

Direction des Études et Synthèses Économiques

G 2010 / 08

**Méthodes économétriques
pour l'évaluation de politiques publiques**

Pauline GIVORD

Document de travail



Institut National de la Statistique et des Études Économiques

INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES

*Série des documents de travail
de la Direction des Études et Synthèses Économiques*

G 2010 / 08

Méthodes économétriques pour l'évaluation de politiques publiques

Pauline GIVORD*

JUIN 2010

Ce document reprend dans ses grandes lignes un cours assuré à l'Ensaë avec Bruno CRÉPON depuis 2006, il s'est enrichi des questions souvent stimulantes des élèves qui ont assisté à ce cours, ainsi que des discussions avec les membres du Département des Études Économiques d'Ensemble. Je remercie tout particulièrement Magali BEFFY, Nila CECI-RENAUD, Laurent DAVEZIES, Xavier D'HAULTFOEUILLE et Lionel WILNER qui m'ont permis par leur lecture attentive et leurs commentaires constructifs d'améliorer significativement les premières versions de ce document.

* Département des Études Économiques d'Ensemble - Division « Marchés et Stratégies d'Entreprise »
Timbre G230 - 15, bd Gabriel Péri - BP 100 - 92244 MALAKOFF CEDEX

Méthodes économétriques pour l'évaluation de politiques publiques

Résumé

Ce document présente les principales méthodes économétriques qui peuvent être utilisées pour l'évaluation *ex post* de l'impact d'une politique publique. Il met l'accent sur les problèmes de sélection qui se posent pour une évaluation (en particulier distinguer ce qui relève de l'effet de la politique des spécificités éventuelles de ces bénéficiaires). Il trace les différentes étapes d'une feuille de route d'un tel exercice (données nécessaires, hypothèse d'identification, interprétation des résultats), illustrées par les exemples issus d'articles récents de la littérature économique. Il commence par rappeler le cadre classique dit « de Rubin » pour définir une inférence causale, puis décrit en détail les quatre principales méthodes empiriques : différences de différences, variables instrumentales, régression sur discontinuités et sélection sur observables.

Mots-clés : Évaluation des politiques publiques, régressions sur discontinuités, matching, variables instrumentales, différences-de-différences

Econometric Methods for Public Policies Evaluation

Abstract

This document presents an overview of econometric tools recently developed for empirical *ex post* policy evaluation. The emphasis is laid on selection problems (how to disentangle the real effect of a policy from the personal features of the affected people), and on the practical issues of a policy evaluation (access to relevant data, identification hypotheses, interpretation of the results) illustrated by examples provided by recent empirical articles. It presents first the Rubin framework for causal inference, and details then the four main empirical designs for evaluation: selection on observables, differences in differences, instrumental variables and regression on discontinuity.

Keywords: Public Policies Evaluation, Regression Discontinuity Design, Matching, Instrumental Variables, Differences-in-Differences

Classification JEL : C1

1 Introduction

La demande d'évaluation est de plus en plus présente dans le débat public. Quel est l'impact d'un programme de suivi des chômeurs? Du classement de certaines zones urbaines en zones franches? D'une réduction de la taille des classes ou de la suppression de la carte scolaire? Autant de questions sur lesquelles le statisticien pourrait être sollicité pour apporter des éléments de réponse. À ces questions, des indicateurs de suivi (nombres de bénéficiaires, coût budgétaire de la mesure) apportent des éléments de cadrages indispensables mais pas de réponse sur l'*efficacité* d'une politique. Le coût social d'une mesure ne se limite pas à son coût budgétaire initial. Une politique de formation des chômeurs, si elle est efficace pour faciliter le retour à l'emploi des chômeurs qui en bénéficient, peut permettre de réduire le nombre de chômeurs indemnisés. Bien que coûteuse en soi, elle peut donc être en partie autofinancée. Une *étude d'impact* est essentielle pour estimer la rentabilité économique de la mesure.

Cette analyse est cependant complexe. Il s'agit de déterminer si les éventuelles améliorations constatées peuvent être attribuées à la mise en place de cette mesure. Une mesure est en général ciblée sur des populations particulières, dont les caractéristiques socio-économiques auraient de tout façon été différentes de celle du reste de la population, que la mesure ait été mise en place ou non. Par exemple, les classes à effectif réduit regroupent généralement les enfants les plus en difficulté. Il n'est pas alors surprenant que les résultats moyens soient plus faibles dans ces classes. En conclure qu'augmenter la taille des classes est efficace pour lutter contre l'échec scolaire serait cependant abusif... Ces effets de sélection, ou de population, rendent l'évaluation difficile.

De nombreuses méthodes économétriques ont été développées pour traiter ces questions sur les deux dernières décennies. Elles se distinguent des usages classiques de l'économétrie appliquée par au moins deux points. D'une part, une grande attention aux hypothèses nécessaires pour identifier un effet causal, qui pousse à plus de rigueur dans l'analyse empirique. D'autre part, une attention particulière est portée sur le fait que les effets puissent être différents selon les personnes. Cette littérature est encore en développement, et il n'existe pas toujours de consensus sur certains points.

Ce document présente les principes généraux de ces méthodes, qui constituent autant d'outils dont dispose l'économiste pour une évaluation *ex post* d'une réforme. Il tente en particulier de présenter simplement leur mise en œuvre pratique, et fournit des illustrations à l'aide d'exemples issus de la littérature empirique récente. Notons que les résultats sur les propriétés statistiques des différents estimateurs ne seront pas présentés en détails ni démontrés, mais les références dont ils sont issus sont systématiquement détaillées. Pour une présentation plus formalisée, on peut en particulier se référer à Imbens et Wooldridge (2008).

Enfin, une remarque peut être utile en préambule. Il existe deux approches assez différentes pour traiter de ces questions : la première, qualifiée de "structuraliste", consiste

à expliciter un modèle complet du comportement des agents économiques. Ce type d'approche est particulièrement utile *ex ante*, pour prévoir le résultat d'une politique qui n'a jamais été appliquée auparavant. Une autre approche, qui est plus "empirique", consiste à tester directement l'effet d'une mesure sur ce qui intéresse l'économètre, sans forcément expliciter tous les mécanismes à l'œuvre. C'est clairement à ce deuxième courant que se rattachent les méthodes étudiées ici.

Ces deux approches sont plus complémentaires que concurrentes. Pour "calibrer" un modèle théorique, c'est-à-dire estimer les paramètres clés du modèle (comme par exemple les élasticités-prix des agents), il est nécessaire de recourir aux estimations empiriques. Inversement, pour être généralisés, les résultats d'une étude empirique demandent souvent de passer par une explicitation des comportements économiques des agents concernés.

1.1 La formalisation du problème d'évaluation : le cadre de Rubin

Dans toute la suite, nous allons nous placer dans le cadre dit "de Rubin", du nom du statisticien qui l'a popularisé au cours des années 70. Le vocabulaire de cette approche emprunte beaucoup au champ de l'expérimentation médicale dont il est issu. Ainsi, on parle souvent de "traitement" pour désigner la variable dont on cherche à mesurer l'effet.

En pratique, on s'intéresse donc à l'évaluation d'une mesure T . Dans la version la plus simple, on distingue les personnes qui bénéficient de cette mesure ($T = 1$) de celles qui n'en bénéficient pas ($T = 0$). On s'intéresse alors à l'effet de cette mesure sur une grandeur d'intérêt, qu'on appellera souvent dans la suite le *revenu*. L'originalité ici est de considérer que chacun a "virtuellement" deux revenus potentiels, selon qu'il bénéficie ou non de la mesure. On note donc Y_0 le revenu associé à $T = 0$ (pas de traitement) et Y_1 le revenu associé à $T = 1$ (traitement).

Remarque : La définition de la variable d'intérêt n'est jamais évidente et demande de s'interroger sur les objectifs que l'on assigne à la politique qu'il s'agit d'évaluer : par exemple, s'intéresse-t-on à l'effet d'un stage de formation sur le retour à l'emploi ? Le salaire ? La récurrence d'épisodes de chômage ? Si l'on diminue la taille des classes cherche-t-on à augmenter les résultats moyens ? Ceux des plus faibles ? Ou plutôt à réduire les inégalités ? Un bilan complet d'une mesure ne peut être fourni par un chiffre unique, et l'analyse de ces multiples facettes est importante.

Le problème fondamental de l'inférence causale Pour une personne, on peut alors définir l'effet propre (ou causal) de la mesure Δ_i , à partir de ces deux revenus potentiels : $\Delta_i = Y_{i1} - Y_{i0}$.

Le problème est que l'on n'observe pas simultanément Y_{i0} et Y_{i1} : pour un bénéficiaire de la mesure, on observe Y_{i1} mais pas Y_{i0} ; pour un non-bénéficiaire on observe Y_{i0} mais pas Y_{i1} . Par exemple, il est possible de savoir si un chômeur a finalement obtenu un emploi après un stage de formation. Il sera difficile de déterminer si cette reprise d'emploi est due

au stage en lui-même : pour cela, il faudrait connaître les chances de retour à l'emploi de ce chômeur, *s'il n'avait pas suivi ce stage*.

Tout l'esprit des méthodes d'évaluation sera justement d'estimer cette situation *contre-factuelle*, i.e. ce qui aurait été observé si la mesure n'avait pas été mise en place. Pour cela, il suffit de trouver des personnes totalement comparables à celles qui ont bénéficié de la mesure, mais qui n'ont pas été affectées par elle. En pratique, il s'agit de trouver des personnes les plus comparables possible. On parle de *groupe de contrôle*.

1.1.1 Définition des paramètres d'intérêt : les effets moyens du traitement

L'estimateur le plus naturel est l'effet moyen de la mesure sur ceux qui en ont effectivement bénéficié, soit l'effet du traitement sur les traités (*Average Treatment effect on the Treated* ou *ATT* dans la suite). Formellement, il s'écrit :

$$\Delta^{ATT} = E(Y_{i1} - Y_{i0} | T_i = 1)$$

Au-delà de l'effet de la mesure sur les seuls bénéficiaires, on peut se demander ce que serait l'effet de cette mesure si elle était étendue à l'ensemble de la population. Il s'agit alors d'estimer l'effet moyen du traitement sur l'ensemble de la population (*Averaged Treatment Effect*), soit formellement :

$$\Delta^{ATE} = E(Y_{i1} - Y_{i0})$$

Ces effets sont susceptibles de dépendre des caractéristiques des personnes : les jeunes chômeurs peuvent par exemple plus bénéficier d'un programme de formation que des chômeurs plus âgés (ou inversement), les élèves issus de milieu défavorisé être plus sensibles à la taille des classes... Pour ces raisons, il peut être pertinent d'estimer des effets moyens à caractéristiques observables données (par exemple : l'âge, le genre, la nationalité...). En pratique, cela revient à estimer des effets moyens conditionnels à ces caractéristiques observables. Nos deux paramètres s'écriront donc plutôt :

$$\Delta^{ATT}(x) = E(Y_{i1} - Y_{i0} | T_i = 1, X_i = x)$$

et

$$\Delta^{ATE}(x) = E(Y_{i1} - Y_{i0} | X_i = x)$$

1.1.2 Effets de sélection, hétérogénéité de l'effet du traitement

L'effet du traitement sur les traités correspond à la différence entre le revenu moyen avec traitement $E(Y_1 | T = 1)$, qu'il faut comparer avec le revenu qu'ils auraient eu s'ils n'avaient pas été traités $E(Y_0 | T = 1)$. Si le premier est connu, le deuxième ne l'est évidemment pas, et il faudra trouver un estimateur crédible de ce revenu potentiel contrefactuel.

L'estimateur le plus simple, qu'on peut qualifier de "naïf", de l'effet moyen du traitement consiste à comparer les revenus des bénéficiaires avec celui des non-bénéficiaires $E(Y_0 | T = 0)$. Comme souligné plus haut, le problème est qu'en général, les bénéficiaires ne le sont

pas par hasard : le fait qu'ils ont été choisis pour en bénéficier est probablement lié aux revenus qu'ils pouvaient espérer, avec ou sans traitement. On peut supposer par exemple que les chômeurs qui choisissent de ne pas suivre un stage de formation sont justement ceux qui ont des chances de retour à l'emploi (sans ce stage) suffisamment élevées pour se passer d'un tel stage. Dans ce cas l'estimateur naïf sous-estimerait l'impact du stage. On peut imaginer à l'inverse que ceux qui choisissent de le suivre sont plus motivés, et auraient de toute façon plus de chances de trouver un emploi, et dans ce cas l'estimateur naïf sur-estime l'effet du stage¹...

Si on appelle Y le revenu observé, alors l'estimateur naïf $E(Y|T = 1) - E(Y|T = 0)$ est *a priori* biaisé :

$$\begin{aligned} E(Y|T = 1) - E(Y|T = 0) &= E(Y_1|T = 1) - E(Y_0|T = 0) \\ &= E(Y_1|T = 1) - E(Y_0|T = 1) \\ &\quad + E(Y_0|T = 1) - E(Y_0|T = 0) \end{aligned}$$

$E(Y_1|T = 1) - E(Y_0|T = 1)$ est ce qu'on cherche à mesurer, i.e. la variation de revenu qui est due au traitement pour les personnes traitées. Il y a donc un biais, égal à $E(Y_0|T = 1) - E(Y_0|T = 0)$. Ce biais correspond donc à la différence de revenu moyen qui serait enregistrée sans la mesure, entre ceux qui ont choisi d'en bénéficier et les autres. Il n'est pas nul dès lors que les personnes choisissent de bénéficier de la mesure en fonction de leur revenu, car dans ce cas les non-bénéficiaires ne sont pas représentatifs des bénéficiaires. En termes statistiques, le revenu sans traitement Y_0 n'est pas indépendant du fait de suivre le traitement :

$$Y_0 \not\perp T \tag{1}$$

Le problème est encore plus sévère lorsque l'on veut estimer l'effet moyen du traitement sur toute la population (ATE). Dans ce cas, on voudrait connaître la situation contrefactuelle des bénéficiaires s'ils n'avaient pas été affectés par la mesure (comme précédemment), mais également la situation contrefactuelle des non-bénéficiaires, s'ils avaient été affectés par cette mesure. Une décomposition simple montre en effet que :

$$\begin{aligned} \Delta^{ATE} &= P(T = 1)[E(Y_1|T = 1) - E(Y_0|T = 1)] \\ &\quad + (1 - P(T = 1))[E(Y_1|T = 0) - E(Y_0|T = 0)] \end{aligned}$$

L'effet moyen sur la population correspond à la moyenne de l'effet observé de la mesure sur les bénéficiaires et de celui, hypothétique, sur les non-bénéficiaires pondérée par leurs proportions respectives : pour que l'effet de la mesure puisse être mesuré à partir des bénéficiaires, il faut que l'effet moyen sur les bénéficiaires soit le même que celui qu'on aurait enregistré en moyenne si on avait appliqué le traitement aux non-bénéficiaires. Cette décomposition montre également qu'on ne peut estimer, au mieux, qu'un effet moyen, alors que l'effet d'une mesure peut différer selon les individus. Le fait de les estimer sur

¹Une formalisation classique de ce problème de sélection est fournie par le modèle "de Roy" (Heckman et Honore 1990).

une sous-population plutôt qu'une autre (par exemple, sur les bénéficiaires plutôt que les non-bénéficiaires) n'est donc pas neutre, si cette sous-population a des caractéristiques différentes du reste de la population. L'attention portée sur cette hétérogénéité est un des apports de cette littérature. Nous aurons l'occasion d'y revenir par la suite. On trouvera également en annexe une comparaison de ce cadre avec le cadre classique de l'économétrie linéaire.

1.1.3 Une hypothèse importante : pas d'externalités

Une hypothèse importante du modèle de Rubin est qu'il n'existe pas d'externalités. Le traitement de l'individu i n'a d'effet que sur son revenu, mais pas sur ceux des autres personnes. Cette hypothèse est appelée dans la littérature *SUTVA*, pour *Stable Unit Treatment Value Assumption*.

Cette hypothèse est assez naturelle même si elle exclut des cas importants. Par exemple, une campagne de vaccination a un impact sur la propagation de la maladie, et peut donc indirectement protéger des personnes qui n'ont pas été vaccinées. Une politique de soutien scolaire aux plus faibles peut avoir des effets d'entraînement au sein d'une classe. Lorsqu'une mesure d'aide à l'emploi des chômeurs est mise en œuvre, il est difficile de croire que le marché de l'emploi est indéfiniment extensible et pourra absorber à court terme l'ensemble des personnes aidées, même si le programme est individuellement efficace. Selon l'ampleur du dispositif, il sera plus ou moins plausible d'ignorer ces effets de bouclages.

Une solution peut être d'"internaliser les externalités", en raisonnant à un niveau plus agrégé quand cela a un sens. Pour étudier l'impact de traitements anti-parasitaires fournis à des enfants scolarisés dans certaines écoles du Kenya (désignées au hasard au cours d'une expérience contrôlée), Miguel et Kremer (2004) comparent l'état de santé moyen de l'ensemble des enfants scolarisés dans les écoles ayant bénéficié du programme (même si tous les enfants n'en ont pas effectivement bénéficié, par exemple parce qu'ils étaient absents le jour du traitement) par rapport à l'état de santé moyen des enfants scolarisés dans les écoles n'en ayant pas bénéficié².

1.1.4 Méthodes empiriques d'évaluation

Les différentes méthodes empiriques qui sont présentées dans la suite de ce document apportent, chacune à leur manière, une réponse aux problèmes d'autosélection et d'hétérogénéité auxquels sont confrontées les évaluations. L'enjeu principal réside dans la possibilité de trouver des données, des situations, permettant de neutraliser l'effet de sélection.

Les expériences contrôlées La manière la plus directe consiste à se placer dans le cadre d'expériences aléatoires contrôlées. Dans de telles expérimentations *ex ante* d'une politique,

²À l'inverse, certains économistes tentent justement de mettre en évidence et de mesurer ces "effets de pairs" (par exemple l'impact des autres élèves sur la réussite scolaire individuelle...). Cela pose des problèmes d'identification difficiles, voir Manski (2000). L'analyse de cette littérature sort du cadre de ce document.

les personnes qui bénéficieront du traitement (par exemple un accompagnement renforcé pour l'aide à la recherche d'emploi pour des chômeurs) et celles qui n'en bénéficieront pas (le groupe de contrôle) déterminées aléatoirement. Si la répartition aléatoire est bien vérifiée, le choix d'être traité ou non est indépendant du résultat potentiel escompté, et l'estimateur :

$$\begin{aligned}\Delta(X) &= E(Y|T = 1, X) - E(Y|T = 0, X) \\ &= E(Y_1|T = 1, X) - E(Y_0|T = 0, X) \\ &= E(Y_1|T = 1, X) - E(Y_0|T = 1, X)\end{aligned}$$

est un estimateur sans biais de l'effet du traitement sur les traités.

Les expériences aléatoires constituent le paradigme des méthodes économétriques d'évaluation. Elles sont présentées en détail dans Duflo, Glennerster, et Kremer (2006). Cependant, même si leur utilisation commence à se développer en France, elles restent encore rares. De fait, elles nécessitent une intervention très en amont de la mise en place de la politique qu'il s'agit d'évaluer. Souvent, le statisticien est plutôt conduit à estimer *ex post* des politiques déjà mises en place. Plusieurs méthodes ont été développées, que nous présentons ici.

Un premier ensemble tente d'utiliser des variations exogènes de l'environnement économique, des réformes de législation par exemple qui créent une situation presque expérimentale. On parle d'"expériences naturelles", ou de "quasi-expériences". Les trois méthodes présentées ici sont les différences de différences, les variables instrumentales, les régressions sur discontinuités. Elles se distinguent par les hypothèses sous lesquelles on estime correctement l'impact causal de la mesure qu'on souhaite évaluer. Le fait d'appliquer l'une ou l'autre sera souvent tributaire des données.

L'estimateur de *différences de différences* consiste à mesurer la différence de revenu entre population traitée et non traitée avant et après la mesure et à effectuer la différence de ces deux évolutions. Sous l'hypothèse que la variable d'intérêt potentielle sans traitement des bénéficiaires aurait évolué de la même manière que dans le groupe de contrôle, comparer les évolutions des deux groupes permet théoriquement d'éliminer les effets de composition liés à la sélection et de ne mesurer que l'impact propre de la mesure.

L'hypothèse nécessaire pour estimer par *variables instrumentales* est différente. La définition de telles variables est bien connue : il s'agit de variables qui expliquent le fait d'être traité, mais qui ne sont pas corrélées aux composantes inobservées du revenu, et qui permettent donc de s'affranchir du problème de sélection.

Enfin, les *régressions sur discontinuités* utilisent le fait que de nombreux dispositifs comportent des seuils. Par exemple, la carte scolaire contraint les parents sur le choix de l'école que leurs enfants peuvent fréquenter ; certaines allocations sont réservées aux personnes dont les revenus se situent en dessous d'un certain plafond... L'idée est qu'autour de la discontinuité, des personnes très proches peuvent avoir des chances très différentes de bénéficier d'une mesure ou d'une aide. L'hypothèse ici consiste à dire que les personnes

autour de ces seuils ont des caractéristiques inobservables identiques et donc que pour ces personnes dont les caractéristiques sont très proches, le fait d'être sélectionné peut être considéré au premier ordre comme indépendant des caractéristiques des individus. On pourra alors par exemple comparer les parents habitant de chaque côté de la rue servant de frontière à la carte scolaire, ou les individus juste en dessous ou au-dessus du plafond de ressources.

Le choix entre ces différentes méthodes est en général conditionné par les données dont on dispose. Souvent, aucune de ces méthodes n'est applicable. Pour réduire le biais de population, le minimum que l'économètre doit faire est de contrôler des différences observables entre les bénéficiaires et les non-bénéficiaires. Les estimateurs par appariement (ou *matching*) consistent ainsi à appairer chaque personne traitée avec un "jumeau" non traité possédant les mêmes caractéristiques. Ils permettent d'identifier l'impact réel du traitement sous l'hypothèse, forte, que l'ensemble d'observables est suffisamment riche pour que conditionnellement à ces observables, le fait de choisir de suivre ou non le traitement est indépendant des caractéristiques inobservables.

De l'estimation à l'interprétation L'ensemble des méthodes passées en revue jusqu'ici tente donc d'estimer l'effet d'une politique, d'une réforme sur une population. Une étape importante sera de se demander dans quelle mesure les données dont on dispose, les estimations qu'on a pu mettre en œuvre nous renseignent effectivement sur cette question.

En premier lieu, tous ces estimateurs reposent sur des hypothèses précises, et leur validité est donc tributaire de ces hypothèses. Il convient donc de s'interroger sur leur pertinence, pour juger de la *validité interne* de notre estimateur. Malheureusement, comme on le verra, ces hypothèses ne sont pratiquement jamais testables, même si quelques tests indirects de leur plausibilité sont parfois possibles.

Par ailleurs, il faut s'interroger si des résultats obtenus sur des données particulières, une population ou une période spécifiques sont généralisables. Des effets de bouclage macroéconomique peuvent également intervenir et compromettre la *validité externe* des estimateurs. La discussion de ces effets ne peut se faire qu'au cas par cas.

2 Les estimateurs de Différences de Différences

2.1 Principe

Le principe d'une évaluation est d'estimer un contrefactuel, la situation qui aurait prévalu en l'absence du programme que l'on souhaite évaluer. Comme on l'a vu, comparer les revenus des bénéficiaires d'une mesure avec les non-bénéficiaires risque de donner des résultats biaisés par la présence d'effets de sélection. Comparer les revenus à ce qu'ils étaient avant la mise en place de la réforme (estimateur avant/après) n'est pas plus satisfaisant. De nombreux facteurs sont susceptibles d'expliquer des variations du revenu dans le temps, indépendamment des effets de la réforme : conjoncture économique plus ou moins favorable, profil de revenu croissant avec l'expérience...

L'idée, très simple, des estimateurs de différences de différences est de mélanger ces deux approches. Dans le cas le plus simple, on dispose de deux groupes qu'on observe avant et après l'introduction de la mesure qu'on souhaite évaluer ; cette réforme ne concerne qu'un des groupes. L'estimateur consiste alors à comparer l'évolution des revenus des bénéficiaires avant et après le traitement à celle des revenus des personnes du groupe de contrôle sur la même période.

Son nom résume bien la méthode : on fait la différence entre les deux groupes des évolutions respectives (différences temporelles) de leurs revenus sur la période.

- l'une des différences doit permettre d'éliminer les différences systématiques entre le groupe des bénéficiaires et les autres ;
- l'autre différence doit permettre d'éliminer l'évolution temporelle, supposée identique pour les deux groupes en l'absence de la mesure.

Cet estimateur permet d'identifier l'effet propre, causal, de la mesure qu'on veut évaluer, sous l'hypothèse que les revenus des deux groupes auraient effectivement évolué de manière identique en l'absence de la mesure.

2.2 Hypothèse identifiante

Dans le cas caricatural présenté sur la figure 1, les évolutions des revenus sans traitement des deux groupes sont parfaitement identiques. Du fait de la mesure, le revenu du groupe traité ($T = 1$) a cependant évolué plus vite que celui du groupe de contrôle ($T = 0$). L'estimateur naïf qui compare les revenus des traités et des non-traités à l'issue de la réforme (donné par la distance AC) est biaisé par des effets de sélection, car les traités sont intrinsèquement différents des non-traités (différence *ex ante* de revenu AB). L'effet réel du traitement est donné par BC.

Ceci peut s'écrire de manière plus formelle. Plaçons-nous dans un cadre plus général, dans lequel on observe plusieurs groupes d'individus, dont seulement certains bénéficient du dispositif qu'on souhaite étudier. Par exemple, ces groupes représentent l'ensemble des entreprises situées dans les zones de la politique de la ville. On suppose qu'on observe la variable d'intérêt (toujours appelée revenu) sur T périodes.

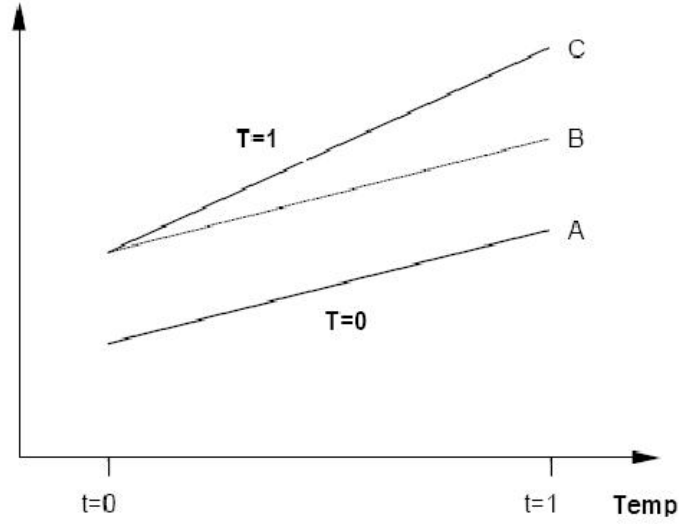


FIG. 1: Présentation schématique de l'évolution des revenus des deux groupes

On suppose que les revenus potentiels se décomposent en :

$$\begin{cases} y_{0igt} = e_t + e_g + u_{it} \\ y_{1gt} = y_{0gt} + \delta \end{cases}$$

i.e. une composante e_g liée au groupe g qui ne varie pas au cours du temps, une composante temporelle e_t qui est la même selon les groupes, et δ mesure l'effet moyen du traitement. Dans le cas des données de panel, la composante e_g correspond à l'effet fixe c_i . Cette hypothèse de séparabilité repose donc sur une double condition : la stabilité de la composition des groupes au cours du temps ; des effets conjoncturels identiques dans tous les groupes.

Rappelons que dans ce cas le revenu observé s'écrit :

$$y_{it} = e_t + e_g + \delta T_{it} + u_{it}$$

Une condition suffisante pour identifier sans biais l'effet du traitement δ est que l'indicatrice de traitement T_{it} soit strictement exogène, i.e. qu'elle ne soit corrélée avec aucun des effets individuels u_{it} . Le fait de bénéficier du traitement ou non ne doit pas dépendre des chocs sur les revenus contrefactuels passés et/ou futurs. Cette hypothèse n'est pas vérifiée lorsque le choix d'en bénéficier ou pas se fait en fonction des revenus précédents (voir infra).

2.3 Mise en œuvre : groupe de contrôle et estimation

2.3.1 Déterminer un groupe de contrôle

Expériences naturelles La question consiste alors à déterminer un groupe de contrôle plausible, i.e. susceptible d’avoir effectivement connu des évolutions temporelles comparables à celles du groupe traité. Il faut insister sur le fait que pour évaluer un dispositif avec libre entrée, le groupe des “ayant droits” qui ont pourtant choisi de ne pas participer ne constitue pas *a priori* un bon groupe de contrôle. Il est possible que les problèmes classiques de sélection réapparaissent : le fait qu’ils aient renoncé au dispositif peut être lié aux bénéfices qu’ils anticipaient pouvoir en retirer... Si on pense qu’ils ont fait ce choix en connaissance de cause, cela signifie que l’évolution de leur revenus potentiels anticipés se distingue de l’évolution des revenus potentiels des bénéficiaires du dispositif. La double différenciation ne permettra pas *a priori* de régler ce problème³.

En revanche, des réformes ciblées sur certaines sous-populations si le ciblage n’est pas déterminé directement par le bénéfice que ces personnes peuvent en tirer ou des modifications de l’environnement économique offrent des situations de “quasi-expérience”. En voici quelques exemples.

