



Secrétariat général
Service de la
coordination des
politiques culturelles
et de l'innovation
Département
des études,
de la prospective
et des statistiques

182, rue Saint-Honore, 75003 Paris cedex 01
☎ 01 40 15 73 17 – ☎ 01 40 15 73 99

Téléchargeable sur le site <http://www.culturecommunication.gouv.fr/Etudes-et-Statistiques> 2013-1

culture méthodes

Évaluation des politiques publiques : trois illustrations dans le champ culturel

Thibault Brodaty*

Public Policy Evaluation:
Three Illustrations in the Cultural Field

Les politiques publiques ont des objectifs précis et un coût supporté, au moins en partie, par les contribuables. Afin de garantir une utilisation optimale de la dépense publique, il semble légitime de s'assurer que les objectifs ont été atteints et que les bénéfices de la mesure sont supérieurs à son coût. Évaluer une mesure d'intervention publique est une tâche méthodologiquement délicate. Il s'agit en effet de comparer la situation observée une fois la mesure mise en place avec celle que l'on observerait en l'absence d'intervention publique. Or cette dernière, appelée « situation contrefactuelle », est inobservable. Tout l'enjeu des méthodes d'évaluation est de proposer une estimation crédible de cette situation virtuelle.

La pratique de l'évaluation se développe en France. Inscrite dans la loi constitutionnelle 2008-724 du 23 juillet 2008 de modernisation des institutions de la V^e République, la fonction d'évaluation des lois est confiée au Parlement, qui « vote la loi. [Il] contrôle l'action du Gouvernement [et] évalue les politiques publiques » (article 24). La Direction interministérielle pour la modernisation de l'action publique « est chargée de l'animation et de la coordination des travaux [...] d'évaluation des politiques publiques » (décret n° 2012-1198 du 30 octobre 2012 portant création du Secrétariat général pour la modernisation de l'action publique).

L'évaluation peut être de différentes natures : administrative ou économique (évaluation d'impact). L'évaluation administrative consiste à vérifier que la politique publique est mise en place conformément à la réglementation et à mesurer le nombre de bénéficiaires de la mesure, une information indispensable pour déterminer le coût réel de la mesure. Le nombre de bénéficiaires n'informe cependant pas sur le caractère efficace ou non de la politique publique et ne permet pas d'évaluer ce qu'aurait été la situation des bénéficiaires s'ils n'avaient pas profité de la mesure. Considérons par exemple le crédit d'impôt adressé aux sociétés de production phonographiques, dont le but est de favoriser l'émergence de nouveaux talents. Comptabiliser le nombre de sociétés bénéficiaires est insuffisant pour juger de l'efficacité de cette mesure. Par un effet d'aubaine, on peut en effet se retrouver dans une situation dans laquelle un grand nombre de sociétés ont bénéficié du crédit d'impôt, mais où peu d'entre elles ont produit plus de nouveaux talents que si elles n'en avaient pas bénéficié.

Il faut donc réaliser des évaluations d'impact pour juger de l'efficacité d'une mesure de politique publique. Elles sont de deux types :

- les évaluations prospectives ou *ex ante* sont réalisées avant que la politique publique soit mise en place, et servent à aider les pouvoirs publics à décider de la mise en place éventuelle de la mesure. Il s'agit

* Université Paris-Est Créteil, Université de la Sorbonne.

Évaluation des politiques publiques : trois illustrations dans le champ culturel

Public Policy Evaluation: Three Illustrations in the Cultural Field

Thibault Brodaty

Éditeur : Département des études, de la prospective et des statistiques

Lieu d'édition : Paris

Année d'édition : 2013

Date de mise en ligne : 21 septembre 2015

Collection : Culture méthodes

ISBN électronique : 9782111397996



<http://books.openedition.org>

Édition imprimée

Date de publication : 30 novembre 2013

Nombre de pages : 12

Référence électronique

BRODATY, Thibault. *Évaluation des politiques publiques : trois illustrations dans le champ culturel*.

Nouvelle édition [en ligne]. Paris : Département des études, de la prospective et des statistiques, 2013 (généré le 25 avril 2021). Disponible sur Internet : <<http://books.openedition.org/deps/336>>. ISBN : 9782111397996.

© Département des études, de la prospective et des statistiques, 2013

Creative Commons - Attribution - Pas d'Utilisation Commerciale 3.0 non transposé - CC BY-NC 3.0

Évaluation des politiques publiques : trois illustrations dans le champ culturel

Thibault Brodaty*

Public Policy Evaluation: Three Illustrations in the Cultural Field

Les politiques publiques ont des objectifs précis et un coût supporté, au moins en partie, par les contribuables. Afin de garantir une utilisation optimale de la dépense publique, il semble légitime de s'assurer que les objectifs ont été atteints et que les bénéfices de la mesure sont supérieurs à son coût. Évaluer une mesure d'intervention publique est une tâche méthodologiquement délicate. Il s'agit en effet de comparer la situation observée une fois la mesure mise en place avec celle que l'on observerait en l'absence d'intervention publique. Or cette dernière, appelée « situation contrefactuelle », est inobservable. Tout l'enjeu des méthodes d'évaluation est de proposer une estimation crédible de cette situation virtuelle.

La pratique de l'évaluation se développe en France. Inscrite dans la loi constitutionnelle 2008-724 du 23 juillet 2008 de modernisation des institutions de la V^e République, la fonction d'évaluation des lois est confiée au Parlement, qui « vote la loi. [II] contrôle l'action du Gouvernement [et] évalue les politiques publiques » (article 24). La Direction interministérielle pour la modernisation de l'action publique « est chargée de l'animation et de la coordination des travaux [...] d'évaluation des politiques publiques » (décret n° 2012-1198 du 30 octobre 2012 portant création du Secrétariat général pour la modernisation de l'action publique).

L'évaluation peut être de différentes natures : administrative ou économique (évaluation d'impact). L'évaluation administrative consiste à vérifier que la politique publique est mise en place conformément à la réglementation et à mesurer le nombre de bénéficiaires de la mesure, une information indispensable pour déterminer le coût réel de la mesure. Le nombre de bénéficiaires n'informe cependant pas sur le caractère efficace ou non de la politique publique et ne permet pas d'évaluer ce qu'aurait été la situation des bénéficiaires s'ils n'avaient pas profité de la mesure. Considérons par exemple le crédit d'impôt adressé aux sociétés de production phonographiques, dont le but est de favoriser l'émergence de nouveaux talents. Comptabiliser le nombre de sociétés bénéficiaires est insuffisant pour juger de l'efficacité de cette mesure. Par un effet d'aubaine, on peut en effet se retrouver dans une situation dans laquelle un grand nombre de sociétés ont bénéficié du crédit d'impôt, mais où peu d'entre elles ont produit plus de nouveaux talents que si elles n'en avaient pas bénéficié.

Il faut donc réaliser des évaluations d'impact pour juger de l'efficacité d'une mesure de politique publique. Elles sont de deux types :

- les évaluations prospectives ou *ex ante* sont réalisées avant que la politique publique soit mise en place, et servent à aider les pouvoirs publics à décider de la mise en place éventuelle de la mesure. Il s'agit

* Université Paris-Est Créteil, laboratoire ERUDITE.

notamment des méthodes de microsimulation ; elles s'appuient sur un modèle formalisé de l'économie qui permet d'évaluer les effets de la mesure, en tenant compte des changements induits dans les comportements des agents économiques. Ces changements sont néanmoins des hypothèses du modèle, et il est donc nécessaire de tester l'éventail des changements possibles ;

- les évaluations rétrospectives ou *ex post*, qui permettent d'évaluer l'effet de la mesure après sa mise en place.

Trois illustrations d'évaluations *ex post* sont présentées ici, associée chacune à une méthode d'évaluation différente. L'effet de la politique est défini par un critère que l'on appelle « estimateur ». Sa valeur est calculée grâce aux données recueillies, elle est appelée « estimation ». Ce critère sera supposé unique, sinon l'évaluateur devra effectuer autant d'évaluations distinctes que de critères s'il est amené à retenir plusieurs critères d'évaluation. Alors que les méthodes de statistiques descriptives permettent de décrire les données et d'établir des corrélations, l'économétrie permet de mettre en évidence non pas seulement des corrélations mais des relations de cause à effet.

