Document généré le 3 juin 2025 14:11

L'Actualité économique

Identification partielle : méthodes et conséquences pour les applications empiriques

Thierry Magnac

Volume 89, numéro 4, décembre 2013

URI : https://id.erudit.org/iderudit/1026444ar
DOI : https://doi.org/10.7202/1026444ar

Aller au sommaire du numéro

Éditeur(s)

HEC Montréal

ISSN

0001-771X (imprimé)
1710-3991 (numérique)

Découvrir la revue

Citer cet article

Magnac, T. (2013). Identification partielle : méthodes et conséquences pour les applications empiriques. *L'Actualité économique*, *89*(4), 233–258. https://doi.org/10.7202/1026444ar

Tous droits réservés © HEC Montréal, 2014

Ce document est protégé par la loi sur le droit d'auteur. L'utilisation des services d'Érudit (y compris la reproduction) est assujettie à sa politique d'utilisation que vous pouvez consulter en ligne.

https://apropos.erudit.org/fr/usagers/politique-dutilisation/

Cet article est diffusé et préservé par Érudit.

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche.

https://www.erudit.org/fr/

L'Actualité économique, Revue d'analyse économique, vol. 89, nº 4, décembre 2013

IDENTIFICATION PARTIELLE : MÉTHODES ET CONSÉQUENCES POUR LES APPLICATIONS EMPIRIQUES*

Thierry MAGNAC
École d'Économie de Toulouse (TSE)
thierry.magnac@tse-fr.eu

« *Yet, there is enormous scope for fruitful inference using weaker and more credible assumptions that partially identify parameters.* »

Charles Manski (2005)**,** *Partial Identification in Econometrics,* New Palgrave Dictionary.

« *The partial identification approach advocates a more fluid view of identification and hence provides the empirical researcher with methods to help study the spectrum of information that we can harness about a parameter of interest using a menu of assumptions.* »

Elie Tamer (2010)**,** « Partial Identification in Econometrics », *Annual Review of Economics*, 2 : 167-195

Introduction

La démarche empirique (en économie et ailleurs) est une démarche interprétative puisque les données économiques observationnelles ou expérimentales ne sont pas interprétables de façon brute. On interprète ces données très couramment de façon implicite, ou plus rarement de façon explicite, en utilisant des modèles économiques. Ces modèles, qu'ils soient sous forme structurelle ou réduite, sont écrits en fonction de *paramètres* qui peuvent inclure de façon générale des paramètres fonctionnels comme des distributions de termes d'hétérogénéité inobservable des agents. Dans

* Cette revue fait suite à une présentation invitée à la réunion de la Société canadienne de science économique à Québec en mai 2013. Je remercie Bruce Shearer, Christian Bontemps, Marc Henry, Eric Maurin et les participants à cette conférence pour leurs remarques. La responsabilité des erreurs et omissions de ce texte m'incombe entièrement.

la plupart des cas, seul un sous-ensemble de ces paramètres est interprétable économiquement et constitue les paramètres dits d'intérêt.

Au niveau abstrait d'une population de taille infinie dans laquelle sont tirés les échantillons de travail, les *données* sont décrites sous forme réduite par des fonctions de répartition des variables économiques, ces variables étant choisies par l'économétre appliqué. La démarche d'*identification* consiste alors en la mise en relation de ces fonctions de répartition avec celles qui sont construites à l'aide des modèles structurels pour en déduire les valeurs des paramètres d'intérêt. Si cette relation est univoque, on dit que les paramètres sont identifiés ponctuellement et c'est cette présentation qu'on trouve dans les manuels avancés d'économétrie (par exemple Greene, 2008).

L'importance de cette notion d'identification s'est vue remise en cause dans les 20 dernières années par Manski et ses coauteurs (1989 et années subséquentes) dans des modèles de traitements, de sélection ou de censure et cette remise en cause a donné lieu à une littérature qui s'amplifie dans les années récentes. Des travaux précurseurs sur l'identification ensembliste ou partielle avaient été développés dès Gini (1921) et Frisch (1934) pour le modèle de régression simple à erreurs de mesure, Marschak et Andrews (1944) pour les modèles à équations simultanées, Hoeffding (1943) et Fréchet (1955) pour les bornes sur les distributions jointes de variables dont seules les distributions marginales sont observables (dans deux enquêtes différentes par exemple) et par Klepper et Leamer (1984) et Leamer (1987), par exemple pour le modèle de régression multiple à erreurs de mesure sur toutes les variables. Ces travaux étaient restés peu connus et peu utilisés jusqu'aux travaux menés par Manski et que lui-même a repris dans un livre (Manski, 2003) ou plus brièvement dans un article dont la citation en exergue de cette revue est extraite. De nombreux étudiants de Manski contribueront à développer cette littérature et parmi eux, Elie Tamer et Francesca Molinari pour ne citer qu'eux (par exemple la revue de Tamer, 2010, et l'article fondateur de Beresteanu, Molinari et Molchanov, 2011).

Le raisonnement général qui mène à l'*identification partielle* part de la notion d'*incomplétude* des données ou des modèles. En premier lieu, les données peuvent être *incomplètes* à cause de mécanismes de censure, de l'utilisation de deux bases de données différentes ou de deux états exclusifs. Pour l'évaluation de politiques publiques, les données observationnelles sont incomplètes puisque les individus ne peuvent jamais être observés simultanément en traitement et hors du traitement. Les modèles structurels peuvent aussi être *incomplets* s'ils ne spécifient pas de solutions univoques. Un exemple classique de ce cas de figure est fourni par les équilibres multiples dans les jeux (Tamer, 2003, par exemple). Le modèle économique ne spécifie pas quel est le mécanisme de sélection (stochastique ou non) du seul équilibre observé dans les données.

La façon de procéder la plus commune dans la littérature appliquée est de faire des hypothèses pour compléter les données ou pour rendre complets les modèles car par construction, leurs paramètres deviennent alors identifiables ponctuellement.

Par exemple, on spécifiera des variables latentes complémentaires ou leurs distributions, pour compléter les données comme dans les modèles de censure ou de traitement, ou des mécanismes qui rendent univoque la solution des modèles économiques, comme un mécanisme de sélection d'équilibre dans un jeu. Néanmoins, cette procédure est arbitraire et l'identification ponctuelle devient peu crédible.

Cette démarche fournit pourtant l'intuition première de l'identification partielle. L'analyse des données se conduit en examinant TOUTES les hypothèses arbitraires acceptables, c'est-à-dire compatibles avec nos *a priori*, et en rassemblant dans un *ensemble identifié* toutes les valeurs des paramètres ponctuellement identifiées auxquelles chacune de ces hypothèses conduit. L'acceptabilité d'une hypothèse dépendra des applications et ces hypothèses seront ensemblistes (par exemple, une probabilité de sélection d'équilibre appartient à l'intervalle [0,1], ou des valeurs censurées sont bornées) ou fonctionnelles (monotonicité, concavité, *etc.*). La puissance identifiante de différentes hypothèses pourra être comparée en termes de taille de l'ensemble identifié.

L'identification partielle semble donc être fort différente du cadre traditionnel. Néanmoins, quand on revient aux autres étapes du travail empirique d'estimation et de construction d'intervalles de confiance, cette construction s'adapte naturellement. Pour l'estimation, on remplace des estimateurs ponctuels du paramètre d'intérêt par des estimateurs ensemblistes par exemple bâtis à partir de l'estimation de la frontière de l'ensemble. Pour l'inférence, la présentation des intervalles de confiance ne change pas puisque les résultats s'expriment naturellement en termes d'intervalles ou régions de confiance. Seule leur interprétation change puisque ces régions de confiance ne sont plus uniquement le résultat de la variabilité de l'échantillonnage mais aussi de l'incertitude radicale sous-jacente quant à l'identification des paramètres.

Cette présentation restera peu technique et s'appuiera sur des exemples du mieux qu'elle pourra. Le but de cet article est de passer en revue les articles présentant les méthodes (mais certains restent assez arides) pour en extraire les principes généraux et les articles rapportant les résultats d'applications empiriques dont je fais une liste partielle et brève. Je reviens d'abord sur la question de l'identification dans la section 1 pour définir plus précisément les notions évoquées dans cette introduction. Les méthodes d'inférence sont présentées dans la section 2 qui est la plus technique de cette revue. J'étudie à la fin de la section 2 des cas plus simples d'inférence quand les ensembles identifiés sont convexes, puis je donne un catalogue bien incomplet d'applications empiriques dans la section 3. La dernière section conclut cette présentation.

1. Identification

Après des définitions plus précises des concepts d'identification présentés dans l'introduction, je présente deux exemples, l'un concernant la prédiction linéaire d'une variable dépendante observée avec censure par intervalles (par exemple,

Manski et Tamer, 2002), et un jeu d'entrée entre deux entreprises dont les réalisations sont observées sur plusieurs marchés (Tamer, 2003). Il est important de souligner deux concepts généraux dans ces exemples. Premièrement, l'identification partielle a pour conséquence que les équations estimantes peuvent devenir des restrictions de moments à l'*inégalité* au lieu d'être uniquement des restrictions de moments à l'égalité. On définit aussi le concept d'identification « exacte »[1] qui permet d'affirmer que tous les points de l'ensemble identifié correspondent bien à des hypothèses acceptables ou crédibles de complétion du modèle ou des données.

