Essai sur quelques problèmes d'identification en économie

Xavier d'Haultfoeuille

ECOLE D'ECONOMIE DE PARIS
UNIVERSITE PARIS-I PANTHEON-SORBONNE
U.F.R. de Sciences Economiques

N^o attribué par la bibliothèque

2	0	0	9	P	A	0	1	2	0	0	1

Thèse pour obtenir le grade de docteur ès Sciences Economiques

présentée et soutenue publiquement par

Xavier D'HAULTFOEUILLE

le 12 juin 2009

sous le titre

ESSAI SUR QUELQUES PROBLEMES D'IDENTIFICATION EN ECONOMIE

Directeur de Thèse

M. Jean-Marc ROBIN, Professeur à l'Ecole d'Economie de Paris

Jury

Rapporteurs : M. Christian BONTEMPS, Professeur à l'Université Toulouse I
M. Bernard SALANIE, Professeur à l'Université de Columbia

Examinateurs : M. Andrew CHESHER, Professeur à l'Université College de Londres
Mme Jacqueline PRADEL, Professeur à l'Ecole d'Economie de Paris

L'Université Paris-1 Panthéon-Sorbonne n'entend ni approuver, ni désapprouver les opinions particulières du candidat : ces opinions doivent être considérées comme propres à leur auteur.

Remerciements

Contrairement à ce que pourrait laisser accroire la première de couverture, cette thèse est un ouvrage collectif.

Un ouvrage collectif par son existence même, tout d'abord. Je remercie ainsi Jean-Marc Robin d'avoir accepté de m'encadrer. Je lui dois infiniment pour sa relecture attentive, ses conseils toujours avisés et ses encouragements bienvenus. Je lui suis également reconnaissant de n'avoir jamais relevé mes écarts au sujet initial de cette thèse...

Un ouvrage collectif par son contenu, ensuite. Merci en premier lieu à mes co-auteurs, Laurent Davezies, Philippe Février et Denis Fougère, d'avoir accepté de m'accompagner dans des projets qui s'assimilent si souvent à des parcours du combattant, encore plus souvent à des voies sans issue. Les chercheurs solitaires forcent l'admiration mais pas l'envie, car rien n'est plus agréable que l'échange fructueux d'idées et la saine émulation entre coauteurs. Je remercie aussi tous ceux qui ont accepté de prendre du temps pour m'écouter ou me lire, et m'ont souvent inspiré par leurs remarques pertinentes. Merci ainsi, sans malheureusement prétendre à l'exhaustivité, à Steve Berry, Stéphane Bonhomme, Nicolas Brunel, Marine Carrasco, Arthur Charpentier, Nicolas Chopin, Elise Coudin, Bruno Crépon, Fabien Dell, Jean-Claude Deville, Paul Doukhan, Romuald Elie, Jean-Pierre Florens, Eric Gautier, Stéphane Grégoir, Johan Hombert, Nicolas Lerner, Steve Machin, Thierry Magnac, Charles Manski, Xavier Mary, Emmanuel Massé, Arnaud Maurel, Amine Ouazad, Martin Pesendorfer, Georges Pavlov, Robert Porter, Christian Yann Robert, Jean-Charles Rochet, Mathieu Rosenbaum et Bernard Salanié.

Un ouvrage collectif, encore, car sans la question certes éprouvante "Alors, ta thèse, elle en est où ?", ce document n'existerait probablement pas. Le délai effectif correspondant aux "quelques jours de finition" d'une thèse a tendance à conforter la théorie de la relativité, et ces petites piqûres se révèlent alors tout à fait indispensables. Merci donc à ma famille et mes amis pour n'avoir jamais hésité à jouer les infirmières zélées, avec une mention particulière à Cécile, Emmanuel et Romuald.

Un ouvrage collectif, enfin, car sans les examinateurs et membres du jury, cette thèse n'aurait pas pu se conclure. Je remercie ainsi Christian Bontemps et Bernard Salanié d'avoir accepté de prendre le temps d'examiner ce document. Ma gratitude va également à Andrew Chesher et Jacqueline Pradel, qui ont accepté d'être membres du jury de cette thèse.

Résumé

Cette thèse présente trois sujets de recherche indépendants, liés néanmoins par la question de l'identification de modèles économiques.

Le premier chapitre est consacré à l'identification de modèles non-paramétriques instrumentaux. J'étudie tout d'abord la condition de complétude, qui a été utilisée récemment dans la régression non-paramétrique instrumentale ou dans le cadre de modèle à erreurs de mesure. Cet essai considère un modèle non-paramétrique additivement séparable entre les deux variables, avec une condition de large support. Dans ce cadre, différentes versions de la condition de complétude sont obtenues selon les conditions de régularité imposées au modèle. La deuxième partie du chapitre développe une nouvelle méthode pour résoudre la sélection endogène. Celle-ci s'appuie sur l'indépendance entre les instruments et la sélection et la condition précédente de complétude de la variable dépendante pour l'instrument. Une méthode d'estimation et une application sont également proposées.

Le deuxième chapitre se concentre sur les modèles d'économie industrielle empirique. Le premier essai considère l'identification non-paramétrique des modèles d'enchères à valeur commune. L'hypothèse identifiante principale est que le support de la distribution des signaux conditionnellement à la valeur du bien varie avec cette valeur. L'intérêt de cette approche est que, hormis cette condition, elle ne repose pas sur des restrictions fonctionnelles, contrairement à la littérature existante. Le deuxième essai étudie le modèle de sélection adverse. Celui-ci est défini par la fonction objectif du principal, l'utilité des agents et la distribution de leur type. Nous prouvons que l'identification de ce modèle nécessite la connaissance d'au moins l'une de ces trois fonctions. Nous montrons également que des changements exogènes dans la fonction objectif du principal sont suffisants pour obtenir une identification partielle ou totale. Une méthode d'estimation non-paramétrique est proposée et utilisée pour tester l'optimalité des contrats.

Le troisième chapitre, enfin, se focalise sur les modèles d'effets de pairs. Alors que ces modèles sont généralement considérés comme non-identifiés, nous montrons qu'une légère modification du modèle linéaire standard permet en général d'identifier les paramètres structurels, en utilisant les variations de taille de groupe. Ces résultats sont étendus à une version binaire de ce modèle.

Abstract

This PhD thesis presents three independent research topics, related however by the issue of the identification of economic models.

The first chapter is dedicated to the identification of nonparametric instrumental models. I first study the completeness condition, which has been recently used in nonparametric instrumental regression or in measurement error models. For that purpose, I suppose that a nonparametric, additively separable model holds between the two variable, together with a large support condition. In this framework, different versions of completeness are obtained, depending on the regularity conditions imposed on the model. The second part of this chapter develops a new method for dealing with endogenous selection. This method is based on the independence between the instruments and the selection and relies on a completeness condition between the outcome and the instrument. An estimation procedure and an application are also proposed.

The second chapter studies two empirical industrial organization models. The first part considers the nonparametric identification of the common value auction model. The main identifying assumption is that the support of the distribution of the signals, conditional on the value of the good, varies with this value. The advantage of our approach is that, apart from this condition, it does not rely on functional restrictions, contrarily to the existing literature. The second part focuses on the adverse selection model. This model is defined by the objective function of the principal, the agent's utility and the distribution of their type. We show that the identification of the model requires the knowledge of at least one of the three functions. We also show that exogenous changes in the objective function of the principal enable to identify fully or partially the model. A nonparametric estimation is proposed and used to test the optimality of contracts.

The third chapter focuses on peer effect models. Whereas these models are usually considered nonidentified, we show that a slight modification of the standard linear-in-means model enables in general to identify the structural parameters, by using the group size variations. These results are extended to a binary version of the model. Parametric estimation of the model is also considered, as well as the finite distance properties of the estimator.

Table des matières

Chapitre 1

Introduction

En tant que science, la discipline économique se doit non seulement de développer des modèles théoriques mais aussi d'utiliser l'observation pour déterminer les paramètres inconnus de ces modèles et tester leur validité[1]. L'économétrie peut se définir précisément comme l'interface entre modèles théoriques économiques et "données brutes"[2]. Ainsi, lors de son discours inaugural de la société d'économétrie en 1931, Ragnar Frisch lui attribuait pour objet de « favoriser les études de caractère quantitatif qui tendent à rapprocher le point de vue théorique du point de vue empirique dans l'exploration des problèmes économiques ». L'économétrie peut se résumer aux trois questions essentielles que sont l'identification, l'estimation et les tests d'un modèle. La première s'intéresse à la possibilité (théorique) de déterminer les paramètres inconnus d'un modèle à partir des observations. La deuxième se concentre sur les fonctions des données à utiliser pour approcher au mieux ces paramètres, et sur leurs propriétés statistiques. Enfin, la troisième question revient à étudier la possibilité de rejeter le modèle à partir des observations. Comme on le verra ci-dessous, ces questions sont liées mais non incluses l'une dans l'autre[3].

Cette thèse se concentre essentiellement sur la question de l'identification, à tra-

[1] A la suite de Popper (1968), on définit en effet fréquemment le caractère scientifique d'un modèle par son caractére réfutable.

[2] Même si cette dichotomie entre données "brutes" et théorie est contestable, cf. par exemple Bachelard (1934) pour une discussion générale sur ce sujet.

[3] Ces notions ne sont pas spécifiques à la science économique. De fait, la question de l'estimation est essentiellement statistique. L'identification et le test de la théorie sont également des problèmes cruciaux de disciplines telles que la science physique. L'économétrie se distingue cependant par le recours systématique aux probabilités, qui affecte la possibilité de mesurer et de tester les modèles théoriques sous-jacents.

vers l'étude indépendante de plusieurs modèles économiques. La première partie s'intéresse aux modèles non-paramétriques instrumentaux, plus particulièrement à la régression non-paramétrique instrumentale et aux modèles non-paramétrique de sélection endogène. La deuxième se penche sur l'identification de modèles structurels avec asymétrie d'information. Enfin, la troisième considère la question de l'identification des effets de pairs dans les modèles d'interaction sociale. Après avoir défini formellement les principales notions liées à l'identification et à la testabilité d'un modèle, l'introduction passera en revue la littérature sur ces questions et détaillera les résultats nouveaux obtenus ici, en reprenant le plan général de la thèse.

1.1 Identification et testabilité en économie

On considère une variable aléatoire X de $(\Omega, \mathcal{A}, P)$ à valeurs dans un espace mesuré $(E, \mathcal{E})$. On suppose que sa mesure de probabilité P^X appartient à $\mathcal{P}$, qui désigne l'ensemble des mesures de probabilités possibles sur $\mathcal{E}$ (ou un sous ensemble de ce-dernier). On définit un modèle statistique sur X comme suit :

Définition 1.1.1 *On appelle modèle statistique tout ensemble* $\{(\theta, P_\theta),\, \theta \in \Theta\}$ *où* Θ *est un ensemble quelconque*[4] *et pour tout* $\theta \in \Theta$, $P_\theta \in \mathcal{P}$. *On dit que* X *suit le modèle statistique* $M = \{(\theta, P_\theta),\, \theta \in \Theta\}$ *s'il existe* $\theta_0 \in \Theta$ *tel que* $P^X = P_{\theta_0}$.

Nous nous intéresserons dans la suite à deux propriétés primordiales d'un modèle statistique que sont l'identification et la testabilité. Pour cela, on considère la fonction

$$\begin{array}{rccc} \varphi_M : & \Theta & \to & \mathcal{P} \\ & \theta & \mapsto & P_\theta \end{array}$$

Définition 1.1.2 *Un modèle statistique* M *est dit identifiable lorsque* φ_M *est injective.*

En pratique, on prouve souvent qu'un modèle est identifié en exhibant une fonction h telle que $\theta = h(P_\theta)$, comme le mettent en évidence les exemples suivants.

[4]Le modèle est dit paramétrique lorsque Θ est un espace vectoriel de dimension finie, non-paramétrique sinon.

Exemple 1 *Modèle gaussien.* $X \sim \mathcal{N}(m, \sigma^2)$ *et on pose* $\theta = (m, \sigma^2)$, $\Theta = \mathbb{R} \times \mathbb{R}^{*+}$. *Le modèle est identifiable. En effet,*

$$(m,\ \sigma^2) = (E_\theta(X),\ V_\theta(X)),$$

donc $\theta = (m, \sigma^2)$ *est une fonction de* P_θ.

Exemple 2 *Régression instrumentale. On observe* $(X, Y, Z) \in \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^q$ *et l'on suppose que* $\mathcal{P}$ *est l'ensemble des lois de probabilité de* (X, Y, Z) *telles que* $rg(E(ZX')) = p$. *On considère alors le modèle suivant :*

$$Y = X'\beta + \varepsilon$$

où $\theta = (\beta, H)$, H *étant la loi de* (ε, X, Z), *et* $\Theta = \mathbb{R}^k \times \{H / E_H(Z'\varepsilon) = 0$ *et* $rg(E(ZX')) = p\}$. *Remarquons que* H *est une fonction de* β *et de* $P^{X,Y,Z}$ *puisqu'il s'agit de la loi de* $(Y - X'\beta, X, Z)$. *Il suffit donc de montrer que* β *est fonction de* P_θ. *On a*

$$E_\theta(ZX')\beta = E_\theta(ZY) \tag{1.1.1}$$

Comme $rg(E(ZX')) = p$, *il existe une seule solution à cette équation, d'où le résultat.*

L'identification peut être interprétée de la manière suivante. Si l'on dispose d'un échantillon infini de données i.i.d., P^X est observable par la loi des grands nombres[5]. Il s'agit alors de savoir si, lorsque X suit le modèle M, la connaissance de P^X implique celle du vrai paramètre. Si tel n'est pas le cas, le vrai paramètre (c'est-à-dire l'un des éléments de $\varphi_M^{-1}(P^X)$) n'est pas accessible à partir des données et du modèle considéré. De fait, l'identification est une condition nécessaire à l'existence d'un estimateur universellement convergent[6]. Notons en revanche que cette condition n'est pas suffisante. Ainsi, on peut montrer que sous cette hypothèse et certaines conditions de régularité, il existe un estimateur universellement convergent de $g(\theta)$ si et seulement si g est une fonction de première classe de Baire (cf. LeCam & Schwartz (1960), proposition 3).

[5]Plus formellement, si l'on note $(X_i)_{i \in \mathbb{N}}$ une suite i.i.d. de loi P^X, on a, pour tout $A \in \mathcal{E}$,

$$\lim_{n \to +\infty} \frac{1}{n} \sum_{i=1}^{n} 1\{X_i \in A\} = P^X(A).$$

[6]C'est-à-dire d'un estimateur convergent pour toute valeur possible de $P^X \in \{P_\theta,\ \theta \in \Theta\}$ et du vrai paramètre.

Dans une perspective purement statistique, l'identification d'un modèle peut être considéré comme un simple problème d'indexation. En particulier, il est toujours possible de modifier un modèle statistique M pour le rendre identifiable. Il suffit en effet pour cela de considérer le modèle $\dot{M} = \{(\dot{\theta},\, P_{\dot{\theta}}), \dot{\theta} \in \dot{\Theta}\}$, où $\dot{\Theta}$ est l'ensemble des classes d'équivalence de la relation R telle que

$$\theta R \theta' \Longleftrightarrow P_\theta = P_{\theta'}.$$

Cependant, le choix de l'index n'est pas anodin dans un modèle théorique. Dans ce cadre en effet, θ est généralement un paramètre structurel permettant par exemple d'évaluer l'efficacité d'une mesure ou d'effectuer de la prévision. En ce sens, et comme évoqué précédemment, l'identification est une notion à la frontière entre la théorie et les données, qui permet de quantifier l'information apportée par le modèle et les données. Cette information se traduit par la notion de région d'identification.

Définition 1.1.3 *La région d'identification de θ est $R_\theta = \varphi_M^{-1}(P_\theta)$.*

R_θ correspond à l'ensemble des paramètres cohérents avec les données observées lorsque celles-ci sont issues de la loi P_θ. Cette région est évidemment réduite au singleton θ lorsque le modèle est identifiable. Dans le cas contraire, on parlera d'identification partielle (cf. Manski, 2003). Deux cas de figures importants conduisent à l'identification partielle des paramètres. Le premier est la présence de données manquantes, due par exemple à la non-réponse (cf. Manski, 2003), à des erreurs de mesure (cf. Horowitz & Manski, 1995), à l'utilisation de source différentes qui ne peuvent être appariées (cf. le problème d'inférence écologique abordée par exemple par Cross & Manski, 2002), ou au fait que le contrefactuel ne peut être observé lorsqu'on s'intéresse à l'effet d'un traitement (cf. Manski, 1997 et Manski & Pepper, 2000). Le deuxième exemple important est celui des modèles incomplets. Dans ces modèles, θ n'est pas associé une seule probabilité P_θ, mais à plusieurs. Cette situation se produit notamment dans des modèles de jeux admettant des équilibres multiples, et où le mécanisme de sélection de l'équilibre n'est pas précisé (voir Jocanovic (1989), 1989, pour une discussion générale sur le sujet, et par exemple Andrews et al., 2003, et Ciliberto & Tamer, 2006 pour une application aux modèles d'entrée sur des marchés oligopolistiques)[7].

[7]Notons qu'il est possible de rendre complet ce modèle en adjoignant à θ les probabilités de sélection (inconnues) de chaque équilibre.

Intuitivement, la raison pour laquelle les paramètres ne sont que partiellement identifiés dans ces deux cas de figure est que le modèle n'apporte dans ces deux cas pas suffisamment d'information pour garantir l'identification ponctuelle des paramètres. De façon générale, la taille de la région d'identification dépendra des hypothèses imposées au modèle. Généralement, celle-ci sera d'autant plus petite que les hypothèses seront restrictives. Parfois, cependant, des hypothèses apparemment très proches conduisent à des résultats d'identification opposés[8].

Par ailleurs, on peut s'intéresser à l'identification d'une fonction du paramètre et non du paramètre θ lui-même. Ceci est le cas si le paramètre d'intérêt est $g(\theta)$ et non θ lui-même, comme dans l'exemple 4.

Définition 1.1.4 *$g(\theta)$ est identifiable lorsque $g(R_\theta) = \{g(\theta)\}$.*

Exemple 3 *Modèle gaussien modifié : $X \sim \mathcal{N}(m_1 + m_2, \sigma^2)$ et on pose $\theta = (m, m', \sigma^2)$, $\Theta = \mathbb{R}^2 \times \mathbb{R}^{*+}$. Dans ce cas le modèle n'est pas identifiable car $P_\theta = P_{\theta'}$ pour tout $\theta = (m_1, m_2, \sigma^2) \neq \theta' = (m_1', m_2', \sigma^2)$ tels que $m_1 + m_2 = m_1' + m_2'$. En revanche, $g(\theta) = (m_1 + m_2, \sigma^2)$ est identifiable pour tout θ.*

Exemple 4 *Effet de traitement. Soit (Y_0, Y_1), $D \in \{0, 1\}$ un traitement binaire et $Y = DY_1 + (1 - D)Y_0$. Soit $\theta = P^{Y_0, Y_1}$. On observe (D, Y) et on suppose que $D \perp\!\!\!\perp (Y_0, Y_1)$. Alors, θ n'est pas identifiable mais $\pi_i(\theta) = P^{Y_i}$ ($i \in \{0, 1\}$) l'est. En particulier, l'effet moyen du traitement $E(Y_1 - Y_0)$ est identifiable.*

On considère maintenant la notion de testabilité d'un modèle. La définition donnée ici rejoint celle de Koopmans & Reiersøl (1950).

Définition 1.1.5 *Un modèle statistique M est dit testable lorsque φ_M est non surjective.*

En d'autres termes, un modèle est testable lorsqu'on peut, théoriquement du moins[9], le rejeter à partir des données. En ce sens, la testabilité peut être considérée

[8] Un exemple frappant est celui des modèles binaires $Y = 1\{X\beta + \varepsilon \geq 0\}$. En effet, la région d'identification de β est non bornée lorsqu'on suppose $E(\varepsilon|X) = 0$, alors que β est identifié ponctuellement (sous des conditions sur le support des X) lorsque $\text{med}(\varepsilon|X) = 0$ (cf. Manski, 1988).

[9] Comme pour l'identification, la testabilité n'implique pas qu'il existe un test convergent de niveau fixé de la validité du modèle. Par exemple si le modèle est simplement $E(X) = m$ pour une valeur m fixée, il n'existe pas de test de niveau fixé convergent permettant de valider le modèle (cf. Bahadur & Savage, 1956, et Dufour, 2003, pour une discussion sur la possibilité d'implémenter des tests).

comme un critère popperien de scientificité du modèle (Popper, 1968).

Les notions d'identification et de testabilité sont bien distinctes, comme le montrent les exemples suivants. Lorsqu'un modèle est identifiable mais non testable, on parle de modèle juste identifié.

Exemple 3 *(suite) Le modèle gaussien modifié est non identifiable mais testable, car il implique par exemple qu'il existe u tel que $E((X-u)^{2k+1}) = 0$ pour tout $k \in \mathbb{N}^*$.*

Exemple 5 *Modèle non-paramétrique : θ est une mesure de probabilité quelconque, $\Theta = \mathcal{P}$ et $\varphi_M = Id$. Le modèle est identifiable et non testable.*

De nombreux modèles, cependant, sont identifiables et testables. Ces modèles sont dits suridentifiés.

Définition 1.1.6 *Un modèle statistique M est suridentifié lorsqu'il est identifiable et testable.*

Intuitivement, un modèle est suridentifié lorsque la connaissance d'une information plus réduite que P_θ (i.e., $h(P_\theta)$ où h est non injective) permet de retrouver θ. La proposition suivante confirme cette intuition.

Proposition 1.1.7 *M est suridentifié si et seulement s'il existe h de $\mathcal{P}$ dans H (ensemble quelconque) non injective telle que $h \circ \varphi_M$ soit injective.*

Preuve : condition suffisante : φ_M est injective puisque $h \circ \varphi_M$ est injective. Supposons que le modèle ne soit pas suridentifié. Alors φ_M est surjective, et donc bijective. Ceci implique que $h = (h \circ \varphi_M) \circ \varphi_M^{-1}$ est injective, ce qui est absurde.

Condition nécessaire : si M est suridentifié, alors $\mathcal{P} \backslash \{P_\theta,\, \theta \in \Theta\}$ est non vide. Soit alors $\theta_0 \in \Theta$ et h de $\mathcal{P}$ dans Θ définie par

$$h(P) = \left| \begin{array}{ll} \theta & \text{s'il existe } \theta \text{ tel que } \ P = P_\theta \\ \theta_0 & \text{pour tout } P \in \mathcal{P} \backslash \{P_\theta,\, \theta \in \Theta\} \end{array} \right.$$

Alors h est non injective mais $h \circ \varphi_M$ l'est. □

Exemple 2 *(suite) si $E(ZX')$ est inversible, le modèle est juste identifié. En effet, pour tout $P^{X,Y,Z}$ dans $\mathcal{P}$, si l'on pose $\beta = E(ZX')^{-1}E(ZY)$, on a*

$$E(Z(Y - X'\beta)) = 0.$$

Donc il existe toujours θ tel que $P = P_\theta$. En revanche, si $Z = (Z_1, Z_2)$ où (Z_1, Z_2) sont non colinéaires, $Z_2 \in \mathbb{R}$ et $E(Z_1X')$ inversible, le modèle est suridentifié. En effet, considérons η le résidu de la régression de Z_2 sur Z_1 et (β_0 étant quelconque) $Y = X'\beta_0 + \eta$. Alors $P^{X,Y,Z}$ ne satisfait pas les contraintes du modèle : il n'existe pas de paramètres β satisfaisant les contraintes du modèle. Supposons en effet le contraire. Alors $E(Z_1(Y - X\beta)) = 0$ et donc nécessairement $\beta = E(Z_1X')^{-1}E(Z_1Y) = \beta_0$. Mais alors $E(Z_2(Y - X'\beta_0)) = E(Z_2\eta) \neq 0$, d'où la contradiction.

1.2 Sur l'identification de modèles non-paramétriques instrumentaux

1.2.1 Motivation

Un des problèmes phares en science est de réussir à distinguer corrélation et causalité. Ce problème se pose sans doute de manière encore plus aiguë en économétrie, où l'expérimentation reste l'exception plutôt que la règle. Ainsi, il est souvent délicat d'interpréter les variations d'une variable explicative comme exogènes. Les variations dans le nombre d'années d'étude, par exemple, sont issues en partie au moins de choix différents d'agents hétérogènes. Cette hétérogénéité est susceptible d'expliquer les différences de salaires et donc de masquer partiellement l'effet causal du diplôme[10]. Par ailleurs, du fait par exemple de l'autosélection des individus ou de la non-réponse, la population sur laquelle les données sont observées ne correspond pas toujours à la population d'intérêt. Ainsi, les salaires observés sur le marché du travail sont ceux des individus ayant choisi de travailler, et leur distribution diffère a priori de celle des salaires de l'ensemble de la population (cf. Heckman, 1974).

[10]Sur la question des rendements de l'éducation, voir par exemple la revue de littérature de Card (2001).

Une des solutions majeures proposée par les économètres pour identifier des causalités est le recours aux variables instrumentales. En général, il s'agit de variables influant la variable explicative ou la sélection mais pas directement la variable expliquée. Bien que ce principe d'exclusion soit général, les variables instrumentales ont longtemps été cantonnées au cadre linéaire. Cette hypothèse est pourtant restrictive. Supposer que les rendements marginaux de l'éducation sont constants, par exemple, semble peu réaliste. Par ailleurs, elle exclut d'emblée les modèles non-linéaires tels que les modèles à variables dépendantes limitées, ou la sélection endogène, qui induit par nature l'existence de non-linéarités. Jusqu'à récemment, les solutions proposées dans la littérature pour résoudre ce problème ont été essentiellement paramétriques. La limite principale de cette approche est évidemment la sensibilité des résultats à la forme fonctionnelle choisie. Malgré leur intérêt évident, et sans doute du fait des difficultés techniques qu'elles posent, en particulier en termes d'identification, les solutions non-paramétriques aux problèmes d'endogénéité ou de sélection endogène n'ont été développées que récemment.

1.2.2 Etat de l'art

Régression non-paramétrique instrumentale

Le problème étudié ici peut se résumer ainsi. Soit Y une variable expliquée et X les variables explicatives, vérifiant

$$Y = \varphi(X, \varepsilon) \tag{1.2.1}$$

où ε est une variable aléatoire inobservée appartenant à un espace quelconque. Une partie des variables X est endogène dans le sens qu'elle n'est pas indépendante de ε. On dispose cependant d'un instrument Z indépendant de ε[11]. Il s'agit alors d'exhiber des conditions permettant d'identifier φ, ou du moins une composante de cette fonction.

Les différentes méthodes d'identification de modèles non-paramétriques peuvent se comprendre comme des extensions des stratégies instrumentales du modèle linéaire avec instruments (cf. l'exemple 2 de la section précédente). Dans ce cadre, une

[11]L'hypothèse d'indépendance n'est pas toujours nécessaire, comme le montre l'exemple du modèle linéaire instrumental. Plus généralement, on suppose que $M(Z, \varepsilon) = 0$, où M est une fonctionnelle connue.

première méthode est de s'appuyer sur l'équation (1.1.1). Cette équation identifie β dès que la matrice $E(ZX')$ est de plein rang. Considérons maintenant une version additive de (1.2.1) :

$$Y = \varphi(X) + \varepsilon \tag{1.2.2}$$

avec ε réel satisfaisant $E(\varepsilon|Z) = 0$. Dans ce cadre, on aura, de manière similaire à l'équation (1.1.1),

$$E(Y|Z) = E(\varphi(X)|Z) \tag{1.2.3}$$

Cette approche est développée par Newey & Powell (2003), Hall & Horowitz (2005) et Darolles et al. (2007). L'équivalent de la condition de rang est la condition de complétude suivante :

$$E(h(X)|Z) = 0 \text{ p.s.} \Longrightarrow h(X) = 0 \text{ p.s.} \tag{1.2.4}$$

pour toute fonction h telle que $E[|h(X)|] < +\infty$. Le nom de cette condition provient de la statistique classique (cf. par exemple Lehmann, 1983), où une statistique est dite complète par rapport à un paramètre θ lorsque pour tout $\theta \in \Theta$, $E_\theta(h(T)) = 0$ implique $h(T) = 0$ presque sûrement. Newey & Powell (2003) donnent des conditions nécessaires et suffisantes lorsque le support de (X, Z) est fini, et montrent également que la condition est satisfaite lorsque la densité conditionnelle de X sachant Z suit un modèle exponentiel.

Une deuxième approche consiste à remarquer que dans le modèle linéaire, X n'est plus corrélé à ε dès lors que l'on ajoute η, résidu de la régression de X sur Z. On parle alors d'approche par "variables de contrôle". D'un point de vue non-paramétrique, si l'on pose

$$X = \psi(Z, \eta) \tag{1.2.5}$$

et si l'on suppose que $Z \perp\!\!\!\perp (\eta, \varepsilon)$, alors X sera indépendant de ε conditionnellement à η. De plus, η est identifié à une fonction strictement croissante près dès que η est réel et $\psi(z, .)$ est une fonction strictement croissante, puisqu'alors $F_\eta(\eta) = F_{X|Z}(X|Z)$. Dans le modèle (1.2.2), on aura alors

$$E(Y|X, \eta) = \varphi(X) + E(\varepsilon|\eta).$$

Cette équation montre que $\varphi(.)$ est identifiée. Cette approche est suivie par Newey et al. (1999). Il est également possible de relâcher l'hypothèse d'additivité dans le modèle (1.2.1). Imbens & Newey (2008) montrent ainsi que sous des conditions

assez restrictives de rang et de large support (voir leur hypothèse 1), les quantiles de $\varphi(x, \varepsilon)$ sont identifiés, pour tout x. Sous des conditions plus faibles, on peut identifier par exemple l'effet marginal moyen $E(\partial\varphi/\partial x(X, \varepsilon))$, ou obtenir des bornes sur les quantiles précédents et sur la "fonction de moyenne structurelle" $\int \varphi(x, e)dP^{\varepsilon}(e)$.

La structure sous-jacente à l'approche par les variables de contrôle, à savoir l'indépendance entre Z et (ε, η) et la monotonicité de $\psi(z, .)$, n'est pas anodine. Florens et al. (2008) considèrent par exemple un modèle de choix éducatif où le lien entre revenus Y et niveau d'études X est quadratique et où Z est une variable affectant le coût des études. Ils montrent que si le coefficient du terme quadratique est aléatoire, l'hypothèse de monotonicité n'est pas satisfaite. Une autre approche a été proposée par Chernozhukov & Hansen (2005), s'appuyant sur la restriction d'un résidu ε réel et d'une fonction $\varphi(x, .)$ strictement croissante dans l'équation (1.2.1). On peut dans ce cas normaliser la distribution de ε à une loi uniforme et les auteurs montrent que si Z est indépendant de ε, alors pour tout $\tau \in [0, 1]$,

$$\Pr(Y < \varphi(X, \tau)|Z) = \tau. \tag{1.2.6}$$

Cette équation admet une seule solution dès que la condition suivante est vérifiée :

$$E(F_{Y|X,Z}(h(X))|Z = z) = E(F_{Y|X,Z}(\varphi(X, \tau))|Z = z) \Longrightarrow h(X) = \varphi(X, \tau) \text{ p.s.} \tag{1.2.7}$$

pour tout $\tau \in [0, 1]$ et toute fonction h. Cette condition est proche de la condition de complétude précédente[12]. Cependant, mis à part les cas où (X, Z) est à support fini et où X, conditionnellement à Z suit un modèle exponentiel d'une certaine forme (cf. Chernozhukov et al. 2007), il n'existe pas pour le moment de condition suffisante à cette hypothèse.

L'hypothèse de stricte monotonicité supposée par Chernozhukov & Hansen (2005) ne s'applique pas lorsque Y est discret. Dans un modèle binaire $Y = \mathbb{1}\{X\beta + \varepsilon \geq 0\}$, par exemple, Y est simplement une fonction croissante de ε. Dans ce cas, Chesher (2008) montre que l'équation (1.2.6) est remplacée par la paire d'inégalités

$$\begin{aligned} \Pr(Y \leq \varphi(X, \tau)|Z) &\geq \tau \\ \Pr(Y < \varphi(X, \tau)|Z) &< \tau. \end{aligned}$$

[12]On peut ainsi montrer que lorsque X est exogène (ou, plus précisément, si ε est indépendant de (X, Z)), cette condition est équivalente à une condition de complétude sur l'espace des fonctions h bornées.

A partir de ces inégalités, Chesher (2008) propose une méthode pour construire les bornes minimales sur cette fonction. Bien sûr, la taille de la région d'identification dépend du lien entre les instruments et les variables endogènes. Mais l'intérêt de cette approche est que les bornes peuvent être calculées dans tous les cas, sans condition de rang non-paramétrique telle que (1.2.4) ou (1.2.7).

Modèles de sélection avec instruments

Le problème de la sélection, proche bien que distinct de l'endogénéité, a également fait l'objet d'une abondante littérature. On suppose ici que l'on observe une variable binaire D (et, éventuellement, des covariables X et des instruments Z) et une variable d'intérêt Y seulement lorsque $D = 1$. Notons que ce cadre inclut aussi bien la non-réponse partielle que le problème du contrefactuel dans la littérature sur l'évaluation. Dans ce dernier cas, le problème d'observation est double puisqu'on n'observe Y_1 (résultat lorsque l'individu est traité) que lorsque le traitement a effectivement eu lieu ($D = 1$) et Y_0 (résultat en l'absence de traitement) que lorsque $D = 0$[13]. La question posée est celle de l'identification de la distribution de Y (ou de paramètres liés à cette distribution).

L'approche la plus courante pour obtenir l'identification de cette distribution est l'hypothèse de sélection sur observables :

$$Y \perp\!\!\!\perp D|X \tag{1.2.8}$$

Cette condition a été discutée en détail par exemple par Little & Rubin (1987) dans le cadre de la non-réponse et par Imbens (2004) dans le cadre de l'évaluation de politiques publiques. Sous l'hypothèse additionnelle que $P(D = 1|X) > 0$ presque sûrement, cette condition assure l'identification ponctuelle de la distribution jointe de (D, Y, X). L'hypothèse (1.2.8) est cependant restrictive puisqu'elle interdit la possibilité d'une sélection sur des variables inobservables corrélées à la variable d'intérêt. Une approche instrumentale a été développée pour résoudre ce problème (cf. par exemple Heckman, 1974, Angrist et al., 1996, ou Heckman & Vytlacil, 2005). Elle repose sur l'hypothèse suivante :

$$Y \perp\!\!\!\perp Z|X \tag{1.2.9}$$

[13]Cependant, si l'on ne s'intéresse qu'à l'effet du traitement sur les traités, $E(Y_1 - Y_0|D = 1)$, le problème de sélection ne se pose que sur Y_0.

Z, par ailleurs, est supposée avoir un impact sur D ($P(D = 1|X, Z) \neq P(D = 1|X)$). L'idée est que Z permet de randomiser la sélection D, comme dans le cadre d'une expérience. Notons cependant que l'hypothèse ne permet pas en général d'identifier ponctuellement la distribution de Y (cf. Manski, 2003). On ne peut obtenir un tel résultat que s'il existe z tel que $P(D = 1|Z = z) = 1$ ou si l'on peut faire tendre cette probabilité vers 1. Dans le cas contraire, on peut cependant estimer des paramètres tels que des effets de traitement locaux (cf. Angrist & Imbens, 1994) ou les effets de traitement marginaux (cf. Heckman & Vytlacil, 2005).

La troisième approche permettant de résoudre le problème de la sélection consiste, plutôt que de s'appuyer sur des relations d'exclusion comme (1.2.8) ou (1.2.9), à s'appuyer sur des restrictions fonctionnelles. Ainsi, Chamberlain (1986) montre qu'un modèle de sélection généralisée peut être identifié sans instrument supplémentaire en utilisant la condition à l'infini issue de la restriction linéaire sur la dépendance entre Y et X. Plus récemment, Lewbel (2007) a montré qu'on pouvait résoudre la sélection endogène sans cette hypothèse de linéarité. Il s'appuie pour ce faire sur une forme semi-paramétrique sur le modèle de sélection, et sur l'existence d'un régresseur continu fortement exogène (c'est-à-dire indépendant des erreurs du modèle de sélection, conditionnellement aux covariables) et satisfaisant une condition de large support. Notons que la forme semi-paramétrique du modèle de sélection impose que la probabilité de sélection tende vers 0 ou 1 lorsque le régresseur en question tend vers l'infini. Dans un cadre de non-réponse non-ignorable, Fay (1986), Baker & Laird (1988) et Park & Brown (1994) montrent que la distribution de variables qualitatives peut être identifiée en utilisant des restrictions du modèle log-linéaire entre les variables dépendantes et explicatives. Enfin, l'attrition endogène dans des données de panel peut être traitée en imposant des restrictions semi-paramétriques (voir par exemple Rotnitzky et al., 1998, Scharfstein et al., 1999, ou Hirano et al., 2001).

1.2.3 Résultats nouveaux

Je présente dans le chapitre 1 deux essais sur les problèmes de régression non-paramétrique instrumentale. Le premier s'intéresse à la condition de complétude

(1.2.4)[14]. Cette hypothèse nécessaire à l'identification de plusieurs modèles non-paramétriques restait en effet relativement abstraite, puisque seuls les cas où le support de (X, Z) est fini et où la densité du couple s'écrit sous forme exponentielle avaient été considérés dans la littérature (cf. Newey & Powell, 2003). Si l'on note $X = (X_0, W)$ et $Z = (Z_0, W)$ pour tenir compte des éléments communs éventuels, je considère ici un modèle entre X et Z de la forme

$$X_0 = \mu(\nu(Z) + \varepsilon).$$

où Z_0 est indépendant de ε conditionnellement à W et où $\nu(., W)$ satisfait une condition de large support. Ce modèle comprend les modèles non-paramétriques additifs mais aussi un grand nombre de modèles non-linéaires standards (probit, tobit, modèles de durée...). Dans ce cadre, plusieurs versions de la complétude sont obtenues, en fonction des hypothèses de régularité considérées sur ε. La condition de complétude standard est satisfaite sous des conditions restrictives, alors que la condition de complétude sur l'espace des fonctions h bornées l'est sous des conditions beaucoup plus faibles. Je montre également, sous des conditions relativement faibles, la condition de complétude sur l'espace des fonctions h dominées par un polynôme. La condition de large support peut être relâchée, au prix d'hypothèses supplémentaires sur ε. Enfin, ces résultats sont appliqués à l'identification de la régression non-paramétrique instrumentale. L'identification est obtenue grâce à une structure triangulaire du modèle, et sous des hypothèses de séparabilité additive. La méthode permet de traiter le cas de régresseurs limités, une situation où l'approche par les fonctions de contrôle est mise en échec.

Le deuxième essai développe une nouvelle méthode pour traiter de la sélection endogène[15]. Quand la sélection dépend directement de la variable d'intérêt, il peut être difficile d'exhiber un instrument influant la sélection mais pas directement la variable d'intérêt, et donc vérifiant la condition (1.2.9). L'article montre qu'une autre stratégie instrumentale, basée sur l'indépendance entre les instruments et la sélection conditionnellement à la variable d'intérêt, est possible. Cette condition a été utilisée précédemment par Chen (2001), Tang et al. (2003) et Smith & Ramalho (2007) dans le cadre de modèles paramétriques entre des covariables et une variable dépendante affectée par une non-réponse non-ignorable. L'essai présent généralise

[14]Ce papier est à paraître dans *Econometric Theory*.

[15]Ce papier est en révision à *Journal of Econometrics*.

leur résultat à un cadre non-paramétrique, en mettant en évidence la condition clé de validité de l'instrument, qui est la complétude de la variable dépendante pour cet instrument. En adaptant les résultats de la partie précédente, je montre qu'un modèle de sélection où la variable d'intérêt peut se décomposer de manière additive, et où l'instrument vérifie une condition de large support, est identifié non-paramétriquement.

La condition instrumentale retenue peut, dans certaines applications, sembler trop restrictive. L'essai montre cependant qu'à la différence de l'hypothèse de sélection exogène, cette condition est testable. De plus, si on l'affaiblit en une hypothèse de montonicité des effets, il est possible d'obtenir des bornes sur un ensemble de paramètres de la distribution de la variable d'intérêt. Comme dans le cadre de Manski & Pepper (2000), ces bornes ne nécessitent pas que la variable d'intérêt soit elle-même bornée.

L'essai considère également l'estimation non-paramétrique du modèle. La probabilité conditionnelle de sélection, qui permet par la suite d'identifier le modèle, est solution d'une équation intégrale, c'est-à-dire d'un problème inverse mal posé. L'estimation non-paramétrique est alors basée sur une méthode de régularisation de Tikhonov. Enfin, la méthode est appliquée à l'évaluation des effets du redoublement à l'école primaire en France.

1.2.4 Perspectives

Le premier essai présente une classe relativement générale de modèles où la condition de complétude (1.2.4) est vérifiée. La condition (1.2.7) de Chernozhukov & Hansen (2005) étant proche, on peut se demander si les résultats obtenus ici peuvent être transposables à leur modèle. Par ailleurs, l'essai ne considère pas de test de cette hypothèse, analogue au test de nullité de la corrélation entre variable endogène et instrument dans le modèle linéaire. Si un test peut être facilement mené lorsque X et Z sont à support fini[16], cette question reste posée dans le cas général.

[16]Il suffit en effet d'effectuer un test de rang sur la matrice aléatoire de terme général $\widehat{P}(X = x_i|Z = z_j)$, où $\widehat{P}(X = x_i|Z = z_j)$ est la fréquence observée de x_i parmi les données telles que $Z = z_j$ (voir Robin & Smith, 2000, pour l'implémentation de tels tests de rang).

Le deuxième essai, quant à lui, étudie une situation où le choix dépend directement de la variable dépendante. Une limite de ce modèle est que l'individu peut, au moment du choix, ignorer la valeur précise de cette variable, du fait de chocs inobservables futurs affectant cette variable. Une question d'importance est donc de savoir si les méthode utilisées ici peuvent s'étendre à un modèle de Roy généralisé où la sélection dépend des anticipations des agents sur cette variable dépendante. Par ailleurs, je montre que lorsque la condition d'indépendance est affaiblie en une condition de monotonicité, des bornes finies peuvent être calculées sur un ensemble de paramètres vérifiant certaines conditions. Ces conditions sont cependant abstraites, et il paraît souhaitable, à l'avenir, de les étudier plus attentivement.

1.3 L'identification des modèles de contrats

1.3.1 Motivation

Nous considérons maintenant l'identification des modèles de contrats. Depuis le papier fondateur d'Akerlof (1970), une attention considérable a été dévolue aux asymétries d'information et à leurs conséquences économiques. Celle-ci s'est traduite par l'émergence de la théorie des contrats, qui modélise le comportement d'un agent et d'un principal lorsqu'il existe une différence informationnelle entre eux. Les trois modèles principaux de cette théorie sont le modèle de sélection adverse, où le principal ignore le type de l'agent, le modèle d'aléa moral, où l'action de l'agent est inconnue du principal, et le modèle de signal, où le principal dispose d'une information privée qu'il souhaite divulguer à l'agent. Cette théorie s'applique à de nombreux domaines comme la régulation (cf. Laffont & Tirole, 1993), la tarification non-linéaire (Wilson, 1993), la taxation optimale (Diamond, 1998), les choix éducatifs (Spence, 1973) etc.

Si les applications empiriques de ces théories sont maintenant nombreuses (voir Chiappori & Salanié, 2002, pour une revue de littérature), il existe encore très peu de résultats d'identification généraux de ces modèles, mis à part le cas des enchères (voir ci-dessous). Les enjeux sont pourtant importants. Si un modèle n'est pas identifié non paramétriquement, les résultats obtenus, et par conséquent les préconisations de politique publique, peuvent être sensibles à la forme paramé-

trique retenue. De même, dans la mesure où la présence d'asymétries d'information, ainsi que leur nature exacte (i.e., asymétries sur le type ou sur l'action de l'agent), ont des conséquences en termes d'intervention publique[17], il est important de savoir si l'on peut tester non-paramétriquement par exemple la présence d'asymétrie d'information ou la sélection adverse par rapport à l'aléa moral. On ne peut en effet distinguer, dans un test paramétrique, ce qui relève de la spécification paramétrique de la validité des hypothèses théoriques sous-jacentes.

1.3.2 Etat de l'art

Les modèles d'enchères[18]

Une grande partie des résultats d'identification et de testabilité des modèles de contrats a été obtenue sur les modèles d'enchères[19]. Ceci provient sans doute du fait que les enchères sont un exemple idéal de jeu aux règles claires et simples et dont la théorie microéconomique est relativement balisée. On peut, pour faire simple, distinguer deux modèles théoriques d'enchères. Dans le modèle à valeurs privées (cf. Vickrey 1961), chaque enchérisseur accorde une valeur différente à l'objet. Cette valeur est connue d'eux seulement ; seule la distribution des valeurs individuelles est connaissance commune. Dans le modèle à valeur commune (cf. par exemple Rothkopf, 1969 ou Wilson, 1969), au contraire, l'objet a une valeur intrinsèque (liée à l'exploitation de ce bien ou à un marché de revente par exemple) inconnue des enchérisseurs au moment de l'enchère. Ceux-ci observent uniquement un signal sur la valeur de ce bien. Les signaux reçus sont indépendants d'un joueur à l'autre[20].

Il existe par ailleurs quatre principaux types d'enchères[21] Dans l'enchère anglaise, un commissaire-priseur augmente le prix du bien jusqu'à ce qu'il ne reste plus

[17] Ainsi, l'existence d'une franchise médicale est justifiée en présence d'aléa moral mais pas de sélection adverse.

[18] Cette discussion est inspirée de Février (2003). Pour une revue récente de la littérature, voir également Athey & Haile (2007).

[19] Les enchères peuvent en effet être considérées comme un contrat passé entre un vendeur et des acheteurs potentiels. Le contrat passé correspond au type de l'enchère (i.e., enchères anglaises, hollandaises, au premier et au second prix, cf. ci-dessous).

[20] Il existe en fait un troisième modèle qui englobe les deux autres, le modèle à valeur affiliée (cf. Wilson, 1977, et Milgrom & Weber, 1982).

[21] La discussion porte ici uniquement sur les enchères où une seule unité indivisible est mise

qu'un enchérisseur. Le bien est alors acquis par ce dernier au prix atteint. L'enchère hollandaise est en quelque sorte l'inverse de cette dernière : le prix descend jusqu'à ce qu'un joueur se propose pour la somme correspondante. Dans l'enchère au premier prix, chaque enchérisseur consigne son offre sous un pli scellé. Le bien est acquis par l'enchérisseur ayant indiqué la somme la plus élevée, à ce prix. Enfin, l'enchère au second prix est également une enchère sous plis scellés, mais le prix payé correspond à la deuxième meilleure offre. Notons que les informations disponibles varient d'un type d'enchère à l'autre. Seule l'offre la plus élevée est observée pour les enchères anglaise et hollandaise, alors que toutes les offres sont a priori observables pour les enchères au premier et au second prix.

La recherche a surtout été active sur les enchères à valeurs privées. L'identification du modèle avec enchères au second prix, anglaise ou hollandaise et joueurs neutres au risques (symétriques ou asymétriques) a été établie par Athey & Haile (2002)[22]. L'identification de l'enchère au premier prix, avec joueurs neutres au risques symétriques ou asymétriques a été obtenue quant à elle par Guerre et al. (2000). Ces résultats s'appuient sur l'hypothèse forte de neutralité au risque des joueurs. Guerre et al. (2008) prouvent qu'en l'absence d'une telle hypothèse, le modèle n'est pas identifié en général. L'identification non-paramétrique peut néanmoins être rétablie par le biais de variations exogènes, comme celle du nombre d'enchérisseurs.

La recherche a été moins active sur le modèle à valeur commune. Ceci s'explique sans doute par les résultats négatifs de Laffont & Vuong (1996) et d'Athey & Haile (2002). Le premier ne souligne en fait qu'un problème de spécification du modèle : les signaux reçus par les joueurs étant inobservés, le modèle reste inchangé par toute transformation strictement croissante de ces signaux. Une normalisation de ces signaux est donc nécessaire. Athey & Haile (2002) approfondissent ce résultat en montrant que même lorsque cette normalisation est effectuée, le modèle n'est pas identifié en général dans l'enchère anglaise[23]. Li et al. (2000) se restreignent à un modèle semi-paramétrique d'enchère à valeur commune. Ils supposent que le signal de chaque joueur peut se décomposer comme le produit d'une compo-

en vente. Les enchères multi-unités ont également été étudiées d'un point de vue empirique (cf. Athey & Haile (2007) pour une revue de littérature).

[22]Dans l'enchère anglaise, le modèle n'est toutefois que partiellement identifié lorsqu'il existe des incréments minimaux pour réenchérir (cf. Haile & Tamer 2003).

[23]Il n'existe toutefois pas de résultat de la sorte dans les enchères au premier et deuxième prix ou dans l'enchère hollandaise.

sante commune (la valeur du bien) et d'une composante idiosyncratique (le signal spécifique). A l'aide de restrictions supplémentaires, ils montrent que leur modèle est identifiable et proposent une procédure non-paramétrique en deux étapes pour estimer la densité des deux composantes. Récemment, Février (2007) a également proposé une alternative non-paramétrique pour une classe particulière de densités conditionnelles du signal. Il obtient l'identification du modèle en réduisant la dimension du problème d'une fonction de deux variables à deux fonctions d'une seule variable.

Les modèles de contrat

Si l'économétrie des enchères a fait l'objet d'une abondante littérature, il n'existe encore que peu de papiers sur les modèles structurels de contrats. Les modèles de régulation structurels ont été estimés notamment par Wolak (1994) dans un cadre paramétrique et par Lavergne & Thomas (2005) dans un cadre semi-paramétrique. Perrigne & Vuong (2004) montrent que le modèle de régulation de Laffont & Tirole (1993) est en fait identifié non-paramétriquement, principalement sous la condition que la fonction de coût a une forme séparable. Plusieurs travaux se sont également intéressés aux contrats passés entre les employeurs et les employés. Ferrall et Shearer (1999) étudient ainsi l'optimalité des contrats de mineurs de Colombie Britannique, en estimant un modèle structurel paramétrique de sélection adverse[24]. Paarsch & Shearer (2000) estiment un modèle similaire à partir de données sur des planteurs de bois de Colombie Britannique. Enfin, Perrigne et Vuong (2007) s'intéressent à l'identification et à l'estimation d'un modèle de tarification non-linéaire. Ils montrent en particulier que lorsque le programme du principal prend une certaine forme, le modèle est identifié non-paramétriquement. Ils appliquent leurs résultats aux contrats passés entre les pages jaunes américaines et les annonceurs.

[24]Ils estiment en effet un modèle de "faux aléa moral", suivant la terminologie de Laffont & Martimort (2002). Dans ce modèle, la production finale dépend de l'effort de l'agent et d'une composante inobservée par le principal. Il s'agit bien d'un modèle de sélection et non d'aléa moral car l'agent observe ex ante cette composante, et ne subit donc pas de risque dans sa production.

1.3.3 Résultats nouveaux

Le deuxième chapitre de la thèse présente deux essais indépendants[25]. Le premier s'intéresse aux enchères à valeur commune sous plis scellés. Il complète ainsi les papiers de Li et al. (2000) et Février (2007) sur le sujet. Nous montrons que l'on peut obtenir l'identification non-paramétrique du modèle si le support de la distribution des signaux conditionnelle à la valeur V du bien varie de façon strictement croissante avec V. Nous renforçons ainsi la propriété de rapport de vraisemblance monotone - condition imposée par Milgrom & Weber (1982) pour obtenir un équilibre symétrique à l'enchère - qui impose que ce support varie de manière croissante. Nous prouvons tout d'abord qu'en utilisant les enchères pour lesquelles l'écart entre les offres maximales et minimales est maximal, il est possible d'identifier ponctuellement la valeur de $g(V)$, où g est une transformation strictement croissante inconnue. En utilisant les variations de la troisième offre, on peut alors identifier la distribution du signal conditionnellement à $g(V)$. Nous montrons ensuite que la distribution de $g(V)$ peut être obtenue de manière unique à partir de l'équation intégrale reliant la densité des signaux à leur densité conditionnelle et à la densité de $g(V)$. Enfin, la condition d'équilibre permet dans un second temps d'identifier la transformation g. Nous montrons que ce résultat d'identification peut être obtenu même si les signaux ne sont pas indépendants conditionnellement à V, comme dans le modèle de droits de minerais, et même si l'un des deux extrema seulement du support du signal varie avec V.