En reprenant l'exemple du crédit d'impôt, de simples moyennes montreront par exemple que les sociétés aidées produisent plus de nouveaux talents que les sociétés non bénéficiaires. Mais cette corrélation ne peut s'interpréter de manière causale : est-ce l'effet du crédit d'impôt ou les sociétés aidées sont-elles intrinsèquement plus innovantes, indépendamment du crédit d'impôt ? Cette statistique descriptive, appelée « estimateur naïf », ne permet pas de répondre à la question, la corrélation ainsi mise au jour ne traduisant pas de relation de cause à effet. Pour obtenir une interprétation causale, il faut raisonner toutes choses égales par ailleurs : ne pas comparer de manière brute les sociétés aidées et les sociétés non aidées, mais comparer des sociétés aidées et non aidées qui ont la même propension initiale à innover. Dès lors, les différences moyennes observées entre bénéficiaires et non-bénéficiaires pourront être attribuées en totalité au crédit d'impôt : raisonner toutes choses égales par ailleurs permet ainsi d'isoler les effets. Lorsque l'on pense disposer de la totalité des données nécessaires, on peut alors utiliser les méthodes économétriques classiques comme celles des moindres carrés ordinaires ou des méthodes plus évoluées comme les méthodes par appariement (*matching*). La plupart du temps cependant, les données ne sont pas suffisantes et il faut recourir à d'autres méthodes qui sont présentées dans ce texte.

Ces méthodes microéconométriques (différences de différences, variables instrumentales et régressions avec discontinuité) ont connu des développements méthodo-

logiques importants ces vingt dernières années et sont aujourd'hui mobilisables. L'objectif est de les présenter ici sans formalisation, en insistant sur les hypothèses qui fondent chacune d'entre elles.

TROIS EXEMPLES DE POLITIQUE PUBLIQUE

Trois dispositifs culturels de nature différente (labellisation, dispositif d'éducation artistique et culturelle et crédit d'impôt) illustrent l'exposé : le label « librairie indépendante de référence » (LIR), le programme « École et cinéma » et le crédit d'impôt audiovisuel (CIA). Ces trois dispositifs ont l'avantage de couvrir le champ des politiques de réglementation (label LIR), des politiques agissant sur le capital humain des individus (École et cinéma) et des aides fiscales (CIA). Plusieurs méthodes peuvent être mobilisées pour évaluer ces politiques publiques, mais par souci de clarté et de concision, une seule méthode sera présentée pour chaque mesure.

Le label LIR

La situation économique des librairies indépendantes, dont plusieurs études¹ ont souligné la fragilité au cours des années 2000, s'est récemment dégradée en raison d'une concurrence démultipliée (grandes surfaces, internet), de la hausse du foncier et de celle des coûts de transports. Le législateur considérant que les librairies indépendantes contribuent à la diversité de l'offre éditoriale et participent de l'aménagement culturel du territoire, le label LIR (Librairie indépendante de référence) a été inscrit dans la loi de finances rectificative du 25 décembre 2007, pour « reconnaître, valoriser et soutenir les engagements et le travail qualitatifs des libraires indépendants ».

Ce label distingue les librairies indépendantes pour la diversité de leur offre et pour leurs efforts en matière d'animation culturelle. Au-delà du signal de qualité ainsi envoyé aux consommateurs, le label LIR ouvre la possibilité d'une exonération de la contribution économique territoriale (CET²) sous réserve d'une délibération expresse des collectivités locales en ce sens, et permet une instruction simplifiée de la demande auprès du Centre national du livre (CNL) pour obtenir une subvention au titre de la mise en valeur des fonds (aide VAL) ; il incite enfin les éditeurs qui ont fait le choix d'intégrer le label LIR parmi leurs critères de calcul de la remise qualitative, à accorder aux librairies labellisées des conditions commerciales plus favorables³.

1. Sophie BARLUET, Rapport Livre 2010 – Pour que vive la politique du livre, Ministère de la Culture et de la Communication, 2007 ; Antoine GALLIMARD, Rapport de la mission de réflexion sur la librairie indépendante, Ministère de la Culture et de la Communication, 2007 ; Rapport sur la situation économique des librairies, Ministère de la Culture et de la Communication, Observatoire de l'économie du livre, 2007.

2. La contribution économique territoriale a remplacé la taxe professionnelle depuis le 1^{er} janvier 2010 (loi de finances 2009-1673 du 30 décembre 2009 pour 2010).

3. Voir le lien ci-dessous pour de plus amples informations :

http://www.centrenationaldulivre.fr/fr/librairie/lr_un_label_de_reference/presentation/

D'un point de vue économique, on peut attendre au moins deux effets du label. D'une part, l'effet signal pourrait augmenter la demande des consommateurs. D'autre part, les éventuels avantages financiers retirés de l'attribution du label pourraient permettre aux librairies labellisées de retrouver une rentabilité plus satisfaisante.

Évaluer ce dispositif ne peut pas se limiter à comptabiliser le nombre de librairies bénéficiant du label LIR. La question sera posée de savoir si le dispositif mis en place a été suffisant pour permettre aux librairies labellisées de faire face aux conditions de marché dégradées. Parmi les différents critères possibles, on peut penser en premier lieu au taux de rentabilité des librairies. Cet indicateur est néanmoins imparfait car il faudrait alors s'intéresser précisément à la manière dont les librairies labellisées utilisent les avantages financiers retirés du label. Mais l'objectif étant ici purement illustratif, on retiendra ce critère simplifié, en gardant à l'esprit ses limites. Pour notre exemple, la question d'évaluation consistera à se demander si le label a permis à certaines librairies d'améliorer leur taux de rentabilité. On cherchera précisément à comparer le taux de rentabilité moyen des librairies labellisées avec celui qu'elles afficheraient si elles n'avaient pas bénéficié du dispositif. Cette mesure servira d'exemple d'application lors de l'exposé de la méthode des différences de différences.

École et cinéma

Initié en 1994, École et cinéma est un dispositif national d'éducation artistique au cinéma. Il est issu d'un partenariat entre le Centre national du cinéma et de l'image animée et le ministère et de l'Éducation nationale, mis en œuvre par l'association Les enfants du cinéma. Ce dispositif permet aux enseignants des écoles primaires d'inscrire dans leur programme pédagogique, durant le temps scolaire, des séances de cinéma pour faire découvrir à leurs élèves des films de qualité, visionnés en salle. Les enseignants ont également l'opportunité de compléter leur formation et bénéficient de ressources pour faciliter l'accès de leurs élèves aux œuvres du catalogue.

Tel que décrit dans le cahier des charges⁴, l'objectif général de cette initiative est de « faire découvrir aux écoliers le cinéma en tant qu'art. [...] ces rencontres avec les films ainsi que les activités que les enseignants proposent favorisent la formation du jugement, du goût et de la sensibilité ».

Déterminer un critère d'évaluation mesurable qui traduise la découverte, la formation du jugement, le goût ou la sensibilité n'est pas aisé. Si l'on s'en tient à l'objectif général du programme qui est de « faire découvrir le cinéma en tant qu'art », on peut considérer que le but principal est atteint si le goût des enfants pour le cinéma

s'en est trouvé développé. Le goût pour le septième art se mesurant par exemple par la fréquentation des salles obscures, l'évaluateur cherchera à comparer la fréquentation moyenne annuelle, à dix-huit ans par exemple, des enfants qui ont bénéficié du programme, avec celle qu'ils auraient eue s'ils n'en avaient pas bénéficié. Ce programme servira à illustrer la méthode des variables instrumentales.

Le crédit d'impôt audiovisuel

Le crédit d'impôt audiovisuel (CIA) a été inscrit dans la loi de finances rectificative du 30 décembre 2004. Géré par le Centre national du cinéma et de l'image animée, son objectif est de favoriser la localisation française des dépenses de production. Il permet à une société de production de déduire de son imposition 20 % de certaines dépenses de production dites dépenses éligibles, sous un plafond de 1 150 euros par minute pour une œuvre de fiction ou un documentaire, et de 1 200 euros par minute pour une œuvre d'animation. Pour être éligible au CIA, l'œuvre doit, entre autres, être réalisée en langue française ou dans une langue régionale, principalement sur le territoire français, et doit contribuer au développement de la création de la production audiovisuelle française et à sa diversité.

