
DOCUMENT
DE TRAVAIL
N° 565

**APPROCHES STRUCTURELLES ET NON STRUCTURELLES EN MICRO-
ÉCONOMÉTRIE DE L'ÉVALUATION DES POLITIQUES PUBLIQUES**

Sébastien Roux

August 2015



**APPROCHES STRUCTURELLES ET NON STRUCTURELLES EN MICRO-
ÉCONOMÉTRIE DE L'ÉVALUATION DES POLITIQUES PUBLIQUES**

Sébastien Roux

August 2015

Les Documents de travail reflètent les idées personnelles de leurs auteurs et n'expriment pas nécessairement la position de la Banque de France. Ce document est disponible sur le site internet de la Banque de France « www.banque-france.fr ».

Working Papers reflect the opinions of the authors and do not necessarily express the views of the Banque de France. This document is available on the Banque de France Website “www.banque-france.fr”.

Approches structurelles et non structurelles en micro-économétrie de l'évaluation des politiques publiques

Sébastien Roux¹,
Banque de France, Ined et Crest

¹ Cet article a été préparé dans le cadre de l'école d'évaluation des politiques publiques organisée à Aussois en mars 2014. L'auteur remercie les trois rapporteurs de la Revue française d'économie pour leurs suggestions d'amélioration et les références qu'ils ont apportées, ainsi que les participants à l'école TEPP d'Aussois. Il reste seul responsable des inexactitudes ou manques.
Email : roux@ensae.fr

Résumé : Cet article vise à présenter et mettre en perspective les approches structurelles et non structurelles en économétrie de l'évaluation des politiques publiques. Si ces approches sont souvent opposées car correspondant à des démarches scientifiques différentes, elles se révèlent complémentaires de par les enseignements qu'elles peuvent apporter. Deux illustrations sont présentées, illustrant cette complémentarité : l'évaluation de la mise en place des 35 heures à partir de 1998 (Crépon, Leclair, Roux [2003]) et l'effet local des radars sur les accidents de la route (Roux, Zamora [2013]).

Mots-Clés : Méthodes d'évaluation, Modèles structurels, Expériences naturelles

Classification JEL : B23, C10, C21, C52

Abstract : This article aims at presenting and comparing structural and atheoretic approaches to micro-econometrics of public policy evaluation. If these approaches are often opposed because they rely on different scientific methodologies, they complement each other in the lessons one can draw from them. Two illustrations are presented: the evaluation of the workweek reduction in 1998 (Crépon, Leclair, Roux [2003]) and the local effect of speed enforcement cameras on road accidents (Roux, Zamora [2013]).

Keywords: Evaluation methods, Structural models, Natural Experiments

JEL classification: B23, C10, C21, C52

Résumé non technique :

Cet article a pour objet de présenter et de comparer les approches structurelle et non structurelle en micro-économétrie de l'évaluation. Chacune des approches a vocation à évaluer les effets d'une mesure de politique économique. Si l'approche structurelle s'appuie sur une représentation explicite de l'économie fondée sur des mécanismes bien identifiés par le chercheur et matérialisés par des paramètres structurels invariants aux politiques économiques mises en œuvre, l'approche non structurelle s'appuie sur la seule variabilité contenue dans les données, s'affranchissant ainsi des hypothèses sur les principaux mécanismes en œuvre.

L'approche structurelle permet ainsi d'évaluer les politiques publiques dans la mesure où elle se fonde sur les mécanismes les plus pertinents pour décrire la situation qu'elle évalue. À la différence de l'approche non structurelle, elle permet aussi d'évaluer les effets d'une politique publique pas encore mise en œuvre. La validité du modèle est alors cruciale, et nécessite d'être argumentée. Faute de validité du modèle, le risque est de se tromper lourdement dans l'évaluation. L'approche non structurelle permet d'évaluer les effets d'une politique donnée dans un contexte précis, en s'appuyant sur la variabilité contenue dans les données. Elle permet de connaître les conséquences pas toujours anticipées d'une politique mise en œuvre. L'interprétation des résultats est alors cruciale pour être en mesure de généraliser à d'autres contextes les conclusions obtenues.

Bien que ces deux approches soient souvent opposées (Deaton [2010], Keane [2010] pour l'approche structurelle et Angrist et Pischke [2010], Imbens [2010] pour l'approche non structurelle), elles ne sont en fait pas si faciles à distinguer (Heckman [2010] et Blundel et Costa-Dias [2004]). Par les méthodes statistiques utilisées, il peut exister un *continuum* entre une approche complètement empirique et « athéorique » et une approche complètement structurelle. Cet article s'appuie notamment sur le modèle structurel de Roy, qui décrit le comportement de participation à un programme d'un individu en fonction du profit qu'il en tire. Il le met en regard du cadre statistique causal de l'évaluation de Rosenbaum et Rubin, qui s'appuie sur la définition d'individus traités comparés à un groupe de contrôle, très utilisé dans les approches non structurelles.

Deux exemples d'application sont présentés, qui montrent que les considérations d'évaluation liées aux données et les réflexions sur les mécanismes en jeu sont utiles pour arriver à des conclusions sur les effets d'une politique publique donnée. Le premier exemple, tiré de Roux et Zamora [2013], étudie les effets locaux sur les accidents corporels de la mise en place de radars fixes. Il illustre comment une approche modélisatrice permet de pallier le manque d'information dans les données et de tester certaines spécifications. Le second exemple, tiré de Crépon, Leclair et Roux [2003], examine les effets dans les entreprises de la réduction du temps de travail mise en œuvre entre 1998 et 2000. Il montre comment une approche structurelle peut être mobilisée pour interpréter des résultats obtenus à partir d'une approche non structurelle.

Les trente dernières années ont connu un développement très important de la micro-économétrie, par le développement de la puissance de calcul des ordinateurs, une plus grande disponibilité des données individuelles, et la plus grande demande d'évaluation de politiques publiques de la part des responsables politiques. D'une façon générale, cette évaluation est conduite en comparant une variable d'intérêt obtenue en la présence de politique publique à la même variable d'intérêt qui aurait été obtenue sans celle-ci, c'est-à-dire en considérant une situation contrefactuelle. Il est fréquent parmi les micro-économètres d'opposer deux approches lorsque l'on est confronté à un problème d'évaluation : l'approche structurelle et l'approche non structurelle.

En simplifiant très fortement, l'approche structurelle s'appuie sur une représentation explicite du fonctionnement de l'économie, se fondant sur une idée *a priori* des mécanismes en œuvre. L'économie est alors modélisée à partir de ces mécanismes et dépend de paramètres qualifiés de structurels, ces paramètres étant supposés ne pas être modifiés en cas de mise en œuvre de la politique publique (par exemple certains paramètres gouvernant le comportement des agents, qui entreraient dans la fonction d'utilité). Les modifications de politique publique, ou d'autres sources de variation exogènes, offrent alors une opportunité de mieux identifier de tels paramètres. Une fois ceux-ci identifiés, l'approche structurelle permet alors de construire des situations contrefactuelles, que l'on aurait observées en l'absence de cette politique (ou avec une autre politique).

L'approche non structurelle vise plus directement à identifier les effets causaux d'une politique mise en œuvre, en se fondant avant tout sur l'information contenue dans les données, et en utilisant des méthodes statistiques plus ou moins sophistiquées (pouvant aller de la simple comparaison de moyennes à l'utilisation de statistique non paramétrique). Cette approche est particulièrement adaptée à l'évaluation de programmes mis en œuvre sur une partie de la population. Elle ne pose pas d'hypothèse précise sur les mécanismes économiques en œuvre et est parfois qualifiée de « athéorique », car les résultats directement observés ne sont pas affectés par un quelconque *a priori* de l'économètre sur le comportement des agents.

Ces démarches sont très différentes et ont des spécificités qui sont des avantages ou des inconvénients selon le contexte pour lequel elles sont mobilisées. Heckman [2010] rappelle qu'il existe trois classes de problème d'évaluation de politique publique :

- Evaluer les impacts de politiques publiques déjà mises en œuvre, en particulier sur le bien-être des individus qui sont concernés, et plus généralement sur la société en général.
- Prévoir les impacts de politiques publiques déjà mises en œuvre dans des environnements donnés sur d'autres environnements.
- Prévoir les impacts de politiques publiques jamais mises en œuvre.

Cette simple distinction entre ces classes de problème montre que seule une approche structurelle permet de répondre aux deuxième et troisième classes de problème. La question qui se pose alors est la capacité à prévoir correctement les effets d'une politique donnée dans un contexte différent. Ainsi, comme le souligne Heckman [2010], de nombreux économistes ont attiré l'attention sur l'importance de la sensibilité des résultats aux hypothèses, parfois très techniques (par exemple relatives à des formes fonctionnelles), qui ont pu susciter des doutes quant à la capacité de cette démarche à produire une information effectivement pertinente pour le décideur public. La complexité des méthodes structurelles, parfois peu appréhendables pour des décideurs, jette en outre un doute sur leur pertinence.

A l'opposé, la démarche non structurelle, qui s'appuie avant tout sur la variabilité des données disponibles et qui peut même parfois chercher à créer les sources de données les plus pertinentes lors de la mise en œuvre de politiques publiques, *via* des expérimentations randomisées, permettra d'apporter des réponses beaucoup plus convaincantes pour des décideurs publics quant aux effets d'une politique mise en œuvre. Néanmoins, la généralisation à des contextes différents de

ceux pour lesquels la démarche a été appliquée repose sur des hypothèses qu'il s'agit d'explicitier et qui ne seront pas forcément vérifiées.

Cet article vise à apporter quelques éléments de réflexion sur l'opposition ou la complémentarité entre ces deux approches, telles qu'elles ont pu être présentées dans des contributions récentes de tenants de l'approche structurelle (Deaton [2010], Keane [2010]), de l'approche non structurelle (Angrist et Pischke [2010], Imbens [2010]) ou de contributeurs cherchant à les articuler l'une par rapport à l'autre (Heckman et Urzua [2010], Heckman [2010] et Blundel et Costa-Dias [2004]). Il adopte le point de vue du praticien économètre à qui il serait demandé de conduire une évaluation de politique publique. Dans une première partie, l'approche structurelle est plus précisément définie, en prenant comme exemple le modèle de Roy [1951]. Dans une seconde partie est présentée l'approche non structurelle, en s'appuyant sur la représentation définie par Neyman [1923] et Cox [1958] et popularisée par Rosenbaum et Rubin [1983]. Dans une troisième partie, les liens entre ces deux approches sont présentés, en précisant comment les différentes méthodes d'évaluation constituent un *continuum* entre ces deux approches. Enfin, deux exemples d'application sont présentés, tirés de travaux auxquels nous avons participé : l'évaluation des effets des 35 heures mises en place à partir de 1998 sur les entreprises, et des effets locaux de la mise en place des radars fixes à partir de 2005 sur les accidents de la route.

Approche structurelle

La démarche idéale en approche structurelle

Un article d'évaluation s'appuyant sur une approche structurelle aura comme première préoccupation d'identifier les principaux mécanismes économiques à prendre en compte. Cela passe par une description précise des stratégies potentielles des acteurs ou une analyse des imperfections de marché (asymétries d'information, frictions) pouvant affecter le comportement des agents. L'identification de ces mécanismes doit s'appuyer sur la réflexion du chercheur et sur la littérature économique qu'il a pu mobiliser sur le sujet qu'il examine.

Un article d'évaluation ne pouvant consister en une seule contribution théorique, il s'agit d'identifier les données disponibles, qui vont permettre d'estimer le modèle. Ces données peuvent être de nature individuelle, mais peuvent aussi consister en des élasticités obtenues à partir de la littérature existante. Le choix des données à considérer dépend alors des mécanismes identifiés comme étant prépondérants : le choix d'une variable particulière peut ainsi être considéré comme crucial, conditionnant les données choisies pour effectuer l'évaluation.

Ensuite vient le temps de la modélisation : il s'agit d'écrire un modèle économique formalisé, ce qui en assure la cohérence et la capacité à lier les données aux mécanismes identifiés précédemment. Lors de cette étape sont définis les paramètres structurels, c'est-à-dire invariants aux politiques publiques envisagées. C'est aussi lors de cette étape que le processus générateur de données est présenté : décrire comment les aléas sont générés, leurs propriétés distributionnelles et comment, au final, les données observées peuvent être générées par le modèle économique formalisé, *via* la spécification de fonctions liant les paramètres structurels, les variables exogènes et les variables endogènes².

Une fois le processus générateur des données précisé, le modèle, et plus particulièrement les paramètres structurels, peuvent être estimés, par maximum de vraisemblance lorsque les distributions des aléas ont été spécifiées. D'autres méthodes peuvent être mobilisées, le plus

²C'est lors de la spécification du modèle qu'est également déterminé le caractère exogène ou endogène des variables considérées. Dans l'exemple du modèle de Roy considéré par la suite, la variable de résultat, ainsi que l'assignation au traitement sont des variables endogènes.

souvent pour des raisons pratiques lorsque la vraisemblance est trop complexe pour être estimée ou pour obtenir des résultats robustes aux hypothèses distributionnelles³. L'étape d'estimation peut nécessiter un recours à une littérature économétrique plus spécialisée, d'une nature différente de celle mobilisée au moment de l'identification des mécanismes économiques à considérer.

A partir des estimations obtenues, il s'agit de vérifier si elles sont crédibles et raisonnables. Pour cela, on peut comparer les résultats obtenus sur certains paramètres à d'autres estimations obtenues dans la littérature. Idéalement, il faut s'assurer de la validité externe des résultats. L'exemple idéal pourrait consister à appliquer les résultats du modèle à des données non utilisées pour l'estimer mais permettant de tester si les prédictions du modèle sont réalisées, notamment en termes de mise en œuvre de la politique publique que l'on cherche à évaluer.