La deuxième partie du chapitre étudie le modèle de sélection adverse. Dans la lignée des premiers travaux sur les enchères, cet essai vise à caractériser les ingrédients principaux permettant l'identification du modèle. Nous considérons donc un modèle simple, inspiré de celui de Laffont & Martimort (2002), mais ayant l'avantage, par rapport par exemple aux travaux de Perrigne & Vuong (2004, 2007), de la généralité. Ce modèle principal-agent est défini par la fonction objectif du principal, la fonction d'utilité des agents et la fonction de répartition des types des agents. Nous montrons qu'en général, l'identification non-paramétrique de ce modèle nécessite la connaissance d'au moins une de ces trois fonctions. Ceci peut être le cas dans des modèles de tarification non-linéaire, de contrats financiers ou de régulation.

[25] Ce chapitre est le fruit d'une collaboration avec Philippe Février.

Nous étudions également le pouvoir identifiant des changements exogènes dans la fonction objectif du principal. Plusieurs papiers empiriques se sont déjà appuyés sur de telles variations pour estimer ou tester des modèles de contrats, mais dans un cadre non-structurel. L'objectif est donc ici de caractériser précisément ce qui est identifié non-paramétriquement à partir de telles variations, à l'instar de plusieurs papiers sur les enchères (cf. par exemple Guerre et al., 2008, Bajari & Hortaçsu (2005), 2005, ou encore Shum & Hu, 2008). Nous montrons qu'avec un changement exogène, le modèle est complètement ou partiellement identifié, suivant que les tarifs marginaux des deux contrats se croisent ou non. Si ces transferts marginaux ne se croisent pas, un changement exogène supplémentaire peut permettre d'identifier complètement le modèle.

Enfin, nous appliquons cette méthode au problème d'incitations dans les firmes. Plus précisément, nous testons l'optimalité des contrats passés entre l'INSEE (Institut National de la Statistique et des Etudes Economiques) et ses enquêteurs. Dans cet exemple, le modèle est partiellement identifié en l'absence de structure sur la fonction objectif du principal. Nous estimons non-paramétriquement les bornes des fonctions structurelles et prouvons que ces bornes sont convergentes. En utilisant la forme de la fonction objectif du principal, nous testons et rejetons dans un second temps l'optimalité des contrats. Nous estimons cependant que l'utilisation de simples contrats linéaires à la place des contrats optimaux n'entraîne une perte que de 10%, ce qui pourrait expliquer pourquoi ces contrats sont si répandus.

1.3.4 Perspectives

Le premier essai montre que le modèle d'enchère à valeur commune peut être identifié en s'appuyant sur la variation de support des signaux. D'autres stratégies sont a priori envisageables. Une piste intéressante serait d'utiliser des résultats récents sur les modèles à erreurs de mesure comme ceux de Hu & Schennach (2008). Ce modèle a en effet des liens très étroits avec le modèle d'enchère à valeur commune. Les résultats de Hu & Schennach (2008) en particulier ont l'avantage d'être très généraux, même s'ils reposent sur des hypothèses relativement abstraites.

Les applications potentielles du deuxième essai sont a priori nombreuses. Une piste prometteuse serait de déterminer les contrats d'assurance santé optimaux et de tester le modèle de sélection adverse dans ce domaine, en utilisant par exemple les

données du Random Health Experiment (cf. Manning et al. 1987). Une autre application intéressante serait l'estimation structurelle, et si possible le test, du modèle de régulation. Enfin, une voie de recherche importante concerne l'identification non-paramétrique du modèle d'aléa moral.

1.4 Le problème d'identification des effets de pairs

1.4.1 Motivation

Nous abordons maintenant la question de l'identification des modèles d'interaction sociale[26]. Cette appellation se réfère, au sens large, à des modèles dans lesquels les actions ou les préférences des autres influent sur notre propre utilité (pour une définition plus précise, cf. par exemple Manski, 2000)[27]. Ces modèles sont particulièrement utiles dans l'étude, par exemple, de la réussite scolaire. En effet la composition optimale des classes - classes de niveau ou collège unique pour caricaturer le débat - dépend de l'existence et de la nature des effets de pairs (cf. Cooley, 2006). De même, ces modèles sont utiles pour comprendre - et éventuellement infléchir - la consommation de tabac ou d'alcool (cf. Krauth, 2006) ou encore des décisions de départ à la retraite (cf. Duflo0 & Saez, 2003).

1.4.2 Etat de l'art

Modèles linéaires

Les modèles linéaires, bien que restrictifs, ont été abondamment étudiés car ils permettent de bien appréhender la nature des problèmes que la prise en compte des interactions sociales soulève en termes d'identification. Manski (1993) est le premier à considérer cette question, en s'appuyant sur le modèle suivant :

$$y = \alpha_0 + x'\beta_{10} + E(y|g)\lambda_0 + E\left(x|g\right)'\beta_{20} + u,\ E(u|g,x) = g'\delta_0. \tag{1.4.1}$$

[26]Pour des revues récentes de la littérature sur cette question, cf. Blume & Durlauf (2005) et Soetevent (2006).

[27]En ce sens, tous les modèles de théorie des jeux par exemple, devraient y être inclus, mais nous verrons ci-dessous que seule une classe assez restreinte de modèles sera en fait retenue. Il semble cependant difficile de proposer une définition précise de cette classe en question.

Dans cette équation, $E(y|g)\lambda_0$ se réfère aux effets de pairs endogènes, $E(x|g)'\beta_{20}$ correspond aux effets de pairs exogènes et $g'\delta_0$ est appelé l'effet contextuel. Manski & Pepper (2000) montre qu'un tel modèle n'est pas identifié. Plus précisément, seuls les paramètres composites $\alpha_0/(1-\lambda_0)$, $(\beta_{20}+\lambda_0\beta_{10})/(1-\lambda_0)$, $\delta_0/(1-\lambda_0)$ et β_{10} peuvent être obtenus, sous l'hypothèse supplémentaire que $[1, E(x|g), g, x]$ sont linéairement indépendants. Le résultat n'est donc pas complètement négatif car le paramètre $(\beta_{20}+\lambda_0\beta_{10})/(1-\lambda_0)$ qui représente l'effet global des pairs présente un intérêt en termes de politiques publiques. De plus, si l'on suppose qu'il existe une variable ayant un effet direct non nul mais pas d'effet social exogène, alors tous les paramètres du modèle sont identifiés. Toutefois, la condition d'indépendance linéaire entre $1, E(x|g), g$ et x est assez forte. Ainsi, si x est une fonction de g, si $E(x|g) = E(x)$ ou si $E(x|g)$ est une fonction linéaire de g (ce qui est le cas dès que g est discret et s'écrit comme un vecteur d'indicatrice), le paramètre composite $(\beta_{20}+\lambda_0\beta_{10})/(1-\lambda_0)$ n'est plus identifié. La dépendance entre x et g doit donc être « modérée » et non-linéaire.

Lorsque les groupes sont finis, il semble discutable de stipuler une dépendance à l'espérance mathématique et non à la moyenne des individus du groupe. Pour prendre en compte cette critique, Graham & Hahn (2005) proposent la variante suivante du modèle précédent :

$$y_{ri} = x'_{ri}\beta_{10} + \overline{y}_r\lambda_0 + \overline{x}_r\beta_{20} + \alpha_r + \varepsilon_{ri},\ E(x'_{ri}\varepsilon_{rj}) = 0 \tag{1.4.2}$$

où r désigne l'indice du groupe, i celui de l'individu, n_r est la taille du groupe et α_r est l'effet fixe du groupe. Ainsi, la variable de groupe est maintenant une collection d'indicatrices. Par construction, $\overline{x}_r$ et $1_{i\in r}$ ne sont pas linéairement indépendants. Le problème est donc plus difficile que précédemment puisque, avec les notations précédentes, $E(x|g)$ est une fonction linéaire de g. Formellement, l'identification de ce modèle est identique à celle d'un modèle linéaire de panel avec effets fixes et variables constantes dans le temps. Une possibilité est de s'appuyer sur une stratégie instrumentale. Ainsi, si l'on dispose d'instruments z_r corrélés à $\overline{x}_r$ mais pas à α_r, l'équation between permet d'identifier $(\beta_{10}+\beta_{20})/(1-\lambda_0)$[28]. Cette condition est remplie par exemple si l'affectation dans les groupes est aléatoire (cf. par exemple Sacerdote, 1996, Katz et al., 2001, ou Ludwig et al., 2001, pour des applications

[28]Notons que, comme toujours dans ces modèles, l'effet direct β_{10} est également identifié, par l'équation within.

empiriques), les variables $\overline{x}_r$ pouvant dans ce cas être utilisées comme instrument. Cependant, même si l'existence d'un instrument est assurée, on ne peut identifier séparément β_{20} et λ_0 sans relation d'exclusion supplémentaire telle que l'existence d'une composante de β_{20} nulle. Ce résultat renforce donc la conclusion négative de Manski. Un autre problème soulevé par Graham et Hahn est celui de la non-réponse : dès qu'un membre du groupe n'est pas observé, $\overline{x}_r$ est mesuré avec erreur, ce qui peut biaiser l'estimation de $(\beta_{10} + \beta_{20})/(1 - \lambda_0)$. Ceci sera en particulier le cas lorsque l'affectation dans les groupes est aléatoire et l'instrument utilisé est $z_r = x_r$. Ainsi, même dans cette situation a priori favorable, l'estimation du paramètre composite d'effet de pair peut être difficile voire impossible.

Dans les modèles précédents, les résultats d'identification reposent sur des relations d'exclusion qui sont par nature contestables. Récemment cependant, plusieurs papiers ont mis en évidence des variantes de ces modèles où, de par l'hétérogénéité des effets d'interactions, les paramètres sont identifiés sans condition supplémentaire. Ainsi, Lee (2007) considère le modèle (1.4.2) où les moyennes sont calculées non pas sur l'*ensemble* des individus du groupe mais sur le groupe privé de l'individu i[29]. Cette modification a priori mineure a d'importantes conséquences en termes d'identification. En effet, les paramètres réduits de l'équation within dépendent alors de la taille des groupes. En utilisant la variation de ces tailles de groupes, il est alors généralement possible de retrouver β_{20} et λ_0.[30].

L'idée de ne pas considérer l'individu i dans le membre de droite a été récemment généralisée par Bramoullé et al. (2009). Ces auteurs s'appuient sur les variations de groupes de références (c'est-à-dire des groupes influençant directement les individus) pour identifier le modèle. Les modèles précédents supposaient que les individus d'un même groupe étaient influencés par les mêmes personnes ou presque (i.e., le reste du groupe). Mais ceci ne représente qu'une situation possible d'in-

[29]Notons également que cette modification avait déjà été considérée par Moffitt (2001). Cependant, ce dernier supposait que la taille des groupes était constante. Dans ce cas, le problème d'identification est similaire à celui de Graham et Hahn (2005).

[30]L'idée d'utiliser les variations (exogènes) de taille de groupe a également été exploitée par Glaeser et al. (1996) et surtout Graham (2008). Graham (2008) montre ainsi que dans un modèle proche de (1.4.2), il est possible d'identifier un effet de pair composite sous l'hypothèse d'homoscédasticité des résidus en la taille des groupe. L'avantage de cette approche est que le test statistique sous-jacent d'absence d'effet de pair est en général beaucoup plus puissant que les tests standards.

teractions sociales. Dans le cas général, Bramoullé et al. (2009) montrent que le modèle sera identifié dès qu'il existe trois individus (i, j, k) tels que i soit influencé par j, j par k tandis que i n'est pas influencée par k (on parle dans ce cas de *triade intransitive*). Même lorsque cette condition n'est pas vérifiée, le modèle est presque toujours identifié dès que le réseau n'est pas un groupe (cf. proposition 3 de Bramoullé et al., 2009).

Ainsi, le résultat négatif de Manski n'est pas robuste à l'hétérogénéité des groupes de références individuels. Cependant, le modèle de Bramoullé et al. (2009) repose sur l'hypothèse que la matrice de liens est connue. Il est donc nécessaire de disposer de données très fines sur les relations sociales des individus. La remarque de Manski (1993) quant à la nécessité, pour progresser dans la compréhension des effets de pairs, d'obtenir des données plus précises sur les liens sociaux, semble donc plus que jamais d'actualité.

Modèles non-linéaires

Les restrictions imposées par le modèle linéaire semblent en partie ad hoc. L'homogénéité des effets de pairs, en particulier, est problématique. Il n'y a en effet aucune raison pour que tous les élèves d'une classe réagissent de la même façon à leur environnement, par exemple. De plus, comme nous l'avons souligné précédemment, le modèle (1.4.1) implique que la composition des groupes n'a pas d'influence sur le résultat moyen. Si l'on souhaite rester a priori agnostique sur cette question, il est nécessaire de considérer un modèle plus général. Le modèle linéaire est également inadapté pour considérer des choix discrets où l'influence des pairs est potentiellement importante, comme les décisions de commencer à fumer, de partir à la retraite ou encore de poursuivre ses études.

Cependant, au vu des résultats précédents (et en particulier ceux de Manski) on peut s'interroger sur la pertinence d'éudier des modèles non-linéaires. Si déjà les modèles linéaires ne sont pas identifiés, pourquoi s'intéresser à des situations plus complexes ? En fait, les modèles linéaires tels que (1.4.1) ou (1.4.2) sont identifiés dès que l'on exclut les effet exogènes. Il est donc légitime de se demander si ce résultat subsiste lorsqu'on lève l'hypothèse de linéarité. Nous abordons dans cette partie l'identification du modèle non-paramétrique général, puis un cas particulier de modèle non-linéaire abondamment étudié : le modèle binaire.

Considérons tout d'abord un modèle dans lequel y dépend de covariables x et du niveau moyen des autres $E(y|g)$ à travers une fonction quelconque f :

$$E(y|g,x) = f(E(y|g),x)$$

Par souci de simplicité, nous supposons comme Manski (1993), outre l'absence d'effet exogène, qu'il n'y a pas d'effets corrélés. Dans ce cadre non-paramétrique, il s'agit de savoir si le contraste $T(e_1,e_0,\xi) = f(e_1,\xi) - f(e_0,\xi)$ est identifié, pour tout ($e_0 \neq e_1$) (resp. ξ) dans le support de $E(y|g)$ (resp. de x). Pour cela, écrivons l'équilibre social du modèle :

$$E(y|g) = \int f(E(y|g),x)dP(x|g)$$

Si cette équation admet une seule solution, Manski (1993) montre que T n'est pas identifié si g est fonction de x, si x est une fonction de g ou si g est indépendant de x. Dans ces trois cas en effet, $E(y|g)$ est une fonction (éventuellement constante) de x. Comme dans le modèle linéaire, une condition nécessaire pour obtenir l'identification de $T(.)$ est donc que la dépendance entre x et g soit « modérée ».

Enfin, si les individus sont face à un choix binaire, et que leur utilité dépend de manière linéaire de leur anticipation (rationnelle) du choix moyen des autres, on est conduit (cf. Brock & Durlauf (2001) au modèle suivant :

$$y = 1\{\alpha_0 + x'\beta_{10} + P(y=1|g)\lambda_0 + g'\delta_0 + \varepsilon \geq 0\}.$$

On suppose par la suite que $-\varepsilon$ est indépendant de (g,x), de fonction de répartition H connue. Avant d'étudier l'identification de ce modèle, on peut s'interroger sur sa *cohérence*. Il s'agit de savoir, en d'autres termes, si pour toute valeur de g et des paramètres $(\alpha_0,\beta_{10},\lambda_0,\delta_0)$, il existe une équilibre social $P(y=1|g)$ solution de l'équation en $p(g)$:

$$p(g) = \int H(\alpha_0 + x'\beta_{10} + p(g)\lambda_0 + g'\delta_0)dP(x|g).$$

Comme l'a mis en évidence Manski (1993), il existe toujours une solution à cette équation, mais il n'y a pas unicité en général[31]. Le modèle est donc cohérent mais

[31]Plus précisément, Brock et Durlauf (2001) montrent que lorsque ε suit une loi logistique, le nombre d'équilibres sociaux vaut 1 ou 3, suivant la valeur des paramètres.

incomplet, dans la mesure où la donnée de (x, ε) ne permet pas nécessairement de calculer la variable d'intérêt y[32].

L'identification du modèle a été étudiée par Brock & Durlauf (2001, 2007)). Leur résultat principal est que le modèle est identifié si g et x sont linéairement indépendants, si aucune des composantes de g n'est à support borné (ce qui exclut en particulier les variables discrètes), et si $P(y = 1|g)$ n'est pas constant en g. Ce résultat est parfois opposé aux conclusions négatives de Manski (1993) et considéré comme une conséquence positive de la non-linéarité du modèle. Cette affirmation est toutefois inexacte car le modèle (1.4.1) est également identifié lorsqu'on suppose $\beta_{20} = 0$. Le résultat de Brock et Durlauf est même plus restrictif que celui de Manski (1993) puisqu'il s'appuie sur une condition de support sur g très limitative, même s'il reste possible que cette condition ne soit en fait pas nécessaire. Notons en revanche qu'il n'est pas nécessaire de supposer H connue pour obtenir l'identification du modèle (cf. Brock et Durlauf, 2004).

1.4.3 Résultats nouveaux

Le chapitre trois s'intéresse, à la suite de Lee (2007), à l'identification des effets exogènes et endogènes de pairs à partir des variations de taille de groupe[33] . L'objectif de ce chapitre est double. Tout d'abord, nous réexaminons et étendons les résultats d'identification de Lee (2007) Nous montrons que dans le modèle linéaire, les hypothèses cruciales sont, d'une part, la connaissance par l'économètre des tailles de groupe et, d'autre part, le fait qu'il existe au moins trois tailles de groupes différentes. En revanche, contrairement à la situation du modèle (1.4.2) il n'est pas nécessaire d'observer tous les membres du groupe. Il n'est pas non plus nécessaire de supposer l'homoscédasticité des résidus en général. Cependant, dans certains cas particuliers, l'identification est perdue et cette dernière hypothèse permet alors de retrouver l'ensemble des paramètres.

Nous étudions également l'identification d'un modèle binaire où seul le signe de y_{ri} est observé. Notre modèle se distingue de celui de Brock & Durlauf (2001) dans

[32]Cette situation survient également dans les modèles de jeux d'économie industrielle. Voir la section 1.5 pour une discussion plus approfondie de cette question

[33]Ce chapitre, qui est le résultat d'une collaboration avec Laurent Davezies et Denis Fougère, est en révision à *Econometrics Journal*.

la mesure où nous supposons que les effets de pairs transitent via les variables latentes plutôt que par les résultats observables.Cette situation est réaliste lorsque le modèle est réellement linéaire, mais que, du fait de la limitation des données, seuls des résultats binaires sont observés. Notons que le modèle présente des similarités avec le probit spatial. A la différence de ce dernier, cependant, le modèle inclut des effets de pairs exogènes et des effets fixes de groupe[34]. Nous obtenons là encore l'identification des effets de pairs exogènes en utilisant les variations de taille de groupe. Ce résultat est de plus obtenu sans hypothèse paramétrique sur la distribution des erreurs. Cependant, du fait de la perte d'information par rapport au modèle précédent, les effets de pairs endogènes ne peuvent être identifiés sans restriction supplémentaire. Nous montrons qu'une hypothèse d'homoscédasticité permet d'obtenir un tel résultat.

D'autre part, nous développons une méthode d'estimation paramétrique du modèle binaire, complétant ainsi le papier de Lee (2007) qui se concentre sur l'estimation paramétrique du modèle linéaire. Nous montrons que sous l'hypothèse de normalité des résidus, une estimation par maximum de vraisemblance peut être facilement implémentée en utilisant l'algorithme GHK.

1.4.4 Perspectives

Le chapitre présente des résultats d'identification de modèles semi-paramétriques d'interactions sociales[35], basés sur la variation des tailles de groupe. La question de savoir dans quelle mesure ces résultats dépendent de la forme fonctionnelle des modèles reste cependant posée. Une perspective de recherche future est donc d'étudier l'identification non-paramétrique de modèles d'interactions sociales, basée sur ces variations de taille de groupe.

La méthode proposée ici possède l'avantage, par rapport aux stratégies instrumentales usuelles, de ne pas s'appuyer sur des relations d'exclusion. Elle repose en revanche sur des restrictions fonctionnelles, qui peuvent être violées empiriquement. Une piste future importante serait donc de tester la validité de ces hypothèses, en s'appuyant par exemple sur des données expérimentales pour lesquelles les effets de pairs peuvent être mesurés simplement.

[34]Une deuxième différence est que nous n'imposons pas de famille de lois sur les résidus.

[35]En effet, les résultats obtenus ne dépendent pas de la loi des résidus des modèles considérés.

Chapitre 2

Identification of nonparametric instrumental models

2.1 On the completeness condition in nonparametric instrumental problems

2.1.1 Introduction

Let X and Z denote two random elements. X will be said to be complete for Z if, for all measurable real functions h such that $\mathbb{E}[|h(X)|] < +\infty$,

$$\Big(\mathbb{E}[h(X)|Z] = 0 \quad \text{a.s.}\Big) \Longrightarrow \Big(h(X) = 0 \quad \text{a.s.}\Big). \tag{2.1.1}$$

X will be bounded complete (resp. P-complete) for Z if the same holds for any bounded h (resp. for any h bounded by a polynomial).[1] Completeness is equivalent to the injectivity of the conditional expectation operator. Thus, not surprisingly, it has appeared to be a key identifying condition in nonparametric instrumental problems. Applications include nonparametric instrumental regression under additive separability (see Newey & Powell, 2003, Darolles et al., 2007 and Blundell

[1]This terminology is in analogy with the notion of complete statistic (see e. g. Lehmann & Scheffé, 1947). Recall that a statistic T is said to be complete (resp. bounded complete) for a statistical model $(P_\theta)_{\theta\in\Theta}$ if for all h (resp. all bounded h), $E_\theta[h(T)] = 0$ for all $\theta \in \Theta$ implies that $h(T) = 0$ a.s. Thus, Z plays the role of θ in equation (2.1.1). Note also that (2.1.1) is sometimes referred to as a strong identification condition (see e. g. Florens et al., 1990).

et al., 2007),[2] local instrumental variables (see Florens et al., 2003) and nonclassical measurement error problems (see Chen & Hu, 2006 and Hu & Schennach, 2008).[3]

This dependence condition is quite abstract though, and a characterization or at least sufficient conditions on the joint distribution of (X, Z) are desirable. Newey & Powell (2003) address the finite support and exponential families cases, but results are still lacking to properly define completeness in terms of dependence between the two variables. The aim of this paper is to go one step in this direction by considering a nonparametric model on (X, Z) for which an additive separability and a large support condition hold. Building on the results of Mattner (1992, 1993) on the completeness of location families, I show that different versions of completeness can be obtained, depending on which regularity conditions are imposed on the error term. Bounded and P-completeness only require mild assumptions, whereas completeness is restrictive. This contrast between the different kinds of completeness is in line with previous results of the statistical literature (see e.g. Hoeffding, 1977, Lehmann, 1986, and Mattner, 1993) and has also been acknowledged by Chernozhukov & Hansen (2005) and Blundell et al. (2007).

Implications for the nonparametric instrumental regressions are also examined. Recent analyses of such models (see e.g. Imbens & Newey, 2008 and Florens et al., 2007) have relied on a control variate approach rather than on a completeness assumption. Conditions for identification are indeed easier to obtain, and the additivity structure of the model can be relaxed. On the other hand, a strict monotonicity assumption is required, which rules out usual models with limited endogenous regressors. The previous result enables to prove the identification of the structural function in a triangular system of simultaneous equations under, roughly, an additive decomposition and a large support condition on the instrumental equation, but without any strict monotonicity condition. This shows that actually, the completeness approach may be more fruitful than the control variate one in some circumstances. Since different versions of completeness provides different identification results, there also appears to be a trade-off in the identification

[2]Chernozhukov & Hansen (2005) also rely on a condition which is close to bounded completeness(see their assumption L_1^*) for the identification of quantile treatment effects with instrumental variables.

[3]Indeed, assumption 2.4 of Chen & Hu (2006) and assumption 2 of Hu & Schennach (2008) are equivalent, under technical conditions, to a completeness condition.

of such models between the regularity condition imposed on the error term of the instrumental equation and the hypothesis on the structural function.

The paper is organized as follows. The main results are given in section two. Section three examines the consequence of these results on the identification of nonparametric instrumental regression. Section four concludes, and the proofs are deferred to section five.

2.1.2 Main results

In the sequel, X and Z belong to $\mathbb{R}^p$ and $\mathbb{R}^q$ respectively, with $p \leq q$. X and Z may share elements in common, and we let W denote these common elements, $W \in \mathbb{R}^r$. For instance, in an instrumental nonparametric regression (see e.g. Newey & Powell, 2003), W corresponds to the exogenous components of X. The remaining elements of X and Z are called respectively X_0 and Z_0, so that $X = (X_0, W)$ and $Z = (Z_0, W)$. In this framework, we will say that X is complete (resp. bounded, P-complete) for Z if (2.1.1) holds for all h such that, for $\mathbb{P}^W-$ almost all w, $h(., w)$ is integrable with respect to $\mathbb{P}^{X_0}$ (resp. bounded, bounded by a polynomial). In the sequel, we suppose that there exists maps μ_1 and ν_1, from respectively $\mathbb{R}^{p-r}$ and $\mathbb{R}^q$ to $\mathbb{R}^{p-r}$, such that

$$X_0 = \mu_1(\nu_1(Z) + \varepsilon_1), \tag{2.1.2}$$

and we consider the following assumptions.

A1. $Z_0 \perp\!\!\!\perp \varepsilon_1 \,|\, W$.

A2. For $\mathbb{P}^W$-almost all w, the measure of $\nu_1(Z_0, w)$ is continuous with respect to the Lebesgue measure and its support is $\mathbb{R}^{p-r}$.

A3. For $\mathbb{P}^W$-almost all w, ε_1 admits a continuous density $f_{\varepsilon_1|W}(., w)$.

Assumption A1 is a conditional independence hypothesis. Because mean-independence can always be achieved by a proper normalization,[4] A1 actually strengthens this

[4]Indeed, if we let $\widetilde{\nu}_1(Z) = \nu_1(Z) + \mathbb{E}(\varepsilon_1|Z)$ and $\widetilde{\varepsilon}_1 = \varepsilon_1 - \mathbb{E}(\varepsilon_1|Z)$, then $\mathbb{E}(\widetilde{\varepsilon}_1|Z_0, W) = 0 = \mathbb{E}(\widetilde{\varepsilon}_1|W)$.

mean-independence into independence. Note that if μ_1 is known, this assumption is testable in the data in general.

A2 is a continuity and large support condition. It may hold as soon as Z has one continuous component. The large support condition is restrictive but widespread in the literature (see e.g. Manski, 1988, or Lewbel, 2000). Moreover, only $\nu_1(Z)$, not necessarily Z, should satisfy this condition. This means that $p - r$ regressors with large support may be sufficient. This assumption, however, may be too strong, and we consider below alternative assumptions (see proposition 2.1.3). Lastly, A3 restricts the analysis to the case of a continuous residual. The continuity condition on its density is satisfied by all usual densities with infinite support.[5]

Despite the apparently strong assumption of an additive decomposition into independent terms, the function μ_1 in (2.1.2) enables to encompass many nonlinear models, beyond the nonparametric additive models with independent errors (for which $\mu_1(x) = x$). Usual ordered choice models correspond to $\mu_1(x) = \sum_{k=1}^{K} k\mathbb{1}_{]\alpha_{k-1};\alpha_k]}(x)$ (where $\mathbb{1}_A(x) = 1$ if $x \in A$, 0 otherwise) for some given thresholds $\alpha_0 = -\infty < \alpha_1 < ... < \alpha_K = +\infty$.[6] Count data models can also be handled by taking $\mu_1(x) = [\exp(x)]$ (where $[a]$ denotes the integer part of a). Simple tobit models correspond to $\mu_1(x) = \max(0, x)$. These three examples underline the fact that X may not be strictly monotonous in ε_1. Lastly, duration models like the accelerated failure time model or the proportional hazard model also fit in this framework. The first corresponds to $\mu_1(x) = \exp(x)$, while in the second, μ_1 is an unknown increasing function and $-\varepsilon_1$ is distributed according to a Gompertz distribution.

To achieve completeness, further restrictions are required.

A4. $\mathbb{P}^W-$almost surely, the conditional characteristic function $\psi_{\varepsilon_1|W}(., w)$ of ε_1 is infinitely often differentiable in $\mathbb{R}^p \backslash A(w)$ for some finite set $A(w)$ and does not vanish on the real line.

[5] It fails for the uniform density but this case is ruled out anyway by assumption A4 below.

[6] Binary choice models are obviously included. Note however that for binary variables X_0, model (2.1.2) is unnecessary since completeness is simply equivalent to non independence between X_0 and Z_0, conditional on W. When X_0 takes more than two values, it can be shown that the completeness condition is equivalent to the positivity of a variance matrix (see Das, 2005, theorem 2.1). However, it is not obvious to check this condition for a given theoretical model.

A5. All the moments of $||\varepsilon_1||$ are finite and there exists B and j such that $||\mu_1(t)|| \leq B||t||^j$ (where $||.||$ is the euclidian norm).

A6. ε_1 is gaussian or satisfies, $\mathbb{P}^W$-almost surely on w and for all $x, y \in \mathbb{R}^{p-r}$, there exists $C(.)$ and $k(.)$ such that

$$f_{\varepsilon_1|W}(x+y,w) \leq C(w)(1+||x||^2)^{k(w)} f_{\varepsilon_1|W}(y,w).$$

Zero-freeness of the characteristic function is a usual assumption in deconvolution problems (see e.g. Devroye, 1989, Fan & Truong, 1993, Li & Vuong, 1998, Schennach, 2004 and 2007) and is satisfied, among others, by gaussian, Student, Laplace and $\alpha-$stable distributions. The only common continuous distributions that fail to satisfy it are the uniform and triangular ones. All standard characteristic functions also satisfy the differentiability condition.

Assumption A5 rules out thick tails on the density of ε_1 and restricts the range of nonlinear models between X_0 and Z. It fails for instance with the previous examples of count data and accelerated failure time models, but holds for all the others aforementioned cases. A similar polynomial growth condition is imposed by Schennach (2007) to identify a nonlinear errors-in-variables model with instruments (on this issue, see also Zinde-Walsh, 2007).

Lastly, assumption A6 is rather restrictive. It imposes in particular that $f_{\varepsilon_1|W}(.,w)$ is either gaussian or has heavy tails.[7] The condition holds for instance for Student and $\alpha-$ stable distributions (see Mattner, 1992).

Theorem 2.1.1 *Suppose that (2.1.2) and A1-A3 hold. Then*
1) if A4 holds, X is bounded complete for Z.
2) If A4 and A5 hold, X is P-complete for Z.
3) If A4 and A6 hold, X is complete for Z.

Theorem 2.1.1 gives conditions under which different versions of completeness hold. The intuition of its proof can be explained as follows. First, one can show that

[7]Put $x = -y$ to see that $1/f_{\varepsilon_1|W}$ must be at most of polynomial order. It can also be shown (see Mattner, 1992) that A6 is implied by the condition $0 < c(w) \leq f_{\varepsilon_1|W}(x,w)(1+||x||)^{\gamma(w)} \leq C(w) < \infty$ for all $x \in \mathbb{R}^{p-r}$ and some real $c(w), C(w)$ and $\gamma(w) > 0$.

completeness is equivalent to the unicity of the following convolution equation in $g(.,w)$ (for almost all w) :

$$\int g(t,w) f_{-\varepsilon_1|W}(u-t,w)dt = 0. \tag{2.1.3}$$

If $g(.,w)$ was integrable, this would imply, by the convolution theorem,

$$\mathcal{F}(g(.,w)) \times \mathcal{F}(f_{-\varepsilon_1|W}(.,w)) = 0. \tag{2.1.4}$$

where $\mathcal{F}$ denotes the Fourier transform. Then, by assumption A4, $\mathcal{F}(g(.,w)) = 0$, and since the Fourier transform is injective, $g(.,w) = 0$. Actually, the problem is more involved because a priori, $g(.,w)$ is not integrable, so that its usual Fourier transform may not exist. To circumvent this issue, I rely on the techniques developed by Ghosh & Singh (1966) and Mattner (1992) to show completeness of location families.

Theorem 2.1.1 shows that in model (2.1.2), bounded completeness holds under rather weak conditions. Many of the usual densities also satisfy the moment condition which ensures P-completeness.[8] Completeness, on the other hand, is obtained under the restrictive hypothesis A6. As theorem 2.1.1 only provides sufficient conditions, one may wonder whether completeness actually holds under milder conditions. If it seems difficult to provide a full characterization of completeness, the following proposition shows that it really imposes stringent condition on the distribution of ε_1.

Proposition 2.1.2 *Suppose that (2.1.2), A1-A3 hold and $\mu_1(t) = t$. Assume also that, for $\mathbb{P}^W-$almost all w, ε_1 is not normal conditional on $W = w$ and there exists $\delta_1, \delta_2 > 0$ such that $\mathbb{E}\left(\exp\left(\delta_1||\varepsilon_1||^{1+\delta_2}\right)|W = w\right) < +\infty$. Then X is not complete for Z.*

Hence, if $f_{\varepsilon_1|W}$ has light tails, X cannot be complete for Z. On the other hand, X can still be bounded or P-complete for Z in such situations.

As mentioned above, the large support assumption A2 is rather strong. It is possible, though, to relax it, at the cost of imposing regularity on the distribution of ε_1. For the sake of simplicity, we restrict here to the case where X_0 is real ($p - r = 1$).

[8] Moreover, a density which does not fulfill condition A5 has heavy tails and thus is likely to satisfy A6.

A2'. For $\mathbb{P}^W$-almost all w, the measure of $\nu_1(Z_0, w)$ is continuous with respect to the Lebesgue measure.

A7. There exists $(a_k(w))_{k\in\mathbb{N}}$ such that, for all $x \in \mathbb{R}$ and for $\mathbb{P}^W$-almost all w,

$$f_{\varepsilon_1|W}(x, w) = \sum_{k=0}^{\infty} a_k(w)x^k. \tag{2.1.5}$$

Moreover, there exists $r_0(w) > 0$ such that $f_{\varepsilon_1|W}(., w)$, as a function on $\mathbb{C}$ defined by (2.1.5), is bounded on $\{z/|\mathrm{Im}(z)| < r_0(w)\}$.

Assumption A2' will generally hold if Z_0 contains a continuous regressor. The first part of assumption A7 states that $f_{\varepsilon_1|W}(., w)$ is entire. Examples of entire functions include the polynomials, the exponential function and all compositions of these functions (including gaussian densities). On the other hand, all densities with support different from $\mathbb{R}$ are not entire. Other counterexamples include the Cauchy and Student distributions. The second part of A7 is a technical condition which is satisfied for instance by gaussian densities.

Proposition 2.1.3 *Suppose that (2.1.2), A1, A2', A3, A4 and A7 hold. Then X is bounded complete for Z.*

Proposition 2.1.3 shows that the large support condition can be dropped, but at the price of restricting the range of the densities of ε_1.

The easiest way to interpret (2.1.2) is that Z causes X. However, it may be convenient sometimes to suppose instead that X causes Z. In the measurement error models of Chen & Hu (2006) and Hu & Schennach (2008) for instance, their condition on the injectivity of operators can be restated into completeness of the unobserved variable X_0 for the measure Z_0. In this case, the model (2.1.2) is unnatural since one would prefer to write the measure as a function of the unobserved variable and an independent error, i.e. a model of the form

$$\mu_2(Z) = \nu_2(X) + \varepsilon_2, \tag{2.1.6}$$

where μ_2 and ν_2 are maps from $\mathbb{R}^q$ (resp. $\mathbb{R}^p$) to $\mathbb{R}^{q-r}$. The standard measurement error model, for instance, corresponds to $\mu_2(z_0, w) = z_0$ and $\nu_2(x_0, w) = x_0$. When

$\mu_2(.,w)$ is one-to-one, the model writes $Z_0 = \mu_2^{-1}(\nu_2(X) + \varepsilon_2)$ and is similar to (2.1.2). However, in general we cannot simply switch X_0 and Z_0 in (2.1.2) to obtain completeness, as the simple example $Z_0 = 1$ shows. In such a model, indeed, Z_0 would not necessarily be informative enough on X_0 for completeness to hold.

We also assume the following hypotheses, which are close to A1, A2 and A4.

A8. $X_0 \perp\!\!\!\perp \varepsilon_2 \,|\, W$.

A9. For $\mathbb{P}^W$-almost all w, $\nu_2(.,w)$ is a one-to-one mapping on $\mathbb{R}^{q-r}$.

A10. $\mathbb{P}^W-$almost surely, the characteristic function $\psi_{\varepsilon_2|W}$ of ε_2 conditional on W has isolated zeros.

Assumption A8 is exactly equivalent to A1. Assumption A9 is similar but stronger than the large support condition A2. Indeed, $\nu_2(.,w)$ is imposed to be one-to-one, so that here $q = p$. A10, on the other hand, is weaker than A4 and holds for all usual distributions, including the uniform and triangular ones. Actually, it holds for all distribution with exponential tails, because then the corresponding characteristic function is holomorphic on a strip of the complex plane and thus has isolated zeros (see Rudin, 1987, p. 208). The Fejer - de la Vallee Poussin density $x \mapsto (\pi x^2)^{-1}(1-\cos(x))$ is a counterexample of a distribution which violates A10, as its characteristic function is equal to $t \mapsto \max(1-|t|, 0)$.

Proposition 2.1.4 *Suppose that (2.1.6) and A8-A10 hold. Then X is complete for Z.*

Thus, even if the completeness condition is asymmetric in X and Z, to a certain extent the roles of X and Z in model 2.1.2 can be exchanged. The conditions on ν_2 are stronger than the one required for theorem 2.1.1 to hold, but completeness and not only bounded or P-completeness is achieved under weak restrictions on the distribution of ε_2.

2.1.3 Implications for the nonparametric instrumental regression

In this section, we apply theorem 2.1.1 to the identification of nonparametric instrumental regressions. Let us consider the following triangular system :

$$\begin{cases} Y &= \varphi(X) + \eta \\ X_0 &= \mu_1(\nu_1(Z) + \varepsilon_1) \end{cases} \quad \mathbb{E}(\eta|Z) = 0 \tag{2.1.7}$$

In this model, X_0 are the endogenous regressors, W are exogenous covariates and Z_0 denote the instruments. The aim is to recover the structural function φ. This system is close to the one studied by Newey et al. (1999), although we allow for nonlinearity in the instrumental equation. A main restriction is the additive separability assumption of the first equation. This is the price to pay for a rather weak exogenous condition $\mathbb{E}(\eta|Z) = 0$. In particular, heteroscedasticity is permitted in this framework.

Note that it is possible, through a control variate approach, to relax additive separability under full independence between Z and (ε, η). Recent contributions include Chesher (2003), Imbens & Newey (2008) and Florens et al. (2008) (see Chesher, 2007, for a survey). However, strict monotonicity in the error term of the instrumental equation is required to identify this error. Hence, this approach generally rules out limited endogenous regressor and cannot be applied to model (2.1.7) unless μ_1 is one-to-one.[9] The completeness approach, one the other hand, can be applied with virtually no assumption on this function.

Proposition 2.1.5 *Suppose that (2.1.7) and A1-A3 hold. Then φ is identified if one of the following conditions is satisfied :*
1) A4 holds and $\varphi(., w)$ is bounded for $\mathbb{P}^W$-almost all w ;
2) A4-A5 hold and $\varphi(., w)$ is bounded by a polynomial for $\mathbb{P}^W$-almost all w ;
3) A4 and A6 hold.

Proposition 2.1.5 shows that to recover φ, there is a trade-off between the regularity conditions imposed on model (2.1.7) and the assumptions on the function φ itself.

[9] One could redefine the instrumental equation as $X_0 = \widetilde{\mu}(Z, \widetilde{\varepsilon}_1)$ with $\widetilde{\mu}_1$ strictly increasing in $\widetilde{\varepsilon}_1$ but then the independence condition $Z \perp\!\!\!\perp (\widetilde{\varepsilon}_1, \eta)$ would not hold anymore in general (see Florens et al., 2007, for a discussion on this point).

The first condition of the proposition is useful when X_0 has a finite support, but imposes a strong restriction on $\varphi(.,w)$ otherwise. Linear forms, for instance, cannot be handled by this case. The second widens considerably the range of identified models, at the price of the moment condition on ε_1 and the polynomial growth restriction on μ_1. Lastly, if one is reluctant to make any assumption on φ, identification is achieved under strong restrictions on ε_1.

2.1.4 Conclusion

This paper provides general sufficient conditions to achieve varieties of completeness conditions, and apply these results to the nonparametric instrumental regression. Two questions on this topic are left for future research. Firstly, one can wonder whether the assumption of additive decomposition into independent parts could be weaken. Secondly, the adaptation of the results above to the identification condition of Chernozhukov & Hansen (2005) (see their assumption L_1^*) in the context of nonseparable models remains a challenging issue.

2.1.5 Proofs

Theorem 2.1.1

For all h, let $\widetilde{h}(t,w) = h(\mu_1(t),w)$. By A1,

$$\begin{aligned}\mathbb{E}[h(X)|Z] &= \mathbb{E}[\widetilde{h}(\nu_1(Z)+\varepsilon_1,W)|Z] \\ &= \int \widetilde{h}(\nu_1(Z)+u,W) f_{\varepsilon_1|W}(u,W)du \quad \text{a.s.} \\ &= \int \widetilde{h}(t,W) f_{-\varepsilon_1|W}(\nu_1(Z)-t,W)dt \qquad \text{a.s.}\end{aligned}$$

By A2 (and conditional on W), $\nu_1(Z)$ admits a continuous distribution whose support is $\mathbb{R}^{p-r}$. Thus,

$$\mathbb{E}[h(X)|Z] = 0 \text{ a.s.} \Leftrightarrow \int \widetilde{h}(t,w) f_{-\varepsilon_1|W}(u-t,w))dt = 0 \ \lambda \otimes \mathbb{P}^W - \text{a. e. in } (u,w) \tag{2.1.8}$$

where λ denotes the Lebesgue measure. Because $\nu_1(Z)+\varepsilon_1$ also admits a continuous distribution and its support is $\mathbb{R}^{p-r}$, it follows that

$$h(X) = 0 \text{ a.s.} \Leftrightarrow \widetilde{h}(t,w) = 0 \ \lambda \otimes \mathbb{P}^W - \text{a. e. in } (u,w). \tag{2.1.9}$$

Moreover, $h(X)$ integrable (resp. h bounded) implies that $\widetilde{h}(\nu_1(Z)+\varepsilon_1)$ is integrable (resp. $\widetilde{h}$ is bounded). Similarly, by A5, if $h(.,w)$ is bounded by a polynomial, $\widetilde{h}(.,w)$ is also bounded by a polynomial. Hence, to prove completeness (resp. bounded completeness, P-completeness), it suffices to prove that for all g such that $g(\nu_1(Z)+\varepsilon_1)$ is integrable (resp. g is bounded, bounded by a polynomial), $\mathbb{P}^W-$almost surely in w,

$$\int g(t,w)f_{-\varepsilon_1|W}(u-t,w)dt = 0 \quad \text{a.e. in } u \Rightarrow g(t,w) = 0 \quad \text{a.e. in } t \qquad (2.1.10)$$

This statement corresponds to the completeness of the location family with density $f_{-\varepsilon_1|W}$, except that the left part of (2.1.10) holds almost everywhere and not everywhere. But in theorem 1.3 of Mattner (1992) (and hence in his theorem 1.1), the statement also holds almost everywhere, so that we can apply it to obtain part 3 of the theorem.

To show part 1, we adapt the proof of theorem 2.4 of Ghosh & Singh (1966). Let L^1 (resp. L^∞) denote the space of equivalent classes of integrable (resp. essentially bounded) functions with respect to the Lebesgue measure. Let w be such that $g(.,w) \in L^\infty$, $\psi_{\varepsilon|W}(.,w)$ does not vanish anywhere and the left part of (2.1.10) holds (the set of such w being of probability one). Let $f_{w,u}(x) = f_{-\varepsilon_1|W}(u-x,w)$, $\mathcal{P}_w = \text{span}\left\{f_{w,u},\, u \in \mathbb{R}^{p-r} \;/\; \int g(t,w)f_{w,u}(t)dt = 0\right\}$ and $\mathcal{Q}_w = \{f_{w,u} \;/\; u \in \mathbb{R}^{p-r}\}$. Let $\mathcal{R}_w = \{u \;/\; f_{w,u} \in \mathcal{P}_w\}$. Because the Lebesgue measure of ${}^c\mathcal{R}_w$ is zero, there exists a sequence $(u_n)_{n\in\mathbb{N}}$ of elements of $\mathcal{R}_w$ such that $u_n \to u$ for all $u \in {}^c\mathcal{R}_w$. By continuity of $f_{-\varepsilon_1|W}(.,w)$ and Scheffé's theorem (see e.g. van der Vaart, 1998, p. 22), $\int |f_{w,u_n}(t) - f_{w,u}(t)|dt \to 0$. Thus $\mathcal{Q}_w$ is included in the closure of $\mathcal{P}_w$ (for the L^1-norm).

Now, by A4 and Wiener's tauberian theorem (see e.g. Rudin, 1991, p. 229), $\mathcal{Q}_w$ is dense in L^1. Thus, $\mathcal{P}_w$ is dense in L^1. By continuity of the linear form $\phi \mapsto \int g(t,w)\phi(t)dt$ and the Riesz theorem (see e.g. Rudin, 1987, p. 130), $g(t,w) = 0$ for almost every t and almost all w.

Lastly, let us turn to part 2. First, because it is integrable, $f_{-\varepsilon_1|W}(.,w) \in \mathcal{S}'$, the space of tempered distribution (see Rudin, 1991, p. 191, example d). Moreover, $g(.,w)$ is bounded by a polynomial, so that $g(.,w) \in \mathcal{S}'$ (see Rudin, 1991, p. 191, example d). Lastly, the function

$$c(.,w) : u \mapsto \int g(t,w)f_{-\varepsilon_1|W}(u-t,w)dt$$

equals zero almost everywhere. Hence, it is the zero distribution and, as such, is tempered.

Now let $g_n(.,w) = g(.,w) \times \mathbb{1}_{[-n,n]}(.)$. $g_n(.,w)$ is a tempered distribution with compact support, so that it belongs to the space of quickly decreasing distributions (see Schwartz, 1973, p. 244). Let us show that $g_n(.,w)$ converges to $g(.,w)$ in $\mathcal{S}'$. We have to prove that

$$\int g_n(u,w)\phi(u)du \to \int g(u,w)\phi(u)du$$

for all $\phi \in \mathcal{S}$, the space of rapidly decreasing functions (see e.g. Rudin, 1991, p. 161). Let Φ be any bounded set in $\mathcal{S}$, the space of rapidly decreasing functions. There exists (see Schwartz, 1973, p. 235) a continuous function b with $b(x) = o(|x|^{-m})$ as $|x| \to \infty$ and for every m, such that $|\phi(x)| \leq b(x)$ for every $x \in \mathbb{R}$ and every $\phi \in \Phi$. Because $g(.,w)$ is bounded by a polynomial, $g(.,w) \times b$ is integrable. Thus, by dominated convergence,

$$\sup_{\phi\in\Phi}\left|\int \phi(u)(g_n(u,w) - g(u,w))du\right| \leq \int b(u)\mathbb{1}_{c[-n,n]}(u)|g(u)|du \longrightarrow 0.$$

Hence, $g_n(.,w) \to g(.,w)$ in $\mathcal{S}'$.

Let us show similarly that $c_n(.,w) = \int g_n(t,w)f_{-\varepsilon_1|W}(.-t,w)dt$ converges to $c(.,w)$ in $\mathcal{S}'$. Let $D(w)$ and $l(w)$ be such that $|g(t,w)| \leq D(w)(1 + \|t\|^{l(w)})$. We get

$$\begin{aligned}\int |g(t,w)|f_{-\varepsilon_1|W}(u-t,w)dt &= \int |g(u-t,w)|f_{-\varepsilon_1|W}(t,w)dt \\ &\leq D(w)\left(1 + \int \|u-t\|^{l(w)} f_{-\varepsilon_1|W}(t,w)dt\right) \\ &\leq D(w)\left[1 + 2^{l(w)-1}\left(\|u\|^{l(w)} + \int \|t\|^{l(w)} f_{-\varepsilon_1|W}(t,w)dt\right)\right],\end{aligned}$$

where the second inequality follows by convexity. Moreover, by assumption A5,

$$\int \|t\|^{l(w)} f_{-\varepsilon_1|W}(t,w)dt < +\infty.$$

This, together with the previous inequality, implies that $(t,u) \mapsto b(u)g(t,w)f_{-\varepsilon_1|W}(u-t,w)$ is integrable. As a consequence, $u \mapsto b(u)(c_n(u,w)-c(u,w))$ is also integrable. Moreover, by dominated convergence,

$$\int b(u)[c_n(u,w)-c(u,w)]du = \int\int b(u)g(t,w)f_{-\varepsilon_1|W}(u-t,w)\mathbb{1}_{c[-n,n]}(t)dtdu \longrightarrow 0.$$

As previously, this shows that $c_n(.,w) \to c(.,w)$ in $\mathcal{S}'$.

The previous results ensure that we can apply lemma 2.1 of Mattner (1992) to $f_{-\varepsilon_1|W}(.,w)$, $g(.,w)$, $c(.,w)$ and $g_n(.,w)$. As a consequence, we get, for almost all w and everywhere except on $A(w)$,

$$\mathcal{F}(g(.,w)) \times \mathcal{F}(f_{-\varepsilon_1|W}(.,w)) = 0.$$

Thus, by A4, $\mathcal{F}(g(.,w)) = 0$ everywhere except on $A(w)$. Applying the same reasoning as at the end of the proof of theorem 1.3 of Mattner (1992) finally yields $g(t,w) = 0$ almost everywhere in t. Part 2 follows and the proof is complete.

Proposition 2.1.2

We keep the notations of the previous proof. Because $\mu_1(t) = t$, $\widetilde{h} = h$. Hence, by (2.1.8) and (2.1.9), completeness is equivalent to

$$\int h(t,w) f_{-\varepsilon_1|W}(u-t,w)dt = 0 \quad \text{a.e. in } u \Rightarrow h(t,w) = 0 \quad \text{a.e. in } t. \qquad (2.1.11)$$

for $\mathbb{P}^W$-almost all w and all h such that $\mathbb{E}\left[|h(X_0)|\right] < \infty$. But theorem 2.4 of Mattner (1993) implies that this condition is not satisfied.[10] Hence, X is not complete for Z.

Proposition 2.1.3

We still keep the previous notations. Following the same lines as in the proof of theorem 2.1.1, we can show that bounded completeness holds if, for $\mathbb{P}^W-$almost all w and all bounded $g(.,w)$,

$$\int g(t,w) f_{-\varepsilon_1|W}(u-t,w)dt = 0 \text{ for a.e. } u \in \text{Supp}(\nu_1(Z)|W=w) \Rightarrow g(t,w) = 0 \text{ a.e.} \qquad (2.1.12)$$

Suppose that the left hand side holds. Then, by Assumption A2', $c(.,w)$ equals zero on an open set O_w.

[10] Actually, Mattner (1993) shows that $\int h(t,w) f_{-\varepsilon_1|W}(u-t,w)dt = 0$ for every u (and not almost every u) implies $h(t,w) = 0$ for almost every t. However, an inspection of the proofs of his theorem 2.4 and lemma 2.3 shows that "every u" can be replaced by "almost every u" without affecting the result.

Besides, by assumption A7 and the fact that $g(.,w)$ is bounded for $\mathbb{P}^W$-almost all w, the function

$$(t,u) \mapsto g(t,w) f_{-\varepsilon_1|W}(u-t,w)$$

is bounded on $\{(t,u) \in \mathbb{R} \times \mathbb{C} / |\mathrm{Im}(u)| < r_0(w)\}$. Moreover, $u \mapsto g(t,w) f_{-\varepsilon_1|W}(u-t,w)$ is holomorphic on $\{u \in \mathbb{C} / |\mathrm{Im}(u)| < r_0(w)\}$ by assumption A7. Thus (see Rudin, 1987, p. 229), $c(.,w)$ is holomorphic on this same set. An holomorphic function which vanishes on an open set actually equals zero everywhere (see e.g. Rudin, 1987, p.209). Thus, $c(.,w) = 0$ everywhere and the end of the proof of theorem 2.1.1, part 1, can be applied.