Le respect des conditions de création des œuvres audiovisuelles est vérifié au moyen d'un barème de 100 points fixé en référence à l'arrêté du 22 mars 1999 relatif au soutien financier de l'industrie cinématographique. Pour être considérées comme réalisées principalement sur le territoire français, les œuvres doivent obtenir au moins la majorité des points, hors ceux affectés à la langue. Ces barèmes sont différents pour les fictions, les documentaires et les films d'animation. Par exemple pour les œuvres de fiction, le barème porte sur sept thèmes : entreprise de production (10 points), langue de tournage (20 points), auteurs (10 points), artistes interprètes (20 points), techniciens collaborateurs de création (14 points), ouvriers (6 points) et tournage et postproduction (20 points).

L'objectif de ce crédit d'impôt étant de favoriser la localisation française des dépenses de production, on pourrait, dans le cas de la fiction, utiliser le nombre de jours de tournage effectués en France comme critère d'évaluation.

Évaluer le crédit d'impôt audiovisuel reviendrait ainsi précisément à comparer le nombre de jours de tournage effectués en France des sociétés bénéficiant du crédit d'impôt avec le nombre de jours de tournage qu'auraient effectué ces mêmes sociétés en France si elles n'en avaient pas bénéficié. Cette mesure fiscale permettra de présenter la méthode de la régression avec discontinuité.

4. Voir <http://www.cnc.fr/web/fr/ecole-et-cinema>

CADRE THÉORIQUE DE L'ÉVALUATION

Le cadre théorique permet de définir précisément la question d'évaluation, les paramètres d'intérêt auxquels il donne lieu, et d'envisager le biais de sélection qui invalide l'utilisation de l'estimateur naïf. Les expérimentations aléatoires permettent de régler directement ce problème mais leurs limites seront également envisagées.

Le modèle causal

Le problème de l'évaluation a été formalisé par Donald B. Rubin⁵. Pour en présenter le cadre, rappelons quelques définitions illustrées par l'exemple de l'évaluation du label LIR. Dans une « population d'intérêt » composée d'« individus », l'évaluateur s'intéresse à l'« effet causal » d'un « traitement » sur une « variable de résultat ». Dans le cas du label LIR, la population d'intérêt est formée par l'ensemble des librairies indépendantes françaises, un individu est représenté par une librairie indépendante, le traitement désigne le fait de bénéficier du label et la variable de résultat est, dans notre cadre d'analyse simplifié, le taux de rentabilité de la librairie. La population d'intérêt est divisée en deux groupes : la partie de la population qui a reçu le traitement, appelée groupe des traités, et celle qui n'a pas reçu le traitement, appelée groupe des non-traités ou groupe de contrôle ou encore groupe témoin, en référence au vocabulaire médical. Dans le cas du label LIR, les traités sont les librairies qui ont bénéficié du label, et le groupe de contrôle est composé des librairies qui n'en ont pas bénéficié.

L'effet causal du traitement sur un individu est la différence entre son résultat s'il est traité et son résultat s'il ne l'est pas. Comme ils sont conditionnés par la réalisation du traitement, les deux résultats dont on cherche à apprécier la différence sont appelés résultats potentiels. Dans l'exemple du label LIR, les deux résultats potentiels d'une librairie sont d'une part son taux de rentabilité dans un monde où elle aurait été labellisée, et d'autre part son taux de rentabilité dans un monde où elle ne l'aurait pas été. L'effet causal du traitement pour cette librairie, différence entre ces deux résultats potentiels, compare les taux de rentabilité de la librairie dans ces deux mondes qui ne peuvent pas se produire en même temps. Dans le cas d'une librairie qui aura augmenté sa rentabilité grâce au label alors qu'elle aurait stagné ou diminué en l'absence de labellisation, on pourra considérer, selon l'indicateur retenu, que le label aura été efficace. En revanche, pour une librairie dont la labellisation n'aura pas empêché une baisse de rentabilité, le dispositif n'aura pas été efficient. En toute généralité, l'effet causal est hétérogène, c'est-à-dire qu'il dépend de chaque individu : l'effet du label LIR peut dépendre du comportement

et des caractéristiques des librairies, et pourrait ainsi être plus ou moins bénéfique selon les librairies. Lorsque l'effet causal est identique pour tous les individus, on dit qu'il est homogène.

On retient traditionnellement deux paramètres d'intérêt : l'effet moyen du traitement sur les traités (noté TT, *Treatment on the Treated*), et l'effet moyen du traitement dans la population totale (noté ATE, *Average Treatment Effect*). Dans le cas du label LIR, TT représente la différence entre le taux de rentabilité moyen des librairies labellisées et le taux de rentabilité moyen non observé des librairies labellisées si elles ne l'avaient pas été. ATE est composé des mêmes moyennes, mais calculées parmi l'ensemble des librairies indépendantes. Si l'on s'intéresse par exemple à TT, une estimation positive signifiera que les librairies labellisées ont eu en moyenne une meilleure rentabilité que si elles n'avaient pas été labellisées. On pourra alors considérer dans ce cas que cette politique publique est efficace et qu'il faut la poursuivre. Tout l'enjeu de l'évaluation est de trouver des méthodes qui permettent d'estimer correctement le contrefactuel, par nature inobservable : comment déterminer le taux de rentabilité moyen des librairies labellisées si elles n'avaient pas bénéficié du label ?

Estimateur naïf et biais de sélection

La première idée que l'on peut avoir pour estimer les contrefactuels est d'utiliser simplement ce qui est observable. TT est égal par définition à la différence entre le résultat moyen observé des traités et le résultat moyen des traités s'ils n'avaient pas été traités. Ce dernier étant un contrefactuel, inobservable, il est naturel de l'estimer par le résultat observé des non-traités. On définit ainsi l'estimateur naïf par la différence entre le résultat moyen observé des traités et le résultat moyen observé des non-traités. Dans le cas du label LIR, l'estimateur naïf consisterait à comparer le taux de rentabilité moyen des librairies labellisées avec le taux de rentabilité moyen des librairies qui n'ont pas bénéficié du label.

L'estimateur naïf estime correctement TT s'il est réaliste de penser que les traités auraient eu en moyenne, s'ils n'avaient pas été traités, le même résultat que les non-traités. Dans le cas du label LIR, l'estimateur naïf serait correct si, en l'absence de labellisation, les librairies labellisées affichaient le même taux de rentabilité que les librairies non bénéficiaires du label. Cette hypothèse est-elle réaliste ? Les librairies sélectionnées prennent en moyenne plus de risques sur leurs fonds, forment mieux leurs salariés et participent davantage à l'animation culturelle de leur quartier. Or il est possible que ces investissements de long terme fragilisent à court terme les librairies. Selon cette hypothèse, on peut penser qu'en l'absence de labellisation, les librairies seraient en moins bonne santé économique que les librairies non labellisées.

5. Donald B. RUBIN, "Estimating the Causal Effects of Treatments in Randomized and Nonrandomized Studies", *Journal of Educational Psychology*, 1974, 66(5), p. 688-701.

MÉTHODES MICROÉCONOMÉTRIQUES D'ÉVALUATION DES POLITIQUES PUBLIQUES

Le taux de rentabilité dans les librairies non labellisées paraît ainsi être une mauvaise approximation du contre-factuel. Il le surestime, de sorte que l'estimateur naïf sous-estime probablement l'effet moyen du label sur les librairies labellisées.

Par définition, si les traités, dans l'hypothèse où ils n'auraient pas été traités, avaient eu un résultat différent des non-traités, ou, autrement dit, si le résultat contre-factuel des traités est différent du résultat observé des non-traités, on dit qu'il existe un biais de sélection. La présence de ce biais est la situation la plus courante lorsque l'on évalue des politiques publiques. Il faut ainsi systématiquement se poser la question de la présence d'un biais de sélection, qui interdit l'utilisation de l'estimateur naïf.