1.1 *Définition*

À un niveau très abstrait qui présuppose qu'on ait défini des variables économiques d'intérêt, *Y*, qui sont bien sûr multidimensionnelles, les données sont décrites en forme réduite par des fonctions de répartition de ces variables :

$$\Pr(Y \leq y), \forall y \in \mathcal{Y}.$$

Les modèles économiques sont décrits sous leur forme structurelle par des fonctions de répartition dépendantes d'un paramètre θ qui peut être aussi général que l'on veut (par exemple des fonctions de répartition) :

$$F(y;\theta), \forall y \in \mathcal{Y}_F \supset \mathcal{Y}.$$

Cette définition des fonctions de répartition structurelles peut « emprunter » une partie de leur construction aux distributions marginales des observations. Par exemple, on confondra souvent les formes structurelles ou réduites des fonctions de répartition de variables explicatives exogènes.

On peut alors définir ce qu'est la notion d'équivalence observationnelle ponctuelle. On dit que les paramètres θ et θ' sont observationnellement équivalents si et seulement si pour tout $y \in \mathcal{Y}$:[2]

$$F(y;\theta) = F(y;\theta').$$

On fera aussi ici une hypothèse de bonne spécification qui suppose qu'il existe au moins une valeur du paramètre θ_0 tel que pour tout $y \in \mathcal{Y}$:

$$\Pr(Y \leq y) = F(y;\theta_0),$$

même s'il est facile d'étendre cette notion à des modèles mal spécifiés. Néanmoins la notion d'identification partielle se prête moins bien à cette extension (Ponomareva et Tamer, 2011) et c'est pourquoi la quasi-totalité de la littérature que nous passons en revue fait implicitement ou explicitement l'hypothèse de bonne spécification.

L'hypothèse d'identification ponctuelle de θ_0 s'écrit alors comme :

$$\forall \theta, \forall y; \Pr(Y \leq y) = F(y;\theta) = F(y;\theta_0) \Rightarrow \theta = \theta_0$$

---

1. Mauvaise traduction du difficile « sharp ». D'autres propositions pourraient être « nette » ou « franche ».

2. Formellement, il suffirait de dire pour presque tout $y \in \mathcal{Y}$ pour la mesure induite par $\Pr(Y \leq y)$. Nous ne précisons pas ces détails techniques dans la suite.

et ceci de manière globale ou locale suivant les hypothèses sur le domaine de variation des θ dans cette définition[3].

Quand ces notions sont remises en cause car les modèles sont *incomplets,* on peut adopter une démarche de complétion ou d'augmentation des données en supposant que les observables sont maintenant $z = (y, t)$. Comme on l'a évoqué dans l'introduction, ces variables t complètent les données de façon arbitraire et par exemple décrivent les censures c'est-à-dire la valeur exacte de la variable dépendante dans le cas de censure par intervalle. Ces variables peuvent aussi décrire l'équilibre sélectionné dans le cas de jeux générant des équilibres multiples. Cette complétion permet de réécrire le modèle au moyen d'une nouvelle fonction de répartition $F_\eta(z; \theta)$ écrite en fonction du paramètre inconnu, η et qui est la distribution des variables t conditionnellement aux observables y. Ce paramètre n'est pas spécifié par la théorie économique ou par le modèle statistique. Par définition de la complétion du modèle, on supposera que dans le modèle complet où η est donné, le paramètre θ est identifié ponctuellement.

En l'absence d'autres restrictions, il semble difficile de progresser. On fera donc des restrictions ensemblistes du type, $t \in \mathcal{T}$, un intervalle par exemple, qui entraîne des restrictions ensemblistes pour η, $\eta \in N$. La probabilité conditionnelle (à toutes les variables exogènes) de sélection d'un équilibre dans un jeu à équilibres multiples sera bornée entre 0 et 1 par exemple. On peut alors définir l'ensemble identifié par toutes les valeurs possibles du paramètre ponctuellement identifié quand la complétion des données est décrite par η :

$$\Theta_I = \left\{\theta; \exists \eta \in N, \Pr(Y \leq y, T \leq t) = F_\eta(z;\theta), \forall z = (y,t) \in \mathcal{Y} \times \mathcal{T}\right\}$$

$$= \cup_{\eta \in N} \left\{\theta; \Pr(Z \leq z) = F_\eta(z;\theta), \forall z \in \mathcal{Z}\right\}$$

c'est-à-dire, toute valeur du paramètre d'intérêt qui peut être réconciliée avec les données pour une valeur au moins du paramètre qui décrit l'augmentation des données pour les rendre complètes.

On peut alors analyser différentes restrictions suivant leur degré de crédibilité :

$$\eta \in N_1, \eta \in N_2 \subset N_1, \ldots$$

Les deux exemples qui suivent permettent de mettre en pratique cette construction.

3. On peut aussi s'intéresser uniquement à un sous-ensemble des paramètres θ, par exemple dans le cas d'un modèle de régression. Dans ce cas, on reformule l'identification en utilisant la fonction de répartition, par exemple par l'intermédiaire de moments.

1.2 *Exemple 1 : censure par intervalles*

Manski et Tamer (2002) étudient parmi d'autres modèles, un modèle de prédiction linéaire

$$y_i^* = \beta_0 + \beta_1 x_i + u_i, \qquad E(u_i) = E(u_i x_i) = 0,$$

où la variable dépendante est censurée par intervalles :

$$y_i^* \in \left[y_{iL}, y_{iL} + \delta_i \right]$$

où seules les variables y_{iL}, $\delta_i > 0$ et x_i sont observables[4]. C'est un cas de figure assez courant quand on utilise des données sur les revenus ou la richesse des ménages car certaines enquêtes procèdent par une démarche en deux temps. On demandera d'abord aux ménages ou aux individus, le niveau exact de leurs revenus ou patrimoines puis, si les ménages ne veulent pas répondre pour des raisons de confidentialité, l'enquêteur reposera la même question mais sous la forme d'intervalles. Votre revenu est-il entre 0 et 500 dollars? ou entre 500 et 1000 dollars? *etc*. On notera dans ce cas que le support des variables y_{iL} et δ_i est discret et qu'il est construit en utilisant les bornes inférieures et les largeurs d'intervalles décidées par le concepteur de l'enquête. On notera aussi que même si 0 est une borne naturelle inférieure pour les revenus, une borne supérieure n'est pas aussi clairement définie. La plupart des études font alors une hypothèse arbitraire sur le montant maximum de la variable dépendante, par exemple le revenu dans l'intervalle observé le plus élevé (par exemple Lee, 2009, quand il évalue les bornes les plus conservatrices).

Pour simplifier les notations, on supposera que $Ex_i = 0$ et on ne s'intéressera qu'à l'identification du paramètre β_1. Pour les extensions au modèle de prédiction linéaire à plusieurs variables, on pourra se référer à la section 2.3 où les résultats de Stoye (2007), Beresteanu et Molinari (2008) et Bontemps, Magnac et Maurin (2012) sont brièvement développés.

On complète maintenant les données pour se ramener au cas d'identification ponctuelle. Comme :

$$y_{iL} \leq y_i^* \leq y_{iL} + \delta_i,$$

on augmente les données par :

$$t_i \in E = [0,1],$$

de telle manière qu'on puisse écrire :

$$y_i^* = y_{iL} + \delta_i t_i.$$

Le chercheur est agnostique quant aux valeurs de t_i dans cet intervalle.

4. L'ensemble est fermé à droite et à gauche par simplicité alors qu'il serait naturel par la convention CADLAG de l'ouvrir à droite. On néglige dans cette littérature ces distinctions topologiques ou on les traite dans les annexes techniques. On considère en général la fermeture des ensembles identifiés.

On remarquera que η est alors la distribution conditionnelle de t_i, $\eta = F(t_i \mid y_{iL}, \delta_i, x_i)$. Le cas général inclut donc les cas d'imputation. Par exemple, dans un cas de Probit ou Logit ordonné, on spécifierait que la variable inobservée, y_i^* est normale ou logistique conditionnellement à x_i et donc que η a une distribution normale ou logistique, tronquée à l'intérieur de chaque intervalle. Sous l'une de ces hypothèses, le paramètre d'intérêt β_1 est identifié ainsi que β_0 et la variance de u_i d'ailleurs[5]. Néanmoins, l'ensemble identifié est beaucoup plus grand que ce singleton donné par le Probit ou Logit multinomial puisqu'il correspond à toutes les distributions, η, possibles.

Contrairement à l'intuition, une analyse directe du cas général est pourtant facile. On procède d'abord à l'identification dans le modèle complet puis on considère l'union de toutes les valeurs identifiées.

Comme :

$$y_i^* = y_{iL} + \delta_i t_i = \beta_0 + \beta_1 x_i + u_i,$$

on peut dériver la valeur du paramètre β_1 comme [6] :

$$\beta_1 = \frac{E((y_{iL} + \delta_i t_i)x_i)}{E(x_i^2)} = \frac{E(y_{iL}x_i)}{E(x_i^2)} + \frac{E(\delta_i t_i x_i)}{E(x_i^2)}.$$

Or si $t_i \in [0, 1]$:

$$E\left(\delta_i t_i x_i\right) = E\left(\delta_i t_i x_i 1\{x_i > 0\}\right) + E\left(\delta_i t_i x_i 1\{x_i < 0\}\right)$$
$$\leq E\left(\delta_i x_i 1\{x_i > 0\}\right),$$

symétriquement, on obtient :

$$E(\delta_i t_i x_i) \geq E(\delta_i x_i \mathbf{1}\{x_i < 0\}).$$

L'intervalle identifié pour β_1 est donc l'union de toutes les valeurs possibles et on obtient :

$$\beta_1 \in \left[\frac{E\left(y_{iL}x_i\right) + E\left[\delta_i x_i 1\{x_i < 0\}\right]}{E\left(x_i^2\right)}, \frac{E\left(y_{iL}x_i\right) + E\left[\delta_i x_i 1\{x_i > 0\}\right]}{E\left(x_i^2\right)} \right], \qquad (1)$$

dont la longueur est toujours positive car :

$$\frac{E\left(\delta_i \left| x_i \right|\right)}{E\left(x_i^2\right)} > 0$$

5. Il faut pour cela postuler une condition de stricte définie négativité sur le hessien de la log-vraisemblance en $\theta_0 = (\beta_0, \beta_1, \sigma^2)$ (Rothenberg, 1971).