Les paramètres structurels ayant été estimés et le modèle validé, il peut être alors mobilisé pour effectuer des exercices contrefactuels correspondant à des simulations de mises en œuvre de politiques publiques et éventuellement être testé en confrontant ses prédictions à une politique publique déjà mise en œuvre⁴.

Un exemple illustratif : le modèle de Roy

Pour illustrer la démarche structurelle présentée, on peut s'appuyer sur le modèle de Roy, tel qu'il est présenté par Vitlacyl [2002] et Heckman [2010] et qui pourra également être utilisé pour illustrer le lien avec l'approche non structurelle dans la suite de l'article. L'intérêt de ce modèle est qu'on peut naturellement le rapprocher de l'approche non structurelle décrite dans la deuxième partie de cet article⁵, à la différence d'un grand nombre d'autres modèles qui ne se prêtent pas à cet exercice⁶.

Ce modèle est naturel à mobiliser pour un économètre dont l'objectif est d'évaluer les effets d'un programme qui va s'appliquer à une partie de la population, sur la base d'une participation volontaire de la part des individus.

La première étape est de décrire le contexte, les mécanismes économiques en jeu et les stratégies des agents.

On considère une situation où les individus ont la possibilité de participer à un programme (de formation ou autre), qui se matérialise par une variable $D = 0$ ou 1 . En fonction de leur choix, les agents vont percevoir un résultat Y_1 s'ils ont choisi de participer, Y_0 sinon. Ce résultat peut consister en un revenu (qui peut alors être directement observé), ou une utilité sous forme de gain en bien-être (qui nécessiterait alors une modélisation plus approfondie).

Le problème d'évaluation auquel l'économètre est confronté est qu'il ne peut pas observer simultanément Y_0 et Y_1 pour le même individu, quand bien même celui-ci peut disposer d'une information privée.

L'hypothèse structurelle sur laquelle se fonde le modèle est que les agents font leur choix en comparant Y_0 et Y_1 , valeurs qu'ils connaissent ou pour lesquelles ils disposent d'une information plus précise que l'économètre (ou ont des croyances différentes). Ce comportement postulé par le modèle conduit à une sélection endogène des individus dans le traitement. Plusieurs règles de sélection sont alors possibles :

³ Parmi celles-ci, la méthode d'inférence indirecte estime les paramètres structurels qui permettent au modèle de simuler un ensemble de statistiques agrégées aussi proches que possible de celles observées dans la réalité (Gouriéroux et Monfort [1996]).

⁴ C'est notamment le cas de McFadden [2001], qui a testé l'efficacité prédictive de son modèle structurel en comparant ses prédictions effectuées *ex ante* aux réalisations effectivement observées dans un article examinant les effets de l'introduction du métro à San Francisco.

⁵ Vitlacyl [2002] établit en quoi les hypothèses permettant l'identification du modèle de Roy sont équivalentes à celles permettant l'identification d'un effet causal dans le cadre non structurel de Rosenbaum et Rubin [1983], *cf.* plus bas.

⁶ Par exemple les modèles d'inspiration macro-économique s'appuyant sur un équilibre partiel ou général.

- Sélection simple : $D = (Y_1 > Y_0)$, les individus choisissent de participer si le résultat qu'ils anticipent parfaitement à participer excède celui à ne pas participer.
- Sélection avec coût : $D = (Y_1 - Y_0 > C)$, toujours s'ils anticipent parfaitement les résultats, les individus ne participent que si l'écart entre le résultat à participer excède celui à ne pas participer plus un coût qui peut s'interpréter comme un coût directement lié à participer au programme mais n'affectant pas le résultat.
- Sélection sur anticipation : on définit $I_D = E(Y_1 - Y_0 - C|Inf)$ où *Inf* est l'information détenue par l'agent au moment de son choix. La décision est alors la suivante : $D = I_D > 0$. Les individus choisissent de participer si l'espérance de gain à participer excède l'espérance du coût à participer. Cette spécification est alors plus réaliste car tient bien compte de la difficulté des agents à anticiper parfaitement les conséquences de leurs choix.

Ces considérations définissent les mécanismes économiques en jeu ici. La structure introduite dans ces mécanismes vient de l'hypothèse sur le comportement des agents : il est supposé que leur choix dépend directement des variables d'intérêt Y_0 et Y_1 , que ce soit leurs valeurs effectivement réalisées ou leurs espérances. Les règles de décision supposées requièrent alors que les données disponibles apportent des informations sur les variables de résultat Y , par exemple des salaires, sur le coût à participer au programme, voire sur l'information détenue par les agents au moment de leur choix⁷.

On suppose que la règle de décision à laquelle les agents se conforment consiste en une sélection sur anticipation. Il s'agit maintenant de préciser le modèle économétrique qui va permettre d'évaluer le programme. On peut commencer par proposer une représentation relativement flexible :

$$Y_0 = \mu_0(X) + U_0$$

$$Y_1 = \mu_1(X) + U_1$$

$$C = \mu_C(Z) + U_C$$

Les variables de résultats Y_0 et Y_1 sont supposées dépendre de caractéristiques observables X dont les rendements sont représentés par des fonctions $\mu_d(\cdot)$, $d = 0,1$, qui matérialisent l'effet du programme, et de variables aléatoires U_d , reflétant le caractère incertain ou hétérogène de la variable considérée. On suppose également que le coût est une fonction de variables explicatives Z , parmi lesquelles peuvent figurer les caractéristiques X qui affectent les variables de résultats, à laquelle une variable aléatoire est ajoutée, là aussi pour en faire intervenir l'aspect incertain ou hétérogène.

De cette spécification, on déduit la règle de décision de l'individu :

$$D = (I_D > 0) = (\mu_D(Z) > -V)$$

où $I_D = E(Y_1 - Y_0 - C|Inf) = \mu_D(Z) + V$.

Avec $\mu_D(Z) = E(\mu_1(X) - \mu_0(X) - \mu_C(Z)|Inf)$ et $V = E(U_1 - U_0 - U_C|Inf)$.

La spécification à laquelle on aboutit ici est relativement générale. Elle permet de décrire le problème de sélection des individus en fonction des paramètres structurels du modèle, ici les fonctions $\mu_0(\cdot)$, $\mu_1(\cdot)$ et $\mu_C(\cdot)$, des variables exogènes qui décrivent leur situation et des aléas qui

⁷ D'autres hypothèses comportementales sur les agents ou leur environnement peuvent conduire à l'expression d'autres modèles structurels. Par exemple, les modèles d'investissement en capital humain postulent que les individus investissent dans leur éducation ou leur formation en fonction de leur espérance de revenu futur actualisé (Ben Porath [1967]). C'est également le cas des modèles de recherche d'emploi, dans lesquels il est postulé que les individus adoptent comme stratégie optimale d'accepter un emploi dont le salaire est supérieur au salaire de réserve en fonction de leurs perspectives de gains futurs dans un contexte d'information imparfaite (Mc Call [1970]).

reflètent l'hétérogénéité des comportements individuels, dont les seuls aspects déterministes définis par les paramètres structurels ne peuvent pas rendre compte.

Cette spécification repose, outre l'hypothèse structurelle de comportement évoquée plus haut, sur des choix de spécification qui ne sont pas neutres. Le choix le plus important consiste en celui des variables X qui expliquent les résultats sous un régime ou l'autre et des variables Z qui déterminent le choix d'être traité ou non D . Le fait d'inclure ou non certaines variables reflète des choix effectués par l'économètre parfois contraints par la disponibilité des données. La tentation est grande d'introduire le plus de variables possibles pour augmenter le pouvoir explicatif du modèle. Une telle pratique se heurte à deux écueils : d'une part on risque d'introduire des variables endogènes parmi les variables explicatives, d'autre part l'identification du modèle s'appuie sur la disponibilité de variables d'exclusion, c'est-à-dire expliquant la décision de participer, mais n'affectant pas directement le résultat. Une telle variable pourrait par exemple être le résultat d'une loterie qui conditionnerait la participation de l'individu au programme, mais n'affecterait pas ses résultats. Introduire l'ensemble de l'information disponible pour expliquer les résultats sans discernement empêche *de facto* de choisir des variables d'exclusion car elles sont alors supposées toutes affecter le résultat. Prenons l'exemple de la participation à un programme de formation, dont on veut examiner l'effet sur les salaires. Imaginons que l'on dispose d'une variable indiquant si l'individu a été informé de l'existence de ce programme. *A priori*, on pourrait penser que cette variable est un bon candidat comme variable d'exclusion dans la mesure où on ne voit pas pourquoi cette information devrait affecter le salaire de l'individu, autrement que par le fait qu'il a ou non participé à la formation. Certaines questions restent cependant en suspens : comment l'individu a-t-il été informé de la formation ? La lui a-t-on proposée ou bien l'a-t-il trouvée lui-même ? L'a-t-il cherchée ou bien est-il tombé dessus par hasard ? Selon les réponses à ces questions, l'accès à l'information sur la formation pourrait refléter le fait que l'individu est plus performant que d'autres et expliquer ainsi le salaire qu'il perçoit. En incluant cette variable parmi les variables d'exclusion, l'économètre suppose que ces mécanismes sont négligeables au regard de l'effet que la disponibilité de l'information peut avoir sur le choix et attribue à la formation même l'effet que cette variable pourrait avoir sur les salaires. De façon plus générale, ce type d'hypothèses n'est pas testable sans des hypothèses sur d'autres variables d'exclusion, et nécessite une discussion approfondie dans tous les cas, qui n'est pas toujours présente dans les papiers s'appuyant sur une formalisation structurelle⁸.

Un autre jeu d'hypothèses, moins cruciales, consiste en celles relatives aux formes fonctionnelles et aux distributions des aléas du modèle. Elles sont moins cruciales car il est souvent possible de tester la robustesse des résultats à des spécifications alternatives. Le plus souvent les formes fonctionnelles correspondent à des formes linéaires reflétant le lien entre résultats et participation et variables explicatives. Des approximations fonctionnelles, sous forme de polynômes ou de fonctions par morceaux, peuvent être mises en œuvre pour tester la robustesse de la spécification choisie. Au final, on peut le plus souvent résumer les fonctions à un nombre fini de paramètres structurels qu'il s'agit d'estimer. Concernant l'hypothèse distributionnelle sur les erreurs, on choisit le plus souvent de les tirer dans des lois normales, de moyenne nulle conditionnellement aux explicatives, et dont il faut estimer la matrice de variance-covariance. Des considérations d'identification peuvent amener à restreindre la forme de matrice à une classe particulière⁹.

Selon ces considérations, le modèle de Roy aura la spécification fonctionnelle suivante :

⁸ D'autres méthodes que les variables instrumentales permettent d'identifier non paramétriquement l'effet d'un traitement en corrigeant de son endogénéité. Ainsi, les méthodes « d'identification à l'infini » (cf. Heckman, [1990]) permettent d'arriver à ce résultat en se concentrant sur les observations dont la probabilité de recevoir le traitement conditionnellement à l'instrument est proche de 1 pour les traités et proche de 0 pour les non traités. En pratique, cette méthode est peu utilisée car requiert un instrument expliquant très fortement l'assignation ou la non-assignation au traitement.

⁹ En annexe n°1 est présentée à titre d'illustration l'écriture de la vraisemblance du modèle présenté ici.

$$\begin{cases} Y_0 = X\beta_0 + U_0 \\ Y_1 = X\beta_1 + U_1 \\ D^* = Z\gamma + V \end{cases}$$

où (U_0, U_1, V) suit une loi normale trivariée, de moyenne nulle et de variance

$$\Omega = \begin{bmatrix} \sigma_0^2 & 0 & \rho_0\sigma_0 \\ 0 & \sigma_1^2 & \rho_1\sigma_1 \\ \rho_0\sigma_0 & \rho_1\sigma_1 & 1 \end{bmatrix}$$

La corrélation entre U_0 et U_1 est supposée être nulle car non identifiable. En effet, pour chaque individu, on ne peut l'observer que dans une situation, et pas dans l'autre. On suppose aussi pour des raisons d'identification que la variance de V est supposée égale à 1, car on n'observe pas la variable latente D^* , mais sa conséquence $D = (D^* > 0)$. On déduit de cette structure la probabilité jointe d'observer $(D = 1, Y_1)$ et $(D = 0, Y_0)$, ce qui permet d'en déduire la vraisemblance du modèle et d'en déduire des estimations des paramètres $(\gamma, \beta_0, \beta_1, \sigma_0, \sigma_1, \rho_0, \rho_1)$ qui définissent complètement le modèle (cf. annexe n°1 pour l'écriture de la vraisemblance du modèle).

Les inconvénients de la démarche structurelle

L'exemple du modèle de Roy, tel qu'il est présenté ci-dessus illustre bien l'un des principaux défauts de la démarche structurelle : sa très forte dépendance à la spécification choisie. Dans l'exemple ci-dessus, la spécification conduit à choisir une représentation linéaire pour rendre compte des effets des variables explicatives sur les résultats et une distribution normale des erreurs.

La spécification présentée ci-dessus est relativement simple car n'examine que l'effet d'un programme à un instant donné, sans se poser la question des effets à long terme du programme. On pourrait fortement complexifier la structure en introduisant la possibilité pour les individus de choisir ou non de suivre une formation à différents moments de leur carrière, ce qui pourrait très fortement complexifier le modèle. Dans un autre cadre, un article influent de Keane et Wolpin [1997] modélise les choix d'éducation des individus à partir de leur entrée au collège en fonction des perspectives de salaire qu'ils peuvent en retirer. L'idée est relativement simple : chaque année, les étudiants choisissent ou non de continuer à étudier en fonction des gains associés à leurs choix, qui dépendent eux-mêmes de paramètres que les économètres souhaitent estimer. La formalisation de ces choix se heurte au problème de la « malédiction de la dimensionnalité » (curse of dimensionality) qui correspond au fait que la séquence de décisions permettant d'arriver au gain du travail croît exponentiellement avec le nombre d'années restant avant d'entrer sur le marché du travail. L'approche structurelle que les auteurs choisissent nécessite de prendre en compte toutes les alternatives possibles aux choix observés, afin de correctement estimer la vraisemblance du modèle. Dans leur article, Keane et Wolpin [1997] proposent une solution à ce problème, qui nécessite l'utilisation de ressources informatiques lourdes, tout en faisant des choix de spécification nécessairement simplificateurs par rapport à la complexité effective des choix d'éducation des individus observés.