Proposition 2.1.4

Let h be such that $E[|h(X)|] < \infty$ and $\mathbb{E}[h(X)|Z] = 0$ almost surely. Let also $\nu_2^{-1}(.,w)$ denote the inverse of $\nu_2(.,w)$ and $\widetilde{h}(t,w) = h(\nu_2^{-1}(t,w),w)$. Letting $T = \nu_2(X)$, we have

$$\mathbb{E}[\widetilde{h}(T,W)|\mu_2(Z),W] = 0.$$

Hence, for all $t_1 \in \mathbb{R}^{q-r}$, almost surely,

$$\mathbb{E}[\widetilde{h}(T,W) e^{it_1'(T+\varepsilon_2)}|W] = 0.$$

Then, by assumption A8, almost surely,

$$\mathbb{E}[\widetilde{h}(T,W) e^{it_1'T}|W]\mathbb{E}[e^{it_1'\varepsilon_2}|W] = 0.$$

This implies, by assumption A10, that the function $t_1 \mapsto \mathbb{E}[\widetilde{h}(T,W) e^{it_1'T}|W]$ vanishes everywhere except perhaps on isolated points. Now, because $\mathbb{E}[|\widetilde{h}(T,W)| \,|\, W] < +\infty$, the function $t_1 \mapsto \mathbb{E}[\widetilde{h}(T,W) e^{it_1'T}|W]$ is continuous by dominated convergence. Thus, for all $(t_1,t_2) \in \mathbb{R}^{q-r} \times \mathbb{R}^r$,

$$\mathbb{E}[\widetilde{h}(T,W) e^{i(t_1'T+t_2'W)}] = 0.$$

This implies (see e. g. Bierens, 1982, theorem 1) that $\mathbb{E}[\widetilde{h}(T,W)|T,W] = 0$ almost surely. In other words, $h(X) = 0$ almost surely.

Proposition 2.1.5

$\mathbb{E}(Y|Z) = \mathbb{E}(\varphi(X)|Z)$, so that any candidate φ' for φ satisfies

$$\mathbb{E}[(\varphi' - \varphi)(X)|Z] = 0.$$

If $\varphi(.,W)$ is known to be bounded, any candidate must be also bounded so that $(\varphi' - \varphi')(.,W)$ is bounded. Then by theorem 2.1.1 (part 1), $\varphi' = \varphi$ so that $\varphi(.,W)$ is identified. If $\varphi(.,W)$ is bounded by a polynomial, so is $(\varphi' - \varphi')(.,W)$, and the same conclusion holds by part 2 of the theorem. Lastly, if A6 holds, $\varphi' = \varphi$ by part 3 of the theorem.

2.2 A new method for dealing with endogenous selection

2.2.1 Introduction

Missing observations are very common in micro data, either because of selection, nonresponse or simply because counterfactual variables cannot be observed. Ignoring this issue by making inference on the observed population generally leads to inconsistent estimators. Moreover, without additional assumptions, only bounds on the parameters of interest can be identified (see e.g. Manski, 2003). Several approaches have been followed to retrieve point identification. The first is to suppose independence between response and variables of interest conditional on observed covariates. This is the so-called missing at random hypothesis (see e.g. Little & Rubin, 1987), or the unconfoundedness assumption in the treatment effect literature (see for instance Imbens, 2004). However, this assumption is often considered too stringent because it rules out any correlation between the selection and outcome variables. When such endogenous selection arises, the common practice is to use instruments which determine selection but not outcomes (see e.g. Heckman, 1974, on tobit models, Angrist et al., 1996, or Heckman & Vytlacil, 2005 on treatment effects). However, this assumption does not point identify the distribution of the outcome in general (see Manski, 2003). Moreover, it may be difficult to find such instruments. When selection depends heavily on the dependent variable, in particular, the assumption of conditional independence is difficult to maintain. A third approach relies on functional restrictions rather than exclusion restrictions. For instance, Chamberlain (1986) obtains identification at the infinity by imposing a linear structure. Lastly, using an appealing composite strategy, Lewbel (2007) obtains identification under the existence of a special regressor which is strongly exogenous (i.e., conditionally independent of the errors of the selection model), a large support condition and restrictions on the probability of selection.[11]

In this paper, another instrumental strategy for solving endogenous selection is considered. Nonparametric identification is based on independence between the instruments and the selection variable, conditional on the outcome and possibly on other explanatory variables. This assumption has been also used in the framework of nonignorable nonresponse by Chen (2001), Tang et al. (2003), Hemvanich (2004)

[11]This probability must tend to zero or one when the special regressor tends to infinity.

and Ramalho & Smith (2007).[12] Apart from nonresponse, this assumption may be particularly suitable when selection is directly driven by the dependent variable. Consider for instance a variable which is observed only if it exceeds an unobserved truncation. Finding an instrument which only affects selection is impossible if this truncation variable is purely random. Instead, any variable which affects the dependent variable will satisfy the exclusion restriction considered here. Other examples where this assumption can be useful include Roy models with unobserved sector, one stratum response based samples or truncated count data models. As in usual instrumental regressions, a rank condition between instruments and outcomes is also required to achieve identification. This condition is stated in terms of completeness, and was already considered in several nonparametric instrumental problems (see, among others, Newey & Powell, 2003, Chen & Hu, 2006 and Hu & Schennach, 2008). Under this hypothesis and the conditional independence assumption, the joint distribution of the data is identified nonparametrically.[13] The key point is that in this framework, it is enough to recover the probability of selection conditional on the outcome. This is similar to the unconfoundedness situation, in which the problem reduces to identifying the propensity score. However, whereas the identification of the propensity score is trivial in the latter case, the conditional probability is harder to retrieve in the former. I show that this function satisfies an integral inverse problem, whose solution is unique under the completeness condition.

If only some moments of the instrument are used, and not its full distribution, the joint distribution of the data can still be recovered under a parametric restriction on the selection model. This result may be useful when only aggregated information on the instruments is available, or for the ease of estimation. The idea of using moments of instruments to deal with nonresponse has also been applied in survey sampling (see Deville, 2002). It is also related to the literature on auxiliary information, which has been developed either for efficiency reasons (see Imbens

[12]The difference with these papers is that they focus mainly on parametric and semiparametric estimation issues, whereas the emphasis is put on nonparametric identification here. Chen (2001) and Tang et al. (2003) propose sufficient conditions for identification in parametric models, and Hemvanich (2004) studies identification when the support of the outcome is finite. We extend his result to a general situation here.

[13]In particular, the marginal effect of the instrument on the outcome, or the effet of the selection variable on the outcome, are identified.

& Lancaster, 1994, Hellerstein & Imbens, 1999) or, as here, to provide identification (see Hellerstein & Imbens, 1999, and Nevo, 2002). Our parametric framework extends Nevo's result to the case of endogenous selection.

The fact that the identification strategy relies on an exclusion restriction may seem restrictive in some applications, and is not needed in Lewbel's framework for instance.[14] However, and contrary to the missing at random assumption for instance, this condition is testable. Furthermore, the method appears to be fruitful even if the exclusion restriction fails. The intuition behind is that this condition is the extreme opposite of unconfoundedness. Indeed, selection only depends on the outcome in the first case, and only on covariates in the second. In between, if selection depends monotonically on both the outcome and a given instrument, the identifying equations underlying the two assumptions provide sharp and finite bounds on parameters of the outcome. Thus, even if the dependent variable is unbounded, one can obtain compact interval on parameters of interest. This result is similar to the one of Manski & Pepper (2000) (see their proposition 2, corollary 2) but within a slightly different framework and under other assumptions. Instead of their monotone treatment response condition, which states that outcomes increase with the treatment, the result relies on the existence of an instrument which affects selection in a monotonic way. Such a condition is weak and is likely to be satisfied in many contexts, including the use of data with nonignorable nonresponse and treatment effects estimation. In this latter case in particular, the result should be of practical importance as it enables to go beyond the standard routine of computing matching estimators as point estimates of these effects.

Apart from identification issues, estimation of the model is also considered. Standard GMM can be used in the parametric case or in the nonparametric one with a discrete outcome. In a nonparametric setting with a continuous dependent variable, the parameter is functional and must be estimated through an infinite number of moment conditions. Estimation is based on a Tikhonov regularization method, as in Hall & Horowitz (2005) or Carrasco et al. (2006). The estimator of the condi-

[14] On the other hand, the existence of a special regressor, which may be difficult to find in practice, is not needed here. Indeed, the instrument may be continuous or discrete, the completeness condition only implying that its support has the same number of or more elements than the one of the outcome. Moreover, no restriction is imposed on the conditional probability of selection, except, as usual, that it should be positive everywhere.

tional probability of selection is shown to be consistent. This estimator enables in turn to make valid inference on the whole population, by an inverse probability weighting procedure, in a similar fashion to Horvitz & Thompson (1952), Hellerstein & Imbens (1999), Nevo (2002) or Wooldridge (2005). Finite sample properties of these estimators are investigated through Monte Carlo simulations.

Lastly, the method is used to estimate the effect of grade retention in fifth grade in France on test achievement. Besides the usual counterfactual problem, identification of this effect is complicated by the fact that French students only take standardized tests at the beginning of the third and sixth grades. Thus, the ability at the end of the fifth grade, which is one of the main factor of grade retention, is observed for promoted students, thanks to the sixth test, but not for retained students. Consequently, the problem fits within our framework. Using the third grade test score as an instrument, sharp bounds on the effects of grade retention are computed. Overall, the short term impact of grade retention seems more likely to be positive. This result is in line with the one of Jacob & Lefgren (2004) for third graders in Chicago.

The rest of the paper is structured as follows. Section two is devoted to identification issues. Estimation methods are described in section three. Monte Carlo results are displayed in section four, and the application to grade retention is presented in section five. The appendix contains all proofs.

2.2.2 Identification

The setting and main result

Let D, Y and Z denote respectively the selection dummy variable, the dependent variable, and the instruments. The first assumptions set the selection problem.

Assumption 1 *We observe D and (Y, Z) when $D = 1$. Y is not observed when $D = 0$.*

Assumption 2 *The distribution of Z is identified.*

Assumptions 1 and 2 are satisfied when Y alone is missing, as in selection problems or item nonresponse. It also covers unit nonresponse where (Y, Z) are missing when

$D = 0$. In this latter situation, auxiliary information on Z is needed to satisfy assumption 2. This information typically stems from a refreshment sample, censuses or administrative data. In these two latter cases, supposing the identifiability of the whole distribution of Z may be overly strong, and we will see in subsection 2.4 that it can be weakened to the knowledge of moments of Z, at the price of imposing parametric restrictions.

Assumptions 1 and 2 alone do not enable to point identify the distribution of (D, Y, Z). More structure on the dependence between these variables is needed. If selection directly depends on Y, the usual assumption of exogenous selection will fail, and it may be difficult to find an instrument which affects selection but not the outcome. On the other hand, we may find variables which are related to Y but not to D. More precisely, we assume here the following condition :[15]

Assumption 3 $D \perp\!\!\!\perp Z \,|\, Y$.

This assumption has also been made by Chen (2001), Tang et al. (2003), Hemvanich (2004) and Ramalho & Smith (2007) in the framework of nonresponse. It is also a particular case of assumption (41) of Manski (1994). The condition can be interpreted as follows. The selection equation depends on Y, which is missing when $D = 0$, and thus cannot be identified with the data alone. On the other hand, if an instrument which affects Y but not directly D is available, one can identify this selection equation, in a similar fashion to usual instrumental regressions. For instance, suppose that (D, Y, Z) follow the nonparametric system

$$\begin{cases} Y &= \varphi(Z, \varepsilon) \\ D &= \psi(Y, \eta). \end{cases} \tag{2.2.1}$$

In this setting, we have the following result.

Proposition 2.2.1 *Suppose that system (2.2.1) holds with $\eta \perp\!\!\!\perp (Z, \varepsilon)$. Then assumption 3 holds.*

By letting $\psi(y, u) = \mathbb{1}\{u \leq P(D = 1|Y = y)\}$, we can suppose without loss of generality that η is independent of Y.[16] The exclusion restriction amounts to reinforcing this into a conditional independence between η and (Y, Z).

[15] We could refine this assumption by supposing that $D \perp\!\!\!\perp Z \,|\, Y, X$ where X denote covariates whose distribution is identified. All the subsequent analysis would then hold conditional on X. We do not introduce such covariates until subsection 2.4 for the ease of notations.

[16] In this case, ψ is not necessarily structural.

As indicated previously, a dependence condition between Y and Z is required to achieve identification of the model. I rely afterwards on a completeness condition. For any random variable T and $q > 0$, let L_T^q denote the space of functions g satisfying $E(|g(T)|^q) < +\infty$. Let us also denote $\mathcal{B}$ the set of real functions g such that $g(Y)$ is bounded below almost surely and $g \in L_Y^1$.

Assumption 4 *Y is $\mathcal{B}$-complete for Z, that is for all $g \in \mathcal{B}$,*

$$\Big(E(g(Y)|Z) = 0 \quad a.s.\Big) \Longrightarrow \Big(g(Y) = 0 \quad a.s.\Big). \tag{2.2.2}$$

Assumption 4 is weaker than the usual completeness condition, for which condition (2.2.2) must hold for any $g \in L_Y^1$, but stronger than bounded completeness, for which condition (2.2.2) must hold for bounded functions only (see e.g. Mattner, 1993, for a discussion on the difference between completeness and bounded completeness). The standard completeness condition has been used in the study of nonparametric instrumental regression under additive separability (see Newey & Powell, 2003, Darolles et al., 2007) and in nonclassical measurement error problems (see Chen & Hu, 2006 and Hu & Schennach, 2008),[17] while the bounded completeness condition has been used for instance by Chen & Hu (2006).

Completeness can be easily characterized when Y and Z have finite supports. Indeed, letting $(y_1, ..., y_s)$ and $(z_1, ..., z_t)$ denote these supports, this assumption amounts to

$$P(\text{rank}(M) = s) = 1, \tag{2.2.3}$$

where M is the random matrix of typical element $P(Y = y_i|Z = z_j)$ (see Newey & Powell, 2003). Hence, the support of Z must be at least as rich as that of Y ($t \geq s$) and the dependence between the two variables must be strong enough for s distinct conditional distributions $P(Y = .|Z = z_j)$ to exist. In this case, completeness is equivalent to bounded completeness. Completeness or bounded completeness are much more difficult to characterize when the support of Y or Z is infinite, and only sufficient conditions have been obtained yet. Both hold when the density of Y conditional on Z belongs to an exponential family (see Newey & Powell, 2003). As proposition 2.2.2 shows, assumption 4 is also satisfied under an

[17] Indeed, assumption 2.4 of Chen & Hu (2006) and assumption 2 of Hu & Schennach (2008) are equivalent, under technical conditions, to a completeness condition.

additive decomposition, a large support assumption and technical restrictions on ε in system (2.2.1).

Proposition 2.2.2 *Consider system (2.2.1) with $Y \in \mathbb{R}$ and suppose that*

1. *(additive decomposition) $\varphi(Z, \varepsilon) = \mu(\nu(Z) + \varepsilon)$ and $Z \perp\!\!\!\perp \varepsilon$.*
2. *(large support) The measure of $\nu(Z)$ is continuous with respect to the Lebesgue measure and the support of $\nu(Z)$ is $\mathbb{R}$ almost surely.*
3. *(regularity conditions on ε) The distribution of ε admits a continuous density f_ε with respect to the Lebesgue measure. Moreover, $f_\varepsilon(0) > 0$ and there exists $\alpha > 2$ such that $t \mapsto t^\alpha f_\varepsilon(t)$ is bounded. Lastly, the characteristic function of ε does not vanish and is infinitely often differentiable in $\mathbb{R} \backslash A$ for some finite set A.*

Then Y is $\mathcal{B}-$complete for Z.

The additive decomposition and the large support condition are identical to the assumptions A1 and A2 made in the previous section to study completeness and bounded completeness.[18] The regularity conditions on ε are satisfied for many distributions such as the normal, the student with degrees of freedom greater than one[19] or the stable distributions with characteristic exponent greater than one. Interestingly, these regularity conditions are hardly stronger than the one needed to achieve bounded completeness, namely, the zero freeness of the characteristic function of ε (see theorem 2.1.1 in the previous section). Hence, in this framework at least, $\mathcal{B}$-completeness appears to be almost equivalent to bounded completeness.

Because identification is based on inverse probability weighted moment conditions, we also suppose that the conditional probability $P(Y) \equiv P(D = 1|Y)$ is positive almost surely. This assumption is similar to the common support condition in the

[18]The additive decomposition considered here encompasses many nonlinear models, beyond the nonparametric additive models for which $\mu(x) = x$. Usual ordered choice models correspond to $\mu(x) = \sum_{k=1}^{K} k \mathbb{1}_{]\alpha_{k-1};\alpha_k]}(x)$ (where $\mathbb{1}_A(x) = 1$ if $x \in A$, 0 otherwise) for some given thresholds $\alpha_0 = -\infty < \alpha_1 < ... < \alpha_K = +\infty$. Simple tobit models correspond to $\mu(x) = \max(0, x)$. Duration models like the accelerated failure time model (for which $\mu(x) = \exp(x)$) or the proportional hazard model (for which μ is an unknown increasing function and $-\varepsilon$ is distributed according to a Gompertz distribution) also fit in this framework.

[19]See e.g. Mattner (1992) for a proof that the conditions on the characteristic function of student distributions are indeed satisfied.

treatment effects literature. It does not hold if D is a deterministic function of Y, as in simple truncation models where $D = \mathbb{1}\{Y \geq y_0\}$, y_0 denoting a fixed threshold. It also fails for random truncation models of the form $D = \mathbb{1}\{Y \geq \eta\}$ if η is strictly greater than the infimum of Y. In example 2 below, for instance, this would be the case if the reservation wage η of individuals is always greater than the lowest potential wage Y.

Assumption 5 *$P(Y) > 0$ almost surely.*

Theorem 2.2.3 *Suppose that assumptions 1-5 hold. Then the distribution of (D, Y, Z) is identified.*

Basically, the result stems from the fact that under assumption 3 and 4, the equation in Q

$$E\left(\frac{D}{Q(Y)}\middle| Z\right) = 1 \tag{2.2.4}$$

admits a unique solution, P. Identification of P follows because the left hand side is identified for any given Q. Then it is easy to show that the knowledge of P enables to identify the distribution of (D, Y, Z). We now present several potential applications of this framework.

Example 1 : nonignorable nonresponse

In this case, an outcome Y is observed only if the individual answers the survey or a given question in the questionnaire ($D = 1$). The aim is to recover the full distribution of Y, given that nonresponse directly depends on Y. For instance, consider the variable $Y = 1$ if the individual has used drugs at least once during the month, 0 otherwise. Accepting to answer the question "have you used drugs at least once during the last month ?" is likely to depend on the answer Y itself. The method can be applied if an instrument affects Y but not directly D. In the drugs example, local drug prices affect the fact of using drugs but are unlikely to play directly on response on drug use. Note that in this example where Y is binary, the completeness condition is easy to check, since it is equivalent to a nonzero correlation between Y and the instrument.

Example 2 : Roy model with an unobserved sector

In this example, Y (resp. η) denotes the wage an individual can obtain in sector 1 (resp. in sector 0). The individual chooses the sector that provides him with the

better wage. Y is observed if sector 1 is chosen but η is never observed. Thus, in this case $D = \mathbb{1}\{Y \geq \eta\}$.[20] For instance, Y may represent the potential wage of an individual, which is observed only if the person enters the labor market, while η denotes his reservation wage. The aim is to recover the distribution of Y, or the effects of covariates X on Y. The usual exclusion restriction requires the existence a variable which affects η but not Y. On the other hand, the strategy above can be applied if there is an instrument Z which affects the potential wage but not directly the reservation wage, so that η is independent of Z conditional on Y (or conditional on (X, Y) if one adds covariates). A possible example of such an instrument is the local unemployment rate (see Haurin & Sridhar, 2003, for evidence that the local unemployment rate does not affect the reservation wage).[21]

Example 3 : Sample from one response stratum

In this example, a researcher seeks to study the effects of Y on a binary variable D, but observes Y only for the stratum $D = 1$.[22] Our instrumental strategy relies on the existence of an instrument Z which affects Y but not D directly, and whose distribution is identified. Suppose for instance that one wants to study the efficiency of vaccination in a developing country, but data on ill people only are available, and the vaccination rate in the population is unknown. In this case D is the dummy variable of being ill, while Y is the dummy of being vaccinated. If there has been an important vaccination campaign after a given date, one can use the dummy of being born after this date as an instrument.[23] Once more, the completeness condition is satisfied as soon as the correlation between Y and the instrument is not zero.

This example also covers truncated count data models. In this case, the aim is to recover the effect of Y on an integer valued variable N, given that Y is observed

[20] Following the previous discussion, assumption 5 will be satisfied if η can be lower than any value of Y, with a positive probability.

[21] No statistical test for completeness conditions has been developed yet in the case where Y is continuous. Thus, assumption 4 has to be maintained in this example. However, one can test implications of assumption 4 by checking for instance that $E(Y|Z)$ is not a constant function.

[22] In this case, Y is a covariate rather than an outcome. The notation Y is maintained however to ensure consistency with assumption 1.

[23] If age is a factor of the disease as well, one can use only individuals born just before and just after the beginning of the campaign, as in the regression-discontinuity approach.

only when $N > 0$.[24] Consider for instance the estimation of the price elasticity of a good through the use of retail data.[25] If we observe the quantities sold N and the sales $N \times Y$, but not directly prices Y, then these prices can be deduced only when the quantities sold are positive. The framework can be applied if there is an instrument whose distribution is identified and which affects prices but not directly the demand. Production cost shifters such as prices of the inputs may be good candidates for that.

Testability

In some contexts, the conditional independence assumption 3 may seem overly strong. An interesting feature of this assumption, yet, is that it is refutable, contrary to the usual missing at random assumption. Firstly, equation (2.2.4) may have no solution. This is especially clear when (Y, Z) has a finite support. If indeed Y and Z take respectively s and t distinct values, with $t > s$, (2.2.4) can be written as a system of t equations with s unknown parameters, so that the model is overidentified.

But even when $s = t$, the model is testable since the solution Q of equation (2.2.4) must be a positive probability, i.e. $Q(y) \in]0, 1]$ for all y.[26] As an illustration, consider the simple case where $(Y, Z) \in \{0, 1\}^2$. Let $p(y, z) = P(D = 1, Y = y|Z = z)$, $\alpha = 1/Q(0)$ and $\beta = 1/Q(1)$. Then, as soon as $p(0, 0)p(1, 1) \neq p(0, 1)p(1, 0)$, that is to say under the completeness condition, equation (2.2.4) is equivalent to

$$\begin{aligned} \alpha &= \frac{p(1,1) - p(1,0)}{p(0,0)p(1,1) - p(0,1)p(1,0)} \\ \beta &= \frac{p(0,0) - p(0,1)}{p(0,0)p(1,1) - p(0,1)p(1,0)}. \end{aligned}$$

Hence, when $p(1, 1) - p(1, 0)$ and $p(0, 0) - p(0, 1)$ have opposite signs, for instance, assumption 3 is rejected. Basically, this happens when $z \mapsto P(D = 1|Y = y, Z = z)$ varies too much compared to $z \mapsto P(Y = y|Z = z)$.

[24] Hence, $D = \mathbb{1}\{N > 0\}$ here and recovering $P(N = k|Y)$ for all $k \in \mathbb{N}$ amounts to identifying $P(D = 1|Y)$. Note that this example differs from the simple truncation model $D = \mathbb{1}\{Y \geq s\}$ described above. In particular, assumption 5 will hold as soon as $P(N = 0|Y) < 1$ almost surely.

[25] As discussed by Grogger & Carson (1991), truncated counts arise more generally with data from surveys which ask participants about their number of participations, or administrative records in which inclusion in the database depends on having engaged in the activity of interest.

[26] If the completeness condition does not hold, Q may not be unique. Then at least one of the solution must belong to $]0, 1]$.

Now, when a solution $Q \in]0,1]$ of equation (2.2.4) does exist, one can expect that assumption 3 cannot be rejected, since intuitively, this equation makes use of all the available information. Theorem 2.2.4 formalizes this idea.

Theorem 2.2.4 *Suppose that assumption 1, 2 and 5 hold. Then assumption 3 can be rejected if and only if there exists no solution to equation (2.2.4) which belongs to* $]0,1]$.

When Y is discrete and takes values in $\{y_1, ..., y_s\}$, a statistical test of assumption 3, under the maintained assumption 4, can be developed as follows. First, we can estime $f = 1/P$ by GMM using (2.2.4). Then testing assumption 3 amounts to making a test of the multiple inequality constraints $f(y_j) \geq 1$ for $j = 1...s$ (see e.g. Gouriéroux & Monfort, 1995, section 21.4, for the implementation of such tests). The situation is more involved when Y is continuous. Under assumptions 1-5 and additional technical conditions, a consistent nonparametric estimator $\widehat{f}$ of f is developed in subsection 3.2. This estimator is constrained to belong to $[1, M]$ with $M > 1$. It should be possible to build a consistent, unconstrained estimator $\widetilde{f}$ of f. Then, under the maintained assumptions 1, 2, 4 and 5, a test of assumption 3 could be based on the distance between $\widehat{f}$ and $\widetilde{f}$. Indeed, under assumption 3, $\widetilde{f}(y)$ should be greater than one for most values of y, so the distance between the two should be close to zero.[27]

Set identification without conditional independence

A second interesting feature of equation (2.2.4) is that it provides an informative bound on parameters of interest under monotonicity conditions, which are far weaker than the conditional independence condition of assumption 3. Because monotonicity conditions are meaningful in ordered sets only, we restrict here to the case where $(Y, Z) \in \mathbb{R}^2$. Besides, let $\widetilde{Z}$ denote a variable which may differ from Z and whose distribution is also identified. Assumption 3 is replaced by the following ones.

Assumption 3' *Almost surely,* $z \mapsto P(D = 1|Y, Z = z)$ *is increasing.*

Assumption 6 *Almost surely,* $y \mapsto P(D = 1|Y = y, \widetilde{Z})$ *is increasing.*

[27] The critical region of such a test would depend on the asymptotic distribution of $(\widehat{f}, \widetilde{f})$, whose derivation is beyond the scope of the paper.

Assumption 3' weakens the conditional independence between selection and instrument set in assumption 3 into a monotone dependence. It is also a variant of the usual instrumental condition which assumes that the instrument affects the probability of selection but is independent of the outcome. Here, the effect on the probability of selection is restricted to be monotonic, but no independence condition between Y and Z is needed. Assumption 6 weakens the missing at random hypothesis of independence between selection and outcome into a monotone dependence.

Theorem 2.2.5 below provides bounds on parameters of the form $E(h(Y))$ for $h \in H_Y$ or $h \in H_{YZ}$, where we let

$$\begin{aligned} H_T &= \{h \in L_T^1 \text{ and } h \text{ is increasing}\} \ (T = Y \text{ or } Z), \\ H_{YZ} &= \{h \in L_Y^1 / \exists \widetilde{h} \in H_Z / h(Y) = E(\widetilde{h}(Z)|D = 1, Y)\}. \end{aligned}$$

The set H_Y includes, among others, functions of the form $h(y) = \lambda y$ with $\lambda > 0$ and indicator functions $h_u(y) = \mathbb{1}\{y \geq u\}$, so that parameters of the form $E(h(Y)), h \in H_Y$, include the survival function of Y at each point. The set H_{YZ} is more abstract. In an informal way, H_{YZ} will increase as the dependence between Y and Z becomes stronger. As a simple illustration, this set only includes constant functions when Y and Z are independent (conditional on $D = 1$) but is equal to H_Y when $Y = Z$. More formally, H_{YZ} is a subset of the range of the conditional expectation operator $g \mapsto (y \mapsto E(g(Z)|D = 1, Y = y))$, which itself is linked to the null space of this operator. Indeed, when (Y, Z) has finite support, the dimension of the range will increase as the dimension of the null space decreases. Thus, at least in finite dimension, H_{YZ} will be maximal if the conditional expectation operator is injective, that is to say under a completeness condition on Y and Z.[28]

It seems difficult to test formally that $h \in H_{YZ}$ for a given, increasing, function h. On the other hand, we can test the stronger condition :

$$E(Z|D = 1, Y) = \alpha + \beta h(Y), \quad \beta > 0 \tag{2.2.5}$$

Test of such functional forms are described for instance by Yatchew (1998, subsection 4.2).

[28] If (Y, Z) has infinite support and the conditional expectation operator is injective, one can show that the dimension of H_{YZ} is infinite.

We suppose afterwards that equation (2.2.4) has a solution, and, as in the previous subsection, we let Q denote such a solution. More precisely, if the constant function $P(D=1)$ is a solution, we let $Q(Y) = P(D=1)$ but otherwise Q can be any of the solutions. We do not impose it neither to lie in $]0,1]$ nor to be unique, so that cases where the completeness condition 4 fails can also be handled.

Theorem 2.2.5 *Suppose that $P(D=1) > 0$ and assumptions 1 and 2 hold for Z and $\widetilde{Z}$. Then :*
a) Under assumption 6, $E[h(Y)] \leq E\left[E(h(Y)|\widetilde{Z}, D=1)\right]$ for all $h \in H_Y$. Moreover, this upper bound is sharp ;
b) Under assumptions 3', $E[Dh(Y)/Q(Y)] \leq E[h(Y)]$ for all function $h \in H_{YZ}$. Moreover, this lower bound is sharp provided that at least one solution Q lies in $]0;\ 1]$.
c) For all function $h \in L^1_Y$, these three expectations are equal when $D \perp\!\!\!\perp (Y, Z, \widetilde{Z})$ or when $Z = \widetilde{Z} = Y$.

Part a) of theorem 2.2.5 is not specific to the methodology developed here, and is rather straightforward. Part b), on the other hand, shows that the moment condition used here leads to a sharp lower bound on this parameter. This lower bound does not depend on the choice of the solution Q of equation (2.2.4), so that no completeness condition is required. The bound also holds even if no solution Q lies in $]0;\ 1]$. In this case however, the bound may not be sharp because one could exploit the fact that the conditional independence assumption 3 is rejected by the data.

An important consequence of theorem 2.2.5 is that for all functions $h \in H_Y \cap H_{YZ}$, we can obtain a compact interval on $E(h(Y))$. This is so even if $h(Y)$ is unbounded. In this sense, the result is similar to proposition 2, corollary 2 of Manski & Pepper (2000), under a different set of assumptions. In particular, we do not rely on the monotone treatment response condition, which is difficult to adapt to the context of selection models or nonresponse. Moreover, the monotone treatment response assumption can be strong in the context of treatment effects. In the Roy model with an unobserved sector developed in example 2, it asserts that almost surely, $Y_1 \geq Y_0$ (or $Y_0 \geq Y_1$), so that only one sector would be chosen at equilibrium, a rather unrealistic situation. Instead of this condition, assumption 3' supposes the existence of an instrument such that the probability of selection increases with

this instrument. This assumption is rather weak and should be satisfied in many contexts, including treatment effects estimation, or estimation of parameters with nonignorable missing data. In example 2, one could use standard instruments such as non-wage income or the number of children for instance.

As part c) shows, the interval can be reduced to a point if D is fully missing at random. Hence, the length of the interval can be interpreted as a measure of the severity of the selection problem. Because the interval is also reduced to a point when $Z = \widetilde{Z} = Y$, its length also reflects the quality of the chosen instruments. As the dependence between $(Z, \widetilde{Z})$ and Y increases, the knowledge of the distribution of the instruments enables to better predict parameters of the distribution of Y. Besides, the upper (resp. lower) inequality turns into an equality whenever $Y \perp\!\!\!\perp D|\widetilde{Z}$ (resp. $Z \perp\!\!\!\perp D|Y$). Hence, Z and $\widetilde{Z}$ must be chosen according to different logics. $\widetilde{Z}$ intends to reduce selection on inobservables correlated with the outcome, whereas Z should be as independent of the selection (conditional on Y) as possible.

As noted before, H_{YZ} increases as the dependence between Y and Z becomes stronger. Hence, the quality of the instrument also matters for the range of applicability of the lower bound. If it seems difficult, without further restrictions, to describe the set $H_Y \cap H_{YZ}$ of functions h such that an interval can be built on $E[h(Y)]$, this set will contain at least all functions $h(y) = \lambda y$ with $\lambda > 0$ under the testable linear condition that $E(Z|D = 1, Y) = \alpha + \beta Y$ (with $\beta > 0$). In this case in particular, $E[Y]$ can be bounded below and above. Besides, if Y and Z exhibit a positive dependence, the following proposition states that the set $H_Y \cap H_{YZ}$ will be equal to H_{YZ}.

Proposition 2.2.6 *Suppose that for all z, $y \mapsto F_{Z|Y=y,D=1}(z)$ is decreasing. Then $H_{YZ} \subset H_Y$.*

Parametric identification

Nonparametric identification stems from the uniqueness of a functional equation. However, one may be reluctant to use nonparametric estimators in practice, because of the curse of dimensionality for instance. Furthermore, assumption 2 may be too strong in some circumstances. Suppose for instance that instruments are observed only when $D = 1$ (as with unit nonresponse or attrition in a panel), but auxiliary information is available on these instruments. This auxiliary information

may however not be sufficient to identify the full distribution of Z. If Z is multivariate and its different components are observed through different sources which cannot be matched, only the marginal distributions will be identified. If the instruments are measured with a zero mean error in these auxiliary data, only $E(Z)$ can be recovered.

In such situations, assumption 2 fails but intuitively, information on Z can provide identification, at least in a parametric setting. Theorem 2.2.5 gives a rigorous treatment to this idea. It generalizes the framework of Nevo (2002) to the case where $Y \neq Z$. It is also very similar to the theory of generalized calibration developed by Deville (2001) in a survey sampling framework to handle nonignorable nonresponse with instruments. Deville (2001), however, does not consider the issue of identification of P.

As we consider a parametric framework here, we add explicitly covariates X. In the following, we suppose that $V = (X', Y')' \in \mathbb{R}^p$ and $W = (X', Z')' \in \mathbb{R}^q$. The identification result is based on the following assumptions.

Assumption 2' *$E(W)$ is known. Moreover, $P(D = 1|V) = F(V'\beta_0)$ where F is a known, differentiable and strictly increasing function from $\mathbb{R}$ to $]0, 1[$, and V is almost surely linearly independent conditional on $D = 1$.*

Assumption 3' $D \perp\!\!\!\perp Z|V$.

Assumption 4' *rank*$(E(DWV'F'(V'\beta_0)/F^2(V'\beta_0))) = p$.

Assumption 4" $E(Z|D = 1, V) = \Gamma_1 X + \Gamma_2 Y$ *where Γ_2 is full rank.*

Assumption 2' weakens assumption 2 on data availability, at the price of imposing a parametric restriction on P. The condition $P(D = 1|V) = F(V'\beta_0)$ with a known F is satisfied for instance if the selection equation is a logit or probit model. Like assumption 4 in the nonparametric setting, assumption 4' is the rank condition. As usually, this condition implies that $q \geq p$. Lastly, assumption 4" is a particular case of assumption 4', which restricts the nonparametric regression of Z on V to a linear form.

Theorem 2.2.7 *Suppose that assumptions 1, 2' and 3' are satisfied. then*
a) β_0 is locally identified if and only if assumption 4' holds.
b) if assumption 4" holds, β_0 is globally identified.

Local identification is obtained under a condition which is very similar to the rank condition in linear regressions with instruments. Theorem 2.2.7 also provides a sufficient and testable condition which ensures the global identification of β_0.

2.2.3 Estimation

We now turn to the parametric and nonparametric estimation of P. The first assumption describes the sampling process.

Assumption 7 *We observe a sample* $((D_1, X_1, Y_1^*, Z_1), ..., (D_n, X_n, Y_n^*, Z_n))$ *of independent copies of* (D, X, Y^*, Z)*, with* $Y^* = DY$*.*

Assuming that the data are i.i.d. is standard in estimation, although this condition can be weakened without affecting consistency or rate of convergence. We also suppose, for the sake of simplicity, that Z is always observed in the data.

Parametric estimation

When Y has a finite support $\{y_1, ..., y_s\}$, the equation

$$E\left(\frac{D}{\sum_{k=1}^{s} P(y_k) 1\{Y = y_k\}} - 1 \middle| Z\right) = 0$$

provides identification of the parameters $(P(y_k))_{1 \leq k \leq s}$ if assumptions 3, 4 and 5 hold, by theorem 2.2.3. Hence, consistent and asymptotically normal estimators can be obtained by GMM in this case. Similarly, if P satisfies the restrictions of assumption 2', then

$$E\left[\left(\frac{D}{F(V'\beta_0)} - 1\right) W\right] = 0. \tag{2.2.6}$$

Moreover, the proof of theorem 2.2.7 (see equation (2.2.27)) ensures that under assumption 4", β_0 is identified globally by these conditions. Thus GMM can also be used in this framework.

Nonparametric estimation

When Y has continuous components and one is reluctant to rely on parametric restrictions on P, the situation is more involved because a function, and not only parameters, must be estimated. This issue is similar to the one of nonparametric instrumental regression (see e.g. Newey & Powell, 2003, Hall & Horowitz, 2005,

Darolles et al., 2007 and Horowitz & Lee, 2007). For the sake of simplicity, we assume that there is no covariate X and that $(Y, Z) \in [0,1]^2$. Moreover, since the paper is mainly focused on identification, we only prove consistency here. The analysis of the rate of convergence could be led by adapting the arguments of Hall & Horowitz (2005).

Let us denote $f = 1/P$ and T be the linear operator defined by

$$T\,\phi(z) = E(D\phi(Y^*)|Z = z).$$

Then (2.2.4) may be written as

$$T\,f = 1.$$

We rely on this equation for estimating f. Because the problem is ill-posed,[29] regularization is needed to ensure consistency of the estimator. We adopt here a Tikhonov regularization, as Hall & Horowitz (2005), Darolles et al. (2007) and Horowitz & Lee (2007). First, let us consider the kernel estimator of T :

$$\widehat{T}\,\phi(z) = \frac{\sum_{i=1}^n D_i\phi(Y_i^*)K_{h_n}(z - Z_i)}{\sum_{i=1}^n K_{h_n}(z - Z_i)}$$

For any $1 < M < \infty$, let us define D_M as the subset of real measurable functions ϕ defined on $[0,1]$ and such that $M \geq \phi(Y) \geq 1$ almost surely. For any square integrable function ϕ defined on $[0,1]$, let also $||\phi||^2 = \int_0^1 \phi(u)^2 du$. Our estimator of f satisfies

$$\widehat{f} \in \arg\min_{\phi \in D_M} \left\| \widehat{T}\,\phi - 1 \right\|^2 + \alpha_n \left\|\phi\right\|^2$$

where α_n is a regularization parameter which, basically, enables to rule out unstable solutions (see e.g. Carrasco et al., 2006, for a discussion on regularization in ill-posed inverse problems). Under the assumptions below, such a solution will always exist but may not be unique (see Bissantz et al., 2004). If not, $\widehat{f}$ is any of the solutions. The consistency result relies on the following assumptions.

Assumption 8 *(a) $f \in D_M$. (b) The distribution of (Y, Z) is continuous with respect to the Lebesgue measure and the marginal densities f_Y and f_Z satisfy $\sup_{y\in[0,1]} f_Y(y) < +\infty$ and $\inf_{z\in[0,1]} f_Z(z) > 0$.*

[29] Indeed, T is unknown and can be estimated only by a finite range estimator. The situation is similar to the one of Gagliardini & Scaillet (2006) in the framework of functional minimum distance.

Assumption 9 *For all $h > 0$ and $u \in \mathbb{R}$, $K_h(u) = K_1(u/h)$ where K_1 is positive, $\int K_1(u)du = 1$ and $\int uK_1(u)du = 0$.*

Assumption 10 *$\alpha_n \to 0$, $h_n^2 + 1/(nh_n) \to 0$ and $(h_n^2 + 1/(nh_n))/\alpha_n \to 0$.*

Assumption 8-(a) strengthens assumption 5. Assumption 9 is weak and standard in nonparametric estimation. Assumption 10, which is identical to assumption 3 of Horowitz & Lee (2007), is also standard. It implies that the bandwidth h_n tends to zero at a slower rate than $1/n$, and that the regularization parameter α_n tends to zero at a slower rate than h_n^2.[30]

Theorem 2.2.8 *Under assumptions 3-4 and 7-10,*

$$\lim_{n\to\infty} E\left(\left\|\widehat{f} - f\right\|^2\right) = 0$$

Theorem 2.2.8 implies that $\left\|\widehat{f} - f\right\|^2$ converges in probability to zero. With $\widehat{f}$ in hand, inverse probability weighting procedures can be used to estimate parameters on the whole population. Let $\widehat{f}^{-i}$ denotes the estimator of f obtained with the sample $(D_j, Y_j^*, Z_j)_{j\neq i}$. For any $g \in L^2_{Y,Z}$ and $\theta = E(g(Y, Z))$, define

$$\widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n} D_i \widehat{f}^{-i}(Y_i^*) g(Y_i^*, Z_i).$$

Corollary 2.2.9 ensures that $\widehat{\theta}$ is consistent.

Corollary 2.2.9 *Suppose that assumptions 3, 4 and 7-10 hold. Then*

$$\lim_{n\to\infty} E\left(|\widehat{\theta} - \theta|\right) = 0$$

2.2.4 Monte Carlo simulations

In this section, we investigate the finite sample properties of parametric and non-parametric estimators of P and inverse probability weighted estimators of $E[g(Y)]$. Let us consider the following model :

$$\begin{cases} Y &= \Lambda(\Lambda^{-1}(Z) + \varepsilon) \\ D &= \mathbb{1}\{P(Y) \geq \eta\} \end{cases} \tag{2.2.7}$$

[30] We suppose here that α_n is a deterministic sequence. See e.g. Gagliardini & Scaillet (2006) for a data driven selection procedure.

where $\Lambda(x) = 1/(1+\exp(-x))$ is the logistic cumulative distribution function, $P(y) = 1 - 0.6/(1+19y^2)$, (Z, ε, η) are mutually independent, $Z \sim U[0,1]$, $\varepsilon \sim N(0,1)$ and $\eta \sim U[0,1]$.[31] The function P was chosen to match the estimator of P in the application (see figure 2below). Within this framework, assumptions 3 and 4 are satisfied by proposition 2.2.1 and 2.2.2. Assumption 8 also holds, with in particular $f(y) = 1/P(y) \leq 2.5$. Lastly, $P(D=1) \simeq 0.8$ so that approximately 20% of the Y are missing.

Estimator	Statistic	$n = 100$	$n = 200$	$n = 500$	$n = 1,000$
$\widehat{f}_1$	MISE	0.1978	0.1532	0.1058	0.0791
$\widehat{f}_2$		0.2010	0.1478	0.1017	0.0758
$\widehat{f}_3$		1.9343	0.3697	0.0673	0.0286
$\widehat{\theta}_1$	RMSE	0.0330	0.0252	0.0155	0.0120
	(bias)	(-0.0081)	(-0.0066)	(-0.0061)	(-0.0058)
$\widehat{\theta}_2$		0.0373	0.0275	0.0174	0.0139
		(-0.0191)	(-0.0130)	(-0.0099)	(-0.0091)
$\widehat{\theta}_3$		0.0316	0.0238	0.0140	0.0104
		(0.0001)	(-0.0005)	(-0.0002)	(-0.0001)
$\widehat{\theta}_4$	RMSE	0.0338	0.0253	0.0154	0.0112

The results were obtained with 1,000 simulations for each sample size. The bias of $\widehat{\theta}_4$ is not indicated as this estimator is unbiased.

TAB. 2.1 – Performances of parametric and nonparametric estimators.

We consider two nonparametric estimators $\widehat{f}_1$ and $\widehat{f}_2$ of f. For both, the regularization parameter is the same, $\alpha_n = 0.05 \times n^{-1/5}$, but their bandwidths differ, namely $h_{1n} = 0.03 \times n^{-1/5}$ and $h_{2n} = 0.02 \times n^{-1/5}$. We also consider a parametric estimator $\widehat{f}_3$ defined by supposing that f belongs to the following flexible parametric family

$$f(y;\beta) = 1 + \exp\left(-\beta_0 - \sum_{k=1}^{4} y \mathbb{1}\{y \geq a_k\}\beta_k\right) \tag{2.2.8}$$

where $\beta = (\beta_0, ..., \beta_4)$, $a_1 = -\infty$ and (a_2, a_3, a_4) are the estimated quartiles of the distribution of Y conditional on $D = 1$. β is estimated through GMM, using

[31] The model amounts to assuming a linear dependence between $\Lambda^{-1}(Y)$ and $\Lambda^{-1}(Z)$. We work with (Y, Z) rather than $(\Lambda^{-1}(Y), \Lambda^{-1}(Z))$ to be consistent with the previous assumption that $(Y, Z) \in [0,1]^2$.

as instrumental variables 1 and $(Z\mathbb{1}\{Z \geq c_i\})_{1\leq i\leq k}$, where $c_1 = -\infty$ and the (c_2, c_3, c_4) are the estimated quartiles of Z. We measure the accuracy of the three estimators of f by the usual mean integrated square error (MISE) :

$$\text{MISE}(\widehat{f}) = E\left(\int_0^1 (\widehat{f}(u) - f(u))^2 du\right).$$

We also consider inverse probability weighted estimators of $\theta = E(Y) = 1/2$. Let us define, for $k \in \{1, 2, 3\}$,

$$\widehat{\theta}_k = \frac{1}{n}\sum_{k=1}^{n} D_i \widehat{f}_k(Y_i^*) Y_i^*.$$

We also compute the infeasible estimator

$$\widehat{\theta}_4 = \frac{1}{n}\sum_{k=1}^{n} D_i f(Y_i^*) Y_i^*.$$

The accuracy of each estimator is described through their bias and root mean square error (RMSE). Results are displayed in table 2.1. On average, $\widehat{f}_2$ outperforms $\widehat{f}_1$, and also the parametric estimator $\widehat{f}_3$ for small sample sizes. $\widehat{f}_3$ is indeed somewhat erratic for small n, but becomes far more accurate than the nonparametric estimators for moderately large n. It seems, in this design, that the bias due to the parametric misspecification is negligible compared to the accuracy gains stemming from the parametric procedure. The corresponding estimator $\widehat{\theta}_3$ is also the most precise one, even for small samples. $\widehat{\theta}_1$ outperforms $\widehat{\theta}_2$, confirming the idea that a better first-step nonparametric estimator does not necessarily yield a better second-step estimator. Lastly, $\widehat{\theta}_4$ is less accurate than $\widehat{\theta}_3$ and comparable with $\widehat{\theta}_1$. It may be that estimating f in a first step actually yields a lower asymptotic variance than the one of $\widehat{\theta}_4$, as for instance with Lewbel's estimator in binary choice models (see Lewbel, 2000 and Magnac & Maurin, 2007).

2.2.5 Application

Introduction

In this section, the strategy developed above is used to estimate bounds on the short term effects of grade retention among fifth grade students in France. Whereas most countries have almost completely given up grade retention as an educational

policy,[32] the level of grade retention in France is still high. In 2002, for instance, a quarter of students have repeated at least once in primary school (see Troncin, 2004). Yet, and despite the controversy on its effects in other countries,[33] there has been no serious attempts to measure its impact in the French educational system.[34]

The study is based on a panel of the French "Ministère de l'éducation Nationale" which follows 9,641 children who entered the first grade of primary school in 1997. Among others, the panel reports the trajectories of children and their results in standardized tests at the beginning of the third grade (variable Z) and sixth grade (variable Y for the 2002 test and Y_1 for the 2003 test).[35] Because the sixth grade test scores are reported in the database only for pupils who reached this grade in 2002 or in 2003, the initial sample comprises 7,175 students who were in fifth grade in 2001 and in sixth grade either in 2002 or in 2003.[36] 23.8 percent of this sample was excluded because of missing data on the standardized test scores in either third or sixth grade. The final sample consists in 5,467 children. Among them, 2.2% were retained in 5th grade ($D = 0$), 6.7% in 6th grade ($D = 1$ and $D_1 = 0$) while the others never repeated ($D = 1$ and $D_1 = 1$). Table 2.2 displays the average scores on this sample. The 2002 6th grade score is missing for children retained in 5th grade since they only entered this grade in 2003. Similarly, the 2003 6th grade score is not observed for children who never repeated, since they

[32] A notable exception is the United States. Indeed, several states have reintroduced this policy by tying promotion on a state or district assessment (see Jacob & Lefgren, 2004).

[33] Positive effects include the possibility for disadvantaged children to catch up (see e.g. Jacob & Lefgren, 2004) and the incentive for every student to increase their school efforts (see Jacob, 2005). On the other hand, most educational and sociological studies underline its harmful effects on the motivation of children (see e.g. Crahaye, 1996), drop outs (see Jimerson et al., 2002) and even academic performances (see e.g. the meta-analyses of Holmes, 1989, or Jimerson, 2001). However, usually, these studies rely on very few controls (see e.g. Lorence, 2006, for a discussion on the studies considered in the meta-analyses of Holmes and Jimerson), so that they probably underestimate the true effects of grade retention.

[34] Troncin (2005) measures the effects of grade retention in the first grade of primary school using a propensity score matching approach, but he relies on data from one school only. Cosnefroy & Rocher (2004) study the effects in third grade on the same data as here, using a linear regression approach.

[35] Tests corresponding to a given grade differ partly from year to year. The scores considered here are built using common items only. The three scores are also standardized on the final sample.

[36] Other situations correspond to missing data on the trajectories, grade-advanced pupils, pupils retained before the fifth grade and students in special classrooms.

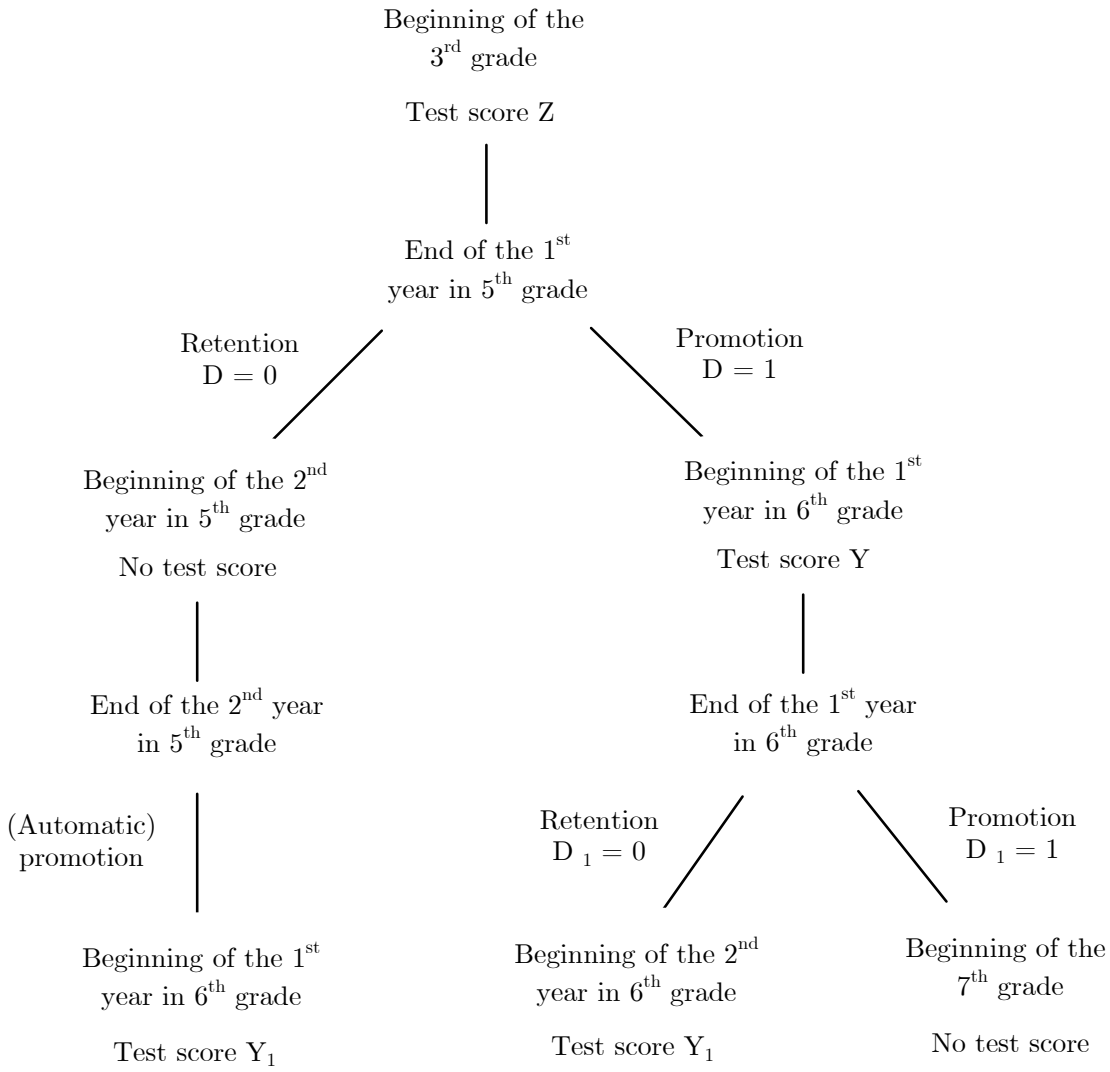

FIG. 2.1 – Promotion, retention and available test scores.

were in 7th grade in 2003. As expected, differences between retained and promoted pupils in terms of test achievement are large. On average, the fifth (resp. sixth) grade repeaters were already, in the 3rd grade test, more than 1.5 (resp. more than 1) standard deviations below the students who never repeated. The table also displays the progression of students retained in 6th grade during their first year in this grade. This progression is available because these students take the test twice, at the beginning of their first and second year in sixth grade (see figure 1). This feature of the sample will be useful in the following.