- Card (1990) utilise l’afflux massif et rapide d’immigrants cubains à Miami suite à l’exode de Mariel⁴ pour évaluer l’impact de l’immigration sur le marché du travail, en comparant avec des villes américaines “proches”.
- Card et Krueger (1994) évaluent l’impact sur l’emploi non qualifié d’une hausse du salaire minimum intervenue dans l’état du New Jersey, en comparant l’emploi dans les *fast-foods* de cet État et de l’État voisin de Pennsylvanie, juste avant et six mois après l’augmentation.
- Durbin, Meyer, et Viscusi (1995) s’intéressent aux modifications des seuils pour le versement des indemnités suite à des accidents du travail pour évaluer l’impact de ces versements sur la durée des arrêts maladie. Sur la figure 2, le groupe de traitement est constitué des personnes disposant de salaires élevés (*high earnings group*), dont le montant de remboursement a augmenté brutalement suite à la réforme (relèvement du plafond de remboursement) ; en revanche, pour les salariés du bas de la distribution (*low earnings group*), le niveau de remboursement est resté identique avant et après la réforme, et ces derniers constituent donc un groupe de contrôle.

Ces exemples constituent des expériences “naturelles”, encore appelées des “quasi-expériences”. On utilise un choc sur l’environnement économique exogène pour identifier un effet causal (l’impact de l’afflux d’une main-d’œuvre étrangère dans le cas de Card (1990), l’effet désincitatif des taux de remplacement sur le retour à l’emploi après un accident

³Plus formellement, on peut supposer que les éligibles qui ne recourent pas au dispositif diffèrent de ceux qui y recourent, en particulier dans le gain qu’ils peuvent espérer du traitement $Y_1 - Y_0$: l’hypothèse $E(y_{1gt}) = E(y_{0gt}) + \delta$ n’est pas forcément vérifiée.

⁴Le 20 avril 1980, suite à des mouvements de protestation à Cuba, Fidel Castro a autorisé ceux qui le souhaitent à quitter l’île. Plus de 100 000 Cubains ont alors embarqué de la baie de Mariel, face aux côtes de Floride.

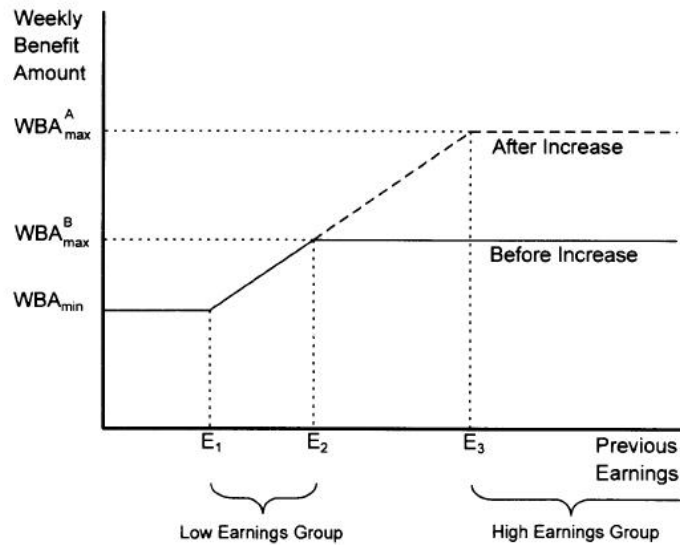


FIG. 2: Taux de remboursement des allocations en cas d'accident du travail selon les revenus, Durbin, Meyer et Viscusi (1995)

pour maladie pour Durbin, Meyer, et Viscusi (1995)). On s'approche ainsi des conditions des expériences contrôlées, au sens où on peut raisonnablement supposer que le fait que certaines populations soient affectées ou non n'est pas lié à leurs caractéristiques propres (ainsi, ce n'est pas parce que le marché de l'emploi était particulièrement dynamique qu'il y a eu une vague d'immigration à Miami).

Ces quasi-expériences proviennent souvent de dispositifs ciblés sur une tranche d'âge (on peut alors comparer avec des âges différents), des zones géographiques, des tranches de revenus... Insistons encore sur le fait que le ciblage ne doit pas être directement lié aux caractéristiques des populations visées.

Construire un contrefactuel La plupart des études qui appliquent une méthode de différences de différences définissent le groupe de contrôle de manière *ad hoc*. En général, on compare la situation avec un groupe considéré comme semblable, mais il existe toujours une part d'arbitraire dans le choix de ce semblable. Une autre solution peut être de constituer un groupe de contrôle en appariant les bénéficiaires avec des non-bénéficiaires qui présentent des caractéristiques observables identiques. Nous verrons cela en détail dans la section sur le *matching*. Une autre méthode est proposée par Abadie, Diamond, et Hainmueller (2007), dans le cas où l'on dispose de données agrégées, sur une région affectée par un choc ou une politique particulière par exemple. Elle consiste à construire un contrefactuel à partir de l'ensemble des non traités.

Pour fixer les idées, supposons que l'on s'intéresse à l'effet d'une réforme intervenue dans une région donnée. Il existe J autres régions, indicées par $2, \dots, J + 1$, qui n'ont pas

été concernées par la réforme et sont donc des groupes de contrôle potentiels. L’approche de différences de différences classique consisterait à “choisir” une de ces régions comme groupe de contrôle. L’approche d’Abadie, Diamond, et Hainmueller (2007) rend ce choix plus transparent, en utilisant comme contrefactuel un indicateur synthétique : une combinaison linéaire de tous les revenus des groupes de contrôle potentiels pondérés par leur distance à l’entité traitée. Cette distance est calculée selon un certain nombre de caractéristiques (revenu avant traitement, ainsi que des déterminants du revenu après traitement).

Soient donc X_i les caractéristiques observables des régions non affectées par la réforme, qui sont des déterminants de l’évolution des revenus, et y_{it} le revenu du groupe i à la date t , t valant 1 (avant la réforme) et 2 (après la réforme). L’idée est de construire un contrefactuel comme combinaison linéaire de tous les états non traités, en choisissant les poids w_j tels qu’ils minimisent la distance entre (y_{11}, X_1) et $\sum_{j=2}^{J+1} w_j (y_{j1}, X_j)$ (avec une métrique adaptée).

L’effet de la mesure peut alors être mesuré par : $y_{12} - \sum_{j=2}^{J+1} w_j y_{j2}$. Si la pondération permet de rendre négligeable les différences de consommations avant la mesure, cet estimateur correspond à celui de différences de différences.

Abadie, Diamond, et Hainmueller (2007) appliquent cette méthode pour évaluer l’impact de la “proposition 99” en Californie (première loi anti-tabac en 1988). La difficulté de cette évaluation est que la consommation de tabac en Californie différait déjà nettement de celle des autres États américains avant même la proposition 99. Comme décrit plus haut, le contrefactuel est une moyenne pondérée des consommations par tête dans ces autres États. Les pondérations sont inversement proportionnelles aux écarts constatés entre les consommations antérieures de tabac entre l’État et la Californie, ainsi qu’entre des prédicteurs de la consommation (revenu moyen, proportion de jeunes, prix du paquet avant la réforme, consommation de bière par tête).

La figure 3 présente le contrefactuel obtenu. Il reproduit très bien la consommation passée. Le décrochage de la consommation en Californie, après le passage de la proposition 99, fournit donc une bonne mesure de l’impact de cette mesure. Il reste à déterminer plus précisément si ces différences sont statistiquement différentes de zéro. Nous verrons cela dans la section 2.3.6.

2.3.2 Estimation : quelles données, quelles méthodes ?

Considérons le cas le plus simple avec deux groupes et deux périodes. L’estimateur de différences de différences correspond à :

$$\Delta = E(y^{t_1} - y^{t_0} | T = 1) - E(y^{t_1} - y^{t_0} | T = 0) \quad (2)$$

où y^{t_1} représente le revenu observé après l’introduction de la mesure, y^{t_0} celui avant son introduction.

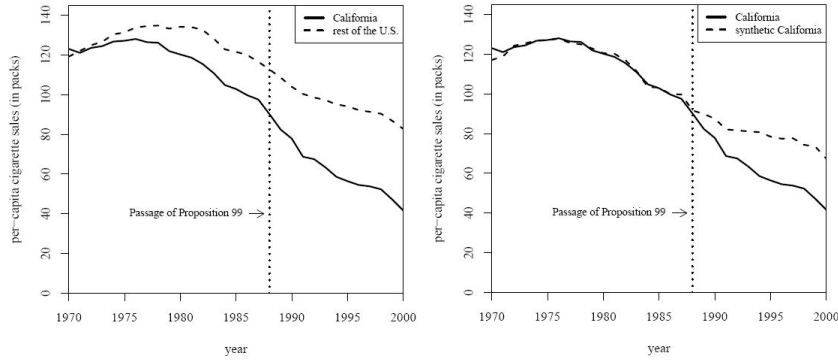


FIG. 3: Abadie, Diamond et Hainmuller (2007) : Évolution de la consommation de tabac en Californie et en moyenne dans les autres états américains (graphique de gauche), construction d'un état contrefactuel (graphique de droite)

En pratique, cet estimateur correspond à l'estimateur des moindres carrés ordinaires dans la régression :

$$y_{it} = \beta_0 + \beta_1 1_{t_1} + \beta_2 1_{T_i=1} + \delta 1_{t_1} 1_{T_i=1} + u_{it}$$

où l'indice t_1 à la date après la mise en place de la réforme.

Cette estimation présente l'avantage de fournir directement la précision de ces estimations. Néanmoins, comme nous le verrons dans la section 2.3.6, l'inférence peut parfois être plus compliquée que la simple application de la précision obtenue dans le cadre d'un modèle linéaire homoscédastique.

Dans le cas général de plusieurs groupes et plusieurs périodes, l'estimation correspond de même à :

$$y_i = e_t + e_g + \delta T_{gt} + u_{it} \quad (3)$$

où T_{gt} est une indicatrice que le groupe g a été traité à la date t ; les e_t sont des effets fixes temporels, qui captent les évolutions de la conjoncture affectant tous les groupes, et les e_g sont des effets fixes groupe rendant compte des différences permanentes entre les groupes.

En pratique, cela signifie qu'on met une indicatrice pour tous les groupes, une indicatrice pour chacune des périodes, et une indicatrice croisée pour indiquer que le groupe g a été traité à t .

Si on dispose de plus d'informations sur les groupes concernés, il peut être judicieux (et surtout plus efficace) de contrôler l'effet de ces variables observables, qui peuvent varier entre les groupes et au cours du temps. On estimera donc :

$$y_{gt} = e_t + e_g + \delta T_{gt} + X_{gt}\beta + u_{gt} \quad (4)$$

Il n'est pas nécessaire de disposer de données de panel (i.e. des données individuelles sur les mêmes personnes aux différentes dates). Des données agrégées sur les différents groupes

aux différentes périodes suffisent. Cependant, pour que l'hypothèse identifiante ait des chances d'être vérifiée, il faut que la composition des groupes au sein des échantillons dont on dispose soit stable au cours du temps.

Notons également que cette spécification n'exclut pas le cas d'une mesure qui serait appliquée à des dates différentes selon les groupes.

Utilisation de données de panel Lorsque l'on dispose de données de panel, il est dommage de ne pas utiliser l'information supplémentaire qu'elles apportent. Typiquement, on peut mieux contrôler l'hétérogénéité individuelle en utilisant des données répétées pour chaque individu. En pratique, cela revient à remplacer les effets moyens groupes par des effets fixes individuels :

$$y_{it} = e_t + c_i + \delta T_{it} + u_{it} \quad (5)$$

où T_{it} est une indicatrice valant 1 lorsque l'individu i bénéficie de la mesure à la date t .

Dans ce cas, on peut estimer δ par des procédures standard d'économétrie des panels, comme des estimations en différences ou des estimateurs à effets fixes (*within*). Rappelons que la première revient à différencier les variables selon la dimension temporelle tandis que la seconde utilise les écarts à la moyenne temporelle, ces deux opérations permettant d'éliminer les effets fixes c_i .

2.3.3 Validité de l'hypothèse identifiante et tests de robustesse

L'hypothèse identifiante est impossible à tester, puisque cela signifierait qu'on pourrait observer l'évolution contrefactuelle du groupe traité en l'absence de traitement... Il peut être cependant utile de faire plusieurs tests, pour juger de sa vraisemblance.

Comparaison des évolutions passées du revenu Le premier test consiste à vérifier que sur les périodes antérieures, les groupes traités et de contrôle ont bien connu des évolutions similaires. Un graphique des évolutions des revenus observés pour les différents groupes peut permettre de se faire une idée sur la crédibilité de l'hypothèse identifiante. Un cas d'école où la condition n'est pas vérifiée est donné par Ashenfelter (1978). Ce dernier s'intéressait à l'impact d'une formation sur les revenus de salariés. Une analyse détaillée de l'évolution des revenus avant la formation montrait que l'année précédant la formation, les salariés formés avaient enregistré une croissance de leur revenu plus faible qu'auparavant. Il est vraisemblable que c'est la raison pour laquelle ils ont choisi de se former. Il est dès lors difficile de savoir si l'effet sur les salaires qu'on observe est bien l'impact de la formation, ou celui d'un "rattrapage" des revenus sur un sentier de croissance moyen, après un choc négatif. Cet effet est passé à la postérité sous le terme d'"Ashenfelter's Dip".

Tests "Placebo" De manière plus systématique, on peut appliquer la même procédure de différences de différences, mais autour d'une date où il ne s'est rien passé. Si on mesure malgré tout un impact significatif de cette mesure "placebo", on peut craindre qu'il existe

une différence significative entre les deux groupes et que le groupe de contrôle n'est pas adapté.

Dans le cas où l'on dispose de séries temporelles longues, et où il s'agit de tester une mesure appliquée à des dates différentes selon les groupes, Autor (2003) propose une procédure inspirée des tests de causalité au sens de Granger. L'idée est de vérifier que le fait d'être traité dans le futur n'a pas d'impact sur le revenu actuel. En effet, toute la question de la sélection est de savoir si c'est la mesure qui modifie le revenu auquel on s'intéresse, ou si c'est pour répondre à des évolutions du revenu que la mesure est mise en place. Dit autrement, que les conséquences ne précèdent pas les causes.

En pratique, on peut estimer une relation du type :

$$y_{igt} = e_t + e_g + \sum_{t=0}^n \alpha_\tau T_{gt-\tau} + \sum_{t=1}^m \delta_\tau T_{gt+\tau} \quad (6)$$

On s'attend à ce que les coefficients δ_τ , qui correspondent à l'impact de la mesure après sa mise en place, soient significatifs⁵. En revanche, il serait inquiétant d'observer que faire partie d'un futur groupe traité (coefficients α_τ) a un effet significatif sur le revenu.

Triples différences Enfin, lorsque l'on dispose de plusieurs groupes de contrôle possibles, il est judicieux de vérifier que l'effet mesuré est identique selon qu'on utilise l'un ou l'autre de ces groupes.

Notons que disposer de plusieurs groupes de contrôle possibles peut conduire à mettre en œuvre des estimateurs de triples différences. Le principe de ces estimateurs est le suivant : supposons qu'on s'intéresse à l'effet d'une politique ciblée sur le suivi des chômeurs de plus de cinquante ans, dans une région donnée. Une première solution serait de comparer l'insertion des chômeurs âgés à celle des chômeurs plus jeunes, avant et après la réforme, par un estimateur de différences de différences classique. Le problème est qu'il est très vraisemblable que la dynamique de l'insertion des chômeurs ne soit pas la même dans les différents groupes d'âge : en général, les chômeurs âgés ne profitent pas avec la même rapidité d'une reprise économique que les autres. Une autre solution, si la réforme a été appliquée dans une seule région, serait d'utiliser comme groupe de contrôle les chômeurs âgés des régions voisines. Le problème cette fois est qu'on peut craindre des effets de conjoncture régionale. Une solution consiste à combiner ces deux approches, en appliquant une procédure de triples différences (DDD) : on retranche à l'évolution avant et après la réforme des chômeurs âgés dans la région concernée à la fois l'évolution des chômeurs plus jeunes dans cette région, et celle des chômeurs âgés dans la région où cette politique n'a pas été appliquée.

L'hypothèse est que le revenu potentiel peut se décomposer comme :

$$y_0 = e_{age} + e_t + e_{reg}$$

⁵ils peuvent être différents si la mesure met du temps à produire ses pleins effets, par exemple.

L'estimateur peut s'écrire :

$$\hat{\delta} = [(\bar{y}_{V,R,t_1} - \bar{y}_{V,R,t_0}) - (\bar{y}_{J,R,t_1} - \bar{y}_{J,R,t_0})] \\ - [(\bar{y}_{V,N,t_1} - \bar{y}_{V,N,t_0}) - (\bar{y}_{J,N,t_1} - \bar{y}_{J,N,t_0})]$$

où les indices V et J correspondent respectivement aux chômeurs âgés et plus jeunes, R à la région concernée par la réforme et N à celle qui ne l'est pas. Les deux premiers termes correspondent à un estimateur de différences de différences classique, où le groupe de contrôle est constitué des jeunes chômeurs de la région concernée. Plus précisément, le deuxième terme permet de contrôler le fait que la région où a été introduite cette réforme a pu connaître une conjoncture particulière ; les deux derniers du fait que le taux de chômage des seniors ne réagit pas de la même manière à la conjoncture que celui des plus jeunes. De même, on aurait un estimateur de différences de différences classique en utilisant seulement les premier et troisième termes : dans ce cas, le groupe de contrôle serait constitué des chômeurs âgés de la région voisine.

En pratique, cet estimateur s'obtient en régressant les revenus sur l'ensemble des indicatrices croisées :

$$y = \beta_0 + \beta_1 1_{t>t_0} + \beta_2 1_{i \in V} + \beta_3 1_{i \in R} + \beta_4 1_{t>t_0} 1_{i \in V} \\ + \beta_5 1_{t>t_0} 1_{i \in R} + \beta_6 1_{i \in V} 1_{i \in R} + \delta 1_{i \in B} 1_{i \in V} 1_{t>t_0} + u$$

Comme précédemment, cette procédure tient à l'hypothèse, forte, que les biais se décomposent de manière additive entre les différents termes.

2.3.4 Contrôler par l'état initial ?

L'hypothèse identifiante suppose implicitement que les différences de revenu (sans traitement) entre les bénéficiaires et les non-bénéficiaires sont stables dans le temps. Ce n'est pas toujours une hypothèse crédible. Typiquement, dans l'exemple d'Ashenfelter, les déterminants de l'inscription à un programme de formation dépendent justement de chocs temporels sur le revenu spécifiques aux bénéficiaires.

On peut donc préférer spécifier un modèle dynamique qui prennent ces évolutions en compte : précisément en conditionnant par les revenus observés les périodes antérieures. On peut ainsi estimer :

$$Y_{it} = \alpha + \gamma Y_{it-1} + e_t + \delta T_{it} + X_{it}\beta + v_{it} \quad (7)$$

On pourrait être tenté de mixer les deux approches : mettre des effets fixes et corriger des évolutions passées du revenu. Malheureusement, ces deux spécifications reposent sur des hypothèses non cohérentes entre elles, et il est donc difficile d'identifier un modèle avec à la fois des effets fixes et de l'auto-corrélation. Il s'agit d'un problème bien connu en économétrie des panels concernant les modèles dynamiques. On peut s'en rendre compte avec le modèle simple suivant :

$$Y_{it} = c_i + \gamma Y_{it-1} + e_t + \delta T_{it} + X_{it}\beta + v_{it} \quad (8)$$

qui ne diffère de celui présenté dans l'équation (7) par le fait qu'on introduit maintenant des effets fixes individuels c_i à la place de la constante. On adopte des hypothèses standard sur les résidus v ($E(v) = 0$, $V(v) = \sigma_v^2$). La présence d'effets fixes complique évidemment l'estimation. Dans ce cadre linéaire, il se résout simplement, par exemple en utilisant des différences premières :

$$\Delta Y_{it} = \gamma \Delta Y_{it-1} + \Delta X_{it} \beta + e_t + \delta \Delta T_{it} + \Delta v_{it} \quad (9)$$

Le problème est que le terme résiduel $\Delta v_{it} = v_{it} - v_{it-1}$ est par construction corrélé avec le terme $\Delta Y_{it-1} = Y_{it-1} - Y_{it-2}$ car $cov(Y_{it-1}, v_{it-1} | X_{it}, T_{it}, t) = \sigma_v^2 \neq 0$.

Notons que cela ne signifie pas qu'une spécification soit supérieure à une autre, mais que les deux ne sont pas cohérentes, et qu'il faut nécessairement choisir l'une ou l'autre. En général, la démarche la plus sage est de tester les deux, en étant conscient qu'elles reposent sur des hypothèses d'identification différentes, et de vérifier qu'elles ne conduisent pas à estimations contradictoires. On peut montrer (voir par exemple Angrist et Pischke (2009)) qu'estimer un modèle avec une spécification à effets fixes, alors qu'il s'agit en réalité d'un modèle dynamique, tend à sur-estimer l'impact de la mesure étudiée. À l'inverse, estimer à tort un modèle dynamique alors qu'il s'agit d'un modèle à effets fixes tend à sous-estimer l'impact. Les deux estimations peuvent être alors vues comme des bornes de l'effet réel.

2.3.5 Extension : différences de différences dans un cadre non linéaire

L'hypothèse de séparabilité linéaire est forte. Athey et Imbens (2006) proposent une extension du modèle de différences de différences dans un cadre non linéaire.

Dans le cadre classique, avec deux périodes et deux groupes, ils montrent qu'il est possible d'identifier non seulement l'effet moyen du traitement, mais également toute la distribution contrefactuelle du revenu pour les bénéficiaires $F_{Y_0|g=1,t=1}$.

Supposons que le revenu potentiel Y_0 s'écrive sous une forme très générale :

$$Y_{i0} = h_0(u_i, t_i) \quad (10)$$

où t_i est une indicatrice de période ($t_i = 0, 1$), et u_i est une variable aléatoire qui représente toutes les caractéristiques inobservables de l'individu i stables au cours du temps⁶. La seule hypothèse sur la dépendance du revenu potentiel est que la fonction h_0 est strictement croissante en u pour les deux périodes $t = 0, 1$ ⁷. On suppose que la distribution de ces inobservables au sein des groupes est stable avec le temps :

$$F_{U|G,t} = F_{U|G} \quad (11)$$

⁶Dans le modèle en différences de différences classique, cette fonction est séparable dans ces deux arguments.

⁷Cette hypothèse de stricte monotonie écarte donc le cas des variables discrètes. En cas de monotonie faible (typiquement lorsque la variable expliquée est binaire), Athey et Imbens (2006) montrent cependant qu'on peut obtenir des bornes sur le revenu.

Athey et Imbens montrent alors qu'on peut identifier la distribution contrefactuelle du revenu des bénéficiaires de la mesure. Notons que la distribution du revenu Y_1 conditionnellement au fait de faire partie du traitement $F_{Y_{i1}|g_i=1,t_i=1}$ est directement observée dans les données.

Sous une hypothèse supplémentaire que le support de la distribution des inobservables du groupe des bénéficiaires $F_{u_i|g_i=1}$ est inclus dans le support de la distribution des inobservables du groupe des non-bénéficiaires $F_{u_i|g_i=0}$, on peut identifier la distribution contrefactuelle $F_{Y_{i0}|g_i=1,t_i=1}$.

On note $F_{g,t}^0(y)$ (resp. $F_{g,t}^1(y)$) la distribution du revenu potentiel en l'absence de la mesure (resp. en présence de la mesure), conditionnelle au fait de faire partie du groupe g à la période t et $F_{gt}(y)$ la distribution des revenus observés conditionnelle au fait d'appartenir au groupe g à la période t . Rappelons que $F_{g,t}^0(y)$ et $F_{g,t}^1(y)$ ne sont jamais observées simultanément, seule $F_{g,t}(y) = TF_{g,t}^1(y) + (1 - T)F_{g,t}^0(y)$ est observée. Cependant, Athey et Imbens (2006) montrent que sous les conditions qui précèdent,

$$F_{11}^0(y) = F_{10}(F_{00}^{-1}(F_{01}(y))) \quad (12)$$

F_{00}^{-1} est l'inverse de la distribution de revenu observé dans le groupe de contrôle à la date $t = 0$, F_{10} la distribution de revenu observée dans le groupe des bénéficiaires à la date $t = 0$, et F_{01} la distribution de revenu observée dans le groupe de contrôle à la date $t = 1$. Notons que la distribution de revenu potentiel Y_1 du groupe traité $g = 1$ et à la période $t = 1$ correspond à la distribution du revenu observé pour ce groupe à cette date :

$$F_{11}^1(y) = F_{11}(y)$$

On peut donc estimer l'impact de la mesure en tout point de la distribution.

L'effet moyen du traitement s'écrit alors :

$$\begin{aligned} \delta &= E(y_{11}^1) - E(y_{11}^0) \\ &= E(y_{11}) - E(F_{01}^{-1}(F_{00}(y_{10}))) \end{aligned}$$

Et un estimateur peut être donné par :

$$\hat{\delta} = N_{11}^{-1} \sum_1^{N_{11}} y_{11,i} - N_{10}^{-1} \sum_1^{N_{10}} \hat{F}_{01}^{-1}(\hat{F}_{00}(y_{10,i}))$$

où on utilise les distributions \hat{F}_{01} et \hat{F}_{00} estimées dans le groupe de contrôle avant et après⁸.

⁸Rappelons qu'on obtient simplement les estimations empiriques à partir d'un échantillon de taille n par $\hat{F}_n(y) = 1/n \sum_{i=1}^n 1_{y_i < y}$. Pour la fonction inverse (qui correspond au calcul des différents quantiles), l'estimation peut être plus complexe, mais elle est fournie de manière standard par les différents logiciels.

2.3.6 Inférence

Nous avons montré différentes méthodes pour estimer un paramètre. Il faut maintenant s'interroger sur l'inférence, i.e. ce que l'on peut dire sur la vraie valeur du paramètre d'intérêt, à partir de l'estimation obtenue.

La méthode classique consiste à estimer la distribution statistique du paramètre estimé, ce qui permet par exemple d'estimer un intervalle de confiance, et de faire des tests de significativité. Des problèmes peuvent se poser si il existe des corrélations temporelles et/ou entre les individus appartenant aux mêmes groupes. L'application des tests classiques fournit alors des diagnostics faussés.

Dans le cas le plus simple, le modèle est homoscedastique, i.e. la variance des termes d'erreurs est identique pour toutes les observations, et ils ne sont pas corrélés entre eux : $V(u_{igt}|g, t, X, D_{gt}) = \sigma^2 I_N$. Si cette dernière hypothèse n'est pas vérifiée (i.e. si les termes d'erreurs u_{igt} ne sont pas indépendants entre eux), l'estimateur classique de sa variance est biaisé. Cela signifie que les tests appliqués à partir de cet estimateur peuvent fournir des résultats incorrects. En particulier, une application simple du test de Student peut conduire à rejeter à tort l'hypothèse de nullité du coefficient (et donc considérer qu'une mesure a un impact alors qu'il est en réalité négligeable).

Dans le cas des estimations par différences de différences, deux problèmes différents peuvent conduire à obtenir une matrice de variance-covariance non diagonale. Le premier est que les individus au sein d'un groupe subissent probablement des chocs communs. Le second est que ces chocs sont également corrélés dans le temps.

Données groupées L'existence de chocs communs apparaît si l'on utilise pour les estimations des données individuelles, alors qu'on s'intéresse à des différences entre des groupes d'individus : typiquement, des différences entre deux états ou deux régions, pour lesquelles on dispose d'échantillons d'individus. Il est cependant vraisemblable que tous les revenus des individus d'un même groupe subissent des chocs communs. Ce sera par exemple le cas si le groupe est constitué des individus d'une même région exposés à la même conjoncture régionale.

Formellement, dans le cadre du modèle de différences de différences, cela revient à considérer que les termes d'erreurs individuels comportent un terme groupe*temps :

$$u_{igt} = \nu_{g,t_i} + \epsilon_i$$

où ν_{g,t_i} est une composante groupe*temps. Elle signifie que certains chocs affectent simultanément l'ensemble des individus d'un groupe. On a dans ce cas à chaque date une corrélation entre les résidus $u_{igt} : E(u_{igt}u_{i'gt}) \neq 0$ pour deux individus différents i et i' du même groupe g , à la période t . Dans ce cas, les résidus ne sont plus indépendants, et la matrice de variance covariance n'est plus diagonale⁹. Ceci introduit un premier biais, dit

⁹si le seul problème est l'existence de ce choc commun à l'intérieur des groupes, elle est bloc diagonale si on regroupe les individus d'un même groupe et d'une même période.

de “données groupées” (*cluster* en anglais) et soulevé en particulier par Moulton (1990).

Avec seulement deux groupes et deux périodes, le problème est sérieux, puisque l’estimateur n’est pas convergent (Donald et Lang 2007). L’intuition est que si un choc purement conjoncturel affecte l’ensemble des personnes d’une région dans laquelle est mise en place une politique, il sera difficile de déterminer si l’écart de revenu est lié à la mesure ou à ce choc.

Formellement :

$$\begin{aligned} y_{igt} &= c_g + c_t + \delta \mathbf{1}_{igt} + u_{igt} \\ &= c_g + c_t + \delta \mathbf{1}_{g_i=1t_i=1} + \nu_{g_i,t_i} + \epsilon_i \end{aligned}$$

L’estimateur correspond à :

$$\begin{aligned} \hat{\delta} &= \bar{y}_{11} - \bar{y}_{10} - (\bar{y}_{01} - \bar{y}_{00}) \\ &= \delta + \hat{\nu}_{11} - \hat{\nu}_{10} - \hat{\nu}_{01} + \hat{\nu}_{00} \end{aligned}$$

On peut par exemple supposer que les termes ν_{gt} suivent une loi normale $N(0, \sigma_\nu^2)$ ¹⁰. On suppose pour l’instant qu’ils sont non corrélés entre eux. On a alors :

$$E(\hat{\delta}) \rightarrow \delta + \nu_{11} - \nu_{10} - \nu_{01} + \nu_{00}$$

L’estimateur n’est donc pas convergent¹¹. Si on pense que les chocs groupés sont importants, cela est problématique car il n’est pas possible d’estimer le biais. Avec seulement deux périodes et deux groupes, on ne peut obtenir un estimateur correct de σ_ν^2 .