Les expérimentations aléatoires ou contrôlées

Il existe un cas dans lequel l'évaluateur n'a pas à faire face au problème de sélection, celui des expérimentations aléatoires, aussi appelées expériences contrôlées. Il s'agit de s'inspirer par exemple de la médecine, et de tester l'efficacité d'une politique publique comme on teste l'efficacité d'un médicament. Tirer au hasard des individus et leur affecter le traitement, en tirant d'autres et ne pas leur affecter le traitement. En raison du tirage aléatoire, le résultat moyen des traités, s'ils n'avaient pas été traités, aurait été le même que celui des non-traités. Il n'y a donc pas de biais de sélection et l'estimateur naïf peut être tout simplement utilisé pour estimer l'effet du traitement. Dans le cas du label LIR, il s'agirait ainsi de faire bénéficier du label certaines librairies tirées au hasard puis de comparer le taux de rentabilité des librairies labellisées avec celui des librairies non labellisées. Cette méthode est couramment utilisée en économie du développement par exemple. Les pouvoirs publics français ont également financé un certain nombre d'évaluations fondées sur cette méthode dans le cadre du Fonds d'expérimentation pour la jeunesse⁶, créé en 2009. On peut citer à titre d'exemple l'évaluation du programme « Bien vu ! », dont l'objectif était de « permettre à des jeunes éloignés des structures culturelles d'apprendre à décrypter les images pour que chaque jeune citoyen puisse se forger un regard personnel et construit sur les images qui les entourent ».

Si la méthode est séduisante car elle permet de s'affranchir de manière satisfaisante du biais de sélection, elle comporte néanmoins des inconvénients non négligeables. Elle induit notamment des coûts d'enquête et de gestion importants et elle prive des individus d'un traitement qui pourrait les aider, pouvant générer un sentiment d'injustice. Elle reste malgré tout une alternative crédible dans le cadre d'un test, lorsque l'on souhaite mesurer les effets d'une politique publique avant de prendre la décision de la généraliser.

Trois méthodes microéconométriques classiques permettant de prendre en compte le biais de sélection sont présentées ici. La première, les différences de différences, s'appuie sur l'idée qu'il est possible de suivre les mêmes individus avant et après l'introduction de la politique publique. La deuxième, les variables instrumentales, profite de la disponibilité d'une variable corrélée avec le traitement, mais pas directement avec la variable de résultat. La troisième, la régression avec discontinuité, utilise l'existence d'un seuil au-delà duquel la probabilité d'être traité change brusquement.

Différences de différences et label Librairie indépendante de référence

Toute la difficulté des différences de différences est de trouver un groupe de contrôle crédible qui permette de rendre réaliste l'hypothèse qui fonde cette méthode. Comme on l'a vu précédemment, le biais de sélection vient en effet du fait que les traités et les non-traités sont, pour des raisons pour partie inobservables, différents vis-à-vis de la variable de résultat.

Pour éviter d'avoir à comparer directement des traités et des non-traités, une première solution consiste à comparer les mêmes individus à des dates différentes. Il faut pour cela observer les résultats des traités avant et après l'introduction du traitement. On peut alors penser à estimer le contre-factuel (le résultat moyen des traités s'ils n'avaient pas été traités) par leur résultat moyen avant l'introduction du traitement, moment auquel les individus n'ont, par définition, pas encore reçu le traitement. On définit ainsi « l'estimateur en simple différence » par la différence entre les résultats moyens des traités, avant et après l'introduction du traitement. Si le résultat moyen observé des traités avant l'introduction du traitement estime correctement le résultat moyen qu'ils auraient en l'absence de traitement, alors l'estimateur en simple différence estime correctement l'effet moyen du traitement sur les traités. Les premiers labels LIR ont été attribués en 2009. Supposons que l'on observe en 2008 et 2010 le taux de rentabilité des librairies labellisées en 2009. L'estimateur en simple différence revient alors à calculer l'évolution du taux de rentabilité des librairies labellisées entre 2008 et 2010. Cette simple différence estime correctement l'effet moyen du label LIR sur les librairies labellisées s'il est raisonnable de supposer qu'en l'absence de labellisation, les librairies labellisées auraient eu le même taux de rentabilité en 2010 qu'en 2008. Or la conjoncture économique est différente en 2008 et en 2010. L'estimateur en simple différence

6. Voir www.experimentation.jeunes.gouv.fr

mélange ainsi de manière générale l'effet du traitement et l'effet de la conjoncture. Il faut donc purger de cet estimateur l'effet de la conjoncture pour estimer correctement l'effet moyen du traitement sur les traités. Pour cela, on utilise les individus non traités : ces individus non traités n'ont pas été affectés par le traitement, ils n'ont été soumis qu'à l'évolution de la conjoncture. Pour prendre en compte la conjoncture, l'idée est alors d'utiliser l'évolution du résultat moyen des non-traités pour estimer l'évolution du résultat moyen des traités s'ils n'avaient pas été traités. Si cette hypothèse est réaliste, on peut alors estimer l'effet moyen du traitement sur les traités par double différence, en retranchant l'évolution observée du résultat moyen des non-traités de celle des traités : cet estimateur est appelé « estimateur en double différence ».

Dans le cas du label LIR, l'estimateur des doubles différences consisterait à retrancher l'évolution observée du taux de rentabilité moyen des librairies non labellisées entre 2008 et 2010 de celle des librairies labellisées. Cet estimateur estime-t-il correctement l'effet moyen du label LIR sur les librairies labellisées ? Oui s'il est réaliste de supposer qu'en l'absence de labellisation, l'évolution du taux de rentabilité entre 2008 et 2010 dans les librairies labellisées aurait été identique à celle observée dans les librairies non labellisées.

Or, à nouveau, le label est accordé à des librairies dont on a fait l'hypothèse que la situation économique est plus fragile à court terme. Si la conjoncture s'est détériorée entre 2008 et 2010, on peut supposer que les librairies labellisées auraient moins bien résisté que les librairies non labellisées, et que leur taux de rentabilité se serait plus dégradé que pour les librairies non labellisées entre 2008 et 2010. L'hypothèse sur laquelle se fonde la méthode des doubles différences paraît alors elle aussi discutable lorsque l'on considère l'ensemble des libraires non labellisées comme groupe de contrôle.

Il faut donc essayer de restreindre le groupe des non-traités pour rendre réaliste cette hypothèse. En 2011, le dispositif du label a été complété par la création d'un label « Librairie de référence » qui permet à des librairies non indépendantes en termes de capital mais indépendantes en termes de gestion d'être labellisées. Le label Librairie de référence offre l'ensemble des avantages du label LIR à l'exception de la possibilité d'être exonéré de contribution économique territoriale. Ce groupe de librairies a donc bénéficié du label en 2011 mais n'était pas labellisé avant. Également sélectionnées et donc potentiellement ressemblantes aux librairies labellisées en 2009, il est légitime de considérer ces librairies comme un groupe de contrôle possible. Cependant, elles font bien souvent partie de grands groupes. Leur degré de résistance à la crise est donc potentiellement différent de celui des librairies labellisées en l'absence de label. Il n'est donc en fait pas souhaitable d'utiliser ce groupe comme groupe de contrôle.

Il serait préférable d'utiliser, par exemple, les librairies qui se sont vues refuser le label en 2009 mais l'ont obtenu après 2010. En effet, ces librairies ressemblent probablement davantage aux librairies labellisées en

2009 que celles qui ne l'ont jamais été, et il deviendrait alors plus raisonnable de supposer que le taux de rentabilité des librairies labellisées en 2010 aurait évolué, sans labellisation, comme celui de ces librairies. En pratique, le choix du groupe de contrôle est validé en s'assurant qu'avant la mise en place de la mesure, les traités et les membres du groupe de contrôle évoluent en moyenne de manière similaire.