6. Rappelons que $Ex_i = 0$ et donc la variance de x_i est $E(x_i^2)$.

si les intervalles de départ sont de longueur positive. Réciproquement, on peut montrer par un argument constructif que tout point de cet intervalle correspond bien à une distribution possible de $t_i \in [0, 1]$ ce qui permet d'affirmer que cet intervalle est identifié de manière exacte (par exemple, Magnac et Maurin, 2008).

Notons finalement que le paramètre β_1 se définit de façon équivalente par des inégalités de moment :

$$\begin{aligned} &E\left(y_{iL}x_i\right)+E\left(\delta_i x_i 1\{x_i<0\}\right)-\beta_1 E\left(x_i^2\right)\le 0,\\ &\beta_1 E\left(x_i^2\right)-E\left(y_{iL}x_i\right)-E\left(\delta_i x_i 1\{x_i>0\}\right)\le 0. \end{aligned} \tag{2}$$

1.3 *Exemple 2 : jeu d'entrée à information complète*

Deux entreprises notées 1 et 2 jouent un jeu d'entrée simultanée sur différents marchés, *m*, et leurs profits, π_{mj}, dépendent négativement de l'entrée sur le marché ou non, $y_{mj'} \in \{0, 1\}$, de leur rival :

$$\pi_{mj}=\beta_0-\beta_1 y_{mj'}+\varepsilon_j \text{ pour } \left(j,j'\right)\in\left\{(1,2),(2,1)\right\},\beta_1>0.$$

On suppose que les deux entreprises observent les variables ε_1 et ε_2 et le jeu est donc à information complète. L'économètre n'observe pas ces variables. Par simplicité, on s'abstiendra de considérer des variables explicatives qui permettent pourtant, sous des conditions de support infini et de restrictions d'exclusion de certaines variables, l'identification ponctuelle des paramètres (Tamer, 2003).

On étudie les résultats de ce jeu en *stratégies pures* mais ces résultats s'étendent facilement au cas de stratégies mixtes. Le graphique 1 décrit les différentes régions du plan en fonction des équilibres du jeu. Dans cette figure, on voit qu'il existe une région du plan (ε_1, ε_2), la région centrale, où deux équilibres sont possibles (0, 1) et (1, 0). Le modèle est dit *incohérent* (Heckman, 1978; Gouriéroux, Laffont et Monfort, 1980) car la vraisemblance du modèle est mal définie à cause de la multiplicité des équilibres.

On complète le modèle par la donnée d'une variable additionnelle, la variable $t_m \in E=[0, 1]$ qui décrit la sélection de l'équilibre (1, 0) dans la région centrale où les équilibres du jeu sont (0, 1) ou (1, 0). Si cette variable était observée, le modèle serait complet et ses paramètres sont identifiables sous certaines conditions (voir plus bas). La distribution de cette variable additionnelle est le paramètre η qui décrit la probabilité de sélection de l'équilibre (1, 0) dans la région centrale.

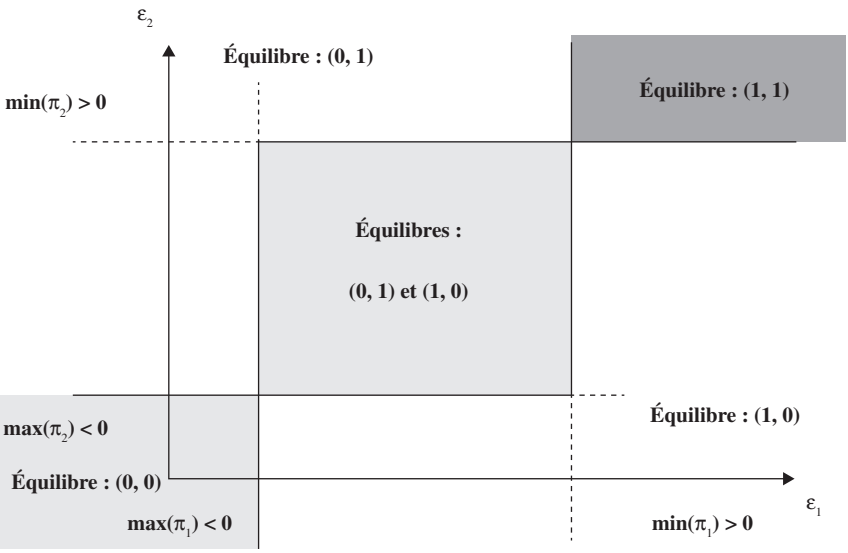
Certaines probabilités sont indépendantes de η :

$$\begin{aligned} &P\left(\left(y_j,y_{j'}\right)=(0,0)\right)=\Pr\left(\varepsilon_1\le-\beta_0,\varepsilon_2\le-\beta_0\right)\\ &P\left(\left(y_j,y_{j'}\right)=(1,1)\right)=\Pr\left(\varepsilon_1>\beta_1-\beta_0,\varepsilon_2>\beta_1-\beta_0\right) \end{aligned} \tag{3}$$

mais d'autres en sont dépendantes :

GRAPHIQUE 1

JEU D'ENTRÉE



$$P\left(\left(y_j, y_{j'}\right) = (0,1)\right) = \Pr\left(\varepsilon_1 > -\beta_0, \varepsilon_2 \le \beta_1 - \beta_0\right) - \eta \Pr(\text{carré central, graphique 1})$$

$$P\left(\left(y_j, y_{j'}\right) = (1,0)\right)$$

$$= \Pr\left(\varepsilon_1 \le \beta_1 - \beta_0, \varepsilon_2 > -\beta_0\right) - (1-\eta)\Pr \text{ (carré central, graphique 1)} \qquad (4)$$

où le carré central, graphique 1 est déterminé par les contraintes :

$$-\beta_0 < \varepsilon_1 \le \beta_1 - \beta_0, -\beta_0 < \varepsilon_2 \le \beta_1 - \beta_0.$$

Sans restrictions sur la fonction de répartition des variables ε_i en fonction de covariables (par exemple l'indépendance) les paramètres β_1 et β_0 et les fonctions de répartitions des ε ne peuvent être identifiées et pour affirmer que le modèle est complet on raisonnera directement sur les probabilités des régions définies par le graphique 1. Ces probabilités sont identifiées par les équations ci-dessus si le paramètre η est donné. Il nous faut alors identifier quatre probabilités, par exemple : $\Pr\left(\varepsilon_1 \le -\beta_0, \varepsilon_2 \le -\beta_0\right)$, $\Pr\left(\varepsilon_1 > \beta_1 - \beta_0, \varepsilon_2 > \beta_1 - \beta_0\right)$, $\Pr\left(\varepsilon_1 > -\beta_0, \varepsilon_2 \le \beta_1 - \beta_0\right)$ et $\Pr\left(\varepsilon_1 \le \beta_1 - \beta_0, \varepsilon_2 > -\beta_0\right)$[7]. Les deux premières sont directement identifiées par

7. Par la formule des probabilités totales, on obtient la probabilité du carré central en sommant ces quatre probabiltés et en soustrayant 1.

les deux premières équations (3) et les deux dernières obéissent aux contraintes dérivées des équations (4) sous les contraintes, $\eta \in [0, 1]$ et $0 \leq \Pr(\text{carré central, graphique 1}) \leq \min\left(\Pr\left(\varepsilon_1 > -\beta_0, \varepsilon_2 \leq \beta_1 - \beta_0\right), \Pr\left(\varepsilon_1 \leq \beta_1 - \beta_0, \varepsilon_2 > -\beta_0\right)\right)$.

De la même façon que dans l'exemple précédent, ces contraintes s'écrivent comme des inégalités de moment. De plus, tout vecteur de probabilités appartenant à l'ensemble identifié correspond à une probabilité de sélection d'équilibres réalisables, $\eta \in [0, 1]$, ce qui permet de dire que cet intervalle est identifié de manière exacte. Ces résultats s'étendent aussi à des jeux à plusieurs joueurs (voir par exemple, Galichon et Henry, 2011) mais d'une façon plus sophistiquée dont nous discutons maintenant.

1.4 *Discussion*

Les principaux résultats obtenus dans ces exemples s'étendent au cas d'ensembles identifiés de dimension quelconque. Ces résultats expriment les restrictions identifiantes sur les paramètres comme des contraintes à l'inégalité sur des fonctions linéaires ou non linéaires des coefficients et conduisent donc à un ensemble fini ou infini d'inégalités de moments. Une question plus difficile est de montrer que la caractérisation par ces inégalités de moment est équivalente à la caractérisation par les restrictions ensemblistes. Dans ce cas, le modèle est identifié partiellement mais exactement (c'est-à-dire *sharply*).

Ces extensions nécessitent pourtant des outils plus sophistiqués que ceux que nous avons utilisés. Dans les modèles structurels et en particulier ceux dérivés de la théorie des jeux, Galichon et Henry (2009, 2011) expliquent comment utiliser des outils comme les capacités de Choquet et les méthodes de transport optimal pour résoudre la question d'identification exacte et construire l'ensemble des inégalités de moment qui sont nécessaires et suffisantes pour l'identification exacte des paramètres. Alternativement, Beresteanu, Molchanov et Molinari (2012) expliquent comment des méthodes d'ensembles aléatoires convexes permettent aussi la résolution de ces questions. Ces outils sont aussi applicables aux modèles réduits tenant compte de la censure des variables comme ceux développés dans Horowitz et Manski (1995), Manski et Pepper (2000) ou plus généralement le livre de revue par Manski (2003). D'autres encore, combinent modèles structurels et censure des données (Galichon et Henry, 2013).