Le problème n’est pas lié au nombre d’observations totales, mais au nombre de groupes, c’est-à-dire aux unités pour lesquelles on peut supposer que les chocs qu’ils subissent ne sont pas (trop) corrélés. Disposer de nombreux individus pour chaque groupe permet d’obtenir une meilleure estimation du revenu moyen \bar{y}_g du groupe. Mais si la composante groupe ν_g constitue une part importante (et indiscernable) de ce revenu moyen, cette estimation du revenu moyen est peu exploitable pour estimer l’effet du traitement.

En revanche, disposer de nombreuses observations peut conduire à une inférence incorrecte, si on “oublie” de spécifier ces effets groupés. La précision estimée augmente en effet mécaniquement avec le nombre d’observations. En réalité, c’est le nombre de groupes (et non le nombre d’observations par groupe) qui fournit l’information “utile”.

Auto-corrélation des revenus À ce problème des données groupées s’ajoute dans le cas d’un modèle de différences de différences l’auto-corrélation temporelle possible des revenus. Reprenons l’équation :

$$y_i = e_t + e_g + \delta T_{gt} + X_i \beta + u_{igt} \tag{13}$$

¹⁰si le modèle est bien spécifié les termes d’erreur sont centrés : $E(\nu_{gt}) = 0$.

¹¹Sa distribution sera $\mathbf{N}(\delta, 4\sigma_\nu^2)$.

On peut agréger les données au niveau du groupe pour éviter le problème de données groupées exposé précédemment. Mais si la conjoncture économique d'une région est mauvaise une année donnée, il est probable qu'elle le sera encore l'année suivante. Cela signifie qu'il existe de la corrélation temporelle entre les termes d'erreur des deux revenus. On peut donc avoir : $E(\bar{u}_{gt}\bar{u}_{gt'}) \neq 0$ pour deux périodes différentes. Là encore, les tests d'inférence classiques ne sont pas corrects.

Ces questions ne sont pas anecdotiques, comme le montrent Bertrand, Duflo, et Mullainathan (2004). Pour cela, ils estiment l'effet de lois *placebos* à partir de données réelles : une année de "réforme" est tirée aléatoirement, ainsi qu'un ensemble de 25 États américains considérés comme "traités". Ils estiment alors l'impact de cette réforme fictive sur le revenu des États "traités" à partir d'une procédure de différences de différences. L'effet réel est ici, par construction, zéro. Ils font ainsi 200 simulations de ce type. Si les écart-types fournis par les MCO sont corrects, le test classique de significativité à 5 % doit aboutir à rejeter (à tort) l'hypothèse nulle dans seulement 5 % des cas (c'est le principe même du test). Cependant, Bertrand, Duflo, et Mullainathan (2004) rejettent l'hypothèse nulle dans 67 % des cas ! C'est-à-dire que dans plus de deux tiers des cas, une étude naïve conduirait à penser qu'une loi qui n'a pas eu lieu a pourtant eu un effet réel. Lorsqu'ils agrègent les données, pour éviter le problème de données groupées décrit plus haut, ils rejettent à tort encore près d'une fois sur deux l'hypothèse nulle¹². Il est donc vraisemblable qu'il subsiste un problème d'auto-corrélation important.

Corrections Plusieurs méthodes sont proposées dans la littérature pour corriger ces deux problèmes, même s'il n'existe cependant pas (encore ?) de réel consensus sur la question¹³.

Elles convergent toutes sur un point : la nécessité de disposer de suffisamment de groupes et / ou de périodes. Comme on l'a déjà dit, il est inutile de disposer de beaucoup d'observations individuelles par groupe, mais de beaucoup de groupes différents. Si le nombre de groupes est trop faible, on ne pourra aller beaucoup plus loin. La distribution pertinente de la statistique de Student est la loi de Student à $G - 2$ degrés de liberté, et non à $M_1 + \dots + M_G - 2$ (où M_g représente le nombre d'observations dans le groupe g).

Citons deux méthodes générales :

1. estimateurs de la matrice de variance-covariance

Tous les logiciels statistiques proposent des estimateurs de la matrice de variance-covariance, en général fondés sur une généralisation de l'estimateur de White classique. En pratique, on peut par exemple agréger les données par groupes (pour traiter le problème des données groupées), puis faire des estimations en groupant les données temporelles au niveau de ces groupes, pour traiter l'auto-corrélation. Encore une fois, cela n'est possible que lorsque le nombre de groupes est suffisant.

¹²Ce résultat demeure lorsqu'ils utilisent des données simulées plutôt que des données réelles, afin de ne pas risquer de mesurer par hasard l'effet d'une loi ayant effectivement eu lieu.

¹³On en trouvera une synthèse dans le chapitre 8.2 d'Angrist et Pischke (2009).

2. *Block Bootstrap*

Rappelons que le principe du *bootstrap* est d’approcher la distribution d’un estimateur en simulant des échantillons fictifs à partir de notre échantillon observé : en pratique, si on dispose d’un échantillon de N observations, on obtient un nouvel échantillon en tirant *avec remise* N individus dans notre échantillon initial. Répéter cette opération permet de simuler la distribution de notre estimateur, et donc d’estimer sa variance dans les cas où elle n’est pas calculable simplement. Dans notre cas, on doit appliquer une méthode de *block bootstrap*, i.e. on respecte la structure de corrélation des états, en tirant simultanément toutes les observations d’un groupe en même temps (i.e toute la série temporelle des revenus du groupe), afin de maintenir la structure d’auto-corrélation interne aux états. Là encore, ceci n’est valide qu’à condition de disposer de suffisamment de groupes. Par exemple, en utilisant cette méthode en groupant les observations par état américain, Bertrand et al. montrent qu’on ne rejette cette fois-ci à tort l’hypothèse nulle que dans 6.5 % des cas (soit une proportion assez proche du taux de rejet théorique de 5 %), mais que si on n’utilise les données que de dix états ce taux de rejet monte à 23 %. Cameron, Gelbach, et Miller (2008) proposent des estimateurs qui donnent des résultats satisfaisants même pour un nombre de groupes restreint.

Une méthode originale Nous nous sommes intéressés jusqu’à maintenant dans cette section à la question de l’inférence statistique. Celle-ci répond fondamentalement à un problème d’échantillonnage : il s’agit de s’interroger sur la représentativité de l’échantillon sur lequel est menée l’estimation¹⁴. Sous certaines hypothèses, on peut déterminer la distribution de probabilité du paramètre estimé. Pour *inférer* le “vrai” effet de la mesure à partir de nos estimations, il faut cependant répondre à une autre question : dans quelle mesure le groupe de contrôle qu’on utilise permet d’estimer un bon contrefactuel des revenus des bénéficiaires s’ils n’avaient pas bénéficié de la mesure. Comme noté par Abadie, Diamond, et Hainmueller (2007), une observation exhaustive de la population (fournie par exemple par des données administratives) permet théoriquement de s’affranchir des problèmes d’échantillonnage, mais n’assure en aucun cas que l’estimation du contrefactuel des bénéficiaires sera correcte. Partant de ce constat, ils proposent une approche complètement différente de l’inférence statistique classique pour fournir un intervalle de confiance de leur estimateur.

Rappelons que leur cadre est celui de l’estimation de l’effet d’une mesure qui affecte un État, pour lequel on dispose de plusieurs groupes de contrôle (les autres États américains). Leur contrefactuel correspond à une combinaison linéaire des revenus de l’ensemble des groupes de contrôle disponibles, suivant une pondération liée à la distance au groupe précédent. Le “test” qu’ils proposent consiste à répliquer systématiquement la même procédure à l’ensemble des États du groupe de contrôle. L’effet de ces lois *placebos* fournit l’équivalent

¹⁴Typiquement, on estime un paramètre sur un échantillon de la population, et on souhaite savoir si ce paramètre représente bien le paramètre qui aurait été estimé sur l’ensemble de la population. L’inférence classique tente alors de déterminer la distribution de l’estimateur qu’on obtiendrait si on tirait un grand nombre d’autres échantillons à la place de celui dont on dispose.

d'un intervalle de confiance, des écarts que l'on peut observer dans la consommation de tabac entre les états, même en l'absence de législation spécifique. Dans leur exemple, l'impact estimé de la loi anti-tabac en Californie (courbe en noir sur le graphique) dépasse les écarts systématiques qui sont révélés par ces *placebos* (courbes grises). Il est donc raisonnable de penser que la loi a réellement eu un impact sur la consommation de tabac, qui dépasserait un pur effet fortuit. Cependant, cette méthode "agnostique" ne fournit aucun résultat sur la distribution de probabilités de l'estimateur obtenu. Il n'est donc pas possible de quantifier le risque de se tromper en rejetant l'hypothèse nulle que la législation n'aurait aucun impact.

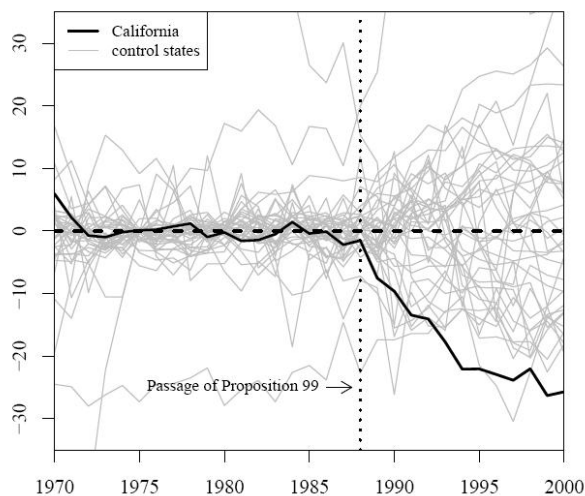


FIG. 4: Abadie, Diamond et Hainmuller (2007) : Évolution contrefactuelle de la consommation de tabac en Californie pour l'ensemble des États américains

2.4 En résumé

En résumé, pour utiliser une méthode de différences de différences pour évaluer une mesure de politique publique, il faut :

1. Déterminer un groupe de contrôle :
il ne sera “crédible” que si on peut supposer que les évolutions des revenus sont les mêmes que celles qu’auraient connu les bénéficiaires en l’absence de la mesure. Attention, il ne peut être simplement constitué des non-bénéficiaires, dès lors qu’il y a (auto) sélection dans le dispositif. Voir section 2.3.1
Disposer de données cohérentes avant et après la mise en place de la mesure est indispensable. Il est préférable que les données soient relativement proches dans le temps, pour éviter de capter d’autres évolutions qui ne seraient pas liées à la mesure.
2. Estimer : en pratique, cela peut se faire très simplement par une procédure des moindres carrés ordinaires (voir section 2.3) Le calcul de la précision de l’estimation est nécessaire pour borner statistiquement la vraie valeur de l’impact de la mesure. Attention, lorsqu’on utilise des données individuelles, il faut tenir compte de l’auto-corrélation temporelle et/ou des chocs “groupés”. Pour cela, on peut utiliser des procédures de *bootstrap* par bloc (voir section 2.3.6).

3 Variables instrumentales

3.1 Rappel sur les variables instrumentales

L'utilisation des variables instrumentales pour traiter des problèmes d'endogénéité est très ancienne, et plus générale que son utilisation pour des questions d'évaluation des politiques publiques. On fait généralement remonter à Wright (1928) l'intuition de ce qui correspond à un instrument.

Rappelons que l'on se place dans le cadre standard de l'équation linéaire :

$$y_i = \alpha + \delta T_i + u_i \quad (14)$$

où l'on soupçonne le régresseur T_i d'être endogène, c'est-à-dire d'être corrélé avec les résidus u_i . C'est typiquement le cas lorsqu'il existe des effets de sélection.

Un instrument est alors simplement une variable Z qui remplit deux conditions (éventuellement conditionnellement à des observables X) :

1. elle est corrélée avec la variable endogène T : $cov(T, Z) \neq 0$
2. mais pas avec les résidus u : $cov(u, Z) = 0$

Si une telle variable existe, il est possible d'obtenir une estimation convergente du paramètre d'intérêt δ . L'estimateur le plus classique est celui des doubles moindres carrés. Il s'obtient en régressant le revenu Y non plus sur le régresseur endogène T , mais sur sa prédiction \hat{T} issue de la régression de première étape de T sur Z :

$$\hat{\delta}^{IV} = (\hat{T}'\hat{T})^{-1}\hat{T}'Y \quad (15)$$

On peut donc l'obtenir en deux étapes, d'où le nom de doubles moindres carrés : on régresse le traitement sur l'instrument ; puis le revenu sur la valeur prédite par cette première régression. Attention, il est fortement déconseillé de procéder ainsi : les écarts-types fournis par la deuxième régression ne sont pas corrects (voir par exemple Robin (2002)). Tous les logiciels de statistiques standard permettent cependant de mettre en œuvre cette procédure (par exemple sous *SAS* la procédure *syslin*, option *2sls*) ; ils fournissent en particulier une estimation correcte de la précision de cet estimateur.

Remarque : lorsque l'instrument est binaire, un estimateur courant dans la littérature de l'évaluation est l'estimateur de Wald :

$$\hat{\delta}_{IV} = \frac{E(Y|Z=1) - E(Y|Z=0)}{E(T|Z=1) - E(T|Z=0)} \quad (16)$$

Sa contrepartie empirique est juste la différence des moyennes empiriques des revenus du groupe des individus pour lesquels $Z = 1$ et du groupe des individus pour lesquels $Z = 0$, rapporté à l'équivalent pour l'instrument. On peut montrer simplement qu'il correspond exactement à la variation des revenus avec l'instrument (moyenne des revenus pour les observations telles que $Z = 1$ moins la moyenne des observations pour celles telles que $Z = 0$), rapporté à la variation du traitement avec l'instrument.

Il correspond exactement à l'estimation des doubles moindres carrés ordinaires dans l'équation (14) (sans variables supplémentaires).

3.2 Quels instruments ?

Pour simple qu'elle soit, la définition d'une variable instrumentale peut laisser fort démuni un économètre averti. Un instrument valide est une variable qui est liée au fait de bénéficier d'un programme, pour une raison que l'économètre peut expliciter et expliquer, mais qui n'ait par ailleurs pas d'impact direct sur les autres déterminants du revenu. Il est parfois d'usage d'utiliser les valeurs ou différences retardées. La validité de ces instruments peut être parfois sujette à caution. C'est pourquoi l'attention s'est déplacée sur les dernières années sur des instruments dont l'exogénéité était plus transparente.

3.2.1 Affectation aléatoire

Les plus crédibles sont ceux fournis par à une affectation aléatoire, par exemple dans le cadre d'une expérience contrôlée. En général, les personnes affectées au groupe de traitement se voient proposer de participer à un programme, ou sont plus incitées que celles du groupe de contrôle, mais elles peuvent ensuite choisir de participer ou non. Ceci introduit à nouveau de la sélection, puisqu'il est probable que celles qui choisissent *in fine* de participer en anticipent un bénéfice plus élevé que les autres. Néanmoins, on dispose d'un instrument "idéal", l'affectation au traitement Z (i.e. le fait d'affecter une personne au groupe des participants, auquel le programme qu'on souhaite évaluer sera proposé, ou au groupe de contrôle). Par définition de l'expérimentation, celle-ci est en effet aléatoire, donc :

$$Y_1, Y_0 \perp Z \quad (17)$$

Si l'expérience est bien menée, on peut légitimement espérer que cette affectation soit corrélée avec la participation réelle :

$$\text{cov}(Z, T) \neq 0 \quad (18)$$

les conditions pour disposer d'un bon instrument sont donc remplies. On peut donc estimer l'impact du programme à évaluer par une méthode classique de doubles moindres carrés, ou par un estimateur de Wald.

3.2.2 Expériences naturelles

En l'absence d'expérience aléatoire, de bons instruments pourront être fournis par des "expériences naturelles". Comme décrit plus haut (voir section 2), il s'agit d'utiliser des variations exogènes de l'environnement des individus qui peuvent expliquer des probabilités plus ou moins élevées de participer. Par exemple, Angrist et Evans (1998) utilisent la "préférence pour la différence" des parents américains pour évaluer l'impact de la fécondité sur la participation des femmes au marché du travail. On observe en général que les mères de famille nombreuses (plus de trois enfants) sont moins souvent actives sur le marché du travail. Il est difficile de déterminer le sens de la causalité : est-il plus difficile pour ces mères de travailler (par exemple du fait d'une prise en charge insuffisante de la petite enfance, ou de discrimination de la part des employeurs), ou ces femmes privilégient-elles

simplement la vie de famille à la vie professionnelle? Angrist et Evans (1998) proposent un instrument original pour déterminer la probabilité d’avoir une famille nombreuse. Sur les générations récentes, on observe plus souvent des familles nombreuses (plus de trois enfants) lorsque les deux aînés sont du même sexe. Le fait d’avoir deux aînés du même sexe est donc un prédicteur du fait d’avoir au moins trois enfants. En revanche, il s’agit d’un événement totalement hors de portée de la volonté des parents, et donc indépendant des caractéristiques éventuelles des femmes concernées.

Autre exemple célèbre, l’utilisation du trimestre de naissance pour évaluer les rendements de l’éducation (surcroît de revenu assuré par une année d’études supplémentaire) par Angrist et Krueger (1991). En effet, du fait de la législation américaine sur la scolarité obligatoire, les enfants entrent à l’école en septembre de l’année de leur six ans (soit entre cinq ans et neuf mois et six ans et huit mois selon leur mois de naissance). En revanche, l’obligation scolaire cesse le jour des seize ans, quel que soit le moment dans l’année scolaire. Les enfants nés en début d’année sont donc théoriquement astreints à une scolarité obligatoire plus courte que ceux nés en fin d’année. De fait, pour les générations nées dans les années trente et quarante (pour lesquelles la scolarité obligatoire était vraiment contraignante), être né le premier trimestre réduit de 0,6 année la durée d’étude finale. Là encore, on peut supposer que le trimestre de naissance n’est en revanche pas corrélé avec les autres déterminants du revenu, ce qui en fait un bon instrument potentiel (nous revenons plus loin sur cet exemple).

D’autres études utilisent la dispersion régionale de certains dispositifs : par exemple, pour évaluer l’efficacité d’un nouveau protocole de prise en charge des infarctus, McClellan, McNeil, et Newhouse (2008) utilisent la proximité de la résidence des patients aux hôpitaux utilisant ce nouveau protocole¹⁵.

3.3 Interprétation : un estimateur local

Que mesure l’estimateur des variables instrumentales? Dans cette partie, nous nous plaçons dans le cadre simple d’un instrument binaire (i.e. valant 0 ou 1), tout comme le traitement.

Comme l’ont montré Angrist, Imbens, et Rubin (1996) dans un papier influent, il ne permet pas de mesurer l’effet moyen sur l’ensemble des personnes “traitées” (le paramètre Δ^{ATE} que nous avons défini en introduction), mais uniquement sur une sous-population, les *compliers*. Il s’agit des personnes pour lesquelles l’instrument a un impact sur le fait de suivre ou non le traitement.

L’exemple d’Angrist, Imbens, et Rubin (1996) repose sur la “loterie” utilisée pour la conscription durant la guerre au Vietnam, qu’ils utilisent pour instrumenter l’effet d’avoir été vétéran sur les revenus ultérieurs. Il s’agit *a priori* d’un instrument “parfait” : il est

¹⁵Pour estimer l’effet de faire des études supérieures sur le revenu, Card (1993) utilise un instrument proche comme la proximité dans l’enfance à un ville universitaire. Cet instrument est cependant à utiliser avec précaution : il n’est pas sûr que le choix de localisation des parents soit totalement déconnecté des possibilités futures de faire des études.

aléatoire, donc non corrélé avec les déterminants inobservés du traitement. Et selon le numéro obtenu à la loterie, les jeunes des tranches d’âge concernées ont eu plus ou moins de chances d’échapper à la conscription. Notons que cela ne signifie pas, loin s’en faut, que le numéro obtenu corresponde exactement au fait d’être parti au Vietnam. Certains ont pu s’engager alors qu’ils avaient un numéro leur permettant d’en être dispensés, tandis que d’autres ayant reçu un “mauvais” numéro ont été réformés, pour raisons physiques par exemple.

Les individus adaptent (ou non) leur comportement à la valeur de l’instrument. Plus exactement, on a une fonction de réaction “potentielle” $T(Z)$. Dans l’exemple de la loterie du Vietnam, il s’agit de la manière dont le choix de s’engager ou pas dépend du numéro attribué. Évidemment, on est confronté au même problème observationnel que pour les revenus potentiels : on observe le choix d’un individu pour une valeur particulière de l’instrument, mais pas pour l’autre. Dans le cas où l’instrument et le traitement sont binaires, on a la partition de la population fournie par le tableau 1.

TAB. 1: Partition de la population selon le comportement par rapport à l’instrument

	$T = 1 Z = 1$	$T = 0 Z = 1$
$T = 1 Z = 0$	Always Takers	Defiers
$T = 0 Z = 0$	Compliers	Never Takers

Lecture : les personnes qui choisissent toujours le traitement quelle que soit la valeur de l’instrument, donc telles que $P(T = 1|Z = 1) = P(T = 1|Z = 0) = 1$ sont appelées *always takers*.

Plus précisément, on désigne par *compliers* la population pour laquelle l’instrument a un effet (ils font ce qu’on leur a dit de faire). Mais pour d’autres, l’instrument n’a pas d’effet. Soit qu’ils ne choisissent jamais le programme, $P(T = 1|Z = 1) = P(T = 1|Z = 0) = 0$ (Angrist, Imbens, et Rubin (1996) les nomment les *never takers*), soit qu’ils le choisissent systématiquement $P(T = 1|Z = 1) = P(T = 1|Z = 0) = 1$ (*always takers*). Enfin, on pourrait en toute rigueur imaginer qu’il existe des “rebelles” (*defiers*), i.e. des personnes pour lesquelles l’instrument a un effet opposé à celui couramment attendu. Dans le cas de la loterie du Vietnam, il s’agirait des cas improbables de personnes qui se seraient engagées si elles avaient reçu un numéro leur permettant de ne pas partir et qui à l’inverse se seraient faites réformer en cas de numéro les soumettant à la conscription...

Angrist, Imbens, et Rubin (1996) montrent que l’estimateur des doubles moindres carrés permet d’estimer l’effet du traitement uniquement pour la sous-population des *compliers*, c’est-à-dire ceux pour qui l’instrument a un impact sur le traitement final. Dans l’exemple du Vietnam, il s’agit des personnes qui se sont engagées (ou non) uniquement en fonction du numéro qui leur a été attribué.

L’effet estimé est donc dit local : Angrist, Imbens, et Rubin (1996) l’appelle le *LATE*, soit *Local Average Treatment Effect*.

Ils se placent en effet dans le cadre d'une affectation aléatoire de l'instrument Z , indépendant du revenu (tout l'effet passe par le traitement), et supposent que les *defiers* n'existent pas. Plus précisément, ils supposent la relation *monotone* suivante :

$$P(T_i = 1|Z_i = 1) \geq P(T_i = 1|Z_i = 0) \quad (19)$$

Ceci permet d'écrire :

$$E(Y_i|Z_i = 1) = E[Y_{i0} + T_i(Y_{i1} - Y_{i0})|Z_i = 1] \quad (20)$$

$$= E[Y_{i0} + T_i(1)(Y_{i1} - Y_{i0})] \quad (21)$$

d'après la condition d'exclusion¹⁶.

Donc :

$$\begin{aligned} E(Y_i|Z = 1) - E(Y_i|Z = 0) &= E[(T_i(1) - T_i(0))(Y_{i1} - Y_{i0})] \\ &= E(0 * (Y_{i1} - Y_{i0})|(T_i(1) - T_i(0)) = 0)P(T_i(1) - T_i(0) = 0) \\ &\quad + E(1 * (Y_{i1} - Y_{i0})|(T_i(1) - T_i(0)) = 1)P(T_i(1) - T_i(0) = 1) \\ &\quad + E(-1 * (Y_{i1} - Y_{i0})|(T_i(1) - T_i(0)) = -1)P(T_i(1) - T_i(0) = -1) \\ &= E(1 * (Y_{i1} - Y_{i0})|(T_i(1) - T_i(0)) = 1)P(T_i(1) - T_i(0) = 1) \end{aligned}$$

car d'après l'hypothèse de monotonie $P(T_i(1) - T_i(0) = -1) = 0$ (pas de *defiers*).

De même, on montre que $E(T|Z = 1) - E(T|Z = 0) = P(T(1) - T(0) = 1)$ et donc que :

$$\delta_{IV} = E[(Y_{i1} - Y_{i0})|T_i(Z = 1) - T_i(Z = 0) = 1] \quad (22)$$

L'estimateur ne mesure donc que l'effet du traitement pour ces *compliers*. Si l'effet du traitement est homogène, $Y_{i1} - Y_{i0} = cst$, ce n'est pas un problème. Sinon, on identifie uniquement un effet local.

L'effet n'est donc identifié que sur une sous-population particulière. Deux remarques sont importantes : la première est que cette population n'est pas connue : *a priori*, il n'est pas possible de déterminer si tel ou tel individu appartient à la population des *compliers*. La seconde est que l'estimation dépendra de l'instrument utilisé. Nous discutons plus avant ces deux points ici.

3.3.1 L'estimation se fait sur une sous-population qui n'est pas connue

Il n'existe pas de possibilité de caractériser la population des *compliers* (pour cela, il faudrait observer leur comportement pour deux valeurs différentes de l'instrument). Notons que sous l'hypothèse de monotonie, on peut cependant identifier certaines populations à partir de l'observation du couple (T_i, Z_i) . Le tableau 2 présente la répartition des différents types selon les valeurs prises par la variable de traitement et l'instrument.

¹⁶On a $T = T(1)1_{Z=1} + T(0)1_{Z=0}$ donc $E(T|Z = 1) = E(T(1)|Z = 1) = E(T(1))$, la dernière égalité provenant de l'hypothèse d'indépendance de Z et T .

TAB. 2: Partition de la population selon les valeurs prises par l'instrument et l'indicatrice de traitement

	$Z = 0$	$Z = 1$
$T = 0$	Never-Takers / Compliers	Never-Takers (/ Defiers)
$T = 1$	Always-Takers (/ Defiers)	Compliers / Always-Takers

Si on exclut la présence de *defiers*, on pourra donc identifier certains *never takers* et certains *always takers*. Par exemple, les premiers sont ceux qui choisissent de ne pas participer alors qu'ils seraient plutôt incités à le faire par la valeur de l'instrument.

Cela permet, par exemple, d'estimer la taille de la population sur laquelle est estimée ce LATE.

Plus précisément, on a :

1. $\pi_a = P(T_i = 1|Z_i = 0)$ proportion d'*always takers* dans la population
2. $\pi_n = P(T_i = 0|Z_i = 1)$ proportion de *never takers* dans la population

Ces égalités viennent du fait que l'instrument est aléatoire et donc indépendant du type. La proportion d'*always takers* parmi $Z_i = 1$ est la même que dans l'ensemble de la population. On peut donc en déduire la proportion de *compliers* dans la population totale $\pi_c = 1 - \pi_a - \pi_n$.

Par exemple, Oreopoulos (2006) s'intéresse à l'estimation des rendements de l'éducation fondée sur l'augmentation de l'âge scolaire. L'idée est que cette réforme contraint certains, de manière exogène, à prolonger leur scolarité. L'instrument est alors lié à l'année de naissance : suivant qu'une personne fait partie ou non des générations impactées par la réforme, elle fera des études plus ou moins longues. Cependant, dans la plupart des pays développés, ces réformes ont accompagné un mouvement général. De nombreux élèves prolongeaient déjà spontanément leur scolarité au-delà du nouvel âge légal de scolarité obligatoire : ce sont des *always takers* au sens de Angrist, Imbens, et Rubin (1996). Toujours en utilisant ce registre, les *compliers* sont ceux qui prolongent leur scolarité du fait de la réforme, mais qui se seraient arrêté avant sinon (et les *never takers* ceux qui n'en tiennent pas compte et s'arrêtent avant). L'estimateur obtenu (le LATE) sera une moyenne des rendements de l'éducation pour ceux ayant a priori une scolarité plus faible.

Si les rendements d'une année d'étude supplémentaire sont décroissants (plus on est déjà diplômé, et moins une année supplémentaire représente de surcroît salarial), les rendements de l'éducation de ces compliers (ce que mesure le LATE obtenu) sera donc une mesure plus élevée que celle des rendements de l'éducation, en moyenne sur l'ensemble de la population. Pour tester cette hypothèse, Oreopoulos utilise le fait que les proportions

sont très différentes selon les pays. L'augmentation de la scolarité obligatoire de 14 à 15 ans aux États-Unis n'a eu d'impact que sur 10 % des générations concernées (c'est-à-dire qu'au moment où la réforme est passée, 90 % des élèves poursuivaient déjà leur scolarité au-delà de 15 ans). En revanche, lors de la dernière réforme similaire intervenue au Royaume-Uni (en 1947), 60 % des adolescents sortaient du système scolaire avant 15 ans. Dans ce cas, la population des "compliers" est donc sur une population beaucoup plus nombreuse. Oreopoulos obtient néanmoins pour le Royaume-Uni des résultats proches de ceux obtenus sur les États-Unis. Cela signifie donc que l'hypothèse de rendements de l'éducation décroissants est erronée... ou simplement qu'ils sont très différents entre les États-Unis et le Royaume-Uni.

