institutionnelle et appelle de ses vœux l'incarnation de l'évaluation, *ex ante* d'une part et *ex post* d'autre part, par deux

Direction générale du Trésor | Téléchargé le 27/10/2026 sur https://shs.cairn.info par Courtier Théophile (IP: 81.1.6.66)

instances appartenant au monde de l'administration mais bénéficiant d'une indépendance fonctionnelle reconnue. Selon lui, les chercheurs ont leur place en tant que contributeurs à l'EPP en dépit de différences notables entre le fonctionnement du monde académique et les besoins en termes d'EPP, mais leurs contributions doivent être coordonnées par de telles instances. En outre, l'évaluation *ex ante* doit être assez rapide pour être compatible avec le temps politique. À cet égard, les deux articles illustrent, sans prétention d'exhaustivité, la diversité des points de vue sur les formes d'organisation que devrait prendre l'EPP.

Les deux articles se rapprochent, avec des nuances, sur l'idée que l'interdisciplinarité (Étienne Wasmer) et la pluridisciplinarité (Stéphane Gregoir) constituent des caractéristiques hautement souhaitables de l'EPP, ainsi que sur le caractère jugé fondamental de l'indépendance des EPP comme facteur assurant leur fiabilité. Sans être mises systématiquement en avant en tant que telles, l'interdisciplinarité et la pluridisciplinarité sont des éléments de la pluralité, souvent évoquée sous ce terme ou à travers des concepts apparentés parmi les grands principes de l'EPP. *A contrario*, si le principe d'indépendance constitue l'un des trois grands piliers de l'EPP selon le rapport Viveret (1989) (avec, notamment, la transparence), il ne figure pas sur la liste des principaux principes de l'évaluation de la Charte de l'évaluation établie par la Société française de l'évaluation (SFE, 2015), à savoir : pluralité, distanciation, compétence, respect des personnes, transparence, opportunité et responsabilité. De même, dans son guide de bonnes pratiques, la Commission européenne (2015) met l'accent sur la conduite collective et la transparence des évaluations mais ne mentionne pas l'indépendance dans sa liste de grands principes.

La multiplicité des avis sur l'EPP[17] est probablement à mettre en regard du "mille-feuille" institutionnel décrit par Stéphane Gregoir et à rapprocher du développement relativement récent en France des évaluations microéconométriques d'impact fondées notamment sur des expériences naturelles ou contrôlées, la convergence des points de vue et l'élaboration d'une vision de l'EPP largement partagée nécessitant du temps. Plus positivement, on peut voir dans la diversité de ces avis et dans la passion entourant les débats sur l'EPP un signe positif de vitalité et de dynamisme d'une discipline et d'une pratique en mouvement. Nul doute que de nombreuses autres publications traitant d'évaluation des politiques publiques sont à venir et qu'au-delà des approches existantes, des méthodes nouvelles émergeront, permettant à l'économiste de répondre à de nouveaux défis[18].

Hélène Erkel-Rousse

Rédactrice en chef d'*Économie et Prévision*

Direction générale du Trésor | Téléchargé le 27/10/2026 sur https://shs.cairn.info par Courtier Théophile (IP: 81.1.6.66)

Notes

(1) *Cf.* Guesnerie (2004).

(2) Le prix Nobel d'économie de Jean Tirole en 2014 atteste de la reconnaissance internationale de la qualité de la recherche en France et engendre une dynamique dont l'EPP ne peut que bénéficier.

(3) En témoignent notamment les publications de nombreux numéros spéciaux consacrés à l'évaluation des politiques publiques et à des problématiques connexes dans les revues scientifiques d'économie françaises – *cf.* notamment *Annales d'économie et statistique* (2002), *Économie et Prévision* (2006, 2008a,b et ce numéro) et *Revue française d'économie* (2011, 2012, 2013, 2014, 2015).

(4) Notamment par des organismes comme la Société française de l'évaluation (SFE, 2006, Fouquet, 2009 et 2013) et du fait des actions en faveur du développement de l'EPP menées par les organismes internationaux (mise en ligne de guides de bonnes pratiques de l'EPP, etc. – *cf.* par exemple OCDE, 2014, et *Commission européenne*, 2015).

(5) *Cf.* notamment Trosa (dir.) (2009), Studer (dir.) (2013), ainsi que les références citées en note (3).