Dans un cadre d'inégalités de moment, les méthodes d'estimation sont dérivées par analogie ou en construisant des critères de distance à l'ensemble identifié (qui seront présentés plus loin) puisqu'ils servent aussi dans les méthodes d'inférence. Pour estimer un intervalle comme celui présenté dans l'exemple 1, on peut se contenter d'estimer directement les deux bornes, supérieure et inférieure de l'intervalle (par exemple, Imbens et Manski, 2004). Pour un ensemble, c'est plus difficile sauf si cet ensemble est convexe et ce cas sera analysé dans la section 2.3.

Certains auteurs omettent l'estimation « ensembliste » et passent directement à l'estimation par régions de confiance et donc à l'inférence que nous présentons maintenant. Notons toutefois un point final relatif à l'identification. De manière générale, il semble important pour l'inférence de contrôler la dimension de la fonction inconnue à qui s'applique la restriction ensembliste. Nous en verrons des exemples plus loin.

2. Méthodes d'inférence

Les techniques générales d'inférence reposent soit sur la construction d'une fonction critère et l'inversion d'un test basé sur cette fonction (Chernozhukov, Hong et Tamer, 2007) soit sur la caractérisation directe de la frontière (Beresteanu et Molinari, 2008; Bontemps *et al.*, 2012). On peut résumer ces deux méthodes par :

- En notant Θ_I l'ensemble identifié, on construit un test de niveau α pour toute valeur de θ :

  $$H_0 : \Theta_I \ni \theta, \qquad H_a : \theta \notin \Theta_I$$

  et on inverse ce test. Tous les θ qu'on ne rejette pas, appartiennent à l'ensemble de confiance de degré de confiance $1 - \alpha$.

- On estime la frontière de l'ensemble identifié Θ_I et on utilise sa variabilité dans l'échantillon pour construire les régions ou intervalles de confiance. On remarquera que cette démarche repose aussi *in fine* sur l'inversion d'un test comme la méthode précédente mais elle est rendue plus facile par l'estimation de la frontière. Cette démarche est typique quand l'ensemble est convexe (voir section 2.3).

Dans le premier cas, on se posera aussi des questions sur la sélection des moments à l'inégalité qui sont pertinents (Andrews et Soares, 2011; Andrews et Barwick, 2012; Andrews et Shi, 2013). D'autre part, les méthodes caractérisant la frontière peuvent amener à des inégalités de moment qui sont des intersections de bornes (Chernozhukov, Lee et Rosen, 2013).

C'est cet ensemble de questions que nous étudions dans cette section. En guise de préliminaire, on discutera deux points qui distinguent les objets des recherches. D'abord, il semble souvent raisonnable d'imposer que l'inférence soit robuste à la variation de la vraie loi de probabilité des données puisque celle-ci est inconnue. On considérera alors que la vraie loi de probabilité des données varie dans un ensemble large de lois de probabilités sous-jacentes et on construira la procédure d'inférence en tenant compte de cette variation. On dira alors dans ce cas que l'inférence est uniforme (par rapport à l'ensemble considéré de lois de probabilités). Deuxièmement, on se demande si l'objet est de couvrir par une région de confiance, soit un point, c'est-à-dire une seule vraie valeur des paramètres, soit un ensemble de valeurs distinctes.

2.1 *Couverture d'un point ou d'un ensemble et uniformité de l'inférence*

La question de la couverture d'un point ou d'un ensemble est la question la plus simple à exposer même si elle nécessite quelques notations. Supposons que la loi des données observables est P et dénotons $\Theta_I(P)$ l'ensemble identifié c'est-à-dire toutes les valeurs compatibles avec P (et implicitement toutes les autres restrictions structurelles que nous avons faites). Si nous voulons couvrir un point unique θ par un intervalle ou une région de confiance I_n à un niveau de confiance asymptotique au moins égal à $1 - \alpha$, il nous faut alors chercher I_n comme solution de :

$$\lim_{n\to\infty} \inf \left(\inf_{\theta \in \Theta_I(P)} \Pr(\theta \in I_n) \right) \geq 1 - \alpha. \tag{5}$$

On voit que l'identification partielle a consisté à remplacer dans cette expression la vraie valeur d'un paramètre ponctuellement identifié $\theta_0(P)$ par toute les valeurs appartenant à l'ensemble identifié, $\Theta_I(P)$.

Dans le premier exemple développé plus haut de censure de la variable dépendante par intervalles, l'intervalle de confiance prendra la forme, $I_n = \left[\hat{\beta}^L_{1,n} - \hat{c}^L_n, \hat{\beta}^U_{1,n} + \hat{c}^U_n \right]$ où $\hat{\beta}^L_{1,n}$ et $\hat{\beta}^U_{1,n}$ sont les estimateurs, respectivement des bornes inférieure et supérieure des quantités définies par l'équation (1) et où $\hat{c}^L_n$ et $\hat{c}^U_n$ sont des estimateurs dépendant de la loi jointe des estimateurs des bornes et d'une valeur critique qui s'ajuste par l'équation (5). Cet ajustement se fait pour toutes les valeurs possibles de θ dans l'intervalle identifié défini dans l'équation (1) au lieu de ne le faire que pour une seule valeur du paramètre comme dans le cas d'identification ponctuelle. Cette construction est exposée de manière précise par Imbens et Manski (2004).

Cette construction couvre un point, la supposée unique valeur vraie des paramètres. Mais maintenant que l'ensemble identifié a une « épaisseur », on pourrait vouloir couvrir des régions ou intervalles I au lieu de singletons $\{\theta\}$. On cherche alors des régions I_n qui satisfont la condition de niveau de confiance asymptotique au moins égal à $1 - \alpha$,

$$\lim_{n\to\infty} \inf \left(\inf_{I \subset \Theta_I(P)} \Pr(I \subset I_n) \right) \geq 1 - \alpha .$$

La plupart des applications économétriques s'attachent à couvrir un point mais il y a un certain nombre d'opinions dissidentes et la littérature est partagée entre ces deux présentations. Par exemple, Romano et Shaikh (2008, 2011) étudient les deux cas dans deux articles différents. Remarquons pourtant que la deuxième condition est plus contraignante que la première puisque les singletons sont des intervalles dégénérés (voir par exemple, Henry et Onatski, 2012). Les régions de confiance couvrant des ensembles sont donc généralement plus grands que ceux couvrant des points, $\theta \in \Theta_I(P)$.

D'autre part, la question de l'*uniformité* peut s'aborder dans un cadre simple d'inférence à une seule dimension comme dans l'exemple 1. L'intervalle identifié est décrit par une borne supérieure et par une borne inférieure comme dans l'équation (1) et on suppose que ces bornes sont des fonctions de moments dans les données. Comme nous l'avons dit plus haut, l'estimation de ces bornes est effectuée directement par la construction de contreparties empiriques à ces fonctions des moments. Les intervalles de confiance pour les bornes inférieures et supérieures « combinent » alors des éléments tenant compte de la distribution jointe des estimateurs des bornes. Néanmoins, les deux bornes estimées peuvent être suffisamment proches (au sens de la métrique données par leurs écarts-types) et leurs intervalles de confiance peuvent se chevaucher. La résolution de cette difficulté fût proposée par Imbens et Manski (2004) et étendue par Stoye (2009). Les auteurs construisent des intervalles de confiance dont les propriétés statistiques sont robustes au fait que le vrai intervalle identifié est aussi petit que l'on veut (y compris un point).

En retournant au cas général, admettons que nous voulions couvrir un point de façon uniforme. Supposons que le vrai processus générateur de données appartient à une famille $\mathcal{P}$. Dans l'exemple 1 de variable dépendante censurée par intervalles, on inclura par exemple dans cette famille les cas où il n'y a pas de censure ce qui fait que la largeur des intervalles d'observation peut être nulle, $\delta_i = 0$. Dans ce cas, le paramètre β_1 est identifié. Puis on cherchera un intervalle de confiance I_n à un niveau de confiance asymptotique au moins égal à $1 - \alpha$ qui satisfait à :

$$\liminf_{n\to\infty}\left(\inf_{P\in\mathcal{P},\theta\in\Theta_I(P)} \Pr(\theta \in I_n)\right) \geq 1-\alpha\,. \tag{6}$$

Ici aussi, la condition est plus contraignante que dans le cas non uniforme et les intervalles de confiance uniformes sont donc plus grands que ceux qu'on avait définis auparavant. C'est pourtant ce cas qui parait le plus intéressant puisqu'on n'a pas toujours une idée bien précise sur la vraie loi des données P et son domaine de variation $\mathcal{P}$. L'uniformité a d'ailleurs des définitions aussi variées que la variation de $\mathcal{P}$.