3.3.2 L'estimateur dépend de l'instrument

Des estimateurs utilisant des instruments différents sont susceptibles de donner des résultats différents. Pour donner un sens aux résultats obtenus, il faut encore une fois tenter de caractériser la sous-population sur laquelle l'effet est identifié. La théorie économique peut permettre d'apporter des pistes d'interprétation. Par exemple, Rosenzweig et Wolpin (2000) proposent un modèle de choix d'éducation qui permet d'interpréter les résultats obtenus dans la littérature sur les rendements de l'éducation.

3.4 Instruments faibles

Revenons maintenant à une discussion plus classique sur la qualité des instruments. En général, la discussion porte sur la validité de l'instrument : il s'agit de justifier que l'instrument n'est pas corrélé avec les déterminants inobservés du revenu. Cette propriété est essentielle : elle assure que l'estimateur obtenu par variables instrumentales sera asymptotiquement sans biais. Il n'existe, cependant, aucun test statistique permettant de valider rigoureusement cette propriété. Rappelons que les tests dits de Sargan, ou J-test, ou de sur-identification qui sont utilisés lorsque l'on dispose de plus d'instruments que de variables endogènes ne testent que la cohérence globale des instruments.

Cependant, le premier critère (corrélacion de l'instrument avec la variable endogène) en doit pas être négligé, comme l'ont montré des études récentes. De fait, la recherche d'instruments dont l'exogénéité ne puisse être mise en cause les rend parfois faiblement corrélés avec le traitement. On parle dans ce cas d'instruments faibles. Dans ce cas, les propriétés des estimateurs obtenus ne sont pas satisfaisantes : les estimations risquent d'être très imprécises, voire plus biaisées que les estimateurs par mco "naifs". Par ailleurs, les tests classiques (intervalle de confiance, significativité) sont erronés.

Rappelons en effet que les propriétés classiques des estimateurs des doubles moindres carrés ne sont qu'asymptotiques. L'estimateur des doubles moindres carrés est biaisé à distance finie : plus précisément, il tend vers l'estimateur des moindres carrés ordinaires. En outre sa distribution peut s'éloigner très nettement de la distribution normale sur laquelle la plupart des tests statistiques sont fondés. En général, on considère que disposer de

grands échantillons permet d'obtenir des approximations satisfaisantes de ces distributions asymptotiques. Cependant, dans le cas d'instruments faibles, Bound, Jaeger, et Baker (1995) ont montré que les biais de distance finie étaient encore présents même sur de très grands échantillons.

3.4.1 Qu'est-ce qu'un instrument faible ?

Plaçons nous dans notre modèle classique dans lequel on s'intéresse à l'effet d'un traitement T sur le revenu Y . T est endogène, mais nous supposons que nous disposons d'un vecteur Z de K instruments tel que $E(Z'u) = 0$, lié au traitement endogène T par la relation (dite "régression de première étape").

$$\begin{cases} Y = T\delta + u \\ T = Z\Pi + v \end{cases} \quad (23)$$

Remarque : Pour simplifier on a supposé que toutes nos variables sont centrées, i.e. de moyenne nulle.

L'estimateur usuel dans ce cas est celui des doubles moindres carrés, obtenu en régressant le revenu non plus sur T mais sur sa projection \hat{T} sur le vecteur des instruments Z : $\hat{T} = Z\hat{\Pi}$.

Π correspond à la mesure de l'intensité de la corrélation entre l'instrument et la variable endogène. À la limite, si $\Pi = 0$ (aucune corrélation entre la variable endogène et l'instrument), on remplace T par 0 et il est facile de comprendre que l'on sera confronté à un problème... En général, on n'a jamais $\Pi = 0$, mais il peut être très proche de 0 et on parle alors d'instrument faible. En pratique, pour mesurer la force de l'instrument un indicateur sans dimension, on peut utiliser la "concentration", définie comme le rapport entre la part de la variable endogène expliquée par l'instrument et celle qui ne l'est pas :

$$\mu^2 = \Pi'Z'Z\Pi/\sigma_v^2 \quad (24)$$

3.4.2 Risque des instruments faibles : des estimateurs imprécis

Le premier problème des instruments faibles est que les estimateurs risquent d'être très imprécis. Pour s'en convaincre, on rappelle que l'estimateur des doubles moindres carrés s'exprime comme :

$$\hat{\delta}_{IV} = (\hat{T}'\hat{T})^{-1}\hat{T}'Y$$

La précision de cet estimateur est mesurée par sa variance. Dans un modèle homoscedastique, elle tend vers :

$$Avar(\hat{\delta}_{IV}) = \sigma_u^2(\hat{T}'\hat{T})^{-1} \quad (25)$$

$$= \sigma_u^2/Var(P_ZT) \quad (26)$$

L'intuition est simple : si T n'est que faiblement corrélé avec les instruments Z , la projection \hat{T} sera petite. À la limite, si Π tend vers 0, \hat{T} est nul et la variance de l'estimateur est infinie.

Il peut être intéressant de la comparer à la variance asymptotique de l'estimateur des moindres carrés ordinaires. Rappelons que celle-ci s'écrit :

$$Avar(\hat{\delta}_{MCO}) = \sigma_u^2 Var(T)^{-1}$$

Dans le cas d'un unique instrument ($K = 1$), on a :

$$\begin{aligned} Avar(\hat{\delta}_{IV}) &= \sigma_u^2 E(T'Z)E(Z'Z)^{-1}E(Z'T)^{-1} \\ &= \sigma_u^2 \frac{var(Z)}{cov(Z, T)^2} \\ &= \sigma_u^2 \frac{var(Z)var(T)}{cov(Z, T)^2 var(T)} \\ &= \sigma_u^2 \frac{1}{corr(Z, T)^2 var(T)} \\ &= \frac{Avar(\hat{\delta}_{MCO})}{corr(Z, T)^2} \end{aligned}$$

On voit qu'une faible corrélation entre le traitement et l'instrument peut faire "exploser" la variance¹⁷.

Cette imprécision est en elle-même un problème : les estimations obtenues par les doubles moindres carrés risquent d'être plus difficiles à exploiter. Cependant, comme on va le voir, le risque est beaucoup plus sérieux que cela. Lorsque les instruments sont faibles, l'estimateur peut être bien plus biaisé que l'estimateur des moindres carrés ordinaires.

3.4.3 Un remède pire que le mal ?

Deux cas doivent être distingués : le premier lorsque l'instrument n'est pas parfaitement valide (il est seulement "moins" corrélé aux endogènes que la variable explicative) ; le second, *a priori* plus favorable, lorsqu'il est parfaitement valide. Comme on le verra, même dans ce cas, l'estimateur obtenu en utilisant des instruments faibles peut conduire à des résultats fallacieux.

¹⁷Dans le cadre plus général d'instruments multiples, un peu de manipulation nous permet d'aboutir à l'expression : $Avar(\hat{\delta}_{IV}) = Avar(\hat{\delta}_{MCO})(1 + \frac{1}{\mu^2})$, et on aboutit aux mêmes conclusions (rappelons que μ^2 est une mesure de la force de l'instrument). En effet,

$$E(T'Z)E(Z'Z)^{-1}E(Z'T) = E(\Pi'Z'Z\Pi) = \mu^2 \sigma_v^2$$

si on est prêt à faire l'hypothèse $E(Z'v) = 0$.

De même $Var(T) = E(T'T) = \mu^2 \sigma_v^2 + \sigma_v^2$ et on a le résultat.

3.4.4 Cas 1 : l'instrument n'est pas parfaitement valide.

Par définition, un bon instrument n'est pas corrélé avec les déterminants inobservés du revenu. En pratique, il existe toujours un risque qu'une corrélation subsiste (rappelons qu'il n'est pas possible de tester la validité de ces hypothèses¹⁸). L'économètre non averti peut être tenté de négliger ce risque, dans la mesure où l'instrument sera *a priori* moins corrélé que la variable endogène, et où il peut donc espérer que le biais sera plus faible. En réalité, comme on va le voir, dans ce cas l'estimation par VI peut être au contraire bien plus biaisée que celui des moindres carrés ordinaires.

Pour s'en convaincre, prenons encore le cas d'un instrument unique. Dans ce cas, on a :

$$\begin{aligned} \text{plim}\hat{\delta}_{OLS} &= \delta + \frac{\text{cov}(T, u)}{\text{var}(T)} \\ \text{plim}\hat{\delta}_{IV} &= \delta + \frac{\text{cov}(Z, u)}{\text{cov}(Z, T)} \end{aligned}$$

Le biais relatif de l'estimateur à VI par rapport à l'estimateur des MCO s'écrit :

$$\frac{\text{plim}\hat{\delta}_{IV} - \delta}{\text{plim}\hat{\delta}_{OLS} - \delta} = \frac{\frac{\text{cov}(Z, u)}{\text{cov}(T, u)}}{\frac{\text{cov}(Z, T)}{\text{var}(T)}} \quad (27)$$

Cette expression fait apparaître deux termes :

- le dénominateur, $\frac{\text{cov}(Z, u)}{\text{cov}(T, u)}$, peut s'interpréter comme la "réduction" d'endogénéité apportée par l'instrument : le moins que l'on puisse attendre de Z est que ce terme soit inférieur à 1
- le numérateur : $\frac{\text{cov}(Z, T)}{\text{var}(T)}$ est la part de la variance de la variable endogène T "expliquée" par l'instrument. Si ce terme est très petit (ce qui est par définition le cas lorsque l'instrument est faible), il y a un risque que l'estimateur à variable instrumentale soit plus biaisé que l'estimateur des moindres carrés. On remarquera que ce rapport est estimé par la valeur du R^2 (ou, ce qui est identique, par la statistique de Fischer du test de nullité jointe de tous les paramètres) dans l'estimation dite "de première étape" de la variable endogène sur l'instrument.

3.4.5 Cas 2 : l'instrument est parfaitement valide

Ce cas apparaît théoriquement plus favorable : l'estimateur des variables instrumentales est asymptotiquement sans biais. Cependant, même sur des échantillons très grands, l'estimateur obtenu par un instrument faible continue de souffrir du biais de distance finie.

¹⁸Les tests de sur-identification supposent qu'ils existent de nombreux instruments. Tester la validité d'un instrument particulier suppose que les autres le sont, ce qui n'est que reporter le problème... En fait, ce test permet plutôt de vérifier la cohérence interne des différents instruments.

On peut montrer que le biais relatif de l'estimateur des doubles moindres carrés par rapport à celui de l'estimateur des moindres carrés ordinaire est inversement proportionnel à μ^2/K , où μ est le paramètre de concentration défini plus haut et K le nombre d'instruments.

De plus, même lorsque les termes d'erreur u et v sont tous deux normaux, l'estimateur des doubles moindres carrés n'aura une distribution "approximativement" normale que pour des valeurs élevées de ce paramètre de concentration (voir Stock, Wright, et Yogo (2002)). Cela signifie en particulier que les tests classiques (de significativité en particulier) ne sont pas utilisables dans ce cas, puisque la distribution de l'estimateur des moindres carrés n'est pas connue.

En effet, on a :

$$\hat{\delta}_{IV} = (T'P_Z Y)/(T'P_Z T) = \delta + (T'P_Z u)/(T'P_Z T) \quad (28)$$

Un peu de manipulation permet à Stock, Wright, et Yogo (2002) d'exprimer cette expression en fonction de variables ayant des distributions connues.

$$\begin{cases} T'P_Z u &= v'P_Z u + \Pi'Z'u \\ T'P_Z T &= \Pi'Z'Z\Pi + 2\Pi'Z'v + v'P_Z v \end{cases} \quad (29)$$

On définit $z_u = \Pi'Z'u/\sigma_u\sqrt{\Pi'Z'Z\Pi}$ et $z_v = \Pi'Z'v/\sigma_v\sqrt{\Pi'Z'Z\Pi}$, qui suivent toutes deux une loi normale centrée réduite du fait de la normalité de u et v . De même, on définit $S_{uv} = (v'P_Z u)/(\sigma_u\sigma_v)$ et $S_{vv} = (v'P_Z v)/\sigma_v^2$ qui sont des formes quadratiques de loi normale. Leurs distributions sont en particulier indépendantes de la taille T de l'échantillon. Rappelons enfin que $\Pi'Z'Z\Pi$ est proportionnelle au carré de la concentration μ^2 défini en (24).

Avec ces définitions, on peut transformer (28) de la façon suivante :

$$\mu(\hat{\delta}_{IV} - \delta) = \frac{\sigma_u}{\sigma_v} \frac{z_u + S_{uv}/\mu}{1 + 2z_v/\mu + S_{vv}/\mu^2}$$

Cette formulation montre que la distribution de $\hat{\delta}_{IV}$ ne dépend de la taille qu'à travers le paramètre de concentration μ . Si μ est élevé, le terme dominant sera z_u qui suit une distribution normale. Une bonne approximation de la distribution de $\hat{\delta}_{IV}$ pourra être une loi normale $N(0, \frac{\sigma_u^2}{\sigma_v^2})$. En revanche, si μ est faible, la distribution pourra être très différente d'une loi normale.

Le fait que l'estimateur ne soit biaisé qu'à "distance finie" peut conduire à négliger ce risque, dès lors qu'on dispose d'échantillons de taille suffisante. En fait, si les instruments sont vraiment faibles, la taille des échantillons disponibles ne permet pas de régler le problème. Bound, Jeager, et Baker (1995) montrent ainsi que les résultats d'Angrist et Krueger (1991) sont probablement biaisés du fait de l'utilisation d'instruments faibles... alors qu'ils utilisent des données issus de recensement, avec un échantillon de plus de 300 000 individus !

3.4.6 Pire qu'un instrument faible : plusieurs instruments faibles

Le problème est d'autant plus sérieux que le nombre d'instruments est élevé. Il s'agit encore ici d'un résultat plus général que celui des instruments faibles. À distance finie, les estimations obtenues avec plusieurs instruments sont plus précises mais aussi plus biaisées vers l'estimateur des moindres carrés ordinaires¹⁹.

Lorsque l'on ne dispose que d'un seul instrument, on peut être tenté d'en créer artificiellement de nouveaux à partir des exogènes du modèle. Si l'instrument est faible, il s'agit d'une mauvaise idée. Là encore, les biais de distance finie perdurent même lorsque l'on utilise de très grands échantillons. Bound, Jaeger, et Baker (1995) le montrent en répliquant l'étude de Angrist et Krueger (1991). Rappelons que ces derniers estiment les rendements de l'éducation en instrumentant le nombre d'années d'études par le trimestre de naissance. Angrist et Krueger (1991) utilisent, pour obtenir plus de variabilité, le trimestre de naissance croisé avec des indicatrices temporelles et de régions. Ils obtiennent ainsi un ensemble de 170 instruments, mais comme le montrent Bound, Jaeger, et Baker (1995), ces instruments n'apportent que très peu d'information. L'estimateur des doubles moindres carrés a tendance à se rapprocher de l'estimateur des moindres carrés ordinaires au fur et à mesure que l'on ajoute des variables instrumentales.

Bound, Jaeger, et Baker (1995) répliquent cette étude en utilisant non plus le véritable trimestre de naissance, mais un trimestre affecté aléatoirement. Leurs simulations aboutissent à des résultats (qui n'ont *a priori* aucun sens) étonnamment proches de ceux d'Angrist et Krueger. De fait, les estimateurs sont biaisés vers les moindres carrés ordinaires : il n'est dès lors pas surprenant que les résultats soient identiques... Une autre leçon de ces résultats est qu'il est inutile voire dangereux de vouloir multiplier les instruments (ce qu'on peut être tenté de faire pour produire des tests de sur-identification par exemple), s'ils sont faibles.

3.5 Pratique

Ces résultats montrent l'importance pour juger de la qualité des résultats obtenus par variables instrumentales de toujours faire figurer les résultats détaillés de l'équation de première étape.

On peut montrer que le paramètre de concentration défini plus haut est lié au test de $\Pi = 0$, soit celui de la nullité jointe de tous les paramètres dans la régression de première étape. Plus précisément \hat{F} , la statistique de Fisher de ce test converge vers : $E(F) \simeq \mu^2/K + 1$ ²⁰. \hat{F} nous permet donc d'avoir une approximation de la mesure de la

¹⁹on pourra en voir un exemple dans le chapitre 7 de Davidson et MacKinnon (1993).

²⁰On a :

$$KF = \frac{\hat{\Pi}'Z'Z\hat{\Pi}}{\sigma_v^2} \frac{\sigma_v^2}{\hat{\sigma}_v^2}$$

Sous l'hypothèse de normalité de v , le premier terme suit une loi du χ^2 à K degrés de liberté, non centré d'espérance $\hat{\Pi}'Z'Z\hat{\Pi}/\sigma_v^2 + K = \mu^2 + K$. Le dernier terme tend asymptotiquement vers 1. On rappelle que

concentration (et donc de la force de l'instrument).

Un instrument sera dit faible si la valeur de μ^2/K (donc de \hat{F}) est telle que l'inférence basée sur les distributions normales usuelles n'est pas valide. Les seuils privilégiés dépendront des priorités de l'économètre. On peut souhaiter que μ^2/K soit suffisamment élevé pour que le biais relatif de l'estimateur des doubles moindres carrés par rapport à celui des MCO soit au plus de 10 % des cas (par exemple); ou encore que μ^2/K soit tel qu'un test de significativité à 5 % ne conduise à rejeter à tort la nullité que dans plus de 10 % (par exemple). On trouvera dans Stock, Wright, et Yogo (2002) les différentes valeurs correspondant à ces deux critères. Dans le cas d'un seul instrument, une valeur de la statistique de Fischer de 10 est nécessaire.

l'espérance d'une variable X qui suit une loi du χ^2 non centrée d'espérance ν à p degrés de liberté est $E(X) = p + \nu$, d'où le résultat.

3.6 En résumé

Pour mettre en œuvre une méthode instrumentale pour évaluer l'impact d'une mesure, il faut donc :

1. Disposer d'un instrument, c'est-à-dire une variable qui explique le fait que certains aient une plus grande chance d'être affectés par cette mesure et d'autres non, mais qui n'ont pas d'impact direct sur les résultats de cette mesure. Une connaissance approfondie de la législation et de ses variations peut par exemple permettre d'obtenir de telles variables
2. S'assurer de sa validité et de sa force. La première n'est pas testable statistiquement, mais doit demander une attention particulière. La seconde (corrélation entre l'instrument et la variable de traitement) demande de s'intéresser à la régression de première étape du traitement sur l'instrument. Comme montré dans la section 3.4, des valeurs importantes de la statistique de Fisher sont nécessaires pour éviter les problèmes d'instruments faibles.
3. Estimer : en pratique, une méthode de doubles moindres carrés peut être appliquée. Mieux vaut utiliser une procédure préprogrammée. Attention, les propriétés présentées ici n'ont été établies que pour des estimations linéaires (en général avec des revenus continus). Il est incorrect d'utiliser une procédure en deux étapes lorsque la variable d'intérêt Y binaire est modélisée par une fonction paramétrique (logit ou probit par exemple).
4. Interpréter les résultats obtenus, au regard en particulier du fait qu'il s'agit d'une estimation locale. Elle mesure les effets de la mesure en moyenne pour une population particulière, celle qui réagit à l'instrument (voir section 3.3). Il n'est pas évident que cette estimation locale puisse être étendue à l'ensemble de la population, mais il n'existe pas de moyens systématiques de répondre à cette question. Il est cependant possible de déterminer la proportion de cette sous-population de "compliers".

4 Régression sur discontinuités

4.1 Introduction

4.1.1 Principe

Les premières utilisations de techniques correspondant à une procédure de régression sur discontinuités (*Regression Discontinuity Design*) remontent aux travaux précurseurs de Thistlewaite et Campbell (1960), en psychologie. Ces derniers étudiaient l’impact de la réussite scolaire sur la carrière future en comparant le devenir d’étudiants ayant raté de peu un examen et de ceux l’ayant juste réussi. Ce n’est cependant qu’à la fin des années 90 que cette technique a été utilisée dans des études économiques, suite à des études influentes, comme celle d’Angrist et Krueger (1999) sur l’impact de la taille des classes sur la réussite scolaire, de Black (1999) sur l’impact de la carte scolaire sur les prix des logements, et à l’article d’Hahn, Todd, et Van der Klaauw (2001) qui précise rigoureusement les conditions d’identification nécessaires à son utilisation. On trouvera une présentation détaillée de ces méthodes dans Imbens et Lemieux (2008), dont s’inspire en grande part cette section.

Les raisons de ce succès tiennent à sa simplicité : la méthode repose sur l’existence d’une variable de sélection (qu’on notera S) qui a un impact discontinu sur la probabilité d’être traité. Plus précisément, il s’agit d’exploiter le fait que de nombreux dispositifs institutionnels présentent des règles d’affectation “à seuils” : allocation sous condition de ressources par exemple, contrainte d’âge pour certains dispositifs, carte scolaire (dans certaines rues, des enfants dépendent d’écoles plus ou moins “cotées”)... L’intuition est qu’autour de ces seuils, les personnes sont “presque” identiques, et pourtant seules certaines d’entre elles bénéficieront de ce dispositif. Il est donc possible d’identifier l’effet causal du traitement en comparant ces deux types de personnes.

De manière plus formalisée, on s’intéresse encore à l’effet d’un traitement T_i sur une variable de revenu y_i :

$$y_i = \alpha + \Delta_i T_i + u_i \quad (30)$$

On peut remarquer que $\Delta_i = Y_{i1} - Y_{i0}$ représente l’effet du traitement pour un individu i , $\alpha = E(Y_0)$ la moyenne des revenus potentiels sans traitement, et les résidus $u_i = Y_{i0} - E(Y_0)$.

En pratique, on doit distinguer deux cas :

4.1.2 *Sharp design*

Le traitement T dépend de manière déterministe de la variable de sélection S

$$T_i = T(S_i) = 1(S_i > \underline{S}) \quad (31)$$

Exemple :

C’est par exemple le cas de dispositifs dépendant de l’âge. Lalive (2008) s’intéresse ainsi aux indemnités chômage qui sont plus généreuses en Autriche lorsque la perte d’emploi a lieu après cinquante ans. Il étudie l’effet de ces indemnités sur le retour à l’emploi en

tenant compte de cette discontinuité. Celle-ci présente de fait une discontinuité marquée après 50 ans.

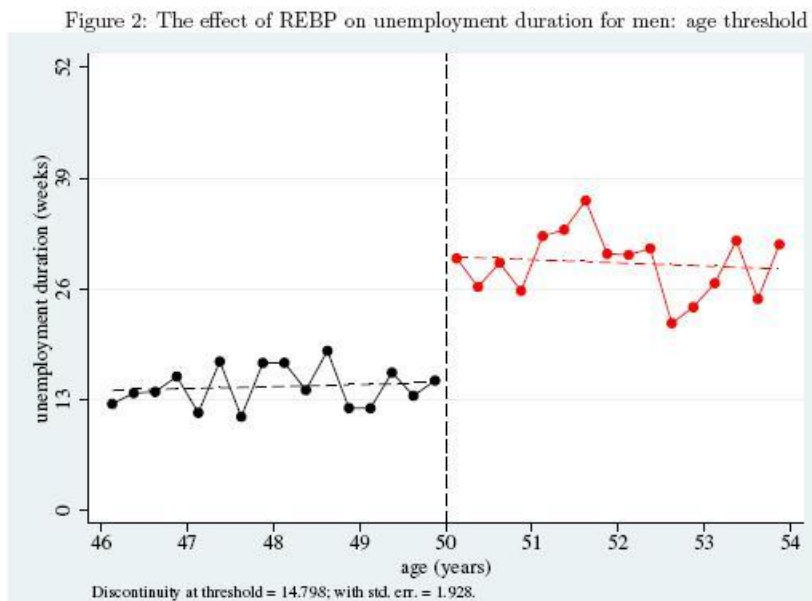


FIG. 5: Lalive (2008), durée du chômage selon l'âge

4.1.3 Fuzzy design

Dans un *fuzzy design*, la variable de sélection S affecte bien la probabilité d'être traité, mais l'affectation n'est plus parfaite. Cette fois, c'est donc seulement $P(T = 1|S)$ qui présente une discontinuité en \underline{S} . Deux exemples sont devenus classiques :

- Van der Klauw (2002) s'intéresse aux montants des bourses offertes aux étudiants par une grande université américaine. Dans ce système universitaire, les universités se font concurrence pour recruter les meilleurs étudiants, et le montant des bourses qu'elles leur allouent constitue l'un des instruments de cette compétition. Van der Klauw (2002) s'intéresse à l'efficacité de cette bourse pour retenir les étudiants. L'estimation directe est difficile, les meilleurs étudiants étant susceptibles d'avoir également des offres intéressantes dans les universités concurrentes.

Pour déterminer l'impact causal de ces bourses, Van der Klauw (2002) utilise le fait qu'en pratique, l'université classe les étudiants dans quatre groupes selon leur score à un examen. Le montant des bourses proposées à un étudiant diffère selon le groupe auquel il appartient. Cela signifie que des étudiants dont le niveau scolaire est très proche, mais qui se situent respectivement en dessous et au-dessus des seuils utilisés pour construire ces groupes, peuvent se voir proposer des montants différents. L'allocation n'est cependant pas parfaite, puisque d'autres variables (revenus des parents, appartenance à une minorité, lettres de recommandation...) interviennent également

pour déterminer le montant des bourses. La figure 6 qui présente le montant réel des bourses en fonction du niveau scolaire, présente de fait une très grande dispersion. La courbe d'interpolation polynomiale moyenne présente cependant des sauts à chaque seuil.

Fig. 3: Financial aid offers — Filers

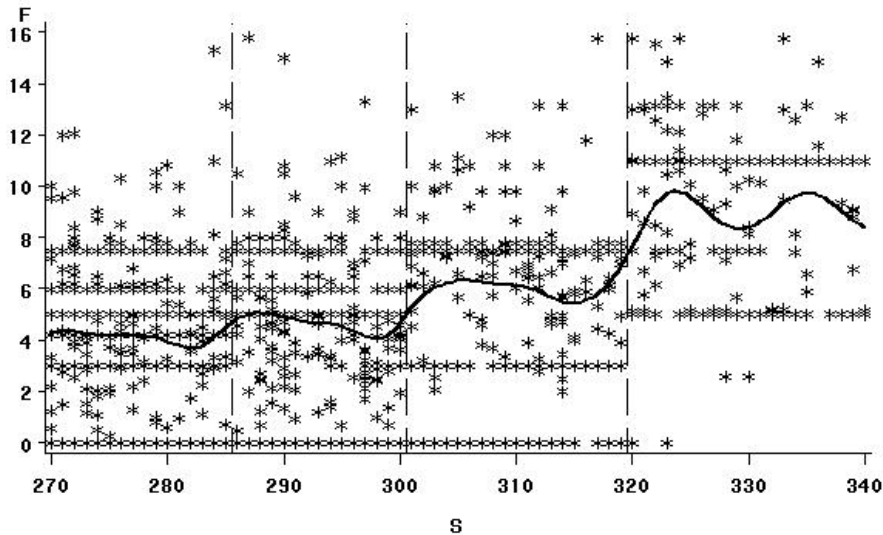


FIG. 6: Van der Klaauw (2002) : montant des bourses offertes selon le niveau scolaire

- Angrist et Lavy (1999) utilisent les limites existantes sur le nombre d'élèves par classe pour étudier l'effet de la taille des classes sur la réussite scolaire : dans une école, les classes sont "remplies" jusqu'à un certain seuil à partir duquel une nouvelle classe est créée, ce qui diminue brusquement le nombre moyen d'élèves par classe dans cette école. En Israël, la règle des *Maimonides* prévoit ainsi que le nombre d'élèves par classe ne doit pas dépasser quarante.

4.1.4 Identification dans le cas d'un *sharp design*

Dans quelle mesure pouvons nous identifier un effet causal dans ces modèles ? Comme on va le voir, les hypothèses sont assez générales, surtout dans le cadre d'un *sharp design*.

Notons que dans le cadre défini plus haut, on peut schématiser simplement le revenu réellement observé (Imbens et Lemieux 2008) (figure 8).

Le revenu réel doit effectivement présenter une discontinuité au point de sélection. De fait, on passe "brutalement" du revenu sans traitement Y_0 au revenu avec traitement Y_1 . Cependant, la discontinuité observée ne permet de mesurer l'effet du traitement que si les revenus potentiels ne présentent pas eux-mêmes de discontinuités à ce point (ce serait le cas si un autre déterminant du revenu dépendait de manière discontinue de la variable

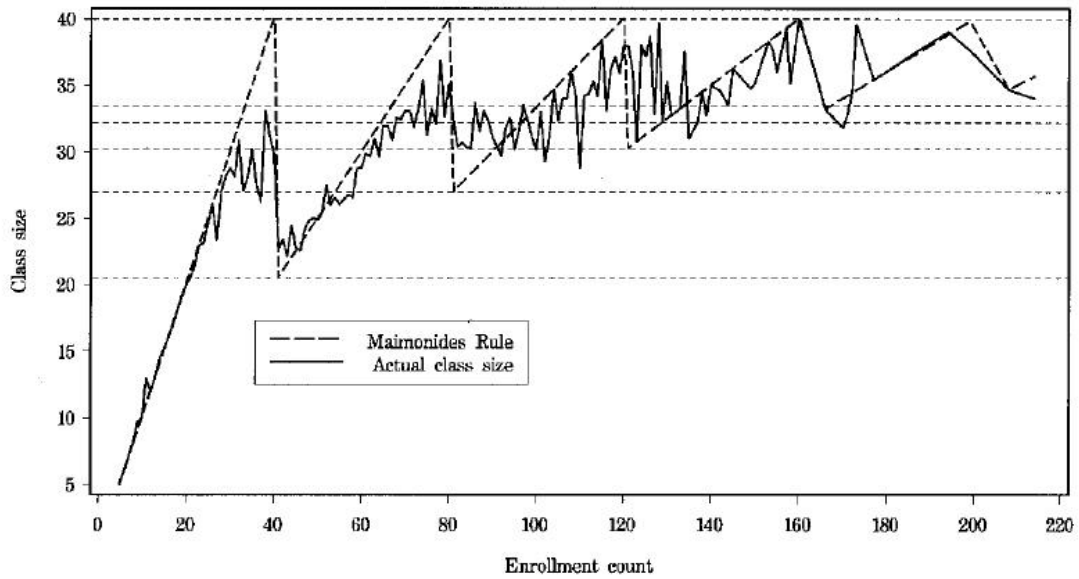


FIG. 7: Angrist et Lavy (1999), nombre d'élèves par classe selon le nombre d'élèves dans l'école

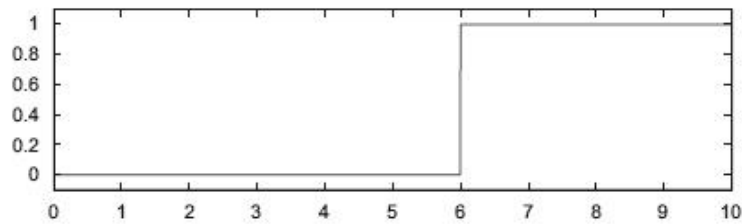


Fig. 1. Assignment probabilities (SRD).