En lien avec la complexité technique de cette approche, le plus gros reproche fait à l'approche structurelle (cf. Angrist et Pischke [2010], et Leamer [2010]) consiste en la « relativement » faible attention portée par l'économètre structurel aux données. Ce reproche n'est pas nécessairement fondé pour tous les articles adoptant cette approche, mais renvoie à une pratique plus diffuse, qui consiste à porter plus d'attention à la spécification du modèle économique qu'à la variabilité réellement identifiante contenue dans les données. L'exemple du modèle de Roy donne une illustration de cette dérive. Comme nous l'avons vu plus haut, il est nécessaire de disposer de variables d'exclusion expliquant la sélection mais pas les variables reflétant les résultats pour identifier le modèle de façon générale. Or, formellement, il est possible d'identifier les paramètres

du modèle sans variable d'exclusion, en ne s'appuyant que sur l'hypothèse de normalité des erreurs. Ce faisant, l'identification devient très fragile car elle repose sur l'hypothèse distributionnelle concernant la variable latente, qui ne peut pas être testée. Le choix d'une distribution alternative peut alors fortement modifier les résultats obtenus. Ce risque illustre le plus important problème que les économètres structurels peuvent rencontrer : la forte dépendance des résultats à des hypothèses considérées comme peu cruciales (ou en tout cas qui devraient le rester). Le problème n'est alors pas tant cette dépendance, mais le fait qu'elle puisse être ignorée par l'économètre, qui produirait alors des résultats peu robustes, mais rendus statistiquement significatifs par la structure qu'il a lui-même apportée dans les données. Or, comme le souligne Keane [2010] lui-même, trop peu d'articles adoptant l'approche structurelle s'interrogent sur la robustesse des résultats ou sur la pertinence du modèle structurel qu'ils utilisent.

Approche non structurelle

La démarche idéale en approche non structurelle

La démarche non structurelle en économétrie de l'évaluation consiste avant tout à identifier l'effet causal de la politique publique à évaluer. A la différence de l'approche structurelle, l'économètre non structurel va se poser en priorité la question de la variabilité exogène présente dans les données permettant d'identifier un effet causal. Les données jouent ici un rôle crucial et peuvent conditionner complètement la problématique de l'article. Ainsi, les économètres non structurels mettront-ils l'accent sur la recherche ou la constitution de bases de données permettant d'avoir des instruments ou variables d'exclusion crédibles en allant jusqu'à organiser des expériences contrôlées sur le terrain avec comme finalité directe de constituer des données adaptées à l'évaluation de programmes donnés. C'est notamment le cas de toutes les expériences de terrain organisées en économie du développement, à la diffusion et au succès desquelles a grandement contribué Esther Duflo [2009]. Au-delà du cas polaire de ces données construites spécifiquement pour effectuer des évaluations, l'économètre non structurel se fondera sur toutes les données qu'il pourra utiliser, enquêtes statistiques, données administratives ou collectées sur internet, en exploitant au mieux les possibilités d'appariement entre ces différentes sources.

Comme l'expliquent Angrist et Pischke [2010], un article seul adoptant une telle démarche a un apport limité. Il est nécessaire de situer la démarche non structurelle dans un dessein plus large, nécessitant de répliquer les expériences dans les contextes les plus variés possibles, pour renforcer le caractère général des conclusions auxquelles les différentes études peuvent arriver. Angrist et Pischke remarquent l'essor très important au cours des vingt dernières années des études s'appuyant sur des quasi-expériences¹⁰, voire des expériences contrôlées. Ils appellent ce phénomène « Credibility Revolution » dans le sens où l'examen plus attentif des données et la volonté d'en inférer des causalités auraient permis de remettre les questions économiques au centre de l'attention des économistes empiriques.

Une étude adoptant une démarche non structurelle s'inscrira souvent dans le champ d'une littérature abondante visant à donner un caractère général aux effets identifiés, ou alors à illustrer en quoi les résultats obtenus peuvent différer de ceux observés dans d'autres contextes. Certes, certains articles seront innovants car réussiront à identifier de façon crédible pour la première fois les effets de certaines politiques publiques. Mais ils n'auront d'influence que s'ils sont abondamment

¹⁰ Angrist et Pischke [2010] appellent quasi-expériences ou expériences naturelles des chocs indiscutablement exogènes par rapport à la variable de résultat. Ces chocs peuvent par exemple résulter de circonstances politiques (modification législative) ou naturelles (tremblement de terre, phénomène climatique).

cités et suscitent des répliques dans des contextes différents de celui initialement envisagé. Si ces papiers, par leurs résultats, permettent d'identifier des mécanismes par lesquels les politiques publiques peuvent jouer, ces mécanismes sont rarement modélisés en tant que tels et sont présentés de façon littéraire comme des résultats de la recherche, et non comme des présupposés, à la différence de l'approche structurelle.

Le cadre causal de l'évaluation de Rosenbaum et Rubin

Rosenbaum et Rubin [1983] ont popularisé le cadre général de l'évaluation de programmes sur la base de la comparaison entre deux populations données : l'une affectée par le programme, dite traitée, l'autre non affectée, dite de contrôle.

On souhaite évaluer l'effet d'un traitement D sur une variable de résultat Y . On définit Y_0 le résultat obtenu en l'absence de traitement et Y_1 le résultat observé lorsque l'individu est traité. Le résultat observé au niveau d'un individu est alors $Y = DY_1 + (1 - D)Y_0$. Ce cadre très simple a l'avantage de définir la notion de « résultat potentiel ». Chaque individu a en fait deux résultats possibles, l'un correspondant à la situation où il est traité, et l'autre où il ne l'est pas. On met ici l'accent sur la distinction entre le groupe traité, indicé par 1, et le groupe non traité, indicé par 0. Ce cadre permet de définir pour chaque individu sa situation contrefactuelle, c'est-à-dire celle qui aurait été observée s'il avait été traité lorsqu'il ne l'a pas été, et *vice versa*.

Imbens et Wooldridge [2009] détaillent les avantages de ce cadre. En premier lieu, il permet de définir formellement l'effet causal du traitement pour chaque individu : il s'agit de la différence $Y_1 - Y_0$, qui est l'écart entre le résultat que l'individu obtiendrait en étant traité et celui en ne l'étant pas. Cet écart ne peut pas être directement observé par l'économètre (ni peut-être même par l'individu qui en a cependant probablement une meilleure idée) car un individu ne peut être à la fois traité et non traité. L'introduction de la notion de résultat potentiel en lieu et place du résultat effectivement réalisé a néanmoins comme avantage d'amener l'économètre à se focaliser sur les moyens d'observer chacun des résultats et de réfléchir à mettre en place des stratégies d'estimation, voire des expériences permettant d'identifier l'effet causal. Le second avantage de ce cadre est qu'il distingue les résultats potentiels du mécanisme d'assignation au traitement, matérialisé par la variable D . Cela permet aussi à l'économètre d'apporter son attention à ce mécanisme d'assignation, crucial quand il s'agit d'examiner si l'effet qu'on observe est effectivement un effet causal. Comme troisième avantage, ce cadre permet de formuler des hypothèses probabilistes en termes de variables potentiellement observables, plutôt que sur des variables inobservables (par exemple sur les termes d'erreur d'une équation linéaire) qui peuvent dépendre d'autres hypothèses (notamment fonctionnelles). Enfin, ce cadre permet de mieux caractériser l'incertitude sur les estimateurs, en distinguant celle portant sur les résultats potentiels de celle portant sur l'assignation au traitement.

Le mécanisme d'assignation va directement conditionner le choix de la méthode statistique, sa complexité et la capacité de l'économètre à en inférer des résultats crédibles. On peut distinguer trois grandes classes de mécanismes d'assignation dans la littérature :

- Les individus peuvent être aléatoirement assignés au programme dans le cadre d'une expérience contrôlée. Dans ce cas, la simple comparaison des moyennes entre le groupe traité et le groupe non traité s'interprète comme l'effet moyen causal du traitement parmi les traités. Cette situation se rencontre très rarement dans la nature, seules des expériences dites « randomisées », organisées directement par des économistes, permettent d'être dans cette situation. Le mécanisme d'assignation est alors pleinement maîtrisé ou au moins connu par l'économètre.
- L'assignation des individus au groupe traité ou non traité est aléatoire, conditionnellement à des variables explicatives X . Dans ce cas, on ne connaît pas exactement le mécanisme d'assignation, il suffit simplement que l'assignation obéisse à l'hypothèse dite « unconfoundness » qui formellement s'écrit $D \perp (Y_1, Y_0) | X$. Cette hypothèse implique que, si l'économètre est capable

d'identifier des variables explicatives qui épuisent la corrélation entre la variable de traitement et les résultats potentiels, alors il peut se ramener au cas où l'assignation au traitement devient aléatoire. Dans ce cadre, la méthode statistique la plus souvent utilisée est la méthode d'appariement qui se fonde sur le score de propension (cf. Rubin [1974] ou Legendre [2013] pour une présentation récente). On peut rencontrer ce cas dans la nature, par exemple lorsque l'on veut évaluer les effets d'une politique publique si on dispose de suffisamment d'informations reflétant la propension pour certains individus à être traités et si, une fois ces informations prises en compte, l'accès au traitement est rationné : c'est-à-dire que certains individus disposant des mêmes caractéristiques seront choisis pour participer tandis que d'autres ne le seront pas, indépendamment des raisons les poussant à être traités. Cette condition permet alors d'avoir des individus traités et non traités réellement comparables.

- L'assignation des individus dépend directement des variables de résultat potentiel : on est alors dans le cas d'une sélection sur inobservables, c'est-à-dire que l'on n'est pas en mesure de randomiser l'assignation. C'est typiquement le cas du modèle de Roy présenté dans la partie précédente. La sélection sur inobservables est, par défaut, la situation que l'on a de plus de chances de rencontrer dans la nature. Pour pouvoir évaluer l'effet causal du traitement, l'économètre doit alors trouver une source de variabilité exogène expliquant l'assignation et non le résultat (*i.e.* un instrument, cf. section suivante).

Les variables d'intérêt définies dans ce cadre ne sont plus alors les coefficients que les économistes structurels ont l'habitude de rencontrer, car les mécanismes économiques ne sont plus explicités. Ainsi définit-on l'effet du traitement sur la population comme l'espérance de l'écart entre le résultat obtenu en cas de traitement et en son absence, $\tau_{ATE} = E(Y_1 - Y_0)$, l'effet du traitement sur la population traitée comme l'écart, parmi les traités, entre le résultat en cas de traitement et en son absence, $\tau_{ATT} = E(Y_1 - Y_0 | D = 1)$, ... Ces variables d'intérêt ne sont pas directement observables par l'économètre car on ne peut observer chaque individu que dans une situation et pas dans l'autre. Cependant, le fait de les définir rigoureusement incite l'économètre à préciser quels estimateurs peuvent être identifiés et sous quelles hypothèses ils peuvent être considérés comme fiables.

Dans chacun des cas ci-dessus, pour pouvoir interpréter les variables d'intérêt comme des effets causaux du traitement, d'autres hypothèses ne portant pas directement sur le mécanisme d'assignation doivent être faites. En premier lieu, il faut que le résultat des individus ne soit pas affecté par le ou les traitements des autres individus. Cette hypothèse est appelée SUTVA pour Stable Unit Treatment Value Assumption. Ensuite, il faut que les variables explicatives introduites soient bien exogènes. Ces restrictions s'appliquent aussi à l'estimation du modèle de Roy présenté dans la partie précédente. Elles font alors partie des hypothèses du modèle structurel.

L'utilisation d'instruments

La plupart de ces paramètres d'intérêt sont aisés à obtenir lorsque l'on se situe dans le cas d'une assignation aléatoire au traitement, conditionnelle ou non. Lorsque l'on est en présence d'une sélection sur inobservables, une façon d'estimer ces variables d'intérêt consiste à s'appuyer sur des instruments. Ceux-ci sont rarement disponibles : ils doivent expliquer l'assignation au traitement, et ne pas directement affecter les variables de résultat potentiel. Si la première condition peut être facilement testée, la seconde l'est beaucoup plus difficilement car il s'agit d'une condition d'exclusion qui ne peut être fondée que sur une analyse des mécanismes expliquant l'assignation. De telles variables expliqueraient ainsi l'assignation sans directement affecter les résultats (cf. section précédente). L'exemple idéal d'un bon instrument est donné par Angrist [1990] pour examiner l'effet du service militaire au moment de la guerre du Vietnam aux Etats-Unis sur les carrières futures des individus y étant soumis. La participation au service militaire était alors conditionnée par une loterie

mise en place dans le cadre d'une conscription. La loterie n'était pas le seul élément affectant la sélection des individus au service militaire, mais il s'agissait de la seule variable dont on pouvait être certain qu'elle était réellement aléatoire et indépendante des autres caractéristiques des individus.

La disponibilité d'un tel instrument permet d'examiner ce que l'économètre est en mesure d'identifier comme effet causal à partir des données à sa disposition et surtout de préciser les hypothèses qui lui permettent d'interpréter certaines statistiques comme un effet causal. Considérons un instrument binaire $Z_i = 0,1$. Selon la valeur prise par l'instrument, chaque individu participe ou non au traitement : $D_i = D_i(0)(1 - Z_i) + D_i(1)Z_i$, où $D_i(Z_i)$ est la participation au traitement, conditionnellement à Z_i ¹¹. Deux types d'hypothèses sont alors posées :

- L'exogénéité de l'instrument : il est bien généré aléatoirement, indépendamment des variables de résultat : $Z_i \perp (Y_{1i}, Y_{0i}, D_i(0), D_i(1))$. C'est le cas de la variable de loterie qui consiste en un tirage aléatoire. On peut noter que le fait de participer, D_i , dépend de l'instrument Z_i . L'hypothèse d'indépendance entre Z_i et les variables $D_i(0)$ et $D_i(1)$ signifie ici que le fait de participer (quelle que soit la valeur qui serait prise par l'instrument Z_i) n'affecte pas l'instrument lui-même.