	Retained in 5th grade ($D=0$)	Retained in 6th grade ($D=1, D_1=0$)	Promoted in both grades ($D=1, D_1=1$)
Number of observations	120	365	4982
3rd grade score Z	-1.48 (0.91)	-1.02 (0.90)	0.11 (0.94)
2002 6th grade score Y	-	-1.32 (0.81)	0.12 (0.93)
2003 6th grade score Y_1	-0.90 (0.87)	-0.64 (0.79)	-

TAB. 2.2 – Summary statistics.

We focus here on the average effects of retention in fifth grade on test score achievement one year after. Let $Y_1(1)$ (resp. $Y_1(0)$) denote the 2003 sixth grade test score a student would have obtained if he had been promoted in sixth grade (resp. retained in fifth grade). Then the parameter of interest is defined by

$$\Delta^{TT} = E(Y_1(0) - Y_1(1)|D=0) \tag{2.2.9}$$

When $D=0$, $Y_1(0)$ is observed by Y_1, but $Y_1(1)$ is unobserved. Because there is no exogenous rule acting on grade retention decisions in France, it seems difficult to rely on an instrumental strategy to overcome this counterfactual issue.[37] Rather, we suppose that the progressions of retained students had they been promoted in sixth grade can be bounded in the following way :

$$0 \leq E(Y_1(1) - Y|D=0, Y) \leq E(Y_1(1) - Y|D=1, D_1=0, Y). \tag{2.2.10}$$

The lower bound simply asserts that on average, retained students would not have regressed during one year, had they been promoted. The upper bound states that on average, their progression would have been smaller than the one of students with same initial test score and who were promoted in sixth grade and retained the year after. The idea behind this bound is that, on average, teachers do not make mistakes by retaining pupils who would have benefited more from the sixth grade than some of the promoted students. The two bounds somewhat represent two extreme situations. The lower bound corresponds to perfect decisions of retention, in that retained students would not have taken any advantage of being promoted.

[37] As an evidence of the discretionary nature of grade retention in France, a Bill of the Minister of Education in 2005 asserts that grade retention should be taken by teachers after discussion with parents, according to the ability of the student and his progression during the year.

The upper bound corresponds to a fully randomized choice among students who would have equally benefited from being promoted.

Under condition (2.2.10), we get

$$E(Y_1|D=0) - E\left[h(Y)|D=0\right] \leq \Delta^{TT} \leq E(Y_1|D=0) - E(Y|D=0), \quad (2.2.11)$$

where $h(Y) = E(Y_1(1)|D = 1, D_1 = 0, Y)$. Students retained in sixth grade take the standardized test twice. Thus, we observe both Y and $Y_1(1)$ for them ($Y_1(1) = Y_1$ in this case), and h is identified. On the other hand, Y is unobserved for students retained in fifth grade, so that $E[h(Y)|D = 0]$ and $E(Y|D = 0)$ are not identified without further restrictions. Nonetheless, we can use the method developed previously to point or set identify them. Indeed, Y, the main factor of D, is unobserved when $D = 0$. Besides, the third grade standardized test score Z is observed for both values of D and correlated with Y. We now consider the two cases corresponding respectively to the independence assumption $D \perp\!\!\!\perp Z|Y$ and the monotonicity conditions considered in subsection 2.3.

Empirical strategies

First strategy : conditional independence

First, let us suppose that grade retention in fifth grade is independent of the third grade test score conditional on Y, i.e. a model of the form :

$$\begin{cases} Y &= \varphi(Z, \varepsilon) \\ D &= \psi(Y, \eta) \end{cases}$$

where $\eta \perp\!\!\!\perp (Z, \varepsilon)$. The completeness condition is also supposed to hold. Informally, both will be satisfied if the third grade score affects the ability at the end of the fifth grade, measured by Y, but not directly grade retention. Under these assumptions, theorem 2.2.3 applies and letting $p = P(D = 0)$, we can identify $E(h(Y)|D = 0)$ by

$$\begin{aligned} E\left[h(Y)|D=0\right] &= \frac{1}{p}\left(E[h(Y)] - (1-p)E[h(Y)|D=1]\right) \\ &= \frac{1}{p}\left((1-p)E\left[\frac{h(Y)}{P(Y)}|D=1\right] - (1-p)E[h(Y)|D=1]\right) \\ &= \frac{1-p}{p}E\left[\frac{1-P(Y)}{P(Y)}h(Y)|D=1\right]. \end{aligned}$$

$E(Y|D=0)$ can be identified similarly. Then, using (2.2.10), we obtain the following lower and upper bounds on Δ^{TT} :

$$\underline{\Delta}_1^{TT} = E[Y_1|D=0] - \frac{1-p}{p} E\left[\frac{1-P(Y)}{P(Y)} h(Y)|D=1\right] \quad (2.2.12)$$

$$\overline{\Delta}^{TT} = E[Y_1|D=0] - \frac{1-p}{p} E\left[\frac{1-P(Y)}{P(Y)} Y|D=1\right] \quad (2.2.13)$$

To estimate these bounds, h and P have to be estimated first. A kernel estimator was used for h, with a gaussian kernel and a bandwidth estimated by cross validation (see figure 2). P was estimated by the flexible parametric form $P(y;\beta) = 1/f(y;\beta)$ with $f(y;\beta)$ defined by (2.2.8) and the same instruments as in the Monte Carlo simulations, except that the thresholds (a_2, a_3, a_4) and (c_2, c_3, c_4) correspond to the estimated quantiles of order 8, 16 and 24 of Y conditional on $D=1$ and Z respectively.[38] The estimator $P(.;\widehat{\beta})$ is displayed in figure 2.[39]

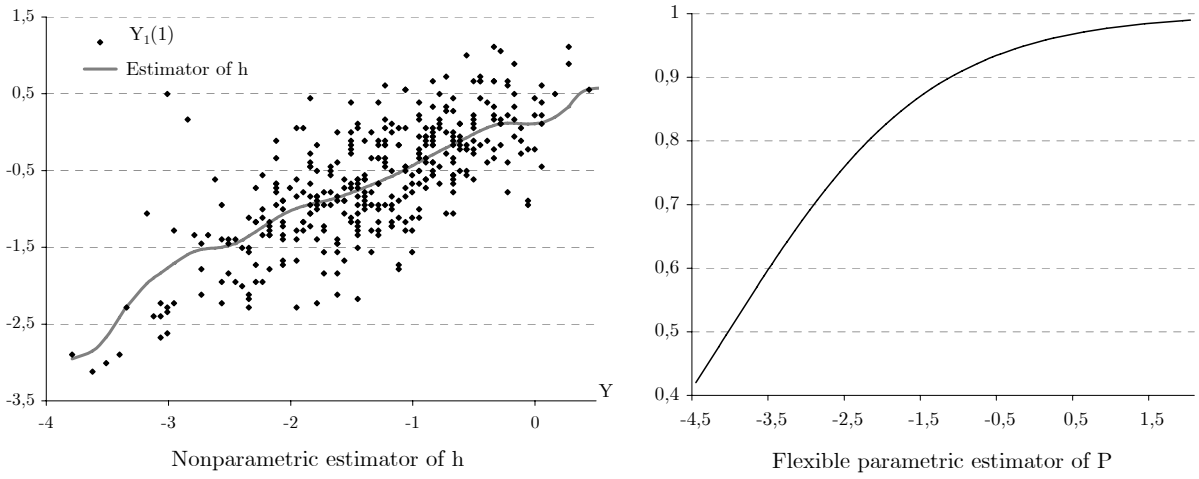

FIG. 2.2 – Estimation of h and P.

The estimator of $\underline{\Delta}_1^{TT}$ and $\overline{\Delta}^{TT}$ are then defined as being the empirical analog of

[38] Several specifications have been tried. Final results are unsensitive to the choice of these thresholds.

[39] This plot corresponds to $\widehat{\beta}_0 = 3.07$, $\widehat{\beta}_1 = 0.75$, $\widehat{\beta}_2 = 4.13$,$\widehat{\beta}_3 = 34.3$, $\widehat{\beta}_4 = 0.42$.

(2.2.12) and (2.2.13) :

$$\widehat{\underline{\Delta}_1^{TT}} = \frac{1}{n_0}\left[\sum_{i/D_i=0} Y_{1i} - \sum_{i/D_i=1} \frac{P(Y_i;\widehat{\beta})}{1-P(Y_i;\widehat{\beta})}\widehat{h}(Y_i)\right],$$

$$\widehat{\overline{\Delta}^{TT}} = \frac{1}{n_0}\left[\sum_{i/D_i=0} Y_{1i} - \sum_{i/D_i=1} \frac{P(Y_i;\widehat{\beta})}{1-P(Y_i;\widehat{\beta})}Y_i\right],$$

where n_0 denotes the number of pupils who repeated their fifth grade.

Second strategy : monotonicity

Basically, the conditional independence condition holds if Y is a perfect measure of ability at the end of fifth grade and if teachers only take into account the current ability when deciding whether to retain a student or not. If the second statement is rather plausible given that teachers usually do not observe children's ability before they enter their grade, the first statement seems too restrictive. Past scores probably bring additional information on the current ability and thus explain part of grade retention. On the other hand, it seems very plausible in this case that the dependence in both variable is monotonic, i.e., assumption 3' and 6 hold. To provide empirical evidence on this assumption, a logit model on D_1 among students who were promoted in sixth grade was estimated. For these students indeed, both Y and Z are known. The results, which are displayed table 2.3, confirm the monotonicity in both variables. As expected, we also observe a far smaller effect of the third grade test score.

Variable	Estimate (std. err.)
2002 6th grade score Y	1.31 (0.08)
3rd grade score Z	0.23 (0.07)

TAB. 2.3 – Logit estimation on the probability of promotion in sixth grade.

To apply theorem 2.2.5 and obtain bounds on $E(h(Y)|D=0)$, we also need to check that $h \in H_Y \cap H_{YZ}$. That h is increasing is apparent from figure 2.To check that $h \in H_{YZ}$, we implemented, as suggested in subsection 2.3, the specification test of the form (2.2.5).[40] We obtain a positive and significant slope coefficient in

[40]More precisely, the simple differencing test suggested by Yatchew (1998, p. 701) was implemented, with the kernel estimator $\widehat{h}$ instead of h.

(2.2.5) and do not reject, at the level of 1%, the linear specification. Hence, we do not reject the assumption that $h \in H_{YZ}$.

Under assumptions 3' and 6, and the condition $h \in H_Y \cap H_{YZ}$, we can apply theorem 2.2.5 to obtain the following bounds on $E(h(Y)|D=0)$:

$$\frac{1-p}{p} E\left[\frac{1-Q(Y)}{Q(Y)} h(Y)|D=1\right] \leq E\left[h(Y)|D=0\right] \leq E\left[E(h(Y)|Z,D=1)|D=0\right]$$

where Q denotes a solution of $E(D/Q(Y)-1|Z)=0$.[41]

To get bounds on $E(Y|D=0)$, we also check that the identity function belongs to H_{YZ}. This is true if $E(Z|D=1,Y)=\gamma+\lambda Y$ with $\lambda>0$. The specification test was not rejected at the level of 5%, so that we accept that the identity function belongs to $H_Y \cap H_{YZ}$. Under these assumptions, we get the same upper bound on Δ^{TT} as under conditional independence, but another lower bound, which satisfies

$$\underline{\Delta}_2^{TT} = E[Y_1|D=0] - E\left[E(h(Y)|Z,D=1)|D=0\right] \tag{2.2.14}$$

Moreover, $\underline{\Delta}_2^{TT}$ and $\overline{\Delta}^{TT}$ are sharp by theorem 2.2.5.

To estimate $\underline{\Delta}_2^{TT}$, a kernel estimator $\widehat{g}$ of $g(z)=E(h(Y)|Z=z,D=1)$ was first estimated, and then plugged into the empirical analog of (2.2.14) :

$$\widehat{\underline{\Delta}_2^{TT}} = \frac{1}{n_0}\left[\sum_{i/D_i=0} Y_{1i} - \sum_{i/D_i=1} \widehat{g}(Z_i)\right].$$

Results

The final results are displayed in table 2.4. Under the assumption of a fully valid instrument, the interval only ranges positive values, so that grade retention leads to positive short terms effect even in the least favorable case.[42] The pattern is less clear if one weakens the instrumental exclusion restriction into a monotonicity condition. Under the extreme case where grade retention only depends on the third grade test score, this policy would be harmful in terms of test achievement. This assumption does not seem very credible, though. As emphasized previously, the effects of Y on D is probably much more important than the one of Z. Thus, even in the worst case, the true effect is more likely to be close to $\widehat{\underline{\Delta}_1^{TT}}$, that is to say around zero.

[41] We do not use P here to emphasize the fact that the solution of this equation is not $P(D=1|Y)$ anymore. However, both P and Q are estimated with $P(.;\widehat{\beta})$.

[42] Indeed, the null hypothesis that the lower bound is negative is rejected at 5%.

Estimator	Value	95% Confidence interval
$\widehat{\overline{\Delta}^{TT}}$	1.17 (0.24)	[0.75,1.67]
$\widehat{\underline{\Delta}_1^{TT}}$	0.29 (0.16)	[0.02,0.65]
$\widehat{\underline{\Delta}_2^{TT}}$	-0.43 (0.06)	[-0.53,-0.30]

Standard errors were obtained through bootstrap with 1,000 replications. Effects are measured in standard deviations terms.

TAB. 2.4 – Bounds on Δ^{TT} under different assumptions.

In conclusion, and even if uncertainty is rather important,[43] the conclusion on short term effects of grade retention is rather positive. This result is in line with the results of Jacob & Lefgren (2004) for third graders in Chicago, but more optimistic than theirs on the sixth graders. This difference could reflect the opposition on grade retention decision rules in the two cases. Letting teachers and parents decide on the basis of their observation of the students during the whole year, and not on two tests only as in Chicago, may reduce measurement errors on the ability of children. On the other hand, such a discretionary process is likely to favour or penalize systematically some subpopulations of students, no matter of their ability, and thus decrease the efficiency of grade retention. The results suggest that the former effects overcome the latter.

2.2.6 Conclusion

This paper considers the issue of endogenous selection with instruments. The key assumption for identification, which contrasts with the usual ones in selection problems, is the independence between instruments and selection, conditional on the dependent variables. A general nonparametric identification result is obtained under a completeness condition. This framework can be applied to a broad class of selection models, including Roy models with an unobserved sector, nonignorable nonresponse or binary models with data taken from one response stratum.

[43] This uncertainty is rather due to the endogenous selection on grade retention than on the true effect of the instrument on fifth grade retention. The former effect, which prevents us from recovering the counterfactual progression of retained students, accounts indeed for 55% of the width of the set.

Set identification is also considered when the conditional independence condition fails. Under weaker conditions of monotonicity indeed, there exist sharp and finite bounds on parameters of interest. This result is used to estimate bounds on the effect of grade retention in France.

The paper raises two challenging issues. First, we may wonder whether the ideas developed here could be adapted to generalized Roy model. In these models, selection depends on prediction on the dependent variable rather than on the dependent variable itself. Thus, the conditional independence condition breaks down but the structure of the model may provide information for point or at least set identification. Second, the sharp upper bounds are obtained on a set of parameters which is rather abstract. Further characterizations of this set appear desirable, for both theoretic and practical reasons.

Appendix : proofs

Proposition 2.2.1

let $\mathcal{A}_y = \{u/\psi(y,u) = 1\}$ and $\mathcal{C}_{y,z} = \{u \in \mathbb{R}/\varphi(z,u) = y\}$. We get, for all (y,z),

$$\begin{aligned} P(D=1|Y=y,Z=z) &= P(\eta \in \mathcal{A}_y|Y=y,Z=z) \\ &= P(\eta \in \mathcal{A}_y|\varepsilon \in \mathcal{C}_{y,z}(y,z),Z=z) \\ &= P(\eta \in \mathcal{A}_y) \\ &= P(\eta \in \mathcal{A}_y|Y=y) \\ &= P(D=1|Y=y), \end{aligned}$$

where the third and fourth equalities stem from the condition $\eta \perp\!\!\!\perp (Z,\varepsilon)$. Thus, assumption 3 holds □

Proposition 2.2.2

The proof proceeds in three steps.

1. First, we show that there exists positive c_1, c_2 and $0 < \alpha' < \alpha - 2$ such that

$$c_1 \leq (f_\varepsilon \star f_{\alpha'})(x) \times (1+|x|)^{\alpha'+1} \leq c_2, \tag{2.2.15}$$

where $f_{\alpha'}$ denote the density of an α'-stable distribution of characteristic function $\exp(-|t|^{\alpha'})$ and $\star$ denotes the convolution product.

To prove (2.2.15), note that $f_{\alpha'}$ satisfies, for well chosen $c < C$ (see e.g. Mattner 1992, p. 146),

$$c \leq f_{\alpha'}(x) \times (1+|x|)^{\alpha'+1} \leq C \tag{2.2.16}$$

Let $I = [a,\, b] \subset [-1,\, 1]$ denote an interval such that $\inf_{x\in I} f_\varepsilon(x) = m > 0$ (such an interval exists by the regularity conditions). For all x and $t \in I$,

$$\begin{aligned} 1+|x-t| &\leq 1+\max(|x-a|,|x-b|) \\ &\leq 1+|x|+\max(|a|,|b|) \\ &\leq 2(1+|x|). \end{aligned}$$

Thus,

$$\begin{aligned}(f_\varepsilon \star f_{\alpha'})(x) &\geq \int_I f_\varepsilon(t) f_{\alpha'}(x-t)dt \\ &\geq mc \int_I \frac{dt}{(1+|x-t|)^{\alpha'+1}} \\ &\geq \frac{mc(b-a)}{2^{\alpha'+1}(1+x)^{\alpha'+1}}.\end{aligned}$$

This shows the first inequality of (2.2.15). To prove the second one, remark that by the regularity conditions, there exists M such that

$$(1+|t|)^\alpha f_\varepsilon(t) \leq M \tag{2.2.17}$$

Moreover, for all $x \geq 0$ and $t < x/2$, we get $1+|x-t| \geq (1+x)/2$. Thus, using both (2.2.16) and (2.2.17), we get

$$\begin{aligned}\int_{-\infty}^{x/2} f_\varepsilon(t) f_{\alpha'}(x-t)dt &\leq \frac{2^{\alpha'+1}MC}{(1+x)^{\alpha'+1}} \int_{-\infty}^{x/2} \frac{dt}{(1+|t|)^\alpha} \\ &\leq \frac{2^{\alpha'+1}MC}{(1+x)^{\alpha'+1}} 2\int_{-\infty}^{0} \frac{dt}{(1-t)^\alpha} \\ &\leq \frac{2^{\alpha'+2}MC}{(\alpha-1)(1+|x|)^{\alpha'+1}}.\end{aligned} \tag{2.2.18}$$

Moreover, because $f_{\alpha'}(x-t) \leq C$ and $\alpha - 1 > \alpha' + 1$,

$$\begin{aligned}\int_{x/2}^{+\infty} f_\varepsilon(t) f_{\alpha'}(x-t)dt &\leq MC \int_{x/2}^{+\infty} \frac{dt}{(1+t)^\alpha} \\ &\leq \frac{2^{\alpha-1}MC}{(1+x)^{\alpha-1}} \\ &\leq \frac{2^{\alpha-1}MC}{(1+x)^{\alpha'+1}}.\end{aligned}$$

This, together with (2.2.18), shows that for all $x \geq 0$, there exists a constant C' such that $(f_\varepsilon \star f_{\alpha'})(x) \times (1+|x|)^{\alpha'+1} \leq C'$. The same reasoning can be applied to any $x < 0$, and the second inequality of (2.2.15) follows.

2. Now let us show that for any $g \in \mathcal{B}$ such that $E[g(Y)|Z] = 0$ a.s., we get, almost everywhere (a.e. for short),

$$(g \circ \mu) \star \phi = 0, \tag{2.2.19}$$

where $\phi = f_{-\varepsilon} \star f_{\alpha'}$.

By definition of $\mathcal{B}$, there exists K such that $g(Y) \geq K$ almost surely. Let $\widetilde{g}(u) = g(\mu(u)) - K$. Using the additive decomposition, we get

$$\begin{aligned}\mathbb{E}[g(Y) - K|Z] &= \mathbb{E}[\widetilde{g}(\nu(Z) + \varepsilon)|Z] \\ &= \int \widetilde{g}(\nu(Z) + u) f_{\varepsilon}(u) du \\ &= \int \widetilde{g}(u) f_{-\varepsilon}(\nu(Z) - u) dt\end{aligned}$$

This implies, by the large support assumption, that

$$\mathbb{E}[g(Y)|Z] = 0 \text{ a.s.} \Leftrightarrow \int \widetilde{g}(u) f_{-\varepsilon}(t - u) dt = -K \text{ a.e.} \tag{2.2.20}$$

In other words, $\widetilde{g} \star f_{-\varepsilon} = -K$. Let α' and $f_{\alpha'}$ be defined as previously. We get, a.e.,

$$(\widetilde{g} \star f_{-\varepsilon}) \star f_{\alpha'} = -K.$$

Because $\widetilde{g}, f_{-\varepsilon}$ and $f_{\alpha'}$ are nonnegative functions, we can apply Fubini's theorem, so that $\widetilde{g} \star (f_{-\varepsilon} \star f_{\alpha'}) = -K$ a.e. Equation (2.2.19) follows.

3. Finally, let us prove that the location family generated by ϕ is complete. This proves the result because then, $g \circ \mu = 0$ a.e. and thus $g(Y) = 0$ almost surely. For this purpose, we check the conditions of theorem 1.1 of Mattner (1992). First, ϕ satisfies condition (i) of this theorem by (2.2.15) and proposition 1.2 of Mattner (1992). Second, the characteristic function Ψ_{ϕ} corresponding to the density ϕ writes as

$$\Psi_{\phi}(t) = \Psi_{\varepsilon}(-t) \times \exp(-|t|^{\alpha'}) \tag{2.2.21}$$

where Ψ_{ε} denotes the characteristic function of ε. Thus, by the regularity conditions, Ψ_{ϕ} is infinitely differentiable on $\mathbb{R} \backslash (A \cup \{0\})$ and condition (ii) of Mattner's theorem holds. Lastly, by (2.2.21) and the regularity conditions once more, Ψ_{ϕ} does not vanish anywhere. Thus theorem 1.1 in Mattner (1992) can be applied, and the proof is finished $\square$

Theorem 2.2.3

By assumption 3 and the definition of P,

$$\begin{aligned} P(D=1|Z)E\left[\frac{1}{P(Y)}|D=1,Z\right] &= E\left(\frac{D}{P(Y)}\middle|Z\right) \\ &= E\left(\frac{E(D|Y,Z)}{P(Y)}\middle|Z\right) \\ &= E\left(\frac{E(D|Y)}{P(Y)}\middle|Z\right). \end{aligned}$$

Hence,

$$E\left(\frac{D}{P(Y)}-1\middle|Z\right)=0 \tag{2.2.22}$$

By assumption 2, $P(D=1|Z)$ can be identified from the data. Thus, for any function R, $E[D/R(Y)-1|Z]$ can be computed from the data. Hence, any candidate for P must satisfy equality (2.2.22). Now let Q be such a candidate and let $g=P/Q-1$. g is bounded below by -1. Moreover, Q must satisfy $E[D/Q(Y)]=1$, which can also be written as $E[P(Y)/Q(Y)]=1$. This implies that

$$E\left[|g(Y|\right] \leq E\left[\frac{P(Y)}{Q(Y)}\right]+1<\infty.$$

Hence, $g\in\mathcal{B}$. Moreover,

$$\begin{aligned} 0 &= E\left(\frac{D}{Q(Y)}-1\middle|Z\right) \\ &= E\left(\frac{P(Y)}{Q(Y)}-1\middle|Z\right) \\ &= E\left(g(Y)|Z\right). \end{aligned} \tag{2.2.23}$$

This together with assumption 4 imply that $g(Y)=0$ a.s., so that $Q(Y)=P(Y)$ a.s. Thus, P is identified.

To finish the proof, let $f_{D,Y,Z}$ denote the density of (D,Y,Z) with respect to an appropriate measure. $f_{D,Y,Z}(1,y,z)$ is identified by $f_{Y,Z|D=1}(y,z)P(D=1)$. Moreover, by assumption 3,

$$\begin{aligned} P(y) &= P(D=1|Y=y,Z=z) \\ &= \frac{f_{D,Y,Z}(1,y,z)}{f_{Y,Z}(y,z)} \end{aligned}$$

Similarly,

$$1-P(y)=\frac{f_{D,Y,Z}(0,y,z)}{f_{Y,Z}(y,z)}$$

Thus,

$$f_{D,Y,Z}(0,y,z) = \left[\frac{1-P(y)}{P(y)}\right] f_{D,Y,Z}(1,y,z).$$

Hence, the joint distribution of the data is identified □

Theorem 2.2.4

Part "if" of the theorem is trivial. To prove the "only if" implication, let us consider a solution Q which belongs to $]0,1]$. Define also a function $g_{D,Y,Z}$ by

$$g_{D,Y,Z}(d,y,z) = \left[\frac{1-Q(y)}{Q(y)}\right]^{1-d} f_{Y,Z|D=1}(y,z)P(D=1).$$

$g_{D,Y,Z}$ is a density (with respect to a convenient measure λ), as it is nonnegative and integrates to one. Indeed,

$$\begin{aligned}
& \int \left[g_{D,Y,Z}(0,y,z) + g_{D,Y,Z}(1,y,z)\right] d\lambda(y,z) \\
= & \int \frac{f_{Y,Z|D=1}(y,z)P(D=1)}{Q(y)} d\lambda(y,z) \\
= & E\left\{E\left[\frac{E(D|Y,Z)}{Q(Y)}|Z\right]\right\} \\
= & 1.
\end{aligned}$$

Moreover,

$$g_{D,Y,Z}(1,y,z) = f_{Y,Z|D=1}(y,z)P(D=1) \tag{2.2.24}$$

and

$$\begin{aligned}
g_Z(z) &= f_Z(z) \int \frac{f_{Y,Z|D=1}(y,z))P(D=1)}{Q(y)f_Z(z)} dy \\
&= f_Z(z) E\left[\frac{E(D|Y,Z)}{Q(Y)}|Z\right] \\
&= f_Z(z).
\end{aligned}$$

This last equality, together with (2.2.24), ensures that $g_{D,Z}(d,z) = f_{D,Z}(d,z)$. Thus, $g_{D,Y,Z}$ is coherent with the observed data. Lastly, because $g_Y(y) = f_{Y|D=1}P(D=1)/Q(y)$, we get after straightforward manipulations :

$$\begin{aligned}
g_{D,Z|Y}(1,z,y) &= Q(y) f_{Z|Y,D=1}(z,y), \\
g_{D,Z|Y}(0,z,y) &= (1-Q(y)) f_{Z|Y,D=1}(z,y).
\end{aligned}$$

In other words, the corresponding distribution of (D,Y,Z) satisfies the independence condition of assumption 3. To conclude, if there exists a solution Q to

equation (2.2.4) which lies in $]0,1]$, one can rationalize the observed data by a distribution which satisfies the independence condition □

Theorem 2.2.5

The result uses the following standard result, which is proved for the sake of completeness.

Lemma 2.2.1 *Let T denote a real random variable and $(h_1, h_2) \in (L_T^2)^2$ be increasing functions. Then cov$(h_1(T), h_2(T)) \geq 0$.*

Proof : let (T_1, T_2) denote two independent copies of T. Then, because both h_1 and h_2 are increasing,

$$(h_1(T_1) - h_1(T_2)) \times (h_2(T_1) - h_2(T_2)) \geq 0.$$

Thus, taking expectation and using the fact that (T_1, T_2) are i.i.d, we get

$$2\left\{E\left[h_1(T)h_2(T)\right] - E\left[h_1(T)\right]E\left[h_2(T)\right]\right\} \geq 0.$$

The result follows □

a) By lemma 2.2.1 and assumption 6,

$$\text{cov}(h(Y), P(D=1|Y,\widetilde{Z})|\widetilde{Z}) \geq 0.$$

Thus,

$$E(h(Y)|\widetilde{Z})P(D=1|\widetilde{Z}) \leq E\left(h(Y)D|\widetilde{Z}\right).$$

This implies that

$$E(h(Y)|\widetilde{Z}) \leq E\left(h(Y)|D=1,\widetilde{Z}\right).$$

Hence, by integration,

$$E[h(Y)] \leq E\left[E\left(h(Y)|D=1,\widetilde{Z}\right)\right].$$

Moreover, this upper bound is sharp because the two terms are identical under the untestable assumption that $D \perp\!\!\!\perp Y|\widetilde{Z}$.

b) Let $h \in H_{YZ}$ and $\widetilde{h} \in H_Z$ be such that $h(Y) = E[\widetilde{h}(Z)|D = 1, Y]$. We get

$$\begin{aligned} E\left[\frac{Dh(Y)}{Q(Y)}\right] - E[h(Y)] &= E\left[\frac{DE(\widetilde{h}(Z)|D=1,Y)}{Q(Y)}\right] - E[h(Y)] \\ &= E\left[\frac{D\widetilde{h}(Z)}{Q(Y)}\right] - E[h(Y)] \\ &= E\left[\widetilde{h}(Z)E\left(\frac{D}{Q(Y)}|Z\right)\right] - E[h(Y)] \\ &= E\left[\widetilde{h}(Z)\right] - E[h(Y)] \\ &= E\left[E\left(\widetilde{h}(Z)|Y\right) - E\left(\widetilde{h}(Z)|D=1,Y\right)\right]. \end{aligned}$$

Now, because $\widetilde{h}$ and $z \mapsto P(D = 1|Y, Z = z)$ are increasing with probability one, we have, similarly to a),

$$E(\widetilde{h}(Z)|D = 1, Y) \geq E(\widetilde{h}(Z)|Y). \tag{2.2.25}$$

Thus,

$$E\left[\frac{Dh(Y)}{Q(Y)}\right] \leq E[h(Y)]. \tag{2.2.26}$$

Moreover, by theorem 2.2.4, if there exists a solution Q to equation (2.2.4) which lies in $]0, 1]$, one cannot reject that (2.2.25) and (2.2.26) are actually equalities. This implies that $E[Dh(Y)/Q(Y)]$ is a sharp lower bound of $E[h(Y)]$.

c) If $D \perp\!\!\!\perp (Y, \widetilde{Z})$, by independence,

$$E\left[E(h(Y)|D = 1, \widetilde{Z})\right] = E\left[E(h(Y)|Z)\right] = E\left[h(Y)\right].$$

Moreover, because $P(D = 1)$ is a solution to (2.2.4), $Q(Y) = P(D = 1)$, so that

$$E\left[\frac{Dh(Y)}{Q(Y)}\right] = E(h(Y)).$$

Now, if $Y = \widetilde{Z}$,

$$E\left[E(h(Y)|D = 1, \widetilde{Z})\right] = E\left[h(Y)\right].$$

Moreover, because $Y = Z$, equation (2.2.4) is equivalent to $Q(Y) = P(D = 1|Y)$. Hence,

$$E\left[\frac{Dh(Y)}{Q(Y)}\right] = E\left[\frac{E(D|Y)h(Y)}{Q(Y)}\right] = E\left[h(Y)\right] \ \square$$

Proposition 2.2.6

Let $\widetilde{h}$ denote an increasing function. We have

$$E(\widetilde{h}(Z)|D=1,Y)=\int \widetilde{h}(z)dF_{Z|Y,D=1}(z).$$

Because $\widetilde{h}$ is increasing, there exists a positive measure μ such that for all $z \leq z_1$,

$$\widetilde{h}(z_1)-\widetilde{h}(z)=\int_z^{z_1} d\mu(u).$$

Thus, for all y and all $M \in \mathbb{R}$,

$$\begin{aligned} E(\widetilde{h}(Z)|D=1,Y=y) &= \int_M^{\infty}\int_M^{z} d\mu(u)dF_{Z|Y=y,D=1}(z)-\int_{-\infty}^{M}\int_z^{M} d\mu(u)dF_{Z|Y=y,D=1}(z) \\ &\quad +2\widetilde{h}(M). \end{aligned}$$

Hence, by Fubini's theorem on nonnegative functions,

$$E(\widetilde{h}(Z)|D=1,Y=y)=\int_M^{\infty}(1-F_{Z|Y=y,D=1}(u))d\mu(u)-\int_{-\infty}^{M}F_{Z|Y=y,D=1}(u)d\mu(u)+2\widetilde{h}(M).$$

Consequently, we get, for all $y \leq y_1$,

$$\begin{aligned} & E(\widetilde{h}(Z)|D=1,Y=y_1)-E(\widetilde{h}(Z)|D=1,Y=y) \\ = & \int [F_{Z|D=1,Y=y}(u)-F_{Z|D=1,Y=y_1}(u)]d\mu(u). \end{aligned}$$

By assumption, the right hand side is nonnegative. The result follows.

Theorem 2.2.7

a) β_0 satisfies

$$E\left(\frac{DW}{F(V'\beta_0)}\right)=E\left(\frac{W}{F(V'\beta_0)}E(D|Z,V)\right)=E\left(\frac{W}{F(V'\beta_0)}E(D|V)\right)=E(W),$$

where the second equality stems from assumption 3. Local identification only requires that the differential of $\beta \to E(DW/F(V'\beta))$ is full rank at $\beta=\beta_0$. This differential is $-E(DWV'F'(V'\beta_0)/F^2(V'\beta_0))$, so the result follows from assumption 4'.

b) Suppose that there exists β such that

$$E\left(\frac{DW}{F(V'\beta)}\right)=E(W)=E\left(\frac{DW}{F(V'\beta_0)}\right). \tag{2.2.27}$$

Then

$$E\left(\left(\frac{1}{F(V'\beta_0)}-\frac{1}{F(V'\beta)}\right)W(\beta_0-\beta)\middle|D=1\right)=0.$$

Thus

$$E\left(\left(\frac{1}{F(V'\beta_0)}-\frac{1}{F(V'\beta)}\right)E(W|V,D=1)(\beta_0-\beta)\middle|D=1\right)=0.$$

Now, by assumption 4",

$$E(W|V,D=1)=\begin{pmatrix} I_r & 0 \\ \Gamma_1 & \Gamma_2 \end{pmatrix}\begin{pmatrix} X \\ Y \end{pmatrix}\equiv \Gamma V,$$

where I_r is the identity matrix of size r =dim(X). Moreover, because Γ_2 is full rank, Γ is also full rank. Hence

$$E\left(\left(\frac{1}{F(V'\beta_0)}-\frac{1}{F(V'\beta)}\right)(V'\beta_0-V'\beta)\middle|D=1\right)=0.$$

Because F is strictly increasing, for any $x\neq y$, $(x-y)(1/F(x)-1/F(y))<0$, so that

$$V'\beta_0=V'\beta \quad P^{D=1}\text{ a.s.}$$

Because V is linearly independent almost surely, $\beta=\beta_0$ □

Theorem 2.2.8.

As Horowitz & Lee (2007), we adapt the proof of theorem 2 of Bissantz et al. (2004). By definition of $\widehat{f}$,

$$\max\left(\left\|\widehat{T}\,\widehat{f}-1\right\|^2,\alpha_n\left\|\widehat{f}\right\|^2\right)\leq\left\|\widehat{T}\,\widehat{f}-1\right\|^2+\alpha_n\left\|\widehat{f}\right\|^2\leq\left\|\widehat{T}\,f-1\right\|^2+\alpha_n\|f\|^2 \tag{2.2.28}$$

Let $\delta_n=h_n^2+1/nh_n$. Because $E(\left\|\widehat{T}\,f-1\right\|^2)=O(\delta_n)$ (see e.g. Gyorfi et al., 2002) and $\delta_n/\alpha_n\to 0$, we get

$$\limsup E(\left\|\widehat{f}\right\|^2)\leq\|f\|.$$

Inequalities (2.2.28) and $\delta_n/\alpha_n\to 0$ also implies that $E(\left\|\widehat{T}\,\widehat{f}-1\right\|^2)\to 0$. Besides, D_M is weakly closed as a closed and convex set (see Bissantz et al., 2004). Moreover, for all $\phi\in D_M$, by Jensen's inequality,

$$(T\,\phi)^2\leq E(\phi(Y)^2\,|Z).$$

Hence,

$$\begin{aligned}\|T\,\phi\|^2 &\leq \int\left[\int \phi(y)^2 f_{Y|Z}(y|z)dy\right]dz \\ &\leq \int \phi(y)^2\left[\int \frac{f_{Z|Y}(z|y)f_Y(y)}{f_Z(z)}dz\right]dy \\ &\leq \frac{\sup_{y\in[0,1]} f_Y(y)}{\inf_{z\in[0,1]} f_Z(z)}\int \phi(y)^2 dy,\end{aligned}$$

where the second inequality follows from Fubini's theorem and Bayes' theorem. Hence, by assumption 8-(b), there exists $A < +\infty$ such that

$$\|T\,\phi\|^2 \leq A\,\|\phi\|^2\,.$$

This inequality and the linearity of T proves that it is continuous. Hence, T is weakly continuous. This and the fact that D_M is weakly closed ensures that T is weakly sequentially closed (see Bissantz et al., 2004). Consequently, we can apply the end of the proof of theorem 2 of Bissantz et al. (2004), and the result follows

□

Corollary 2.2.9

By the triangular inequality,

$$|\widehat{\theta}-\theta| \leq \frac{1}{n}\sum_{i=1}^{n} D_i|g(Y_i^*,Z_i)||\widehat{f}^{-i}(Y_i^*)-f(Y_i^*)| + \left|\frac{1}{n}\sum_{i=1}^{n} D_i g(Y_i^*,Z_i)f(Y_i^*)-\theta\right| \tag{2.2.29}$$

By assumption 8, $|D_i f(Y_i^*)| \leq M$. Hence, $E[|D_i g(Y_i^*,Z_i)f(Y_i^*)|^2] < \infty$ and by the weak law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n} D_i g(Y_i^*,Z_i)f(Y_i^*) \xrightarrow{L^2} E(Dg(Y^*,Z)f(Y^*)).$$

Moreover,

$$\begin{aligned}E(Dg(Y^*,Z)f(Y^*)) &= E(Dg(Y,Z)f(Y)) \\ &= E(E(D|Y,Z)g(Y,Z)f(Y)) \\ &= \theta.\end{aligned}$$

Thus the second term of the r.h.s. of (2.2.29) tends to zero in quadratic mean.

Now, because the $(\widehat{f}^{-i}(Y_i^*))_i$ are identically distributed, the first term T_1 of the r.h.s. of (2.2.29) satisfies

$$\begin{aligned} E\left(|T_1|\right) &= E\left(D_1|g(Y_1^*, Z_1)||\widehat{f}^{-1}(Y_1^*) - f(Y_1^*)|\right) \\ &\leq \sqrt{E\left(|g(Y_1^*, Z_1)|^2\right) E\left(|\widehat{f}^{-1}(Y_1^*) - f(Y_1^*)|^2\right)}, \end{aligned}$$

by the Cauchy-Schwartz inequality. Now, by independence between Y_1^* and $\widehat{f}^{-1}$,

$$E(|\widehat{f}^{-1}(Y_1^*) - f(Y_1^*)|^2) \leq \sup_{y\in[0,1]} f_Y(y) E\left(||\widehat{f}^{-1} - f||^2\right).$$

Thus, the left hand side tends to zero by theorem 2.2.8. As a consequence, $E(|T_1|)$ also tends to zero. This yields the announced result □

Chapitre 3

Identification of two asymmetric information models

3.1 Nonparametric Identification of Common Value Auctions Models

3.1.1 Introduction

Structural econometric approaches have been successfully applied during the last decade to study auction data. The aim of such analyzes is to recover the structural parameters of a theoretical model from the data using econometric methods. In the case of auctions, the econometrician is interested in estimating the distributions of the value of the good for each participant from the observed bids. It relies on the equilibrium that defines how bids depend on these distributions.

Previous studies mostly focused on the private value paradigm (PV) (Laffont et al., 1995, Donald & Paarsch, 1996, Elyakime et al., 1994, 1997, Guerre et al., 2000). In these models, each bidder knows his own private value for the auctioned good but does not know others' valuations.
The "opposite" case is known as the common value paradigm (CV). In this model, the value of the auctioned good is unknown but the same for each bidder and each participant receives a signal correlated with this value. It turns out that the econometrics of CV models is more complicated than IPV models. The main reason behind these difficulties comes from the nonparametric identification of

these models from observed bids (Laffont & Vuong, 1996, Athey & Haile, 2002). As a consequence, one has to impose some further restrictions to obtain identification results. Paarsch (1992) proposes a parametric approach, whereas Li et al. (2000) develop a semiparametric one. In their paper, the authors assume a multiplicative decomposition of the signals into a common component (the value of the good) and an indiosyncratic component (a specific signal) for each bidder. Adding some further restrictions, Li et al. show that the CV model is identifiable and propose a two-step nonparametric procedure to estimate the densities of both components. Recently Février (2007) proposed an alternative nonparametric approach for a particular class of signals' distributions.

In this paper, we analyze the econometrics of CV models under the assumption that the supports of the distribution functions of the signals conditional on the value V of the good are bounded and vary with V. We first study the Mineral Rights Model (or pure common value model) in which the signals are drawn independently conditional on V. We then prove that our results are robust to any dependence structure of the signals. An important feature of our conclusion is that no restriction on the link between the signals and the value are needed.

Our main identification theorem is that these CV models are nonparametrically identified. Such a result obtains by exploiting all the information contained in the observed data. The intuition is the following. The values V of the good that are coherent with a given signal are bounded from above, which in turn defines an upper bound for the signals received by other bidders. Hence, by observing two signals that are the furthest apart, the econometrician infers the underlying value of the good in this auction. Using this idea, we show that it is possible to recover the conditional distributions of the signals up to a transformation. Then, the joint distribution of the signals and the value of the good is identified, up to the same transformation, using the integral equation that relates it with the marginal distribution of the signals. Lastly, the integral equation given by the equilibrium condition allows us to identify the transformation itself.

The paper is organized as follow. Section 2 presents the Common Value Model. In Section 3, we analyze its nonparametric identification. Section 4 is devoted to our estimation procedure. Monte Carlo simulations are presented in Section 5. Section 6 concludes.

3.1.2 The Common Value Model

In the Common Value Model (Milgrom & Weber, 1982), a single and indivisible good is auctioned to n bidders. The value V of the good, unknown to the bidders, is distributed following a distribution function $F_V(.)$ (and a density function $f_V(.)$) on the support $[\underline{V}, \overline{V}]$ with $(\underline{V}, \overline{V}) \in \mathbb{R}^{+^2}$. Each bidder i receives a private signal S_i. We note $F_{S_1,...,S_n|V}(.,...,.|.)$ (resp. $f_{S_1,...,S_n|V}(.,...,.|.)$) the distribution function (resp. the density function) of the signals given V. In the general setting, the signals can be correlated conditional on V, but we impose symmetry by supposing that they are exchangeable. We also make the hypothesis that the densities are continuously differentiable on their support. These supports may vary with V and are denoted $[\underline{S}(V), \overline{S}(V)]^n$ with $(\underline{S}(V), \overline{S}(V)) \in \mathbb{R}^{+^2}$. Each player knows his private signal as well as the distribution functions. He does not know however the private signals of other bidders. Finally, we also impose that $(V, S_1, ..., S_n)$ are affiliated.[1].

A special case, known as the Mineral Rights Model (Rothkopf, 1969 ; Wilson, 1977), is obtained when the signals are conditionally independent given the common value V. In such a case, we will note $F_{S|V}(.|.)$ (resp. $f_{S|V}(.|.)$) the distribution function (resp. the density function) of the signals given V. The affiliation property is obtained by imposing that $f_{S|V}(.|.)$ satisfies the monotone likelihood ratio property.[2]

We consider a first price auction. Each bidder submits a bid and the winner is the one who submits the highest bid. He obtains the object and pay his bid.

A strategy for a player i is a function $b_i(.)$ that associates to each signal S_i the amount $b_i(S_i)$ that player i wants to bid. As shown by Milgrom and Weber (1982), a symmetric equilibrium exists in first price common value auctions. To describe this equilibrium, it is useful to introduce the following functions. We note $Y_i = \max_{j \neq i} S_j$ and $F_{Y_i|S_i}(.|.)$ (resp. $f_{Y_i|S_i}(.|.)$) its distribution function (resp. density function) conditional on the signal S_i of player i. We also introduce the function $V(s, y) = E[V|S_i = s, Y_i = y]$ that is the expected value of the good conditional on the signal S_i of player i and the highest signal Y_i of other players.

[1] See Milgrom and Weber (1982) for an extensive discussion of the importance of this concept in auctions

[2] The density $f_{S|V}$ has the monotone likelihood ratio property if for all $s' > s$ and $v' > v$, $f_{S|V}(s|v)/f_{S|V}(s|v') \geq f_{S|V}(s'|v)/f_{S|V}(s'|v')$.

Proposition 3.1.1 *Milgrom-Weber (1982). In a common value first price auction, a symmetric equilibrium strategy exists and is given by :*

$$b(s) = V(s,s) - \int_{\underline{S}(\underline{V})}^{s} L(\alpha|s) dV(\alpha,\alpha)$$

where $L(\alpha|s) = exp[-\int_{\alpha}^{s} f_{Y_i|S_i}(u|u)/F_{Y_i|S_i}(u|u)du]$.

3.1.3 Nonparametric Identification

In this section, we turn to the key issue of identification. This amounts to determine whether the structural elements of the model, i.e. the joint distribution of $(S_1, ..., S_n, V)$, are uniquely defined by the observations and the restrictions of the model. Of course, the nature of the observations matters and we suppose in this paper that all bids are available for all auctions.

As explained in Laffont & Vuong (1996) (see also Athey and Haile, 2002), the common value model is not nonparametrically identified. Indeed, one has to remark that, from an economic point of view, the scaling of the signals is arbitrary. The reason is that the common value model is intrinsically "not well" defined. If v is the value of the good, a model in which the signals s are distributed uniformly on $[V-1, V+1]$ is exactly the same than the model in which the signals $s' = 2s$ are distributed uniformly on $[2V-2, 2V+2]$. The same amount of information is contained in s and s' about the value v of the good. More generally, any monotonic transformation of the bids $s' = g(s)$ leads to the "same" model. For this reason (see Athey and Haile, 2002), it is possible to normalize the model and a convenient normalization is given in the following assumption.

Assumption 11 *(normalization)* $b(s) = s$.

As a consequence, the observation of the bids allow the econometrician to identify the joint distribution of the signals. Unfortunately, the equilibrium condition still does not allow to recover the joint distribution of the signals and the value. Indeed, the first order condition depends only on the moments $V(s,s)$, which are insufficient to identify the entire joint distribution. Hence, the model remains unidentified.

One can note however that the monotone likelihood ratio property imposes restrictions on this joint distribution. It implies in particular that $\underline{S}(.)$ and $\overline{S}(.)$ are increasing functions. Our key assumption to recover nonparametric identification is to strengthen this condition into a strict monotonicity.

Assumption 12 *(Strict monotonicity of the support functions)* $\underline{S}(.)$ *and* $\overline{S}(.)$ *are differentiable functions such that for all* $v \in [\underline{V}, \overline{V}]$, $\underline{S}'(v) > 0$ *and* $\overline{S}'(v) > 0$.

Assumption 12 states that the support of S conditional on $V = v$ varies strictly with v. It implies that the support of the signals conditional on the value is compact. This assumption is common in the empirical auctions literature. The main consequence of assumption 12 is that the observation of the bids provide bounds on the possible values of V. Actually, we will show that the bounds can be made arbitrarily close by focusing on the auctions such that the difference between the smallest and the largest bids is maximal. Then the variations of the other bids provide identification of the structural functions. Hence, the method requires at least three auctionneers.

Assumption 13 *(Minimal number of auctioneers)* $n \geq 3$.

We show in the next subsection how these assumptions enables to recover the joint distribution of $(S_1, ..., S_n, V)$ in the particular case of the Mineral Rights Model where $(S_1, ..., S_n)$ are i.i.d. conditional on V. We then prove that the result is actually robust to any form of dependence, and discuss some extensions of our main result.

Identification of the Mineral Rights Model

We focus here on the Mineral Rights Model, which is the most usual restriction of the common value model (see e.g. Li et al., 2000 or Février, 2006).

Assumption 14 *(Mineral Rights Model)* $(S_1, ..., S_n)$ *are i.i.d. conditional on* V.

The proof of identification will proceed in four steps. We first show that the data enables to recover $\overline{S} \circ \underline{S}^{-1}(.)$ and $\underline{S}(\overline{V})$. We then show that the distribution of S given $W = \underline{S}(V)$ is identified. Then we prove the identification of the distribution of W. Lastly, the identification of $\underline{S}^{-1}(.)$ is established.

Identification of $\overline{S} \circ \underline{S}^{-1}(.)$ *and* $\underline{S}(\overline{V})$

Before identifying the conditional distribution of the signals, it will be convenient to identify the support of S conditional on $W = w$ and the support of W itself. These two supports depend respectively on $\overline{S} \circ \underline{S}^{-1}(.)$ and $\underline{S}(\overline{V})$.Suppose that bidder 1 obtains a signal $S_1 = s \leq \underline{S}(\overline{V})$. The value of V compatible with this signal has to be smaller than $\underline{S}^{-1}(s)$ because $S_1 \geq \underline{S}(V)$.[3] As a consequence, the signal S_2 of bidder 2 has to be smaller than $\overline{S} \circ \underline{S}^{-1}(s)$. Moreover, this value can be reached by letting $V = \underline{S}^{-1}(s)$. Hence, $\overline{S} \circ \underline{S}^{-1}(s)$ is the maximum of the support of S_2 conditional on $S_1 = s$, for all $s \leq \underline{S}(\overline{V})$.

Now, for all $s \geq \underline{S}(\overline{V})$, this maximum is $\overline{S}$. Indeed, $V = \overline{V}$ can be rationalized by the data in this case. Both situations are depicted in figure 3.1.

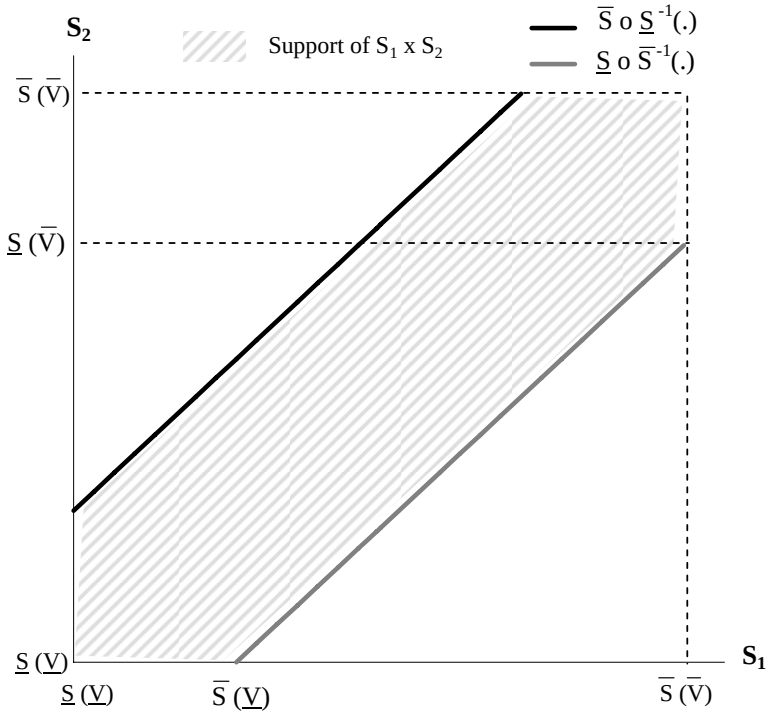

FIG. 3.1 – Support of $S_1 \times S_2$, $\overline{S} \circ \underline{S}^{-1}(.)$ and $\underline{S}(\overline{V})$.