2.2 *Techniques générales d'inférence*

On décrit rapidement ici différentes techniques d'inférence proposées dans la littérature récente. Les premiers auteurs à utiliser une fonction critère positive qui prend une valeur nulle pour les points se situant à l'intérieur de l'ensemble identifié, ont été Chernozhukov *et al.* (2007) et ils ont été suivis par Rosen (2008), Andrews et Soares (2010), Romano et Shaikh (2010) et bien d'autres. Le point de départ de cette littérature est l'étude d'un nombre fini de conditions de moment à l'inégalité. La littérature la plus récente l'étend à un cadre où il y a une infinité de moments à l'inégalité. L'article d'Andrews et Shi (2013) s'intéresse en particulier à la transformation de moments conditionnels à l'inégalité en un nombre fini mais croissant avec la taille d'échantillon, de moment non conditionnels. Une autre méthode est fournie par Chernozhukov, Lee et Rosen (2013) qui développent une méthode

d'inférence pour une borne dans le cas où il y a une infinité de conditions de moments contenant cette borne. D'autres auteurs comme Menzel (2008), Ponomareva (2010) ou Armstrong (2011) étudient aussi ces cas en détail. Notons que l'utilisation du *bootstrap* a des aspects délicats puisque le *bootstrap* « naf » ne marche pas (Andrews et Guggenberger, 2010). Des *bootstraps* sophistiqués sont disponibles par exemple dans Bugni (2010), Andrews et Barwick (2012) et dans Henry, Meango et Queyranne (2013) ou Romano, Shaikh et Wolf (2013).

2.2.1 *Utilisation d'une fonction critère*

On suppose qu'il existe une fonction critère $Q(\theta) \geq 0$ qui décrit l'ensemble identifié Θ_I par la condition :

$$\theta \in \Theta_I \Leftrightarrow Q(\theta) = 0.$$

Par exemple, si on suppose que les restrictions forment un système fini de conditions de moments à l'inégalité :

$$E\left(h_j(Y,X,\theta)\right) \leq 0 \text{ pour } j = 1,.,J$$

alors on peut définir une version de la fonction $Q(\theta)$ comme :

$$Q(\theta) = \sum_{j=1}^{J} \left[Eh_j(Y,X,\theta) \right]^2 1\left\{ Eh_j(Y,X,\theta) > 0 \right\}.$$

La valeur du critère pour tout point hors de l'ensemble Θ_I est ainsi quadratique en la distance à 0 des moments en ce point. Cette construction s'applique directement au premier exemple étudié plus haut. Les équations (2) utilisent en effet deux fonctions :

$$h_1\left(y_{iL}, x_i, \delta_i, \beta_1\right) = y_{iL} x_i + \delta_i x_i 1\left\{x_i < 0\right\} - \beta_1 x_i^2,$$
$$h_2\left(y_{iL}, x_i, \delta_i, \beta_1\right) = \beta_1 x_i^2 - y_{iL} x_i - \delta_i x_i 1\left\{x_i > 0\right\}.$$

Une infinité de telles fonctions critères peuvent être définies et leurs propriétés comparées (Andrews et Shi, 2013 par exemple). Ici on utilise une fonction du type Cramer-von Mises qui somme les écarts quadratiques mais on pourrait utiliser une approche de type Kolmogorov en retenant le maximum de ces écarts. D'après les travaux d'Andrews et de ses coauteurs, la deuxième construction semble dominée par la première (par exemple, Andrews et Barwick, 2012). Une pondération des conditions de moments en fonction de leur matrice de covariance est aussi recommandée par ces auteurs comme dans une approche de méthode généralisée de moments (GMM) pour des restrictions de moment à l'égalité.

La contrepartie empirique de cette fonction critère dans un échantillon de taille n et d'observations $(y_i, x_i)_{i=1,..,n}$ est construite à partir des contreparties empiriques des moments, par exemple $h_{jn}(\theta) = \frac{1}{n}\sum_{i=1}^{n} h_j\left(y_i, x_i, \theta\right)$ et en notant :

$$Q_n(\theta) = \sum_{j=1}^{J} h_{jn}(\theta)^2 1\{h_{jn}(\theta) > 0\}.$$

Chernozhukov *et al.* (2007) proposent alors la construction de l'estimateur de l'ensemble comme :

$$\hat{\Theta}_n = \{\theta; Q_n(\theta) < \tau_n\}$$

où τ_n est un paramètre de lissage qui respecte les conditions aux limites :

$$\tau_n / \sqrt{n} \to 0, \sqrt{\ln \ln n} / \tau_n \to 0.$$

Ils proposent aussi directement l'estimation d'une région de confiance :

$$\hat{\Theta}_n^C = \left\{\theta; Q_n(\theta) < c_n\left(\hat{\Theta}_n, 1-\alpha\right)\right\}$$

à un niveau $1 - \alpha$. On peut voir cette construction comme l'inversion du test d'appartenance du paramètre θ à l'ensemble identifié Θ_I.

Le *point délicat* est le calcul de la valeur critique c_n qui se fait par sous-échantillonnage (« *subsampling* », Chernozhukov *et al.*, 2007; Romano et Shaik, 2010) ou par *bootstrap* (Bugni, 2010) en montrant que la région de confiance ainsi construite respecte la condition de couverture asymptotique donnée par l'équation (5) ou l'équation (6). La difficulté de ces papiers vient de la technicité des démonstrations asymptotiques.

2.2.2 *Sélection de moments*

Quand on teste une vraie valeur du paramètre θ_0, on peut améliorer l'inférence, c'est-à-dire le pouvoir du test conduisant à la construction de la région de confiance pour la vraie valeur (ou du vrai ensemble) du paramètre au niveau $1 - \alpha$ si on reconnaît que les conditions de moments à l'inégalité peuvent ne pas être toutes saturées en même temps. Les moments d'indice *j* tels que

$$E(h_j(Y_i, X_i, \theta_0)) < 0$$

ne « comptent » pas. Seuls les moments d'indice *k*, $E\left(h_k\left(Y_i, X_i, \theta_0\right)\right) = 0$ comptent mais le point délicat est qu'on ne sait quels moments comptent ou ne comptent pas.

Andrews et Soares (2010) dans un cadre de conditions en nombre fini et son extension, Andrews et Shi (2013), dans le cadre d'un système de conditions de moments conditionnels proposent la méthode suivante qu'ils appellent *sélection généralisée de moments* (GMS). Pour le test d'appartenance du paramètre θ à l'ensemble identifié Θ_I on n'utilisera pour la construction de la valeur critique c_n associée à ce test que les conditions de moments qui sont proches de la saturation. Les auteurs proposent que ce soient les données qui décident en fonction de la distance à la saturation de chaque moment. Pour calculer c_n, ces auteurs comparent aussi différentes techniques asymptotiques, par calcul de la variance asymptotique, *bootstrap* ou par *subsampling*. La dernière technique proposée en premier par

Chernozhukov *et al.* (2007) et reprise par Romano et Shaikh (2010) semble dominée par les autres propositions qui ont un meilleur comportement à distance finie, comme le montrent les travaux plus récents de Andrews et coauteurs.

D'autre part, pour assurer là aussi, un meilleur comportement à distance fini (mais dont les propriétés ne reposent pour l'instant que sur des expériences de Monte-Carlo), Andrews et Barwick (2012) construisent une *sélection affinée de moments* (*Refined Moment Selection*). La sélection de moments se fait alors en utilisant un critère plus lâche de sélection et qui ne varie pas avec le nombre d'observations tout en corrigeant le niveau du test à cause du caractère conservateur d'une telle méthode. Une simplification de cette méthode très intensive numériquement et en particulier dans le cas où le nombre de moments est important est proposée par Romano, Shaikh et Wolf (2012) au prix d'un caractère légèrement conservateur de la procédure et donc d'une légère perte de puissance.

2.2.3 *Intersection de bornes*

Quand il y a des moments surnuméraires, on peut exploiter plusieurs ou une infinité de conditions de moments. Par exemple, il existe de nombreux exemples (par exemple, Manski et Pepper, 2000) où les conditions estimantes sont écrites comme :

$$\theta \leq E\left(h\left(Y_i, X_i, Z_i\right) \mid Z_i = z\right), \text{ pour tout } z$$

et donc satisfont nécessairement :

$$\theta \leq \inf_z \left[E\left(h\left(Y_i, X_i, Z_i\right) \mid Z_i = z\right)\right].$$

Si on note $h_n(z)$ un estimateur dans un échantillon de taille n, par exemple un estimateur non paramètrique, de $E\left(h\left(Y_i, X_i, Z_i\right) \mid Z_i = z\right)$, estimer cette borne par la contrepartie empirique :

$$\inf_z \left(h_n(y_i, x_i, z)\right) \tag{7}$$

est sévèrement biaisée vers le bas à distance finie puisqu'on ne contrôle pas la variabilité de l'échantillonnage et donc la variation de la variance de $h_n(z)$ en fonction de z. L'argument de l'infimum de la fonction estimée dans l'équation (7) a une forte tendance à être un point z pour lequel l'estimation est très bruitée.

Chernozhukov, Lee et Rosen (2013) proposent de résoudre ce problème en construisant l'intervalle de confiance de la borne en utilisant l'estimateur :

$$\inf_z \left[h_n(z) + c_n v_n(z)\right]$$

où v_n est un estimateur de la variance de la contrepartie empirique en z. On rajoute donc dans la fonction objectif un terme qui pénalise les régions où la variable par rapport à laquelle on prend l'infimum conduit à des variances importantes de l'objectif. La difficulté là-aussi est le calcul de la valeur critique c_n. La plus grande partie de l'article est consacrée à cette construction et à la justification asymptotique de cette construction.

2.3 *Structure additionnelle : convexité*

Dans certains cas et en particulier dans les cas où les moments sont linéaires, l'ensemble identifié est convexe. L'idée est alors de remplacer l'estimation et l'inférence construites pour un ensemble par l'estimation et l'inférence pour une fonction, la fonction support. En effet, cette fonction caractérise complètement l'ensemble identifié convexe par ses coordonnées polaires (par exemple, Rockafellar, 1970). Beresteanu, Molchanov et Molinari (2012) développent le cadre général dans lequel la théorie des ensembles aléatoires peut être utilisée et qui fournit une autre justification de la fonction support.