(6) *Cf. Cahiers de l'évaluation* (2008) pour une présentation des principes de cette approche, qui bénéficie d'une longue tradition d'expertise française. Voir aussi *Économie et Prévision* (2006) pour une analyse de l'évolution du calcul économique (article de R. Guesnerie) et pour plusieurs illustrations (articles de Bonnafous, Jensen et Roy et de Maurice, Quinet et Sauvant).

(7) Cette diversité apparaît fortement à la lecture de Studer (dir.) (2002), ainsi que des références citées dans la note (3).

(8) On trouvera des exemples d'approches macroéconomiques, structurelles ou non, dans *Annales d'économie et statistique* (2002), *Économie et Prévision* (2006) (article de C. Gianella sur les 35 heures) et *Économie et Prévision* (2008b).

(9) Soit en tant que telle, soit indirectement. Par exemple, des paramètres d'un modèle microéconomique structurel mobilisé dans le cadre d'une évaluation de politique *ex ante* peuvent être calibrés sur la base des résultats d'expériences contrôlées réalisées dans le cadre d'une évaluation antérieure – *cf.* notamment Fougère (2010) pour des références de ce type. N.B. : les expériences contrôlées relèvent de l'approche microéconométrique athéorique d'évaluation d'impact – pour plus de détails, voir *infra*.

(10) Le lecteur trouvera des développements très intéressants sur ces débats dans les numéros spéciaux publiés par la *Revue française d'économie* (2011, 2012, 201, 2014, 2015) sur l'EPP – *cf.* notamment Roux (2015).

(11) Le cadre de Rubin est parfois critiqué pour sa vision, jugée quelque peu réductrice, des politiques comme des traitements administrés à des populations. Le cadre de Rubin a néanmoins le grand avantage de poser le problème de l'évaluation d'impact d'une manière claire et en mettant bien en évidence certaines de ses difficultés essentielles en matière d'inférence (*cf. infra*).

(12) Pour une introduction très pédagogique à ces problèmes et aux manières d'y remédier, le lecteur pourra se référer à Dupont-Courtade *et alii* (2015), qui s'appuient sur de nombreux exemples très éclairants.

(13) L'attrition désigne le fait que certains individus disparaissent de l'échantillon après l'assignation aléatoire. Si elle affecte différemment le groupe test et le groupe témoin, elle a pour conséquence de rendre les deux groupes imparfaitement comparables, alors que l'expérience avait précisément pour objet de les rendre parfaitement comparables pour neutraliser le biais de sélection. Les effets de déplacement (ou, encore, effets de diffusion ou externalités) correspondent à la situation où les individus du groupe témoin sont affectés indirectement par la mesure à évaluer du fait de son application aux individus du groupe test. Une configuration typique de ce type est celle où un vaccin appliqué à une sous-population (le groupe test) induit une moindre contamination des individus non vaccinés (le groupe témoin) du fait de leur moindre exposition au virus due à la vaccination des personnes du groupe test.

(14) Pour des compléments sur les intérêts et limites de ces méthodes, on pourra se reporter notamment à Banerjee et Duflo (2009a,b), L'Horty et Petit (2011), Fougère (2010, 2012), *Conseil d'analyse économique* (2013), Legendre (2013) et Dupont-Courtade *et alii* (2015).

Direction générale du Trésor | Téléchargé le 27/10/2026 sur https://shs.cairn.info par Courtier Théophile (IP: 81.1.6.66)

(15) Seules, c'est-à-dire en tant que telles et non au sens envisagé dans la note (9). Cette note explique en effet comment les méthodes microéconométriques athéoriques peuvent contribuer indirectement au traitement des questions 2 et 3, par exemple en justifiant le calibrage de certains paramètres de modèles structurels.

(16) Ainsi, Langot (2011) considère l'approche microéconométrique athéorique comme excessivement limitée au regard de ce que peut traiter l'approche structurelle. Fouquet (2009), quant à elle, décrit l'analyse contrefactuelle par expérience contrôlée comme une méthode qui « ne réduit pas la complexité du réel, elle l'ignore. Elle ne s'applique qu'à des cas très spécifiques (un dispositif réduit à une mesure et non une politique ; un contexte de rareté qui justifie que la mesure soit réservée à certains ; une causalité simple ou sans intérêt pour l'action et la décision ». Cette critique pourrait porter aussi bien sur les analyses contrefactuelles menées sur la base d'expériences naturelles.