Because the maximum of S_2 given S_1 is observed, $\overline{S} \circ \underline{S}^{-1}(.)$ is identified as soon as $\underline{S}(\overline{V})$ is identified. Similarly, the function $\underline{S} \circ \overline{S}^{-1}(.)$ is identified using the minimum of S_2 conditional on S_1. In particular, $\underline{S} \circ \overline{S}^{-1}(\overline{S}(\overline{V})) = \underline{S}(\overline{V})$ is identified as the

[3]Note that $\underline{S}^{-1}(s)$ is well defined for all $s \in [\underline{S}(\underline{V}), \underline{S}(\overline{V})]$.

minimum of S_2 given that S_1 takes the largest possible value.[4]

Identification of the distribution of S conditional on W

We now turn to the identification of the distribution of S conditional on W. As mentioned previously, the idea here is that the conditional on distribution function of S_3 given (S_1, S_2) will provide information on the conditional distribution of S. More precisely, let $\mathbf{w} = (w, \overline{S} \circ \underline{S}^{-1}(w))$ and suppose that $(S_1, S_2) = \mathbf{w}$. As previously explained, the value of S_1 implies that $V \leq \underline{S}^{-1}(w)$ and the one of S_2 entails $V \geq \underline{S}^{-1}(w)$. As a consequence, the only way to rationalize the data is that $W = \underline{S}(V) = w$. Thus, informally, the distribution of S_3 conditional on $(S_1, S_2) = \mathbf{w}$ is equal to the distribution of S_3 conditional on $W = w$.

This statement is not rigorous, however, because the distribution of S_3 conditional on $(S_1, S_2) = \mathbf{w}$ is not defined. Indeed, the joint density of (S_1, S_2) at this point is zero.[5] However, we can use a continuity argument by conditioning on the fact that (S_1, S_2) belongs to a neighborhood of $\mathbf{w}$. This approach is valid as soon as the distribution of S conditional on $W = w$ is a continuous function of w.

Assumption 15 *(continuity of $F_{S|W}$ in W) For all s, the function $w \mapsto F_{S|W}(s|w)$ is continuous on $[\underline{S}(\underline{V}), \underline{S}(\overline{V})]$.*

Continuity in W is a weak restriction, which is satisfied by all usual conditional distribution. It especially includes, in the case of common value auctions, the multiplicative model of Li and Vuong (2000), $S = V \times \varepsilon$ and the model of Février (2006).

For all $\varepsilon > 0$ and $w \in [\underline{S}(\underline{V}), \underline{S}(\overline{V})]$, let $\underline{w}_\varepsilon = \max(w - \varepsilon, \underline{S}(\underline{V}))$ and $\overline{w}_\varepsilon = \min(w + \varepsilon, \underline{S}(\overline{V}))$. We define the set $A_\varepsilon(w)$ by

$$A_\varepsilon(w) = [\underline{w}_\varepsilon; \overline{w}_\varepsilon] \times \left[\overline{S} \circ \underline{S}^{-1}(\underline{w}_\varepsilon); \overline{S} \circ \underline{S}^{-1}(\overline{w}_\varepsilon)\right].$$

[4] Because the maximum of the support of S_2 conditional on S_1 is strictly increasing for all $s \leq \underline{S}(\overline{V})$ and flat after, $\underline{S}(\overline{V})$ is also characterized by

$$\underline{S}(\overline{V}) = \inf\{s : \max(\text{Support}(S_2|S_1 = s)) = \overline{S}(\overline{V})\}.$$

[5] To see this, note that

$$f_{S_1,S_2}(s, \overline{S} \circ \underline{S}^{-1}(s)) = \int_s^s f_{S_1|W}(s|w) f_{S_1|W}(\overline{S} \circ \underline{S}^{-1}(s)|w) f_W(w) dw = 0.$$

Then, when $(S_1, S_2) \in A_\varepsilon(w)$, W belongs to the interval $[\underline{w}_\varepsilon,\ \overline{w}_\varepsilon]$. For all $\delta > 0$, assumption 15 ensures the existence of $\varepsilon > 0$ such that $|F_{S_3|W}(s|u) - F_{S_3|W}(s|w)| < \delta$ for all u such that $|u - w| < \varepsilon$. Hence,

$$\begin{aligned} & \left|F_{S_3|(S_1,S_2)\in A_\varepsilon(w)}(s) - F_{S_3|W}(s|w)\right| \\ = & \left|\int_{\underline{w}_\varepsilon}^{\overline{w}_\varepsilon} \left(F_{S|W}(s|u) - F_{S|W}(s|w)\right) f_{W|(S_1,S_2)\in A_\varepsilon(w)}(u)du\right| \\ \leq & \int_{\underline{w}_\varepsilon}^{\overline{w}_\varepsilon} \left|F_{S|W}(s|u) - F_{S|W}(s|w)\right| f_{W|(S_1,S_2)\in A_\varepsilon(w)}(u)du \\ < & \ \delta, \end{aligned}$$

where the second line stems from the independence between S_1 and (S_2, S_3) conditional on W. Hence, for all $w \in [\underline{S}(\underline{V}), \underline{S}(\overline{V})]$ and all $s \in [w, \overline{S} \circ \underline{S}^{-1} w]$,

$$\lim_{\varepsilon \to 0} F_{S_3|(S_1,S_2)\in A_\varepsilon(w)}(s) = F_{S|W}(s|w).$$

As a consequence, the distribution of S conditional on W is identified under assumptions 11 to 15. Assumption 12 is crucial for this result. The bounds are used to identify the unique value W that is compatible with two extreme signals. A third signal is then required to identify the distribution itself.

Identification of the distribution of W

We now turn to the identification of the distribution of W. Because $F_{S|W}$ is identified, the density[6] f_W is a solution of the integral equation

$$F_{S_1,...,S_n}(s_1, ..., s_n) = \int \prod_{i=1}^{n} F_{S|W}(s_i|u) f_W(u)du. \tag{3.1.1}$$

This equation should overidentify f_W in general. Indeed, the unknown function is a mapping of one variable and one relies on an equation of n variables to recover it. However, to the best of our knowledge, there is no general result on such equations which garantees the uniqueness of the solution. To secure identification here, we exploit once more the variation of the support of W with $(S_1, ..., S_n)$. We rely for that purpose on the following assumption.

Assumption 16 *(positive densities)*
– $\forall w \in [\underline{S}(\underline{V}), \underline{S}(\overline{V})]$, $\forall s \in]w, \overline{S} \circ \underline{S}^{-1}(w)[$, $f_{S|W}(s|w) > 0$.

[6]The existence of the density of W is ensured by assumption 12.

– $\forall w \in]\underline{S}(\underline{V}), \underline{S}(\overline{V})[,\ f_W(w) > 0.$
– *There exists* $k \in \mathbb{N}$ *such that for all* $w \in [\underline{S}(\underline{V}), \underline{S}(\overline{V})]$, $\frac{\partial^k f_{S|W}}{\partial s^k}(w|w) \neq 0$ *or* $\frac{\partial^k f_{S|W}}{\partial s^k}(\overline{S} \circ \underline{S}^{-1}(w)|w) \neq 0$.

The main restriction of condition 16 is the existence of a k-th derivative of $f_{S|W}$ which is strictly positive at one of its boundary. This assumption is satisfied, for instance, by all the beta distributions, or more generally by all distribution whose density is polynomial. The following proposition states the desired result. Its proof is deferred in appendix.

Proposition 3.1.2 *Suppose that assumptions 11-16 hold. Then the distribution of* W *is identified.*

Identification of $\underline{S}^{-1}$

Using support variations, we have shown up to now that the data alone enable to identify the joint distribution of $(S_1, ..., S_n, W)$, where $W = \underline{S}(V)$. However, we cannot identify $\underline{S}(.)$, and thus the distribution of $(S_1, ..., S_n, V)$, without further condition. Indeed, no restriction on V arises from the data alone, so that any strictly increasing transformation would be possible. On the other hand, the theoretical equilibrium does provide information on V. As stated in proposition 3.1.1, the equilibrium strategy is

$$b(s) = V(s, s) - \int_{\underline{S}(\underline{V})}^{s} L(\alpha|s) dV(\alpha, \alpha)$$

Taking the derivative of this equation and using $b(s) = s$, we obtain :

$$V(s, s) = s + \frac{F_{Y|S}(s|s)}{f_{Y|S}(s|s)}$$

The right part of this equation is observed in the data, so that $V(s, s)$ is identified. Furthermore, by definition,

$$V(s, s) = \frac{\int_{\overline{S}^{-1}(\max(s, \underline{V}))}^{\min(\underline{S}^{-1}(s), \overline{V})} v(n-1) f_{S|V}^2(s|v) F_{S|V}^{n-2}(s|v) f_V(v) dv}{f_{Y,S}(s, s)}$$

After a change in variables, this equation becomes

$$s f_{Y,S}(s, s) + f_S(s) F_{Y|S}(s|s) =$$
$$\int_{\underline{S} \circ \overline{S}^{-1}(\max(s, \overline{S}(\underline{V})))}^{\min(s, \underline{S}(\overline{V}))} \underline{S}^{-1}(w)(n-1) f_{S|W}^2(s|w) F_{S|W}^{n-2}(s|w) f_W(w) dw$$

Hence, for $s \in [\overline{S}(\underline{V}), \overline{S}(\overline{V})]$, one obtains an equation of the form :

$$\int_{l(s)}^{\min(s,\underline{S}(\overline{V}))} \underline{S}^{-1}(u)h(s,u)du = k(s) \tag{3.1.2}$$

where $l(.)$, $h(.,.)$ and $k(.)$ are known. This is again an integral equation in $\underline{S}^{-1}$. Proposition 3.1.3 ensures that this equation has a unique solution.

Proposition 3.1.3 *Under assumptions 11-16, $\underline{S}^{-1}(.)$ is identified as the unique solution of equation (3.1.2).*

As a consequence, we are able to identify the joint distribution of $(S_1, ..., S_n, V)$, since $F_V(v) = F_W(\underline{S}(v))$ and $F_{S|V}(s|v) = F_{S|W}(s|\underline{S}(v))$.

Summary

Theorem 3.1.4 sums up our previous finding. Its proof follows directly from the previous discussion.

Theorem 3.1.4 *Suppose that assumptions 11-16 hold. Then the joint distribution of $(S_1, ..., S_n, V)$ is identified.*

The determination of the full joint distribution of signals and values that we obtained is essential for addressing policy questions (extent of the winner's curse, optimal reserve prices,...) or for quantifying the bidders' uncertainty about the value after they observed their signals. Hence, tis result is important as it extends previous studies (Li et al., 2000, and Février, 2006) to a fairly large class of Common Value auctions.

As previously mentioned, the mineral rights model is defined up to a transformation of the signals and we made the assumption that $b(s) = s$. This normalization is arbitrary and one can prefer to stay agnostic. In this case, our proof would lead to the identification of the functions $F_V(.)$, $F_{S|V}(b^{-1}(.)|.)$, $\underline{S} \circ b^{-1}(.)$ and $\overline{S} \circ b^{-1}(.)$.

Lastly, it is quite clear that the mineral rights model is overidentified. Indeed we observe a function of n variables, $F_{S_1,...,S_n}$, and seek to identify a function of two variable $F_{S_1|V}$ and a function of one variable, F_V. Hence, it will theoretically be possible to test if the structure it imposes on the data is satisfied by observed bids.

3.1.4 Extensions

We propose here to extend our result to second price auctions, and to show that nonparametric identification can still be achieved when some of the previous assumptions are relaxed.

Second price auctions

In second price auctions, each bidder submits a bid. The winner is the one who submits the highest bid. He obtains the object but pays only the second highest bid. Milgrom and Weber (1982) proved that a symmetric equilibrium strategy exists and is given by $b(s) = V(s, s)$.
The arguments developed to study first price auctions apply to second price auctions. As previously, the observed data allow the econometrician to recover the joint distribution function of $(S_1, ..., S_n, V)$ up to a transformation on V. Then, using the normalization $b(s) = s$, the first order condition takes the form of (3.1.2), with $k(s) = sf_{Y,S}(s, s)$. Because the result of proposition 3.1.3 does not depend on the function $k(.)$, nonparametric identification of $\underline{S}^{-1}$ is still achieved. The reasoning used in the case of first price auctions applies and we obtain the identification result for second price auctions. We state this result as a proposition.

Proposition 3.1.5 *Suppose that assumptions 11-16 hold. Then the mineral rights model is nonparametrically identified with second price auctions.*

The i.i.d. assumption

In the mineral rights model, the signals are independently and identically distributed conditional on the value of the good. This assumption does not play a key role to derive our result. The next proposition states that an extended model in which the dependence of the signals is defined using a copula is nonparametrically identified. The proof is deferred in appendix.

Proposition 3.1.6 *Suppose $n \geq 3$ and that the distribution of the signals conditional on V given by $F(s_1, ..., s_n|v) = C_v(F_{S|V}(s_1|v), ..., F_{S|V}(s_n|v))$ where $C(., ..., .)$ is a known copula. Given 12-6, the mineral rights model is nonparametrically identified.*

This proposition proves that the identification of the mineral rights model does not depend on the structure of the dependence of the signals. No restrictions are imposed on the copula $C_v(.,...,.)$ and our identification result is therefore valid whatever the dependence is.

The strict monotonicity of the support functions

In assumption 12, we suppose that both functions $\underline{S}(.)$ and $\overline{S}(.)$ are strictly increasing in V. We consider here a weakened version of this assumption where only one of these functions depends strictly on V.

Assumption 17 *(strict monotonicity, weak version)* $\underline{S}$ *and* $\overline{S}$ *are differentiable and for all* $v \in [\underline{V}, \overline{V}]$*, either* $\underline{S}'(v) > 0$ *or* $\overline{S}'(v) > 0$*.*

As proposition 3.1.7 shows, the result is still valid in this case. In particular, if the lower bound of the signals is strictly increasing in V, one can suppose that the upper bound is infinite. The proof of proposition 3.1.7 is displayed in appendix.

Proposition 3.1.7 *Suppose that assumptions 11, 13-16 and 17. Then the joint distribution of* $(S_1, ..., S_n, V)$ *is identified.*

Variations in at least one of the bounds are nevertheless needed. To derive this result, we use the fact that a change in s has two effects on the density $f_S(s) = \int_{\max(\overline{S}^{-1}(s),\underline{V})}^{\min(\underline{S}^{-1}(s),\overline{V})} f_{S|V}(s|v) f_V(v) dv$. Indeed, such a change has a direct effect on the probability $f_{S|V}(s|v)$ of observing s if the value of the good is v. It also has an indirect effect as it changes the set of values V that are "compatible" with the signal s. It is possible, as shown in the proof, to separate these two effects and to recover the structural functions when there is variation in only one bound.

Lastly, the key assumption 12 is testable. By the previous discussion, it is indeed sufficient to plot the maximum and minimum of the support of S_2 conditional on $S_1 = s$. If these functions vary with s, assumption 12 is valid. If not, we cannot reject the hypothesis that only one of the functions $\underline{S}(.)$ or $\overline{S}(.)$ depends on v.[7]

[7]More precisely, one can also test if equation (3.1.3) in the appendix is equal to zero or not. Indeed, the difference between the partial derivatives is strictly positive only if there are some variations in at least one of the bounds.

3.1.5 Conclusion

We prove in this paper that the common value model is nonparametrically identified as soon as one allows some variations in the support of the conditional distribution of the signals. This result is important as it extends previous studies (Li et al. (2000) and Fevrier (2006)) to a fairly large class of Common Value auctions and gives a positive answer concerning the nonparametric identification of Common Value auctions that are known to be, under the most general setting, non identifiable.

Appendix : proofs

Proof of proposition 3.1.2

We first consider the case where $k = 0$ in assumption 16, and we suppose that $f_{S|W}(w|w) \neq 0$ (the proof is similar if $f_{S|W}(\overline{S} \circ \underline{S}^{-1}(w)|w) \neq 0$). By definition, if $s_1 \leq s_2 \leq s_3$,

$$f(s_1, s_2, s_3) = \int_{\underline{S}\circ\overline{S}^{-1}(\max(s_3, \overline{S}(\underline{V})))}^{\min(s_1, \underline{S}(\overline{V}))} f_{S|W}(s_1|w) f_{S|W}(s_2|w) f_{S|W}(s_3|w) f_W(w) dw$$

Taking the derivative in s_1 when $s_1 \leq \underline{S}(\overline{V})$, we obtain

$$\begin{aligned}\frac{\partial f}{\partial s_1}(s_1, s_2, s_3) =& f_{S|W}(s_1|s_1) f_{S|W}(s_2|s_1) f_{S|W}(s_3|s_1) f_W(s_1) \\ &+ \int_{\underline{S}\circ\overline{S}^{-1}(\max(s_3, \overline{S}(\underline{V})))}^{s_1} \frac{\partial f_{S|W}}{\partial s}(s_1|w) f_{S|W}(s_2|w) f_{S|W}(s_3|w) f_W(w) dw\end{aligned}$$

Similarly, by taking the derivative in s_2, we find

$$\frac{\partial f}{\partial s_2}(s_1, s_2, s_3) = \int_{\underline{S}\circ\overline{S}^{-1}(\max(s_3, \overline{S}(\underline{V})))}^{s_1} f_{S|W}(s_1|w) \frac{\partial f_{S|W}}{\partial s}(s_2|w) f_{S|W}(s_3|w) f_W(w) dw$$

Hence, if $s' \geq s \geq \underline{S}(\underline{V})$ and $\underline{S}(\overline{V}) \geq s$, we get

$$\frac{\partial f}{\partial s_1}(s, s, s') - \frac{\partial f}{\partial s_2}(s, s, s') = f^2_{S|W}(s|s) f_W(s) f_{S|W}(s'|s) \tag{3.1.3}$$

Because the density $f_{S|W}$ is identified and positive by assumption, f_W is identified on its support by

$$f_W(s) = \frac{\frac{\partial f}{\partial s_1}(s, s, s') - \frac{\partial f}{\partial s_2}(s, s, s')}{f^2_{S|W}(s|s) f_{S|W}(s'|s)}$$

When $k > 0$, we can instead use the following equation :

$$\frac{\partial^{2k+1} f}{\partial s_1^{k+1} \partial s_2^k}(s, s, s') - \frac{\partial^{2k+1} f}{\partial s_1^k \partial s_2^{k+1}}(s, s, s') = \left(\frac{\partial^k f_{S|W}}{\partial s^k}\right)^2 (s|s) f_W(s) f_{S|W}(s'|s)$$

Hence, the distribution of W is still identified in this case.

Proof of proposition 3.1.3

Suppose that two functions $\underline{S}_1^{-1}(.)$ and $\underline{S}_2^{-1}(.)$ satisfy (3.1.2) and let $\phi = \underline{S}_1^{-1} - \underline{S}_2^{-1}$. We shall prove that $\phi = 0$. By definition, ϕ satisfies for every $s \in [\overline{S}(\underline{V}), \overline{S}(\overline{V})]$,

$$\int_{l(s)}^{\min(s,\underline{S}(\overline{V}))} \phi(u)h(s,u)du = 0 \tag{3.1.4}$$

Taking the derivative of this expression in s and using $h(s,s) = 0$ ($F_{S|V}(s|\underline{S}^{-1}(s)) = 0$), one obtains

$$-l'(s)\phi(l(s))h(s,l(s)) + \int_{l(s)}^{\min(s,\underline{S}(\overline{V}))} \phi(u)\frac{\partial h}{\partial s}(s,u)du = 0$$

Hence, because $h(s,l(s))$ is strictly positive for all $s \in [\overline{S}(\underline{V}), \overline{S}(\overline{V})]$, we can rewrite this equation as :

$$l'(s)\phi(l(s)) = \int_{l(s)}^{\min(s,\underline{S}(\overline{V}))} \phi(u)\frac{\frac{\partial h}{\partial s}(s,u)}{h(s,l(s))}du$$

$\left|\frac{\frac{\partial h}{\partial s}(s,u)}{h(s,l(s))}\right|$ is continuous and thus bounded over $[\overline{S}(\underline{V}), \overline{S}(\overline{V})] \times [\underline{S}(\underline{V}), \underline{S}(\overline{V})]$. If we note M a majorant, we have

$$\begin{aligned} l'(s)\left|\phi(l(s))\right| &\leq M \int_{l(s)}^{\min(s,\underline{S}(\overline{V}))} \left|\phi(u)\right| du \\ &\leq M \int_{l(s)}^{\underline{S}(\overline{V})} \left|\phi(u)\right| du \end{aligned}$$

Noting $\psi(s) = \int_{l(s)}^{\underline{S}(\overline{V})} \left|\phi(u)\right| du$, we finally obtain

$$\psi^{'}(s) + M\psi(s) \geq 0$$

and

$$e^{Ms}(\psi'(s) + M\psi(s)) \geq 0$$

We conclude that $e^{Ms}\psi(s)$ is increasing in s. Because $\psi(\overline{S}(\overline{V})) = 0$, $\psi(.)$ has to be negative for all $s \in [\overline{S}(\underline{V}), \overline{S}(\overline{V})]$. However, by definition, $\psi(.)$ is also a positive function. Hence $\psi(.)$ is null everywhere.

We conclude that $\phi(.)$ is also equal to 0 over the interval $[\underline{S}(\underline{V}), \underline{S}(\overline{V})]$. Proposition 3.1.3 follows.

Proof of proposition 3.1.6

The proof is similar to previously, except that we have to take into account for the copula $C(u_1, ..., u_n)$ that defines the dependence structure. Equation (3.1.3) takes the following form :

$$\begin{aligned}\frac{\partial f}{\partial s_1}(s,s,s') - \frac{\partial f}{\partial s_2}(s,s,s') =& (\underline{S}^{-1})'(s) f(s,s,s'|\underline{S}^{-1}(s)) f_V(\underline{S}^{-1}(s)) \\ =& (\underline{S}^{-1})'(s) f^2_{S|V}(s|\underline{S}^{-1}(s)) f_V(\underline{S}^{-1}(s)) f_{S|V}(s'|\underline{S}^{-1}(s)) \\ & \frac{\partial^3 C}{\partial u_1 \partial u_2 \partial u_3}(0,0,F_{S|V}(s'|\underline{S}^{-1}(s),1,...,1)\end{aligned}$$

Hence, for all $s \in [\underline{S}(\underline{V}), \underline{S}(\overline{V})]$ and $s' \in [s, \overline{S} \circ \underline{S}^{-1}(s)]$, we observe

$$\frac{\frac{\partial^2 C}{\partial u_1 \partial u_2}(0,0,F_{S|V}(s'|\underline{S}^{-1}(s),1,...,1)}{\frac{\partial^2 C}{\partial u_1 \partial u_2}(0,0,1,1,...,1)} = \frac{\int_s^{s'} \frac{\partial f}{\partial s_1}(s,s,t) - \frac{\partial f}{\partial s_2}(s,s,t)dt}{\int_s^{\overline{S}\circ\underline{S}^{-1}(s)} \frac{\partial f}{\partial s_1}(s,s,t) - \frac{\partial f}{\partial s_2}(s,s,t)dt}$$

Because $C(.,...,.)$ is known, the function $F_{S|V}(s'|\underline{S}^{-1}(s)$ is identified. $F_V(\underline{S}^{-1}(s))$ is obtained as a byproduct, as well as the functions $\underline{S} \circ \overline{S}^{-1}$ and $\overline{S} \circ \underline{S}^{-1}$.

The first order condition can be written as

$$\begin{aligned}V(s,s) f_{Y,S}(s,s) = & \int_{\max(\underline{S}\circ\overline{S}^{-1}(s), \underline{S}(\underline{V}))}^{\min(s,\underline{S}(\overline{V}))} \underline{S}^{-1}(u)(n-1) f^2_{S|V}(s|\underline{S}^{-1}(u))(\underline{S}^{-1})'(u) f_V(\underline{S}^{-1}(u)) \\ & \frac{\partial^2 C}{\partial u_1 \partial u_2}(F_{S|V}(s|\underline{S}^{-1}(u), F_{S|V}(s|\underline{S}^{-1}(u), F_{S|V}(s|\underline{S}^{-1}(u), ..., F_{S|V}(s|\underline{S}^{-1}(u))du\end{aligned}$$

Hence, for $s \in [\overline{S}(\underline{V}), \overline{S}(\overline{V})]$, one obtains an equation similar to equation (3.1.2) :

$$\int_{l(s)}^{\min(s,\underline{S}(\overline{V}))} \underline{S}^{-1}(u) h(s,u) du = k(s).$$

Therefore, $\underline{S}$ is identified using the same arguments. This ends the proof of proposition 3.1.6 and show that the identification result does not depend on the structure of the dependence of the signal.

Proposition 3.1.7

Suppose for example that $\underline{S}(.)$ is strictly increasing but that $\overline{S}(V) = \overline{S}$ is constant for all $V \in [\underline{V}, \overline{V}]$. Let us start from

$$f(s_1,s_2,s_3) = \int_{\underline{V}}^{\min(\underline{S}^{-1}(s_1),\overline{V})} f_{S|V}(s_1|v) f_{S|V}(s_2|v) f_{S|V}(s_3|v) f_V(v) dv$$

Equation (3.1.3) used to prove proposition 3.1.2 is still valid here. This enables to identify $f_{S|W}$ by

$$f_{S|W}(s'|s) = \frac{\frac{\partial f}{\partial s_1}(s,s,s') - \frac{\partial f}{\partial s_2}(s,s,s')}{\int \frac{\partial f}{\partial s_1}(s,s,t) - \frac{\partial f}{\partial s_2}(s,s,t)dt}$$

for all $\overline{S} \geq s' \geq s$ and $s \in [\underline{S}, \overline{S}]$. Hence, using equation (3.1.3) once more, we can identify f_W over its support $[\underline{S}, \underline{S}(\overline{V})]$.

The equilibrium condition allows us to derive an expression similar to equation (3.1.4) : for all $s \in [\underline{S}(\underline{V}), \overline{S}]$,

$$\int_{\underline{S}(\underline{V})}^{\min(s,\underline{S}(\overline{V}))} \phi(u)h(s,u)du = 0$$

and we have to prove that such a function $\phi(.)$ is null everywhere.
Deriving this equation $n-1$ times and using $F_{S|V}(s|\underline{S}^{-1}(s)) = 0$, we obtain for all $s \in]\underline{S}(\underline{V}), \underline{S}(\overline{V})[$

$$\phi(s)\frac{\partial^{n-2}h}{\partial s^{n-2}}(s,s) + \int_{\underline{S}(\underline{V})}^{s} \phi(u)\frac{\partial^{n-1}h}{\partial s^{n-1}}(s,u)du = 0$$

where $\frac{\partial^{n-2}h}{\partial s^{n-2}}(s,s) = (n-1)!f_{S|V}^{n}(s|\underline{S}^{-1}(s))(\underline{S}^{-1})'(s)f_V(\underline{S}^{-1}(s)) \neq 0$.
The same technique as the one used in proposition 3.1.3 allows us to conclude that $\phi(.)$ is equal to zero everywhere and that $\underline{S}(.)$ is identified.
The mineral rights model is therefore nonparametrically identified even if only one of the bounds is strictly increasing in V.

3.2 Identification and Estimation of Incentive Problems : Adverse Selection

3.2.1 Introduction

Since the seminal work of Akerlof (1970), extensive attention has been devoted to asymmetries of information and their consequences in economics. A canonical example where these asymmetries play a fundamental role is the adverse selection model. This model is helpful, for instance, to better understand nonlinear pricing, regulation, financial contracts or taxation theory (Wilson, 1993 ; Laffont & Tirole, 1993 ; Freixas & Rochet, 1997 ; Diamond, 1998). However, and as pointed out by Chiappori & Salanié (2002), few econometric work has been done to estimate structurally these models. Regulatory contracts have been studied by Wolak (1994), Gagnepain & Ivaldi (2002) and Perrigne (2002), while Ivaldi & Martimort (1994) and Miravette (2002) estimated nonlinear pricing models. All these papers adopt a parametric framework. Lavergne & Thomas (2005) are more flexible and specify a semiparametric model to study regulation. Perrigne & Vuong (2004) are the only ones who follow a nonparametric approach. They study the Laffont & Tirole (1986) regulation model in which ex-post costs are observed, and show that such a model is nonparametrically identified.

In this paper, we study the empirical content of the canonical adverse selection model (Laffont & Martimort, 2002, Salanié, 2005). This model is characterized by the objective function of the principal, the distribution of agents' types and the utility function of the agents. The econometrician is supposed to observe the contract and the associated trades. In the most general setting, the model is not identified. However, under a separability assumption on the utility of the agents, we prove that the knowledge of one of the structural functions is sufficient to obtain full identification. Hence, in a regulation context for instance, if ex-post costs are observed, the utility function can be recovered and full identification is achieved. Similarly, if the regulatory maximizes the sum of the firms' profits and of the consumers' surplus, the objective function of the principal is known and the model is nonparametrically identified.

The identification of the model can also be achieved by observing a change in the contracts between the principal and the agents, under the exclusion restriction that

the utility function and the types of the agents are not affected by this change. As described by Chiappori & Salanié (2002) such exclusion restrictions arise naturally in experiments or natural experiments (see e.g. Manning et al., 1987 ; Ausubel, 1999 ; Lazear, 2000, or Shearer, 2004). We characterize what is identified under these conditions. If the marginal transfer functions defined in the two contracts cross, the model is fully identified. If not, nonparametric bounds can be recovered on the utility function and the distribution of the types. Furthermore, two changes can be sufficient to obtain full nonparametric identification of the model.

To prove these results, we extensively use the first order condition which defines the optimal choices of the agents, and the observed distribution of the trades. The first equation allows us to define what we call horizontal transformations whereas the second one yields to vertical transformations. These transformations are identified in the data and are combined to define recursively the functions of interest. An important feature of our identification procedure is that the utility function and the distribution of the agents' types are recovered using the agent's program solely. This is convenient when the optimality of the principal is questionable. For instance, the common knowledge assumption on the distribution function of the agents' types or their cost function may fail to hold, the principal may also be risk averse (see Lewis & Sapppington, 1995 ; Gence-Creux, 2000) and the costs of implementing nonlinear contracts may modify significantly his program (see Ferrall & Shearer, 1999). Our results are not affected by these problems.

Beyond identification, we also examine some tests of the model. First, our identification results can be used to test in a structural and nonparametric way the optimality of the observed contracts. This question has received considerable attention in the empirical literature (see Prendergast, 1999, for a survey) and the results are rather mixed, the authors finding few evidence that contracts vary with the relevant parameters. However, no nonparametric structural test of the model has been proposed so far.[8] Second, we examine the possibility of testing the asymmetric model versus the symmetric one. We show that implementing such a test is difficult. Without prior knowledge on the structural functions, both models cannot be distinguished nonparametrically, even with exogenous variations in the contracts. This result contrats with previous papers on the topic (see e. g. Puelz & Snow, 1994 ; Chiappori & Salanié, 2000, in which auxiliary variables are used.

[8]Ferrall & Shearer (1999) implement a structural but parametric test.

We also develop a nonparametric estimation procedure based on our recursive identification method and prove that the estimators are consistent. This method is implemented on contract data between the French National Institute of Economics and Statistics (Insee) and its interviewers[9] to study incentives in firms in the spirit of Ferrall & Shearer (1999) and Paarsch & Shearer (2000). Thanks to a change in the bonus paid to interviewers, we recover bounds on the utility function of the agents and the distribution of their type. Then, using the objective function of the principal, we are able to perform a nonparametric test of the contracts' optimality. We reject that Insee's linear contracts are optimal and concludes that it does not fully take into account agents' responses when writing his contracts. However, when estimating the cost of using linear contracts instead of the optimal ones, we find that Insee's loss is about 9%. This result contrasts with the previous literature in which linear contracts were thought to be quite inefficient but simple and easy to implement (Ferrall and Shearer, 1999). Our application points out on the contrary that the loss is quite small, which may explain the wide use of linear contracts. We also recover what Insee's surplus would have been under complete information and find that the estimated expected surplus under incomplete information are 84% of full information surplus. Overall, the cost due to the asymmetry of information are twice the costs associated with the bonus system. Our last results are obtained using parametric specifications in line with our nonparametric results.

The paper is organized as follows. Section 2 recalls the main theoretical results for a principal-agent model with adverse selection. Section 3 is devoted to the nonparametric identification of this model. Our nonparametric estimation method and its application are presented in section 4, and section 5 concludes.

3.2.2 Adverse selection model

Following Laffont & Martimort (2002) and Salanié (2005*a*), we consider a basic adverse selection model where a principal trades y with some agents and provides them with a monetary transfer t. Agents are heterogeneous with a quasi-linear utility function $U(t, y, \theta) = t - C(y, \theta)$.[10] The monetary cost $C(y, \theta)$ of implementing

[9] Insee hires interviewers for its household surveys.

[10] The convention, here, is that y is produced by the agents as in the regulatory model. Equivalently, we could assume that the agents consume y and that the utility function takes the form

y depends on their type θ which is unobserved by the principal. We suppose that θ is real and nonnegative, $\theta \in \Theta = [\underline{\theta}, \overline{\theta}]$, so that we can interpret it as a measure of the agent's intrinsic efficiency. $\underline{\theta}$ is the most efficient agent's type whereas $\overline{\theta}$ is the least efficient. We denote by $F_\theta(.)$ (resp. $f_\theta(.)$) the distribution function (resp. density function) of θ and suppose it to be common knowledge. Lastly, the principal is assumed to be risk neutral. His objective function is quasi linear and takes the form $W(t, y, \theta) = S(y, \theta) - t$. The following regularity conditions are imposed.

Assumption 18 *(regularity conditions)* $f_\theta(.) > 0$ *;* $\partial S/\partial y(.,.) > 0$ *and* $\partial^2 S/\partial y^2(.,.) < 0$ *;* $\partial C/\partial\theta(.,.) > 0$*,* $\partial C/\partial y(.,.) > 0$*,* $\partial^2 C/\partial y^2(.,.) > 0$ *and* $\partial^2 C/\partial\theta\partial y(.,.) > 0$*.*

The objective function of the principal is increasing and concave. The cost increases with inefficiency and with the level of y. Moreover, it is convex as a function of y. Lastly, the positivity of its cross derivative is the Spence-Mirrlees condition, which indicates that a more efficient type is also more efficient at the margin.

The firm proposes to the agent a set of contracts of the form $[(y, t); y \in \mathbb{R}^+, t \in \mathbb{R}^+]$. The agent of type θ can either refuse all contracts or accept one of them. If he accepts a contract (y, t), the agent delivers y and receives a transfer t. If he refuses, he obtains his outside opportunity utility level normalized to zero.

Without asymmetry of information between the principal and the agent, the firm makes a take it or leave it offer to the agent of type θ that implements first-best trade levels. More precisely, the firm proposes to an agent of type θ a contract (y^*_θ, t^*_θ) defined by :

$$\frac{\partial S}{\partial y}(y^*_\theta, \theta) = \frac{\partial C}{\partial y}(y^*_\theta, \theta) \tag{3.2.1}$$

$$t^*_\theta = C(y^*_\theta, \theta) \tag{3.2.2}$$

The optimal trade level is the quantity that equalizes marginal gain and marginal cost. The transfer function is such that the agent accepts the offer but makes zero profit.

If θ is the agent's private information, the complete information optimal contracts can no longer be implemented. The problem arises because of the asymmetric

$U(t, y, \theta) = U(y, \theta) - t$ as in the price discrimination model.

information. In particular, efficient agents mimic inefficient ones and prefer to trade less to have a positive utility.

The optimal menu of contracts is more complex in this case. It may be the case, for example, that the contracts targeted for different types coincide. For those contracts, we say that there is bunching of types. The general theory can be found in Laffont & Martimort (2002) but we only describe the optimal menu of contracts without bunching.

Proposition 3.2.1 *Laffont-Martimort (2002)*
Under assumption 18 and if there is no bunching at equilibrium, the optimal menu of contracts, of the form $(y(\theta), t(\theta); \theta \in \Theta)$*, entails :*
– *A downward output distortion for all types except the most efficient :*

$$\frac{\partial S}{\partial y}(y(\theta), \theta) = \frac{\partial C}{\partial y}(y(\theta), \theta) + \frac{F_\theta(\theta)}{f_\theta(\theta)} \frac{\partial^2 C}{\partial \theta \partial y}(y(\theta), \theta) \tag{3.2.3}$$

– *A positive information rent for all types except the less efficient :*

$$t(\theta) = C(y(\theta), \theta) + \int_\theta^{\overline{\theta}} \frac{\partial C}{\partial \theta}(y(\tau), \tau) d\tau \tag{3.2.4}$$

The firm has to leave a positive information rent to the agents for them to reveal their types (the term $\int_\theta^{\overline{\theta}} \partial C/\partial\theta(y(\tau), \tau)d\tau$ in equation (3.2.4)). This information rent increases with the efficiency of the agent and creates inefficiencies in production (the term
$F_\theta(\theta)/f_\theta(\theta) \times \partial^2 C/\partial\theta\partial y(y(\theta), \theta)$ in equation (3.2.3)).

3.2.3 Nonparametric identification

The general case

We now discuss the empirical content of the model. In the sequel, we suppose that the econometrician observes the trades at equilibrium $(y(\theta_i))_{i\in\mathbb{N}}$ for an infinite sample of agents indexed by i (the types θ_i being independently drawn from F_θ). Agents are supposed to be homogenous except for their unknown types ; if they differ by observed characteristics, our results below must be understood conditionally on these characteristics. We also suppose that the corresponding transfers $t(y(\theta_i))$

are observable.[11] The trades and transfers enable one to identify the cumulative distribution function of y, $F_y(.)$ and the transfer function $t(.)$ on the support $\mathcal{Y}$ of $y(.)$. The question is whether $C(.,.)$, $F_\theta(.)$ and $S(.,.)$ can be recovered from these functions and the model.

Without further assumption, we can always replace θ by $F_\theta(\theta)$ and change $C(.,.)$, F_θ and $S(.,.)$ accordingly. Consequently, F_θ is not identified and we can suppose θ to be uniformly distributed on $[0,1]$. Hence, we focus on $C(.,.)$ and $S(.,.)$ only.

Furthermore, we analyse in this section the case where there is no bunching at equilibrium. Under assumption 18, the existence of bunching is equivalent to F_y admitting at least one mass point. Hence, it is easily testable in the data, by checking whether at least two observed trades are identical or not. The analysis of bunching is deferred to subsection 3.4.

Assumption 19 *There is no bunching at the equilibrium.*

Our identification results are based on three equations.

First, by the Spence-Mirrlees condition and assumption 19, $y(.)$ is strictly decreasing. Thus it admits an inverse $\theta(.)$ which satisfies, for all $y \in \mathcal{Y}$,

$$F_y(y) = \mathbb{P}(y(\theta) < y) = \mathbb{P}(\theta > \theta(y)) = 1 - \theta(y) \tag{3.2.5}$$

The first equality stems from the fact that the distribution of y is atomless and the second from $\theta(.)$ being strictly decreasing. Because $F_y(.)$ is observed, this equation shows that $\theta(.)$ is identified .

Secondly, the agent chooses his production according to the first order condition, which writes as

$$\frac{\partial C}{\partial y}(y, \theta(y)) = t'(y) \tag{3.2.6}$$

In other terms, $\partial C/\partial y$ is identified on $L = \{(y, \theta(y)), y \in \mathcal{Y}\}$. Note also that $\partial C/\partial y$ is not identified elsewhere because no (y, θ) is observed outside equilibrium. Given the price schedule, we can only recover for each agent's type θ, the marginal utility at his optimal choice of production.

[11] This assumption may be strong (see Wolak, 1994, and Ferrall & Shearer, 1999, for examples where the transfers are unknown).

Lastly, the third equation is the first order condition of the principal (3.2.3), which can also be written as :

$$\widetilde{S}(y) = \frac{\partial C}{\partial y}(y, \theta(y)) + \theta(y)\frac{\partial^2 C}{\partial \theta \partial y}(y, \theta(y)) \tag{3.2.7}$$

where $\widetilde{S}(y) = \partial S/\partial y(y, \theta(y))$. From an identification point of view, this equation does not impose conditions on $S(.,.)$ directly, but rather on $\widetilde{S}(.)$, and we thus focus on this function hereafter.[12] Because no (y, θ) is observed outside equilibrium, $\partial^2 C/\partial\theta\partial y$ is not identified. Thus, $\widetilde{S}(.)$ is not identified either.

Note that even if $\widetilde{S}(.)$ is known by the theory, we cannot recover $\partial C/\partial y(.,.)$ outside L and predict for instance what would be the optimal contract under another objective function of the principal.

To circumvent this nonidentification result, we impose a restriction on the utility function by supposing that the cost function is separable. This hypothesis is often made in the theoretical adverse selection literature (see Wilson, 1993 or Laffont & Tirole, 1993). It is also assumed in empirical research (see Wolak, 1994, Ferrall & Shearer, 1999, Lavergne & Thomas, 2005) and in the nonparametric analysis of Perrigne & Vuong (2004). Other functional restrictions on $C(.,.)$ are possible. We present one of them, which stems from the false moral hazard model, in subsection 3.4.

Assumption 20 *(cost separability)* $C(y, \theta) = \theta C(y)$.

Under assumption 20, the uniform normalization of F_θ is not possible anymore. On the other hand, we can replace $(\theta, C(.))$ by $(\alpha\theta, C(.)/\alpha)$ and leave the model unchanged. Thus, another normalization is necessary and for a given $y_0 \in \mathcal{Y}$, we can choose θ_0 such that $\theta(y_0) = \theta_0$. In other words, assumption 20 reduces the dimensionality of the problem by replacing a function of two variables $C(.,.)$ by two functions of one variable $C(.)$ and $F_\theta(.)$ and the structural parameters are now $(C', F_\theta, \widetilde{S})$. Identification is based on the same equations than previously, and the proof of proposition 3.2.2 is deferred to appendix A.

Proposition 3.2.2 *Under assumptions 18, 19 and 20,* $(C', F_\theta, \widetilde{S})$ *are not identified jointly. On the other hand, if one of these three functions is known, the other two can be identified.*

[12] In general, the knowledge of $\widetilde{S}(.)$ does not enable to recover $S(.,.)$. However, it will (up to an additive constant) in some applications where $S(.,.)$ does not actually depend on θ.

This result states that even under the separability assumption, the model remains unidentified. Basically, only three equations are available whereas we seek to recover four functions : $(C'(.), F_\theta(.), \widetilde{S}(.), \theta(.))$. Hence, we can fix one of the four functions and deduce the others from the data and the model. Actually, this function cannot be chosen completely arbitrarily. Indeed, the three structural functions $(C'(.), F_\theta(.), \widetilde{S}(.))$ must satisfy assumption 18 and be such that the second order conditions of the principal's and the agent's programs hold. Some choices can be discarded according to these criterions and bounds on parameters of interest may be obtained through these constraints (see Salanié, 2005b, for an approach of this kind).

On the other hand, if one of the three structural functions is known, we can recover the other functions of interest. In particular, and contrarily to the previous general case, the knowledge of $\widetilde{S}(.)$ enables to recover the cost function everywhere. Another application of our result is regulation with ex-post observable costs. Suppose indeed that $\theta(y)C(y)$ is observed. Because $\theta(y)C'(y) = t'(y)$ is also identified, $C'/C(.)$ is identified. Then C can be recovered up to a multiplicative constant, which is given by the normalization $\theta(y_0) = \theta_0$. As a consequence, F_θ and $\widetilde{S}$ are identified. This result is in line with Perrigne & Vuong (2004)'s one.[13]

We review in the following several classical settings where this model is useful and show how identification can be obtained.

Quality and Price Discrimination

In Mussa & Rosen (1978), the principal is a firm that produces a good of quality q at a cost $H(q)$. Agents have heterogenous preferences for quality θ (distributed following F_θ) and have a utility $U = \theta q - t$ if they pay t for a good of quality q.

In this setting, the objective function of the principal is unknown and depends on $H(.)$ ($S(q, \theta) = H(q)$ in the previous notation) whereas the utility of the agents is specified ($C(q) = q$). Our proposition implies that this model is identified nonparametrically.[14]

The same model can be used to study nonlinear pricing by a monopoly (Maskin

[13] Actually, the problem of Perrigne & Vuong (2004) is more involved because they consider the regulation model of Laffont & Tirole where the regulated firm can make a costly and unobserved effort to reduce its cost. Thus, they have to deal both with adverse selection and moral hazard.

[14] In this example, the principal minimizes his objective function instead of maximizing it.

& Riley, 1984). Our result shows that if the utility function of agents is specified, the model is identified.

Financial contracts

In Freixas & Laffont (1990) framework, the principal is a lender who provides a loan y to a borrower and has a utility $S(y) = t - Ry$ where R is the risk-free interest rate. Agents are firms with profit $U = \theta f(y) - t$. $\theta f(y)$ is the production of the firm, y represents the units of capital and θ is a productivity index.

Here, the objective function of the principal is known because R, the risk-free interest rate is observed. Our proposition implies that the production function $f(.)$ and the distribution of the types $F_\theta(.)$ are identified.

Regulation

In the Baron & Myerson (1982) model, the regulator maximizes a weighted sum of the consumers' surplus and the regulated firms defined by heterogenous cost functions of the form $\theta C(y)$. In our notation, we have :

$$S(y,\theta) - t(y) = (1-\alpha)\left[\int_0^y p(u)du - t(y)\right] + \alpha\left[t(y) - \theta C(y)\right] \tag{3.2.8}$$

From this equation, we derive $\widetilde{S}(y) = (1-\alpha)p(y) + \alpha t'(y)$. Hence, when the price function $p(y)$ is observed (which is usually the case in the empirical literature), $\widetilde{S}(.)$ is known up to the parameter α and our proposition proves that $C(.)$ and $F_\theta(.)$ are identified up to this parameter.

Identification under exclusion restrictions

The setting

Proposition 3.2.2 states that the model is identified provided that one of the three structural functions is known. This condition may however be restrictive in certain settings. In the regulation problem, one has to assume that the weighting parameter α is known. In the price discrimination case, identification is based on the assumption of linearity of U in q. Moreover, one of the major empirical question in contract theory is whether the observed contracts are optimal compared to the theoretical ones (Chiappori & Salanié, 2002). Answering this question implies estimating $\widetilde{S}$ and compare it with the theoretical one. In this case, one can obviously not rely on the theoretical $\widetilde{S}$ to identify the model.

In this subsection, we examine the identification of the model when several menus of contracts are available but none of the functions $(F_\theta, C', \widetilde{S})$ is known. More precisely, we suppose that we observe variations in the menus of contracts under the exclusion restriction that the utility function and the distribution function of θ are not affected by these changes. We suppose in particular that there is no selection effect (see the discussion on this issue in subsection 3.4).

Exogenous variations in the menus can be observed for different reasons. The first and ideal case is experiments : if different contracts are proposed to people in a random way, then endogenous changes or selection problems are not a concern. For instance, the Rand Health Insurance experiment (see Manning et al., 1987) randomly assigned families who participate in the experiment to 14 different insurance plans. Similarly, Ausubel (1999) analyses the market for bank credit by using randomized mailed solicitations. The propositions vary in the interest rates and the duration of the loan. The econometrician may also use natural experiments where the objective function S of the principal changes for an exogenous reason.[15] Laws modifications are often good candidates for this purpose. Many examples have been already studied in the literature, especially in moral hazard situations (see e.g. Dionne & Vanasse, 1996 ; Chiappori, Durand & Geoffard, 1998a ; Chiappori, Geoffard & Kyriadizou, 1998b ; Banerjee et al., 2002). In the regulation context, one could also use changes in the government, which may induce variation of the parameter α, while the structural parameters (F_θ, C') remain constant (Gagnepain & Ivaldi, 2007). In a monopoly price discrimination model, the price of one input may increase, inducing a change in the cost function of the monopoly and thus in S. However, in a partial equilibrium framework, this increase does not affect the utility function of the customer. In the delegation of production to agents, the firm may restructure their wage schedule for an exogenous reason such as, for instance, a managerial change. Another example on the French National Institute will be developed in section four (see also Lazear, 2000).

In the sequel, we denote by λ the index of these different menus. We suppose that there are K different indices and we denote their set by $\Lambda = \{\lambda_1, ..., \lambda_K\}$. $t(., \lambda)$ (resp. $\widetilde{S}(., \lambda)$) is the transfer function (resp. the marginal objective function

[15] Our analysis could be adapted to the case where C' (resp. F_θ) alone varies, $(\widetilde{S}, F_\theta)$ (resp. $(\widetilde{S}, C')$) remaining constant. We focus on the variations of the principal's objective function because we believe it to be the most common situation.

at the optimum) corresponding to λ. The production chosen at equilibrium also depends on λ and we denote it by $y(\theta, \lambda)$, its distribution function being $F_y(., \lambda)$. Its inverse function is $\theta(y, \lambda)$. We now suppose that the econometrician has access to data of the form $(t(y_i, \lambda_i), y_i, \lambda_i), i \in \{1, ..., N\}$. Here, λ_i does not necessarily have an economic meaning : observing λ_i only indicates that the type of contract of individual i is known. As previously, $t(., \lambda)$ and $F_y(., \lambda)$ are identified for all $\lambda \in \Lambda$, and our aim is to recover the structural parameters $\left(C', F_\theta, (\widetilde{S}(., \lambda))_{\lambda \in \Lambda}\right)$.

We start from the first order condition and the monotonicity condition :

$$\frac{\partial t}{\partial y}(y, \lambda) = \theta(y, \lambda) C'(y) \tag{3.2.9}$$

$$F_y(y(\theta, \lambda), \lambda) = 1 - F_\theta(\theta) \tag{3.2.10}$$

These two equations, together with (3.2.3), show that it suffices to identify $(y(., \lambda))_{\lambda \in \Lambda}$ or equivalently $(\theta(., \lambda))_{\lambda \in \Lambda}$ to recover the structural parameters. The idea is thus to focus on these functions in order to derive our identification results. To do so, we first need to introduce two types of transforms that will be at the basis of our identification method.

The horizontal and vertical transforms

Equation (3.2.10) implies that for all $\theta \in \Theta$,

$$F_y(y(\theta, \lambda_i), \lambda_i) = F_y(y(\theta, \lambda_j), \lambda_j).$$

The ranks of $y(\theta, \lambda_i)$ and $y(\theta, \lambda_j)$ are identical in their respective distribution. Hence, letting $H_{ij}(y) = F_y^{-1}[F_y(y, \lambda_i), \lambda_j]$ denote the quantile-quantile transformation between the distribution of $y(\theta, \lambda_j)$ and $y(\theta, \lambda_i)$, we get

$$y(\theta, \lambda_j) = H_{ij}(y(\theta, \lambda_i)) \tag{3.2.11}$$

The horizontal transforms on figure 3.2 are identified and we can recover point (1) for instance if we know point (0).

Besides, let $\mathcal{Y}_i$ be the support of $y(\theta, \lambda_i)$. For $i \neq j$, suppose that $\mathcal{Y}_i \cap \mathcal{Y}_j \neq \emptyset$ and let $y \in \mathcal{Y}_i \cap \mathcal{Y}_j$. The first order condition implies that

$$\frac{\frac{\partial t}{\partial y}(y, \lambda_i)}{\theta(y, \lambda_i)} = \frac{\frac{\partial t}{\partial y}(y, \lambda_j)}{\theta(y, \lambda_j)}.$$

If we define the vertical transform $V_{ij}(.,.)$ by $V_{ij}(\theta, y) = \partial t/\partial y(y, \lambda_j) \times \theta/\partial t/\partial y(y, \lambda_i)$, we get

$$\theta(y, \lambda_j) = V_{ij}(\theta(y, \lambda_i), y) \tag{3.2.12}$$

Because $V_{ij}(.,.)$ is identified on $\mathbb{R} \times \mathcal{Y}_i \cap \mathcal{Y}_j$, the knowledge of $\theta(y, \lambda_i)$ implies the knowledge of $\theta(y, \lambda_j)$. In particular, it suffices to identify $\theta(., \lambda_1)$ to recover the other functions $\theta(., \lambda_k)_{2 \leq k \leq K}$. Starting from point (1) for instance on figure 3.2, we can identify point (2).

To conclude, starting from $(y_0, \theta(y_0, \lambda_1))$, we can identify $(y_1, \theta(y_1, \lambda_1))$ where $y_1 = H_{12}(y_0)$ and $\theta(y_1, \lambda_1) = V_{21}(\theta(y_0, \lambda_1), y_1)$. By induction, we identify all the black points in figure 3.2.