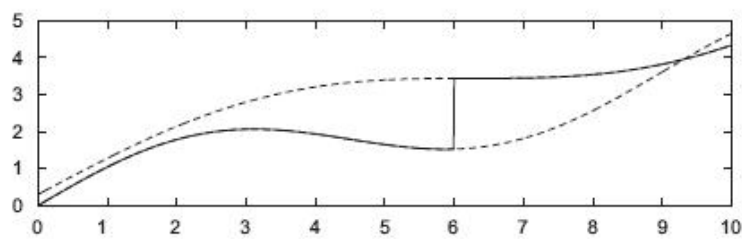


Fig. 2. Potential and observed outcome regression functions.

FIG. 8: Revenus potentiels avec et sans traitement, cas d'un *sharp design*, Imbens et Lemieux (2008)

de sélection). Dit autrement, il est nécessaire que les personnes juste en dessous ou juste au-dessus du seuil soient vraiment comparables.

L'hypothèse identifiante est donc : les revenus potentiels Y_1 et Y_0 sont continus autour

du point de discontinuité de T . Cela signifie que la composante inobservée du revenu $E(u_i/S)$ est continue en \underline{S} , ainsi que l'effet moyen du traitement $E(\Delta_i/S)$.

L'effet du traitement au point de discontinuité correspond alors simplement à :

$$E(\Delta_i|S = \underline{S}) = \lim_{S \searrow \underline{S}^+} E(Y|S) - \lim_{S \nearrow \underline{S}^-} E(Y|S) \quad (32)$$

car

$$\lim_{S \searrow \underline{S}^+} E(u_i|S) = \lim_{S \nearrow \underline{S}^-} E(u_i|S) \quad (33)$$

et

$$\lim_{S \searrow \underline{S}^+} E(\Delta_i T_i|S) = E(\Delta_i|S = \underline{S}) \quad (34)$$

Il faut bien noter qu'on n'estime qu'un effet local du traitement, au point de discontinuité. Si le traitement n'est pas constant dans la population, l'interprétation de cette estimation sera limitée.

4.1.5 Identification dans le cas d'un *fuzzy design*

Cette hypothèse est-elle suffisante dans le cas d'un *fuzzy design*? Oui si l'effet du traitement est constant au voisinage de \underline{S} .

D'après

$$\lim_{S \searrow \underline{S}^+} E(Y|S) - \lim_{S \nearrow \underline{S}^-} E(Y|S) = \lim_{S \searrow \underline{S}^+} E(\Delta_i T_i|S) - \lim_{S \nearrow \underline{S}^-} E(\Delta_i T_i|S) \quad (35)$$

$$+ \lim_{S \searrow \underline{S}^+} E(u_i|S) - \lim_{S \nearrow \underline{S}^-} E(u_i|S) \quad (36)$$

Si l'effet du traitement est constant au voisinage du point de discontinuité, il est donné par :

$$\Delta = \frac{\lim_{S \searrow \underline{S}^+} E(Y|S) - \lim_{S \nearrow \underline{S}^-} E(Y|S)}{\lim_{S \searrow \underline{S}^+} E(T|S) - \lim_{S \nearrow \underline{S}^-} E(T|S)} \quad (37)$$

Cependant, si l'effet du traitement varie avec les individus, il faut faire une hypothèse supplémentaire.

La première peut être l'indépendance "locale" du traitement (Hahn, Todd, et Van der Klaauw 2001). Conditionnellement à S , T est indépendant de Δ_i autour de \underline{S} :

$$T_i \perp \Delta_i | S_i \quad (38)$$

Le ratio plus haut permet d'estimer l'effet local du traitement

$$E(\Delta_i|S = \underline{S})$$

Si on n'est pas prêt à faire cette hypothèse, Hahn, Todd, et Van der Klaauw (2001) montrent qu'une hypothèse moins restrictive suffit. Il s'agit d'une hypothèse de monotonie : $T_i(S)$ est non décroissante en S , pour tout individu i , autour du point de discontinuité.

Celui-ci est directement l’analogie de celui démontré par Angrist, Imbens, et Rubin (1996) pour les variables instrumentales (voir section 3). Elle signifie qu’il n’existe pas de *defiers*, i.e. de personnes qui choisissent le traitement si elles sont en dessous du seuil, mais pas le traitement si elles sont au-dessus. En revanche, on n’exclut pas l’existence de personnes qui sont toujours, ou jamais traitées (*never taker*, ou *always taker* dans les termes d’Angrist, Imbens, et Rubin (1996)).

Qu’estime alors l’estimateur ci-dessus ? Ici aussi, il s’agit d’un LATE, i.e. un effet local :

$$\lim_{e \searrow 0} E(\Delta_i | T_i(\underline{S} + e) - T_i(\underline{S} - e) = 1) \quad (39)$$

c’est-à-dire l’effet du traitement pour les *compliers*, ceux pour qui le fait d’être en dessous ou au-dessus du seuil fait une différence dans l’acceptation du traitement.

4.1.6 Manipulation du seuil

Attention, pour que l’hypothèse soit valide, il faut être sûr que les individus ne peuvent pas “manipuler” la valeur du seuil, de manière à disposer ou non du traitement. Sinon, l’hypothèse de continuité a peu de chances d’être respectée (les individus ne seront plus comparables au niveau du seuil).

4.1.7 “Tests” graphiques

L’identification de l’effet repose sur deux hypothèses essentielles :

1. existence d’une variable de sélection pour laquelle la probabilité de traitement est discontinue
2. continuité des autres composantes du revenu en ce point.

Même si elle n’a pas statut de validation, une analyse graphique peut permettre de vérifier que la stratégie est bien adaptée. Plus exactement, on sera amené à vérifier plusieurs points :

- la variable de traitement présente bien une discontinuité au point prévu (si ce n’est pas le cas - par exemple parce que le dispositif institutionnel “officiel” n’est finalement pas appliqué en pratique - il est inutile de poursuivre)
- la variable de revenu présente une discontinuité au même point. Il est également plus rassurant qu’elle ne présente pas d’autres sauts d’ampleur comparable en-dehors de ceux correspondant aux discontinuités de la variable de traitement. Dans le cas contraire, s’il n’existe pas de justification plausible, on peut craindre qu’une discontinuité du revenu ne résulte pas uniquement de la discontinuité du traitement. L’estimateur fournira donc une mesure biaisée de l’effet causal.
- il est également judicieux de vérifier que les autres déterminants éventuels du revenu ne présentent pas de discontinuité au point \underline{S} .

4.2 Estimation

L'esprit de la méthode de régression sur discontinuités requiert d'estimer des valeurs *locales* au point de discontinuité : cela signifie en pratique de se concentrer sur les observations dont les valeurs de la variable de sélection sont proches de ce point de discontinuité. Nous présentons dans une première section les méthodes semi-paramétriques qui permettent une estimation rigoureuse. Parfois cependant, du fait d'un trop faible nombre d'observations autour de ce point de discontinuité, certaines études utilisent des points plus éloignés, voire l'ensemble de l'échantillon. Pour tenir compte du fait que les composantes inobservées du revenu peuvent cependant varier en fonction de la variable de sélection (ce qui n'est pas un problème localement, mais le devient dès lors qu'on utilise des points plus éloignés dans l'échantillon), certains choisissent alors une forme paramétrique de cette dépendance, que nous présentons dans un deuxième temps.

4.2.1 Estimation semi-paramétrique

On cherche à estimer les limites telles que $\lim_{S \searrow \underline{S}^+} E(Y/S)$. Pour cela, on peut utiliser des estimateurs à noyau (kernel estimator). En pratique, on pourra estimer $\lim_{S \searrow \underline{S}^+} E(Y/S)$ par :

$$\frac{\sum_i 1_{(S_i > \underline{S})} k_h(S_i - \underline{S}) Y_i}{\sum_i 1_{(S_i > \underline{S})} k_h(S_i - \underline{S})} \quad (40)$$

avec k_h une fonction à noyau (on trouvera une présentation plus détaillée dans la section 5) dont h est la taille de la fenêtre.

Il faut noter qu'en pratique, cette formule peut se révéler très simple. En utilisant une fonction à noyau uniforme : $k_h(S) = 1_{|S| < h}$, il s'agit simplement de faire la moyenne des points situés à moins de h du point de discontinuité. Nous discutons plus bas du choix éventuel de la taille de la fenêtre.

Néanmoins, comme montré par Porter (2002), ces estimateurs de la valeur du revenu ne convergent pas rapidement autour de points où cette variable est discontinue. Hahn, Todd, et Van der Klaauw (2001) proposent donc des estimateurs semi-paramétriques plus efficaces : plus précisément, ils corrigent de l'éventuelle dépendance de la variable d'intérêt en S par une fonction linéaire ou polynômiale.

Estimation semi-paramétrique dans un *sharp design* Plus précisément, pour estimer la valeur du revenu de la limite à droite au point de discontinuité $\lim_{S \searrow \underline{S}^+} E(Y|S)$, on cherche les paramètres (a_-, b_1, \dots, b_p) tels que :

$$\min_{(a_-, b_1, \dots, b_p)} \sum_i 1_{(S_i > \underline{S})} k_h(S_i - \underline{S}) (Y_i - a_- - b_1(S_i - \underline{S}) - \dots - b_p(S_i - \underline{S})^p)^2$$

Un estimateur de $\lim_{S \searrow \underline{S}^+} E(Y|S)$ sera alors :

$\hat{\mu}(\underline{S}) = \hat{a}_- + \hat{b}_1(\underline{S} - \underline{S}) + \dots = \hat{a}_-$. On peut obtenir de même un estimateur de la limite à

gauche $\lim_{S \nearrow \underline{S}^-} E(Y|S)$. L'effet causal du traitement dans le cas d'un *sharp design* est donc estimé par

$$\hat{\delta} = \hat{a}_+ - \hat{a}_- \quad (41)$$

Notons encore une fois que cette technique se met très facilement en pratique : dans le cas le plus simple (régression linéaire, noyau uniforme), il s'agit simplement de mettre en œuvre une régression linéaire en se restreignant aux observations à voisinage immédiat du point de discontinuité $[\underline{S} - h, \underline{S} + h]$. En toute rigueur, on fixe une taille de la fenêtre "optimale" (voir plus bas) pour chacun des termes. Par souci de simplicité, on peut choisir de fixer une taille de la fenêtre identique pour les deux estimateurs, et d'estimer simultanément les paramètres :

$$\min_{(\Delta, b_1, \dots)} \sum_i (Y_i - a - \Delta 1_{(S_i > \underline{S})} - b_1(S_i - \underline{S}) - b'_1(S_i - \underline{S}) 1_{(S_i > \underline{S})} - \dots)^2 \quad (42)$$

Estimation semi-paramétrique dans un *fuzzy design* Le principe est le même dans ce cas, mais il faut en plus estimer $\lim_S E(T|S)$ à gauche et à droite du point de discontinuité. L'estimateur de l'effet causal du traitement sera alors :

$$\frac{\hat{a}_{y+} - \hat{a}_{y-}}{\hat{a}_{t+} - \hat{a}_{t-}} \quad (43)$$

avec \hat{a}_{y+} correspondant au programme de minimisation

$$\min_{(a_+, b_1, \dots, b_p)} \sum_i 1(S_i > \underline{S}) k_h(S_i - \underline{S}) (Y_i - a_+ - b_1(S_i - \underline{S}) - \dots - b_p(S_i - \underline{S})^p)^2 \quad (44)$$

et de manière équivalente pour les trois autres termes \hat{a}_{y-} , \hat{a}_{t+} , \hat{a}_{t-} .

Remarque : Ici encore, on peut choisir des fenêtres identiques pour les quatre minimisations. En pratique, cela signifie qu'avec un noyau uniforme et une dépendance linéaire dans la variable de sélection (ce qui est une hypothèse tout à fait plausible *localement*), l'estimateur correspondra exactement à celui obtenu par une procédure de doubles moindres carrés où on instrumente la variable de traitement par le seuil.

Choix de la fenêtre Comment choisir la taille de la fenêtre h pour ces estimations ? Celle-ci doit répondre au classique arbitrage biais/variance. Pour estimer la valeur de la variable d'intérêt au point de discontinuité, on privilégie une petite fenêtre ; mais trop restreindre celle-ci risque de faire reposer l'estimation sur de très petits échantillons.

Dans la littérature, le choix est souvent fait de manière *ad hoc*, quitte à vérifier la sensibilité des résultats à différents choix de fenêtre.

Une méthode plus rigoureuse est d'utiliser une procédure de validation croisée (*cross validation*), déjà présentée dans la section 5. Rappelons qu'il s'agit de choisir une fenêtre qui permette d'obtenir des estimations correctes des valeurs obtenues. On va donc comparer les revenus estimés et observés sur un ensemble de points. Plus précisément, on

définit $\mu_-(s) = \lim_{z \nearrow s} E(Y_i | S_i = z)$. A h donné, un estimateur de $\mu_-(s)$ est fourni par $\hat{\mu}(s) = \hat{a}_-(s)$, avec \hat{a}_- qui correspond à :

$$\min_{(a_-, b_1, \dots, b_p)} \sum_i 1_{(S_i < s)} k_h(S_i - s) (Y_i - a_- - b_1(S_i - s) - \dots - b_p(S_i - s))^p \quad (45)$$

On va choisir h qui minimise

$$CV_Y(h) = \frac{1}{N} \sum_{i=1}^N 1_{(S_i < \underline{S})} (Y_i - \hat{\mu}(S_i))^2 \quad (46)$$

$$h_{CV}^{opt} = \operatorname{argmin} CV_Y(h) \quad (47)$$

Remarques :

- cette méthode peut être évidemment adaptée à un choix de fenêtre identique pour l'estimation à gauche et à droite
- dans le cas d'un *fuzzy design*, choisir une taille identique pour l'estimation du revenu et de traitement suppose de déterminer entre les deux tailles optimales. Pour minimiser les biais potentiels, on prendra le plus petit des deux.
- On peut choisir d'exclure les points trop loin du point de discontinuité. Par exemple, Ludwig et Miller (2007) proposent d'exclure une part de la distribution de la variable de sélection S , à droite et à gauche.

4.2.2 Estimation paramétrique

Les estimations précédentes supposent des estimations locales autour du point de discontinuité. Les données ne sont parfois pas assez détaillées pour disposer de suffisamment d'observations autour du point de discontinuité, par exemple parce que la variable de sélection n'est mesurée que par des valeurs discrètes.

De nombreux papiers empiriques pallient ces manques en utilisant des échantillons plus larges, quitte à utiliser des observations plus éloignées du point de discontinuité. On contrôle de la dépendance de la variable d'intérêt dans la variable de sélection par une spécification paramétrique, généralement polynômiale.

Là encore, les spécifications sont différentes selon qu'on est dans un *sharp design* ou dans un *fuzzy design*.

Estimation paramétrique dans un *sharp design* L'idée est ici de rendre compte de la dépendance du revenu dans la variable de sélection S . À partir des hypothèses précédentes, on peut écrire le modèle sous une forme linéaire classique :

$$y_i = m(S_i) + T_i \Delta + e_i$$

avec $E(e_i | T_i, S_i) = 0$, $\Delta = E(\Delta_i | S_i = \underline{S})$ et le traitement $T_i = 1_{(S_i > \underline{S})}$.

À partir de l'équation (30), on a $e_i = y_i - E(y_i | T_i, S_i)$ ce qui assure par construction

l'hypothèse d'identifiabilité $e_i \perp T_i$. La fonction $m(S)$ est définie comme $m(S_i) = \alpha + E(u_i|S_i) + T_i(E(\Delta_i|S_i) - E(\Delta_i|S_i = \underline{S}))$ ²¹. D'après l'hypothèse de continuité locale, $m(S)$ est continue en S . Si on spécifie correctement la fonction $m(S)$, on peut estimer l'effet moyen du traitement par une simple régression linéaire, puisque par construction $e_i \perp T_i$, ce qui assure l'exogénéité de T_i dans la régression MCO. En pratique, on approxime souvent $m(S)$ par des fonctions polynômiales.

Estimation paramétrique dans un *fuzzy design* De la même façon, toujours avec $\Delta = E(\Delta_i|\underline{S})$:

$$y_i = m(S_i) + E(T_i|S_i)\Delta + e_i$$

Avec $e_i = y_i - E(y_i|S_i)$, $m(S_i) = E(u_i|S_i) + E(T_i\Delta_i|S_i) - E(T_i|S_i)E(\Delta_i|\underline{S})$. On peut estimer l'effet voulu par une procédure en deux étapes :

1. $E(T_i|S_i) = f(S_i) + \gamma 1_{(S_i \geq \underline{S})} + v_i$ avec f fonction paramétrique de S continue en \underline{S} .
2. on utilise l'estimation de $E(T_i|S_i)$ obtenue en première étape dans l'équation

En fait, si $f(S)$ et $m(S)$ sont identiques, il s'agit exactement d'une procédure de doubles moindres carrés, avec une variable endogène T et un instrument $1_{(S_i \geq \underline{S})}$.

L'estimation paramétrique permet d'utiliser un échantillon plus large. Attention cependant, les estimations risquent d'être moins robustes, puisqu'elles varient avec la pertinence de la spécification retenue. Black, Galdo, et Smith (2007) comparent ainsi les résultats obtenus par une affectation aléatoire de personnes sans emploi à un dispositif d'aide à l'emploi, et ceux qu'on obtiendrait par une procédure de régression sur discontinuités. Ils montrent que la méthode de régression sur discontinuités fournit des résultats assez convaincants, au sens où elle permet de bien reproduire les résultats de l'expérience aléatoire, à *condition de se limiter aux valeurs proches* des points de discontinuités. Les estimations qui utilisent plus de points, en se reposant sur une spécification paramétrique pour corriger de la dépendance du revenu dans la variable de sélection, se révèlent en revanche peu robustes : conformément à l'intuition, elles sont sensibles à la forme retenue qui peut s'éloigner de la "vraie" fonction inconnue.

4.3 Discussion

4.3.1 Variable instrumentale ou RDD ?

Pour évaluer un dispositif, plusieurs méthodes peuvent être mobilisées. Comme on vient de le voir, les techniques de régressions sur discontinuités peuvent être utilisées lorsque dispositif présente des seuils dans les modalités d'attribution. On peut cependant s'interroger sur le fait d'utiliser d'autres méthodes. En particulier, on peut utiliser la variable de sélection comme un instrument : typiquement, on pourrait utiliser $1_{(S > \underline{S})}$. De fait, comme on l'a

²¹On la déduit directement de l'équation 30. On a par définition $m(S_i) = E(y_i|T_i, S_i) - T_i\Delta = \alpha + E(u_i|T_i, S_i) + E(\Delta_i T_i|S_i) - T_i E(\Delta_i|S_i = \underline{S})$. Comme $T = 1_{S > \underline{S}}$, toute l'information est apportée par S .

vu, les estimateurs utilisés dans un cadre de régression sur discontinuités ressemblent beaucoup “techniquement” à des estimateurs par variables instrumentales au sens où ils peuvent être calculés par des procédures de type doubles moindres carrés. Dans la littérature empirique, la frontière est assez floue entre les deux méthodes... on appelle aujourd’hui régression sur discontinuités des méthodes d’identification qui auraient été appelées instrument il y a quelques années. La différence essentielle tient à l’hypothèse sous laquelle on identifie un effet causal : même si l’estimateur est parfois le même, on n’est clairement pas dans un cadre où les variables instrumentales sont indiquées. On ne fait pas *a priori* l’hypothèse que les revenus potentiels sont indépendants de la variable de sélection. Dit autrement, les hypothèses nécessaires pour justifier que la procédure proposée mesure bien un effet causal sont moins fortes que pour un instrument. Cette méthode permet également d’insister sur le fait que l’estimateur ainsi obtenu est encore plus local. Dans le cadre d’un *fuzzy design*, on n’identifie l’effet que sur les personnes pour lesquelles la variable de sélection a un impact (ce sont les mêmes *compliers* que dans le cadre de variables instrumentales), mais également parce qu’on identifie l’effet seulement pour une valeur particulière des revenus (celles qui correspondent au point de discontinuité).

4.4 En résumé

Pour résumé, une procédure de régression sur discontinuités peut être utilisée pour évaluer une procédure ou une mesure dont le mode des désignations des bénéficiaires présente une discontinuité. Les principales étapes peuvent se décliner comme précédemment :

1. Disposer de données détaillées et d'échantillons suffisamment importants pour pouvoir se restreindre aux observations proches du point de discontinuité.
2. Vérifier la pertinence de la procédure. Cela peut se faire par exemple graphiquement : il faut que la probabilité de bénéficier de la mesure présente une discontinuité visible dans la variable de sélection ; il est également utile de vérifier que le revenu étudié est par ailleurs continu en dehors des points de discontinuités utilisés.
3. Estimer : en pratique, on privilégiera les estimations locales, c'est-à-dire en se restreignant aux observations dont les valeurs se trouvent autour du point de discontinuité (cf. 4.2.1). La manière la plus simple consiste à se limiter aux observations autour de la discontinuité, et de faire une régression linéaire sur ce sous-échantillon (simple régression linéaire pour un sharp design, doubles moindres carrés pour un fuzzy design). Idéalement, on peut choisir la taille du voisinage par une procédure de validation croisée (cf. 4.2.1). En pratique, l'immense majorité des études définit un voisinage de manière *ad hoc*, quitte à vérifier la sensibilité en le faisant varier. Si on élargit la taille de ce voisinage, il faut tenir compte de la dépendance selon le revenu, et les estimations risquent d'être moins robustes car sensibles à la spécification retenue (cf. 4.2.2).
4. Interpréter : il s'agit d'un effet local, obtenu pour les individus ayant une valeur particulière pour la variable de sélection. Savoir si ces résultats sont généralisables à l'ensemble de la population demande, comme souvent, une discussion au cas par cas.

5 Contrôler des observables : le matching

Les différentes méthodes vues jusqu'à présent demandent de disposer de caractéristiques particulières (instrument, groupe de contrôle exogène pour appliquer des différences de différences, processus ayant une discontinuité...). Elles ne peuvent être utilisées qu'au cas par cas. Lorsque les conditions de validité d'aucune de ces méthodes n'est applicable et qu'il existe de la sélection probable dans les dispositifs, on peut corriger, au moins en partie, cet effet de sélection en contrôlant des différences observables entre les bénéficiaires et les non-bénéficiaires. Les méthodes d'appariement (ou *matching*) permettent cela. Elles ne permettent d'identifier l'effet causal de la mesure qu'on veut évaluer que sous l'hypothèse très forte que les différences observables entre les bénéficiaires et les non-bénéficiaires captent l'ensemble des déterminants de la sélection des bénéficiaires.

Formellement, l'hypothèse identifiante pour évaluer cet effet causal est que conditionnellement à des variables observables X , l'affectation au traitement est indépendante des revenus escomptés :

$$Y_{0i}, Y_{1i} \perp T_i | X_i \quad (48)$$

Cette hypothèse est souvent appelée *CIA*, pour *Conditional Independence Assumption*. Supposons que nous observions deux individus avec des caractéristiques observables identiques, mais dont l'un est traité et l'autre pas. Cette hypothèse signifie que le fait qu'un individu soit traité et pas un autre n'est pas dû à des différences escomptées dans les revenus potentiels. Le revenu de la personne non traitée est donc un bon contrefactuel du revenu de la personne traitée, si elle n'avait pas été traitée. Et inversement... La comparaison des deux permet donc d'obtenir un estimateur non biaisé de l'effet du traitement (conditionnel à ces observables).

Si l'on ne souhaite évaluer que l'effet de la mesure sur les seuls bénéficiaires, on peut se contenter d'une hypothèse un peu moins forte. Il "suffit" en effet que :

$$Y_{0i} \perp T_i | X_i \quad (49)$$

C'est-à-dire que, conditionnellement à ces caractéristiques observables, le revenu des personnes qui n'ont pas bénéficié de la mesure fournit un bon contrefactuel du revenu potentiel des bénéficiaires, s'ils n'en avaient pas bénéficié.

Par exemple, pour évaluer l'impact du passage en Zone Franche Urbaine (ZFU) sur l'emploi et la création d'entreprises, Rathelot et Sillard (2009) se réfèrent aux critères utilisés pour classer ces zones de la politique de la ville. Ces zones étant par définition moins propices au développement économique, comparer directement l'évolution de l'emploi (par exemple) avec celle observée dans d'autres quartiers risquent de donner une évaluation biaisée. En fait, les ZFU ont été définies à partir de certaines caractéristiques socio-économiques (taux de chômage, potentiel fiscal de la commune...), mais également sur des considérations locales laissées à l'appréciation du préfet. Pour réduire ce biais, Rathelot et Sillard (2009) comparent l'évolution des zones avant leur passage en ZFU avec celles d'autres quartiers en difficulté, présentant des caractéristiques socio-économiques comparables. L'hypothèse

est que le fait que deux quartiers présentant des caractéristiques proches, mais dont l'un a bénéficié du statut de zones franches et pas l'autre, n'est pas lié à l'anticipation des effets escomptés de la zone.

Cette hypothèse identifiante est très forte. Elle signifie qu'en dehors des observables il n'existe pas d'autres caractéristiques qui influent à la fois sur les revenus potentiels et le choix du traitement. Le risque est donc grand qu'elle ne soit pas vérifiée. Néanmoins, il s'agit souvent de la seule possible, ce qui explique que les méthodes s'appuyant sur cette hypothèse sont très répandues.

Par ailleurs, ces méthodes ne peuvent être appliquées que dans certains cas : pour pouvoir comparer un individu traité avec un non traité identique... il faut qu'il existe ! Si pour certaines caractéristiques X les individus sont certains d'être traités, il ne sera pas possible de trouver un non-traité comparable. Dans l'exemple des Zones Franches Urbaines ci-dessus, c'est le fait que la création des zones ne soit pas déterminée exactement par les caractéristiques socio-démographiques qui permet de les utiliser comme variables de conditionnement. Si cela n'était pas le cas, il ne serait pas possible de trouver des quartiers comparables aux zones franches. Formellement, cela suppose que les méthodes d'appariement ne sont valables que sous l'hypothèse de l'existence d'un support commun, c'est-à-dire que pour toutes les valeurs des observables on puisse comparer des traités et des non-traités :

$$0 < P(T_i = 1|X_i) < 1 \quad (50)$$

En pratique, sous l'hypothèse d'indépendance conditionnelle, l'effet peut être simplement estimé par une régression linéaire, mais de nombreuses autres méthodes se sont développées sur les deux dernières décennies. Les plus courantes sont les méthodes d'appariement (ou *matching*), qui tentent d'apparier chaque bénéficiaire à un non-bénéficiaire ayant exactement les mêmes caractéristiques observables. Il n'existe cependant pas encore de cadre commun à l'ensemble de ces méthodes, qui se présente plus comme une collection de techniques que comme une théorie clairement édifée et unifiée.

5.1 Estimation par matching

5.1.1 Limites de la régression linéaire

La méthode la plus simple sous hypothèse d'indépendance conditionnelle du traitement est la régression linéaire. Supposons par exemple que la distribution du revenu potentiel sans traitement conditionnelle aux caractéristiques observables X soit linéaire :

$$E(Y_{i0}|X_i) = \alpha + \beta X_i \quad (51)$$

Si on suppose également que l'effet de la mesure $Y_{i1} - Y_{i0}$ est constant (noté Δ), le revenu observé s'écrit :

$$Y_i = \alpha + \Delta T_i + \beta X_i + \epsilon_i \quad (52)$$

avec $\epsilon_i = Y_{i0} - E(Y_{i0}|X_i)$. On peut estimer l'effet du traitement Δ par une régression des moindres carrés ordinaires, qui "contrôle" des observables. Notons aussi qu'il est possible de

relâcher l'hypothèse d'un effet constant de la mesure. On peut supposer qu'il dépend des caractéristiques observables en ajoutant au modèle des termes croisés entre le traitement et les observables. Le calcul de la précision de ces estimateurs est standard.

La régression linéaire présente des avantages indéniables. Elle a par ailleurs un fondement théorique clair et son application pratique est bien balisée, tant pour l'estimation que pour l'inférence statistique. Même si la distribution conditionnelle des revenus n'est pas exactement une fonction linéaire des observables, la régression linéaire fournit la meilleure approximation *linéaire*.