- L'effet de l'instrument sur la participation : on suppose que l'instrument a un effet toujours croissant (ou décroissant) sur la participation, $D_i(1) \geq D_i(0), \forall i$. Cette hypothèse signifie qu'il ne faut pas qu'il y ait de « defiers », c'est-à-dire d'individus qui souhaiteraient participer parce qu'ils n'ont pas été tirés au sort. Dans le cas de la loterie de la guerre du Vietnam, cette hypothèse semble raisonnable. Il faut cependant garder à l'esprit que cette hypothèse ne peut pas être testée, mais seulement argumentée. Il faut également qu'il y ait des « compliers », c'est-à-dire des individus qui participeront parce qu'ils ont été tirés au sort, *i.e.* tels que $D_i(1) > D_i(0)$.

Angrist introduit alors la notion d'effet local du traitement (LATE pour « Local Average Treatment Effect ») sur la population des « compliers » :

$$\tau_{LATE} = E(Y_{1i} - Y_{0i} | D_i(0) = 0, D_i(1) = 1) = \frac{E(Y_i | Z_i = 1) - E(Y_i | Z_i = 0)}{E(D_i | Z_i = 1) - E(D_i | Z_i = 0)}$$

Cette formule dont la preuve est présentée en annexe n°2 appelle plusieurs remarques. En premier lieu, elle fait le lien entre une variable d'intérêt faisant intervenir des variables contrefactuelles et l'expression d'espérances estimables simplement par des moyennes sur des populations identifiées par la valeur de l'instrument. Les hypothèses faites sur les instruments permettent d'établir ce lien. En second lieu, si l'effet local du traitement est bien estimé sur les « compliers », on ne sait pas qui, dans la population, en est effectivement un, car cette définition repose sur la différence de participation selon la valeur de l'instrument tirée, qui n'est pas observable. Le point qu'Angrist souligne ici est que, si l'effet causal est correctement estimé pour les « compliers » sous les hypothèses explicitées ci-dessus, d'autres hypothèses sont nécessaires pour généraliser à d'autres populations ce que l'on peut inférer du résultat. Si on reprend l'exemple utilisé par Angrist sur la guerre du Vietnam, on peut imaginer que l'effet du service militaire n'est pas le même sur les « volontaires » (c'est-à-dire participant quel que soit le tirage de la loterie) ou les « déserteurs » (ceux ayant échappé à la conscription malgré un tirage allant dans ce sens). Cette restriction est importante lorsque l'on envisage d'étendre un programme à des populations non concernées jusqu'alors, par exemple l'ensemble de la population.

Lorsque l'instrument n'est pas binaire, mais continu, on est en mesure d'identifier les effets d'une faible variation de la participation, on parle alors d'effet marginal du traitement. L'utilisation de cette méthode requiert une analyse approfondie du caractère exogène et pertinent de la variable

¹¹ La notation $D_i(Z_i)$ est ici adoptée pour distinguer le rôle différent de l'instrument Z_i sur la participation D_i de celui de la participation D_i sur le traitement Y_{D_i} .

instrumentale. Au final, ces méthodes permettent de mieux comprendre l'identification de l'effet causal.

Limites de la démarche non structurelle : identification vs interprétation

L'analyse d'Angrist sur l'usage des instruments dans le cadre d'évaluation présentée ci-dessus illustre très bien la limite de l'approche non structurelle. La recherche de la causalité pure restreint l'estimation des variables d'intérêt à un sous-groupe de la population qui n'est pas nécessairement pertinent pour le décideur public. On se trouve alors confronté au problème de la généralisation des résultats à des populations différentes, auxquelles on souhaiterait appliquer le traitement évalué. Keane [2010] a une critique de cette démarche encore plus radicale. Contrairement à ce qu'avance un grand nombre d'économistes adeptes de la démarche non structurelle, celle-ci fait des hypothèses économiques implicites, en particulier pour l'interprétation des résultats et, partant, leur capacité à les généraliser. Keane mobilise Rosenzweig et Wolpin [2000] qui reviennent sur l'instrument de loterie utilisé par Angrist [1999] pour examiner l'effet du service militaire sur les carrières futures des individus. Il ne remet pas en question l'effet négatif du service militaire sur les carrières futures mais son interprétation directe, sur laquelle Angrist s'avance peu. S'agit-il d'une perte en capital humain liée à un éloignement temporaire du marché du travail ? L'effet s'explique-t-il par le caractère traumatisant de la participation à des opérations de guerre ? Une autre interprétation pourrait consister en une baisse d'investissement en capital humain pour ceux ayant fait un mauvais tirage, ce qui violerait l'hypothèse selon laquelle l'instrument ne devrait pas directement affecter les résultats (restriction d'exclusion). Ces différentes interprétations ne sont pas testables avec le seul recours aux données. Seul le recours à des considérations externes (mode d'organisation de la loterie, autres enquêtes faites sur les individus, explorations de mécanismes théoriques, ...) permet d'apporter des arguments allant dans le sens de telle ou telle interprétation.

Keane imagine même des situations dans lesquelles l'utilisation d'instruments est impossible. Il prend l'exemple de la mesure de l'élasticité de l'offre de travail au salaire dans un modèle classique d'offre de travail. On peut imaginer une situation dans laquelle toute variable affectant les salaires va directement affecter le rendement du capital humain et ainsi les décisions d'offres de travail (cf. Imai et Keane [2004] pour une présentation complète de ce modèle). Si l'offre de travail est produite selon ce mécanisme, il n'est pas possible d'isoler une variable affectant le salaire et non l'offre de travail. Seul le recours à une démarche structurelle permet alors d'estimer les paramètres pertinents du modèle.

Keane (comme Heckman [2012]) distingue donc le problème d'identification de celui de l'interprétation. En se focalisant sur les questions d'identification de l'effet d'une politique, le risque est de laisser en suspens la question de son interprétation ou en tout cas de l'examiner avec moins d'attention. En outre, en se focalisant sur les questions d'identification sans considérer réellement les mécanismes en jeu, l'approche non structurelle seule ne permet pas de se prononcer dans les situations où le groupe de contrôle ou l'ensemble de l'économie est affecté par le programme, comme c'est le cas dans les effets de bouclage, lorsque l'effet agrégé du traitement conduit à modifier le comportement de l'ensemble des agents, au-delà de son effet direct sur les agents initialement seuls concernés¹², c'est-à-dire lorsque l'hypothèse SUTVA n'est pas vérifiée.

¹² Cf. par exemple plus bas l'effet de la réduction du temps de travail sur le Smic horaire qui a affecté l'ensemble de l'économie.

Lien entre les approches

Entre l'approche structurelle pure, qui spécifie tout le processus générateur des données presque indépendamment des données effectivement observées, et l'approche non structurelle, qui se focalise sur ce qui peut être identifié à partir des données disponibles, de nombreuses approches intermédiaires existent. Elles se reflètent dans les méthodes statistiques d'évaluation utilisées qui se caractérisent par la source de variabilité présente dans les données (en particulier l'exogénéité des variables explicatives) et le caractère général que l'on veut donner aux résultats.

Lien entre le modèle de Roy et le cadre causal de l'évaluation

Blundell et Costa Dias ([2009], JHR) présentent les principales méthodes d'évaluation, de la moins structurelle à la plus structurelle et présentent le lien entre le modèle économique, de type Roy, et les variables d'intérêt décrites dans le cadre causal de l'évaluation de Rubin. L'idée est de considérer des données générées à partir d'une version du modèle de Roy, tel qu'il est défini dans la première partie et d'appliquer à ces données les différentes techniques d'évaluation statistique présentées dans la seconde partie.

On réécrit ainsi la spécification structurelle du modèle de Roy :

$$\begin{cases} Y_{0i} = \beta + U_{0i} \\ Y_{1i} = \beta + \alpha_i + U_{1i} \\ D_i^* = Z_i\gamma + V_i \end{cases}$$

Par rapport à la spécification générale présentée plus haut, on autorise l'effet du traitement α_i à être individuel, et on omet les variables explicatives X_i pour simplifier. Par rapport aux variables d'intérêt de l'évaluation, on peut définir :

- L'effet moyen du traitement sur l'ensemble de la population, $\tau_{ATE} = E(\alpha_i)$
- L'effet moyen du traitement sur les traités : $\tau_{ATT} = E(\alpha_i | Z_i\gamma + V_i \geq 0)$
- Pour deux valeurs distinctes de Z , Z^* et Z^{**} , on peut définir l'effet local du traitement lorsque Z varie entre Z^* et Z^{**} , $\tau_{LATE} = E(\alpha_i | D_i(Z^{**}) = 1, D_i(Z^*) = 0) = E(\alpha_i | -Z^{**}\gamma \leq V_i \leq -Z^*\gamma)$.

Si le traitement avait le même effet sur tous les individus, *i.e.* $\alpha_i = \alpha$, ces différents effets seraient identiques entre eux. Surtout, on pourrait inférer de l'effet local du traitement, qui est le seul à être directement estimable, l'effet du traitement sur l'ensemble des individus.

L'ensemble de la spécification peut se réduire en une seule équation :

$$Y_i = \beta + \tau_{ATE}D_i + (U_i + D_i(\alpha_i - \tau_{ATE})) = \beta + \tau_{ATE}D_i + e_i$$

Cela ramène alors à une spécification classique pour les économètres, où toutes les variables sont directement observées et ne sont plus des résultats potentiels. Dans ce cadre, une sélection non aléatoire va se traduire par une corrélation non nulle entre e_i et D_i . Cette corrélation peut avoir deux sources. Elle peut être liée à une corrélation directe entre U_i et D_i , auquel cas on l'interpréterait comme un biais de variable omise qui affecterait à la fois la participation et les résultats. On peut alors examiner l'écart entre le coefficient d'intérêt que l'on souhaite identifier τ_{ATE} et le résultat d'une estimation par les moindres carrés ordinaires τ_{OLS} :

$$E[\tau_{OLS}] = \tau_{ATE} + E(U_i | D_i = 1) - E(U_i | D_i = 0)$$

En cas d'effet hétérogène du traitement, la corrélation entre α_i et D_i peut aussi contribuer à celle entre e_i et D_i . Elle s'interprète par le fait que les agents participent au programme en fonction de ce qu'ils en attendent. Ce mécanisme conduit à modifier le biais :

$$E[\tau_{OLS}] = \tau_{ATE} + E(U_i|D_i = 1) - E(U_i|D_i = 0) + E(\alpha_i - \tau_{ATE}|D_i = 1)$$

Le terme rajouté reflète que les individus traités ont un effet attendu du traitement supérieur à l'espérance sur l'ensemble des individus.

Les méthodes d'estimations : le rôle des données

Dans leur article, Blundell et Costa-Dias distinguent plusieurs cadres qui conditionnent l'utilisation des méthodes, en fonction de l'information des données disponibles. Moins la variabilité contenue dans les données permet d'identifier un effet causal, plus on rajoute de la structure pour pallier ce problème d'identification, ce qui augmente la complexité de l'estimation.

Le cadre idéal de l'évaluation se produit lorsque la participation au programme dépend d'une assignation aléatoire, ce que l'on ne rencontre que dans le cadre d'expériences contrôlées. Dans ce cas, les différentes variables d'intérêt peuvent être obtenues en comparant directement les groupes traités et non traités.

Ce cadre idéal ne se rencontre pas dans les situations se déroulant dans le monde réel. Néanmoins, on retrouve souvent dans le monde réel des situations dites de quasi-expériences ou d'« expériences naturelles », qui sont des situations où une population clairement identifiée se trouve « traitée » par exemple par un changement de législation. Si on dispose dans les données d'une caractérisation précise de la population traitée et d'une mesure de l'évolution des grandeurs d'intérêt, on peut utiliser la méthode de différences de différences, qui croise (le plus souvent) une dimension temporelle avec le traitement. Cette méthode repose sur une hypothèse plus faible de l'exogénéité de l'assignation par rapport à la variable de résultat : $E(U_{it_1} - U_{it_0}|D_i = 1) = E(U_{it_1} - U_{it_0}|D_i = 0)$. Cette hypothèse n'est parfois pas respectée, en particulier si les individus *a priori* concernés peuvent anticiper le traitement et modifier leur comportement. Cet effet a été mis en évidence par Ashenfelter [1978] qui examinait l'effet du programme de formation sur les salaires et qui avait noté une baisse des salaires des individus formés juste avant leur entrée en formation. Cette méthode peut aussi être faussée lorsque des tendances différenciées préexistent entre les groupes traités et non traités. L'examen de l'effet des radars locaux sur les accidents de la route (Roux et Zamora [2013]) illustre clairement cet écueil (*cf.* plus loin).

Un autre type d'expérience naturelle peut être utilisé, lorsque les données s'y prêtent : l'exploitation des discontinuités dans l'assignation des individus au traitement. On applique alors une « régression par discontinuité ». Ce contexte apparaît lorsqu'il existe un seuil au-delà duquel un individu est traité avec une probabilité plus élevée. On fait la distinction entre le « sharp design », où la probabilité passe de 0 à 1 au seuil et le « fuzzy design », où elle augmente de façon discontinue. L'existence de discontinuités offre un cadre adapté pour estimer un effet local du traitement autour du seuil et lui donner une interprétation causale. Néanmoins, ce n'est qu'autour du seuil que l'effet est estimé et une identification précise requiert beaucoup d'observations autour.