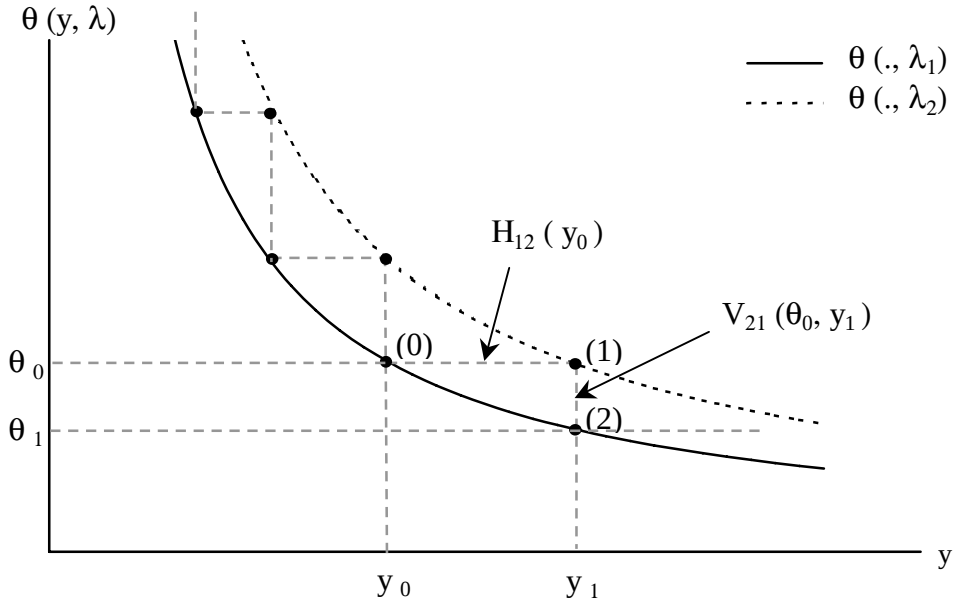

FIG. 3.2 – The horizontal and vertical transforms.

Identification results for $K = 2$.

Figure 3.2 corresponds to a situation where $\theta(y, \lambda_1) < \theta(y, \lambda_2)$ for all $y \in \mathcal{Y}_1 \cap \mathcal{Y}_2$. This does not hold in general, since $\theta(., \lambda_1)$ and $\theta(., \lambda_2)$ may cross. Actually, because $\theta(y, \lambda_1) < \theta(y, \lambda_2)$ is equivalent to $\partial t/\partial y(y, \lambda_1) < \partial t/\partial y(y, \lambda_2)$, the two cases can be distinguished by the data. As they lead to different results in terms of identification, we consider them separately.

Figure 3.2 suggests that without crossing, $\theta(.,\lambda_1)$ can be identified on some points but not everywhere. This implies partial identification of the model. Theorem 3.2.3 formalizes this idea.

Theorem 3.2.3 *Suppose that $K = 2$, assumptions 18-20 hold and $\partial t/\partial y(.,\lambda_1) < \partial t/\partial y(.,\lambda_2)$. Then $C'(.)$ and $F_\theta(.)$ are identified on two sequences. Upper and lower bounds can be obtained elsewhere. The functions $\left(\widetilde{S}(.,\lambda)_{\lambda\in\Lambda}\right)$ are not identified.*
Bounds on C' and F_θ can be recovered when $\partial t/\partial y(.,\lambda_1) < \partial t/\partial y(.,\lambda_2)$. Full nonparametric identification is not achieved but what is recovered can be sufficient to test parametric restrictions on F_θ or C'. In contrast with these positive results, the proposition shows that nothing can be learnt on the principal's value function when $K = 2$. Recovering $\widetilde{S}$ is indeed more demanding than identifying (C', F_θ) on some points, since it requires to recover f_θ (see equation (3.2.3)). Here, f_θ is unidentified because only isolated points of F_θ can be obtained.[16]

To recover full nonparametric identification, more structure on the model and on the functions of interest must be imposed. However, we need not fix or observe entirely one of the structural functions anymore as in the case $K = 1$. Usually, a parametric restriction on one of these functions will be sufficient to identify (and even overidentify) the model. Consider for instance the case of regulation, where the exogenous change of the transfer function is due to a modification of the unknown parameter α in (3.2.8) and let us call α_1 (resp. α_2) this parameter in the first (resp. second) sample. On the one hand, using proposition 3.2.3, $C'(.)$ is identified on a sequence $(y_n)_{n\in\mathbb{Z}}$. On the other hand, fixing $\alpha = \alpha_1'$ is equivalent to specifying the principal objective function. This defines, by proposition 3.2.2, a unique function $C'_{\alpha_1'}$. In general, the sequences $(C'(y_n))_{n\in\mathbb{Z}}$ and $(C'_{\alpha_1'}(y_n))_{n\in\mathbb{Z}}$ are equal for a unique value α_1', which ensures the identification of this parameter. We identify similarly α_2. By proposition 3.2.2, the whole model is identified.

We now turn to the case where functions $\partial t/\partial y(.,\lambda_1)$ and $\partial t/\partial y(.,\lambda_2)$ cross. In this case, the model can be fully recovered thanks to the intersection point. The proof is quite different from previously and can be explained as follows. By the normalization, an intersection point (y_c, θ_c) can always be fixed. For any y_0, define the sequence $(\theta_n)_{n\in\mathbb{N}}$ as in figure 3.3 from a given θ_0. We show that $(\theta_n)_{n\in\mathbb{N}}$

[16] To obtain partial identification, more structure is needed. For example, if $S(y,\theta) = S(y)$, we can show that bounds can be recovered for $S(y_{n+1}) - S(y_n)$, where $(y_n)_{n\in\mathbb{N}}$ is the sequence of points defined as in figure 3.2.

converges to θ_c if and only if $\theta_0 = \theta(y_0, \lambda_1)$. This enables to recover θ_0, since θ_c is known. Hence, $\theta(., \lambda_1)$ is fully identified and all the structural functions can be recovered.

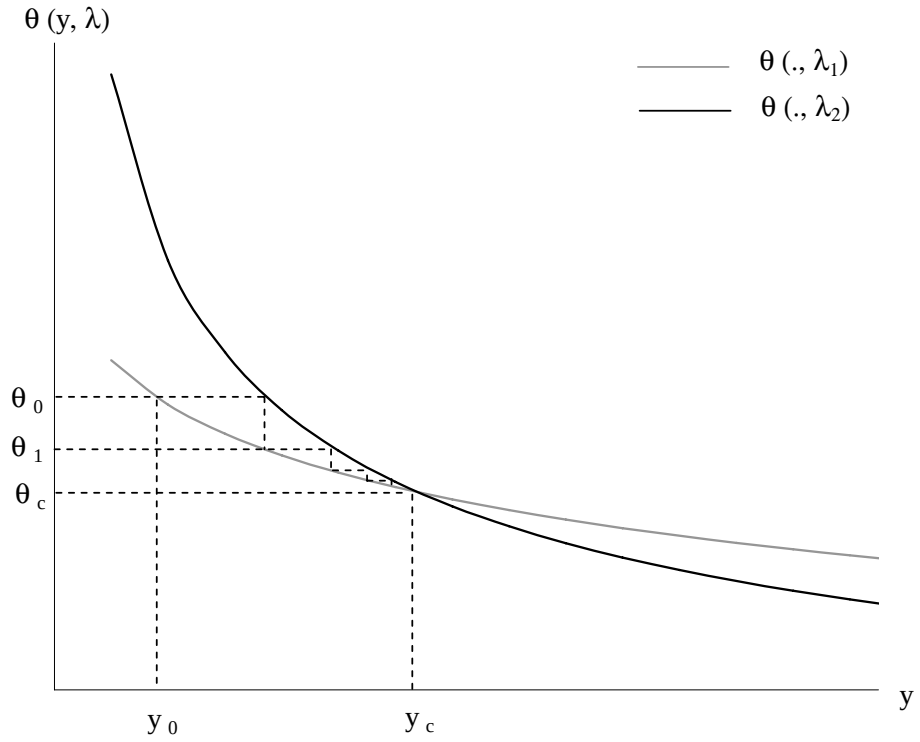

FIG. 3.3 – Identification when $\partial t/\partial y(., \lambda_1)$ and $\partial t/\partial y(., \lambda_2)$ cross.

Theorem 3.2.4 *Suppose that $K = 2$, assumptions 18-20 hold and $\partial t/\partial y(y_c, \lambda_1) = \partial t/\partial y(y_c, \lambda_2)$ for a given y_c in the interior $\overset{\circ}{\mathcal{Y}}_1$ of $\mathcal{Y}_1$. Then C', F_θ, $\widetilde{S}(., \lambda_1)$ and $\widetilde{S}(., \lambda_2)$) are identified on their support.*

Theorem 3.2.4 is reminiscent of the result of Guerre, Perrigne and Vuong (2005) in the context of first-price auctions with risk averse bidders. They also use exogenous variations (namely, the variation in the number of bidders) to obtain identification of the model at the limit, using a converging sequence (see their proposition 1).

Identification results for $K \geq 3$.

In this subsection, we study the noncrossing case with two or more exogenous changes. Here we have in hand not only the transforms H_{12} and V_{12}, but also (when $K = 3$) H_{13}, H_{23}, V_{13} and V_{23}. As a consequence, the set on which $\theta(., \lambda_1)$ is identified is larger. Figure 3.4 gives an example where starting from (y_0, θ_0) on the

curve $\theta(.,\lambda_1)$, we can identify $\theta(.,\lambda_1)$ on y_1 as previously but also between y_0 and y_1 (on y_2 for instance). Proposition 3.2.5 defines the precise set where the function $\theta(.,\lambda_1)$ is point identified.

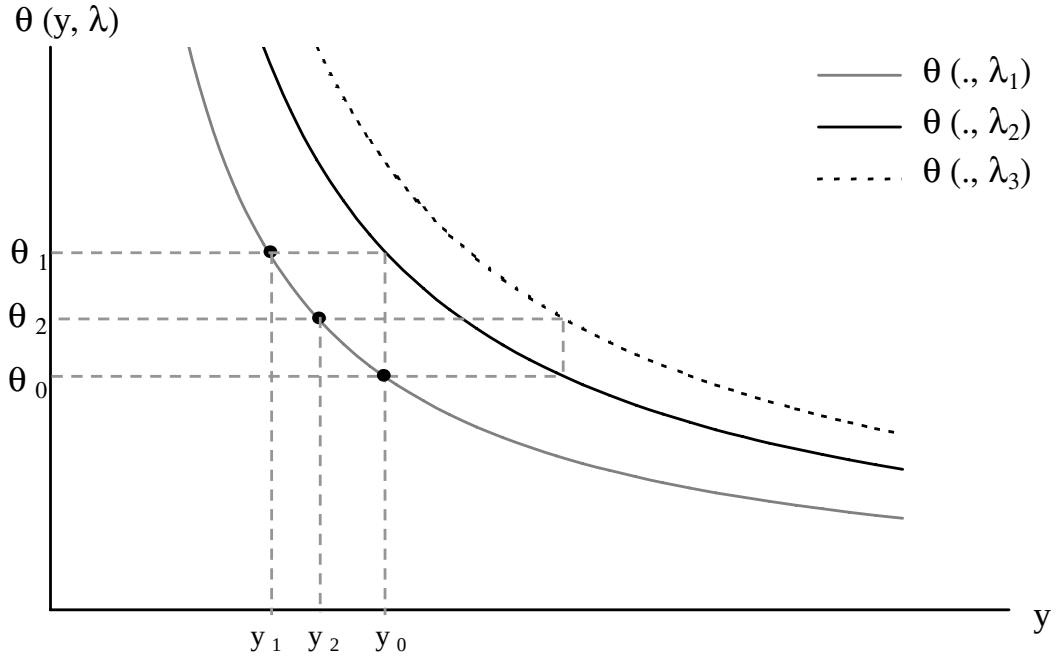

FIG. 3.4 – Identification with $K = 3$ different transfer functions.

Proposition 3.2.5 *Suppose that $K \geq 3$, assumptions 18-20 hold and $\partial t/\partial y(.,\lambda_1) < ... < \partial t/\partial y(.,\lambda_K)$. $\theta(.,\lambda_1)$ is identified on $\overline{\mathbb{Y}} \cap \mathcal{Y}_1$ where $\overline{\mathbb{Y}}$ is the closure of the set $\mathbb{Y}$ defined by :*

$$\begin{cases} y_0 \in \mathbb{Y} \\ \text{For all } (y,i,j) \in \mathbb{Y} \times \{1,...,K\}^2,\ y \in \mathcal{Y}_i \text{ implies } H_{ij}(y) \in \mathbb{Y} \end{cases}$$

It seems difficult to characterize $\mathbb{Y}$ more precisely without further restrictions. It may happen that $\overline{\mathbb{Y}} \cap \mathcal{Y}_1 \neq \mathcal{Y}_1$, so that similarly to the case $K = 2$, only bounds can be obtained on C' and F_θ. However, under the assumptions below, we prove that $\mathbb{Y}$ is dense in $\cup_{i=1,..,K}\mathcal{Y}_i$ and that the model is fully identified.

Assumption 21 *(separability in the transfer function) For all (y,λ),*

$$\partial t/\partial y(y,\lambda) = l(\lambda)m(y).$$

Assumption 22 *(non periodicity) : there exists* $1 \leq i < j < k \leq K$ *such as*

$$\frac{\ln(l(\lambda_i)/l(\lambda_j))}{\ln(l(\lambda_k)/l(\lambda_j))} \notin \mathbb{Q}. \tag{3.2.13}$$

Up to ignoring some elements of Λ*, we can let without loss of generality* $i = 1$*,* $j = 2$ *and* $k = 3$*.*

Assumption 23 *(large support) :* $H_{21}(\mathcal{Y}_1 \cap \mathcal{Y}_3) \cap \mathcal{Y}_2 \neq \emptyset$ *and* $H_{23}(\mathcal{Y}_1 \cap \mathcal{Y}_3) \cap \mathcal{Y}_2 \neq \emptyset$*.*

Assumption 22 is a technical condition which ensures that the identifying sequences are not periodic. Because almost every real are irrational, this assumption should not be seen as restrictive. The large support condition ensures that enough horizontal transforms can be performed to obtain new points where $\theta(., \lambda_1)$ is identified. This assumption can be easily checked in the data. The more restrictive assumption is the separability hypothesis. It includes nevertheless the case of constant marginal transfers and is directly testable in the data.[17] Theorem 3.2.6 shows that the model is fully nonparametrically identified under these assumptions.

Theorem 3.2.6 *If* $K \geq 3$ *and assumptions 18-23 hold,* $C'(.)$*,* $F_\theta(.)$ *and* $\widetilde{S}(., \lambda_i)$*, for all* $i = 1, .., K$*, are identified on their support.*

Theorem 3.2.6 implies that when $K \geq 4$, the model is actually overidentified.[18] Indeed, we can recover (under suitable adaptations of the assumptions) F_θ and C' by using different subsets of Λ. If the different corresponding functions do not coincide, the model is rejected.

Identification results with continuous variations.

It may happen that there is actually a continuum of exogenous variations. In the case of price discrimination for instance, the prices of the input of the monopoly may take any value in an interval, implying also that the value function of the principal changes continuously. As mentioned above, the model with cost separability is now overidentified. Actually, identification can be obtained without this assumption.

[17] Note that it automatically implies $\partial t/\partial y(., \lambda_1) < ... < \partial t/\partial y(., \lambda_K)$, up to a reindexation.

[18] Whether the model is just identified or overidentified when $K = 3$ is unclear to us.

Proposition 3.2.7 *Suppose that* $\Lambda = [\underline{\lambda}, \overline{\lambda}]$*, assumption 1 and 2 hold,* θ *is uniformly distributed on* $[0, 1]$ *(without loss of generality) and* $\partial^2 t/\partial y \partial \lambda(y, \lambda) > 0$*. Then*
– $\partial C/\partial y(.,.)$ *is identified on* $\{(y, \theta)/\theta \in \Theta,\ \exists \lambda \in \Lambda/\theta(y, \lambda) = \theta\}$
– $\widetilde{S}(., \lambda)$ *is identified on* $\{y/\exists(\theta, \lambda) \in \Theta \times \Lambda/\theta(y, \lambda) = \theta\}$.

Testing the model

The empirical foundations of theoretical models from contract theory have been much debated in the literature. In this subsection, we identify three tests of our model, which correspond to well defined economic questions.

Do people react to incentives ?

Incentives are at the core of economic reasoning and many papers have sought evidence on the reactions to these incentives (see Prendergast, and Chiappori & Salanié, for surveys on this issue).[19] In our model and with $K \geq 2$, we can test such reactions by checking if

$$\partial t/\partial y(y, \lambda_1) < \partial t/\partial y(y, \lambda_2) \Longleftrightarrow \theta(y, \lambda_1) < \theta(y, \lambda_2)$$

holds for all $y \in \mathcal{Y}_1 \cap \mathcal{Y}_2$. Using (3.2.10), this condition may be written as

$$\left\{ y/\frac{\partial t}{\partial y}(y, \lambda_1) < \frac{\partial t}{\partial y}(y, \lambda_2) \right\} = \left\{ y/F_y(y, \lambda_1) > F_y(y, \lambda_2) \right\}. \tag{3.2.14}$$

In particular, when $\partial t/\partial y(., \lambda_1) < \partial t/\partial y(., \lambda_2)$, the model implies that $F_y(., \lambda_1)$ is stochastically dominated at the first order by $F_y(., \lambda_2)$. This implication can be straightforwardly tested by the data, for instance through a one-sided Kolmogorov-Smirnov test.

Are contracts optimal ?

Another important empirical issue of contract theory is the optimality of observed contracts. Nonstructural approaches can only bring limited clues on this issue and the structural papers have relied so far on parametric forms (see e.g. Ferrall & Shearer, 1999). In such a framework, the rejection of the model can either discard the parametric hypotheses or the theoretical framework. On the contrary, our nonparametric approach enables to test the theory solely.

[19] Overall, the conclusion of this research is rather positive.

The previous results imply that the model is fully or partially identified when $K \geq 2$ and without the knowledge of the theoretical value function of the principal. Hence, with such an extra assumption, the model is generally overidentified, and we can test whether this theoretical form is coherent with the data.

Does asymmetric information really matter ?

Lastly, a large literature has dealt with the empirical relevance of asymmetries of information (see Chiappori and Salanié, 2002, for a review). Generally speaking, the results of this literature throw doubts on the importance of such asymmetries. In this subsection, we question the possibility of testing nonparametrically the complete information model described in section 2.1 versus the asymmetric information model, when the information set of the principal is unknown. Proposition 3.2.8 sums up our findings.

Proposition 3.2.8 *Suppose that assumption 18, 19 and 20 hold. In general, no test of complete versus asymmetric information can be done if $K = 1$, even if one of the structural functions is known. When $K \geq 2$, a test can be implemented under one of the following conditions :*

–

$$\left\{ y / \frac{\partial t}{\partial y}(y, \lambda_1) < \frac{\partial t}{\partial y}(y, \lambda_2) \right\} \neq \left\{ y / t(y, \lambda_1) < t(y, \lambda_2) \right\} \tag{3.2.15}$$

– *assumption 21 holds and C' is known ;*

– *assumption 21 holds, one function $\widetilde{S}(., \lambda_0)$ is known and $y \mapsto t(y, \lambda_0) \times (1 - F_y(y, \lambda_0))$ is not constant.*

Hence, the possibilities of testing for asymmetric information are rather limited. In particular, the two models cannot be distinguished under assumption 21 without auxiliary information, even when $K \geq 2$. This stems from the fact that under this assumption, the two models lead to the same function $\theta(., \lambda_1)$.[20] This result contrasts with the previous literature on insurance in which nonstructural tests are performed using auxiliary variables (Puelz and Snow, 1994 ; Chiappori & Salanié, 2000).

[20]The overidentification restrictions on $(\theta(., \lambda))_{\lambda \in \Lambda}$ (when $K \geq 4$) are useless to test between the two models because if they fail to hold, both models are rejected.

Discussion and extensions

In this subsection, we come back on assumptions 19 and 20. We also discuss the issue of selection effects.

Bunching

We have maintained up to now the assumption that no bunching occurs at the equilibrium. As mentioned previously, this assumption is testable, by checking if F_y admits a mass point. Bunching is also equivalent to $t(.)$ being continuously differentiable. Many observed contracts do not fulfill this requirement. For instance many production contracts exhibit kinks (see e.g. Ferrall and Shearer, 1999). In this case, bunching provides more information on the model. Figure 3.5 gives the intuition for this result. Fixing (θ_0, y_0) at one extremity of the bunching, it is possible to identify $\theta_1 = (\partial t/\partial y^-(y_0, \lambda_1))/(\partial t/\partial y^+(y_0, \lambda_1))\theta_0$ i.e. the interval $[\theta_0, \theta_1]$. The horizontal and vertical transformations of this interval allows the econometrician to identify $\theta(., \lambda_1)$ and consequently $C'(.)$, $F_\theta(.)$ and the functions $(\widetilde{S}(., \lambda))_{\lambda \in \Lambda}$ on several segments, even when $K = 2$.

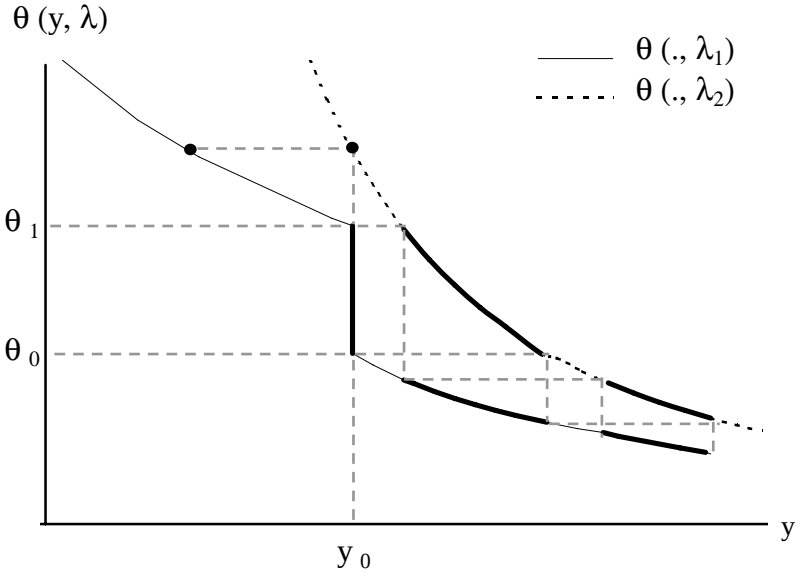

FIG. 3.5 – Identification with bunching. Thick lines (resp. black points) correspond to the identified intervals (resp. points) of $\theta(., \lambda_1)$ and $\theta(., \lambda_2)$.

The false moral hazard model

Our results are based on the cost separability assumption $C(\theta, y) = \theta C(y)$. Other restrictions are however possible. One example is given by the "false moral hazard"

model (see Laffont and Martimort, 2002, p. 287). In this framework, we suppose that the production of the agent depends on θ but also on the level of his effort e, so that $y = g(\theta, e)$. The cost $C(e)$ depends only on the effort e. The agent observes the random term θ before he chooses his effort so that he maximizes $t(g(\theta, e), \lambda) - C(e)$. Though apparently close to the moral hazard model, this model does not share its properties (the trade-off between efficiency and risk insurance especially) and is rather an adverse selection model.

From an econometric point of view, the identification of the false moral hazard model is very close to the previous analysis. The model satisfies $C(y, \theta) = C(g_2^{-1}(\theta, y))$ where $g_2^{-1}(\theta, .)$ is the inverse function of $g(\theta, .)$.As previously, the structural parameters are C', F_θ and the $(\widetilde{S}(., \lambda))_{\lambda \in \Lambda}$. In this framework, proposition 3.2.2 and 3.2.5 are still valid. The equivalent of proposition 3.2.6 also holds if the marginal transfer is constant and the production function is separable, $g(\theta, e) = h_1(\theta)h_2(e)$, as with Cobb-Douglas functions.

Selection effects

We have supposed until now that variations in the transfer functions do not yield any changes in F_θ. However, selection effects can be important. Lazear (2000), for instance, showed that half of the productivity increase observed in a car glass company after moving from constant wages to piece rates could be explained by the arrival of more productive worker. These effects are not taken into account in our model where all types of agent participate. Hence, our analysis is not valid in general when selection occurs.

There is, however, a simple way to detect selection effects when panel data are available. It suffices indeed, as in Lazear (2000) to compare the distributions of the production of the stayers and the entrants. If these distributions differ significantly, selection should be dealt with by modeling the selection process.

3.2.4 Application

In this section, we apply our results on contract data between the French National Institute of Economics and Statistics (Insee) and the interviewers it hires to make household surveys. For each survey, Insee's interviewers receive a random sample of households close to their residence and has to fulfill a maximum number of

face to face interviews. The Insee cannot compel its interviewers to obtain a given number of interviews and it provides them incentives through a linear scheme. Each interviewer receives a basic wage plus a bonus for each interview he achieves.

This application contributes to the literature of provisions of incentive by firms (see Prendergast, 1999, for a survey) which has been interested in studying 1) to what extent agents react to incentives and 2) the optimality of the observed contracts. Our approach is structural and similar to Paarsch and Shearer (2000) and Ferrall and Shearer (1999), but nonparametric. In these models, the production of a worker is a known function of his effort and his type. As explained in subsection 3.4, they are what Laffont and Martimort (2002) call "false moral hazard " models. These models have a structure that are very similar to adverse selection models and our results apply.

Here, for the ease of the presentation, we model the behavior of interviewers without making explicit reference to the underlying effort they produce. In a survey $j \in \{1, ..., J\}$, an interviewer $i \in \{1, ..., N\}$ receives a random sample of housings. The households to interview may be easy or difficult to contact and reluctant or not to accept the questionnaire. We summarize the average difficulty of this sample by a parameter $\theta_{ij} \in \mathbb{R}^+$ which is observed by the interviewer.[21] Letting y denote the response rate of the sample, we assume that the cost of interviewing the households (via his effort) is separable and writes as $\theta_{ij} C_j(y)$. The interviewer receives $w_j + \delta_j$ from Insee when the interview is achieved and w_j otherwise. The program of interviewer i is thus similar to the agent's program we have been looking at and writes as

$$\max_y \delta_j y - \theta_{ij} C_j(y).$$

Finally, Insee is an institute depending on public money and maximizing the social value of each survey. We denote by λ_j the "price" of the information contained in a household's answers, i.e. the social value of an interview in survey j. Hence, the Insee's objective function, associated with an interviewer whose response rate is y, writes as

$$S(y, \lambda_j) = \lambda_j y$$

[21] Before trying to contact the households, interviewers must locate the housings of their sample (in order to identify unoccupied or destroyed housings, for instance). During this phase, the interviewer learns the difficulty θ_{ij} of his sample.

The data

We use data on three household living conditions surveys ("enquête Permanente sur les Conditions de Vie des Ménages", PCV hereafter) which took place in October 2001, 2002 and 2003 ($j = 1, 2, 3$). In 2001 and 2002, the focus of the survey was put respectively on the use of new technologies and participation in associations, while in 2003, the survey studied education practices in the family. As a consequence, almost all households were eligible to the survey in 2001 and 2002 whereas only families were eligible in 2003. Otherwise, the surveys were identical in their designs and rules for the fieldwork.

In all surveys, we restrict our attention to the housings where more than three persons lived at the time of the 1999 census. This information is indeed available to the interviewer before conducting the surveys and is a good proxy for the eligibility of the household in 2003. For these households, the bonus for achieving an interview increased from 20 and 20.2 euros in 2001 and 2002 to 23.4 euros in 2003. To avoid selection effects, we also focus on the interviewers who conduct the three surveys.

We interpret this change as a modification of the principal objective function. We believe that the 2003 survey on education was considered by Insee to be more important than the ones on new technologies and participation to associations. Indeed, there is much debate in France on the relationship between families, education and the emergence of inequalities (see for instance the report of the Haut Conseil de l'Education in 2007 on this topic). More formally, more publications from Insee and other institutions were based on this survey and the questionnaire was slightly longer in 2003. Given these elements, we believe that the social value of an interview was higher in 2003 ($\lambda_1 = \lambda_2 < \lambda_3$). However, because the surveys were drawn in the same way, conducted in the same period, were close in time and had identical rules for the fieldwork, we assume that for all interviewers i, the distributions of θ_{ij}, $j \in \{1, 2, 3\}$ are identical. We also suppose that $C'_1 = C'_2 = C'_3$, even if an interview in 2003 was a little longer. Indeed, the cost of achieving an interview is mainly due to contact tries rather than the length of the interview itself.[22] Under these hypotheses, the variation in the transfer function is exogenous, as defined in subsection 3.2.

Our data consists in the identification number of the interviewer and his response

[22]This claim is supported by the discussions that we had with interviewers.

	Interviewers	Bonus	Response rates	
			Mean	Std error
2001	236	20	79.6%	0.188
2002	236	20.2	80.4%	0.199
2003	236	23.4	83.8%	0.144

TAB. 3.1 – Descriptive statistics in 2001, 2002 and 2003 surveys.

rate, for each survey. The response rate of an interviewer is defined as the ratio of the number of respondents on the number of housings which are in the field of the survey (i.e., excluding secondary, unoccupied and destroyed housings for instance). Table 3.1 summarizes the main information about the surveys. It appears that the 2001 and 2002 surveys are very similar and we aggregate them in the rest of the application to obtain more precise results.[23] The 2003 survey is also similar to the other ones except for a higher response rate. Figure 3.6 displays more precisely the distribution function of the response rates for the 2001-2002 and 2003 surveys. As predicted by the theory, the distribution function of the 2003 survey stochastically dominates the one of 2001-2002, which proves that interviewers react to incentives.[24] We find that on average production increases by 5% when the piece rate increases by 16%. This result suggests a significant incentive effect, but relatively smaller than what the previous literature has found (Lazear, 2000 ; Paarsch and Shearer, 2000).

Figure 3.6 also displays several jumps in the distribution functions. We do not interpret them as evidence of bunching as the transfer functions do not exhibit kinks. These mass points are rather due to finite approximations of the response rates, and we neglect the error term in our estimation.

[23]The difference between the two average response rates is not significant at 5%, contrarily to differences between 2001 (or 2002) and 2003. In the following, we suppose that the two bonuses were identical and equal to 20.2 euros.

[24]The one-sided Kolmogorov-Smirnov test which tests the equality of the distribution functions rejects the null hypothesis at 5%.

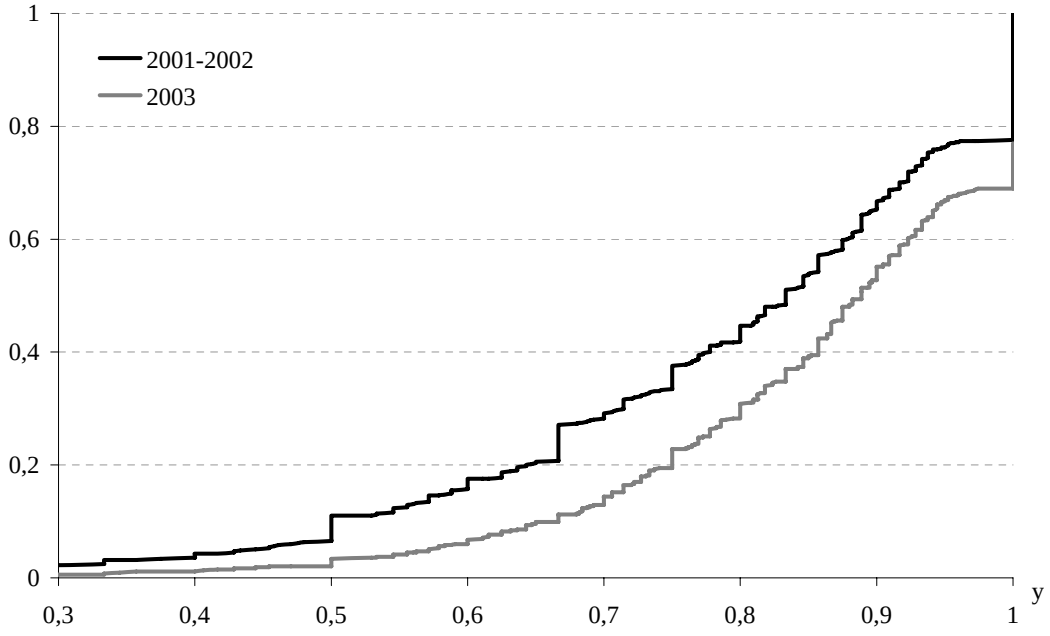

FIG. 3.6 – Distribution functions of the response rates for 2001-2002 and 2003 surveys.

Estimation

In this subsection, we define and study the estimators of the structural parameters in the framework of our application. We suppose that $K = 2$ and assume that $t(y, \lambda_j) = \delta_j y$.[25] Our asymptotic results rely on a standard assumption on the sample $(y_{ij})_{i\in\{1,...,N\},j\in\{1,2\}}$, which does not impose any dependency structure between θ_{i1} and θ_{i2}.[26]

Assumption 24 *(independent sampling)* $(\theta_{11}, ..., \theta_{N1})$ *(resp.* $(\theta_{12}, ..., \theta_{N2})$*) are independently drawn according to* F_θ*.*

By proposition 3.2.3, we identify the function $\theta(., \lambda_1)$ on an increasing sequence $(y_n)_{n\in\mathbb{Z}}$. More precisely (see equation (3.2.20) in the proof of proposition 3.2.3), we identify the sequence

$$y_n = H_{12}^n(y_0)\mathbb{1}_{H_{12}(y_{n-1})\in\mathcal{Y}_1} + y_{n-1}\mathbb{1}_{H_{12}(y_{n-1})\notin\mathcal{Y}_1}, \quad n \in \mathbb{Z},$$

[25] Actually, our results hold for any transfer function.

[26] We suppose for simplicity that the same agents are observed in both contracts.

where $H_{12}^n = H_{12} \circ ... \circ H_{12}$ for any $n \in \mathbb{N}$ (and similarly, $H_{12}^{-n} = H_{12}^{-1} \circ ... \circ H_{12}^{-1}$). Moreover, using $\theta(H_{12}(y), \lambda_1) = V_{21}(\theta(y, \lambda_1), H_{12}(y))$ and by induction, $\theta_n = \theta(y_n, \lambda_1)$ satisfies :

$$\theta_n = (\delta_1/\delta_2)^n \theta_0 \mathbb{1}_{H_{12}(y_{n-1}) \in \mathcal{Y}_1} + \theta_{n-1} \mathbb{1}_{H_{12}(y_{n-1}) \notin \mathcal{Y}_1},$$

where θ_0 is chosen arbitrarily.

These sequences are estimated as follows. For $j \in \{1, 2\}$, let $\widehat{F}_j$ (resp. $\widehat{F}_j^{-1}$) denote the empirical distribution function (resp. empirical quantile function) of the $(y_{ij})_{i \in \{1,...,N\}}$. Our estimator of H_{12} is

$$\widehat{H}_{12}(x) = \widehat{F}_2^{-1} \circ \widehat{F}_1(x).$$

For all $n \in \mathbb{Z}$, we estimate y_n by

$$\widehat{y}_n = \widehat{H}_{12}^n(y_0) \mathbb{1}_{\widehat{H}_{12}(\widehat{y}_{n-1}) \in \widehat{\mathcal{Y}_1}} + \widehat{y}_{n-1} \mathbb{1}_{\widehat{H}_{12}(\widehat{y}_{n-1}) \notin \widehat{\mathcal{Y}_1}},$$

where $\widehat{\mathcal{Y}_1} = [\min_i y_{i1}, \max_i y_{i1}]$, and θ_n by

$$\widehat{\theta}_n = (\delta_1/\delta_2)^n \theta_0 \mathbb{1}_{\widehat{H}_{12}(\widehat{y}_{n-1}) \in \widehat{\mathcal{Y}_1}} + \widehat{\theta}_{n-1} \mathbb{1}_{\widehat{H}_{12}(\widehat{y}_{n-1}) \notin \widehat{\mathcal{Y}_1}}.$$

Now, let us now turn to F_θ and C'. These functions are point identified respectively on $(\theta_n)_{n \in \mathbb{Z}}$ and $(y_n)_{n \in \mathbb{Z}}$ and we focus on the estimation of $F_\theta(\theta_n)$ and $C'(y_n)$. First, using (3.2.10), $F_\theta(\theta_n)$ satisfies $F_\theta(\theta_n) = 1 - F(y_n,\, \lambda_1)$. Hence, we define $\widehat{F_\theta(\theta_n)}$ by

$$\widehat{F_\theta(\theta_n)} = 1 - \widehat{F}_1(\widehat{y}_n).$$

Similarly, by 3.2.9, $C'(y_n) = \delta_1/\theta_n$. Thus, we consider the following estimator :

$$\widehat{C'(y_n)} = \frac{\delta_1}{\widehat{\theta}_n}.$$

The following theorem establishes the consistency of $(\widehat{\theta}_n, \widehat{F_\theta(\theta_n)}))$ and $(\widehat{y}_n, \widehat{C'(y_n)})$. The result relies on assumption 25, which prevents any y_n from being on the boundary of $\mathcal{Y}_1$. Since the measure of this boundary is zero, this assumption is unrestrictive.

Assumption 25 *For all $n \in \mathbb{Z}$, $y_n \notin \mathcal{Y}_1 \backslash \overset{\circ}{\mathcal{Y}_1}$.*

Theorem 3.2.9 *Suppose that $K = 2$, $t(y, \lambda_j) = \delta_j y$ and assumptions 18-20, 24-25 hold. Then, for all $n \in \mathbb{Z}$ and when $N \to +\infty$,*

$$(\widehat{\theta_n}, \widehat{F_\theta(\theta_n)}) \xrightarrow{\mathbb{P}} (\theta_n, F_\theta(\theta_n))$$
$$(\widehat{y_n}, \widehat{C'(y_n)}) \xrightarrow{\mathbb{P}} (y_n, C'(y_n))$$

By the monotonicity of F_θ and C', the estimators of $F_\theta(\theta_n)$ and $C'(y_n)$ enable us to build bounds on the two functions.

Results

We now apply this estimation method on our contract data. Firstly, starting from a middle point $y_0 = 0.6$ (with $\theta_0 = 1$), we estimate $(y_n, \theta_n, F_\theta(\theta_n), C'(y_n))_{n\in\mathbb{Z}}$ as indicated and obtained 12 distinct points which correspond to $n \in \{-3, ..., 8\}$. Figures 3.7 and 3.8 display the estimations of the bounds on $F_\theta(.)$ and $C'(.)$ respectively, and their 95% confidence interval obtained by bootstrap. With twelve points, the bounds on both functions are close and we are able to correctly retrieve their shape. The highly convex form of the cost function shows in particular that incentives are relatively large for small values of the production but sensitively lower for higher ones. This may explain the small average effect of incentives that we have found compared to the previous results of the literature. We also note that, as expected, the errors accumulate in the estimation procedure and that the width of the confidence intervals on the bounds of F_θ (resp. C') increases with $|\theta - 1|$ (resp. $|y - 0.6|$).

Secondly, this information is used to estimate the objective function of the principal $S(y, \lambda_3) = \lambda_3 y$ in 2003. Using equation 3.2.19 in appendix A, it can be shown that λ_3 satisfies for all $n \neq 0$:

$$\lambda_3 = \delta_3 \left(1 + \frac{\ln(\theta_n/\theta_0)}{\ln(F_\theta(\theta_n)/F_\theta(\theta_0))}\right)$$

We then obtain eleven different estimators of λ_3, for $n \in \{-3, 8\}$, $n \neq 0$:

$$\widehat{\lambda}_{3n} = \delta_3 \left(1 + \frac{\ln(\widehat{\theta_n}/\theta_0)}{\ln(\widehat{F_\theta(\theta_n)}/\widehat{F_\theta(\theta_0)})}\right)$$

and the model is overidentified (see the discussion following proposition 3.2.3). Figure 3.9 depicts these estimators and their associated confidence interval calculated by bootstrap. We see that $\widehat{\lambda}_{3n}$ does not appear to be constant in n, which

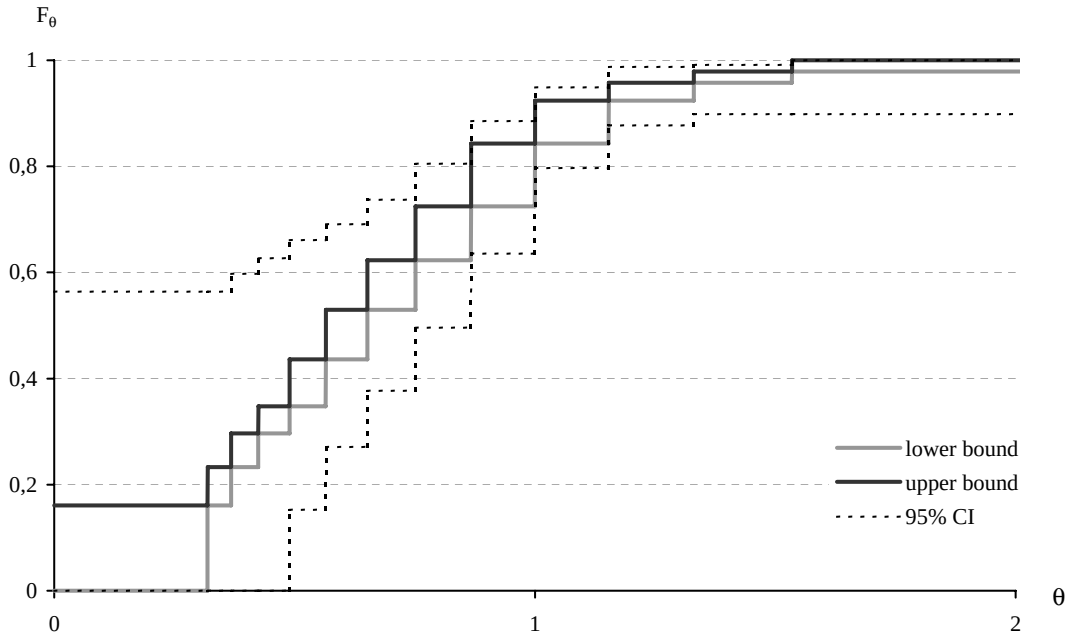

FIG. 3.7 – Estimated bounds on $F_\theta(.)$

contradicts the contracts' optimality. More formally, assuming asymptotic normality, we compute the Wald statistic of the test $\lambda_{-3} = ... = \lambda_8$. This statistic equals 65.5 and should be compared to the 95% quantile of a $\chi^2(11)$, i.e., 19.7. Hence, we clearly reject the contracts' optimality at the 5% level (p-value$\simeq 10^{-9}$).

The rejection of contracts' optimality is also reinforced by the violation of the Informativeness Principle which states that all factors correlated with performance should be included in the contracts (Prendergast, 1999). In our application, contracts are not optimal as the bonus does not depend for instance on the region in which interviewers are working, even if the average response rate in Paris area (0.73% in 2003) is significantly lower than in the rest of France (0.85%).

To understand why Insee designs linear contracts, it is interesting to measure the cost of using simple but inefficient incentives. We thus compare the actual surplus of the institute with the one it would obtain under the use of optimal contracts. We also compute the surplus of Insee under complete information and estimate the cost of asymmetric information. To do so, we need to recover λ_3 in the objective function of Insee. However, under the assumption that Insee maximizes his objective function only in the class of linear contracts, equation (3.2.7) is not

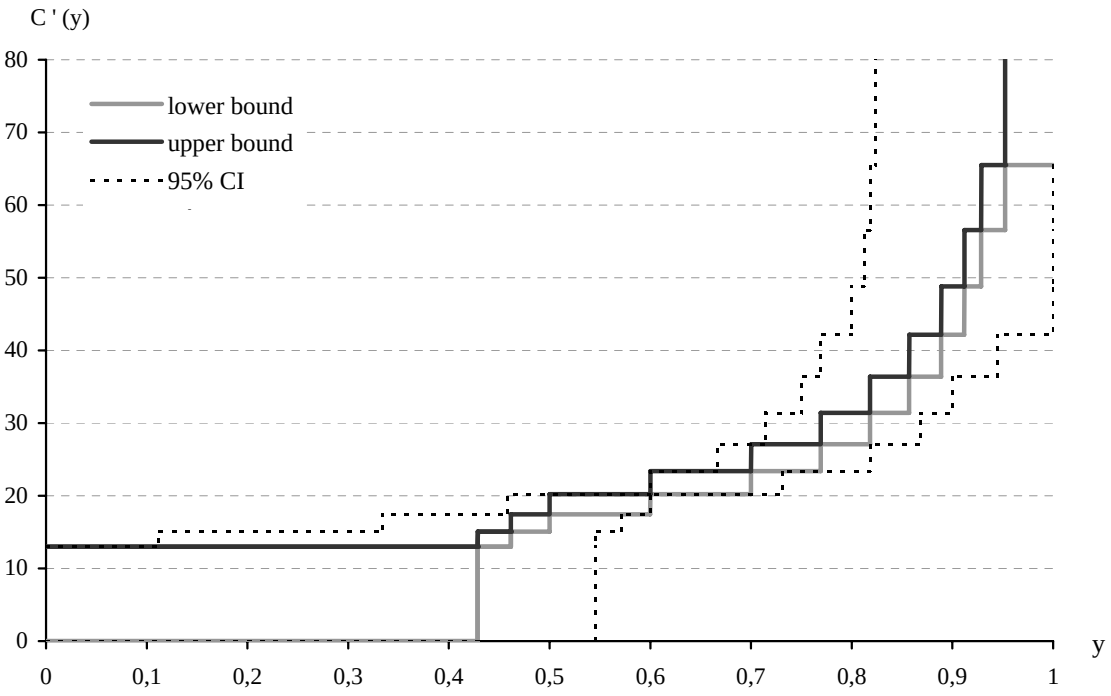

FIG. 3.8 – Estimated bounds on $C'(.)$.

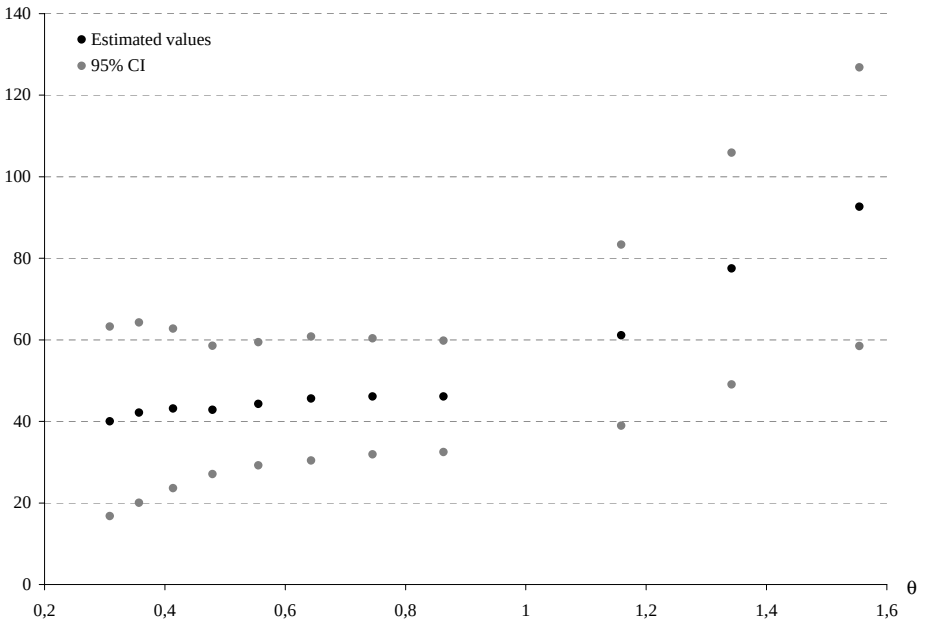

FIG. 3.9 – Estimated $\widehat{\lambda}_{3n}$, $n \in \{-3, 8\}$, $n \neq 0$.

valid anymore. It can be shown, in this case, that λ_3 is not identified even with $K = 2$. We thus adopt a parametric approach and impose a structure coherent with our previous nonparametric analysis. More precisely, we suppose that θ follows a Weibull distribution $F_\theta(\theta) = 1 - \exp(-a\theta^b)$ for all $\theta \in \mathbb{R}^+$ and that the cost function takes the form $C'(y) = \alpha\,(y/1-y)^\beta$ on $[0, 1[$. The parameters of interest are estimated by regressing $\ln(-\ln(\widehat{F}_j(y)))$ on $\ln[(1-y)/y]$ (see appendix B). We obtain $\widehat{a} = 1.87\,(0.03)$, $\widehat{b} = 1.89\,(0.04)$, $\widehat{\alpha} = 16.7\,(0.07)$ and $\widehat{\beta} = 0.46\,(0.01)$. Figure 3.10 shows that these parametric forms perfectly fits the nonparametric estimated points.

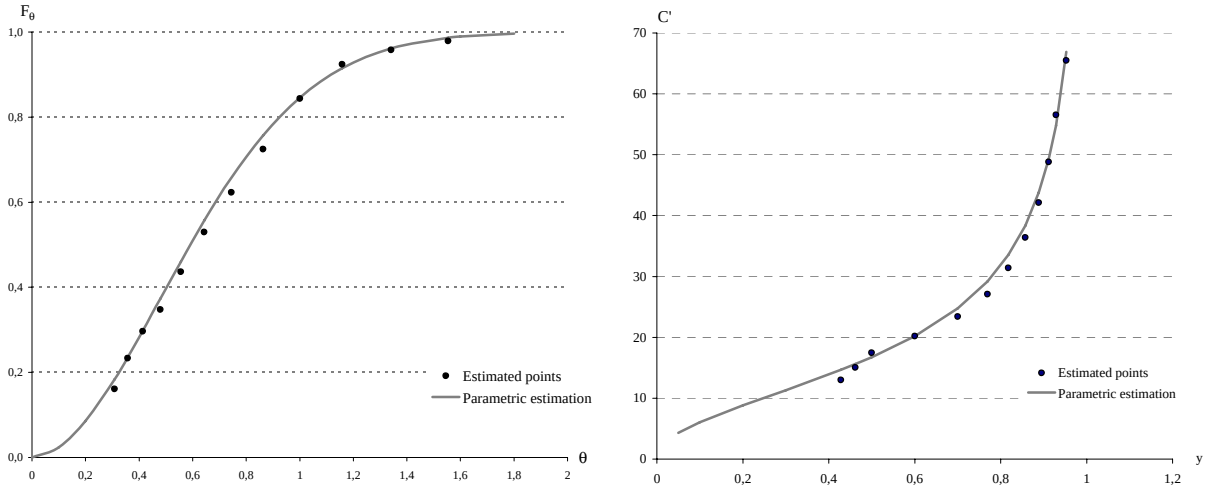

FIG. 3.10 – Parametric estimation of F_θ and C'.

Under these parametric restrictions, we are able (see appendix B) to estimate λ_3 and find that the social value of an interview to be $\widehat{\lambda}_3 = 102.1$ euros. We also compute the expected surplus under full information and the expected surplus under incomplete information with linear or optimal contracts. Table 3.2 summarizes our results. We find that the surplus loss associated with the use of linear contracts is about 9% and that the response rate also decreases by 9% compared to optimal contracts. This result contrasts with the previous literature in which linear contracts were thought to be quite inefficient. Ferrall and Shearer (1999), for instance, evaluate this loss at 50%. Our results point out on the contrary that the cost is quite small and that optimal contracts are not highly nonlinear. This may explain why firms widely use linear contracts compared to nonlinear ones :

they are less costly to implement and almost efficient.

Environment	Pay method	E[surplus]	Relative	E[response rate]
Full information	Optimal contract	86.0	1.00	0.99
Incomplete information	Optimal contract	72.5	0.84	0.92
Incomplete information	Linear contract	65.9	0.77	0.84

TAB. 3.2 – Surplus and response rates under alternative compensation schemes.

Finally, we find moderate cost of incomplete information. The surplus under asymmetric information is 77% of what it could be under complete information. Two third of this loss is due to incomplete information whereas only one third is associated with the simple bonus system.

3.2.5 Conclusion

This work contributes to the recent structural analysis of incentive problems. First, by focusing on the general adverse selection model, we complement Perrigne and Vuong (2004)'s paper. Our result that these models are not fully identified is important to understand what restrictions are needed to recover the functions of interest in different settings such as regulation, nonlinear pricing or taxation. Second, we propose a way to exploit exogenous changes in order to identify or test nonparametrically the model and the contracts' optimality. The recursive method we develop enables us to analyze existing experiments or natural experiments. To our knowledge, such a recursive method is new. A consequence is that the econometric procedure, which is based on this method, is also a novelty. Third, studying the provision of incentives in firms, we test nonparametrically and reject the contracts' optimality proposed by Insee. We also estimate Insee's surplus to be 77% of the full information surplus. Two third of this loss correspond to the cost associated with asymmetric information, whereas the use of inefficient linear contracts only explains one third of it. This result may explain why firms widely use these contracts that are also very easy to implement.

Beyond these estimations, this approach can be useful to firms for determining the optimal contracts that they should implement. By proposing different contracts to random samples of the population, the firm will learn the structural parameters

of its agents and will then be able to design the optimal contract.