Limites La simplicité du modèle linéaire constitue également sa limite. Si la distribution conditionnelle s'éloigne trop d'une fonction linéaire, même la meilleure approximation linéaire restera peu adaptée et les estimateurs risquent d'être peu robustes.

En fait, le problème est particulièrement sensible lorsque les deux échantillons (traitement et contrôle) sont très différents du point de vue des variables de contrôle, car c'est alors que la linéarité de l'impact de ces variables sur le résultat devient une hypothèse forte.

Pour mieux comprendre cela, rappelons que notre problème est d'estimer un contre-factuel du revenu sans traitement pour les bénéficiaires de la mesure. Si l'hypothèse d'indépendance conditionnelle est correcte, la meilleure estimation linéaire est fournie par le revenu moyen du groupe de contrôle, corrigé des différences de composition entre les deux échantillons :

$$\hat{E}(Y_{i0}|T_i = 1) = \bar{Y}_0 + (\bar{X}_1 - \bar{X}_0)\hat{\beta} \quad (53)$$

où \bar{Y}_t et \bar{X}_t sont les moyennes empiriques respectivement de Y et X dans chacun des deux échantillons.

L'estimateur de l'effet du traitement correspond simplement à :

$$\hat{\Delta} = \bar{Y}_1 - \bar{Y}_0 - (\bar{X}_1 - \bar{X}_0)\hat{\beta} \quad (54)$$

Si la différence $\bar{X}_1 - \bar{X}_0$ entre les deux échantillons est importante, la correction le sera aussi, et sera très sensible à des modifications même mineures de la spécification.

Imbens et Wooldridge (2008) suggèrent une mesure de la différence de composition, à partir des différences normalisées :

$$\Delta_X = \frac{\bar{X}_1 - \bar{X}_0}{\sqrt{S_1^2 + S_0^2}} \quad (55)$$

avec $S_j = \sum_{i:T_i=j} (X_i - \bar{X})^2 / (N_j - 1)$ les écarts-types empiriques dans les deux échantillons, et \bar{X}_j les moyennes empiriques dans chacun des deux échantillons ²².

²²Il est donc différent du Student du test de l'hypothèse nulle d'égalité des moyennes, qui s'écrit : $St_X = \frac{\bar{X}_1 - \bar{X}_0}{\sqrt{S_1^2/N_1 + S_0^2/N_0}}$. Celui-ci augmente mécaniquement avec la taille de celui-ci. Si les proportions

Ils proposent comme règle empirique de considérer que les estimations risquent d'être sensibles à la spécification lorsque $\Delta_X > 0,25$.

On en trouvera une illustration de cet effet dans la critique de LaLonde (1986). Dans un article célèbre, celui-ci s'interroge sur la validité des méthodes non expérimentales pour évaluer l'effet de politiques publiques. Pour cela, il estime de plusieurs manières l'impact d'un programme expérimental de formation des chômeurs sur le salaire des personnes en ayant bénéficié. Pour estimer l'impact de ce stage, ces dernières avaient été désignées de manière aléatoire parmi un groupe de chômeurs. Pour permettre l'évaluation, des données avaient également été collectées sur les salaires des personnes n'ayant pas été retenues pour participer à cette formation. L'assignation étant aléatoire, la comparaison de ces deux revenus permet d'obtenir l'effet réel de cette formation. Lalonde compare alors cette "vraie" valeur à celles obtenues en utilisant non les données sur le groupe de contrôle expérimental, mais des données administratives ou d'enquêtes (donc utilisées lorsque l'on ne dispose pas de données expérimentales). Il estime donc l'impact de la formation en contrôlant des caractéristiques observables pouvant affecter le revenu (sexe, âge, race, ancienneté dans le chômage et revenu passé). Les résultats de Lalonde montrent une très forte variabilité des estimations aux données et aux spécifications utilisées, notamment les variables explicatives retenues pour contrôler des différences de composition dans l'échantillon ; ce qui n'est évidemment pas le cas lorsqu'il utilise le groupe de contrôle expérimental.

Cependant, comme le soulignent Imbens et Wooldridge (2008), les échantillons du groupe témoin (ayant reçu la formation) et du groupe de contrôle non expérimental sont très différents. Lalonde s'appuyant uniquement sur des spécifications linéaires, la variabilité qu'il observe est liée en grande partie à ces écarts de composition. S'appuyant sur les mêmes données que Lalonde, Dehejia et Wahba (1999) obtiennent des résultats plus satisfaisants en utilisant une méthode d'appariement (*matching*).

5.1.2 Le matching

Les méthodes d'appariement ont été beaucoup utilisées sur les deux dernières décennies. Il est nécessaire de les évoquer au pluriel, tant sont multiples les méthodes et variantes.

La méthode la plus simple est de comparer chaque individu avec son "jumeau". Pour évaluer l'effet moyen sur les traités, cela revient à comparer le revenu Y de chaque bénéficiaire avec un non-bénéficiaire ayant exactement les mêmes caractéristiques observables. Il est rare de trouver un individu du groupe de contrôle exactement identique, donc on choisit le "plus proche voisin". Cette notion demande de définir une métrique. À la simple métrique euclidienne (la distance entre deux individus est la somme de la distance de tous ces covariates), on préfère en pratique une métrique utilisant des covariables normalisées.

restent identiques dans les deux groupes, simplement augmenter la taille de l'échantillon ne permet pas de régler notre question principale d'estimer un contrefactuel à partir du groupe de contrôle, si celui-ci présente des caractéristiques observables trop éloignées du groupe de traitement (ce que mesure mieux *a priori* les seules différences normalisées (55)).

La plus courante est la distance de Mahalanobis, qui pondère par la matrice de variance-covariances des covariables X :

$$d(x_i, x_j) = (x_i - x_j)' \Sigma^{-1} (x_i - x_j) \quad (56)$$

ou au moins sa version diagonale, où chaque covariable est pondérée par l'inverse de son écart-type.

Pour estimer l'effet moyen de la mesure sur les traités on apparie chaque personne de l'échantillon des bénéficiaires avec son plus proche voisin dans l'échantillon des non bénéficiaires²³. L'effet moyen estimé sur les traités $E(Y_1 - Y_0 | T = 1)$ correspond simplement à la moyenne de ces effets estimés :

$$\hat{\delta} = \frac{1}{N_1} \sum_{E_1} (Y_{i1} - \hat{Y}_{i0}) \quad (57)$$

où N_1 est l'effectif du groupe de bénéficiaires E_1 , Y_{i1} le revenu observé de l'individu i , et \hat{Y}_{i0} le revenu du plus proche voisin de i dans le groupe de contrôle.

Le *matching* peut être fait sans remise (un individu du groupe de contrôle ne peut être apparié qu'une fois avec un individu du groupe de traitement), ou avec remise (on utilise l'ensemble de l'échantillon à chaque fois, ce qui autorise des appariements avec le même individu). Le *matching* sans remise suppose qu'on dispose d'un large échantillon de contrôle. L'un de ses inconvénients est que l'estimation pourra être sensible à l'ordre dans lequel l'appariement est effectué.

L'appariement avec le plus proche voisin est sans doute l'un des estimateurs de *matching* le plus utilisé. Il est assez simple à mettre en œuvre, son principe est intuitif et contrairement aux variantes que nous allons voir, il ne demande pas de choix de paramètres. Deux critiques peuvent lui être adressées. La première est qu'on ne contrôle pas de la qualité de l'appariement, la notion de plus proche voisin est (par nature) relative. Certains "plus proches" voisins pourront être en fait éloignés. Or la méthode du plus proche voisin traite de la même manière des couples proches et moins proches. La seconde critique est qu'apparier avec un seul individu prive de l'information apportée par tous les autres, ce qui réduit *a priori* la précision de l'estimation. Par exemple, certains bénéficiaires peuvent avoir plusieurs jumeaux très proches. On peut considérer qu'il est dommage d'en choisir (plus ou moins arbitrairement) un seul.

Il existe donc plusieurs variantes, qui tiennent à la manière dont on construit le contrefactuel de chaque bénéficiaire. En pratique, il va donc s'agir d'estimer \hat{Y}_{i0} en utilisant plus ou moins de personnes du groupe de contrôle.

- *Plus proches voisins* : Plutôt que se limiter au plus proche voisin, on choisit d'apparier avec un nombre fixe M de plus proches voisins. Le revenu contrefactuel du bénéficiaire

²³l'effet du traitement sur toute la population est estimé en effectuant la même opération pour les individus du groupe de contrôle : on compare les revenus observés à ceux des plus proches voisins parmi le groupe des traités.

i sera alors simplement la moyenne des revenus de ces M voisins.

Une autre variante consiste à exclure les couples trop éloignés : on n'utilise pas les bénéficiaires pour lesquels on ne peut pas trouver un (ou M) jumeau à moins d'une certaine distance d à fixer.

- *Radius*, ou *caliper* On sélectionne toutes les personnes du groupe de contrôle situées dans un proche voisinage fixé (i.e. $\|X_i - X_j\| < h$, pour un certain rayon de voisinage h).
- *Kernel*. Ces méthodes vont au bout de cette logique. Le contrefactuel de l'individu i est ainsi calculé par une estimation à noyau (*Kernel*). Tous les individus du groupe de contrôle sont utilisés, mais pondérés par leur distance au traité :

$$\hat{Y}_{0i} = \frac{\sum_{E_0} K\left(\frac{\|X_i - X_k\|}{h}\right) Y_k}{\sum_{E_0} K\left(\frac{\|X_i - X_k\|}{h}\right)} \quad (58)$$

où K est le noyau utilisé²⁴. On utilise souvent un noyau gaussien, ou d'Eparechnikov. h est la taille de la fenêtre (*bandwidth*) du noyau. Les fonctions à noyau classique sont en effet souvent rapidement décroissantes. La fenêtre mesure donc la taille du voisinage en dehors duquel les poids sont très faibles. Plus la fenêtre est petite, et plus l'estimation du contrefactuel d'un bénéficiaire ne prendra en compte que les personnes du groupe de contrôle dont les caractéristiques observables sont très proches de celui-ci. Il n'existe pas vraiment de règle établie pour le choix de cette fenêtre, dans le cas de ces méthodes d'appariement. En pratique, le choix est souvent fait de manière *ad hoc*, ou à partir de "règles du pouce" classiques pour les estimations fonctionnelles semi-paramétriques dont la validité n'a pas été démontrée dans le cas des méthodes par appariement. Frolich (2004) montre à partir de données simulées que, même sur de petits échantillons, un choix de la fenêtre obtenu par validation croisée peut donner des résultats corrects. Pour présenter de manière intuitive cette méthode, il s'agit de choisir la fenêtre qui permet d'obtenir la meilleure estimation du revenu. Plus précisément, il s'agira ici de :

$$h^{CV} = \underset{i \in E_0}{\operatorname{argmin}} \sum (Y_i - \hat{Y}_{-i})$$

où \hat{Y}_{-i} correspond à l'estimation du revenu de la personne i obtenue par la relation (58) à partir de l'ensemble des personnes du groupe de contrôle *sauf* i .

Chacune de ces méthodes dispose de ses avantages et inconvénients. L'opposition entre la plus simple (appariement avec le plus proche voisin) et la plus complexe (noyau) reflète le dilemme classique entre biais et variance. Ne pas utiliser l'ensemble de l'information disponible, dans le cas de l'appariement avec le plus proche voisin, réduit théoriquement la précision. Frolich (2004) montre que même sur de petits échantillons, les estimations

²⁴Rappelons qu'un noyau est une fonction non négative symétrique définie sur \mathfrak{R} , telle que $\int_{-\infty}^{\infty} K(u) du = 1$, c'est-à-dire en général une densité de probabilité.

par fonction à noyau sont toujours plus précises. On peut cependant craindre qu’elles augmentent aussi les risques de mauvais appariements (et donc les biais). En pratique on recommande de tester la sensibilité des résultats à la méthode utilisée.

Au-delà du choix de la méthode, le *matching* exact est souvent compliqué à mettre en œuvre. *À priori*, pour que l’hypothèse d’indépendance conditionnelle soit vérifiée, on peut souhaiter utiliser le maximum d’information, et donc apparier sur de très nombreuses variables. Il pourra alors s’avérer difficile de trouver un voisin proche. On peut montrer qu’à distance finie les estimateurs sont d’autant plus biaisés que le nombre de variables de conditionnement est élevé. Le problème s’avère encore plus aigu lorsque des variables de conditionnement sont continues. C’est pourquoi la grande majorité des études empiriques préfèrent l’appariement sur le score de propension.

5.1.3 Matching sur le score de propension

Une propriété montrée par Rosenbaum et Rubin (1983) permet de régler théoriquement le problème de dimensionnalité. Ils ont en effet montré que si la propriété d’indépendance conditionnelle (48) est vérifiée, alors le revenu est également indépendant du traitement conditionnellement au *score de propension*, c’est-à-dire la probabilité $p(X) = P(T = 1|X)$ d’être traité conditionnellement aux observables :

$$Y_0 \perp T|X \Rightarrow Y_0 \perp T|p(X) \quad (59)$$

Cette propriété simplifie en grande partie les estimations en diminuant la dimension des comparaisons. Néanmoins, cette solution n’est pas miraculeuse : en pratique, le score de propension exact est rarement connu, et il est donc nécessaire de l’estimer. S’il est toujours possible d’envisager des estimations non paramétriques du score, en pratique, l’immense majorité des études utilisent une spécification de type logit ou probit²⁵.

En fait, le résultat de Rosenbaum et Rubin (1983) est plus général : la relation (59) est vérifiée pour toutes les fonctions b , éventuellement différentes du score de propension p , vérifiant :

$$X \perp T|b(X) \quad (60)$$

On dit qu’il s’agit d’une fonction “équilibrée” (*balanced*) des covariables. Intuitivement, cela signifie que conditionnellement à cette fonction des covariables, la composition des observables est la même dans le groupe des bénéficiaires et des non-bénéficiaires. L’avantage de cette propriété est que même si on n’observe pas le vrai score, l’estimation obtenue en utilisant une spécification \hat{p} de ce score sera convergente dès lors que (60) le sera.

²⁵Chassez le paramétrique par la porte, il reviendra par la fenêtre... Alors que l’une des justifications du *matching* est de ne plus reposer sur des spécifications paramétriques de la distribution conditionnelle du revenu, on en revient finalement à des spécifications paramétriques du score. Cela réduit pour certains l’intérêt de la méthode de l’appariement sur le score de propension.

Matching sur le score Les méthodes décrites plus haut : appariement au(x) plus proche(s) voisin(s), radius, ou *kernel matching* peuvent toutes être appliquées en utilisant uniquement l'estimation du score de propension comme mesure de la distance entre deux observations.

Estimation par strates Un autre estimateur classique proposé par Rosenbaum et Rubin (1983) s'appuie simplement sur une division de l'échantillon en J groupes d'individus ayant approximativement le même score de propension.

Au sein de chaque strate S_j , on peut calculer l'effet du traitement comme la différence entre le revenu moyen des traités et des non-traités :

$$\hat{\delta}_j = \bar{Y}_{1j} - \bar{Y}_{0j} \quad (61)$$

où \bar{Y}_{1j} et \bar{Y}_{0j} sont les moyennes de revenus des personnes appartenant à la strate j respectivement dans l'échantillon des traités et des non-traités. On peut alors calculer l'effet moyen du traitement en pondérant chacun de ses estimateurs par l'effectif de la strate :

$$\hat{\delta} = \sum_j \frac{N_{1j} + N_{0j}}{N} \hat{\delta}_j \quad (62)$$

La variance est estimée en général à partir des variances empiriques au sein de chaque strate. En supposant que l'assignation au traitement est aléatoire au sein de chaque strate, la variance pour chaque strate est obtenue par : $\hat{V}_j = \hat{V}_{0j} + \hat{V}_{1j}$, où \hat{V}_{0j} et \hat{V}_{1j} sont simplement les variances empiriques calculées sur la strate j pour chaque groupe (traité et non traité). \hat{V}_{0j} s'obtient par exemple par $\hat{V}_{0j} = \frac{S_{0j}^2}{N_{0j}}$ avec $S_{0j}^2 = \frac{1}{N_{0j}} \sum_{i \in S_j, T_i=0} (Y_i - \bar{Y}_{0j})^2$. Un estimateur de la variance totale s'obtient à partir de :

$$\hat{V} = \sum_j \hat{V}_j \left(\frac{N_{1j} + N_{0j}}{N} \right)^2 \quad (63)$$

La question est évidemment de savoir le nombre de strates qu'il faut retenir. Si elles sont trop peu nombreuses, elles seront plus "larges", et le score risque de beaucoup varier au sein d'une strate. L'hypothèse sous-jacente (les personnes traitées et non traitées d'une même strate ont des valeurs du score proches) risque d'être prise en défaut. Mais des strates trop fines risquent de ne contenir personne. En pratique, les études qui utilisent cette méthode retiennent souvent de manière *ad hoc* cinq strates.

L'avantage de cette méthode est d'être particulièrement simple à estimer. Son inconvénient est qu'elle risque de fournir des résultats biaisés, si les strates sont de taille trop importante et regroupent donc des personnes avec des caractéristiques finalement disparates. Une extension intéressante consiste à mixer cette approche avec une régression linéaire au sein de chaque strate. On estimera $\hat{\delta}_j$ par une régression linéaire à l'intérieur de la strate S_j :

$$Y_i = \alpha_j + \delta_j T_i + \beta_j X_i + \epsilon_{ij} \quad (64)$$

Cette estimation permet d'obtenir à la fois l'effet moyen du traitement au sein de la strate S_j , et sa variance, qu'on peut réagrèger comme plus haut.

L'avantage de cette approche mixte est qu'elle permet de dépasser les limites des approches linéaires et par strate. La régression linéaire permet de corriger l'hétérogénéité probable au sein d'une strate. Si l'approximation linéaire n'est pas tenable sur l'ensemble de l'échantillon, elle est plus crédible localement.

Estimation pondérée Comme on l'a vu, comparer directement les moyennes des revenus dans les deux échantillons risque de fournir un indicateur biaisé, car les deux échantillons n'ont *a priori* pas la même composition dans les observables. La comparaison des revenus des groupes des traités et des non-traités risque donc de refléter avant tout ces différences de composition. L'idée de l'estimateur pondéré est que pour obtenir un estimateur du revenu potentiel avec traitement moyen sur l'ensemble de la population à partir du seul échantillon des bénéficiaires, on va sur-pondérer les personnes qui avaient peu de chances de l'être, et qui sont donc a priori plus comparables aux non-bénéficiaires. Et inversement, pour obtenir un revenu potentiel sans traitement représentatif de toute la population à partir du seul échantillon des non-bénéficiaires, les revenus individuels observés sont d'autant plus pondérés qu'ils sont proches des bénéficiaires.

Plus formellement, on peut remarquer qu'un estimateur sans biais des revenus potentiels moyens est fourni par :

$$E(Y_{1i}) = E\left[\frac{p(X_i)Y_{1i}}{p(X_i)}\right] = E\left[E\left(\frac{T_i Y_{1i}}{p(X_i)} \mid X\right)\right] = E\left[\frac{T_i Y_{1i}}{p(X_i)}\right] = E\left(\frac{T_i Y_i}{p(X_i)}\right) \quad (65)$$

La deuxième égalité vient de la définition de $p(X) = P(T = 1 \mid X) = E(T \mid X)$, de l'hypothèse identifiante, et de la loi des espérances itérées. Dans la dernière expression, c'est le revenu *observé* qui intervient, ce qui assure qu'il est possible de l'estimer.

De même,

$$E(Y_{0i}) = E\left(\frac{(1 - T_i)Y_i}{1 - p(X_i)}\right) \quad (66)$$

Un estimateur empirique de l'effet moyen du traitement est alors :

$$\hat{\Delta} = \frac{1}{N} \sum_{i=1}^N \left(\frac{T_i Y_i}{p(X_i)} - \frac{(1 - T_i)Y_i}{1 - p(X_i)} \right) \quad (67)$$

Cet estimateur a cependant un défaut important : *a priori*, les deux systèmes de pondérations ne somment pas à un²⁶. De fait, Hirano, Imbens, et Ridder (2003) montrent que paradoxalement, il est plus efficace d'utiliser un estimateur du score dans le système de pondération plutôt que le vrai score. Ceci est plutôt une bonne nouvelle, puisque le vrai score n'est pas connu en général.

²⁶Par exemple, pour l'échantillon des bénéficiaires la somme des pondérations vaut $\frac{1}{N_1} \sum_{E_1} \frac{T_i}{p(X_i)}$. Elle est en général différente de un, même si son espérance est égale à un.

Dans ce cas, un estimateur convergent et efficace de l'effet moyen du traitement est fourni par :

$$\hat{\Delta} = \sum_{i=1}^N \frac{T_i Y_i}{\hat{p}(X_i)} / \sum_{i=1}^N \frac{T_i}{\hat{p}(X_i)} - \sum_{i=1}^N \frac{(1 - T_i) Y_i}{1 - \hat{p}(X_i)} / \sum_{i=1}^N \frac{1 - T_i}{1 - \hat{p}(X_i)} \quad (68)$$

Hirano, Imbens, et Ridder (2003) proposent également un estimateur de l'effet moyen du traitement sur les traités. Ils montrent plus spécifiquement que l'estimateur convergent optimal est donné par :

$$\hat{\Delta}^{ATT} = \sum_{i=1}^N p(X_i) \left[\frac{T_i Y_i}{\hat{p}(X_i)} - \frac{(1 - T_i) Y_i}{1 - \hat{p}(X_i)} \right] / \sum_{i=1}^N p(X_i) \quad (69)$$

Cette expression mixe donc à la fois le vrai score et le score estimé. En pratique, le vrai score $p(x)$ n'étant pas connu, ils montrent que l'estimateur convergent le plus efficace est obtenu par la formule ci-dessus en le remplaçant par son estimation $\hat{p}(x)$.

Les propriétés de ces estimateurs sont asymptotiques. Il n'existe pas à ce jour de littérature sur leurs propriétés à distance finie. Sur des échantillons de taille limitée, on peut en particulier s'interroger sur la dépendance des résultats à l'approximation paramétrique pour estimer le score.

5.2 Estimation du score

En général, le score n'est pas connu, et il est nécessaire de l'estimer. Pour tenir compte de la nature bornée du score (c'est une probabilité...), il est d'usage de le modéliser par une forme logistique (ou un probit).

$$\hat{p}(X) = \frac{\exp(f(X))}{1 + \exp(f(X))} \quad (70)$$

où $f(X)$ est une fonction des observables X . La fonction la plus simple (et la plus courante...) est une forme linéaire $f(X) = X\beta$. Pour être toutefois plus proche de la vraie distribution, il est recommandé d'utiliser une approximation de type polynômial dans les covariables : outre les covariables on peut ajouter les termes de degré plus élevé, y compris donc des interactions. De manière formelle, Hirano, Imbens, et Ridder (2003) proposent d'estimer le score par :

$$\hat{p}(x) = \frac{\exp(\hat{\gamma}_0 h_0(x) + \dots + \hat{\gamma}_K h_K(x))}{1 + \exp(\hat{\gamma}_0 h_0(x) + \dots + \hat{\gamma}_K h_K(x))} \quad (71)$$

où les fonctions $h_k(x)$ sont des fonctions polynomiales des variables explicatives, dont le nombre K dépend de la taille de l'échantillon. En pratique, on se contente en général de coefficients de degré un ou deux ²⁷.

²⁷Précisément, les résultats de convergence sont établis par Hirano, Imbens, et Ridder (2003) pour $K = cste.N^\nu$ avec ν tel que $0 < \nu < 1/9$ si le vrai score est infiniment différentiable.

À notre connaissance, il n'existe pas de règle clairement établie pour le choix optimal des termes. Dehejia et Wahba (1999) proposent une procédure de "test" intéressante. Elle s'appuie sur une propriété importante du score :

$$T \perp X | p(X) \quad (72)$$

i.e. le score permet d'équilibrer dans les observables les deux échantillons de traités et de non-traités. Ils proposent donc de complexifier la modélisation du score jusqu'à obtenir des échantillons équilibrés. Partant d'une spécification parcimonieuse dans les observables ($f(X) = X\beta$), on enrichit la spécification par des termes de degrés plus élevés tant que la condition n'est pas vérifiée. En principe, il faudrait pouvoir tester $\bar{X}_1 = \bar{X}_0$ conditionnellement à chaque valeur de $p(X)$. Dehejia et Wahba (1999) suggèrent de tester par strates du score : on subdivise les échantillons selon leur score ; au sein de chaque strate, on vérifie que les moyennes des observables sont identiques dans l'échantillon des traités et des non-traités, par exemple par un test d'égalité des moyennes. Le problème de cette règle pratique est qu'en cas de rejet, il est difficile de savoir si un rejet du test tient à des strates trop grossières (la propriété n'est censée être vérifiée qu'à $p(X)$ constant), ou parce que le score est effectivement mal spécifié. Becker et Ichino (2002) fournissent un programme Stata implémentant ce test. Plus récemment, Huber (2009) propose une autre procédure de test de cette propriété.

5.3 Restriction au support commun

Les méthodes présentées sous l'hypothèse d'indépendance conditionnelle reposent sur l'hypothèse d'un support commun. Elle signifie qu'il existe à la fois des bénéficiaires et des non-bénéficiaires avec des valeurs identiques des observables. Si ce n'est pas le cas, il sera par exemple impossible de trouver pour chaque bénéficiaire des non-bénéficiaires comparables.

C'est pourquoi il est important de vérifier que le support commun, c'est-à-dire la zone sur lesquelles cette condition est vérifiée, est suffisamment large. Pour cela, la manière la plus simple est de représenter la distribution du score sur les deux sous-échantillons. En pratique, on peut utiliser les histogrammes de la probabilité estimée d'être traité pour l'échantillon des bénéficiaires et des non bénéficiaires. Il est nécessaire de vérifier que le recouvrement est large, c'est-à-dire que pour chaque valeur du score, on a un nombre suffisant de personnes dans les deux sous-échantillons. Cela peut également se faire en traçant les densités. Les graphiques ci-dessous montrent deux cas caricaturaux. Dans le premier, si les modes des deux distributions sont bien séparés, ce qui exprime que la probabilité d'être traité est plus importante dans le sous-échantillon effectivement traité, les supports sont larges. Dans le second, les deux distributions sont pratiquement disjointes, et il sera impossible d'estimer un effet puisqu'on sera amené à comparer des individus très différents en terme d'observables.

L'hypothèse d'indépendance conditionnelle suppose qu'on compare des personnes avec des caractéristiques observables identiques. Si elles n'existent pas pour certaines valeurs

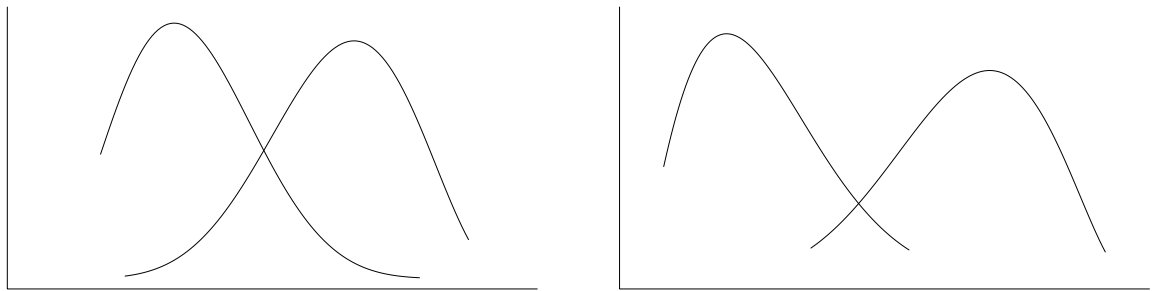


FIG. 9: Distributions du score de propension

des observables, il est incorrect d'utiliser ces individus pour l'estimation. Si on ne se limite pas au support commun, les estimations pourront être biaisées²⁸.

En pratique, plusieurs méthodes ont été proposées pour se restreindre au support commun. Il faut remarquer que ces méthodes changent la nature de ce qui est estimé : on estime un impact sur une partie seulement de la population (i.e. dont les observables sont telles qu'on observe un recouvrement pour les deux sous-échantillons).

- Méthode du min/max. Pour estimer l'effet du traitement sur les traités, Dehejia et Wahba (1999) éliminent les individus du groupe de contrôle dont le score est inférieur au minimum observé sur le groupe de traitement. Pour l'effet moyen sur l'ensemble de la population, il faut également faire l'inverse : éliminer les personnes du groupe de traitement dont la probabilité estimée d'être traité est plus grande que le maximum observé dans le groupe de contrôle.
- Méthode d'écrémage (*trimming*). Elle consiste à exclure les individus dont la probabilité d'être traité est trop forte ou trop faible. En pratique cela revient à se restreindre aux individus tels que :

$$\alpha \leq p(X) \leq 1 - \alpha$$

On trouvera dans Imbens et Wooldridge (2008) une condition sur α pour que l'estimateur de l'effet du traitement soit efficace, sous certaines conditions. Ils proposent comme règle empirique de prendre $\alpha = 0.01$ ²⁹.

²⁸Empiriquement, les estimateurs ne traitent pas vraiment ce problème. Un plus proche voisin d'un bénéficiaire pourra se trouver en dehors du support de la distribution des observables pour les bénéficiaires. L'hypothèse linéaire permet "artificiellement" de comparer des individus pour des valeurs de covariables que l'on n'observe pas.

²⁹Heckman, Ichimura, et Todd (1997) utilisent une méthode beaucoup plus complexe.