Après avoir estimé l'effet moyen du traitement sur les traités, τ , la question se pose d'estimer l'effet moyen du traitement sur toute la population d'intérêt, ATE. On cherche ainsi à mesurer quel est l'effet du label LIR non plus seulement sur les librairies labellisées mais aussi sur les librairies qui n'ont pas bénéficié du label. Mais l'objectif du label étant de consolider les librairies engagées dans une démarche de qualité et d'inciter les autres à les rejoindre, la question de l'effet du label sur les librairies qui ne s'inscriraient pas dans une telle démarche n'est en fait pas pertinente.

Il faut également être attentif au fait que les individus peuvent anticiper l'introduction d'une politique publique. Bien que peu probable dans ce cas, l'inscription du label LIR dans la loi de finances de 2007 pourrait avoir incité certaines librairies à s'engager dans une démarche de qualité au moment de l'annonce de la mise en place du dispositif. Considérer ces librairies comme non traitées en 2008 poserait donc problème puisque avant de recevoir effectivement le label, certaines librairies auraient déjà modifié leur comportement. Une solution serait en définitive de déterminer une date d'observation suffisamment éloignée de la mise en place du traitement.

Variables instrumentales et programme École et cinéma

Rappelons que l'estimateur naïf, qui consiste à comparer le résultat moyen des traités et des non-traités, n'estime pas correctement l'effet moyen du traitement sur les traités s'il existe un biais de sélection : si le résultat moyen des traités et des non-traités est différent, il est impossible de savoir si c'est l'effet du traitement ou si cela traduit le fait que les traités et les non-traités sont différents vis-à-vis de caractéristiques inobservées par l'économètre qui ont également une influence sur la variable de résultat.

Dans le cas du programme École et cinéma, la variable de résultat que nous avons choisie est la fréquentation annuelle des salles obscures et la population d'intérêt est la population française. L'âge et la génération étant potentiellement un déterminant de la fréquentation, il est nécessaire de fixer ces deux variables. Par souci de clarté, l'étude sera restreinte aux individus âgés de 18 ans en 2012, c'est-à-dire nés en 1994, date de lancement du programme. Parmi eux, on trouvera des individus ayant participé au programme et des individus n'y ayant pas participé. L'estimateur naïf consisterait à comparer la fréquentation moyenne à 18 ans des participants et des non-participants. Mais la participation à ce programme n'est pas complètement aléatoire. En effet, par-

ticiper au programme École et cinéma est un projet de classe et dépend donc du rôle de l'enseignant dans la décision de postuler. Un enseignant postulera d'autant plus qu'il a lui-même un goût prononcé pour le cinéma, ainsi les classes bénéficiaires ont-elles probablement un enseignant dont le goût pour le cinéma est plus prononcé que la moyenne. Or le goût de l'enseignant pour le cinéma est une variable potentiellement inobservée par l'économètre, et qui a une influence sur la participation au programme mais aussi potentiellement sur la fréquentation future des salles de cinéma. Si, à dix-huit ans, les bénéficiaires vont plus souvent au cinéma que les non-bénéficiaires, il n'est pas possible de savoir si c'est en raison du programme ou si c'est parce qu'ils ont eu un enseignant qui leur a transmis son goût pour le cinéma. Pour le dire autrement, si les bénéficiaires n'avaient pas participé au programme, ils auraient probablement une fréquentation plus importante que les non-bénéficiaires en raison du goût de leur enseignant pour le cinéma. La présence de biais de sélection est donc probable.

L'idée de la méthode des variables instrumentales est de trouver un artefact, un « objet statistique », un instrument-témoin qui permette de mesurer l'effet de cet éventuel biais de sélection, de témoigner de son existence. Il nous faut donc l'aide d'une « source de variation exogène », c'est-à-dire une variable qui fasse varier le traitement (la fréquence des séances de cinéma pendant l'école), mais qui n'ait pas d'effet direct sur la variable de résultat (la fréquentation ultérieure des salles de cinéma). Cette variable permet alors de comparer des traités et des non-traités, qui ont par ailleurs les mêmes déterminants du résultat. Cette source de variation exogène est appelée « variable instrumentale » ou « instrument » (par exemple des coûts de transport pour aller de l'école au cinéma qui sont soit faibles soit élevés). Elle peut être continue ou discrète, mais pour faciliter la présentation intuitive de cette méthode, nous allons supposer que l'instrument est binaire, c'est-à-dire qu'il ne retient que deux valeurs, 0 et 1. Les groupes d'individus qui ont comme valeur de l'instrument 0 et 1 seront appelés respectivement sous-populations 0 et 1. Cette variable doit vérifier deux propriétés. Elle doit tout d'abord être corrélée au traitement, c'est-à-dire que le pourcentage de traités (les élèves emmenés au cinéma par leur professeur) doit être différent dans les sous-populations 0 et 1. Cette première propriété est facile à vérifier lorsque l'on dispose des données. En revanche, la seconde l'est moins. Il s'agit de s'assurer que l'instrument ne doit pas avoir d'effet direct sur la variable de résultat, c'est-à-dire la fréquentation à l'âge de dix-huit ans. Elle peut avoir un effet indirect, qui passe par le traitement, mais pas d'effet direct. Cette hypothèse ne peut pas toujours être testée. Il faut alors mobiliser des arguments théoriques ou l'expérience des acteurs de terrain pour se convaincre ou non de sa validité.

Intuitivement, la méthode des variables instrumentales revient à comparer les résultats moyens dans les sous-populations 0 (celle pour laquelle l'instrument vaut 0) et 1 (celle pour laquelle l'instrument vaut 1). En effet, si l'on observe une différence de résultat moyen entre ces

deux sous-populations, ce ne sera pas à cause de l'instrument (le coût de transport), qui n'a pas d'effet direct sur le résultat, mais parce que les pourcentages de traités y seront différents.

Dans le cas du programme École et cinéma, l'instrument doit donc être une variable qui explique la participation au programme, mais qui n'explique pas directement la future fréquentation des salles de cinéma. Comme on l'a vu, le goût pour le cinéma des enseignants peut venir expliquer la participation au programme de leurs élèves. Mais cette variable peut aussi expliquer la fréquentation des salles des bénéficiaires à dix-huit ans et ne peut donc pas être utilisée comme instrument. D'autres variables peuvent expliquer néanmoins la participation, par exemple des variables financières. Le déplacement des classes implique un coût financier (billet de cinéma et transport), particulièrement important dans les zones rurales du fait de l'éloignement des cinémas. Lorsqu'il est trop élevé, le coût du transport peut expliquer que certaines classes ne trouvent pas les financements suffisants pour participer au programme. Le coût du transport paraît donc être un bon candidat pour faire office d'instrument. Définissons donc une variable à deux modalités qui informe sur le caractère faible ou important des coûts de transport, et qui définit ce qu'on a appelé les sous-populations 0 et 1. Une mesure de l'effet moyen d'École et cinéma consistera alors à déterminer d'une part la différence de fréquentation moyenne entre les élèves qui ont fait face à des coûts de transport importants et ceux qui ont fait face à des coûts modérés, et d'autre part la différence entre la part des élèves bénéficiaires du programme parmi ceux qui ont fait face à des coûts élevés et la part des élèves bénéficiaires parmi ceux qui ont fait face à des coûts faibles. Il faudra alors former le rapport de ces deux quantités. Cette mesure est appelée estimateur de Wald. Si les individus « à bas coûts » ont une fréquentation moyenne plus importante à 18 ans que les autres, ce ne sera pas directement grâce au coût moins élevé du transport mais parce que ce faible coût leur aura permis de participer au programme. Si l'effet de ce programme est supposé homogène, alors l'estimateur de Wald estime bien ce qu'on cherche à estimer. Si l'on pense au contraire que l'effet est hétérogène entre les élèves, alors cet estimateur estime seulement l'effet moyen sur les enfants qui participent au programme lorsque les coûts sont faibles et qui n'y participent pas lorsque les coûts sont élevés. On ne peut dans ce cas rien dire sur les participants au programme (TT) et il faut alors faire en sorte de mesurer l'effet de l'expérience sur des populations homogènes, parmi tous les enfants. En pratique, l'hétérogénéité de l'effet du traitement sera testée en appliquant distinctement la méthode à des sous-échantillons définis par des caractéristiques exogènes de type sociodémographiques par exemple. Si l'effet s'avère être hétérogène, on prendra alors moins de risque à supposer que l'effet est homogène au sein de chaque sous-échantillon ainsi constitué et on pourra alors y mesurer l'effet de la politique à évaluer.