En l'absence d'expérience naturelle ou de caractéristiques des données permettant d'appliquer directement les méthodes d'évaluation, l'économètre doit se contenter d'hypothèses plus générales sur les données, qui sont plus difficiles à étayer. Ainsi, une méthode très populaire auprès des économètres est celle des méthodes d'appariement, qui reposent sur l'hypothèse d'indépendance conditionnelle $D \perp (Y_1, Y_0)|X$. Elle suppose que, une fois tenu compte des variables explicatives, l'assignation au traitement est aléatoire et indépendante des variables de résultat. Pour estimer l'effet du traitement sur les traités τ_{ATT} , on s'appuie sur le fait que, sous l'hypothèse d'indépendance conditionnelle, $E(U_i|X_i, D_i) = E(U_i|X_i)$, et que $0 < P(D_i = 1|X_i) < 1$ pour s'assurer qu'il existe des individus non traités comparables. On peut remarquer que l'identification de l'effet moyen du traitement τ_{ATE} requiert en plus une hypothèse sur l'hétérogénéité de l'effet du traitement : $E(\alpha_i|X_i, D_i) = E(\alpha_i|X_i)$ et toujours $0 < P(D_i = 1|X_i) < 1$. Cette méthode requiert beaucoup des données : les variables explicatives X_i introduites doivent capter toute la variabilité

liant Y_i et D_i . En outre, comme le montrent Heckman et Navarro-Lozano [2004], l'introduction de variables explicatives non pertinentes risque d'accroître le biais.

L'hypothèse d'indépendance conditionnelle postule implicitement une forme d'assignation des individus au traitement particulière. Alternativement, on peut postuler l'existence de variables instrumentales Z qui expliquent l'assignation, *i.e.* $P(D = 1|Z) \neq P(D = 1)$, mais pas les résultats, *i.e.* $E(U|Z) = E(U)$. Cela ramène au cadre présenté ci-dessus. Lorsque le traitement est homogène, l'estimation du modèle avec variables instrumentales permet d'obtenir $\tau_{ATT} = \tau_{ATE} = \tau_{LATE} = \alpha_i = \alpha$. Cette méthode se heurte à deux écueils. Le premier est pratique : il est difficile de trouver de bons instruments. Non seulement il faut qu'ils correspondent bien à des restrictions d'exclusion, mais il faut aussi qu'ils aient une variabilité suffisante pour obtenir des estimateurs précis, ce qui est rarement le cas en pratique. Le second écueil est plus méthodologique. En cas d'hétérogénéité de l'effet du traitement, on ne peut plus identifier que l'effet local du traitement τ_{LATE} entre deux valeurs de l'instrument si on ajoute comme hypothèse que l'effet du traitement α_i est également indépendant de l'instrument Z . Si l'instrument Z est continu, on peut définir l'effet marginal du traitement comme une limite de l'effet local du traitement lorsque les deux valeurs de Z considérées sont très proches. On peut exprimer l'effet marginal du traitement en fonction de l'effet qu'a Z sur la probabilité d'assignation $P(Z)$:

$$\tau_{MTE}(P(Z)) = E(\alpha|P(Z)) = \frac{\partial E(Y|P(Z))}{\partial P(Z)}$$

La contrepartie empirique de la dernière partie de l'équation peut être estimée si on dispose de suffisamment d'observations autour de la valeur de l'instrument considéré.

Enfin, plutôt qu'utiliser une méthode à variable instrumentale, on peut s'appuyer sur une méthode à fonction de contrôle. Cette méthode consiste à traiter l'endogénéité de la sélection comme un problème de variable omise : la fonction de contrôle introduite comme variable explicative supplémentaire permet de corriger le biais de sélection. Toutefois, pour appliquer cette méthode, il faut faire des hypothèses supplémentaires sur la distribution des erreurs U et V . Ici, on suppose qu'ils sont tirés dans une loi normale et que ρ est leur coefficient de corrélation. Ainsi, sous les hypothèses spécifiées plus haut, peut-on exprimer l'espérance conditionnelle au traitement de chaque résidu : $E[U|D = 1, Z] = \rho\lambda_1(Z\gamma)$, et $E[U|D = 0, Z] = \rho\lambda_0(Z\gamma)$ où $\lambda_j(\cdot)$ est une fonction facilement calculable, qui s'exprime en fonction de la cumulative et de la densité de la loi normale. En utilisant le fait que :

$$E(Y_i|D_i, Z_i) = \beta + D_i E(\alpha_i|D_i = 1, Z_i) + E(U_i|D_i, Z_i)$$

$$E(Y_i|D_i, Z_i) = \beta + D_i \tau_{ATT}(Z_i) + D_i E(U_i|D_i = 1, Z_i) + (1 - D_i) E(U_i|D_i = 0, Z_i)$$

on peut estimer l'équation suivante par moindres carrés ordinaires (si on dispose déjà d'un estimateur de γ obtenu en première étape) :

$$Y_i = \beta + D_i \tau_{ATT}(Z_i) + D_i \rho \lambda_1(Z_i \gamma) + (1 - D_i) \rho \lambda_0(Z_i \gamma) + r_i$$

En fait, cette méthode s'appuie sur les instruments et s'expose donc aux mêmes critiques que celle précédemment présentée. En particulier, on ne peut identifier que l'effet local du traitement sur les traités, et non l'effet sur l'ensemble de la population. En outre, elle s'appuie sur des hypothèses paramétriques concernant la distribution des erreurs qui la rendent très proches d'une estimation purement structurelle. En contrepartie, les estimateurs obtenus avec cette méthode sont plus précis, le gain de précision étant alors apporté par la structure introduite dans le modèle.

Les estimateurs présentés dans cette section requièrent des hypothèses de plus en plus fortes pour pallier l'absence de variabilité exogène dans les données. De façon plus positive, si on

dispose des données adéquates, on peut les utiliser pour vérifier si les hypothèses les plus contraignantes sont effectivement vérifiées. C'est ce que fait Lalonde [1986] qui, à l'aide de la disponibilité alors récente de données randomisées, a pu montrer que les hypothèses sur lesquelles s'appuyaient la plupart des études précédentes pour évaluer les effets des programmes de formation n'étaient pas vérifiées.

Deux approches mélangeant structurel et non structurel

Au-delà des différentes méthodes présentées ci-dessus qui illustrent la continuité qui existe entre les deux approches, Heckman [2010] et Chetty [2009] présentent des méthodes visant à utiliser les deux approches pour répondre au problème d'évaluation.

Heckman [2010] se situe dans le cadre du modèle de Roy et son approche ne prétend pas à être étendue à tous les modèles structurels. En particulier, l'action publique considérée ici correspond à la façon d'assigner des individus à des programmes déjà existants et dont on suppose qu'ils ont fait l'objet d'évaluations.

Son idée principale consiste à considérer comme paramètres structurels des paramètres estimés localement comme l'effet local moyen du traitement ou l'effet marginal du traitement. En effet, ces paramètres étant estimés à probabilité d'assignation donnée, ils sont exogènes par rapport à celle-ci. Si on est en mesure d'estimer l'effet local du traitement entre deux probabilités suffisamment proches $\tau_{LATE}(p_1, p_2)$ pour plusieurs valeurs de (p_1, p_2) possibles, on peut estimer l'effet de différentes politiques publiques en sommant les effets locaux du traitement avec une pondération reflétant le mode d'assignation (sous forme d'une probabilité) prévue par la puissance publique.

Il reste toutefois à trouver les instruments permettant d'estimer suffisamment d'effets locaux du traitement pour reconstruire les effets globaux de la politique à appliquer. En outre, cette approche ne permet pas de traiter d'éventuels effets de bouclage, qui ne peuvent pas être observés au niveau de l'estimation de chaque effet local, mais qui pourraient jouer un rôle important en cas d'assignation massive des individus au traitement.

Malgré ces limites, l'approche proposée par Heckman a l'avantage de s'appuyer sur une approche non structurelle pour identifier des paramètres dits structurels pouvant être utilisés pour répondre à certaines demandes d'évaluation de politique publique. Toutefois, elle conserve comme inconvénient de ne pas s'appuyer sur une interprétation des mécanismes en jeu.

Chetty [2009] décrit une approche alternative, fondée sur les « statistiques suffisantes ». Son idée principale est de se concentrer sur un jeu de politiques publiques particulier, dont les effets peuvent être appréciés à l'aide de paramètres pouvant être estimés à partir de formes réduites (le plus souvent des élasticités). Ces paramètres, appelés statistiques suffisantes, ne sont pas structurels car ils ne sont pas nécessairement invariants à toutes les politiques publiques, mais le sont par rapport à la politique considérée. Par rapport à l'estimation structurelle, l'approche par les statistiques suffisantes a trois avantages (*cf.* Chetty [2009]) : elle est plus facile à mettre en œuvre car requiert moins de données ou de variabilité que l'estimation d'un modèle structurel complet, elle nécessite moins d'hypothèses d'identification ou de spécification qu'un modèle structurel car se focalise sur quelques grandeurs particulières, enfin les mêmes statistiques suffisantes peuvent être compatibles avec différents modèles structurels ce qui leur donne un caractère plus robuste que l'estimation d'un seul modèle structurel. Cette approche a aussi les défauts de ses avantages : ne se fondant pas directement sur des paramètres structurels, la statistique suffisante à considérer ne sera pas la même selon la variable d'intérêt. Dans certains cas, il peut ne pas être possible de construire des statistiques suffisantes. Un autre inconvénient de cette approche consiste en la confiance excessive qu'on peut en avoir : dans la mesure où la statistique suffisante est robuste à des déviations du modèle structurel dont elle dérive, on peut croire à tort qu'il est inutile de vérifier si le modèle structurel est bien cohérent avec les données.

Deux exemples d'utilisation conjointe des approches structurelle et non structurelle

Les éléments présentés dans les parties précédentes restent très théoriques. Dans cette partie, nous nous appuyons sur deux recherches auxquelles nous avons participé pour leur donner une teneur plus concrète. La première vise à évaluer l'effet local des radars fixes sur les accidents de la route (Roux et Zamora [2013]), et la seconde examine les effets de la réduction du temps de travail sur l'emploi et la productivité des entreprises (Crépon, Leclair et Roux [2006]). Ces études ont comme point commun l'utilisation des approches relevant à la fois du structurel et du non structurel.

Les effets locaux des radars fixes sur les accidents de la route

L'objet de l'article de Roux et Zamora [2013] est d'examiner l'effet de l'installation des radars fixes sur les accidents de la route. Nous disposons comme informations le nombre d'accidents corporels par commune et par trimestre, ainsi que les lieux et dates d'installation des radars. Les problèmes auxquels nous sommes confrontés sont les suivants :

- le lieu précis de l'accident dans la commune n'est pas bien renseigné
- l'accident dans une commune donnée est un événement rare, ce qui conduit à beaucoup d'observations de valeur nulle par trimestre
- les sites où les radars ont été déployés sont hétérogènes
- le déploiement des radars n'est pas aléatoire, ils sont *a priori* placés aux endroits où le nombre d'accidents est le plus fréquent.

Le premier point nécessite de se limiter aux communes comme niveau de désagrégation le plus fin des unités géographiques, ce qui a pour conséquences que les estimations seront particulièrement imprécises pour les communes les plus grandes. La rareté de l'événement conduit à appliquer des méthodes économétriques adaptées aux données de comptage. L'hétérogénéité des sites conduit à adopter une stratégie d'identification en différences de différences. Enfin, une discussion sur le caractère potentiellement endogène du déploiement des radars est nécessaire.

L'article s'appuie sur le fichier national des accidents corporels de la sécurité routière : chaque observation correspond à un accident corporel (123 985 en 1998, 81 245 en 2007). La gravité des accidents est renseignée, *via* le nombre de blessés et de tués, comme la date et la commune du lieu de l'accident. Outre ces données, l'information sur les lieux et dates d'installation des radars était disponible, les lieux étant définis par la commune et le type de route où ils étaient installés (autoroute, route nationale, départementale ou communale). A partir de sources annexes (base de données communales de l'Insee), nous disposons également d'informations exogènes sur les villes (population, type d'unité urbaine ...).

Pour examiner l'effet local des radars, deux approches ont été tentées (*cf.* Roux et Zamora [2011]). Dans la première, un modèle explicitant l'effet des radars est construit, qui débouche sur la définition d'un paramètre « structurel » reflétant l'effet de la mise en place du radar, la spécification permettant aussi d'examiner son effet au cours du temps. Dans la seconde, l'approche est fondée sur la définition d'un groupe de contrôle adapté et les méthodes d'évaluation par appariement sont appliquées pour comparer l'évolution des accidents dans les communes traitées par rapport au groupe de contrôle.

Le modèle « structurel » a été développé avant tout pour pallier le manque de précision du lieu de l'accident dans la commune. Dans la mesure où l'information n'est disponible qu'à un niveau relativement agrégé, le modèle avait pour but de comprendre comment cette agrégation pouvait affecter les indicateurs disponibles. Cela nous conduit à adopter des spécifications particulières pour rendre compte de l'effet des radars, dont la forme résulte des hypothèses détaillées ci-dessous. C'est en cela que nous qualifions cette approche de « structurelle ». En revanche, nous ne spécifions pas ici

les raisons qui conduisent certaines communes plutôt que d'autres à s'équiper de radars, à la différence du modèle de Roy présenté plus haut.

Nous nous sommes appuyés sur l'idée selon laquelle le réseau routier communal est constitué de plusieurs lieux potentiellement accidentogènes et que le nombre d'accidents qui s'y produisent par trimestre est le produit d'une caractéristique fixe T_j , reflétant le trafic moyen sur le site, et d'un terme reflétant une tendance temporelle commune à tous les sites λ_t , qui peut s'interpréter comme les progrès de la sécurité routière communs à tous les sites. L'espérance du nombre moyen y_{jt} d'accidents au trimestre t sur le site j s'écrit ainsi $E(y_{jt}) = \lambda_t T_j$. Cette représentation est évidemment une approximation grossière de la réalité, dont on pense néanmoins qu'elle capte les deux principaux phénomènes expliquant les accidents à un endroit donné. Surtout, il nous semble que la spécification multiplicative est plus raisonnable qu'une spécification additive pour refléter l'évolution de la sécurité routière sur l'ensemble du réseau routier.