5.4 Quelles variables de conditionnement ?

L'hypothèse d'indépendance conditionnelle, sous laquelle ces différents estimateurs correspondent à l'effet propre de la mesure étudiée, demande de disposer de suffisamment de caractéristiques observables dans les données pour que, conditionnellement à ces observables, le fait de bénéficier ou non de la mesure ne soit pas lié au bénéfice escompté. La question du choix de ces variables est évidemment primordiale. Comme souvent, il n'y a pas de règle précise pour les déterminer. En pratique, il doit s'agir de variables qui peuvent avoir un impact sur la variable d'intérêt et sur le choix de bénéficiaire du dispositif. Deux points doivent être notés. En premier lieu, il faut faire attention à ne pas utiliser des variables mesurées *après* la mise en place du dispositif, et qui peuvent être également affectées par celui-ci : cela peut créer un problème évident d'endogénéité. Par exemple, supposons qu'on souhaite évaluer l'impact d'une aide aux entreprises sur le niveau d'emploi, il ne serait pas correct de conditionner par des variables de bilan postérieures à l'obtention de l'aide. Celles-ci peuvent en effet être également modifiées par le niveau de l'aide.

En second lieu, pour que l'hypothèse de support commun soit vérifiée, il ne faut pas que des variables explicatives expliquent "trop bien" le fait d'être traité, car cela peut rendre impossible un appariement avec des bénéficiaires très proches. Pour prendre un exemple caricatural, supposons l'existence d'une aide aux entreprises généralisée à l'échelle d'un secteur d'activité. Il serait alors impossible d'utiliser l'appartenance à ce secteur comme variable de conditionnement puisqu'il n'existerait pas d'entreprise non bénéficiaire à laquelle comparer la situation des bénéficiaires. Si on peut raisonnablement penser que le "revenu" ne dépend pas de cette variable, cela ne pose aucun problème. Si ce n'est pas le cas, cela signifie que l'effet mesuré sera biaisé, puisqu'il intégrera les différences systématiques liés à ce déterminant du revenu.

5.5 Inférence

Dans le cadre du modèle linéaire, l'estimation de la précision est tout à fait standard et nous ne revenons pas dessus. Dans les autres cas, les méthodes sont loin d'être aussi balisées. Pour mener l'inférence, de méthodes de bootstrap sont souvent utilisées. Il faut cependant noter que les preuves de leur convergence n'ont pas été établies jusqu'à présent pour l'ensemble des estimateurs présentés ici, et Abadie et Imbens (2008) montrent même qu'ils ne sont pas convergents en général pour les méthodes d'appariement sur un nombre fixe de voisins avec remise.

5.6 Traitements multiples

L'un des avantages de la régression linéaire est de permettre d'évaluer des traitements avec plusieurs modalités, ou continus. Dans le cas du *matching* sur le score de propension, Lechner (2002) propose une extension au cas de plusieurs traitements (par exemple, plusieurs types de formation pour les chômeurs). La solution la plus simple consiste à étudier l'impact des différents traitements deux à deux. Par exemple pour étudier celui de suivre

le traitement T_j plutôt que T_k , on se place sur le sous-échantillon des individus ayant suivi l'un ou l'autre des ces deux traitements, et on applique une des procédures d'appariement vue plus haut. Notons que les résultats peuvent être plus difficiles à présenter et également à interpréter : on ne peut donner l'impact d'un traitement que par rapport à un autre.

5.7 En résumé

En résumé, les étapes pour mettre en œuvre des méthodes sous une hypothèse d'indépendance conditionnelle sont les suivantes :

1. Sélectionner un groupe de contrôle.
2. Sélectionner un ensemble de variables conditionnantes (voir section 5.4). Rappelons qu'en principe, des données en coupe sont suffisantes. Néanmoins, si on dispose de données individuelles antérieures, l'hypothèse d'indépendance conditionnelle est plus crédible lorsque l'on peut contrôler d'effets fixes, en comparant les évolutions des revenus dans le groupe de bénéficiaire et le groupe de contrôle déterminé par appariement.
3. Déterminer une méthode d'estimation. Comme on l'a vu, on peut tenter une spécification linéaire. Cependant, si les compositions des deux échantillons de traités et du groupe de contrôle sont trop différentes et que la dépendance réelle du revenu dans les observables s'éloigne trop de cette approximation linéaire, les estimations risquent d'être peu robustes (voir section 5.1.1). Il peut donc être préférable d'utiliser une procédure moins paramétrique. Les méthodes sont multiples, la plus courante est l'appariement sur le score de propension.
4. Dans ce dernier cas, estimer le score de propension (cf. section 5.2). Cela peut se faire par un logit ou un probit. En théorie, il est nécessaire d'utiliser une estimation du score qui "équilibre" les deux échantillons, c'est-à-dire que les personnes de groupes différents ayant des valeurs proches pour cette fonction présentent des caractéristiques observables identiques.
5. Estimer les effets. L'estimation proprement dite se fait par une des méthodes présentées dans la section 5.1.2.

Toute la crédibilité des estimateurs proposés ici repose sur la validité de l'hypothèse d'indépendance conditionnelle. Il faut être conscient que cette hypothèse est très forte. Elle tient précisément aux observables disponibles et utilisées pour la comparaison.

En particulier, si on dispose de trop peu d'informations dans nos données sur les choix entre les variables, il ne sera pas possible d'éliminer le biais de sélection.

Conclusion

Ce document présente les différentes méthodes d'évaluation. En pratique, un chargé d'étude pourra se demander quelle méthode est la "meilleure". Lorsqu'on se limite à une évaluation *ex post*, il n'existe évidemment pas de réponse générale à cette question : en général, on utilisera la méthode "utilisable" en fonction du dispositif évalué, et des données disponibles. Il faut aussi être conscient qu'il s'agit d'une littérature dont les bases théoriques se posent progressivement. Les usages évoluent aussi en fonction de ces avancées. Très schématiquement, les différences de différences, conceptuellement très simples mais qui reposent sur un choix du groupe de contrôle *ad hoc*, ont été "supplantées" au cours des années quatre-vingt-dix par les variables instrumentales issues d'expériences naturelles dans la lignée des travaux d'Angrist, Krueger et Card par exemple. La limite de ce type d'estimateur, mise en évidence par exemple par Angrist, Imbens, et Rubin (1996), et le papier influent de Heckman, Ichimura, et Todd (1997) qui présente une étude approfondie des biais afférents à l'usage des méthode d'appariement, a beaucoup fait pour populariser ces dernières au tournant de la dernière décennie. Sur la période très récente, ce sont maintenant les régressions sur discontinuités qui sont considérées comme la méthode la plus convaincante, conséquence en particulier de Hahn, Todd, et Van der Klaauw (2001) qui formalise et fournit des fondements théoriques à cette méthode d'identification. Les frontières entre les régressions sur discontinuités et variables instrumentales sont floues : de nombreuses études antérieures à l'article de Hahn, Todd, et Van der Klaauw (2001) utilisaient déjà, sous le terme de variables instrumentales, une méthode d'identification qui serait appelée aujourd'hui régression sur discontinuité. Comme on l'a vu, la mise en œuvre pratique est identique dans certains cas.

Il faut cependant garder à l'esprit que toutes ces méthodes empiriques reposent sur des hypothèses fortes. S'affranchir des effets de sélection est difficile. Cela signifie en particulier qu'il est extrêmement rare de juger de manière indiscutable de l'efficacité d'un dispositif à l'aide d'une telle évaluation. Pour une évaluation rigoureuse et peu contestable sinon incontestable, il est presque impossible de faire l'économie d'une évaluation par une expérimentation aléatoire. Celle-ci a évidemment un coût, et suppose d'être anticipée et mise en place avant l'implémentation d'une mesure sur l'ensemble du territoire. Notons enfin que même ces méthodes prêtent le flanc aux critiques sur les effets d'équilibre général, ou les externalités possibles d'une politique. Le retour à des modèles plus structurels (Lise, Seitz, et Smith 2004), ou la prise en compte explicite des externalités (Miguel et Kremer 2004, Angelucci et De Giorgi 2009) sont des pistes intéressantes pour traiter de ces questions. Enfin, citons les avancées récentes en économétrie sur les régressions de quantile, avec la mise à disposition croissante de procédures adaptées dans les logiciels statistiques standards et d'un corpus théorique fournissant des bases statistiques nécessaires. Ces méthodes, qui dépassent le cadre de l'évaluation, permettent de porter un diagnostic plus précis sur l'impact des différentes politiques. Elles constituent l'une des pistes les plus intéressantes pour la suite³⁰. Au-delà de l'effet moyen d'un dispositif, il s'agit d'évaluer son impact sur l'ensemble de la distribution de la variable de résultat à laquelle on s'intéresse.

³⁰On en trouvera des exemples dans Abadie, Angrist, et Imbens (2002), Firpo (2007), Lamarche (2008).

Annexe A : Cadre de Rubin et économétrie linéaire classique

La régression linéaire est l'un des outils les plus utilisés pour évaluer l'effet d'une variable sur un revenu. À ce titre, il est intéressant de faire le lien entre l'approche ci-dessus et le cadre plus classique de l'économétrie linéaire. Ce dernier ne nécessite pas le recours à cette construction intellectuelle que constituent les revenus potentiels, puisqu'on ne s'intéresse qu'au revenu observé. Notons que celui-ci s'exprime simplement en fonction des revenus potentiels.

Pour tout individu i en effet, on observe son revenu $Y_i = T_i Y_{1i} + (1 - T_i) Y_{0i}$. On peut aussi écrire : $Y_i = Y_{i0} + T_i \Delta_i$ où $\Delta_i = Y_{i1} - Y_{i0}$ représente l'effet individuel du traitement pour la personne i .

On peut réécrire cette expression pour lui donner la forme classique d'une régression linéaire :

$$Y_i = \alpha + T_i \Delta + w_i \quad (73)$$

Il suffit par exemple de noter $\alpha = E(Y_0)$, la variable d'intérêt potentielle moyenne en l'absence de traitement, et d'appeler w_i le terme résiduel :

$$w_i = Y_{i0} - E(Y_0) + T_i [\Delta_i - \Delta]$$

Cette approche est plus condensée, mais c'est sans doute là que réside sa faiblesse. Comme noté par Imbens et Wooldridge (2008), elle oblige à traiter simultanément deux problèmes différents : premièrement, l'hétérogénéité des effets de la mesure, qui implique que Δ_i n'est pas un paramètre constant sur l'ensemble des observations ; deuxièmement, celle de la présence d'auto-sélection des individus, qui introduit un biais potentiel d'estimation de l'effet moyen Δ . Pour bien comprendre ce deuxième problème, il faut se demander ce que représente le coefficient Δ qu'on cherche à estimer, et noter qu'une condition pour que le modèle soit correctement spécifié est que l'espérance du résidu de l'équation soit nulle. Or :

$$E(w_i) = E(Y_{i0}) - E(Y_0) + E[T_i(\Delta_i - \Delta)] \quad (74)$$

$$= P(T = 1)[E(\Delta_i|T_i = 1) - \Delta] \quad (75)$$

Ce terme n'est nul que si $\Delta = E(\Delta_i|T_i = 1)$. Autrement dit, le seul paramètre identifiable *a priori* est l'effet moyen de la mesure sur les bénéficiaires. Pour estimer un effet moyen sur toute la population, il faut faire l'hypothèse :

$$E(\Delta_i|T_i = 1) = E(\Delta_i)$$

Pour pouvoir identifier l'effet sur toute la population, il faut donc que l'effet du traitement soit indépendant du traitement. Cette condition n'est pas nécessaire si on veut se contenter d'estimer l'effet du traitement sur la seule population des personnes traitées. Elle n'est jamais suffisante même dans ce dernier cas.

En effet, on cherche maintenant à estimer :

$$\Delta = E(\Delta_i|T_i = 1)$$

toujours avec l'équation (73). Comme $E(\Delta_i|T_i = 1) = E(\Delta_i)$ (74) implique $E(w_i) = 0$. Pour estimer sans biais le paramètre Δ , il faut cependant qu'on ait la relation classique $E(w_i|T = 1) = E(w_i)$ ($= 0$), i.e. que la distribution *ex ante* des revenus des bénéficiaires soit identique à celles des revenus des non-bénéficiaires. Cette relation est vérifiée si $Y_0 \perp T$ ³¹. On retombe bien sur la condition d'indépendance (1).

Dans le cadre linéaire classique, les hypothèses identifiantes s'appuient sur les propriétés statistiques des composantes inobservées. Ces dernières mélangent des effets liés à la sélectivité à ceux de l'évolution des revenus individuels. Raisonner sur les revenus potentiels permet à l'inverse de dissocier la modélisation des revenus de celle des effets de sélection. En outre, elle souligne le fait que le traitement puisse avoir des effets différents selon les individus alors que l'approche classique fait plus ou moins implicitement l'hypothèse d'un effet constant. Enfin, l'approche par revenus potentiels permet de mieux expliciter les stratégies qui peuvent permettre d'identifier un effet causal (déterminer un groupe de contrôle crédible).

Ces avantages ont aussi un coût. En particulier, ce cadre est plus difficile à transposer lorsqu'on s'intéresse à un traitement continu (montants d'aides par exemple) ou à plusieurs modalités (efficacité relative de plusieurs types d'aides par exemple), même si des études récentes tentent de l'étendre dans ce sens.

³¹On a $E(w|T = 1) = E(Y_0|T = 1) - E(Y_0) + E[T_i\Delta_i|T = 1] - \Delta$. Le deuxième terme est nul car par définition $\Delta = E(\Delta_i|T_i = 1)$, et le premier également si le revenu potentiel Y_0 est indépendant de T .

Références

- ABADIE, A., J. ANGRIST, ET G. W. IMBENS (2002) : “Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings,” *Econometrica*, 70(1), 91–117.
- ABADIE, A., A. DIAMOND, ET J. HAINMUELLER (2007) : “Synthetic Control Methods for Comparative Case Studies : Estimating the Effect of California’s Tobacco Control Program,” *NBER Working Papers*, 12831.
- ABADIE, A., ET G. W. IMBENS (2008) : “On the Failure of the Bootstrap for Matching Estimators,” *Econometrica*, 76(6), 1537–1557.
- ANGELUCCI, M., ET G. DE GIORGI (2009) : “Indirect Effects of an Aid Program : How Do Cash Transfers Affect Ineligibles’ Consumption ?,” *American Economic Review*, 99(1), 486–508.
- ANGRIST, J., ET W. EVANS (1998) : “Children and Their Parents Labor Supply : Evidence from Exogenous Variation in Family Size,” *The American Economic Review*, 88.
- ANGRIST, J., G. W. IMBENS, ET D. RUBIN (1996) : “Identification of causal effects using instrumental variables.,” *Journal of the American Statistical Association*, 91(434), 444–455.
- ANGRIST, J., ET A. KRUEGER (1991) : “Does Compulsory School Attendance Affect Schooling and Earnings,” *Quarterly Journal of Economics*, 106.
- (1999) : *Empirical Strategies in Labor Economics* vol. IIIA, chap. 23. Ashenfelter and Card.
- ANGRIST, J., ET LAVY (1999) : “Using Maimonides’ rule to estimate the effect of class size on scholastic achievement,” *Quarterly Journal of Economics*, 114(2), 533–575.
- ANGRIST, J., ET A. PISCHKE (2009) : *Mostly Harmless Econometrics : An Empiricist’s Companion*. Princeton University Press.
- ASHENFELTER, O. (1978) : “Estimating the effects of Training Programs on Earnings,” *Review of Economic Studies*.
- ATHEY, S., ET G. W. IMBENS (2006) : “Identification and Inference in Nonlinear Difference-in-differences Models,” *Econometrica*, 74, 431–497.
- AUTOR, D. H. (2003) : “Outsourcing at Will : The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing,” *Journal of Labor Economics*, 21(1), 1–42.
- BECKER, S., ET A. ICHINO (2002) : “Estimation of Average Treatment Effects Based on Propensity Scores,” *The Stata Journal*, 2(4), 358–377.
- BERTRAND, M., E. DUFLO, ET S. MULLAINATHAN (2004) : “How much should we trust differences-in-differences estimates?,” *The Quarterly Journal of Economics*, 119, 249–275.
- BLACK, D. A., J. GALDO, ET J. A. SMITH (2007) : “Evaluating the performing of Non-experimental estimators : Evidence from a Randomized Ui program,” *mimeo*.

- BLACK, S. (1999) : “Do better schools matter? Parental valuation of elementary education,” *Quarterly Journal of Economics*, 114(2), 587–599.
- BOUND, J., D. JEAGER, ET R. BAKER (1995) : “Problems with Instrumental Variable Estimation When the Correlation Between the Instruments and the Endogenous is weak,” *Journal of the American Statistical Association*, 90(430), 443–450.
- CAMERON, A. C., J. B. GELBACH, ET D. L. MILLER (2008) : “Bootstrap-Based Improvements for Inference with Clustered Errors,” *The Review of Economics and Statistics*, 90(3), 414–427.
- CARD, D. (1990) : “The Impact of the Mariel Boatlift on the Miami Labor Market,” *Industrial and Labor Relations Review*, 43(2), 245–257.
- CARD, D. (1993) : “Using Geographic Variation in College Proximity to Estimate the Return to Schooling,” *NBER Working Paper*, 4483.
- CARD, D., ET A. KRUEGER (1994) : “Minimum Wages and Employment : A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania,” *American Economic Review*, 84(4), 772–793.
- DAVIDSON, R., ET J. G. MACKINNON (1993) : *Estimation and Inference in Econometrics*. Oxford University Press.
- DEHEJIA, R., ET S. WAHBA (1999) : “Causal effects in non experimental studies : reevaluating the evaluation of training programs,” *Journal of Statistical American Association*, 94(448).
- DONALD, S. G., ET K. LANG (2007) : “Inference with Difference-in-Differences and Other Panel Data,” *The Review of Economics and Statistics*, 89(2), 221–233.
- DUFLO, E., R. GLENNERSTER, ET M. KREMER (2006) : “Using Randomization in Development Economics Research : A Toolkit,” *NBER Working Paper*, T0333.
- DURBIN, D., B. MEYER, ET W. VISCUSI (1995) : “Workers’ Compensation and Injury Duration : Evidence from a Natural Experiment,” *American Economic Review*, 85(3).
- FIRPO, S. (2007) : “Efficient Semiparametric Estimation of Quantile Treatment Effects,” *Econometrica*, 75(1), 259–276.
- FROLICH, M. (2004) : “Finite-Sample Properties of Propensity-Score Matching and Weighting Estimators,” *The Review of Economics and Statistics*, 86(1), 77–90.
- HAHN, J., P. TODD, ET W. VAN DER KLAUW (2001) : “Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design,” *Econometrica*, 69(1), 201–09.
- HECKMAN, J., H. ICHIMURA, ET P. TODD (1997) : “Matching as an econometric evaluation estimator : evidence from evaluating a job training programme,” *Review of Economic Studies*, 64(4), 605–654.
- HECKMAN, J. J., ET B. E. HONORE (1990) : “The Empirical Content of the Roy Model,” *Econometrica*, 58(5), 1121–49.

- HIRANO, K., G. W. IMBENS, ET G. RIDDER (2003) : “Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score,” *Econometrica*, 71(4), 1161–1189.
- HUBER, M. (2009) : “Treatment evaluation in the presence of sample selection,” University of St. Gallen Department of Economics working paper series 2009 2009-07, Department of Economics, University of St. Gallen.
- IMBENS, G. W., ET T. LEMIEUX (2008) : “Regression discontinuity designs : A guide to practice,” *Journal of Econometrics*, 142(2), 615–635.
- IMBENS, G. W., ET J. WOOLDRIDGE (2008) : “Recent Developments in the Econometrics of Program Evaluation,” *NBER Working Paper*, 14251.
- LALIVE, R. (2008) : “How do extended benefits affect unemployment duration ? A regression discontinuity approach,” *Journal of Econometrics*, 142(2), 785–806.
- LALONDE, R. (1986) : “Evaluating the Econometric Evaluation of Training Programs with Experimental Data,” *American Economic Review*, 76, 604–620.
- LAMARCHE, C. (2008) : “Private school vouchers and student achievement : A fixed effects quantile regression evaluation,” *Labour Economics*, 15(4), 575–590.
- LECHNER, M. (2002) : “Program Heterogeneity And Propensity Score Matching : An Application To The Evaluation Of Active Labor Market Policies,” *The Review of Economics and Statistics*, 84(2), 205–220.
- LISE, J., S. SEITZ, ET J. SMITH (2004) : “Equilibrium Policy Experiments and the Evaluation of Social Programs,” NBER Working Papers 10283, National Bureau of Economic Research, Inc.
- LUDWIG, J., ET D. L. MILLER (2007) : “Does Head Start Improve Children’s Life Chances ? Evidence from a Regression Discontinuity Design,” *The Quarterly Journal of Economics*, 122(1), 159–208.
- MANSKI, C. (2000) : “Economic Analysis of Social Interactions,” *Journal of Economic Perspectives*, 14(3), 115–136.
- MCCLELLAN, M., B. J. MCNEIL, ET J. P. NEWHOUSE (2008) : “Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality ? Analysis using instrumental variables,” *NBER Working Paper*, 14251.
- MIGUEL, E., ET M. KREMER (2004) : “Worms : identifying impact on Education and Health in the Presence of Treatment Externalities,” *Econometrica*, 72(1), 159–217.
- MOULTON, B. R. (1990) : “An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Unit,” *The Review of Economics and Statistics*, 72(2), 334–38.
- OREOPOULOS, P. (2006) : “Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws Really Matter,” *American Economic Review*, 96(1), 152–175.
- PORTER, J. (2002) : “Asymptotic Bias and Optimal Convergence Rates for Semiparametric Kernel Estimators in the Regression Discontinuity Model,” Harvard Institute of Economic Research Working Papers 1989, Harvard - Institute of Economic Research.

- RATHELOT, R., ET P. SILLARD (2009) : “Zones Franches Urbaines : quels effets sur l’emploi salarié et les créations d’établissements,” *Economie et Statistique*, 415–416, 81–96.
- ROBIN, J.-M. (2002) : “Endogénéité et variables instrumentales dans les sciences sociales,” *Insee méthodes*, 101.
- ROSENBAUM, P. R., ET D. B. RUBIN (1983) : “The Central Role of the Propensity Score in Observational Studies for Causal Effects,” *Biometrika*, 70, 41–55.
- ROSENZWEIG, M., ET K. WOLPIN (2000) : “Natural “Natural Experiment” in Economics,” *Journal of Economic Literature*, XXXVIII, 827–874.
- STOCK, J., J. WRIGHT, ET M. YOGO (2002) : “A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments,” *Journal of Business and Economic Statistics*, 20(4), 518–29.
- VAN DER KLAUW, W. (2002) : “Estimating the Effect of Financial Aid offers on College Enrollment : A regression discontinuity design,” *International Economic Review*, 43(4), 1249–1487.

G 9001	J. FAYOLLE et M. FLEURBAEY Accumulation, profitabilité et endettement des entreprises		Macro-economic import functions with imperfect competition - An application to the E.C. Trade		françaises : une évaluation empirique des théories de la structure optimale du capital	G 9412	J. BOURDIEU - B. CŒURÉ - B. COLIN-SEDILLOT Investissement, incertitude et irréversibilité Quelques développements récents de la théorie de l'investissement
G 9002	H. ROUSSE Détection et effets de la multicollinéarité dans les modèles linéaires ordinaires - Un prolongement de la réflexion de BELSLEY, KUH et WELSCH	G 9203	I. STAPIC Les échanges internationaux de services de la France dans le cadre des négociations multilatérales du GATT Juin 1992 (1ère version) Novembre 1992 (version finale)	G 9312	L. BLOCH - B. CŒURÉ Q de Tobin marginal et transmission des chocs financiers	G 9413	B. DORMONT - M. PAUCHET L'évaluation de l'élasticité emploi-salaire dépendelle des structures de qualification ?
G 9003	P. RALLE et J. TOUJAS-BERNATE Indexation des salaires : la rupture de 1983	G 9204	P. SEVESTRE L'économétrie sur données individuelles-temporelles. Une note introductive	G 9313	Equipes Amadeus (INSEE), Banque de France, Métrix (DP) Présentation des propriétés des principaux modèles macroéconomiques du Service Public	G 9414	I. KABLA Le Choix de breveter une invention
G 9004	D. GUELLEC et P. RALLE Compétitivité, croissance et innovation de produit	G 9205	H. ERKEL-ROUSSE Le commerce extérieur et l'environnement international dans le modèle AMADEUS (réestimation 1992)	G 9314	B. CREPON - E. DUGUET Research & Development, competition and innovation	G 9501	J. BOURDIEU - B. CŒURÉ - B. SEDILLOT Irreversible Investment and Uncertainty : When is there a Value of Waiting ?
G 9005	P. RALLE et J. TOUJAS-BERNATE Les conséquences de la désindexation. Analyse dans une maquette prix-salaires	G 9206	N. GREENAN et D. GUELLEC Coordination within the firm and endogenous growth	G 9315	B. DORMONT Quelle est l'influence du coût du travail sur l'emploi ?	G 9502	L. BLOCH - B. CŒURÉ Imperfections du marché du crédit, investissement des entreprises et cycle économique
G 9101	Equipe AMADEUS Le modèle AMADEUS - Première partie - Présentation générale	G 9207	A. MAGNIER et J. TOUJAS-BERNATE Technology and trade : empirical evidences for the major five industrialized countries	G 9316	D. BLANCHET - C. BROUSSE Deux études sur l'âge de la retraite	G 9503	D. GOUX - E. MAURIN Les transformations de la demande de travail par qualification en France Une étude sur la période 1970-1993
G 9102	J.L. BRILLET Le modèle AMADEUS - Deuxième partie - Propriétés variantielles	G 9208	B. CREPON, E. DUGUET, D. ENCAOUA et P. MOHNEN Cooperative, non cooperative R & D and optimal patent life	G 9317	D. BLANCHET Répartition du travail dans une population hétérogène : deux notes	G 9504	N. GREENAN Technologie, changement organisationnel, qualifications et emploi : une étude empirique sur l'industrie manufacturière
G 9103	D. GUELLEC et P. RALLE Endogenous growth and product innovation	G 9209	B. CREPON et E. DUGUET Research and development, competition and innovation : an application of pseudo maximum likelihood methods to Poisson models with heterogeneity	G 9318	D. EYSSARTIER - N. PONTY AMADEUS - an annual macro-economic model for the medium and long term	G 9505	D. GOUX - E. MAURIN Persistence des hiérarchies sectorielles de salaires: un réexamen sur données françaises
G 9104	H. ROUSSE Le modèle AMADEUS - Troisième partie - Le commerce extérieur et l'environnement international	G 9301	J. TOUJAS-BERNATE Commerce international et concurrence imparfaite : développements récents et implications pour la politique commerciale	G 9319	G. CETTE - Ph. CUNÉO - D. EYSSARTIER - J. GAUTIÉ Les effets sur l'emploi d'un abaissement du coût du travail des jeunes	G 9505	D. GOUX - E. MAURIN Bis Persistence of inter-industry wages differentials: a reexamination on matched worker-firm panel data
G 9105	H. ROUSSE Effets de demande et d'offre dans les résultats du commerce extérieur manufacturé de la France au cours des deux dernières décennies	G 9302	Ch. CASES Durées de chômage et comportements d'offre de travail : une revue de la littérature	G 9401	D. BLANCHET Les structures par âge importent-elles ?	G 9506	S. JACOBZONE Les liens entre RMI et chômage, une mise en perspective <i>NON PARU - article sorti dans Economie et Prévision n°122 (1996) - pages 95 à 113</i>
G 9106	B. CREPON Innovation, taille et concentration : causalités et dynamiques	G 9303	H. ERKEL-ROUSSE Union économique et monétaire : le débat économique	G 9402	J. GAUTIÉ Le chômage des jeunes en France : problème de formation ou phénomène de file d'attente ? Quelques éléments du débat	G 9507	G. CETTE - S. MAHFOUZ Le partage primaire du revenu Constat descriptif sur longue période
G 9107	B. AMABLE et D. GUELLEC Un panorama des théories de la croissance endogène	G 9304	N. GREENAN - D. GUELLEC / G. BROUSSAUDIER - L. MIOTTI Innovation organisationnelle, dynamisme technologique et performances des entreprises	G 9403	P. QUIRION Les déchets en France : éléments statistiques et économiques	G 9601	Banque de France - CEPREMAP - Direction de la Prévision - Erasme - INSEE - OFCE Structures et propriétés de cinq modèles macroéconomiques français
G 9108	M. GLAUDE et M. MOUTARDIER Une évaluation du coût direct de l'enfant de 1979 à 1989	G 9305	P. JAILLARD Le traité de Maastricht : présentation juridique et historique	G 9404	D. LADIRAY - M. GRUN-REHOMME Lissage par moyennes mobiles - Le problème des extrémités de série	G 9602	Rapport d'activité de la DESE de l'année 1995
G 9109	P. RALLE et alii France - Allemagne : performances économiques comparées	G 9306	J.L. BRILLET Micro-DMS : présentation et propriétés	G 9405	V. MAILLARD Théorie et pratique de la correction des effets de jours ouvrables	G 9603	J. BOURDIEU - A. DRAZNIKES L'octroi de crédit aux PME : une analyse à partir d'informations bancaires
G 9110	J.L. BRILLET Micro-DMS NON PARU	G 9307	J.L. BRILLET Micro-DMS - variantes : les tableaux	G 9406	F. ROSENWALD La décision d'investir	G 9604	A. TOPIOL-BENSAÏD Les implantations japonaises en France
G 9111	A. MAGNIER Effets accélérateur et multiplicateur en France depuis 1970 : quelques résultats empiriques	G 9308	S. JACOBZONE Les grands réseaux publics français dans une perspective européenne	G 9407	S. JACOBZONE Les apports de l'économie industrielle pour définir la stratégie économique de l'hôpital public	G 9605	P. GENIER - S. JACOBZONE Comportements de prévention, consommation d'alcool et tabagie : peut-on parler d'une gestion globale du capital santé ? <i>Une modélisation microéconométrique empirique</i>
G 9112	B. CREPON et G. DUREAU Investissement en recherche-développement : analyse de causalités dans un modèle d'accélérateur généralisé	G 9309	L. BLOCH - B. CŒURE Profitabilité de l'investissement productif et transmission des chocs financiers	G 9408	L. BLOCH, J. BOURDIEU, B. COLIN-SEDILLOT, G. LONGUEVILLE Du défaut de paiement au dépôt de bilan : les banquiers face aux PME en difficulté	G 9606	C. DOZ - F. LENGART Factor analysis and unobserved component models: an application to the study of French business surveys
G 9113	J.L. BRILLET, H. ERKEL-ROUSSE, J. TOUJAS-BERNATE "France-Allemagne Couplées" - Deux économies vues par une maquette macro-économétrique	G 9310	J. BOURDIEU - B. COLIN-SEDILLOT Les théories sur la structure optimal du capital : quelques points de repère	G 9409	D. EYSSARTIER, P. MAIRE Impacts macro-économiques de mesures d'aide au logement - quelques éléments d'évaluation	G 9607	N. GREENAN - D. GUELLEC La théorie coopérative de la firme
G 9201	W.J. ADAMS, B. CREPON, D. ENCAOUA Choix technologiques et stratégies de dissuasion d'entrée	G 9311	J. BOURDIEU - B. COLIN-SEDILLOT Les décisions de financement des entreprises	G 9410	F. ROSENWALD Suivi conjoncturel de l'investissement		
G 9202	J. OLIVEIRA-MARTINS, J. TOUJAS-BERNATE			G 9411	C. DEFEUILLEY - Ph. QUIRION Les déchets d'emballages ménagers : une analyse économique des politiques française et allemande		