La méthode des variables instrumentales repose donc sur la qualité de l'instrument. Au-delà des hypothèses qui

le définissent, il faut également que l'instrument (le coût de transport) explique suffisamment le traitement, c'est-à-dire qu'il faut que la proportion de traités dans la sous-population 0 (les enfants qui vont au cinéma grâce à l'école, parmi ceux qui bénéficient de coûts de transport faibles) soit la plus différente possible de la proportion de traités dans la sous-population 1 (les enfants qui ont des coûts de transport élevés). Dans le cas contraire, l'instrument est dit faible et donne alors lieu à des estimations de mauvaise qualité. Signalons pour finir que cette méthode a été présentée dans le cas particulier où la variable de traitement (bénéficiaire ou non du programme École et cinéma) et l'instrument (prix élevé ou faible du coût du transport au moment du programme) sont des variables binaires. Cette méthode est en fait plus générale et s'applique aussi lorsque ces variables ne sont plus binaires.

Régressions avec discontinuité et crédit d'impôt audiovisuel

Supposons maintenant que l'on observe une variable continue, appelée score. Supposons que le score soit tel que lorsqu'il franchit un certain seuil, la probabilité d'être traité se modifie de manière discontinue, brusque. Un cas particulier, le cas à « discontinuité nette » (*sharp design* en anglais), correspond à la situation dans laquelle, lorsque le score est inférieur au seuil, l'individu n'est pas traité, et lorsque le score est supérieur, l'individu est traité. Dans ce cas, la probabilité d'être traité varie de manière discontinue au passage du seuil, puisqu'elle passe de 0 à 1. Lorsqu'au passage du seuil la probabilité d'être traité varie de manière discontinue, mais sans qu'il soit certain de ne pas être traité en dessous du seuil et certain de l'être au-dessus, on appelle ce cadre le cas de la « discontinuité floue » (*fuzzy design*).

Considérons tout d'abord le cas de la discontinuité nette. L'idée de la méthode de la régression avec discontinuité est de zoomer sur le voisinage immédiat du seuil. Concentrons-nous donc sur les individus qui ont une valeur du score extrêmement proche du seuil. Pour un score qui prendrait ses valeurs entre 0 et 100 avec un seuil de 50, on pourrait par exemple se restreindre aux individus qui ont un score compris entre 48 et 52. Parmi ces individus, certains sont en dessous de 50 et ne sont donc pas traités, les autres sont au-dessus et sont donc traités. Lorsque l'on considère les individus qui ont un score compris entre 48 et 52, il paraît raisonnable de penser que deux individus, un à 49 et l'autre à 51, ont été distribués autour de 50 de manière purement aléatoire. Lorsque cette hypothèse est réaliste, on peut considérer que l'affectation dans le traitement a été faite de manière aléatoire, à l'image d'une expérimentation aléatoire. Le résultat moyen des non-traités dans la fenêtre peut alors être utilisé pour estimer le résultat moyen des traités s'ils n'avaient pas été traités (contrefactuel). L'estimateur naïf fournit alors, dans la fenêtre, une estimation correcte de l'effet moyen du traitement.

L'idée que, dans la fenêtre autour du seuil, l'affectation dans le traitement puisse être considérée comme aléatoire reste l'hypothèse qui fonde la méthode dans le cas de la discontinuité floue. Mais, à l'inverse du cas de discontinuité nette, dans la fenêtre, les individus au-dessus du seuil ne sont pas tous traités et les individus en dessous ne sont pas tous non traités. Ainsi, être au-dessus ou en dessous du seuil explique seulement en partie la probabilité d'être traité, mais n'explique pas en soi le résultat puisque la répartition autour du seuil peut être considérée comme aléatoire. Au regard de la section précédente sur la méthode des variables instrumentales, la variable qui informe sur le fait d'être au-dessus ou en dessous du seuil est ainsi un bon instrument du point de vue du traitement et du résultat. Dans le cas de la discontinuité floue, l'effet moyen du traitement peut ainsi être estimé correctement par la méthode des variables instrumentales, avec l'instrument ainsi défini.

Dans le cas de l'évaluation du crédit d'impôt audiovisuel, l'estimateur naïf consisterait à comparer le nombre de jours de tournage sur le territoire français des œuvres bénéficiant du CIA et de celles qui n'en bénéficient pas. En l'absence de crédit d'impôt audiovisuel, les œuvres aidées auraient-elles été tournées en moyenne autant sur le territoire français que l'ont été les œuvres non bénéficiaires ? Certaines sociétés de production décident de tourner à l'étranger pour réduire les coûts. Si le tournage coûte moins cher, il existe cependant un coût d'entrée, ou un coût d'apprentissage, lors de la ou des premières délocalisations dans la mesure où il faut évoluer dans un environnement inconnu. Les sociétés qui ont déjà franchi ce pas avant l'introduction du crédit d'impôt ont donc déjà payé ce coût d'entrée, sont donc incitées à continuer à délocaliser les tournages et n'ont dès lors peut-être même pas postulé au crédit d'impôt audiovisuel. Inversement, les sociétés qui ont moins l'habitude de délocaliser font face à un coût supplémentaire pour tourner à l'étranger, elles n'ont pas totalement payé le coût d'entrée et n'ont pas encore pu se constituer un réseau stable. Le crédit d'impôt audiovisuel est probablement plus incitatif pour ces sociétés. Il est donc probable qu'en l'absence de CIA, les sociétés aidées auraient plus tourné en France que les entreprises non bénéficiaires du CIA. Le biais de sélection est donc probable, et l'estimateur naïf aura tendance à surestimer l'effet de ce crédit d'impôt.

Or, de manière heureuse pour l'évaluateur, il existe un barème sur lequel s'appuie le comité d'experts pour sélectionner les projets. Une œuvre sera considérée comme réalisée principalement sur le territoire français si elle obtient au moins la majorité des points hors ceux affectés à la langue. Cette condition n'est ni nécessaire ni suffisante : une société qui obtiendrait un score légèrement inférieur au seuil pourrait être rattrapée si le comité de sélection considère que l'œuvre contribue suffisamment au développement de la création cinématographique, et inversement une œuvre au-dessus du seuil pourrait ne pas être sélectionnée si elle n'y contribue pas suffisamment. Si la probabilité d'être sélectionné augmente brusquement au passage du seuil, cette discontinuité peut être utilisée

pour évaluer le CIA. On est alors dans le cadre de la discontinuité floue puisqu'il existe des bénéficiaires du CIA des deux côtés du seuil. L'idée est alors de se restreindre à une fenêtre de taille très réduite autour du seuil et d'utiliser la méthode des variables instrumentales avec, comme instrument, la variable qui renseigne sur le fait d'être au-dessus ou en dessous du seuil. Cette variable est en effet un bon candidat pour faire office d'instrument. D'une part, elle a une influence sur le traitement puisque les proportions de bénéficiaires sont, par hypothèse, différents au-dessus et en dessous du seuil. D'autre part, le fait d'être, dans la fenêtre, sur ou sous le seuil est indépendant des déterminants du nombre de jours de tournage en France. En effet, le barème ne contenant pas d'item à moins de deux points, le « biais de notation » est tel que le comité d'expert ne note pas de manière parfaitement identique deux dossiers équivalents. Ainsi, pour des raisons purement subjectives, indépendantes des déterminants du nombre de jours de tournage en France, les dossiers dont la valeur se situe à environ 50 sont affectés aléatoirement entre, disons, 48 et 52.