The paper also raises several challenging issues. Firstly, the properties of our estimators with three or more different contracts remain to be established. This may be difficult because identification is obtained at the limit with a density argument. Besides, one can wonder whether our strategy could be adapted to the moral hazard setting. Testable implications of this model have already been brought to light (Abbring et al., 2003 ; Chiappori et al., 2006), but its nonparametric identifiability has not been settled yet.

3.2.6 Appendix A : proofs

Proof of proposition 3.2.1

The revelation principle (Myerson, 1981) ensures that there is no loss of generality in restricting our analysis to truthful direct revelation mechanisms $(y(\theta), t(\theta); \theta \in \Theta)$ when studying optimal contracts under asymmetric information. A direct mechanism is such that the principal commits to offer the transfer $t(\tilde{\theta})$ and the production $y(\tilde{\theta})$ if the agent announces a type $\tilde{\theta}$. The direct revelation mechanism is truthful if the agent's best response is to announce his true type :

$$\forall \theta \in \Theta,\ \theta = \arg\max_{\tilde{\theta}}\ t(\tilde{\theta}) - C(y(\tilde{\theta}), \theta)$$

Restricting the analysis to differentiable functions $y(.)$ and $t(.)$, these conditions rewrite as

$$t'(\theta) = y'(\theta)\frac{\partial C}{\partial y}(y(\theta), \theta)$$
$$0 \geq y'(\theta)\frac{\partial^2 C}{\partial \theta \partial y}(y(\theta), \theta)$$

The first equation corresponds to the first order condition. The second inequality corresponds to the second order condition and ensures that the optimum defined by the first equation is a maximum. Under assumption 18, these conditions rewrite as :

$$t'(\theta) = y'(\theta)\frac{\partial C}{\partial y}(y(\theta), \theta)$$
$$y'(\theta) \leq 0$$

Integrating the first equation and using $t(\overline{\theta}) = C(y(\overline{\theta}), \overline{\theta})$ (i.e. the utility of the less efficient type is equal to zero), we obtain :

$$t(\theta) = C(y(\theta), \theta) + \int_{\theta}^{\overline{\theta}} \frac{\partial C}{\partial \theta}(y(\tau), \tau) d\tau \geq C(y(\theta), \theta)$$

$\int_{\theta}^{\overline{\theta}} \frac{\partial C}{\partial \theta}(y(\tau), \tau) d\tau$ is the informational rent that the principal has to give to the agent of type θ for him to reveal his type. This rent decreases with θ.

The optimization program of the principal writes as

$$\max_{y(.),t(.)} \int_{\underline{\theta}}^{\overline{\theta}} [S(y(\theta),\theta) - t(\theta)] f_\theta(\theta) d\theta$$

Using equation (3.2.4), the principal's objective function becomes

$$\max_{y(.)/y'(.)\leq 0} \int_{\underline{\theta}}^{\overline{\theta}} \left[S(y(\theta),\theta) - C(y(\theta),\theta) - \int_{\theta}^{\overline{\theta}} \frac{\partial C}{\partial \theta}(y(\tau),\tau) d\tau \right] f_\theta(\theta) d\theta$$

$$=$$

$$\max_{y(.)/y'(.)\leq 0} \int_{\underline{\theta}}^{\overline{\theta}} \left[S(y(\theta),\theta) - C(y(\theta),\theta) - \frac{\partial C}{\partial \theta}(y(\theta),\theta) \frac{F_\theta(\theta)}{f_\theta(\theta)} \right] f_\theta(\theta) d\theta$$

Ignoring the constraint on $y'(.)$ and maximizing pointwise, we get :

$$\frac{\partial S}{\partial y}(y(\theta),\theta) = \frac{\partial C}{\partial y}(y(\theta),\theta) + \frac{F_\theta(\theta)}{f_\theta(\theta)} \frac{\partial^2 C}{\partial \theta \partial y}(y(\theta),\theta)$$

Actually, the constraint on $y'(.)$ is not binding under assumption 19. Indeed, if $y(.)$ defined by equation (3.2.3) were not decreasing, there would be some bunching in the optimal contract (see Laffont and Martimort, 2002).

The term $\frac{F_\theta(\theta)}{f_\theta(\theta)} \frac{\partial^2 C}{\partial \theta \partial y}(y(\theta),\theta)$ is the deviation from the first best. It can be interpreted as the price paid by the firm to incite the efficient types to reveal their types.■

Proof of proposition 3.2.2

The first order condition of the agent writes

$$C'(y) = \frac{t'(y)}{\theta(y)} \tag{3.2.16}$$

Besides, by monotonicity of $\theta(.)$,

$$F_y(y) = 1 - F_\theta(\theta(y)) \tag{3.2.17}$$

By assumption 20, the first order condition of the principal writes as

$$\widetilde{S}(y) = \theta(y) C'(y) + \frac{F_\theta(\theta(y))}{f_\theta(\theta(y))} C'(y).$$

Thus, by (3.2.16),

$$\widetilde{S}(y) = \left[1 + \frac{F_\theta(\theta(y))}{\theta(y) f_\theta(\theta(y))} \right] t'(y).$$

Moreover, by (3.2.17), $F_\theta(\theta(y)) = 1 - F_y(y)$ and $f_\theta(\theta(y)) = -f_y(y)/\theta'(y)$. Thus,

$$\widetilde{S}(y) = \left[1 - \frac{1 - F_y(y)}{f_y(y)} \frac{\theta'}{\theta}(y)\right] t'(y). \tag{3.2.18}$$

Now, for any strictly decreasing and differentiating function $\theta(.)$, it is possible to define $C'(.)$, $F_\theta(.)$ and $\widetilde{S}$ using respectively equation (3.2.16), (3.2.17) and (3.2.18). For the model to be not identified, it is sufficient to prove that two sets of such functions satisfy assumption 18. Because assumption 18 is satisfied for the true function $\theta^0(.)$, we conclude that the assumption is satisfied also locally around $\theta^0(.)$. Hence, the model is not identified.

If $C'(.)$ (resp. $F_\theta(.)$) is known, equation 3.2.16 (resp 3.2.17) enables to identify $\theta(.)$ on $\mathcal{Y}$. Then, $F_\theta(.)$ (resp C') is identified on Θ (resp. $\mathcal{Y}$) using the previous equations. Finally, $\widetilde{S}$ is identified by (3.2.18).

Lastly, if $\widetilde{S}$ is known, (3.2.18) also writes as

$$\ln(\theta(y)) = \ln(\theta_0) + \int_{y_0}^{y} \left(1 - \frac{\widetilde{S}(u)}{t'(u)}\right) \frac{f_y(u)}{1 - F_y(u)} du \tag{3.2.19}$$

Hence $\theta(.)$ is identified. By (3.2.16) and (3.2.17), C' and F_θ are also identified. ■

Proof of theorem 3.2.3

Firstly, we prove that $\theta(., \lambda_1)$ is identified on the closure of a sequence $(y_n)_{n\in\mathbb{Z}}$. If $\theta(y, \lambda_1)$ is known and if $H_{12}(y) \in \mathcal{Y}_1$ then $\theta(H_{12}(y), \lambda_1) = V_{21}(\theta(y, \lambda_1), H_{12}(y))$ is identified because H_{12} and V_{21} are.
Hence, using $y_0 \in \mathcal{Y}_1$, we deduce by induction that $\theta(., \lambda_1)$ is identified on the increasing sequence[27]

$$y_n = H_{12}^n(y_0)\mathbb{1}_{H_{12}(y_{n-1})\in\mathcal{Y}_1} + y_{n-1}\mathbb{1}_{H_{12}(y_{n-1})\notin\mathcal{Y}_1} \tag{3.2.20}$$

for any $n \geq 0$. Similarly, $\theta(., \lambda_1)$ is identified on the sequence

$$y_n = H_{21}^{-n}(y_0)\mathbb{1}_{H_{12}(y_{-n+1})\in\mathcal{Y}_2} + y_{-n+1}\mathbb{1}_{H_{12}(y_{-n+1})\notin\mathcal{Y}_2}$$

for any $n \leq 0$. By continuity of $\theta(., \lambda_1)$, the function is actually identified on the closure of $\{y_n, n \in \mathbb{Z}\}$.

[27] f^n denotes $f \circ ... \circ f$ and not $f \times ... \times f$ for any function f and $n \in \mathbb{N}$. Similarly, $f^{-n} = f^{-1} \circ ... \circ f^{-1}$.

Then, using the property that $\theta(.,\lambda_1)$ is decreasing, we obtain that for any $y_{n-1} < x < y_n$, $\theta(y_{n+1},\lambda_1) < \theta(x,\lambda_1) < \theta(y_n,\lambda_1)$ and similarly around the bounds.

Using this results and equations (3.2.16) and (3.2.17), we obtain that $C'(.)$ and $F_\theta(.)$ are identified respectively on the sequences $(y_n)_{n\in\mathbb{Z}}$ and $(\theta(y_n,\lambda_1)_{n\in\mathbb{Z}}$ whereas bounds are obtained for other values.

Now let us prove that $\theta(.,\lambda_1)$ is not identified outside of the sequence $(y_n)_{n\in\mathbb{Z}}$. As in proposition 3.2.2, we derive from this result that $C'(.)$ and $F_\theta(.)$ are not identified outside the sequences $(y_n)_{n\in\mathbb{Z}}$ and $(\theta(y_n,\lambda_1)_{n\in\mathbb{Z}}$. Furthermore, $\theta'(.,\lambda_1)$ is identified nowhere. By (3.2.18), this proves that the functions $\widetilde{S}(.,\lambda_1)$ and $\widetilde{S}(.,\lambda_2)$ are not nonparametrically identified.

To prove that $\theta(.,\lambda_1)$ is not identified outside of the $(y_n)_{n\in\mathbb{Z}}$, we show that defining $\theta(.,\lambda_1)$ on $\mathcal{Y}_1$ can be reduced by the first order conditions to defining this function on $]y_0,y_1[$ and that almost no restriction can be imposed on this interval.[28]
Suppose indeed that $\theta(.,\lambda_1)$ is a known and differentiable function on the interval $]y_0,y_1[$. For any y in this interval, we are able to construct a sequence $(\tilde{y}_n)_{n\in\mathbb{Z}}$ with $\tilde{y}_0 = y$, just as we defined the sequence $(y_n)_{n\in\mathbb{Z}}$. Hence, $\theta(.,\lambda_1)$ is defined on all these sequences i.e. on $\mathcal{Y}_1$. Furthermore, this function is differentiable everywhere except eventually at points $(y_n)_{n\in\mathcal{Z}}$. Using $\theta(H_{12}(y),\lambda_1) = V_{21}(\theta(y,\lambda_1),H_{12}(y))$, the differentiability of the function at these points is ensured as soon as

$$\begin{aligned} H'_{12}(y_0)\theta'_-(y_1,\lambda_1) &= H'_{12}(y_0)\theta'_+(y_1,\lambda_1) \\ &= H'_{12}(y_0)\frac{\partial V_{21}}{\partial y}(\theta(y_0,\lambda_1),y_1) + \theta'_+(y_0,\lambda_1)\frac{\partial V_{21}}{\partial \theta}(\theta(y_0,\lambda_1),y_1) \end{aligned}$$

where $\theta'_-(y_1,\lambda_1)$ (resp. $\theta'_+(y_1,\lambda_1)$) is the left (resp. right) derivative of $\theta(.,\lambda_1)$ in $y = y_1$.

Finally, consider all the functions $\theta(.,\lambda_1)$ defined on $]y_0,y_1[$ that can be extended to differentiable function on $\mathcal{Y}_1$. By construction, all these functions are coherent with respect to the horizontal and vertical transforms. Hence, $\theta(.,\lambda_1)$ is not identified and proposition 3.2.3 follows. ■

[28] At least locally around the true function for which assumption 18 is satisfied.

Proof of theorem 3.2.4

By the normalization and the fact that $y_c \in \overset{\circ}{\mathcal{Y}}_1$, we can always fix $0 < \theta_c < \infty$ such that $\theta_c = \theta(y_c, \lambda_1)$. Let y'_c denote the greatest intersection point of $\partial t/\partial y(., \lambda_1)$ and $\partial t/\partial y(., \lambda_2)$ which is smaller than y_c ($y'_c = \inf \mathcal{Y}_1$ if such a point does not exists). We suppose without loss of generality that $\partial t/\partial y(., \lambda_2) > \partial t/\partial y(., \lambda_1)$ on (y'_c, y_c).

Let $y'_c < y_0 < y_c$, and define the increasing sequence $(y_n)_{n\in\mathbb{N}}$ by

$$y_n = H_{12}^n(y_0)\mathbb{1}_{H_{12}(y_{n-1})\in\mathcal{Y}_1} + y_{n-1}\mathbb{1}_{H_{12}(y_{n-1})\notin\mathcal{Y}_1}.$$

We see that $y_n < y_c$ for all $n \in \mathbb{N}$. Indeed, the result is true for $n = 0$. Moreover, if it holds for $n-1$, then $y_n = H_{12}(y_{n-1}) < H_{12}(y_c) = y_c$ since H_{12} is strictly increasing. Hence, for all $n \in \mathbb{N}$, $y_n \in \mathcal{Y}_1$ and $y_n = H_{12}(y_{n-1})$. The sequence is increasing and bounded above by y_c, so it admits a limit y_∞ which satisfies $y_\infty = H_{12}(y_\infty)$. Hence $y_\infty = y_c$.

Now let θ_n be such that $\theta_n = \theta(y_n, \lambda_1)$. Note that because we fix θ_c and not θ_0, we cannot apply the proof of proposition 3.2.3 to show that θ_n is identified. Let us prove that θ_0 is identified.

First,

$$\theta_{n+1} = V_{21}(\theta_n, y_{n+1}) = \frac{\frac{\partial t}{\partial y}(y_{n+1}, \lambda_1)}{\frac{\partial t}{\partial y}(y_{n+1}, \lambda_2)}\theta_n.$$

Thus, by induction,

$$\theta_n = \prod_{i=1}^{n} \left[\frac{\frac{\partial t}{\partial y}(y_i, \lambda_1)}{\frac{\partial t}{\partial y}(y_i, \lambda_2)}\right] \theta_0.$$

Because $(y_n)_{n\in\mathbb{N}}$ converges to y_c and $\theta(., \lambda_1)$ is continuous, the sequence $(\theta_n)_{n\in\mathbb{N}}$ converges to $\theta_\infty = \theta_c$. Hence,

$$\theta_c = \prod_{i=1}^{\infty} \left[\frac{\frac{\partial t}{\partial y}(y_i, \lambda_1)}{\frac{\partial t}{\partial y}(y_i, \lambda_2)}\right] \theta_0.$$

Because $0 < \theta_c < \infty$, the product in the right hand side is strictly positive and finite and we have

$$\theta_0 = \frac{\theta_c}{\prod_{i=1}^{\infty} \left[\frac{\frac{\partial t}{\partial y}(y_i, \lambda_1)}{\frac{\partial t}{\partial y}(y_i, \lambda_2)}\right]}.$$

Hence, θ_0 is identified since the right term can be recovered from the data. y_0 was arbitrary, so $\theta(., \lambda_1)$ is identified on (y'_c, y_c). Starting from θ_0 , we can also identify

an increasing sequence $(\theta'_n)_{n\in\mathbb{N}}$ which converges to θ'_c such that $y(\theta'_c, \lambda_1) = y'_c$. This proves that $\theta(., \lambda_1)$ is actually identified on $[y'_c, y_c]$. The same reasoning can be applied between y'_c and another crossing point y''_c. By repeating this as much as necessary, we can identify $\theta(., \lambda_1)$ on $\mathcal{Y}_1$.

Finally, by proposition 3.2.2, F_θ is identified on Θ and C' and $\widetilde{S}(., \lambda_1)$ are identified on $\mathcal{Y}_1$. Furthermore, by the horizontal transformation, $\theta(., \lambda_2)$ is also identified on $\mathcal{Y}_2$. Hence, $\widetilde{S}(., \lambda_2)$ is identified on $\mathcal{Y}_2$, and C' is identified on $\mathcal{Y}_1 \cup \mathcal{Y}_2$. ■

Proof of proposition 3.2.5

$\widetilde{\mathbb{Y}}$ is the set of $\cup_{i=1,..,K}\mathcal{Y}_i$ such that the functions $\theta(., \lambda_i)$ are identified on $\widetilde{\mathbb{Y}} \cap \mathcal{Y}_i$. This set is defined by induction using the horizontal and vertical transformations.

Suppose that $y \in \widetilde{\mathbb{Y}} \cap \mathcal{Y}_i$ and that the point $(y, \theta(y, \lambda_i))$ is identified.

– For all j, $H_{ij}(y)$ and $\theta(H_{ij}(y), \lambda_j) = \theta(y, \lambda_i)$ are known because H_{ij} is identified. Hence the points $(H_{ij}(y), \theta(H_{ij}(y), \lambda_j))$ are identified.
– For all j such that $y \in \mathcal{Y}_j$, $\theta(y, \lambda_j) = V_{ij}(\theta(y, \lambda_i), y)$ is known because V_{ij} is identified. Hence the points $(y, \theta(y, \lambda_j))$ are identified if $y \in \mathcal{Y}_j$.

This method defines by induction $\widetilde{\mathbb{Y}}$, which corresponds to the set $\mathbb{Y}$ of the proposition. Furthermore, let $y \in \overline{\mathbb{Y}} \cap \mathcal{Y}_i$ and $y_n \in \mathbb{Y} \cap \mathcal{Y}_i$ such that $y_n \to y$. Then, by continuity of $\theta(., \lambda_i)$, we get

$$\theta(y, \lambda_i) = \lim_{n\to\infty} \theta(y_n, \lambda_i).$$

$\theta(., \lambda_i)$ is actually identified on $\overline{\mathbb{Y}} \cap \mathcal{Y}_i$. ■

Proof of theorem 3.2.6

Let us define $\widetilde{\Theta} = \theta(\mathbb{Y} \cap \mathcal{Y}_1, \lambda_1)$, $E_1 = l(\lambda_1)/l(\lambda_2)$, $E_3 = l(\lambda_3)/l(\lambda_2)$ and $\Theta = [\underline{\theta}, \overline{\theta}]$. The result is based on the following two lemmas.

Lemma 3.2.1 *Suppose that*

$$\{E_1^{m_1} E_3^{m_3} \theta_0,\ (m_1, m_3) \in \mathbb{Z}^2\} \cap \Theta \subset \widetilde{\Theta}. \tag{3.2.21}$$

Then $\overline{\widetilde{\Theta}} = \Theta$.

Lemma 3.2.2 *If* $\overline{\widetilde{\Theta}} = \Theta$, *then* $\theta(., \lambda_1)$ *is identified on* $\mathcal{Y}_1$.

Proof of lemma 3.2.1 : let us introduce

$$G = \{m_1 \ln(E_1) + m_3 \ln(E_3), (m_1, m_3) \in \mathbb{Z}^2\}.$$

G is an additive subgroup of $\mathbb{R}$. Thus, it is either of the form $a\mathbb{Z}$ or dense in $\mathbb{R}$. By assumption 22, $\ln(E_1)/\ln(E_3)$ is irrational. Thus G is dense in $\mathbb{R}$. By continuity of the exponential, $\{E_1^{m_1} E_3^{m_3} \theta_0,\ (m_1, m_3) \in \mathbb{Z}^2\}$ is dense in $\mathbb{R}^+$. Hence, by (3.2.21), $\Theta \subset \overline{\widetilde{\Theta}}$. The other inclusion is obvious, so $\overline{\widetilde{\Theta}} = \Theta$. □

Proof of lemma 3.2.2 : by continuity of $y(., \lambda_1)$, the inverse of $\theta(., \lambda_1)$, we get

$$y(\overline{\widetilde{\Theta}}, \lambda_1) \subset \overline{y(\widetilde{\Theta}, \lambda_1)} = \overline{\mathbb{Y}} \cap \mathcal{Y}_1.$$

Thus, if $\overline{\widetilde{\Theta}} = \Theta$, $\mathcal{Y}_1 = y(\overline{\widetilde{\Theta}}, \lambda_1) \subset \overline{\mathbb{Y}}$ and by proposition 3.2.5, $\theta(., \lambda_1)$ is identified on $\mathcal{Y}_1$. □

Now, let us come back to the proof of the theorem. By assumption 23, there exists $y_0 \in \mathcal{Y}_1 \cap \mathcal{Y}_3$ such that $H_{23}(y_0) \in \mathcal{Y}_2$ and $H_{21}(y_0) \in \mathcal{Y}_2$. Let θ_0 satisfy $\theta(y_0, \lambda_2) = \theta_0$. By assumption 21, the vertical transforms write as $V_{2i}(\theta, y) = E_i\theta$ (for $i = 1, 3$) provided that $y \in \mathcal{Y}_2 \cap \mathcal{Y}_i$. Moreover, by definition of y_0, $\theta(y_0, \lambda_i) = E_i\theta_0$ is well defined (i.e., $E_i\theta_0 \in \Theta$). Hence,

$$\theta(H_{31}(y_0), \lambda_1) = E_3\theta_0.$$

In other terms, $E_3\theta_0 \in \widetilde{\Theta}$. Similarly, $E_1\theta_0 \in \widetilde{\Theta}$.

Moreover, $H_{23}(y_0) \in \mathcal{Y}_2 \cap \mathcal{Y}_3$, so that $V_{32}(\theta_0, H_{23}(y_0)) = E_3^{-1}\theta_0$ is well defined. Hence,

$$\theta(H_{21}(H_{23}(y_0)), \lambda_1) = E_3^{-1}\theta_0.$$

Thus, $E_3^{-1}\theta_0 \in \widetilde{\Theta}$ (and similarly for $E_1^{-1}\theta_0$).

Now, for all $n \geq 1$, define

$$A_n = \{E_1^{m_1} E_3^{m_3} \theta_0,\ |m_1| + |m_3| = n\} \cap \Theta$$

Let us show by induction on n that $A_n \subset \widetilde{\Theta}$. The result is true for $n = 1$ by the preceding. Suppose that it is true for n and let $\theta = E_1^{m_1} E_3^{m_3} \theta_0 \in A_{n+1}$, with $\theta \geq \theta_0$ (the case where $\theta < \theta_0$ is similar). We have to show that $\theta \in \widetilde{\Theta}$.

Suppose first that $m_1 < 0$. Because $\theta > E_1\theta \geq E_1\theta_0 \geq \underline{\theta}$, $E_1\theta \in \Theta$. Moreover,

$$E_1\theta = E_1^{m_1+1} E_3^{m_3} \theta_0, \quad |m_1 + 1| + |m_3| = (-m_1 - 1) + |m_3| = n.$$

Thus $E_1\theta \in A_n \subset \widetilde{\Theta}$ by the induction hypothesis. By definition of $\widetilde{\Theta}$, there exists $y \in \mathbb{Y}$ such that $\theta(y, \lambda_1) = E_1\theta$. $\theta = V_{12}(E_1\theta, y)$ is well defined because $\theta \in \Theta$. This implies that $H_{21}(y) \in \mathbb{Y} \cap \mathcal{Y}_1$. Moreover, $\theta(H_{21}(y), \lambda_1) = \theta$. Thus, $\theta \in \widetilde{\Theta}$.

Similarly, if $m_1 \geq 0$ and $m_3 > 0$, the same reasoning applies, using $E_3^{-1}\theta$ instead of $E_1\theta$.[29] $A_{n+1} \subset \widetilde{\Theta}$ and the result is true for all n. Hence, $\cup_n A_n \subset \widetilde{\Theta}$, so that (3.2.21) holds.

Finally, by lemmas 3.2.1 and 3.2.2, $\theta(., \lambda_1)$ is identified on $\mathcal{Y}_1$. We conclude as in theorem 3.2.4. ■

Proof of proposition 3.2.7

By equations (3.2.5) and (3.2.6), $\theta(., \lambda)$ and $\partial C/\partial y(., .)$ are identified on respectively $\mathcal{Y}_\lambda = \{y/\exists\theta \in \Theta/\theta(y, \lambda) = \theta\}$ and $\{(y, \theta(y, \lambda)), y \in \mathcal{Y}_\lambda\}$, for all $\lambda \in \Lambda$. Now, for all $y \in \mathcal{Y}_\lambda$,

$$\frac{\partial^2 C}{\partial y \partial \lambda}(y, \theta(y, \lambda)) = \frac{\partial^2 C}{\partial y \partial \theta}(y, \theta(y, \lambda))\frac{\partial \theta}{\partial \lambda}(y, \lambda).$$

Moreover, $\partial^2 t/\partial y \partial \lambda(y, \lambda) > 0$ implies that $\partial\theta/\partial\lambda(y, \lambda) > 0$. Thus, for all $y \in \mathcal{Y}_\lambda$, $\partial^2 C/\partial y \partial \theta(y, \theta(y, \lambda))$ is identified by

$$\frac{\partial^2 C}{\partial y \partial \theta}(y, \theta(y, \lambda)) = \frac{\partial^2 C/\partial y \partial \lambda(y, \theta(y, \lambda))}{\partial\theta/\partial\lambda(y, \lambda)}.$$

By equation (3.2.7), $\widetilde{S}(., \lambda)$ is therefore identified on $\mathcal{Y}_\lambda$. ■

Proof of proposition 3.2.8

When $K = 1$ and C', F_θ and $\widetilde{S}$ are unknown, neither model implies any testable restriction. If one of these three functions is known, proposition 3.2.2 shows that the asymmetric model is just identified. Using equations (3.2.1), (3.2.2) and equation (3.2.17) in appendix A, it can be shown that the complete information model is also identified. In general, no contradiction arises in either model, except if assumption 18 fails to hold in one model but not in the other.

When $K \geq 2$, in the complete information model, equations (3.2.17) and (3.2.2) show that

$$t(y, \lambda_1) > t(y, \lambda_2) \Longleftrightarrow F_y(y, \lambda_1) < F_y(y, \lambda_2).$$

[29] The case where $m_1 \geq 0$ and $m_3 \leq 0$ is impossible, since we assumed that $\theta \geq \theta_0$.

This condition is slightly different from (3.2.14) and we are able to distinguish the two models if (3.2.15) holds.

Otherwise, the analysis is more involved. The identification of the complete information model with exogenous variations follows the same lines than previously. Using equations (3.2.1) and (3.2.2), we recover $\widetilde{S}(.,\lambda)$ and $C(.)$ from $\theta(.,\lambda_1)$. By the separability hypothesis, this function is in turn identified through the same horizontal and vertical transforms as in the incomplete model. Hence, using an asymmetric information model when the true model is the complete information one (or the contrary) actually leads to the same $\theta(.,\lambda_1)$, and consequently to the same F_θ.

Differentiating the first order condition in the symmetric case leads to

$$t'(y) = \theta(y)C^{S\prime}(y) + \theta'(y)C^S(y).$$

On the other hand, C^A satisfies

$$t'(y) = \theta(y)C^{A\prime}(y).$$

Because $\theta'(y)C^S(y) < 0$, $C^{A\prime} < C^{S\prime}$ and the marginal cost functions differ. Hence, if C' is known, the models are distinguishable.

Now let us suppose that $\widetilde{S}(.,\lambda_0)$ is known. The functions $\widetilde{S}^S(.,\lambda_0)$ and $\widetilde{S}^A(.,\lambda_0)$ obtained respectively with the symmetric and asymmetric model satisfy

$$\widetilde{S}^S(y,\lambda_0) = \theta(y,\lambda_0)C^{S\prime}(y) = t'(y,\lambda_0) - t(y,\lambda_0)\frac{\theta'(y,\lambda_0)}{\theta(y,\lambda_0)}.$$

$$\widetilde{S}^A(y,\lambda_0) = \left[1 - \frac{1-F_y(y,\lambda_0)}{f_y(y,\lambda_0)}\frac{\theta'(y,\lambda_0)}{\theta(y,\lambda_0)}\right] t'(y,\lambda_0).$$

Thus, $\widetilde{S}^S(.,\lambda_0)$ and $\widetilde{S}^A(.,\lambda_0)$ are identical if and only if, for all y,

$$\frac{t'(y,\lambda_0)}{t(y,\lambda_0)} = \frac{f_y(y,\lambda_0)}{1-F_y(y,\lambda_0)}$$

Integrating this expression leads to

$$t(y,\lambda_0) = \frac{K}{1-F_y(y,\lambda_0)}.$$

for a given $K > 0$. Thus, we can distinguish the models if and only if $y \mapsto t(y,\lambda_0) \times (1 - F_y(y,\lambda_0))$ is not constant. ■

Proof of theorem 3.2.9

In the following, we denote $F_j = F(.,\lambda_j)$ and $\theta_j = \theta(.,\lambda_j)$, $(j \in \{1,2\})$. The result is based on the six following lemmas.

Lemma 3.2.3 *Let $j \in \{1,2\}$ and K denote a compact set in $\overset{\circ}{\mathcal{Y}}_j$. Then*

$$\inf_{y\in K} f_j(y) > 0.$$

Lemma 3.2.4 *Let $j \in \{1,2\}$ and K denote a compact set in $(0,1)$. Then*

$$\sup_{y\in K} |\widehat{F}_j^{-1}(y) - F_j^{-1}(y)| \xrightarrow{\mathbb{P}} 0.$$

Lemma 3.2.5 *For all compact set $K \subset \overset{\circ}{\mathcal{Y}}_1$,*

$$\sup_{x\in K} |\widehat{H}_{12}(x) - H_{12}(x)| \xrightarrow{\mathbb{P}} 0.$$

Lemma 3.2.6 *If $\widehat{y}_{n-1}$ converges in probability to y_{n-1}, then*

$$\widehat{H}_{12}(\widehat{y}_{n-1}) \xrightarrow{\mathbb{P}} H_{12}(y_{n-1}).$$

Lemma 3.2.7 *If $\widehat{y}_{n-1}$ converges in probability to y_{n-1}, then*

$$\begin{aligned} \mathbb{1}_{\widehat{H}_{12}(\widehat{y}_{n-1})\in\widehat{\mathcal{Y}_1}} &\xrightarrow{\mathbb{P}} \mathbb{1}_{H_{12}(y_{n-1})\in\mathcal{Y}_1}. \\ \mathbb{1}_{\widehat{H}_{12}(\widehat{y}_{n-1})\notin\widehat{\mathcal{Y}_1}} &\xrightarrow{\mathbb{P}} \mathbb{1}_{H_{12}(y_{n-1})\notin\mathcal{Y}_1}. \end{aligned}$$

Lemma 3.2.8 *For all $n \in \mathbb{Z}$ and when $N \to +\infty$,*

$$(\widehat{\theta}_n, \widehat{y}_n) \xrightarrow{\mathbb{P}} (\theta_n, y_n).$$

Proof of lemma 3.2.3 : let (for instance) $j = 1$ and let $\theta_1 \equiv \theta(.,\lambda_1)$. By equation (3.2.10), we get for all $y \in \overset{\circ}{\mathcal{Y}}_1$,

$$f_1(y) = -f_\theta(\theta_1(y))\theta_1'(y).$$

Now, deriving the first order condition, we get, because $t'(.,\lambda_1) = \delta_1$,

$$\theta_1'(y) = -\frac{\delta_1}{C'^2(y)} C''(y).$$

Thus, by assumption 18, $\theta_1'(y) < 0$ for all $y \in \overset{\circ}{\mathcal{Y}_1}$. Moreover, because K is a compact subset strictly included in $\mathcal{Y}_1$, $f_\theta(\theta_1(y)) > 0$ for all $y \subset K$. Hence, by continuity of f_θ and θ',

$$\inf_{y \in K} f_1(y) = \min_{y \in K} [-f_\theta(\theta_1(y))\theta_1'(y)] > 0 \ \square$$

Proof of lemma 3.2.4 : let $j = 1$ and $\varepsilon > 0$ be such that $E = \{x \in \mathbb{R} / \exists y \in F_1^{-1}(K) / |x - y| \leq \varepsilon\}$ is a subset of $\overset{\circ}{\mathcal{Y}_1}$. E is compact, so by the previous lemma, $m = \inf_{x \in E} f_1(x) > 0$. Moreover, for all $y \in F_1^{-1}(K)$, $F_1(y - \varepsilon) + m\varepsilon \leq F_1(y) \leq F_1(y + \varepsilon) - m\varepsilon$. Consequently,

$$\begin{aligned}
\sup_{y \in K} |\widehat{F}_1^{-1}(y) - F_1^{-1}(y)| > \varepsilon \iff & \exists y \in K \ / \ |\widehat{F}_1^{-1}(y) - F_1^{-1}(y)| > \varepsilon \\
\iff & \exists y \in K \ / \ F_1(F_1^{-1}(y)) > \widehat{F}_1(F_1^{-1}(y) + \varepsilon) \\
& \text{ or } F_1(F_1^{-1}(y)) < \widehat{F}_1(F_1^{-1}(y) - \varepsilon) \\
\implies & \exists y \in K \ / \ F_1(F_1^{-1}(y) + \varepsilon) - \widehat{F}_1(F_1^{-1}(y) + \varepsilon) > m\varepsilon \\
& \text{ or } \widehat{F}_1(F_1^{-1}(y) - \varepsilon) - F_1(F_1^{-1}(y) - \varepsilon) > m\varepsilon \\
\implies & \sup_{x \in \mathcal{Y}_1} |\widehat{F}_1(x) - F_1(x)| > m\varepsilon
\end{aligned}$$

Hence,

$$\mathbb{P}(\sup_{y \in K} |\widehat{F}_1^{-1}(y) - F_1^{-1}(y)| > \varepsilon) \leq \mathbb{P}(\sup_{x \in \mathcal{Y}_1} |\widehat{F}_1(x) - F_1(x)| > m\varepsilon).$$

The right term tends to zero by the Glivenko-Cantelli theorem. Thus, $\sup_{y \in K} |\widehat{F}_1^{-1}(y) - F_1^{-1}(y)|$ converges to zero in probability. $\square$

Proof of lemma 3.2.5 : by the triangular inequality,

$$\sup_{x \in K} |\widehat{H}_{12}(x) - H_{12}(x)| \leq \sup_{x \in K} |\widehat{F}_2^{-1} \circ \widehat{F}_1(x) - F_2^{-1} \circ \widehat{F}_1(x)| + \sup_{x \in K} |F_2^{-1} \circ \widehat{F}_1(x) - F_2^{-1} \circ F_1(x)| \tag{3.2.22}$$

Let us show that the two terms in the right hand side converge to zero. Because $K = [\underline{K}, \overline{K}] \subset \overset{\circ}{\mathcal{Y}_1}$, there exists a compact set $L \subset (0, 1)$ such that $F_1(K) \subset L$. Thus, for all $\varepsilon > 0$, since $\widehat{F}_1$ is increasing,

$$\begin{aligned}
\mathbb{P}(\sup_{x \in K} |\widehat{F}_2^{-1} \circ \widehat{F}_1(x) - F_2^{-1} \circ \widehat{F}_1(x)| > \varepsilon) \ \leq \ & \mathbb{P}(\widehat{F}_1(\underline{K}) \notin L \cup \widehat{F}_1(\overline{K}) \notin L) \\
& + \ \mathbb{P}(\sup_{y \in L} |\widehat{F}_2^{-1}(y) - F_2^{-1}(y)| > \varepsilon).
\end{aligned}$$

The first term of the right hand side converges to zero since $\widehat{F}_1(\underline{K})$ (resp. $\widehat{F}_1(\overline{K})$) converges in probability to $F_1(\underline{K}) \in L$ (resp. $F_1(\overline{K}) \in L$). The second term tends

to zero by the previous lemma. Thus the first term of (3.2.22) converges to zero in probability.

Let us turn to the second term of (3.2.22). F_2^{-1} is uniformly continuous on $F_1(K)$, as a continuous function on a compact. Thus, there exists η such that $|x-y| \leq \eta$ implies $|F_2^{-1}(x) - F_2^{-1}(y)| \leq \varepsilon$. Hence,

$$\mathbb{P}(\sup_{x\in K} |F_2^{-1} \circ \widehat{F}_1(x) - F_2^{-1} \circ F_1(x)| > \varepsilon) \leq \mathbb{P}(\sup_{x\in K} |\widehat{F}_1(x) - F_1(x)| > \eta).$$

The right term converges to zero by the Glivenko-Cantelli theorem, implying the result. □

Proof of lemma 3.2.6 : For all $A, \varepsilon > 0$,

$$\begin{aligned}\mathbb{P}(|\widehat{H}_{12}(\widehat{y}_{n-1}) - H_{12}(y_{n-1})| > A) &\leq \mathbb{P}(|\widehat{H}_{12}(\widehat{y}_{n-1}) - H_{12}(y_{n-1})| > A, |\widehat{y}_{n-1} - y_{n-1}| \leq \varepsilon) \\ &\quad + \mathbb{P}(|\widehat{y}_{n-1} - y_{n-1}| > \varepsilon).\end{aligned}$$

Let K denote a compact set in $\overset{\circ}{\mathcal{Y}}_1$ such that $y_{n-1} \in \overset{\circ}{K}$. Because H'_{12} is continuous on K, $\max_{x\in K} |H'_{12}(x)| = M$ exists. Let $\varepsilon > 0$ be such that $M\varepsilon < A$ and the closed ball of radius ε centered at y_{n-1} is a subset of K. Then, when $|\widehat{y}_{n-1} - y_{n-1}| \leq \varepsilon$,

$$\begin{aligned}|\widehat{H}_{12}(\widehat{y}_{n-1}) - H_{12}(y_{n-1})| &\leq |\widehat{H}_{12}(\widehat{y}_{n-1}) - H_{12}(\widehat{y}_{n-1})| + |H_{12}(\widehat{y}_{n-1}) - H_{12}(y_{n-1})| \\ &< \sup_{x\in K} |\widehat{H}_{12}(x) - H_{12}(x)| + M\varepsilon.\end{aligned}$$

Thus,

$$\begin{aligned}\mathbb{P}(|\widehat{H}_{12}(\widehat{y}_{n-1}) - H_{12}(y_{n-1})| \geq A) &\leq \mathbb{P}\left(\sup_{x\in K} |\widehat{H}_{12}(x) - H_{12}(x)| > A - M\varepsilon\right) \\ &\quad + \mathbb{P}(|\widehat{y}_{n-1} - y_{n-1}| > \varepsilon)\end{aligned}$$

By assumption, the second term tends to zero in probability. $\sup_{x\in K} |\widehat{H}_{12}(x) - H_{12}(x)|$ also converges to zero by lemma 3.2.5. The result follows. □

Proof of lemma 3.2.7 : We consider two cases. First, if $H_{12}(y_{n-1}) \notin \mathcal{Y}_1$, then, because $\mathcal{Y}_1$ is a closed set, $\min_{x\in\mathcal{Y}_1} |H_{12}(y_{n-1}) - x| = a > 0$. Now, because $\widehat{\mathcal{Y}_1} \subset \mathcal{Y}_1$,

$$\begin{aligned}\mathbb{P}(\widehat{H}_{12}(\widehat{y}_{n-1}) \in \widehat{\mathcal{Y}_1}) &\leq \mathbb{P}(\widehat{H}_{12}(\widehat{y}_{n-1}) \in \mathcal{Y}_1) \\ &\leq \mathbb{P}(|\widehat{H}_{12}(\widehat{y}_{n-1}) - H_{12}(y_{n-1})| \geq a)\end{aligned}$$

By lemma 3.2.6, $\mathbb{P}(\widehat{H}_{12}(\widehat{y}_{n-1}) \in \widehat{\mathcal{Y}_1})$ tends to zero.

Secondly, suppose that $H_{12}(y_{n-1}) \in \mathcal{Y}_1$. Then, by assumption 25 and because $H_{12}(y_{n-1}) = y_n$, $H_{12}(y_{n-1}) \in \overset{\circ}{\mathcal{Y}}_1$ and $0 < F_1(H_{12}(y_{n-1})) < 1$. Hence, there exists $a > 0$ such that $F_1(H_{12}(y_{n-1}) + a) < 1$ and $F_1(H_{12}(y_{n-1}) - a) > 0$. Now,

$$\begin{aligned}
\mathbb{P}(\widehat{H}_{12}(\widehat{y}_{n-1}) \notin \widehat{\mathcal{Y}_1}) &\leq \mathbb{P}(\widehat{H}_{12}(\widehat{y}_{n-1}) \notin \widehat{\mathcal{Y}_1}, |\widehat{H}_{12}(\widehat{y}_{n-1}) - H_{12}(y_{n-1})| \leq a) \\
&\quad + \mathbb{P}(|\widehat{H}_{12}(\widehat{y}_{n-1}) - H_{12}(y_{n-1})| > a) \\
&\leq \mathbb{P}\Big[\{\widehat{H}_{12}(\widehat{y}_{n-1}) > \max_k y_{1k}\} \cup \{\widehat{H}_{12}(\widehat{y}_{n-1}) < \min_k y_{1k}\}, \\
&\quad |\widehat{H}_{12}(\widehat{y}_{n-1}) - H_{12}(y_{n-1})| \leq a\Big] + \mathbb{P}(|\widehat{H}_{12}(\widehat{y}_{n-1}) - H_{12}(y_{n-1})| > a) \\
&\leq \mathbb{P}\left[\{H_{12}(y_{n-1}) + a > \max_k y_{1k}\} \cup \{H_{12}(y_{n-1}) - a < \min_k y_{1k}\}\right] \\
&\quad + \mathbb{P}(|\widehat{H}_{12}(\widehat{y}_{n-1}) - H_{12}(y_{n-1})| > a) \\
&\leq \mathbb{P}(H_{12}(y_{n-1}) + a > \max_k y_{1k}) + \mathbb{P}(H_{12}(y_{n-1}) - a < \min_k y_{1k}) \\
&\quad + \mathbb{P}(|\widehat{H}_{12}(\widehat{y}_{n-1}) - H_{12}(y_{n-1})| > a) \\
&\leq [F_1(H_{12}(y_{n-1}) + a)]^N + [1 - F_1(H_{12}(y_{n-1}) - a)]^N \\
&\quad + \mathbb{P}(|\widehat{H}_{12}(\widehat{y}_{n-1}) - H_{12}(y_{n-1})| > a).
\end{aligned}$$

The first two terms converge to zero. The third term tends to zero in probability by lemma 3.2.6. Thus $\mathbb{P}(\widehat{H}_{12}(\widehat{y}_{n-1}) \notin \widehat{\mathcal{Y}_1})$ tends to zero. This proves the lemma. □

Proof of lemma 3.2.8 : We proceed by induction. The result holds for $n = 0$ since $\widehat{y_0} = y_0$ and $\widehat{\theta_0} = \theta_0$. Suppose that it is true for $n - 1 \geq 0$. Using the induction hypothesis, lemma 3.2.6 and 3.2.7, $\widehat{y_n}$ converges to y_n in probability. Similarly, using lemma 3.2.7, $\widehat{\theta_n}$ converges to θ_n. Because convergence in probability of each term implies joint convergence (see e.g. van der Vaart, 1998, theorem 2.7), the result holds for n. Hence, lemma 3.2.8 is true for all $n \geq 0$. The proof is similar for negative values. □

Now, let us come back to the proof of theorem 3.2.9. By the previous lemma, it suffices to prove that $\widehat{F_\theta(\theta_n)}$ (resp. $\widehat{C'(y_n)}$) converges in probability to $F_\theta(\theta_n)$ (resp. $C'(y_n)$). By the triangular inequality,

$$\begin{aligned}
|\widehat{F_\theta(\theta_n)} - F_\theta(\theta_n)| &= |\widehat{F}_1(\widehat{y}_n) - F_1(y_n)| \\
&\leq |\widehat{F}_1(\widehat{y}_n) - F_1(\widehat{y}_n)| + |F_1(\widehat{y}_n) - F_1(y_n)| \\
&\leq \sup_{y \in \mathcal{Y}_1} |\widehat{F}_1(y) - F_1(y)| + |F_1(\widehat{y}_n) - F_1(y_n)|.
\end{aligned}$$

The first term converges to zero by the Glivenko-Cantelli theorem. The second term also converges to zero by the previous lemma and the continuity of F_1. The first convergence follows. Because $\widehat{C'(y_n)} = \delta_1/\widehat{\theta}_n$ is a continuous function of $\widehat{\theta}_n$, the second result stems from lemma 3.2.8 once more. ■

3.2.7 Appendix B : surplus

Estimation of the parameters a, b, α, β

The first order condition of the agent writes as $\theta(y, \lambda_i) = \delta_i / C'(y)$. Using the parametric form of the cost $C'(y) = \alpha \left(\frac{y}{1-y}\right)^{\beta}$, we obtain

$$\theta(y, \lambda_i) = \frac{\delta_i}{\alpha} \left(\frac{1-y}{y}\right)^{\beta}$$

Then, using the Weibull specification, we obtain
$F_y(y, \lambda_i) = 1 - F_\theta(\theta(y, \lambda_i)) = \exp\left(-a \left(\frac{\delta_i}{\alpha}\right)^b \left(\frac{1-y}{y}\right)^{b\beta}\right)$
which rewrites as

$$\ln(-\ln(F_y(y, \lambda_i))) = \ln(a) + b\left(\ln(\delta_i) - \ln(\alpha)\right) + b\beta \ln\left(\frac{1-y}{y}\right).$$

The parameters are estimated by regressing $\ln(-\ln(\widehat{F}_j(y)))$ on $\ln(\frac{1-y}{y})$. As explained in the main text, a normalization is necessary and we impose $C'(y_0) = \delta_1/\theta_0$.
Estimation of λ_3

We suppose here that Insee can only implement linear contracts $t(y, \delta) = \delta y$ and denote by $y(\theta, \delta)$ the response rate chosen by a θ type agent. Insee solves, for the 2003 survey,

$$\max_{\delta} \int (\lambda_3 - \delta) y(\theta, \delta) f_\theta(\theta) d\theta.$$

and the first order condition writes as

$$-\int y(\theta, \delta_3) f_\theta(\theta) d\theta + (\lambda_3 - \delta_3) \int \frac{\partial y}{\partial \delta}(\theta, \delta_3) f_\theta(\theta) d\theta = 0.$$

Using again the first order condition of the agent, we have $y(\theta, \delta) = \frac{1}{1+\left(\frac{\alpha\theta}{\delta}\right)^{1/\beta}}$. Thus,

$$\frac{\partial y}{\partial \delta}(\theta, \delta) = \frac{1}{\beta\delta} y(1-y).$$

and the condition takes the form

$$-E(y(\theta,\delta_3)) + \frac{\lambda_3-\delta_3}{\beta\delta_3}E(y(\theta,\delta_3)(1-y(\theta,\delta_3))) = 0$$

Hence,

$$\lambda_3 = \delta_3\left(1+\beta\frac{E(y(\theta,\delta_3))}{E(y(\theta,\delta_3)(1-y(\theta,\delta_3)))}\right)$$

which is estimated, noting $\overline{y}_3$ (resp. $\overline{y^2}_3$) the empirical mean of the response rate (resp. the empirical mean of the response rate's square) in 2003, by

$$\widehat{\lambda}_3 = \delta_3\left(1+\widehat{\beta}\frac{\overline{y}_3}{\overline{y}_3-\overline{y^2}_3}\right)$$

Linear contracts under incomplete information

The expected surplus of Insee, when linear contracts are used, is $\Pi = (\lambda_3 - \delta_3)E(y(\theta,\delta_3))$ which can be estimated by

$$\widehat{\Pi} = (\widehat{\lambda}_3 - \widehat{\delta}_3)\overline{y}_3$$

Optimal contracts under incomplete information

On one hand, under incomplete information, the optimal contract is defined by

$$\lambda_3 = \left[\theta + \frac{F_\theta(\theta)}{f_\theta(\theta)}\right]C'(y^I(\theta,\lambda_3)).$$

Hence, using the form of the cost function, we obtain

$$y^I(\theta,\lambda_3) = \frac{1}{1+\left(\frac{\alpha}{\lambda_3}\right)^{1/\beta}\left[\theta+\frac{F_\theta(\theta)}{f_\theta(\theta)}\right]^{1/\beta}}.$$

On the other hand, using the first order condition of the agent, we have

$$t'(y^I(\theta,\lambda_3)) = \theta C'(y^I(\theta,\lambda_3)) = \lambda_3\frac{\theta}{\theta+F_\theta(\theta)/f_\theta(\theta)}.$$

Hence, the surplus under incomplete information, which writes as

$$\Pi^I = \lambda_3 E(y^I(\theta,\lambda_3)) - E(t(y^I(\theta,\lambda_3)))$$

can be estimated by Monte-Carlo simulations.

More precisely, to estimate this surplus, we draw 100,000 values of θ in our estimated Weibull distribution $F_\theta(.|\widehat{a},\widehat{b})$ and compute the previous functions using the parameters $\widehat{\lambda}_3$, $\widehat{\alpha}$ and $\widehat{\beta}$.

Optimal contracts under complete information

Under complete information, the first order condition for the agent writes $\lambda_3 = \theta C'(y^C(\theta, \lambda_3))$. Hence, using the parametric form of the cost function, we obtain

$$y^C(\theta, \lambda_3) = \frac{1}{1 + \left(\frac{\alpha\theta}{\lambda_3}\right)^{1/\beta}}.$$

Furthermore, the transfer function is given by

$$t(y^C(\theta, \lambda_3), \lambda_3) = \theta C(y^C(\theta, \lambda_3)) = \alpha\theta \int_0^{y^C(\theta,\lambda_3)} \left(\frac{u}{1-u}\right)^{\beta} du$$

Finally, the expected surplus under complete information is given by

$$\Pi^C = \lambda_3 E(y^C(\theta, \lambda_3)) - E(t(y^C(\theta, \lambda_3))).$$

As previously, to estimate this surplus, we draw 100,000 values of θ in our estimated Weibull distribution $F_\theta(.|\widehat{a}, \widehat{b})$ and compute the previous functions using the parameters $\widehat{\lambda}_3$, $\widehat{\alpha}$ and $\widehat{\beta}$.

3.2.8 Appendix C : discussion on assumption 22

In this appendix, we provide a characterization of assumption 22 under a rational condition. Let $E_1 = l(\lambda_1)/l(\lambda_2)$ and $E_3 = l(\lambda_1)/l(\lambda_3)$. In a linear transfer framework, E_1 and E_3 represent ratios of piece rates. Thus in practice, these two numbers should be rational. The proposition below shows that under this condition, assumption 22 is almost always true.

Proposition 3.2.10 *Suppose that $(E_1, E_3) \in \mathbb{Q}^2$, $E_1 = p_1/q_1$, $E_3 = p_3/q_3$, p_1 and q_1 (resp. p_3 and q_3) being coprime. Let $p_1 = \prod_{i=1}^{n_1} p_{1i}^{\alpha_{1i}}$, $p_3 = \prod_{i=1}^{n_3} p_{3i}^{\alpha_{3i}}$, $q_1 = \prod_{i=1}^{m_1} q_{1i}^{\beta_{1i}}$ and $q_3 = \prod_{i=1}^{m_3} q_{3i}^{\beta_{3i}}$ be the decomposition into prime factors of p_1, p_3, q_1 and q_3. Then assumption 22 fails to hold if and only if $p_{1i} = p_{3i}$, $q_{1i} = q_{3i}$ for all i (with $n_1 = n_3$ and $m_1 = m_3$),*

$$\frac{\alpha_{1i}}{\gcd(\alpha_{11}, ..., \alpha_{1n_1})} = \frac{\alpha_{3i}}{\gcd(\alpha_{31}, ..., \alpha_{3n_1})} \tag{3.2.23}$$

$$\frac{\beta_{1i}}{\gcd(\beta_{11}, ..., \beta_{1n_1})} = \frac{\beta_{3i}}{\gcd(\beta_{31}, ..., \beta_{3n_1})} \tag{3.2.24}$$

and

$$\frac{\gcd(\alpha_{11},...,\alpha_{1n_1})}{\gcd(\alpha_{31},...,\alpha_{3n_1})} = \frac{\gcd(\beta_{11},...,\beta_{1n_1})}{\gcd(\beta_{31},...,\beta_{3n_1})} \tag{3.2.25}$$

Proof : $\ln(E_3)/\ln(E_1)$ is rational if and only if there exists $g \in \mathbb{Q}$ such that

$$\left(\frac{p_1}{q_1}\right) = \left(\frac{p_3}{q_3}\right)^g.$$

This in turn is equivalent to the existence of (k_1, k_3) coprimes satisfying

$$p_1^{k_1} q_3^{k_3} = p_3^{k_3} q_1^{k_1}.$$

By the Gauss theorem, $p_1^{k_1} | p_3^{k_3}$ and conversely, $p_3^{k_3} | p_1^{k_1}$. Thus, $p_1^{k_1} = p_3^{k_3}$. Similarly, $q_1^{k_1} = q_3^{k_3}$. This implies that the primal factors of p_1 and p_3 (and q_1 and q_3) are identical and for all $1 \leq i \leq n_1$,

$$\alpha_{1i} k_1 = \alpha_{3i} k_3 \tag{3.2.26}$$

Once more by the Gauss theorem, there exists u such that $\gcd(\alpha_{31},...,\alpha_{3n_1}) = k_1 u$, and similarly v such that $\gcd(\alpha_{11},...,\alpha_{1n_1}) = k_3 v$. Hence, letting $\alpha_{1i'} = \alpha_{1i}/\gcd(\alpha_{11},...,\alpha_{1n_1})$ (and similarly for α_{3i}), we get for all i,

$$\alpha_{1i'} v = \alpha_{3i'} u$$

Taking the greatest common divisor in both sides, we get $u = v$ and thus (3.2.23) is proved. (3.2.24) can be obtained similarly. Lastly, equation (3.2.26) shows that

$$\frac{k_1}{k_3} = \frac{\gcd(\alpha_{31},...,\alpha_{3n_1})}{\gcd(\alpha_{11},...,\alpha_{1n_1})}.$$

The same holds with the (β_{1i}) and (β_{3i}), implying (3.2.25). One can easily shows that these conditions are sufficient ■

Chapitre 4

Identification of peer effects using group size variation

4.1 Introduction

In a seminal paper, Manski (1993) showed that in a linear-in-expectations model with social interactions, endogenous and exogenous peer effects cannot be separately identified. Only a function of these two types of effects can be identified under some strong exogeneity conditions. In the context of pupils achievement for instance, Hoxby (2000) and Ammermueller & Pischke (2006) reach identification by assuming that variations in time or between classrooms within the same school are random.[1] However, Lee (2007) has recently proposed a modified version of the social interaction model, which corresponds to a linear-in-means model, and which is shown to be identifiable without any of the previous restrictive assumptions, thanks to the group size variation.