L'unité d'observation étant la commune, on peut utiliser notre spécification pour agréger l'ensemble des accidents à son niveau. Pour cela, nous faisons comme hypothèse supplémentaire que le niveau de trafic moyen dans la commune i est homogène, $T_j = T_i$. Cela reflète l'idée que les différences de trafic sont plus fortes entre communes qu'en leur sein sur les sites potentiellement accidentogènes. En outre, les communes peuvent se différencier en fonction du nombre de sites accidentogènes N_i , qui dépend de l'état du réseau routier (supposé fixe sur la période d'observation) et d'autres caractéristiques propres à la commune. Enfin, on suppose qu'équiper un site d'un radar conduit à y diminuer proportionnellement le nombre d'accidents d'un facteur $(1 - \alpha)$. Au final, en appliquant ces spécifications, on est en mesure d'exprimer l'espérance du nombre d'accidents se produisant dans la commune i lors du trimestre t : $E(y_{it}) = \lambda_t T_i (N_i - \alpha R_{it})$ où R_{it} est le nombre de radars en activité. On obtient alors une spécification non linéaire liant le nombre d'accidents dans la commune à deux effets fixes spécifiques à chaque commune caractérisant son nombre de sites accidentogènes et son niveau de trafic. Un tel modèle avec deux effets fixes communaux étant difficile à identifier et probablement peu robuste, nous avons choisi de simplifier la spécification en considérant deux hypothèses polaires. Soit on considère que le nombre de sites accidentogènes est le même entre les différentes communes, qui se différencient alors par rapport au trafic qu'elles connaissent, $N_i = N$. Ce sera le cas notamment lorsqu'on va considérer les petites communes où il y a le plus souvent un seul site accidentogène $N_i = 1$. Soit la variabilité du nombre de sites accidentogènes est plus forte que celle du trafic, auquel cas $T_i = T$. On obtient alors deux spécifications alternatives. Le premier modèle est purement multiplicatif $E(y_{it}) = \lambda_t T_i (1 - \alpha) R_{it}$, car ce modèle s'applique avant tout dans les communes les plus petites où il n'existe qu'un site accidentogène, et où donc un seul radar peut être installé. Le second modèle est mixte car ni multiplicatif, ni additif : $E(y_{it}) = \lambda_t T (N_i + \alpha R_{it})$ et s'applique plutôt aux grandes villes, où ce qui compte est la variabilité du nombre de sites accidentogènes plutôt que celle du trafic.

Chacune des spécifications est estimée par la méthode des moments généralisés appliquée à une forme non linéaire sous l'hypothèse que, une fois pris en compte les effets fixes, l'installation du radar est exogène. Outre ces spécifications, nous avons tenté une spécification additive :

$$y_{it} = \alpha R_{it} + \mu_i + \lambda_t + \varepsilon_{it}$$

qui correspond à une spécification standard dans laquelle des effets fixes « commune » et des effets temporels sont introduits de façon additive. Cette spécification correspond à une différence de différences. L'effet des radars est alors estimé en rapportant la variation du nombre d'accidents moyens avant et après la mise en place du radar. Or cette spécification suppose que la tendance sous-jacente à l'évolution des accidents est additive, c'est-à-dire qu'elle est la même quelle que soit la taille de la commune. Cette hypothèse est testable et a été fortement rejetée par les données. C'est justement cela qui nous a conduits à mieux décrire les mécanismes par lesquels l'installation de radars fixes pouvait impacter les accidents au niveau communal. Nous sommes ici dans une situation

où la méthode de différence de différences, lorsqu'elle s'appuie sur une spécification additive des phénomènes en jeu, ne permet pas de rendre compte de l'effet des radars.

Comme cela a été expliqué précédemment, l'approche structurelle conduit à faire dépendre les résultats de la forme fonctionnelle retenue. Cependant, en explicitant les hypothèses qui permettent d'arriver à une spécification donnée, elle permet de proposer des spécifications alternatives dont on peut comparer les résultats. Ces différentes spécifications peuvent servir à s'assurer de la robustesse des résultats. Elles ont comme finalité première de mieux comprendre les mécanismes en jeu. La présentation du modèle ci-dessus décrit les différentes étapes de sa construction. Elle met en évidence les nombreuses hypothèses simplificatrices qui permettent de construire le modèle. Si elles aboutissent à une forme fonctionnelle relativement simple, la principale source d'identification de l'effet des radars reste la comparaison des tendances d'accidents entre communes traitées et non traitées. Néanmoins, nous avons vu que lorsque la forme fonctionnelle devient trop éloignée du phénomène, comme cela est le cas pour la forme additive, les résultats obtenus peuvent différer largement. Dans notre cas, la baisse tendancielle des accidents de la route s'est appliquée de façon proportionnelle à l'ensemble des communes qu'elles soient traitées ou non. En voulant rendre compte de cette baisse *via* un modèle additif, on surestime la baisse dans les communes traitées (où le nombre d'accidents est plus élevé) par rapport aux communes non traitées (cf. tableau n°1 dans Roux et Zamora [2013]).

Au-delà de la forme fonctionnelle, il subsiste encore un risque de biais de variable omise, lié au fait que l'installation d'un radar pourrait dépendre d'une accidentologie temporairement plus forte qui conduirait justement à cette installation (effet nommé « Ashenfelter's dip » dans la littérature, Ashenfelter [1978]). Le mécanisme de mise en place de nouveaux radars était censé éviter ce type de phénomène. En outre, il est toujours possible que la structure introduite dans le modèle, *via* les hypothèses décrites dans sa construction, conditionne les résultats. Pour se prémunir contre ce risque, nous avons mis en place une stratégie d'identification différente (Roux et Zamora [2011]) ne s'appuyant pas sur des formes fonctionnelles appliquées à l'ensemble des communes, mais plutôt sur la définition d'un groupe de contrôle composé des communes pas encore équipées, mais qui le seront dans un futur proche, à quelques mois d'intervalle. Plus cet intervalle est faible, plus le groupe de contrôle est comparable aux communes équipées, mais moins leur nombre est élevé. Il y a donc un arbitrage à trouver entre la proximité du groupe de contrôle que l'on veut avoir avec le groupe traité et le nombre d'observations que l'on veut conserver dans le groupe de contrôle. L'inconvénient de cette approche est qu'elle risque de donner des résultats peu précis, du fait de la perte d'observations pouvant contribuer à l'identification de l'effet. En outre, à la différence de l'estimation structurelle, on ne peut plus examiner l'effet des radars à plus long terme, puisque l'on ne peut plus définir alors de groupe de contrôle crédible. Il est néanmoins possible de comparer les deux approches sur les effets des radars au cours des six premiers mois de leur installation.

Le tableau n°1 (repris du tableau n°12 de Roux et Zamora [2012]) présente les effets de l'installation des radars en comparant les deux méthodes, structurelle et non structurelle. Il montre que les deux approches aboutissent à des résultats qualitativement similaires : les radars ont un effet négatif significatif sur les accidents dans les six mois suivant leur mise en place. Toutefois, des différences quantitatives apparaissent : l'ampleur de l'effet est plus faible avec la méthode non paramétrique, en particulier lorsque l'on exclut les autoroutes du champ de l'étude. Par ailleurs, l'approche structurelle apporte des résultats plus précis que l'approche non paramétrique.

Tableau 1

Impact de l'installation des radars dans les six premiers mois dans les communes de moins de 6000 habitants ; comparaison des approches structurelle et non paramétrique

	Approche structurelle		Approche non paramétrique	
	Modèle mixte	Modèle multiplicatif	E=6 mois	E=12 mois
Toutes routes				
Estimation sur l'échantillon complet	-0,114 (0,019)	-0,126 (0,029)		
Estimation sur le sous-échantillon des communes qui seront équipées	-0,103 (0,021)	-0,123 (0,031)	-0,070 (0,035)	-0,080 (0,032)
Toutes routes (sauf autoroutes)				
Estimation sur l'échantillon complet	-0,130 (0,021)	-0,149 (0,034)		
Estimation sur le sous-échantillon des communes qui seront équipées	-0,105 (0,023)	-0,113 (0,030)	-0,074 (0,043)	-0,061 (0,035)

Source : Roux et Zamora [2011].

Note : Dans les communes de moins de 6000 habitants, l'installation des radars a fait diminuer en moyenne de 0,114 le nombre d'accidents par trimestre dans les communes venant d'être équipées, d'après le modèle mixte et de 0,126 d'après le modèle multiplicatif, par rapport à une situation où ils n'auraient pas été installés (estimations reconstruites en appliquant le modèle structurel à l'ensemble des communes).

Néanmoins, les deux approches amènent des résultats plus proches lorsque l'on restreint le calcul de l'effet des radars au groupe de contrôle utilisé dans l'approche non paramétrique. Dans ce cas, les différences ne sont plus statistiquement significatives.

Les effets de la réduction du temps de travail sur l'emploi et la productivité des entreprises

L'objet de l'article de Crépon, Leclair et Roux [2004] était d'évaluer les effets de la réduction du temps de travail dont la mise en place a été progressive jusqu'à la loi du 19 janvier 2000, dite « Aubry II » qui a abaissé la durée hebdomadaire légale du travail à 35 heures. Auparavant, la loi Robien (du 11 juin 1996) et surtout la loi dite « Aubry I » (du 13 janvier 1998) offraient un cadre incitatif aux entreprises afin de les amener à réduire leur temps de travail. Fin 2000, les entreprises pouvaient alors se retrouver dans différentes situations :

- Dans le cadre d'accords d'entreprise signés jusqu'au 01/01/2000 prévus par la loi « Aubry I », elles pouvaient procéder à une réduction effective de leur durée du travail d'au moins 10%, sans redéfinition du temps de travail. En contrepartie d'engagement à créer au moins 6% d'emploi, elles pouvaient bénéficier immédiatement d'aides conséquentes.
- D'autres entreprises, dites « Aubry II précurseurs » pouvaient réduire leur temps de travail en le redéfinissant partiellement, ce qui a conduit à une baisse d'environ 6%, non directement observable entreprise par entreprise. Ces entreprises n'ont pas bénéficié d'aides autres que celles prévues par la loi Aubry II.
- Suite à la loi Aubry II, les entreprises pouvaient réduire leur temps de travail, y compris en le redéfinissant, dans le cadre d'accords signés après le 01/01/2000. Elles pouvaient alors bénéficier d'aides structurelles (sous la forme d'une diminution progressive des charges jusqu'à 1,8 Smic + 4000F), ainsi que toutes celles qui avaient déjà procédé à une réduction (Aubry I et Aubry II précurseurs).
- Certaines entreprises pouvaient rester à 39 heures, auquel cas elles ont dû augmenter les salaires car les heures au-delà de 35 heures devaient être considérées comme des heures supplémentaires.

L'ensemble de ces situations n'a concerné que les entreprises de 20 salariés et plus en 2000, les entreprises en dessous de ce seuil étant concernées par une loi mise en place ultérieurement. Par

ailleurs, la loi a aussi limité l'usage des heures supplémentaires en abaissant le plafond autorisé. On peut noter que le passage à 35 heures dans un grand nombre d'entreprises a directement conduit à une hausse des salaires horaires, qui, *via* les mécanismes de revalorisation du Smic, a conduit à sa forte augmentation.

Le contexte de l'évaluation de cette réforme est similaire au modèle de Roy présenté en première partie : les entreprises avaient le choix entre les différentes situations présentées ci-dessus. On peut aussi retrouver le cadre causal de l'évaluation présenté en deuxième partie pour essayer de répondre à la question de l'effet contrefactuel de la réduction du temps de travail.

La demande d'évaluation portait non seulement sur l'emploi et les salaires mais aussi sur la productivité des entreprises. Il s'agissait d'abord d'avoir des données à notre disposition permettant d'observer ces grandeurs. Nous nous sommes appuyés sur les données fiscales sur les comptes des entreprises (BRN) qui donnent une information pour chaque entreprise sur la valeur ajoutée, le stock de capital détenu par les entreprises et l'emploi, sur les déclarations administratives de données sociales, qui apportent une information précise sur la composition des entreprises en termes de qualification d'âge et de sexe, ainsi que sur les bases des accords d'entreprises sur le temps de travail recensés par le ministère du Travail. Du fait que beaucoup d'entreprises avaient modifié la définition de leur durée du travail pour avoir une baisse affichée d'environ 10%, l'information sur le nombre d'heures travaillées que les entreprises reportaient ne pouvait pas être utilisée.

Afin de tenir compte de ces restrictions sur les données, et pour mieux comprendre les objets examinés dans l'article, nous avons écrit un modèle explicitant les canaux par lesquels la réduction du temps de travail pouvait fonctionner. On considère ainsi une fonction de production à deux facteurs, capital K et travail L , où $L = HN$, N étant le nombre d'employés dans l'entreprise et H leur durée de travail. Le temps de travail peut affecter la productivité horaire *via* deux canaux : l'usage du capital, dont le taux d'utilisation $d(H)$ dépend du temps de travail des employés, et l'efficacité horaire des salariés $e(H)$ qui a une forme concave. Si H est trop faible, l'efficacité horaire est croissante, *via* des effets de mise en place par exemple ; si H est trop élevée, l'efficacité horaire devient décroissante à cause de l'effet de fatigue.