(17) Pour diverses conceptions de l'EPP et une tentative de synthèse, le lecteur pourra se reporter par exemple à l'introduction de Trosa "L'évaluation : nécessité ou gadget ?" dans Trosa (dir.) (2009).

(18) À cet égard, voir par exemple Heckman (2000), *summary and conclusions*.

Direction générale du Trésor | Téléchargé le 27/10/2026 sur https://shs.cairn.info par Courtier Théophile (IP: 81.1.6.66)

Bibliographie

Angrist J.D. et Pischke J.-S. (2010). "The Credibility Revolution in Empirical Economics: How Better Research Is Taking the Con out of Econometrics", *Journal of Economic Perspective*, vol. 24, n° 2, pp. 3-30.

***Annales d'économie et de statistique* (2002).** *Économétrie de l'évaluation des politiques*, Numéro spécial, n° 67-68, juillet-décembre, 517 pages.

Banerjee A.V. et Duflo E. (2009a). "The Experimental Approach to Development Economics", *Annual Review of Economics*, vol. 1, April, pp. 151-178.

Banerjee A.V. et Duflo E. (2009b). "L'approche expérimentale en économie du développement", *Revue d'économie politique*, vol. 119, mai, pp. 691-726.

Blundel R. et Costa-Dias M. (2004). "Alternative Approaches to Evaluation in Empirical Microeconomics", *Journal of Human Resources*, vol. 44, n° 3, pp. 565-640.

***Cahiers de l'évaluation* (2008).** "Évaluer aujourd'hui - Méthodologie : analyse coûts-avantages", Dossier *Calculer pour décider*, vol. 2, n° 2, septembre, pp. 9-11.

***Commission européenne* (2015).** *Better Regulation "Toolbox*?, 414 pages (http://ec.europa.eu/smart-regulation/index_en.htm).

***Conseil d'analyse économique* (CAE) (2013).** *Évaluation des politiques publiques*, Les notes du Conseil d'analyse économique, n° 1, février, 12 pages.

Deaton A. (2010). "Instruments, Randomization and Learning about Development", *Journal of Economic Literature*, vol. 48, n° 2, pp. 424-455.

Dupont-Courtade L., Chabé-Ferret S., Treich N. et Perbet M. (2015). "Évaluation des politiques publiques : expérimentation randomisée et méthodes quasi-expérimentales", Document de travail de la Direction générale du Trésor n° 2015/01, février, 63 pages.

***Économie et Prévision* (2006).** *Aide à la décision publique : développements récents*, Numéro spécial, n° 175-176, 2006/4-5, 220 pages.

***Économie et Prévision* (2008a).** *Économie expérimentale, outil d'aide à la décision*, Numéro spécial, n° 182, 2008/1, 172 pages.

***Économie et Prévision* (2008b).** *La nouvelle modélisation macroéconomique appliquée à l'analyse de la conjoncture et à l'évaluation des politiques*, Numéro spécial, n° 183-184, 2008/2-3, 262 pages.

Fougère D. (2010). "Les méthodes économétriques d'évaluation", *Revue Française des Affaires Sociales*, n° 1-2, pp. 105-128.

Fougère D. (2012). "Les méthodes d'expérimentation en question", *Éducation et Formations*, n° 81, mars, pp. 41-47.

Fouquet A. (2009). "L'évaluation des politiques publiques, concepts et enjeux", *in* Trosa (dir.) (2009), pp. 21-33.

Fouquet A. (2013). "L'évaluation des politiques publiques en France : définitions et historique", in Studer (dir.) (2013), pp. 9-16.

Guesnerie R. (2004). "Dupuit hier et aujourd'hui", CNRS-EHESS-ENS, Document de travail du Delta n° 2004-22, 20 pages.

Heckman J.J. (2000). "Microdata, Heterogeneity and the Evaluation of Public Policy", Nobel Prize Lecture, December 8, *Economic Sciences* 2000, pp. 255-322 (également publié dans *The Journal of Political Economy*, 2001, vol. 109, n° 4, August, pp. 673-748).

Heckman J.J. (2010). "Building Bridges between Structural and Program Evaluation Approaches to Evaluating Policy", *Journal of Economic Literature*, vol. 48, n° 2, pp. 356-398.