Le principal intérêt de cette méthode est qu'elle mime l'affectation aléatoire dans le traitement. Comme on vient de le voir, dans le cas de la discontinuité nette, l'estimateur naïf peut être tout simplement utilisé, et, dans le cas de la discontinuité floue, il faut recourir à la méthode des variables instrumentales avec, comme instrument, la variable qui indique si l'individu est au-dessus ou en dessous du seuil. Il existe des inconvénients, de deux ordres. La première limite est d'ordre empirique. Elle vient du fait que l'évaluateur souhaite une taille de fenêtre la plus petite possible pour que l'hypothèse d'affectation aléatoire soit réaliste ; mais il souhaite en même temps une fenêtre la plus grande possible pour disposer d'un échantillon suffisamment large. Cette méthode nécessitera donc une vaste base de données, et il faudra faire varier la taille de la fenêtre pour tester la robustesse des résultats. La deuxième limite est la faible validité externe de la méthode. Dans la mesure où l'estimation de l'effet du traitement est réalisée en n'utilisant que les individus autour du seuil, les résultats ne sont valables qu'autour du seuil : on dit qu'il y a identification locale. Si l'on pense que l'effet du traitement est homogène, alors on peut généraliser le résultat à toute la population, mais si l'on pense au contraire que l'effet est hétérogène, alors il faut sérieusement se questionner avant toute généralisation à la totalité du score.

Dans le cas du CIA, il est possible que des scores élevés soient associés à des entreprises qui auraient de toute façon tourné en France, et que des scores autour du seuil soient au contraire associés à des entreprises qui auraient plus souvent délocalisé les tournages. Si tel était le cas, l'effet du CIA serait hétérogène, moins important pour des valeurs élevées du score qu'autour du seuil. Il serait donc peu recommandé de généraliser les résultats à la

totalité du score. Les résultats de cette méthode restent néanmoins informatifs. Tout d'abord, les pouvoirs publics peuvent être particulièrement intéressés par l'efficacité du programme sur les entreprises qui y sont potentiellement le plus sensibles (autour du seuil dans notre exemple). Ensuite, si les résultats montraient que l'effet est nul sur cette zone du score, on pourrait en déduire qu'il l'est, selon toute vraisemblance, également sur le reste du score. Enfin, les pouvoirs publics seront informés avec précision de l'utilité de déplacer marginalement le seuil, à l'intérieur de la fenêtre qui a servi à réaliser l'évaluation : si la fenêtre retenue a été [48, 52] et que l'évaluation a détecté un effet positif dans cette fenêtre, alors abaisser le seuil de 50 à 48 aura également un effet causal positif sur toutes les entreprises dont le score est compris entre 48 et 50.

CONCLUSION

Le but de cet article est de présenter de manière intuitive, non formalisée, les méthodes économétriques récentes d'évaluation de politiques publiques⁷. Par souci de pédagogie, nous n'avons associé à chaque politique publique qu'une seule méthode d'évaluation. De manière générale, une même mesure peut être évaluée au moyen de plusieurs méthodes, le choix de ces dernières étant dicté par le cadre institutionnel de la mesure, par les données disponibles et par les hypothèses de validité sous-jacentes aux méthodes. L'attribution du label LIR étant également réalisée sur critères, ce label pourrait ainsi être évalué grâce à la méthode de la régression avec discontinuité. De même, le crédit d'impôt audiovisuel ayant été créé en 2004, la méthode des différences de différences pourrait être mobilisée. On peut également signaler que pour renforcer la validité des hypothèses sous-jacentes, ces méthodes peuvent être articulées avec les méthodes économétriques plus classiques : des variables de contrôle peuvent (et doivent) être incluses dans les modèles. C'est le cas notamment de la méthode des différences de différences : pour rendre plus réaliste l'hypothèse identifiante d'évolution similaire des traités et des non-traités en l'absence de traitement, il est utile d'inclure dans le modèle des variables explicatives variant dans le temps qui rendent plus semblables le groupe des traités et le groupe de contrôle.

Au-delà des limites spécifiques à chaque méthode, il existe deux limites supplémentaires communes à chacune d'entre elles.

D'une part il peut exister des effets d'équilibre. Dans l'exemple du programme École et cinéma, si, à prix constant, l'effet instantané du programme est positif, l'augmentation de la demande du bien « cinéma en salle » fera à terme potentiellement augmenter son prix. Certains

7. Pour une présentation plus formalisée, voir par exemple Pauline GIVORD, 2010 (http://www.insee.fr/fr/publications-et-services/docs_doc_travail/G2010-08.pdf), Denis FOUGÈRE, 2010 (http://www.crest.fr/ckfinder/userfiles/files/Pageperso/fougere/fougere_fichiers/ARTICLEFOUGERE_RFAS_1-22010.pdf) ou les chapitres dédiés des manuels d'économétrie récents comme celui de Bruno CRÉPON et Nicolas JACQUEMET, *Économétrie : méthodes et applications*, De Boeck, 2010.

individus pourront alors choisir de consommer un autre bien culturel plutôt que le cinéma devenu trop cher pour eux. Ainsi l'effet total du programme pourra rester positif, mais sera plus limité que son effet instantané. Si l'évaluation se déroule trop tôt pour laisser le temps aux prix de s'ajuster, alors l'effet du programme sera surestimé.

D'autre part il peut exister des externalités, c'est-à-dire un effet du traitement sur les non-traités. Dans le cas du label LIR, on peut assister à un effet de substitution : certains clients peuvent décider de changer de librairie pour passer d'une librairie non labellisée à une librairie labellisée. Ce processus aura un effet positif sur les librairies labellisées et négatif sur les librairies non labellisées. Ces deux limites sont l'objet de travaux de recherche et n'ont donc pas encore de solution définitive. Pour autant, ces limites n'invalident pas les avancées qu'a permises le développement de ces méthodes, qui participent à améliorer la gestion de la dépense publique.

Insistons, enfin, sur la condition sans laquelle toute évaluation est impossible : disposer de données en quan-

tité et en qualité suffisantes. Bien qu'évident, ce point n'en est pas moins crucial. L'exemple du programme École et cinéma est éclairant à cet égard : réussir à enquêter suffisamment d'individus âgés de dix-huit ans qui ont participé au programme dix ans plus tôt est possible mais nécessite un sondage coûteux. Obtenir par ailleurs les coûts de transport en vigueur lors de l'expérience ne peut se faire que si cela a été prévu. Il faut enfin disposer d'informations permettant de constituer des sous-populations homogènes au regard de l'effet du programme, par exemple la catégorie sociale des parents des enfants qui iront au cinéma à 18 ans. Pour qu'une évaluation de politique publique soit envisageable, le recueil des données doit en fait être anticipé lors de la mise en place de la mesure. Si des données administratives existent, il faut favoriser leur disponibilité auprès des évaluateurs. Sinon, il faut réfléchir en amont à la façon la plus efficace de recueillir régulièrement ces données. Une anticipation des méthodes utilisées sera un bon guide dans l'exécution de cette tâche. ■

À lire aussi :



CM-2007-1

Les premiers principes de l'analyse d'impact économique local d'une activité culturelle

Yann NICOLAS

Avril 2007, 8 p.

L'impact économique local d'un équipement ou d'un événement culturel est attribuable aux dépenses de ceux qui les fréquentent. Il ne peut être solidement estimé qu'en respectant, une fois compris ce qu'il mesure, plusieurs premières règles de méthode : utilisation d'un modèle économique local *ad hoc*, exclusion de plusieurs types de dépenses, prise en considération de ses coûts. Cette estimation faite, les résultats ne peuvent toutefois être utilisés qu'avec prudence pour motiver la subvention publique locale à une activité culturelle. Cette note détaille les raisons justifiant cette attitude.



CM-2011-1

Une mesure de la diversité des marchés du film en salles et en vidéogrammes en France et en Europe

Florence LÉVY-HARTMANN

Octobre 2011, 16 p.

La question de la diversité culturelle est abordée sous l'angle de la mesure de la diversité du marché du film en salles et du film en vidéogrammes au cours des années 2000, sur la base d'un échantillon inédit de 5 600 films diffusés en salles et de 6 500 films enregistrés en vidéogrammes. La méthode retenue consiste, pour les deux marchés, à mesurer la diversité à partir des critères de variété, d'équilibre et de disparité. Des éléments d'appréciation de l'offre (films), de la distribution (copies) et de la demande (nombre de billets ou de vidéogrammes vendus) sont proposés pour six pays d'Europe : France, Danemark, Espagne, Pologne, Royaume-Uni, Suède. Le marché du film en vidéo se limite à la mesure du cas français. La méthode testée tend à montrer que la France est le pays où le marché du film en salles est le plus diversifié, mais que le marché du film en vidéo est beaucoup moins diversifié. Elle confirme les différences de diversité selon les circuits de commercialisation et, plus inédit, le niveau élevé de diversité du marché du film en vidéo sur l'internet.