La fonction support d'un ensemble convexe B est définie par :

$$\delta^*(q \mid B) = \sup_{\beta \in B}\left(q^T \beta\right) \text{ pour toutes directions, } q \text{ telles que } \| q \| = 1,$$

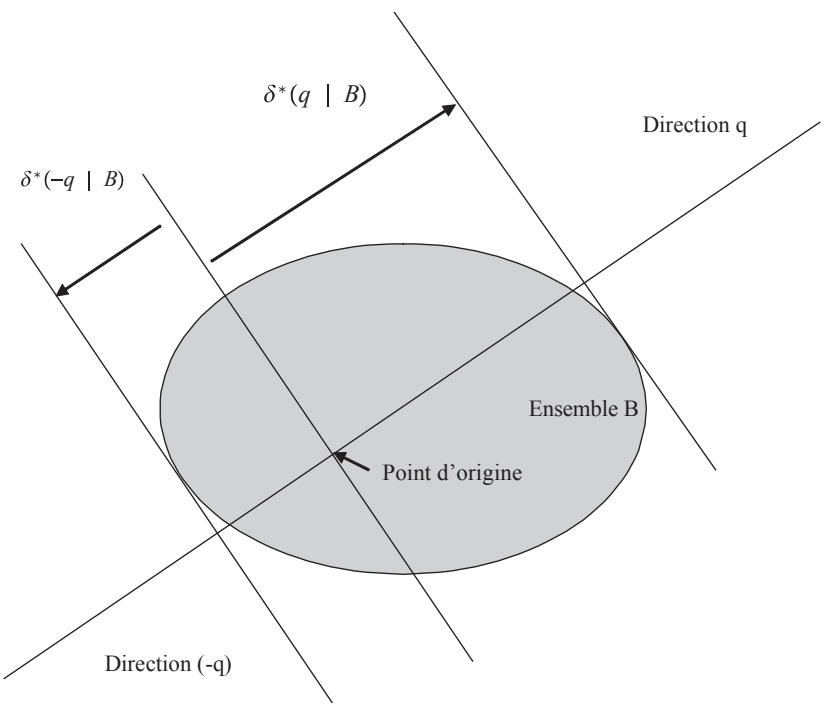
qui caractérise l'ensemble convexe B :

$$\beta \in B \Leftrightarrow \forall q, \| q \| = 1, q^T \beta \leq \delta^*(q \mid B).$$

Cette construction est illustrée dans le graphique 2.

GRAPHIQUE 2

LA FONCTION SUPPORT D'UN ENSEMBLE CONVEXE



Par exemple, dans le cas de censure par intervalles étudié par Stoye (2007), dans un modèle de prédiction linéaire, extension de l'exemple 1 à un cadre multi-dimensionnel, on écrit que :

$$y_i^* = x_i\beta + \varepsilon_i, y_i^* \in \left[y_{iL}, y_{iL} + \delta_i \right].$$

Par notre argumentation habituelle, le vecteur de paramètres β appartient à l'ensemble identifié, Θ_I si et seulement s'il existe une variable t_i de fonction de répartition $\eta = F(.\mid x_i, \delta_i, y_{iL})$ sur [0,1] et qui complète les données par l'intermédiaire de :

$$y_i^* = y_{iL} + \delta_i t_i,$$

comme dans l'exemple 1. Ceci permet alors d'écrire :

$$\beta = \left(E(x_i^T x_i) \right)^{-1} E\left(x_i^T \left(y_{iL} + t_i \delta_i \right) \right),$$

ce qui permet d'affirmer d'abord que l'ensemble est convexe puisque l'intervalle auquel t_i appartient, [0,1] est convexe. Cela permet ensuite d'écrire la fonction support comme :

$$\delta^*(q \mid B) = \sup_{\beta \in B} \left(q^T \beta \right) = \sup_{\eta_i \in 0,1} \left(q^T \left(E(x_i^T x_i) \right)^{-1} E\left(x_i^T \left(y_{iL} + t_i \delta_i \right) \right) \right).$$

Quelques calculs simples permettent d'obtenir la fonction support comme des moments des données (voir Bontemps *et al.*, 2012) :

$$\delta^*(q \mid B) = E\left(z_q w_q \right) = q^T \left(E\left(x^T x \right) \right)^{-1} E\left(x^T \left(1\left\{ q^T E\left(x^T x \right)^{-1} x^T > 0 \right\} \delta + y_L \right) \right).$$

Un estimateur de cette quantité peut être exprimé comme un estimateur MCO dans chaque direction *q*. En effet, l'expression du point de la frontière tangent à l'ensemble identifié et perpendiculaire à la direction *q* (voir le graphique 2) s'écrit :

$$\beta_q = \left(E\left(x^T x \right) \right)^{-1} E\left(x^T \left(1\left\{ x_q > 0 \right\} \delta + y_L \right) \right)$$

et est donc obtenu comme la limite en probabilité de l'estimateur MCO, $\hat{\beta}_q$, dans la prédiction linéaire de la variable dépendante construite en fonction des observables $\left(1\left\{ q^T E\left(x^T x \right)^{-1} x^T > 0 \right\} \delta + y_L \right)$ en fonction des covariables *x*. Notons alors l'estimateur de la frontière comme :

$$\hat{\delta}_n^*(q \mid B) = q^T \hat{\beta}_q.$$

Les méthodes d'inférence sont développées par Beresteanu et Molinari (2008) et Bontemps *et al.* (2012). Sous certaines hypothèses de régularité de la frontière (qui exclut en particulier des variables discrètes), ces auteurs montrent que le processus stochastique $\sqrt{n}\left(\hat{\delta}_n^*(q \mid B) - \delta^*(q \mid B) \right)$ tend uniformément en distribution quand *n* tend vers l'infini vers un processus stochastique gaussien. Bontemps *et al.* (2012) montrent comment l'opérateur de covariance, c'est-à-dire la covariance entre les estimateurs de la fonction support concernant des directions différentes

s'obtient simplement en utilisant les résidus des régressions dans chaque direction que nous avons décrites plus haut pour la construction de $\hat{\beta}_q$. Les tests et donc les intervalles de confiance sont écrits pour chaque direction q. Sous certaines conditions techniques, l'inférence est efficace (Kaido et Santos, 2011).

On peut étendre ces résultats au cas des variables discrètes mais les résultats asymptotiques font apparaître un processus additionnel dans la loi asymptotique de l'estimateur de la frontière précédente même si ceci ne semble pas avoir d'importance à distance finie dans des expériences de Monte-Carlo (Bontemps *et al.*, 2012). Ce dernier article étend aussi le cadre de prédiction linéaire en étudiant des restrictions générales de moments linéaires en nombre fini mais supérieur au nombre de paramètres à estimer (moments surnuméraires) et l'estimation et les méthodes d'inférence associées.

3. Un échantillon d'applications empiriques et autres sujets connexes

Le nombre d'applications empiriques de ces principes reste réduit et n'utilise pas toujours les résultats récents sur l'inférence que nous venons de passer en revue. Nous en donnons ici un échantillon non exhaustif.

Un des premiers articles utilisant des concepts d'identification partielle dans un cadre empirique est celui de Hotz, Mullin et Sanders (1997). C'est un modèle sous forme réduite qui posent des questions d'évaluation d'un traitement dans un cadre où la principale variable instrumentale ne respecte pas pleinement les conditions usuelles de validité. Le paramètre d'intérêt est l'effet causal de l'âge au premier enfant pendant l'adolescence sur les comportements et résultats ultérieurs et la variable instrumentale en question est le fait de subir une fausse couche au moment de la première grossesse. Les fausses couches ne fournissent en effet un instrument valide que pour un sous-échantillon dans la population et il est donc contaminé dans le sens d'Horowitz et Manski (1995). Cette littérature de traitement et sélection inclut aussi l'article de Manski et Pepper (2000) sur les rendements de l'éducation. Les auteurs analysent les rendements de l'éducation à chacun de ses niveaux et ces niveaux sont considérés comme un traitement multiple. De plus, ils utilisent des hypothèses de monotonicité sur l'effet du traitement ou sur l'existence d'une variable qui augmente les revenus. Les mêmes auteurs analysent les effets dissuasifs de la peine de mort aux États-Unis (Manski et Pepper, 2013) et montrent que différentes hypothèses conduisent à des conclusions diamétralement opposées.

Un autre exemple d'estimation sous forme réduite dans un cadre de sélection inclut l'article de Honoré et Lleras-Muney (2006). Ces deux auteurs estiment des bornes sur l'évolution au cours des 40 dernières années aux États-Unis des deux causes de mortalité principales que sont les maladies du coeur et le cancer. Ces causes sont traitées comme des risques concurrents dans un modèle de durée et la corrélation entre ces risques est le paramètre qui n'est pas identifiable. Les auteurs montrent que les progrès dans la lutte contre le cancer semblent avoir été cachés dans les analyses supposant des risques concurrents indépendants, par l'importance des progrès contre les accidents cardiaques.

L'évaluation de politiques publiques comme celles de stages offerts à certaines populations prend une place centrale dans la littérature récente en économétrie appliquée et à la suite de Manski, des auteurs ont développé un certain nombre de résultats sur les bornes. Par exemple, Lee (2009) montre comment surmonter les problèmes de sélection dans l'emploi pour évaluer les effets d'un programme de formation, le *Job Corps* aux États-Unis. Lee utilise des données d'expérimentations contrôlées et une hypothèse de monotonicité de l'effet du traitement sur l'emploi pour en inférer les effets du traitement sur les salaires *conditionnellement* à l'emploi. Le cadre proposé par Manski pour la sélection est aussi appliqué par Blundell, Gosling, Ichimura et Meghir (2007) au cas de l'évolution des rendements de l'éducation et ceci en fonction de la non-participation pour les hommes et pour les femmes au Royaume-Uni sur les 30 dernières années. La littérature empirique dans ces cadres de traitement et de sélection est donc assez bien développée et d'autres références auraient pu être données.