Imbens G.W. (2010). "Better LATE than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)", *Journal of Economic Literature*, vol. 48, n° 1, pp. 399-423.

Keane M.P. (2010). "Structural vs. Atheoric Approaches to Econometrics", *Journal of Econometrics*, vol. 156, n° 1, pp. 3-20.

Lalonde R. (1986). "Evaluating the Econometric Evaluations of Training Programs with Experimental Data", *American Economic Review*, vol. 76, n° 4, pp. 604-620.

Direction générale du Trésor | Téléchargé le 27/10/2026 sur https://shs.cairn.info par Courtier Théophile (IP: 81.1.6.66)

Langot F. (2011). "Évaluer les politiques d'emploi : un plaidoyer pour une approche structurelle", *in Revue française d'économie* (2011), pp. 123-166.

Legendre F. (2013). "Une introduction à la microéconométrie de l'évaluation", *in Revue française d'économie* (2013), pp. 9-42.

L'Horty Y. et Petit P. (2011). "Évaluation aléatoire et expérimentations sociales", *in Revue française d'économie* (2011), pp. 13-48.

OCDE (2014). *OECD Framework for Regulatory Policy Evaluation*, OECD Publishing, 156 pages (http://dx.doi.org/10.1787/9789264214453-en).

***Revue française d'économie* (2011).** *L'évaluation des politiques publiques*, Numéro spécial, Langot F. et Petit P. (dir.), vol. XXVI, n° 1, juillet, 251 pages.

***Revue française d'économie* (2012).** *L'évaluation des politiques publiques*, Numéro spécial, Langot F. et Petit P. (dir.), vol. XXVII, n° 1, juillet, 213 pages.

***Revue française d'économie* (2013).** *L'évaluation des politiques publiques*, Numéro spécial, vol. XXVIII, n° 1, juillet, 208 pages.

***Revue française d'économie* (2014).** *L'évaluation des politiques publiques*, Numéro spécial, vol. XXIX, n° 1, juillet, 231 pages.

***Revue française d'économie* (2015).** *L'évaluation des politiques publiques*, Numéro spécial, vol. XXX, n° 1, juillet, 217 pages.

Rodrik D. (2008). "The New Development Economics: We Shall Experiment, but How Shall We Learn?", Harvard Kennedy School, Faculty Research Working Paper Series, October, n° 08-055, 36 pages.

Roux S. (2015). "Approches structurelles et non structurelles en microéconométrie de l'évaluation des politiques publiques", *in Revue française d'économie* (2015), pp. 13-65.

Rubin D.B. (1974). "Estimating Causal Effect of Treatment in Randomized and Nonrandomized Studies", *Journal of Education Psychology*, vol. 66, n° 5, pp. 688-701.

***Société française de l'évaluation* (SFE) (2006).** *Charte de l'évaluation des politiques publiques et des programmes publics*, version actualisée au 21 juin 2006, 4 pages.

Studer N. (dir.) (2013). *Méthodes microéconométriques d'évaluation des politiques publiques*, Actes du séminaire, Direction de la recherche, des études, de l'évaluation et des statistiques (Drees), collection Études et statistiques, Ministère du travail, de l'emploi et de la santé, Ministère du Budget, des Comptes publics, de la Fonction Publique et de la Réforme de l'État, Ministère des solidarités et de la cohésion sociale, DICOM : 11-025, 9 décembre, 139 pages.

Tetlock P. E. (2005). *Expert Political Judgment. How Good Is It? How Can We Know?*, Princeton University Press, 321 pages.

Trosa S. (dir.) (2009). *Évaluer les politiques publiques pour améliorer l'action publique, une perspective internationale*, Institut de la gestion publique et du développement économique, éditions du Comité pour l'histoire économique et financière de la France, série Recherche, Études, Veille, Ministère de l'Économie, des Finances et de l'Emploi, Ministère du Budget, des Comptes publics, de la Fonction Publique et de la Réforme de l'État, Paris, novembre, 244 pages.

Viveret P. (1989). *L'évaluation des politiques et des actions publiques, Propositions en vue de l'évaluation du Revenu minimum d'insertion,* Rapport au Premier Ministre, Documentation française, juin, 193 pages.

Direction générale du Trésor | Téléchargé le 27/10/2026 sur https://shs.cairn.info par Courtier Théophile (IP: 81.1.6.66)