G 9608	N. GREENAN - D. GUELLEC Technological innovation and employment reallocation
G 9609	Ph. COUR - F. RUPPRECHT L'intégration asymétrique au sein du continent américain : un essai de modélisation
G 9610	S. DUCHENE - G. FORGEOT - A. JACQUOT Analyse des évolutions récentes de la productivité apparente du travail
G 9611	X. BONNET - S. MAHFOUZ The influence of different specifications of wages-prices spirals on the measure of the NAIRU : the case of France
G 9612	PH. COUR - E. DUBOIS, S. MAHFOUZ, J. PISANI-FERRY The cost of fiscal retrenchment revisited: how strong is the evidence ?
G 9613	A. JACQUOT Les flexions des taux d'activité sont-elles seulement conjoncturelles ?
G 9614	ZHANG Yingxiang - SONG Xueqing Lexique macroéconomique Français-Chinois
G 9701	J.L. SCHNEIDER La taxe professionnelle : éléments de cadrage économique
G 9702	J.L. SCHNEIDER Transition et stabilité politique d'un système redistributif
G 9703	D. GOUX - E. MAURIN Train or Pay: Does it Reduce Inequalities to Encourage Firms to Train their Workers?
G 9704	P. GENIER Deux contributions sur dépendance et équité
G 9705	E. DUGUET - N. IUNG R & D Investment, Patent Life and Patent Value An Econometric Analysis at the Firm Level
G 9706	M. HOUEBINE - A. TOPIOL-BENSAÏD Les entreprises internationales en France : une analyse à partir de données individuelles
G 9707	M. HOUEBINE Polarisation des activités et spécialisation des départements en France
G 9708	E. DUGUET - N. GREENAN Le biais technologique : une analyse sur données individuelles
G 9709	J.L. BRILLET Analyzing a small French ECM Model
G 9710	J.L. BRILLET Formalizing the transition process : scenarios for capital accumulation
G 9711	G. FORGEOT - J. GAUTIÉ Insertion professionnelle des jeunes et processus de déclassement
G 9712	E. DUBOIS High Real Interest Rates: the Consequence of a Saving Investment Disequilibrium or of an insufficient Credibility of Monetary Authorities?
G 9713	Bilan des activités de la Direction des Etudes et Synthèses Economiques - 1996
G 9714	F. LEQUILLER Does the French Consumer Price Index Overstate Inflation?
G 9715	X. BONNET Peut-on mettre en évidence les rigidités à la baisse des salaires nominaux ? Une étude sur quelques grands pays de l'OCDE
G 9716	N. IUNG - F. RUPPRECHT Productivité de la recherche et rendements d'échelle dans le secteur pharmaceutique français
G 9717	E. DUGUET - I. KABLA Appropriation strategy and the motivations to use the patent system in France - An econometric analysis at the firm level
G 9718	L.P. PELÉ - P. RALLE Âge de la retraite : les aspects incitatifs du régime général
G 9719	ZHANG Yingxiang - SONG Xueqing Lexique macroéconomique français-chinois, chinois-français
G 9720	M. HOUEBINE - J.L. SCHNEIDER Mesurer l'influence de la fiscalité sur la localisation des entreprises
G 9721	A. MOURougane Crédibilité, indépendance et politique monétaire Une revue de la littérature
G 9722	P. AUGERAUD - L. BRIOT Les données comptables d'entreprises Le système intermédiaire d'entreprises Passage des données individuelles aux données sectorielles
G 9723	P. AUGERAUD - J.E. CHAPRON Using Business Accounts for Compiling National Accounts: the French Experience
G 9724	P. AUGERAUD Les comptes d'entreprise par activités - Le passage aux comptes - De la comptabilité d'entreprise à la comptabilité nationale - A <i>paraître</i>
G 9801	H. MICHAUDON - C. PRIGENT Présentation du modèle AMADEUS
G 9802	J. ACCARDO Une étude de comptabilité générationnelle pour la France en 1996
G 9803	X. BONNET - S. DUCHÈNE Apports et limites de la modélisation « Real Business Cycles »
G 9804	C. BARLET - C. DUGUET - D. ENCAOUA - J. PRADEL The Commercial Success of Innovations An econometric analysis at the firm level in French manufacturing
G 9805	P. CAHUC - Ch. GIANELLA - D. GOUX - A. ZILBERBERG Equalizing Wage Differences and Bargaining Power - Evidence from a Panel of French Firms
G 9806	J. ACCARDO - M. JLASSI La productivité globale des facteurs entre 1975 et 1996
G 9807	Bilan des activités de la Direction des Etudes et Synthèses Economiques - 1997

G 9808	A. MOURougane Can a Conservative Governor Conduct an Accommodative Monetary Policy ?
G 9809	X. BONNET - E. DUBOIS - L. FAUVET Asymétrie des inflations relatives et menus costs : tests sur l'inflation française
G 9810	E. DUGUET - N. IUNG Sales and Advertising with Spillovers at the firm level: Estimation of a Dynamic Structural Model on Panel Data
G 9811	J.P. BERTHIER Congestion urbaine : un modèle de trafic de pointe à courbe débit-vitesse et demande élastique
G 9812	C. PRIGENT La part des salaires dans la valeur ajoutée : une approche macroéconomique
G 9813	A.Th. AERTS L'évolution de la part des salaires dans la valeur ajoutée en France reflète-t-elle les évolutions individuelles sur la période 1979-1994 ?
G 9814	B. SALANIÉ Guide pratique des séries non-stationnaires
G 9901	S. DUCHÈNE - A. JACQUOT Une croissance plus riche en emplois depuis le début de la décennie ? Une analyse en comparaison internationale
G 9902	Ch. COLIN Modélisation des carrières dans Destinie
G 9903	Ch. COLIN Evolution de la dispersion des salaires : un essai de prospective par microsimulation
G 9904	B. CREPON - N. IUNG Innovation, emploi et performances
G 9905	B. CREPON - Ch. GIANELLA Wages inequalities in France 1969-1992 An application of quantile regression techniques
G 9906	C. BONNET - R. MAHIEU Microsimulation techniques applied to inter-generational transfers - Pensions in a dynamic framework: the case of France
G 9907	F. ROSENWALD L'impact des contraintes financières dans la décision d'investissement
G 9908	Bilan des activités de la DESE - 1998
G 9909	J.P. ZOYEM Contrat d'insertion et sortie du RMI Evaluation des effets d'une politique sociale
G 9910	Ch. COLIN - FI. LEGROS - R. MAHIEU Bilans contributifs comparés des régimes de retraite du secteur privé et de la fonction publique
G 9911	G. LAROQUE - B. SALANIÉ Une décomposition du non-emploi en France
G 9912	B. SALANIÉ Une maquette analytique de long terme du marché du travail
G 9912 Bis	Ch. GIANELLA Une estimation de l'élasticité de l'emploi peu qualifié à son coût
G 9913	Division « Redistribution et Politiques Sociales » Le modèle de microsimulation dynamique DESTINIE
G 9914	E. DUGUET Macro-commandes SAS pour l'économétrie des panels et des variables qualitatives
G 9915	R. DUHAUTOIS Evolution des flux d'emplois en France entre 1990 et 1996 : une étude empirique à partir du fichier des bénéficiaires réels normaux (BRN)
G 9916	J.Y. FOURNIER Extraction du cycle des affaires : la méthode de Baxter et King
G 9917	B. CRÉPON - R. DESPLATZ - J. MAIRESSE Estimating price cost margins, scale economies and workers' bargaining power at the firm level
G 9918	Ch. GIANELLA - Ph. LAGARDE Productivity of hours in the aggregate production function: an evaluation on a panel of French firms from the manufacturing sector
G 9919	S. AUDRIC - P. GIVORD - C. PROST Evolution de l'emploi et des coûts par qualification entre 1982 et 1996
G 2000/01	R. MAHIEU Les déterminants des dépenses de santé : une approche macroéconomique
G 2000/02	C. ALLARD-PRIGENT - H. GUILMEAU - A. QUINET The real exchange rate as the relative price of nontradables in terms of tradables: theoretical investigation and empirical study on French data
G 2000/03	J.-Y. FOURNIER L'approximation du filtre passe-bande proposée par Christiano et Fitzgerald
G 2000/04	Bilan des activités de la DESE - 1999
G 2000/05	B. CREPON - F. ROSENWALD Investissement et contraintes de financement : le poids du cycle Une estimation sur données françaises
G 2000/06	A. FLIPO Les comportements matrimoniaux de fait
G 2000/07	R. MAHIEU - B. SÉDILLOT Microsimulations of the retirement decision: a supply side approach
G 2000/08	C. AUDENIS - C. PROST Déficit conjoncturel : une prise en compte des conjonctures passées
G 2000/09	R. MAHIEU - B. SÉDILLOT Equivalent patrimonial de la rente et souscription de retraite complémentaire
G 2000/10	R. DUHAUTOIS Ralentissement de l'investissement : petites ou grandes entreprises ? industrie ou tertiaire ?
G 2000/11	G. LAROQUE - B. SALANIÉ Temps partiel féminin et incitations financières à l'emploi
G2000/12	Ch. GIANELLA Local unemployment and wages
G2000/13	B. CREPON - Th. HECKEL - Informatisation en France : une évaluation à partir de données individuelles

	- Computerization in France: an evaluation based on individual company data	G2002/01	F. MAGNIEN - J.-L. TAVERNIER - D. THESMAR Les statistiques internationales de PIB par habitant en standard de pouvoir d'achat : une analyse des résultats	G2002/16	F. MAUREL - S. GREGOIR Les indices de compétitivité des pays : interprétation et limites	G2004/06	M. DUÉE L'impact du chômage des parents sur le devenir scolaire des enfants
G2001/01	F. LEQUILLER - La nouvelle économie et la mesure de la croissance du PIB - The new economy and the measurement of GDP growth	G2002/02	Bilan des activités de la DESE - 2001	G2003/01	N. RIEDINGER - E. HAUVY Le coût de dépollution atmosphérique pour les entreprises françaises : Une estimation à partir de données individuelles	G2004/07	P. AUBERT - E. CAROLI - M. ROGER New Technologies, Workplace Organisation and the Age Structure of the Workforce: Firm-Level Evidence
G2001/02	S. AUDRIC La reprise de la croissance de l'emploi profite-t-elle aussi aux non-diplômés ?	G2002/03	B. SÉDILLOT - E. WALRAET La cessation d'activité au sein des couples : y a-t-il interdépendance des choix ?	G2003/02	P. BISCOURP et F. KRAMARZ Création d'emplois, destruction d'emplois et internationalisation des entreprises industrielles françaises : une analyse sur la période 1986-1992	G2004/08	E. DUGUET - C. LELARGE Les brevets accroissent-ils les incitations privées à innover ? Un examen microéconométrique
G2001/03	I. BRAUN-LEMAIRE Evolution et répartition du surplus de productivité	G2002/04	G. BRILHAULT - Rétropolation des séries de FBCF et calcul du capital fixe en SEC-95 dans les comptes nationaux français - Retropolation of the investment series (GFCF) and estimation of fixed capital stocks on the ESA-95 basis for the French balance sheets	G2003/03	Bilan des activités de la DESE - 2002	G2004/09	S. RASPILLER - P. SILLARD Affiliating versus Subcontracting: the Case of Multinationals
G2001/04	A. BEAUDU - Th. HECKEL Le canal du crédit fonctionne-t-il en Europe ? Une étude de l'hétérogénéité des comportements d'investissement à partir de données de bilan agrégées	G2002/05	P. BISCOURP - B. CRÉPON - T. HECKEL - N. RIEDINGER How do firms respond to cheaper computers? Microeconomic evidence for France based on a production function approach	G2003/04	P.-O. BEFFY - J. DERUYON - N. FOURCADE - S. GREGOIR - N. LAÏB - B. MONFORT Évolutions démographiques et croissance : une projection macro-économique à l'horizon 2020	G2004/10	J. BOISSINOT - C. L'ANGEVIN - B. MONFORT Public Debt Sustainability: Some Results on the French Case
G2001/05	C. AUDENIS - P. BISCOURP - N. FOURCADE - O. LOISEL Testing the augmented Solow growth model : An empirical reassessment using panel data	G2002/06	C. AUDENIS - J. DERUYON - N. FOURCADE L'impact des nouvelles technologies de l'information et de la communication sur l'économie française - un bouclage macro-économique	G2003/05	P. AUBERT La situation des salariés de plus de cinquante ans dans le secteur privé	G2004/11	S. ANANIAN - P. AUBERT Travailleurs âgés, nouvelles technologies et changements organisationnels : un réexamen à partir de l'enquête « REPONSE »
G2001/06	R. MAHIEU - B. SÉDILLOT Départ à la retraite, irréversibilité et incertitude	G2002/07	J. BARDAJI - B. SÉDILLOT - E. WALRAET Évaluation de trois réformes du Régime Général d'assurance vieillesse à l'aide du modèle de microsimulation DESTINIE	G2003/06	P. AUBERT - B. CRÉPON Age, salaire et productivité La productivité des salariés décline-t-elle en fin de carrière ?	G2004/12	X. BONNET - H. PONCET Structures de revenus et propensions différentes à consommer - Vers une équation de consommation des ménages plus robuste en prévision pour la France
G2001/07	Bilan des activités de la DESE - 2000	G2002/08	J.-P. BERTHIER Réflexions sur les différentes notions de volume dans les comptes nationaux : comptes aux prix d'une année fixe ou aux prix de l'année précédente, séries chaînées	G2003/07	H. BARON - P.O. BEFFY - N. FOURCADE - R. MAHIEU Le ralentissement de la productivité du travail au cours des années 1990	G2004/13	C. PICART Évaluer la rentabilité des sociétés non financières
G2001/08	J. Ph. GAUDEMET Les dispositifs d'acquisition à titre facultatif d'annuités viagères de retraite	G2002/09	F. HILD Les soldes d'opinion résumant-ils au mieux les réponses des entreprises aux enquêtes de conjoncture ?	G2003/08	P.-O. BEFFY - B. MONFORT Patrimoine des ménages, dynamique d'allocation et comportement de consommation	G2004/14	J. BARDAJI - B. SÉDILLOT - E. WALRAET Les retraites du secteur public : projections à l'horizon 2040 à l'aide du modèle de microsimulation DESTINIE
G2001/09	B. CRÉPON - Ch. GIANELLA Fiscalité, coût d'usage du capital et demande de facteurs : une analyse sur données individuelles	G2002/10	I. ROBERT-BOBÉE Les comportements démographiques dans le modèle de microsimulation Destinie - Une comparaison des estimations issues des enquêtes Jeunes et Carrières 1997 et Histoire Familiale 1999	G2003/09	P. BISCOURP - N. FOURCADE Peut-on mettre en évidence l'existence de rigidités à la baisse des salaires à partir de données individuelles ? Le cas de la France à la fin des années 90	G2005/01	S. BUFFETEAU - P. GODEFROY Conditions de départ en retraite selon l'âge de fin d'études : analyse prospective pour les générations 1945 à 1974
G2001/10	B. CRÉPON - R. DESPLATZ Evaluation des effets des dispositifs d'allègements de charges sociales sur les bas salaires	G2002/11	J.-P. ZOYEM La dynamique des bas revenus : une analyse des entrées-sorties de pauvreté	G2003/10	M. LECLAIR - P. PETIT Présence syndicale dans les firmes : quel impact sur les inégalités salariales entre les hommes et les femmes ?	G2005/02	C. AFSA - S. BUFFETEAU L'évolution de l'activité féminine en France : une approche par pseudo-panel
G2001/11	J.-Y. FOURNIER Comparaison des salaires des secteurs public et privé	G2002/12	F. HILD Prévisions d'inflation pour la France	G2003/11	P.-O. BEFFY - X. BONNET - M. DARRACQ-PARIES - B. MONFORT MZE: a small macro-model for the euro area	G2005/03	P. AUBERT - P. SILLARD Délocalisations et réductions d'effectifs dans l'industrie française
G2001/12	J.-P. BERTHIER - C. JAULENT R. CONVENEVOLE - S. PISANI Une méthodologie de comparaison entre consommations intermédiaires de source fiscale et de comptabilité nationale	G2002/13	M. LECLAIR Réduction du temps de travail et tensions sur les facteurs de production	G2004/01	P. AUBERT - M. LECLAIR La compétitivité exprimée dans les enquêtes trimestrielles sur la situation et les perspectives dans l'industrie	G2005/04	M. LECLAIR - S. ROUX Mesure et utilisation des emplois instables dans les entreprises
G2001/13	P. BISCOURP - Ch. GIANELLA Substitution and complementarity between capital, skilled and less skilled workers: an analysis at the firm level in the French manufacturing industry	G2002/14	E. WALRAET - A. VINCENT - Analyse de la redistribution intragénérationnelle dans le système de retraite des salariés du privé - Une approche par microsimulation - Intragenerational distributional analysis in the french private sector pension scheme - A microsimulation approach	G2004/02	M. DUÉE - C. REBILLARD La dépendance des personnes âgées : une projection à long terme	G2005/05	C. L'ANGEVIN - S. SERRAVALLE Performances à l'exportation de la France et de l'Allemagne - Une analyse par secteur et destination géographique
G2001/14	I. ROBERT-BOBÉE Modelling demographic behaviours in the French microsimulation model Destinie: An analysis of future change in completed fertility	G2002/15	P. CHONE - D. LE BLANC - I. ROBERT-BOBÉE Offre de travail féminine et garde des jeunes enfants	G2004/03	S. RASPILLER - N. RIEDINGER Régulation environnementale et choix de localisation des groupes français	G2005/06	Bilan des activités de la Direction des Études et Synthèses Économiques - 2004
G2001/15	J.-P. ZOYEM Diagnostic sur la pauvreté et calendrier de revenus : le cas du "Panel européen des ménages"			G2004/04	A. NABOULET - S. RASPILLER Les déterminants de la décision d'investir : une approche par les perceptions subjectives des firmes	G2005/07	S. RASPILLER La concurrence fiscale : principaux enseignements de l'analyse économique
G2001/16	J.-Y. FOURNIER - P. GIVORD La réduction des taux d'activité aux âges extrêmes, une spécificité française ?			G2004/05	N. RAGACHE La déclaration des enfants par les couples non mariés est-elle fiscalement optimale ?	G2005/08	C. L'ANGEVIN - N. LAÏB Éducation et croissance en France et dans un panel de 21 pays de l'OCDE
G2001/17	C. AUDENIS - P. BISCOURP - N. RIEDINGER Existe-t-il une asymétrie dans la transmission du prix du brut aux prix des carburants ?					G2005/09	N. FERRARI Prévoir l'investissement des entreprises Un indicateur des révisions dans l'enquête de conjoncture sur les investissements dans l'industrie.

G2005/10	P.-O. BEFFY - C. L'ANGEVIN Chômage et boucle prix-salaires : apport d'un modèle « qualifiés/peu qualifiés »
G2005/11	B. HEITZ A two-states Markov-switching model of inflation in France and the USA: credible target VS inflation spiral
G2005/12	O. BIAU - H. ERKEL-ROUSSE - N. FERRARI Réponses individuelles aux enquêtes de conjoncture et prévision macroéconomiques : Exemple de la prévision de la production manufacturière
G2005/13	P. AUBERT - D. BLANCHET - D. BLAU The labour market after age 50: some elements of a Franco-American comparison
G2005/14	D. BLANCHET - T. DEBRAND - P. DOURGNON - P. POLLET L'enquête SHARE : présentation et premiers résultats de l'édition française
G2005/15	M. DUÉE La modélisation des comportements démographiques dans le modèle de microsimulation DESTINIE
G2005/16	H. RAOUI - S. ROUX Étude de simulation sur la participation versée aux salariés par les entreprises
G2006/01	C. BONNET - S. BUFFETEAU - P. GODEFROY Disparités de retraite de droit direct entre hommes et femmes : quelles évolutions ?
G2006/02	C. PICART Les gazelles en France
G2006/03	P. AUBERT - B. CRÉPON - P. ZAMORA Le rendement apparent de la formation continue dans les entreprises : effets sur la productivité et les salaires
G2006/04	J.-F. OUVRARD - R. RATHELOT Demographic change and unemployment: what do macroeconomic models predict?
G2006/05	D. BLANCHET - J.-F. OUVRARD Indicateurs d'engagements implicites des systèmes de retraite : chiffres, propriétés analytiques et réactions à des chocs démographiques types
G2006/06	G. BIAU - O. BIAU - L. ROUVIERE Nonparametric Forecasting of the Manufacturing Output Growth with Firm-level Survey Data
G2006/07	C. AFSA - P. GIVORD Le rôle des conditions de travail dans les absences pour maladie
G2006/08	P. SILLARD - C. L'ANGEVIN - S. SERRAVALLE Performances comparées à l'exportation de la France et de ses principaux partenaires Une analyse structurelle sur 12 ans
G2006/09	X. BOUTIN - S. QUANTIN Une méthodologie d'évaluation comptable du coût du capital des entreprises françaises : 1984-2002
G2006/10	C. AFSA L'estimation d'un coût implicite de la pénibilité du travail chez les travailleurs âgés
G2006/11	C. LELARGE Les entreprises (industrielles) françaises sont-elles à la frontière technologique ?
G2006/12	O. BIAU - N. FERRARI Théorie de l'opinion Faut-il pondérer les réponses individuelles ?
G2006/13	A. KOUBI - S. ROUX Une réinterprétation de la relation entre productivité et inégalités salariales dans les entreprises
G2006/14	R. RATHELOT - P. SILLARD The impact of local taxes on plants location decision
G2006/15	L. GONZALEZ - C. PICART Diversification, recentrage et poids des activités de support dans les groupes (1993-2000)
G2007/01	D. SRAER Allègements de cotisations patronales et dynamique salariale
G2007/02	V. ALBOUY - L. LEQUIEN Les rendements non monétaires de l'éducation : le cas de la santé
G2007/03	D. BLANCHET - T. DEBRAND Aspiration à la retraite, santé et satisfaction au travail : une comparaison européenne
G2007/04	M. BARLET - L. CRUSSON Quel impact des variations du prix du pétrole sur la croissance française ?
G2007/05	C. PICART Flux d'emploi et de main-d'œuvre en France : un réexamen
G2007/06	V. ALBOUY - C. TAVAN Massification et démocratisation de l'enseignement supérieur en France
G2007/07	T. LE BARBANCHON The Changing response to oil price shocks in France : a DSGE type approach
G2007/08	T. CHANEY - D. SRAER - D. THESMAR Collateral Value and Corporate Investment Evidence from the French Real Estate Market
G2007/09	J. BOISSINOT Consumption over the Life Cycle: Facts for France
G2007/10	C. AFSA Interpréter les variables de satisfaction : l'exemple de la durée du travail
G2007/11	R. RATHELOT - P. SILLARD Zones Franches Urbaines : quels effets sur l'emploi salarié et les créations d'établissements ?
G2007/12	V. ALBOUY - B. CRÉPON Aléa moral en santé : une évaluation dans le cadre du modèle causal de Rubin
G2008/01	C. PICART Les PME françaises : rentables mais peu dynamiques
G2008/02	P. BISCOURP - X. BOUTIN - T. VERGÉ The Effects of Retail Regulations on Prices Evidence from the Loi Galland
G2008/03	Y. BARBESOL - A. BRIANT Economies d'agglomération et productivité des

G2008/04	D. BLANCHET - F. LE GALLO Les projections démographiques : principaux mécanismes et retour sur l'expérience française
G2008/05	D. BLANCHET - F. TOUTLEMONDE Évolutions démographiques et déformation du cycle de vie active : quelles relations ?
G2008/06	M. BARLET - D. BLANCHET - L. CRUSSON Internationalisation et flux d'emplois : que dit une approche comptable ?
G2008/07	C. LELARGE - D. SRAER - D. THESMAR Entrepreneurship and Credit Constraints - Evidence from a French Loan Guarantee Program
G2008/08	X. BOUTIN - L. JANIN Are Prices Really Affected by Mergers?
G2008/09	M. BARLET - A. BRIANT - L. CRUSSON Concentration géographique dans l'industrie manufacturière et dans les services en France : une approche par un indicateur en continu
G2008/10	M. BEFFY - É. COUDIN - R. RATHELOT Who is confronted to insecure labor market histories? Some evidence based on the French labor market transition
G2008/11	M. ROGER - E. WALRAET Social Security and Well-Being of the Elderly: the Case of France
G2008/12	C. AFSA Analyser les composantes du bien-être et de son évolution Une approche empirique sur données individuelles
G2008/13	M. BARLET - D. BLANCHET - T. LE BARBANCHON Microsimuler le marché du travail : un prototype
G2009/01	P.-A. PIONNIER Le partage de la valeur ajoutée en France, 1949-2007
G2009/02	Laurent CLAVEL - Christelle MINODIER A Monthly Indicator of the French Business Climate
G2009/03	H. ERKEL-ROUSSE - C. MINODIER Do Business Tendency Surveys in Industry and Services Help in Forecasting GDP Growth? A Real-Time Analysis on French Data
G2009/04	P. GIVORD - L. WILNER Les contrats temporaires : trappe ou marchepied vers l'emploi stable ?
G2009/05	G. LALANNE - P.-A. PIONNIER - O. SIMON Le partage des fruits de la croissance de 1950 à 2008 : une approche par les comptes de surplus
G2009/06	L. DAVEZIES - X. D'HAULTFOEUILLE Faut-il pondérer ?... Ou l'éternelle question de l'économètre confronté à des données d'enquête
G2009/07	S. QUANTIN - S. RASPILLER - S. SERRAVALLE Commerce intragroupe, fiscalité et prix de transferts : une analyse sur données françaises
G2009/08	M. CLERC - V. MARCUS Élasticités-prix des consommations énergétiques des ménages
G2009/09	G. LALANNE - E. POULIQUEN - O. SIMON Prix du pétrole et croissance potentielle à long terme
G2009/10	D. BLANCHET - J. LE CACHEUX - V. MARCUS Adjusted net savings and other approaches to sustainability: some theoretical background
G2009/11	V. BELLAMY - G. CONSALES - M. FESSEAU - S. LE LAIDIER - É. RAYNAUD Une décomposition du compte des ménages de la comptabilité nationale par catégorie de ménage en 2003
G2009/12	J. BARDAJI - F. TALLET Detecting Economic Regimes in France : a Qualitative Markov-Switching Indicator Using Mixed Frequency Data
G2009/13	R. AEBERHARDT - D. FOUGÈRE - R. RATHELOT Discrimination à l'embauche : comment exploiter les procédures de <i>testing</i> ?
G2009/14	Y. BARBESOL - P. GIVORD - S. QUANTIN Partage de la valeur ajoutée, approche par données microéconomiques
G2009/15	I. BUONO - G. LALANNE The Effect of the Uruguay round on the Intensive and Extensive Margins of Trade
G2010/01	C. MINODIER Avantages comparés des séries des premières valeurs publiées et des séries des valeurs révisées - Un exercice de prévision en temps réel de la croissance trimestrielle du PIB en France
G2010/02	V. ALBOUY - L. DAVEZIES - T. DEBRAND Health Expenditure Models: a Comparison of Five Specifications using Panel Data
G2010/03	C. KLEIN - O. SIMON Le modèle MÉSANGE réestimé en base 2000 Tome 1 – Version avec volumes à prix constants
G2010/04	M.-É. CLERC - É. COUDIN L'IPC, miroir de l'évolution du coût de la vie en France ? Ce qu'apporte l'analyse des courbes d'Engel
G2010/05	N. CECI-RENAUD - P.-A. CHEVALIER Les seuils de 10, 20 et 50 salariés : impact sur la taille des entreprises françaises
G2010/06	R. AEBERHARDT - J. POUGET National Origin Differences in Wages and Hierarchical Positions - Evidence on French Full-Time Male Workers from a matched Employer-Employee Dataset
G2010/07	S. BLASCO - P. GIVORD Les trajectoires professionnelles en début de vie active : quel impact des contrats temporaires ?
G2010/08	P. GIVORD Méthodes économétriques pour l'évaluation de politiques publiques