En tenant compte de tous ces mécanismes, la fonction de production d'une entreprise s'écrit de la façon suivante :

$$Y = F(d(H)K, e(H)HN)$$

Au final, on s'appuie sur une forme fonctionnelle de type Cobb-Douglas avec rendements constants pour aboutir à la spécification suivante :

$$Y = AK^{1-\beta}N^{\beta}H^g$$

où g est l'élasticité de la production au temps de travail et β celle à l'emploi. Elle mélange les mécanismes correspondant aux deux canaux décrits ci-dessus. Exprimée en log-différences, on peut estimer l'évolution de la productivité globale des facteurs (PGF) par tête à partir de l'équation précédente

$$\Delta PGF = \Delta \ln Y - \beta \Delta \ln N - (1 - \beta) \Delta \ln K = \Delta \ln A - g \Delta \ln H$$

Dans cette spécification, $-g \Delta \ln H$ correspond à l'effet de la réduction du temps de travail sur la PGF par tête. Comme on ne peut pas observer la réduction effective du temps de travail dans les entreprises, on utilise directement l'information sur le mode de passage à 35 heures. L'équation estimée en forme réduite est alors :

$$\Delta PGF = X\gamma + g_{AI}1_{\text{AubryI}} + g_{AIIp}1_{\text{AubryIIp}} + g_{AII}1_{\text{AubryII}} + \varepsilon$$

X correspond à des variables de contrôle estimées au début de la période, g_{AI} est par exemple l'effet sur la PGF d'avoir réduit la durée du travail *via* un accord Aubry I, $g_{AI} = -g\Delta \ln H_{\text{AubryI}}$. Ce coefficient s'interprète donc comme la perte de productivité par tête induite par la réduction du temps de travail, qui, si cette baisse est inférieure à 10%, peut en fait correspondre à une hausse de la productivité horaire.

Le cadre présenté ci-dessus permet de donner une interprétation à l'évolution de la PGF en fonction du cadre choisi par les entreprises pour réduire leur temps de travail. La PGF est ici interprétée comme un paramètre technologique dont l'évolution n'est affectée que par la baisse du temps de travail et est indépendante des autres facteurs qui affectent la production de l'entreprise.

La même spécification, simple différence de différences, peut être utilisée pour examiner les effets sur d'autres variables du passage à 35 heures. Le tableau n°2 reproduit les résultats obtenus dans l'article. Ces résultats ont été obtenus en appliquant simplement la technique des moindres carrés ordinaires et en ne contrôlant que des variables observables. Ces résultats sont robustes à l'utilisation de méthodes par appariement, en considérant comme groupe de contrôle les entreprises restées à 39 heures.

La modélisation des différents mécanismes a permis ici d'aboutir à une équation dont les paramètres peuvent être interprétés. Comme dans l'exemple précédent, la structure apportée par ces formes fonctionnelles permet de résumer l'effet du traitement (ici la réduction du temps de travail) sur la variable d'intérêt en un paramètre.

Tableau 2

Evolutions relatives des entreprises passées à 35 heures par rapport à celles restées à 39 heures

	Aubry I aidées		Aubry II précurseurs		Aubry II	
Emploi	9,91	(0,33)	3,78	(0,67)	4,89	(0,37)
PGF	-3,68	(0,31)	-1,79	(0,62)	-0,67	(0,34)
Valeur ajoutée	5,02	(0,39)	1,28	(0,80)	3,76	(0,44)
Intensité du capital (K/L)	-5,71	(0,44)	-3,38	(0,89)	-2,34	(0,49)
Salaire hebdomadaire	-2,54	(0,23)	-2,45	(0,46)	-1,49	(0,25)
Coût salarial hebdomadaire	-5,97	(0,23)	-4,26	(0,46)	-1,98	(0,25)
Coût salarial par unité de production	-1,08	(0,24)	-1,77	(0,48)	-0,84	(0,26)
Différence entre TFP et coût salarial	-2,29	(0,25)	-2,47	(0,50)	-1,31	(0,27)

Source : tableau n°2 de Crépon, Leclair et Roux [2004].

Groupe de référence : firmes de 20 salariés et + encore à 39 heures à la fin de 2000.

Variables de contrôle : taille, secteur, proportion de salariés en fonction de leurs qualification et âge, part des coûts salariaux dans la valeur ajoutée en 1997.

Trois phénomènes, qui ne peuvent pas être contrôlés par des caractéristiques observables, sont susceptibles d'affecter les résultats :

- L'hétérogénéité inobservée : l'évolution des variables économiques et le choix d'un dispositif peuvent être liés à des caractéristiques inobservables gouvernant l'une et l'autre des variables.
- L'hétérogénéité de l'effet du traitement : l'effet du traitement peut différer entre entreprises et expliquer en partie leur choix de se placer dans un dispositif de réduction du temps de travail plutôt que dans un autre. Dans ce cas, il n'est pas possible d'inférer les effet de la réduction du temps de travail que l'on observerait si toutes les entreprises étaient obligées de réduire leur temps de travail.
- Les effets de bouclage : la réduction du temps de travail a aussi concerné indirectement les entreprises restées à 39 heures, du fait du caractère massif que cette politique a eu. L'effet le plus direct a consisté en l'augmentation du salaire horaire qui a lui-même augmenté le Smic et donc

directement augmenté les coûts salariaux des entreprises restées à 39 heures. On est donc dans une situation où le groupe de contrôle a été affecté.

Les deux premiers risques de biais sont liés au caractère potentiellement endogène de l'assignation au traitement. Afin d'explorer ces risques dans un cadre simple, nous avons examiné dans l'article plus précisément le cas de la productivité globale des facteurs et le choix entre rester à 39 heures et réduire son temps de travail dans le cadre Aubry I. Dans la mesure où nous interprétons ce paramètre comme étant purement technologique, ses variations peuvent être interprétées comme les conséquences de la réduction du temps de travail et ne sont pas maîtrisables comme d'autres variables telles que l'emploi ou les salaires. Néanmoins, il est toujours possible que certaines variables inobservables expliquent le choix de réduire la durée du travail et l'évolution de la productivité. C'est pourquoi nous avons utilisé deux approches : celle par maximum de vraisemblance, présentée dans le cadre du modèle de Roy dans la première partie de cet article, et celle par la fonction de contrôle, présentée dans la troisième partie.

Dans tous les cas, ces approches requièrent la disponibilité d'instruments (ou variables d'exclusion) qui expliquent le choix de réduire la durée du travail sans affecter théoriquement la variable d'intérêt. Le fait que nous considérons ici l'évolution de la PGF nous guide dans le choix des instruments. Nous utilisons ainsi le niveau des subventions que les entreprises ont ou auraient obtenues en réduisant leur temps de travail. En effet, ce paramètre semble important pour expliquer les raisons poussant les entreprises à réduire leur temps de travail, et il ne devrait pas directement affecter l'évolution de la PGF du fait de son caractère de paramètre technologique¹³. En outre, il peut être estimé *ex ante* pour toutes les entreprises, car nous disposons pour chacune de la distribution des salaires. Un tel instrument ne fonctionne pas si on considère une autre variable telle que l'emploi ou les salaires, dont on ne peut soutenir l'hypothèse selon laquelle ils ne seraient pas directement affectés par l'instrument.

Pour s'assurer de la robustesse des résultats, nous avons considéré deux autres instruments : la part d'entreprises du même secteur (en 114 postes différents) ayant signé un accord de réduction du temps de travail, et une indicatrice d'un ratio de dette sur fonds propres supérieur à 50%. Les résultats obtenus suggèrent qu'il n'y a pas de problème d'hétérogénéité inobservée, mais qu'il y a un problème d'hétérogénéité de l'effet du traitement : les entreprises passées à 35 heures seraient celles pour lesquelles la réduction du temps de travail avait le moins d'effets négatifs sur leur productivité. Selon ce constat, il est possible d'évaluer l'effet qu'a eu la réduction du temps de travail pour les entreprises l'ayant appliquée, mais pas de connaître l'effet qu'elle aurait eue sur les entreprises ne l'ayant pas appliquée. On retrouve ici le résultat selon lequel en cas d'hétérogénéité d'effet du traitement, seul l'effet local du traitement peut être identifié.

Considérer l'emploi comme variable dépendante est plus délicat car la variable instrumentale qui consistait en subventions potentiellement perçues par l'entreprise en cas de réduction du temps de travail a très probablement un effet direct sur l'emploi des entreprises qui les perçoivent. Néanmoins, elle reste pertinente pour les entreprises restées à 39 heures, c'est-à-dire non traitées, qui ne les ont pas perçues. Il est donc possible d'estimer le modèle sur les entreprises restées à 39 heures, ce qui nous a permis de tester l'hypothèse de l'absence de biais d'hétérogénéité inobservable, qui n'a pas été rejetée par les données. En revanche, sans variable d'exclusion crédible pour les entreprises ayant réduit leur temps de travail, nous ne sommes pas en mesure de tester si un biais d'hétérogénéité de l'effet du traitement est susceptible d'affecter les résultats.

Toutefois, un autre problème est susceptible d'affecter les résultats portant sur l'emploi : les effets de bouclage. En effet, la mise en place de la réduction du temps de travail, en augmentant les salaires horaires dans les entreprises concernées, a contribué à augmenter le Smic, ce qui a conduit à augmenter les salaires dans les entreprises restées à 39 heures. Le groupe de contrôle a donc été

¹³ Si le niveau de la PGF est bien corrélé au niveau moyen des salaires et donc au niveau de subvention dont les entreprises ont bénéficié, nous supposons que ce n'est pas le cas de son évolution liée à la réduction du temps de travail.

affecté par la politique conduite. Cet effet peut être évalué en appliquant l'élasticité de l'emploi au coût du travail à l'évolution du Smic induite par la réduction du temps de travail. S'il existe, cet effet est mineur et n'affecte pas les interprétations obtenues.

Au final, nous obtenons des résultats qui reflètent l'évolution des principales grandeurs en présence de RTT, que nous pouvons dans une certaine mesure mettre en regard les unes des autres. Pour l'interprétation de ces effets, il faut réfléchir aux mécanismes économiques. Dans la fin de cet article, nous avons ainsi proposé deux situations possibles, chacune pouvant être testée par les données. Dans la première situation, dite « classique », l'évolution de l'emploi dépend des gains de productivité et négativement des salaires. En reprenant les notations ci-dessus, l'évolution de l'emploi a l'expression suivante :

$$\Delta \ln N = \frac{\Delta PGF - \Delta \ln w}{1 - \beta}$$

Le choix d'emploi des entreprises dépendant alors de la maximisation de leur profit. Cette situation se produit lorsque l'entreprise n'est pas contrainte par la demande qu'elle perçoit. Alternativement, lorsque l'entreprise est contrainte, l'évolution de l'emploi va être directement opposée à celle des gains de productivité

$$\Delta \ln N = - \frac{\Delta PGF}{\beta}$$

Il s'agit du modèle de partage de travail. En comparant les évolutions de l'emploi, des gains de productivité et des coûts salariaux induites par la réduction du temps de travail, on est en mesure de choisir entre l'une ou l'autre des interprétations. Le tableau n°3 présente les effets sur l'emploi de la réduction du temps de travail dans chacun des deux cadres, le modèle classique et le modèle de partage du travail, à partir des estimations de ses effets sur l'emploi, la PGF, les salaires et le coût salarial obtenus dans le tableau n°2. Les élasticités de l'emploi au coût salarial β sont estimées à la médiane par secteur du ratio masse salariale sur valeur ajoutée calculé pour chaque entreprise. La première ligne du tableau présente l'effet sur l'emploi de la réduction du temps de travail, en tenant compte des effets de bouclage¹⁴. Les trois lignes suivantes décomposent les effets sur l'emploi dans le modèle classique en distinguant les effets de productivité, de modération salariale (effet de la RTT sur les salaires) et des aides (réduction du coin socialo-fiscal). La ligne test du modèle compare l'effet sur l'emploi observé à celui prévu par le modèle classique. Les lignes suivantes présentent les effets sur l'emploi selon le modèle du partage du travail et la comparaison avec les effets observés.

D'après ce tableau on peut conclure que le mécanisme de partage du travail est rejeté pour les entreprises ayant réduit leur temps de travail *via* des accords Aubry I ou Aubry II. Il serait en revanche accepté pour les entreprises ayant réduit leur temps de travail dans le cadre Aubry II précurseur, ces entreprises étant minoritaires.

Au final, la plus forte augmentation de l'emploi dans les entreprises ayant réduit leur temps de travail serait principalement liée aux mesures ayant limité l'augmentation des coûts salariaux. En fait, les entreprises passées à 35 heures dans les cadres Aubry I ou Aubry II en auraient bénéficié, car les pertes de productivité (relativement faibles) ont été plus que compensées par la baisse du coût du travail induite par les aides et la modération salariale.

On voit ici l'intérêt d'une démarche modélisatrice, qui permet de mieux comprendre les mécanismes en œuvre au-delà de l'évaluation de l'effet de cette politique publique.

¹⁴ Les effets de bouclage sont contrôlés en ôtant de la différence de croissance d'emploi entre les entreprises ayant réduit leur temps de travail et celles restées à 39 heures la partie attribuée à l'évolution du Smic pour les entreprises restées à 39 heures (qui aurait contribué à réduire leur emploi). La différence avec les effets hors bouclage reste dans tous les cas de second ordre.

Tableau 3

Interprétation des effets sur l'emploi de la réduction du temps de travail

	Aubry I aidé		Aubry II précurseur		Aubry II	
	Coefficient	St. Dev	Coefficient	St. Dev	Coefficient	St. Dev
Croissance de l'emploi	9,6	(0,34)	3,31	(0,73)	4,97	(0,38)
	Modèle classique					
Effet des aides	16,95	(0,28)	9,29	(0,54)	2,65	(0,27)
Modération salariale	11,58	(1,24)	10,52	(2,66)	6,51	(1,34)
Perte de productivité	-18,09	(1,57)	-9,98	(3,37)	-3,4	(1,76)
Test du modèle	-0,84	(1,21)	-6,51	(2,58)	-0,78	(1,37)
	Modèle de partage du travail					
Effet de partage du travail	4,64	(0,39)	-2,48	(0,84)	0,81	(0,45)
Test du modèle	4,96	(0,47)	0,84	(1,03)	4,16	(0,50)

Source : Crépon, Leclair et Roux [2004].