CM-2011-2

Pratiques culturelles, 1973-2008 Questions de mesure et d'interprétation des résultats

Olivier DONNAT

Décembre 2011, 12 p.

L'enquête *Pratiques culturelles des Français*, réalisée depuis 1973 par le ministère de la Culture et de la Communication, mesure les taux de pénétration, au sein de la population française, des principales formes d'accès à la culture. Comparer des données d'enquêtes sociologiques produites depuis trente-cinq ans à intervalle d'une décennie environ requiert plusieurs précautions méthodologiques à mettre en œuvre pour l'analyse des données. Il convient de tenir compte des évolutions structurelles de la société aux plans démographique, économique et social et de distinguer les effets d'époque, de génération et d'offre. Ensuite, les pratiques culturelles elles-mêmes, et plus généralement les modes d'accès à la culture se sont transformés depuis trente-cinq ans, notamment sous l'effet des innovations technologiques. Enfin, le document revient sur l'interprétation des processus souvent convoqués pour analyser l'évolution des pratiques culturelles : féminisation, élitisation ou démocratisation, juvénalisation.



CM-2011-3

Conceptualisation statistique du champ de la culture

Valérie DEROIN

Décembre 2011, 12 p.

La définition du champ statistique de la culture a fait l'objet de travaux européens de 2009 à 2011, après que les nomenclatures nationales (nomenclature d'activités françaises, NAF Rév. 2) et européenne (nomenclature statistique des activités économiques dans la Communauté européenne, NACE Rév. 2) ont été rénovées en 2008. Sous l'égide d'Eurostat, l'Office statistique de l'Union européenne, un groupe de travail européen (*Task Force 1*) a conçu la structure du nouveau cadre statistique européen constitué d'une trentaine d'activités culturelles, dans un souci d'harmonisation des outils méthodologiques et pour garantir une meilleure comparabilité des résultats publiés. Le cadre européen s'insère dans le cadre international défini par l'Unesco pour les statistiques culturelles et croise dix domaines culturels – patrimoine culturel, archives, bibliothèques, livre et presse, arts plastiques, audiovisuel et multimédia, architecture, création publicitaire, artisanat d'art – avec six fonctions économiques : création, production et édition, diffusion et commercialisation, conservation, formation, administration et management culturels. La publication revient sur les différentes nomenclatures existantes en France, en Europe et à l'échelon international et précise leur principe d'emboîtement. Publiée conjointement dans la collection « Culture études » du DEPS, *Approche statistique européenne de la culture* est la synthèse des travaux du réseau européen de statistique ESSnet-Culture et vient compléter utilement le présent document méthodologique.

RÉSUMÉ

L'évaluation des politiques publiques permet de mesurer l'effectivité de la mise en œuvre d'une mesure et son efficacité, évaluée selon les objectifs définis et justifiant l'intervention publique. Les méthodologies d'évaluation des politiques publiques peuvent être de différentes natures et les sciences économiques disposent d'outils permettant une évaluation d'ordre économétrique.

Comparer directement les bénéficiaires et les non-bénéficiaires d'une mesure fournit une évaluation biaisée dans la mesure où ces deux groupes ont en général des caractéristiques différentes. Les méthodes économétriques ont en commun de résoudre ce problème en permettant de raisonner « toutes choses égales par ailleurs », c'est-à-dire d'isoler l'effet de la mesure afin d'en apprécier l'impact propre. Conformément au modèle causal de Rubin, l'évaluation consiste ainsi à comparer la situation que l'on peut observer avec celle que l'on observerait en l'absence d'intervention publique. Cette dernière, appelée situation contrefactuelle, est fondamentalement inobservable et tout l'enjeu des méthodes économétriques est d'en fournir une estimation la plus précise possible. Les trois outils les plus classiques sont présentés ici et illustrés par trois dispositifs d'intervention publique dans le champ de la culture : le label LIR, le dispositif École et cinéma et le crédit d'impôt audiovisuel.

Dans le cas du label LIR, il est proposé d'évaluer si le label a permis à certaines librairies d'améliorer leur taux de rentabilité. Le label LIR permet d'exposer théoriquement la méthode des différences de différences ou estimateur en double différences, qui consiste à comparer avant et après la création du label les librairies labellisées et non labellisées pour évaluer son efficacité.

Avec le dispositif École et cinéma, l'évaluateur se demande si le programme a permis de développer le goût des enfants pour le cinéma en analysant le taux de fréquentation annuelle des salles obscures des jeunes adultes. École et cinéma permet d'illustrer la méthode des variables instrumentales qui consiste à utiliser une source de variation exogène qui n'a pas d'effet direct sur le nombre d'entrées au cinéma des jeunes adultes mais qui explique la participation au programme de ces jeunes à l'école primaire.

Enfin, la méthode de la régression avec discontinuité est proposée pour l'évaluation du crédit d'impôt audiovisuel afin d'analyser l'objectif de localisation française des dépenses de production en mesurant le nombre de jours de tournage sur le sol français des sociétés de production. Cette méthode consiste à profiter de l'existence d'un score de sélection des participants au programme, et de restreindre l'évaluation au voisinage du seuil donnant l'accès au crédit d'impôt afin de mimer une affectation aléatoire des individus dans le programme.

Plutôt qu'une évaluation des dispositifs proposés ici à titre illustratif, cette publication vise à exposer des outils encore peu mobilisés pour l'évaluation de politiques culturelles.

ABSTRACT

The evaluation of public policies enables an assessment to be made of how efficiently a measure has been implemented and how effective it is in itself, evaluated in accordance with defined criteria and justifying public intervention. There is a range of different public policy evaluation methodologies, with economics providing various tools for econometric evaluation.

A direct comparison of the beneficiaries and non-beneficiaries of a given measure will give a biased evaluation, as, generally speaking, both groups will have quite different characteristics. Econometric methods all avoid this problem by taking an "all other things being equal" approach, i.e. by isolating the effect of a given measure in order to assess its own specific impact. In accordance with Rubin's causal model, an evaluation compares an observed situation with that which one might observe in the absence of any public intervention. The latter is referred to as the counterfactual situation; it is fundamentally unobservable, so econometric methods aim to supply as precise an assessment as possible. The three most standard methods are outlined here and are illustrated using three public intervention initiatives from the cultural sphere: the LiR label initiative, the Schools & Cinema initiative and the audiovisual tax credit scheme.

In the case of the LiR label, it aims to determine whether this labelling scheme has enabled certain bookshops to improve their profitability. The LiR label enables a theoretical exposition of the difference in differences or double differences method, which compares labelled and non-labelled bookshops both before and after creation of the label in order to evaluate its efficacy.

For the Schools and Cinema initiative, the assessment looked at whether this scheme had helped develop young people's taste for the cinema by analysing young people's annual cinema attendance rates. School and Cinema provides an illustration of the instrumental variables method which involves taking an exogenous variable which has no direct effect on cinema attendance levels in young adults but which explains the participation of these young people in the programme at primary school level.

Finally, regression discontinuity design is used to assess the audiovisual tax credit scheme, analysing the aim of attracting production expenditure to France by measuring the number of days spent filming on French soil by production companies. This method involves using an existing selection rating for scheme participants, restricting the evaluation to those nearest the threshold for tax credit qualification in order to imitate a random selection of the individuals in the programme.

Rather than an evaluation of the schemes shown here for illustrative purposes, this publication aims to outline those tools which are still under-used in the evaluation of cultural policy.

Tous les documents publiés par le DEPS sont téléchargeables sur
<http://www.culturecommunication.gouv.fr/Politiques-ministerielles/Etudes-et-statistiques>
et sur www.cairn.info

Pour recevoir régulièrement les publications du DEPS et pour toute demande d'information :
contact.deps@culture.gouv.fr