Plus rares sont les applications structurelles. L'un des premiers exemples est celui présenté par Haile et Tamer (2003). Les auteurs développent un modèle structurel dans des enchères ascendantes qui sont un cadre d'enchères dont les paramètres sont notoirement difficiles à identifier car l'information observée y est particulièrement réduite. Les auteurs n'exploitent que certaines contraintes de rationalité sur le comportement des agents et ne font pas d'hypothèses distributionnelles pour estimer les valeurs privées des enchérisseurs. Les auteurs supposent que les acheteurs potentiels n'enchérissent que jusqu'à la valeur qu'ils confèrent à l'objet et ne laissent pas l'objet être vendu à un prix inférieur à cette valeur.

Un autre exemple est fourni par Pakes, Porter, Ho et Ishii (2006) et son extension dans Pakes (2010). Ce dernier développe l'estimation de modèles structurels sous des contraintes de rationalité très générales sur des choix ordonnés (comme le nombre de comptoirs de retraits bancaires, *ATM*) ou dans des jeux non coopératifs entre hôpitaux et *HMOs*. Ces contraintes à l'inégalité s'appuient sur des restrictions sur les choix observés des entreprises et les paramètres d'intérêt sont les fonctions de profit des agents. Les choix faits par les entreprises doivent leur rapporter des profits qui sont supérieurs à ceux qu'elles obtiendraient si elles prenaient d'autres décisions. De la même façon, Ciliberto et Tamer (2009) estiment un modèle structurel dérivé d'un jeu d'entrée sur des lignes aériennes reliant deux aéroports en utilisant des données américaines et des restrictions dérivées de la rationalité des entreprises. Ce modèle généralise le deuxième exemple introduit plus haut et cet article utilise une méthode d'estimation dérivée de Chernozhukov, Hong et Tamer (2007). Henry et Mourifié (2012) étudient la concurrence politique et le modèle spatial de vote et montrent comment, malgré l'identification partielle, tester ce modèle. Les auteurs le rejettent d'ailleurs en utilisant des données américaines.

De nombreux sujets connexes se rattachent au cadre d'identification partielle. Ridder et Moffitt (2007) offrent un panorama complet de modélisations de données venant de sources multiples comme dans deux enquêtes, par exemple, ou deux états du monde mutuellement exclusifs. Les modèles à variation discrète dans un

cadre d'équations simultanées ont fait l'objet de recherches par Chesher (2005, 2010) et Chesher et Rosen (2013), des modèles discrets polytomiques sont traités par Chesher et Smolinski (2012) et les modèles binaires à régresseur « très exogène » dont l'observation est discrète ou censurée par intervalles par Magnac et Maurin (2008). Par ailleurs Davezies et d'Haultfoeuille (2012) traitent de l'attrition. L'identification partielle de paramètres de variance et covariance est étudiée par Horowitz et Manski (2005), Fan et Park (2010), Fan et Wu (2010) et Gomez et Pacini (2012). Nevo et Rosen (2012) et Conley, Hansen et Rossi (2011) introduisent ce que sont des instruments imparfaits. L'extension de méthodes de vraisemblance empirique fait l'objet d'un article par Canay (2010). On peut ainsi voir que les domaines d'application de telles méthodes sont très larges et de nombreux autres articles que nous n'avons pas cités ici en sont la preuve.

Conclusion

De façon générale, on peut décrire la démarche de l'économétre appliqué comme un choix d'hypothèses dite *structurelles* qu'il ou elle applique à un ensemble de données pour en dériver des paramètres économiques fondamentaux. La démarche traditionnelle est de chercher à compléter l'ensemble d'hypothèses de telle manière qu'un seul paramètre soit le résultat de cette démarche. Par exemple en utilisant des données censurées, on peut identifier ponctuellement les paramètres d'intérêt en faisant des hypothèses de normalité comme dans le cas d'un Probit ordonné dans le premier exemple exposé dans cet article. Le concept d'identification partielle permet d'abandonner cette complétion et admet que les structures crédibles soient assez lâches pour conduire seulement à l'identification d'un ensemble de valeurs des paramètres d'intérêt. Au lieu de faire une hypothèse de normalité, on fera des hypothèses d'indépendance vis-à-vis de variables explicatives ou d'indépendance en moyenne. Malgré cette extension du concept d'identification, la présentation des résultats de l'inférence par régions de confiance ne change pas et les raisonnements empiriques habituels des économétres appliqués restent les mêmes.

Il est à noter pourtant que cette démarche semble anti-poppérienne. Popper (2005) suggérait en effet qu'une des qualités d'une théorie est de faire des hypothèses suffisamment contraignantes pour qu'il soit facile de les falsifier ou rejeter. La démarche d'identification partielle semble au contraire se protéger contre tout rejet, en affaiblissant les hypothèses sur un certain nombre d'objets. L'assouplissement sans frein des restrictions semble donner lieu à un phénomène de régression à l'infini et le résultat en est assez peu encourageant puisque sans restrictions, on ne peut rien identifier. Des hypothèses faibles ont aussi le risque d'aboutir à des recommandations de politique économique peu précises au prix d'une robustesse qui pourrait paraître extrême dans la plupart des domaines scientifiques.

C'est pourquoi il faut interpréter cette démarche autrement. Une direction naturelle suggérée par Manski est de comparer des hypothèses de plus en plus contraignantes et qui permettent de réduire la taille de l'ensemble identifié (par exemple, la stratégie empirique utilisée dans Manski et Pepper, 2000). Ce ne seront

pas les données qui justifieront la crédibilité de chaque résultat puisque les données restent les mêmes. Ce sont les hypothèses que les chercheurs doivent justifier. Si la démarche est suffisamment ouverte pour que les lecteurs évaluent la crédibilité de ces restrictions, ils auront le choix de mener des raisonnements empiriques assez riches pour dire que telle hypothèse entraîne telle conclusion empirique ou absence de conclusion empirique. En effet, les bornes des intervalles ou régions identifiées peuvent être larges sous des hypothèses faibles. L'absence de conclusion doit alors provoquer la recherche d'hypothèses crédibles ou la construction de nouvelles données qui devraient permettre de rendre plus précis les résultats des démarches empiriques en économie.

Il faut aussi noter en passant les difficultés pratiques de l'inférence. Celle-ci est assez facile dans le cas convexe et certains programmes en Stata commencent à être écrits, par exemple par Francesca Molinari et ses coauteurs (http://economics.cornell.edu/fmolinari/#Stata_SetBLP). Cela est aussi vrai pour l'application des méthodes dans l'article de Chernozhukov *et al.* (2013) (http://www.cemmap.ac.uk/wps/cwp381313.pdf). Dans le cadre des inégalités de moment, les méthodes générales d'inférence sont innovantes mais elles prennent un certain temps. Ceci n'est qu'un problème technique qui se résoudra par une augmentation de la puissance de calcul.

Il ne me reste donc qu'à formuler un souhait qui est celui de voir un nombre plus important d'applications empiriques qui seul permettra d'évaluer la pertinence et la portée de ce courant récent de recherches théoriques pour la recherche empirique.

BIBLIOGRAPHIE

Andrews, D.W.K et P.J. Barwick (2012), « Inference for Parameters Defined by Moment Inequalities : A Recommended Moment Selection Procedure », *Econometrica*, 80(6) : 2805-2826.

Andrews, D.W.K. et P. Guggenberger (2009), « Validity of Subsampling and Plug-in Asymptotic Inference for Parameters Defined by Moment Inequalities », *Econometric Theory,* 25(3) : 669-709.

Andrews, D.W.K. et P. Guggenberger (2010), « Assymptotic Size and a Problem with Subsampling and with the *m* out of *n* Bootstrap », *Econometric Theory*, 26(02) : 426 – 468.

Andrews, D.W.K et X. Shi (2013), « Inference Based on Conditional Moment Inequalities », *Econometrica,* 81(2) : 609-666.

Andrews, D.W.K et G. Soares (2010), « Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection », *Econometrica,* 78(1) : 119-157

Armstrong, T.B. (2011), « Asymptotically Exact Inference in Conditional Moment Inequality Models », unpublished manuscript.

Beresteanu, A. et F. Molinari (2008), « Asymptotic Properties for a Class of Partially Identified Models », *Econometrica,* 76(4) : 763-814.

Beresteanu, A., I. Molchanov et F. Molinari (2011), « Sharp Identification Regions in Models with Convex Moment Predictions », *Econometrica*, 79(6) : 1785-1821.

Beresteanu, A., I. Molchanov et F. Molinari (2012), « Partial Identification Using Random Set Theory », *Journal of Econometrics*, 166(1) : 17-32.

Blundell, R.W., A. Gosling, H. Ichimura et C. Meghir (2007), « Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds », *Econometrica,* 75(2) : 323-363.

Bollinger, C.R. (1996), « Bounding Mean Regressions when a Binary Regressor is Mismeasured », *Journal of Econometrics*, 73(2) : 387-399.

Bontemps, C., T. Magnac et E. Maurin (2012), « Set Identified Linear Models », *Econometrica*, 80(3) : 1129-1155.

Bugni, F. (2010), « Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities », *Econometrica*, 78(2) : 735-754.