Conclusion

Cette contribution a pour essentiel but de poser les termes d'un débat qui oppose les tenants d'une approche structurelle (Keane, Wolpin ...) et d'une approche « athéorique » (Angrist, Pischke, Imbens ...). Dans un premier temps les démarches structurelle et athéorique ont été présentées et leurs avantages et limites discutés, en s'appuyant sur les contributions de chacun. A la lumière des contributions de Blundel et Costa Dias [2009] et Heckman [2010], il apparaît que ces deux approches ne sont pas si faciles à distinguer. Par les méthodes statistiques utilisées, il peut exister un *continuum* entre une approche complètement empirique et « athéorique » et une approche complètement structurelle. Les deux exemples présentés illustrent le fait que considérations d'évaluation liées aux données et réflexion sur les mécanismes en jeu sont utiles pour arriver à des conclusions sur les effets d'une politique publique donnée. Idéalement, les deux approches devraient être utilisées, ne serait-ce que pour tester leurs conclusions de façon croisée¹⁵.

Au cours des vingt dernières années, l'approche non structurelle s'est fortement développée, avec une plus grande attention portée aux données et plus généralement aux faits stylisés, après un ascendant relatif de l'économie structurelle.

Si on souhaite résumer les deux approches, on peut considérer que l'approche structurelle permet d'évaluer les politiques publiques dans la mesure où elle se fonde sur les mécanismes les plus pertinents pour décrire la situation qu'elle évalue. La validité du modèle est alors cruciale, et nécessite d'être argumentée. Faute de validité du modèle, le risque est de se tromper lourdement dans l'évaluation. L'approche non structurelle permet d'évaluer les effets d'une politique donnée dans un contexte précis, en s'appuyant sur la variabilité contenue dans les données. L'interprétation des résultats est alors cruciale pour être en mesure de généraliser à d'autres contextes les conclusions obtenues.

En caricaturant de façon peut-être extrême, l'approche structurelle cherche à comprendre les effets de la politique publique évaluée, tandis que l'approche non structurelle cherche à les connaître. Les deux sont nécessaires pour que l'évaluation soit utile à la décision publique.

¹⁵ Cf. Todd et Wolpin [2006] qui utilisent des expériences randomisées pour tester la validité d'un modèle structurel.

Annexe 1

Ecriture de la vraisemblance du modèle de Roy

Calcul des probabilités $P(D = 0, Y_0)$ et $P(D = 1, Y_1)$

$$\begin{aligned} P(D = 0, Y_0) &= P(D^* < 0, U_0 = Y_0 - X\beta_0) \\ &= P(Z\gamma < -V, U_0 = Y_0 - X\beta_0) \\ &= P\left(Z\gamma < -\frac{\rho_0}{\sigma_0}(Y_0 - X\beta_0) + V_0, U_0 = Y_0 - X\beta_0\right) \\ &= \left[1 - \Phi\left(\frac{Z\gamma + \frac{\rho_0}{\sigma_0}(Y_0 - X\beta_0)}{\sqrt{1 - \rho_0^2}}\right)\right] \frac{1}{\sigma_0} \varphi\left(\frac{Y_0 - X\beta_0}{\sigma_0}\right) \end{aligned}$$

où $\varphi(\cdot)$ est la densité de la loi normale centrée réduite et $\Phi(\cdot)$ sa cumulative. On a également

$$P(D = 1, Y_1) = \Phi\left(\frac{Z\gamma + \frac{\rho_1}{\sigma_1}(Y_1 - X\beta_1)}{\sqrt{1 - \rho_1^2}}\right) \frac{1}{\sigma_1} \varphi\left(\frac{Y_1 - X\beta_1}{\sigma_1}\right)$$

La vraisemblance s'en déduit :

$$L = \prod_{i=1}^N P(D_i = 1, U_{1i} = Y_{1i} - X_i\beta_{1i})^{D_i=1} P(D_i = 0, U_{0i} = Y_{0i} - X_i\beta_{0i})^{D_i=0}$$

Annexe 2

Dérivation de la formule LATE (cf. Imbens et Angrist [1994])

Sous l'hypothèse qu'il n'y a pas de « defiers » (cf. p11), la quantité $D_i(1) - D_i(0)$ ne peut prendre que deux valeurs, 0 ou 1. Dans ce dernier cas, $D_i(0) = 0$ et $D_i(1) = 1$.

Ainsi,

$$\begin{aligned} \tau_{LATE} &= E(Y_{1i} - Y_{0i} | D_i(0) = 0, D_i(1) = 1) \\ &= E(Y_{1i} - Y_{0i} | D_i(1) - D_i(0) = 1) = \frac{E\left((Y_{1i} - Y_{0i})(D_i(1) - D_i(0))\right)}{P(D_i(1) - D_i(0) = 1)} \end{aligned}$$

Or,

$$E(Y_i | Z_i = 1) - E(Y_i | Z_i = 0) = E(D_i(1)Y_{1i} + (1 - D_i(1))Y_{0i} | Z_i = 1) - E(D_i(0)Y_{1i} + (1 - D_i(0))Y_{0i} | Z_i = 0)$$

Du fait de l'indépendance conditionnelle de Z_i par rapport aux autres variables, cette quantité s'écrit :

$$E\left((D_i(1) - D_i(0))(Y_{1i} - Y_{0i})\right)$$

De même,

$$E(D_i | Z_i = 1) - E(D_i | Z_i = 0) = E(D_i(1) | Z_i = 1) - E(D_i(0) | Z_i = 0) = E(D_i(1) - D_i(0))$$

qui est égal à la probabilité $P(D_i(1) - D_i(0) = 1)$, puisque $D_i(1) - D_i(0)$ ne peut prendre comme valeurs que 0 et 1.

Références

- J. D. Angrist et J-S. Pischke [2010] : *The Credibility Revolution in Empirical Economics : How Better Research is Taking the Con out of Econometrics*, **Journal of Economic Perspective**, vol. 24(2), pp 3-30.
- O. Ashenfelter [1978] : *Estimating the Effect of Training Programs on Earnings*, **Review of Economics and Statistics**, vol. 60(1), pp 47-57.
- Y. Ben-Porath [1967] : *The Production of Human Capital and the Life Cycle of Earnings*, **Journal of Political Economy**, vol. 75(1), pp 352-365.
- R. Blundell et M. Costa Dias [2008] : *Alternative Approaches to Evaluation in Empirical Microeconomics*, **Journal of Human Resources**, vol. 44(3), pp. 565-640.
- R. Chetty [2009] : *Sufficient Statistics for Welfare Analysis: A Bridge between Structural and Reduced-Form Methods*, **Annual Review of Economics**, 1: 451-488.
- D. R. Cox [1958] : **Planning of Experiments**, New York, Wiley.
- B. Crépon, M. Leclair et S. Roux [2004] : *RTT, productivité et emploi : nouvelles estimations sur données d'entreprises*, **Economie et statistique**, n°376-377.
- A. Deaton [2010] : *Instruments, Randomization and Learning about Development*, **Journal of Economic Literature**, 48:2, pp. 424-455.
- E. Duflo [2009] : *Leçon inaugurale au collège de France de la chaire « Expérience, Science et Lutte contre la Pauvreté »*.
- C. Gouriéroux et A. Monfort [1996] : **Simulation-Based Econometric Methods**. Oxford University Press.
- J. J. Heckman [2010]: *Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy*, **Journal of Economic Literature**. 48:2, pp. 356-398.
- J. J. Heckman [1990] : *Varieties of Selection Bias*, **The American Economic Review**, 80, pp. 313-318.
- J. J. Heckman [2010]: *Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy*, **Journal of Economic Literature**, 48:2, pp. 356-398.
- J. J. Heckman et S. Navarro-Lozano [2004] : *Using Matching, Instrumental Variables and Control Functions to Estimate Economic Choice Models*, **Review of Economics and Statistics**, vol. 86(1), pp. 30-57.
- J. J. Heckman et S. Urzua [2010] : *Comparing IV with Structural Models: What Simple IV can and cannot Identify*, **Journal of Econometric**, 156, pp. 27-37.
- G. W. Imbens [2010] : *Better LATE Than nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)*, **Journal of Economic Literature**, 48:1, pp. 399-423.
- G. W. Imbens et J. D. Angrist [1994] : *Identification and Estimation of Local Average Treatment Effects*, **Econometrica**, 62:2, pp. 467-475.
- G. W. Imbens et J. M. Wooldridge [2009] : *Recent Developments in the Econometrics of Program Evaluation*, **Journal of Economic Literature**, 47:1, pp. 5-86.
- M. P. Keane [2010] : *A Structural Perspective on the Experimentalist School*, **Journal of Economic Perspectives**, vol. 24(2), pp. 47-58.

- M. P. Keane [2010] : *Structural vs Atheoretic Approaches to Econometrics*, **Journal of Econometrics**, vol. 156, pp. 3-20.
- M. P. Keane et K. I. Wolpin [1997] : *The Career Decisions of Young Men*, **Journal of Political Economy**, vol. 105(3), pp. 473-522.
- R.J. Lalonde [1986] : *Evaluating the Econometric Evaluation of Training Programs with Experimental Data*, **The American Economic Review**, 76, pp. 604-620.
- F. Legendre [2013] : *Une introduction à la micro-économétrie de l'évaluation*, **Revue française d'économie**, 2013-1, vol. XXVIII, pp. 9-41.
- J. J. Mc Call [1970] : *Economics of Information and Job Search*, **Quarterly Journal of Economics**, 84, pp. 113-126.
- D. McFadden [2001] : *Economic Choices*, **The American Economic Review**, 91(3), pp. 351-378.
- J. Neyman [1923] : *Statistical Problems in Agricultural Experiments*, **Journal of the Royal Statistical Society II** (supplément 2): 107-80.
- P. R. Rosenbaum et D. B. Rubin [1983] : *The Central Role of the Propensity Score in Observational Studies for Causal Effects*, **Biometrika**, 70(1), pp. 41-55.
- S. Roux et P. Zamora [2011] : *Short and Medium-Run Local Effects of Speed Enforcement Cameras on Accidents*, CREST working papers, 2011-01.
- S. Roux et P. Zamora [2013] : *L'impact local des radars fixes sur les accidents de la route : un effet important après l'installation, mais plus réduit à long terme*, **Economie et statistique**, n°460-461.
- D. Roy [1951] : *Some Thoughts on the Distribution of Earnings*, **Oxford Economic Papers**, 3(2), pp. 135-146.
- D. B. Rubin [1974] : *Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies*, **Journal of Educational Psychology**, 66(5), pp. 688-701.
- P. E. Todd et K. I. Wolpin [2006] : *Assessing the Impact of a School Subsidy Program in Mexico : Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility*, **The American Economic Review**, 96(5), pp. 1384-1417.
- E. J. Vitlacyl [2002] : *Independence, Monotonicity, and Latent Index Models: An Equivalence Result*, **Econometrica**, 70(1), pp. 331-341.

Documents de Travail

550. M. Bussière, G. Cheng, M. Chinn, N. Lisack, “For a Few Dollars More: Reserves and Growth in Times of Crises” April 2015
551. N. Berman, V. Rebeyrol and V. Vicard, “Demand learning and firm dynamics: evidence from exporters” May 2015
552. L. Arrondel, P. Lamarche et Frédérique Savignac, “Wealth Effects on Consumption across the Wealth Distribution: Empirical Evidence” May 2015
553. C. Jude and M. I. Pop Silaghi, “Employment effects of foreign direct investment. New Evidence from Central and Eastern European Countries” May 2015
554. F. Koulischer, “Asymmetric shocks in a currency union: The role of central bank collateral policy” May 2015
555. V. Vicard, “Profit shifting through transfer pricing: evidence from French firm level trade data” May 2015
556. O. Loisel, “The Implementation of Stabilization Policy” June 2015
557. P. Aghion, U. Akcigit, A. Bergeaud, R. Blundell and D. Hémou, “Innovation and Top Income Inequality” June 2015
558. A. Monfort, F. Pegoraro, J.-P. Renne and G. Roussellet, “Staying at Zero with Affine Processes: An Application to Term Structure Modelling” June 2015
559. C. Jude and G. Leveuge, “Growth effect of FDI in developing economies: The role of institutional quality” June 2015
560. M. Bas, T. Mayer and M. Thoenig, “From Micro to Macro: Demand, Supply, and Heterogeneity in the Trade Elasticity” June 2015
561. J. Acalin, B. Cabrillac, G. Dufrénot, L. Jacolin and S. Diop, “Financial integration and growth correlation in Sub-Saharan Africa” June 2015
562. L. Gauvin and C. Rebillard, “Towards Recoupling? Assessing the Global Impact of a Chinese Hard Landing through Trade and Commodity Price Channels” July 2015
563. L. Chauvet and H. Ehrhart, “Aid and Growth. Evidence from Firm-Level Data” July 2015
564. M. Bürker and G. A. Minerva, “Civic capital and the Vertical Integration of Service Provision: Evidence from Italy” July 2015
565. S. Roux, “Approches structurelles et non structurelles en micro-économétrie de l'évaluation des politiques publiques” August 2015

Pour accéder à la liste complète des Documents de Travail publiés par la Banque de France veuillez consulter le site : www.banque-france.fr

For a complete list of Working Papers published by the Banque de France, please visit the website: www.banque-france.fr

Pour tous commentaires ou demandes sur les Documents de Travail, contacter la bibliothèque de la Direction Générale des Études et des Relations Internationales à l'adresse suivante :

For any comment or enquiries on the Working Papers, contact the library of the Directorate General Economics and International Relations at the following address :

BANQUE DE FRANCE
49- 1404 Labolog
75049 Paris Cedex 01
tél : 0033 (0)1 42 97 77 24 ou 01 42 92 63 40 ou 48 90 ou 69 81
email : 1404-ut@banque-france.fr