Canay, I. (2010), « EL Inference for Partially Identified Models : Large Deviations Optimality and Bootstrap Validity », *Journal of Econometrics*, 156(2) : 408-425.

Chernozhukov, V., H. Hong et E. Tamer (2007), « Inference on Parameter Sets in Econometric Models », *Econometrica,* 75(5) : 1243-1284.

Chernozhukov, V., S. Lee et A.M. Rosen (2013), « Intersection Bounds : Estimation and Inference », *Econometrica,* 81(2) : 667-737.

Chesher, A. (2005), « Non Parametric Identification under Discrete Variation », *Econometrica,* 73(5) : 1525-1550.

Chesher, A. (2010), « Instrumental Variable Models for Discrete Outcomes », *Econometrica*, 78(2) : 575-601.

Chesher, A. et A. Rosen (2013), « Generalized Instrumental Variable Models », unpublished.

Chesher, A. et K. Smolinski (2012), « IV models of Ordered Choice », *Journal of Econometrics*, 166(1) : 33-48.

Ciliberto, F. et E. Tamer (2009), « Market Structure and Multiple Equilibria in Airline Markets », *Econometrica,* 77(6) : 1791-1828.

Conley T., C. Hansen. et P. Rossi (2012), « Plausibly Exogenous », *The Review of Economic and Statistics*, 94(1) : 260-272.

Davezies, L. et X. d'Haultfoeuille (2012), « Partial Identification with Missing Outcomes », preprint.

Erickson, T. (1993), « Restricting Regression Slopes in the Errors in Variables Model by Bounding the Error Correlation », *Econometrica,* 61 : 959-969.

Fan, Y. et S. Park (2010), « Sharp Bounds on the Distribution of the Treatment Effects and Their Statistical Inference », *Econometric Theory*, 26(3) : 931-951.

Fan, Y. et J. Wu (2010), « Partial Identification of the Distribution of Treatment Effects in Switching Regimes Models and its Confidence Sets », *Review of Economic Studies*, 77(3) : 1002-1041

Fréchet, M. (1951), « Sur les tableaux de corrélation dont les marges sont données », *Annales de l'Université de Lyon*, III° Série Sci. A : 53-77.

Frisch, R. (1934), *Statistical Confluence Analysis by Means of Complete Regression Systems,* Oslo, Norway : University Institute of Economics.

Galichon, A., et M. Henry (2009), « A Test of Non-Identifying Restrictions and Confidence Regions for Partially Identified Parameters », *Journal of Econometrics,* 152(2) : 186-196.

Galichon, A. et M. Henry (2011), « Set Identification in Models with Multiple Equilibria », *Review of Economic Studies,* 78(4) : 1264-1298.

Galichon, A. et M. Henry (2013), « Dilatation Bootstrap : A Methodology for Constructing Confidence Regions with Partially Identified Models », *Journal of Econometrics,* forthcoming.

Gini, C. (1921), « Sull'interpoliazone di una retta quando i valori della variabile indipendente sono affetti da errori accidentali », *Metroeconomica,* 1(1) : 63-82.

Gomez, F. et D. Pacini (2013), « Measuring Inequality from Contaminated Data », preprint Bristol U.

Gouriéroux, C., J.J. Laffont et A. Monfort (1980), « Coherency Conditions in Simultaneous Linear Equation Models with Endogenous Switching Regime's », *Econometrica,* 48 : 675-95.

Greene, W.H. (2008), *Econometric Analysis*, Prentice Hall, 7th Edition.

Haile, P. et E. Tamer (2003), « Inference with an Incomplete Model of English Auctions », *Journal of Political Economy*, 111(1) : 1-51.

Heckman, J.J (1978), « Dummy Endogenous Variables in a Simultaneous Equation System », *Econometrica,* 46 : 931-960.

Henry, M., R. Meango et M. Queyranne (2013), « Combinatorial Bootstrap Inference in Partially Identified Incomplete Structural Models », unpublished manuscript.

Henry, M. et I. Mourifié (2012), « Set Coverage and Robust Policy », *Economics Letters,* 115(2) : 256-257.

Hoeffding, W. (1940), « Masstabinvariante Korrelationstheorie », *Shriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin,* 5(3) : 179-233.

Honoré, B. et A. Lleras-Muney (2006), « Bounds in Competing Risks Models and the War on Cancer », *Econometrica,* 74(6) : 1675-1698.

Horowitz, J.L. et C.F. Manski (1995), « Identification and Robustness with Contaminated and Corrupted Data », *Econometrica*, 63(6) : 281-302.

Hotz, V. J., C. H. Mullin, et S. G. Sanders (1997), « Bounding Causal Effects using Data from a Contaminated Natural Experiment : Analysing the Effects of Teenage Childbearing », *The Review of Economic Studies*, 64(4) : 575-603.

Imbens, G. et C.F., Manski (2004), « Confidence Intervals for Partially Identified Parameters », *Econometrica,* 72(6) : 1845-1859.

Kaido, H. et A. Santos (2011), « Asymptotically Efficient Estimation of Models Defined by Convex Moment Inequalities », preprint.

KLEPPER S. et E. E. LEAMER (1984), « Consistent Sets of Estimates for Regressions with Errors in all Variables », *Econometrica,* 52 : 163-183.

LEAMER, E.E. (1987), « Errors in Variables in Linear Systems », *Econometrica,* 55(4) : 893-909.

LEE, D.S. (2009), « Training, Wages, and Sample Selection : Estimating Sharp Bounds on Treatment Effects », *Review of Economic Studies,* 76(3) : 1071-1102.

MAGNAC, T. et E. MAURIN (2008), « Partial Identification in Monotone Binary Models : Discrete Regressors and Interval Data », *Review of Economic Studies,* 75(3) : 835-864.

MANSKI, C.F. (1989), « Anatomy of the Selection Problem », *Journal of Human Resources,* 24(3) : 343-360.

MANSKI, C.F. (2003), *Partial Identification of Probability Distributions*, Springer-Verlag : Berlin.

MANSKI, C.F. (2005), « Partial Identification in Econometrics », *New Palgrave Dictionary of Economics.*

MANSKI, C.F. et J.V. PEPPER (2000), « Monotone Instrumental Variables : With an Application to the Returns to Schooling », *Econometrica,* 68(4) : 997-1010.

MANSKI, C.F. et J.V. PEPPER (2013), « Deterrence and the Death Penalty : Partial Identification Analysis Using Repeated Cross Sections », *Journal of Quantitative Criminology*, 29(1) : 123-141.

MANSKI, C.F. et E. TAMER (2002), « Inference on Regressions with Interval Data on a Regressor or Outcome », *Econometrica,* 70(2) : 519-546.

MARSCHAK, J. et W.H. ANDREWS (1944), « Random Simultaneous Equations and the Theory of Production », *Econometrica*, 12 : 143-205.

MCFADDEN, D. (1994), « Contingent Valuation and Social Choice », *American Journal of Agricultural Economics*, 76(4) : 689-708.

MENZEL, K. (2008), « Estimation and Inference with many Moment Inequalities », Preprint, Massachussetts Institute of Technology.

NEVO, A. et A. M. ROSEN (2012), « Identification with Imperfect Instruments », *Review of Economics and Statistics*, 94(3) : 659-671.

PAKES, A. (2010), « Alternative Models for Moment Inequalities », *Econometrica*, 78(6) : 1783-1822.

PAKES, A., J. PORTER, K. HO et J. ISHII (2006), « Moment Inequalities and Their Applications », working paper.

PONOMAREVA, M. (2010), « Inference in Models Defined by Conditional Moment Inequalities with Continuous Covariates », Unpublished manuscript, Department of Economics, University of Western Ontario.

PONOMAREVA, M. et E. TAMER (2011), « Misspecification in Moment Inequality Models : Back to Moment Equalities », *Econometrics Journal,* 14(2) : 186-203.

POPPER, K. (2005), *The Logic of Scientific Discovery*, London and New York : Routledge/Taylor & Francis.

RIDDER, G. E. et R. MOFFITT (2007), « The Econometrics of Data Combination », *in* HECKMAN, J.J. et E. E. LEAMER (éds), *Handbook of Econometrics*, vol. 6, North-Holland, Amsterdam, chap. 75.

ROCKAFELLAR, R.T. (1970), *Convex Analysis,* Princeton University Press : Princeton.

ROMANO, J.P. et A.M. SHAIKH (2008), « Inference for Identifiable Parameters in Partially Identified Econometric Models », *Journal of Statistical Planning and Inference*, 138(9) : 2786-2807.

ROMANO, J.P. et A.M. SHAIKH (2010), « Inference for the Identified Set in Partially Identified Econometric Models », *Econometrica*, 78(1), 169-211.

ROMANO, J.P., A.M. SHAIKH et M. WOLF (2012), « A Simple Two-step Method for Testing Moment Inequalities with an Application to Inference in Partially Identified Models », available at SSRN 2137550.

ROSEN, A.M. (2008), « Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities », *Journal of Econometrics*, 146(1) : 107-117.

ROTHENBERG, T. J. (1971), « Identification in Parametric Models », *Econometrica,* 39 : 577-591.

STOYE, J. (2007), « Bounds on Generalized Linear Predictors with Incomplete Outcome Data », *Reliable Computing*, 13(3) : 293-302.

STOYE, J. (2009), « More on Confidence Intervals for Partially Identified Parameters », *Econometrica,* 77(4) : 1299-1315.

TAMER, E. (2003), « Incomplete Simultaneous Discrete Response Model with Multiple Equilibria », *The Review of Economic Studies*, 70(1) : 147-165.

TAMER, E. (2010), « Partial Identification in Econometrics », *Annual Review of Economics*, 2 : 167-195