The aim of our paper is threefold. Firstly, we reexamine the identification of this linear-in-means model when group sizes do not depend on the sample size.[2] We believe that, in practice, such an assumption is virtually always satisfied. For instance, there is no reason why the mean classroom size should depend on the size of the sample. Moreover, this extra assumption enables to clarify the sources of

[1] In the following, we will often consider the example of peer effects at school, although the model could also be applied to other topics like smoking (see e.g. Krauth, 2006, productivity in teams (see Rees et al., 2008) or retirement (Duflo0 & Saez, 2003).

[2] This is approximately the scenario with small group interactions of Lee (2007).

identification in this model.[3] More precisely, we show that in his linear-in-means model, the crucial assumptions for identification are 1) the knowledge of the group sizes, and 2) the fact that group sizes take at least three different values. Parametric assumptions on the error term are not needed. In general, homoskedasticity is not required either. It is useful however when peer effects cancel each others, since in this case identification is lost without such a restriction.

Secondly, we extend these results to a model where only binary outcomes are observed. Identification of discrete models with social interactions has already been studied by, e.g., Brock & Durlauf (2001, 2007) and Krauth (2006). Our model is slightly different, though, as we assume that social interactions may affect individuals through peers' latent variables rather than through their observable outcomes. This is convenient when only binary outcomes are observable, because of data limitation. This model is close to spatial discrete choice models (see e.g. Case, 1992, McMillen, 1992, Pinkse & Slade, Beron & Vijverberg or Klier & McMillen). The difference is that we allow for both exogenous peer effects and fixed group effects. The attractive feature of our result is that it does not rely on any functional assumption concerning the errors. Once more, the exogenous peer effects can be identified through group size variation. On the other hand, due to the loss of information, endogenous peer effects cannot be identified without further restrictions. We show that an homoskedasticity condition is sufficient for this purpose.

Thirdly, we develop a parametric estimation of the binary model, complementing the methods proposed by Lee (2007) for the model with continuous outcome. We show that under a normality assumption on the residuals and a linear specification à la Mundlak (1961) on the fixed effect, a simulated maximum likelihood estimator can be implemented using the GHK algorithm (Geweke, 1989, Keane, 1994, and Hajivassiliou et al., 1996). Thus, this estimator is close to Beron & Vijverberg's one on spatial probit models. We investigate its finite sample properties by Monte Carlo simulations. The results stress the determining effect of average group size for the accuracy of the inference, in line with Lee (2007)'s result on the linear model.

The paper is organized as follows. In the first section, we present the theoretical model of social interactions. Section two considers the identification of the model,

[3]Under his more general setting, Lee (2007) provides sufficient conditions for identification, but they are rather difficult to interpret (see his assumption 6.1 and 6.2).

in the continuous and in the discrete case. The fourth section discusses the parametric estimation method of the discrete model. Section six displays Monte Carlo simulations. Section seven concludes. Proofs are given in the appendix.

4.2 A theoretical model of social interactions

We consider the issue of individual choices with social interactions in groups. Let e_i denote the continuous choice variable of an individual i who belongs to the group $(1, ..., m)$, x_i be his exogenous characteristics and ε_i correspond to an idiosyncratic individual term. We suppose that the utility of i when choosing e_i, while other choose $(e_j)_{j\neq i}$, takes the following form :

$$\begin{aligned}\mathcal{U}_i(e_i,(e_j)_{j\neq i}) = \; & e_i\Bigg[x_i\beta_{10} + \left(\frac{1}{m-1}\textstyle\sum_{j=1,\,j\neq i}^{m} e_j\right)\lambda_0 \\ & + \left(\frac{1}{m-1}\textstyle\sum_{j=1,\,j\neq i}^{m} x_j\right)\beta_{20} + \alpha + \varepsilon_i\Bigg] - \frac{1}{2}e_i^2\end{aligned} \tag{4.2.1}$$

In this framework, the marginal returns of individual i depends on his peers' choices, on their exogenous characteristics and on a group fixed effect α. In a classroom for instance, the utility of a student will depend on his effort e_i and the effort of others, because of spillovers in the learning process. It will also possibly depend on the characteristics of others. There has been indeed empirical evidence of spillovers to peers' race, sex or parental education (see e. g. Hoxby, 2001 or Cooley, 2007). Lastly, the outcome can depend on a fixed classroom effect, due for instance to teachers' quality. This model is close to the one considered by Calvó-Armengol et al. (2008) to study the effect of peers on education. An important difference is that they consider network of friends, whereas our model is better suited when all classmates potentially affect the result of a student.
Assuming that α and the $(x_i, \varepsilon_i)_{1\leq i\leq m}$ are observed by all the individuals of the group, the Nash equilibrium of the game $(y_1^*, ..., y_m^*)$ writes as

$$y_i^* = x_i\beta_{10} + \left(\frac{1}{m-1}\sum_{j=1,j\neq i}^{m} y_j^*\right)\lambda_0 + \left(\frac{1}{m-1}\sum_{j=1,j\neq i}^{m} x_j\right)\beta_{20} + \alpha + \varepsilon_i. \tag{4.2.2}$$

This model is identical to Lee's model (2007) of social interactions. Following the terminology introduced by Manski (1993), the second term in the right hand side

corresponds to the endogenous peer effect, the third refers to the exogenous peer effects and α is a contextual (group-specific) effect. This model departs from the one considered by Manski (1993) or by Graham and Hahn (2005) by replacing, on the right hand side, the expectations relative to the whole group by the means of outcomes and covariates in the group of peers.[4] Interestingly, one can show that Manski's model is actually the Bayesian Nash equilibrium of the game when player i does not observe the characteristics $(x_j, \varepsilon_j)_{j \neq i}$ of his peers, if the $(\varepsilon_i)_i$ are mutually independent and independent of (x_i, α, m). This framework seems more realistic in large groups, whereas the hypothesis that the characteristics of others are observed is likely to hold in small ones.

4.3 Identification

We now turn to the identification of model (4.2.2). First, as a benchmark, we suppose that the y_i^* are observed directly. This case corresponds to Lee (2007)'s results, but we will investigate it in a slightly different framework. Then, we study the situation where only rough measures of them, namely $y_i = \mathbb{1}\{y_i^* \geq 0\}$, are available. In both cases, we implicitly assume that the econometrician knows the groups of interaction of each individual. In the previous example, this assumption is mild if students really interact within the classroom, since the classroom identifier is usually known. It can be restrictive otherwise, but to our best knowledge, this assumption is also maintained by all papers studying identification of peer effects, including Manski (1993), Brock & Durlauf (2001), Lee (2007), Graham (2008) and Bramoullé et al. (2009). This stems from the fact that, in Manski's model at least, very little can be inferred from the data and the model if the peer group is not known (see Manski, 1993, subsection 2.5).

4.3.1 The benchmark : the linear model

In this section, we clarify and extend the results of Lee (2007), in the case where m does not depend on the size of the sample.[5] We believe that, in practice, such

[4] Graham & Hahn (2005) makes the further restriction that $\beta_{20} = 0$, i.e. that there are no exogenous peer effects.

[5] This is approximately the scenario with small group interactions of Lee (2007).

an assumption is virtually always satisfied. For instance, there is no reason why the mean classroom size should depend on the size of the sample. Moreover, this restriction enables us to show what is identified from the usual exogeneity condition (see assumption A4 below) and when homoscedasticity is necessary (see theorem 2 below).

It is quite common to observe some but not all members of groups in samples, and we take this into account for identification. On the other hand, we maintain the assumption that the size of the group is observed.[6] Let n denote the number of sampled individuals in the group. We denote by $\widetilde{Y}^*$ (respectively, $\widetilde{X}$) the vector of outcomes y_i^* (respectively, of covariates) of the individuals sampled in the group. Let $F_{m,n}$ denote the distribution function of (m,n) and $F_{\widetilde{Y}^*,\widetilde{X}|m,n}$ denote the distribution of $(\widetilde{Y}^*,\widetilde{X})$ conditional on (m,n). Lastly, for any random variable T, we let $\text{Supp}(T)$ denote its support. We rely on the following definition of identification.

Definition 4.3.1 *$(\beta_{10},\beta_{20},\lambda_0)$ is identified if there exists a function φ such that*

$$(\beta_{10},\beta_{20},\lambda_0)=\varphi((F_{\widetilde{Y}^*,\widetilde{X}|m=u,n=v})_{u,v\in Supp(m,n)},F_{m,n}).$$

This definition states that the structural parameters are identified if they can be obtained through the distribution of the data. Implicit in the definition is the fact that our asymptotic is in the number of groups, as in standard panel data models.[7] Now, the key point for identification of the parameters when the y_i^* are observed is to focus on the within-group equation, which writes as

$$W_n\widetilde{Y}^*=W_n\widetilde{X}\left(\frac{(m-1)\beta_{10}-\beta_{20}}{m-1+\lambda_0}\right)+W_n\frac{\widetilde{\varepsilon}}{1+\lambda_0/(m-1)}, \tag{4.3.1}$$

where $\widetilde{\varepsilon}$ is the vector of unobserved residuals for individuals sampled in the group and W_n denotes the within-group matrix of size n, that is to say the matrix with $1-1/n$ on the diagonal and $-1/n$ elsewhere. To identify the structural parameters, we use the variation in the slope coefficient $\beta(m)=((m-1)\beta_{10}-\beta_{20})/(m-1+\lambda_0)$. For this purpose, we make the following assumptions.

A1. $\Pr(n\geq 2)>0$.

[6] This assumption is realistic in our leading example. In French panels of students, for instance, classroom sizes are observed while only a fraction of pupils within classrooms is sampled.

[7] Indeed, when the number of groups tend to infinity, we are able to estimate consistently $(F_{\widetilde{Y}^*,\widetilde{X}|m=u,n=v})_{u,v\in\text{Supp}(m,n)}$ as well as $F_{m,n}$.

A2. Supp(m) contains at least three values.

A3. For all $1 \leq i, j \leq m$, $E\left[x_i' \varepsilon_j \mid m, n\right] = 0$.

A4. $E\left[\widetilde{X}' \, W_n \, \widetilde{X} \mid m, n\right]$ is almost surely nonsingular.

A5. $1 > \lambda_0 > 1 - \min\left(\text{Supp}(m)\right)$.

Assumption A1 simply states that the within-group approach is possible. Assumption A2, which is the cornerstone of our approach, ensures that there is sufficient variation in group sizes. Assumption A1 and A3 and A4 are standard in linear panel data models, except that conditional expectations depend here both on the number of observed individuals in each group and on the group size. Conditioning by n does not cause any trouble if, for instance, the observed individuals are drawn at random from the group. Finally, assumption A5 ensures that $\beta(u)$ exists for all $u \in \text{Supp}(m)$.[8]

Theorem 1. *Under assumptions A1-A5, β_{10} is identified. Moreover,*
- if $\beta_{20} \neq -\lambda_0 \beta_{10}$, then λ_0 and β_{20} are identified;
- if $\beta_{20} = -\lambda_0 \beta_{10}$, then λ_0 is not identified and β_{20} is identified only up to a constant.

Theorem 1 states that all parameters are generally identified provided that there is sufficient variation in the group sizes. As a notable exception, identification is lost in the absence of endogenous and exogenous peer effects, since then $\beta_{20} = -\lambda_0 \beta_{10} = 0$. One can always rationalize such a model with any $\lambda_0' \neq 0$ and $\beta_{20}' = -\lambda_0' \beta_{10}$. Intuitively, one cannot distinguish in the data, using the first order conditions alone, peer effects which cancel each other from no peer effects at all. Below, we provide a method which yields identification in this case, but it relies on a stronger assumption of homoskedasticity. In any case, one can check whether identification is lost or not, since this amounts to test whether $\beta(.)$ is constant or not.
Contrary to the reduced form approach, we do not need to know the means $(\overline{x}_r)_{1 \leq r \leq R}$ on the whole groups to identify the parameters. Thus the problem of measurement error of $\overline{x}_r$, which appears when some individuals in the group are unobserved, does not arise in our framework. Here the crucial assumption is the

[8] This assumption could be weakened to $\lambda_0 \notin -\text{Supp}(m-1)$ without affecting theorem 1. However, it will be necessary in this form in theorems 2, 3, 4 and lemma 1.

knowledge of the group size. If it is unknown but can be estimated, the measurement error problem comes back in a nonlinear way. The issue of identification in this case is left for future research.[9]

Another identifying assumption lies in the nature of the group size effect. Indeed, m may be correlated with α in a general way, but we cannot add interaction terms between the indicators $\mathbb{1}\{m = u\}$ (with $u \in \text{Supp}(m)$) and the covariates to the list of regressors, since then assumption A4 would fail. Another way to see this is that if β_{10} and β_{20} depend on m in an unspecified way, we can still identify $\beta(m)$ but not go back to the structural parameters. On the other hand, identification of these structural parameters can still be achieved if the dependency of β_{10} and β_{20} in m is parametric.[10] Of course, identification requires more than three different values of m in these cases. This also implies that the basic model where β_{10}, β_{20} and λ_0 are constant over group sizes is overidentified as soon as there is at least four different group sizes. A simple way to test this restriction is to estimate $\beta(.)$ by within estimators for each group size, and then use the overidentification test of classical minimum distance estimators (see e. g. Wooldridge, 2002, p. 444).

If $\beta_{20} = -\lambda_0\beta_{10}$, then λ_0 and β_{20} cannot be identified. However they can be identified by studying variance variation under an homoskedasticity condition (assumption A6 below). More precisely, the conditional variance of the residuals should not depend on the group size. This hypothesis is quite weak since it does not restrict the relationship between the residuals ε_{ri} and the covariates x_{ri}. Moreover, under A6, one needs less variation on the group sizes than previously and we can replace assumption A2 by A2'.

A2'. Supp(m) contains at least two values.

A6. $\text{Var}\left(\widetilde{\varepsilon} \mid n, m\right) = \sigma^2 I_n$ where I_n is the identity matrix of size n.

Theorem 2. *Under assumptions A1, A2' and A3-A6,* $(\beta_{10}, \lambda_0, \beta_{20})$ *are identified.*

The idea of using second order moments to identify peer effects has already been used by Glaeser et al. (1996) and Graham (2008). Graham, in particular, deve-

[9] Following Schennach (2004), the model would still be identified if two independent measures of m were available. The remaining issue is whether the model is identified with only one measure, as it is (under weak conditions) in a linear model (see, e.g., Lewbel (1997)).

[10] For instance we can let these parameters write as affine functions of m. This is equivalent to adding interaction terms between $\widetilde{X}$ and m.

lops a framework where composite peer effects can be identified through such a restriction. In his model however, endogenous peer effects cannot be identified.

4.3.2 The binary model

We now investigate whether the parameters are still identified when one cannot observe directly the outcome variable y_i^* but only a rough binary measure of it, namely $y_i = \mathbb{1}\{y_i^* \geq 0\}$.[11] For instance, to study pupil achievement, only grade retention rather than test scores may be available. Similarly, in violence studies, only criminal (that is, sufficiently violent) acts can be observed by the econometrician. Note that these models remain essentially linear because the underlying model is linear. One could also study the case where y_i^* depends on y_j rather than y_j^*. Such models, which have been studied by Brock & Durlauf (2001, 2007) , Tamer (2003), Krauth (2006) and Bayer & Timmins (2007), are more complex because in general multiple equilibria arise.

When the outcome is a binary variable, the reduced-form equation (3) is useless for identification because $W_n\widetilde{Y}^*$ has no observational counterpart. Instead, we rely on equation (4.3.2) below.

Lemma 1. *Suppose that $y_i = \mathbb{1}\{y_i^* \geq 0\}$ with y_i^* satisfying equation (1), and that assumption A5 holds. Then the model is observationally equivalent to*

$$\begin{aligned} y_i \;=\; \mathbb{1}\Bigg\{ & x_i\left(\beta_{10} - \frac{\beta_{20}}{m-1}\right) + \left[\overline{x}\frac{m}{m-1}\left(\beta_{20} + \frac{\beta_{10}+\beta_{20}}{1-\lambda_0}\lambda_0\right) + \alpha\left(1+\lambda_0(m)\right)\right] \\ & + \overline{\varepsilon}\,\lambda_0(m) + \varepsilon_i \geq 0\Bigg\} \end{aligned} \tag{4.3.2}$$

where $\lambda_0(m) = \frac{m}{m-1}\frac{\lambda_0}{1-\lambda_0}$.

The term into brackets corresponds to a group-specific effect. Thus we are led back to a binary model for panel data. Identification of such a model has been considered, among others, by Manski (1987), and our analysis relies on his paper. In the sequel, we let x_j^k denote the k-th covariate of individual j. The following assumptions are needed for identification :

[11] The definition of identification here is similar to previously, except that $\widetilde{Y}^*$ in definition 1 has to be replaced by $\widetilde{Y}$, the vector of y_i of the individuals sampled in the group.

A7. $(\varepsilon_1, ..., \varepsilon_m)$ are exchangeable conditional on $(m, x_1, ..., x_m, \alpha)$. The support of $\varepsilon_1 + \overline{\varepsilon}\frac{m}{m-1}\frac{\lambda_0}{1-\lambda_0}$ conditional on $(m, x_1, ..., x_m, \alpha)$ is $\mathbb{R}$ almost surely.

A8. Let $z = x_2 - x_1$.[12] The support of z is not contained in any proper linear subspace of $\mathbb{R}^K$ (where K is the dimension of x_{ri}).

A9. There exists k_0 such that z^{k_0} has everywhere a positive Lebesgue density conditional on $(m, z^1, ..., z^{k_0-1}, z^{k_0+1}, ..., z^K)$ and $\beta_{10}^{k_0} = 1$. Without loss of generality we set $k_0 = 1$.

The first part of assumption A7 holds for instance if, conditional on m and α, the $(\varepsilon_i)_{i=1,...,m}$ are exchangeable and independent of the $(x_i)_{i=1,...,m}$. In particular, A7 is satisfied if the $(\varepsilon_i)_{i=1,...,m}$ are i.i.d. and independent of $(x_1, ..., x_m, m, \alpha)$. The second part of assumption A7 is a technical condition, which is identical to the second part of assumption 1 set forth by Manski (1987). Assumption A8 ensures that z varies enough within a group. As usually in binary models, one parameter must be normalized and this is the purpose of A9. However, a small difficulty arises here, because the reduced form does not allow us to identify the sign of the structural parameters. A sufficient condition is to fix one parameter (and not only its absolute value) : thus we set $\beta_{10}^1 = 1$.[13]

Theorem 3. *Suppose that assumptions A1-A2, A5 and A7-A9 hold. Then β_{10} is identified. Moreover,*
- if $\beta_{20} \neq \beta_{20}^1\beta_{10}$, then β_{20} is identified,
- if $\beta_{20} = \beta_{20}^1\beta_{10}$, β_{20}^1 is not identified and the other β_{20}^k are identified up to β_{20}^1.
On the other hand, λ_0 is not identified.

If fewer parameters than in model (1) are identified, theorem 4 shows that the main attractive features of the method remain. Without any exclusion restriction and even if only two members of the groups are observed, β_{10} and β_{20} are generally identified. Similarly to the result of theorem 1, identification of β_{20} is lost when there is no exogenous effect, because in this case $\beta_{20} = \beta_{20}^1\beta_{10} = 0$. That λ_0 cannot be identified is not surprising as this parameter only appears in the fixed effect and the residuals (see equation (4.3.2)). Heuristically, without any hypothesis imposed on these terms, any λ_0 can be rationalized by changing accordingly α and the $(\varepsilon_i)_{1\leq i\leq m}$.

[12] Without loss of generality, we assume here that individuals 1 and 2 are observed.
[13] Obviously, theorem 3 also holds with $\beta_{10}^1 = -1$.

Thus, stronger assumptions are needed for identifying λ_0. One possibility is to observe $\overline{x}$ and to restrict the dependence between the residuals and the covariates.

A2''. The support of m conditional on $\overline{x}$ has at least three elements with a positive probability.

A9'. $\overline{x}$ is observed.

A10. $(\varepsilon_1, ..., \varepsilon_m, \alpha) \perp\!\!\!\perp (x_1, ..., x_m) \mid m, \overline{x}$.

A11. $\text{Var}(\varepsilon_1, ..., \varepsilon_m, \alpha \mid \overline{x}, m) = \begin{pmatrix} \text{Var}(\varepsilon_1 \mid \overline{x}) I_m & 0 \\ 0 & \text{Var}(\alpha \mid \overline{x}) \end{pmatrix}$.

A12. Given $(\overline{x}, m)$, the support of $\left\{x_1(\beta_{10} - \frac{\beta_{20}}{m-1}), x_2(\beta_{10} - \frac{\beta_{20}}{m-1})\right\}$ is $\mathbb{R}^2$.

A2'' is slightly more restrictive than A2 but should hold most of the time. It is satisfied for instance under a multinomial logit (or probit) model on m conditional on $\overline{x}$. As mentioned above, assumption A9' is a restrictive condition as it imposes either to observe all individuals in the group or to consider only the covariates for which the means are known. Assumption A10 is in the same spirit than assumption A7. It restricts the dependence between α and the covariates to a dependence on the mean. Assumption A11 is the assumption of homoskedasticity in m; it is very similar to assumption A6. The difference between both assumptions stems from the identifying equation we use in both cases. In the discrete model, α remains in expression (4) and thus its variance must be modeled as well as its covariance with the $(\varepsilon_i)_{1 \leq i \leq m}$.[14] Lastly, assumption A12 is a condition of large support. It especially implies that $m \geq 3$. Otherwise, indeed, the two variables belong to a line in $\mathbb{R}^2$.

Theorem 4. *Under assumptions A1, A2'', A5, A9', and A7-A12 and if* $\beta_{20} \neq \beta_{20}^1 \beta_{10}$, λ_0 *is also identified.*

4.4 Estimation

In this section we restrict to the case where only $\mathbb{1}\{y_i^* \geq 0\}$ is observed, as the continuous case is analyzed in full details in Lee (2007). We also restrict ourselves

[14] The assumption of no covariance is not restrictive. Indeed, if the correlation between ε_i and α is not zero and independent of i, we can always reparametrize the model in order to make them uncorrelated.

to the following parametric setting with homoscedasticity.

A13. The $(\varepsilon_i)_{1\leq i\leq m}$ are i.i.d. and $\varepsilon_i \sim \mathcal{N}(0,1)$.

A14. $\alpha|\overline{x}, m \sim \mathcal{N}(\gamma_0(m) + \delta_0(m)\overline{x}, \sigma_0^2)$.

Assumption A13 imposes the normality of the residuals. This assumption is also imposed by Lee (2007) when developing his conditional maximum likelihood estimator, or by McMillen (1992) and Beron and Vijverberg (2004), among others, when studying spatially dependent discrete choice models. Contrarily to the previous section, we adopt here the usual normalization by supposing that the variance of the residuals is equal to one. Assumption A14 can be divided into two parts. First, it strengthens assumption A10 and A11 to a linear dependence à la Mundlak (1961) between α and $\overline{x}$, conditional on m. Note that on the other hand, we remain very flexible on the dependence between α and m. Second, it imposes the normality of the residual term, in a similar way to the standard random effect probit.

Under these conditions, the model is fully identified, as in theorem 5 but in a more direct way. Indeed, β_{10} and β_{20} can be identified through group size variations. Moreover, the model can be written in this case as

$$y_i = \mathbb{1}\left\{\gamma_0'(m) + x_i\left(\beta_{10} - \frac{\beta_{20}}{m-1}\right) + \overline{x}\,\delta_0'(m) - u_i \geq 0\right\} \tag{4.4.1}$$

where $\gamma_0'(m)$ and $\delta_0'(m)$ depend on $\gamma_0(m)$, $\delta_0(m)$ and the parameters of the model, and the error term u_i combines the $(\varepsilon_i)_{i=1\ldots m}$ with the residual $\alpha - \gamma_0(m) - \delta_0(m)\overline{x}$. The vector $(u_i)_{i=1,\ldots,m}$ is normal and exchangeable, with

$$\begin{aligned} V(u_i) &= 1 + \sigma_0^2 + \lambda_0(m)(2+\lambda_0(m))\left(\sigma_0^2 + \tfrac{1}{m}\right), \\ Cov(u_i, u_j) &= \sigma_0^2 + \lambda_0(m)(2+\lambda_0(m))\left(\sigma_0^2 + \tfrac{1}{m}\right) \qquad \forall\, i \neq j. \end{aligned} \tag{4.4.2}$$

One can show that making m vary enables to separate in the covariances (or in the variance) λ_0 from σ_0^2.

Now, we suppose to observe a sample of R groups where, for the sake of simplicity, all members of each groups are observed (actually, we only need observing $\overline{x}$). Hence, for group r, we observe m_r, the vector of outcomes $\widetilde{Y}_r = (y_{r1}, ..., y_{rm_r})$ and the vector of characteristics $\widetilde{X}_r = (x_{r1}, ..., x_{rm_r})$. We suppose that the $(m_r)_{1\leq r\leq R}$ are i.i.d., and that $(\widetilde{X}_r, \alpha_r, \widetilde{\varepsilon}_r)_{1\leq r\leq R}$ are independent and distributed according to

$F_{\widetilde{X},\alpha,\widetilde{\varepsilon}|m,n}$. In the previous example of peer effects in the classrooms, this condition imposes that there is no spillovers between classrooms.

Let $\theta = (\beta_1, \beta_2, \lambda, \sigma^2, (\gamma'(u), \delta'(u))_{u\in \text{Supp}(m)})$ denote the vector of all parameters. Under the previous i.i.d. assumption, the likelihood of the whole sample satisfies

$$L(\widetilde{Y}_1, ..., \widetilde{Y}_R | m_1, ..., m_R, \widetilde{X}_1, ...\widetilde{X}_R, \theta) = \prod_{r=1}^{R} L(\widetilde{Y}_r | m_r, \widetilde{X}_r, \theta).$$

where $L(\widetilde{Y}_r | m_r, \widetilde{X}_r, \theta)$ denotes the likelihood for group r. Moreover, by (4.4.1), this likelihood writes as

$$L(\widetilde{Y}_r | m_r, \widetilde{X}_r, \theta) = P\left[(2y_{r1} - 1)u_{r1} \leq (2y_{r1} - 1)\left(\gamma'(m_r) + x_{r1}\left(\beta_1 - \frac{\beta_2}{m_r - 1}\right) + \overline{x}_r\delta'(m_r)\right),\right.$$

$$\left. ..., (2y_{rm_r} - 1)u_{rm_r} \leq (2y_{rm_r} - 1)\left(\gamma'(m_r) + x_{rm_r}\left(\beta_1 - \frac{\beta_2}{m_r - 1}\right) + \overline{x}_r\delta'(m_r)\right)\right]$$

This is the probability that a multivariate normal vector belongs to an hyper-rectangle in $\mathbb{R}^{m_r}$. Such a probability can be estimated for instance by the GHK algorithm (Geweke, 1989, Keane, 1994 and Hajivassiliou et al., 1996). Thus, the model can be estimated by simulated maximum likelihood.

4.5 Monte Carlo simulations

In this section, we investigate the finite sample performance of our estimator. The sample data are generated with one regressor $x_{ri} \sim \mathcal{N}(0, 4)$, the (x_{ri}) being independent for all r and i. The true parameters are $\beta_{10} = 1$, $\beta_{20} = 0.2$, $\lambda = 0.1$, $\sigma_0^2 = 0.1$, $\gamma(m) = 0$ for all m and $\delta(m) = 0.1$ for all m. As Lee (2007), we consider a case where the average size group is small, and another where it is relatively large. In the first one, the group sizes vary from 3 to 8 the number of groups of each size being the same. In the relatively large case, they range from 15 to 25. The first case could be realistic for groups of good friends or roommates for instance, whereas the second one rather corresponds to groups of students in a classroom. In each case, we consider different sample sizes from 330 to 21,120. In the GHK algorithm, we used Halton sequences instead of standard uniform random numbers as they improve, on average, the accuracy of the integral estimation (see e.g. Sándor & András, 2004). In the small group case where the dimension of the integral is low,

we rely on 25 replications, whereas we utilize 50 replications in the large group case.

The first striking point is that sample sizes must be quite large to obtain satisfactory results on the estimations. If we compare the results of our small groups scenario with the one of Lee (see Lee, 2007, table 1, model SG-SX), it seems that observing a binary measure of y_i^* instead of y_i^* itself leads to rather large biases for even moderately large sample sizes.[15] In particular, in both the small and large groups scenario, the bias on λ_0 is systematically negative for small to moderate sample size. The second striking result is the influence of the group sizes. The accuracy of the estimator of β_{20} in large groups is approximately the same as the one in small groups, but with a sample four times larger. This is not surprising, since identification of peer effects becomes weak as the sample size increases (see Lee, 2007). The parameter λ_0 is also better estimated with small groups, but the difference between the two designs seems to reduce as the sample size grows. On the other hand, and quite surprisingly, the estimator of β_{10} is more precise in large groups.

4.6 Conclusion

This paper considers identification and estimation of social interaction models using group size variation. Provided that the size of the group is known and varies sufficiently, endogenous and exogenous effects can be identified without any exclusion restriction in the linear model. The result is extended to a binary model. In this case, exogenous peer effects are also identified under weak assumptions. Identification of endogenous peer effects is more stringent, as it requires an homoscedastic condition and restrictions on the dependence between the fixed group effects and covariates.

Our paper has two main limitations. First, the size of the group is assumed to be known. However, as emphasized by Manski & Pepper (2000), it is often difficult to

[15] Note that it is difficult to compare our large group scenario with the one of Lee, since he considers a model with two independent covariates x_{1i} and x_{2i} such that x_{1i} has only a direct effect on y_i (i.e., $\beta_{20}^1 = 0$), while x_{2i} affects y_i only through exogenous peer effects (so that $\beta_{01}^2 = 0$).

Sample		Small groups		Large groups	
size	Parameter	Mean	Std. err.	Mean	Std. err.
660	β_{10}	0.9975	0.2254	1.0128	0.1658
	β_{20}	0.8956	0.8877	1.4445	2.7601
	λ_0	-0.0304	0.5688	-0.3801	0.6600
1320	β_{10}	1.0029	0.1198	1.0025	0.0865
	β_{20}	0.9823	0.4885	0.9780	1.4712
	λ_0	0.1158	0.3458	-0.0026	0.3093
2640	β_{10}	0.9936	0.0951	0.9978	0.0678
	β_{20}	0.9378	0.3739	1.0761	0.8625
	λ_0	0.1831	0.1405	0.1247	0.1833
5280	β_{10}	0.9904	0.0664	1.0001	0.0419
	β_{20}	0.9744	0.2425	1.0264	0.5747
	λ_0	0.1927	0.0678	0.1620	0.1167
10560	β_{10}	0.9914	0.0451	1.0014	0.0285
	β_{20}	0.9708	0.1690	1.0240	0.4303
	λ_0	0.2000	0.0389	0.1788	0.0513
21120	β_{10}	0.9911	0.0295	0.9984	0.0180
	β_{20}	0.9872	0.1065	0.9777	0.2847
	λ_0	0.1897	0.0284	0.1950	0.0311

TAB. 4.1 – Results of the Monte Carlo simulations.

define groups on an a priori background. This criticism is common to all models of social interactions, but may be especially problematic here. Indeed, ignoring the boundaries of the group leads (among other difficulties) to measurement errors on the group size, which could prevent identification. Second, we do not consider a fully nonparametric regression. The issue of whether group size variation has an identifying power in this general case remains to be settled.

Appendix : proofs

Theorem 1

First, under assumption A3, $E(\widetilde{X}_1' W_{n_1} \widetilde{\varepsilon}_1 \mid n_1, m_1) = 0$ and thus, by assumption A4, $\beta(m)$ is identified for all $m \in \text{Supp}(m_1)$. We now prove that the knowledge of $m \mapsto \beta(m)$ allows in general to identify the structural parameters.

Let $(m_1^*, m_2^*) \in \text{Supp}(m_1)^2$, then

$$\frac{(m_1^* - 1)\beta_{10} - \beta_{20}}{m_1^* - 1 + \lambda_0} = \frac{(m_2^* - 1)\beta_{10} - \beta_{20}}{m_2^* - 1 + \lambda_0}$$

is equivalent to

$$-\lambda_0 \beta_{10} \left(\frac{1}{m_1^* - 1} - \frac{1}{m_2^* - 1} \right) = \beta_{20} \left(\frac{1}{m_1^* - 1} - \frac{1}{m_2^* - 1} \right).$$

Hence, if $\beta_{20} = -\lambda_0 \beta_{10}$, $\beta(.)$ is constant, and if not, $\beta(.)$ is a one-to-one mapping. In the first case, $\beta(m) = \beta_{10}$ for all m. Thus β_{10} is identified, but λ_0 cannot be identified by $\beta(.)$. Because $\beta_{20} = -\lambda_0 \beta_{10}$, β_{20} is identified up to a constant.

Now suppose that $\beta_{20} \neq -\lambda_0 \beta_{10}$ and let (m_0^*, m_1^*, m_2^*) be three different values in $\text{Supp}(m_1)$. We will prove that the knowledge of $\beta(m_0^*), \beta(m_1^*)$ and $\beta(m_2^*)$ permits to identify $(\beta_{10}, \lambda_0, \beta_{20})$. This amounts to show that the system

$$\begin{cases} \beta(m_0^*)\lambda_0 - (m_0^* - 1)\beta_{10} + \beta_{20} &= -\beta(m_0^*)(m_0^* - 1) \\ \beta(m_1^*)\lambda_0 - (m_1^* - 1)\beta_{10} + \beta_{20} &= -\beta(m_1^*)(m_1^* - 1) \\ \beta(m_2^*)\lambda_0 - (m_2^* - 1)\beta_{10} + \beta_{20} &= -\beta(m_2^*)(m_2^* - 1) \end{cases}$$

has a unique solution. Using the matrix form, we can rewrite the system as $A\zeta_0 = B$ where $\zeta_0 = (\lambda_0, \beta_{10}, \beta_{20})'$. If $\det(A) \neq 0$, ζ_0 is identified. Suppose that $\det(A) = 0$. Then $com(A)'B = 0$ where $com(A)$ denotes the comatrix of A. By using the first line of this equation and the expression of $det(A)$, we get

$$\begin{cases} (m_2^* - m_1^*)\beta(m_0^*) + (m_0^* - m_2^*)\beta(m_1^*) + (m_1^* - m_0^*)\beta(m_2^*) = 0 \\ (m_0^* - 1)(m_2^* - m_1^*)\beta(m_0^*) + (m_1^* - 1)(m_0^* - m_2^*)\beta(m_1^*) + (m_2^* - 1)(m_1^* - m_0^*)\beta(m_2^*) = 0. \end{cases}$$

Hence,

$$\begin{cases} (m_2^* - m_1^*)\beta(m_0^*) = -(m_0^* - m_2^*)\beta(m_1^*) - (m_1^* - m_0^*)\beta(m_2^*) \\ \\ m_0^*(m_2^* - m_1^*)\beta(m_0^*) + m_1^*(m_0^* - m_2^*)\beta(m_1^*) + m_2^*(m_1^* - m_0^*)\beta(m_2^*) = 0. \end{cases}$$

Thus,

$$\begin{cases} (m_2^* - m_1^*)\beta(m_0^*) + (m_0^* - m_2^*)\beta(m_1^*) + (m_1^* - m_0^*)\beta(m_2^*) = 0 \\ \\ \beta(m_1^*)(m_0^* - m_2^*)(m_1^* - m_2^*) + \beta(m_0^*)(m_2^* - m_1^*)(m_0^* - m_2^*) = 0. \end{cases}$$

Because $m_1^* \neq m_2^*$ and $m_0^* \neq m_2^*$, this implies that $\beta(m_1^*) = \beta(m_0^*)$, which is in contradiction with the fact that $\beta(.)$ is a one-to-one mapping. Thus $\det(A) \neq 0$ and ζ_0 is identified.

Theorem 2

Because $m \mapsto \beta(m)$ is identified, $\text{Var}\left(\dfrac{W_{n_1}\widetilde{\varepsilon}_1}{1 + \dfrac{\lambda_0}{m_1 - 1}} \mid n_1, m_1 \right)$ is known. Thus, under assumption A6,

$$\text{Var}\left(\frac{W_{n_1}\widetilde{\varepsilon}_1}{1 + \frac{\lambda_0}{m-1}} \mid n_1, m_1 \right) = \frac{\sigma^2}{\left(1 + \frac{\lambda_0}{m_1 - 1}\right)^2} W_{n_1}.$$

Hence, for $m_1^* \neq m_2^*$,

$$C \equiv \frac{\left(1 + \frac{\lambda_0}{m_1^* - 1}\right)^2}{\left(1 + \frac{\lambda_0}{m_2^* - 1}\right)^2}$$

is identified. Under assumption A5, $\left(1 + \frac{\lambda_0}{m-1}\right) > 0$ for all $m \in \text{Supp}(m_1)$. Thus

$$\left(\frac{\sqrt{C}}{m_1^* - 1} - \frac{1}{m_2^* - 1} \right) \lambda_0 = 1 - \sqrt{C}.$$

It is clear that $\left(\frac{\sqrt{C}}{m_1^* - 1} - \frac{1}{m_2^* - 1} \right) \neq 0$. Otherwise $C = 1$ and then $m_1^* = m_2^*$, which contradicts the assumption. Thus λ_0 is identified.

Then, because $m \mapsto \beta(m)$ is identified, $\beta_{10} - \dfrac{\beta_{20}}{m - 1}$ is known for all $m \in \text{Supp}(m_1)$. Taking two different values for m permits to identify β_{20}, and then β_{10}.

Lemma 1

Applying the between-group operator to (1) gives

$$\overline{y^*} = \overline{x}\left(\frac{\beta_{10}+\beta_{20}}{1-\lambda_0}\right) + \frac{\alpha}{1-\lambda_0} + \frac{\overline{\varepsilon}}{1-\lambda_0},$$

since $1/(1-\lambda_0)$ exists according to assumption A5. Consequently, replacing $\overline{y^*}$ in equation (1), we obtain

$$\begin{aligned} y_i^*\left(1+\tfrac{\lambda_0}{m-1}\right) = \;& x_i\left(\beta_{10}-\tfrac{\beta_{20}}{m-1}\right) + \overline{x}\tfrac{m}{m-1}\left(\beta_{20}+\tfrac{\beta_{10}+\beta_{20}}{1-\lambda_0}\lambda_0\right) \\ & +\alpha\left(1+\tfrac{m}{m-1}\tfrac{\lambda_0}{1-\lambda_0}\right) + \overline{\varepsilon}\tfrac{m}{m-1}\tfrac{\lambda_0}{1-\lambda_0} + \varepsilon_i. \end{aligned}$$

Note that this equation is equivalent to (1). Now, under assumption A5, $1+\lambda_0/(m-1) > 0$, so that $y_i^* \geq 0$ if and only if $y_i^*\left(1+\frac{\lambda_0}{m-1}\right) \geq 0$. Thus, under assumption A5, $y_i = 1\{y_i^* \geq 0\}$, where y_i^* satisfies equation (1), is observationally equivalent to y_i satisfying equation (4).

Theorem 3

Assumption A7 implies that the conditional distribution of $\overline{\varepsilon}\frac{m}{m-1}\frac{\lambda_0}{1-\lambda_0} + \varepsilon_i$ is identical for every i. Thus assumption 1 in Manski (1987) is satisfied and, using A8 and A9, we can apply directly his result to identify $\frac{(m-1)\beta_{10}-\beta_{20}}{\left|m-1-\beta_{20}^1\right|}$. The first term of the vector, $\frac{(m-1)\beta_{10}^1-\beta_{20}^1}{\left|m-1-\beta_{20}^1\right|}$, is also identified. By assumption A9,

$$\widetilde{\beta}(m) \equiv \frac{(m-1)\beta_{10}-\beta_{20}}{m-1-\beta_{20}^1} = \frac{\frac{(m-1)\beta_{10}-\beta_{20}}{\left|m-1-\beta_{20}^1\right|}}{\frac{(m-1)\beta_{10}^1-\beta_{20}^1}{\left|m-1-\beta_{20}^1\right|}},$$

so that $\widetilde{\beta}(m)$ is identified as the ratio of two known terms. The rest of the proof of identification of (β_{10}, β_{20}) follows the same line than the one of Theorem 1, λ_0 being replaced by $-\beta_{20}^1$.

However, λ_0 cannot be identified. Indeed, let $\lambda_0' \neq \lambda_0$ and define

$$\varepsilon_i' = \varepsilon_i + \overline{\varepsilon}\frac{m(\lambda_0-\lambda_0')}{(m-1+\lambda_0')(1-\lambda_0)}.$$

Finally let

$$\alpha' = \frac{m\overline{x}(\beta_{10}+\beta_{20})(\lambda_0-\lambda_0') + \alpha(m-1+\lambda_0)(1-\lambda_0')}{(m-1+\lambda_0')(1-\lambda_0)}.$$

Then $(\lambda_0', \alpha', \varepsilon_1', ..., \varepsilon_m')$ are observationally equivalent to the initial model. Indeed, we can check that they lead to (4.3.2) as well. Moreover, conditioning on $(m, x_1, ..., x_m, \alpha')$ is equivalent to conditioning on $(m, x_1, ..., x_m, \alpha)$, and conditional exchangeability of $(\varepsilon_1, ..., \varepsilon_m)$ implies conditional exchangeability of the $(\varepsilon_1', ..., \varepsilon_m')$. Furthermore, letting F_u denote the c.d.f. of any random variable u,

$$F_{\varepsilon_1' + \overline{\varepsilon'} \frac{m}{m-1} \frac{\lambda_0'}{1-\lambda_0'} | m=m^*, x_1=x_1^*, ..., x_m=x_m^*, \alpha'=\alpha'^*} = F_{\varepsilon_1 + \overline{\varepsilon} \frac{m}{m-1} \frac{\lambda_0}{1-\lambda_0} | m=m^*, x_1=x_1^*, ..., x_m=x_m^*, \alpha=\alpha^*},$$

where

$$\alpha^* = \frac{(m - 1 + \lambda_0')(1 - \lambda_0)\alpha'^* - m\overline{x}(\beta_{10} + \beta_{20})(\lambda_0 - \lambda_0')}{(m - 1 + \lambda_0)(1 - \lambda_0')}.$$

Thus the second part of assumption A7 also holds with $(\lambda_0', \alpha', \varepsilon_1', ..., \varepsilon_m')$. This shows that λ_0 is not identified.

Theorem 4

Let $\theta_0 = \dfrac{\lambda_0}{1 - \lambda_0}$ and

$$v_i = \left[\overline{x} \frac{m}{m-1} \left[\beta_{20} + \theta_0(\beta_{10} + \beta_{20})\right] + \alpha \left(1 + \frac{m}{m-1}\theta_0\right)\right] + \overline{\varepsilon} \frac{m}{m-1}\theta_0 + \varepsilon_i.$$

Note that $F_{v_1,...,v_m|x_1,...,x_m,m} = F_{v_1,...,v_m|\overline{x},m}$. Indeed

$$\begin{aligned}
&F_{v_1,...,v_m|x_1,...,x_m,m}(v_1^*, ..., v_m^* | x_1^*, ..., x_m^*, m^*) \\
&= \int F_{v_1,...,v_m|x_1,...,x_m,m,\alpha}(v_1^*, ..., v_m^* | x_1^*, ..., x_m^*, m^*, \alpha^*) dF_{\alpha|x_1,...,x_m,m}(\alpha^* | x_1^*, ..., x_m^*, m^*) \\
&= \int F_{v_1,...,v_m|\overline{x},\alpha,m}(v_1^*, ..., v_m^* | \overline{x^*}, \alpha^*, m^*) dF_{\alpha|\overline{x},m}(\alpha^* | \overline{x^*}, m) \\
&= F_{v_1,...,v_m|\overline{x},m}(v_1^*, ..., v_m^* | \overline{x^*}, m^*),
\end{aligned}$$

where the third line stems from assumption A10 and the fact that, given $x_1, ..., x_m, m, \alpha$, $(v_1, ..., v_m)$ is a deterministic function of $(\varepsilon_1, ..., \varepsilon_m)$. Now

$$\begin{aligned}
&\Pr(y_1 = 0, y_2 = 0 \mid x_1 = x_1^*, x_2 = x_2^*, \overline{x} = x, m = m^*) \\
&= \Pr\left\{v_1 \leq -x_1^* \left(\beta_{10} - \tfrac{\beta_{20}}{m-1}\right),\ v_2 \leq -x_2^* \left(\beta_{10} - \tfrac{\beta_{20}}{m-1}\right) \mid x_1 = x_1^*, x_2 = x_2^*, \overline{x} = x, m = m^*\right\} \\
&= F_{v_1,v_2|\overline{x},m}\left(-x_1^* \left(\beta_{10} - \tfrac{\beta_{20}}{m^*-1}\right), -x_2^* \left(\beta_{10} - \tfrac{\beta_{20}}{m^*-1}\right) | x, m^*\right).
\end{aligned}$$

Because, by theorem 4, (β_{10}, β_{20}) is identified, $x_1^* \left(\beta_{10} - \frac{\beta_{20}}{m^*-1}\right)$ and $x_2^* \left(\beta_{10} - \frac{\beta_{20}}{m^*-1}\right)$ are known. Moreover, $\overline{x}$ is observed so that the first term is identified on the whole support of (x_1, x_2). Thus, by assumption A12, making (x_1, x_2) vary allows us to

identify the whole conditional distribution of (v_1, v_2) given $\overline{x}$ and m. Thus, using assumption A11,

$$\text{Cov}(v_1, v_1 - v_2 \mid \overline{x}, m) = \text{Cov}\left(\overline{\varepsilon}\frac{m}{m-1}\theta_0 + \varepsilon_1, \varepsilon_1 - \varepsilon_2 \mid \overline{x}, m\right) = \text{Var}(\varepsilon_1 \mid \overline{x}),$$

so that the right term is identified. Moreover, a little algebra shows that

$$\begin{aligned}(m-1)^2\text{Cov}(v_1, v_2|\overline{x}, m) &= m^2\left[(1+\theta_0)^2\text{Var}(\alpha|\overline{x})\right] + m\Big[-2(1+\theta_0)\text{Var}(\alpha|\overline{x}) \\ &\quad +\theta_0(2+\theta_0)\text{Var}(\varepsilon_1|\overline{x})\Big] + \left[\text{Var}(\alpha|\overline{x}) - 2\theta_0\text{Var}(\varepsilon_1|\overline{x})\right].\end{aligned}$$

Conditional on $\overline{x}$, this is a regression of the (known) left term on $(m^2, m, 1)$. By $A2''$, there exists a set A of positive probability such that m can take three different values with positive probability conditional on $\overline{x} = x^*$, for all $x^* \in A$. Thus, the coefficients (a, b, c) of this regression can be identified.[16] We will show that the knowledge of these coefficients implies that θ_0 is identified. The conclusion will follow because θ_0 is one-to-one with λ_0.

First, let $\phi_0 = 1+\theta_0$ and $\rho_0 = \frac{\text{Var}(\alpha|\overline{x})}{\text{Var}(\varepsilon|\overline{x})}$. Let also $a' = a/\text{Var}(\varepsilon|\overline{x})$, $b' = b/\text{Var}(\varepsilon|\overline{x})+1$ and $c' = c/\text{Var}(\varepsilon|\overline{x}) - 2$. Then a', b' and c' are identified, and

$$\begin{cases} \phi_0^2\rho_0 &= a' \\ -2\phi_0\rho_0 + \phi_0^2 &= b' \\ \rho_0 - 2\phi_0 &= c'. \end{cases}$$

Replacing ρ_0 by $c' + 2\phi_0$ in the first and second equation leads to

$$\begin{cases} \phi_0^3 + c'/2\phi_0^2 - a'/2 &= 0 \\ \phi_0^2 + 2c'/3\phi_0 + b'/3 &= 0 \\ \rho_0 - 2\phi_0 &= c'. \end{cases} \tag{4.6.1}$$

This system admits at most two solutions in (ρ, ϕ). Suppose that there are two different solutions, and let (ρ_1, ϕ_1) denote the second one. Then we can write the polynomial of the first equation as a product in which one factor is the polynomial of the second equation. Hence, there exists x such as, for all $\phi \in \mathbb{R}$,

$$\phi^3 + c'/2\phi^2 - a'/2 = (\phi^2 + 2c'/3\phi^2 + b'/3)(\phi + x).$$

[16] These coefficients depend on $\overline{x}$ but for the sake of simplicity, we let this dependency implicit from now on.

Thus

$$\begin{cases} x &= -c'/6 \\ 2c'x &= -b' \\ 2b'x &= -3a'. \end{cases}$$

Hence $c'^2 = 3b'$. Replacing b' and c' by their expression gives

$$3(-2\phi\rho + \phi^2) = (\rho - 2\phi)^2,$$

which must hold for (ρ_0, ϕ_0) and (ρ_1, ϕ_1). But this statement is equivalent to $\phi+\rho = 0$. Replacing ρ by $-\phi$ in c' gives $\phi_0 = \phi_1 = -c'/3$ and thus also $\rho_0 = \rho_1$. This contradicts $(\rho_0, \phi_0) \neq (\rho_1, \phi_1)$. Thus ϕ_0 is identified by 4.6.1 and the conclusion follows.

Bibliographie

Abbring, J. H., Heckman, J. J., Chiappori, P. A. & Pinquet, J. (2003), 'Adverse selection and moral hazard in insurance : Can dynamic data help to distinguish ?', *Journal the European Economic Association* **1**, 512–521.

Akerlof, G. (1970), 'The market for lemons : Quality uncertainty and the market mechanism', *The Quarterly Journal of Economics* **84**, 488–500.

Ammermueller, A. & Pischke, J. S. (2006), Peer effects in european primary schools : Evidence from pirls. IZA Discussion Paper No. 2077.

Andrews, D., Berry, S. & Jia, P. (2003), Placing bounds on parameters of entry games in the presence of multiple equilibria. Mimeo.

Angrist, J. D. & Imbens, G. W. (1994), 'Identification and estimation of local average treatment effects', *Econometrica* **62**, 467–475.

Angrist, J. D., Imbens, G. W. & Rubin, D. B. (1996), 'Identification of causal effects using instrumental variables', *Journal of the American Statistical Association* **91**, 444–455.

Athey, S. & Haile, P. (2007), Nonparametric approaches to auctions, *in* 'Handbook of econometrics', Vol. 6, North Holland.

Athey, S. & Haile, P. A. (2002), 'Identification of standard auction models', *Econometrica* **70**, 2107–2140.

Ausubel, L. M. (1999), Adverse selection in the credit card market. Mimeo.

Bachelard, G. (1934), *Le nouvel esprit scientifique*, Presses Universitaires de France.

Bahadur, R. R. & Savage, L. J. (1956), 'The nonexistence of certain statistical procedures in nonparametric problems', *Annals of Mathematical Statistics* **27**, 1115–1122.

Bajari, P. & Hortaçsu, A. (2005), 'Are structural estimates of auction models reasonable ? evidence from experimental data', *Journal of Political Economy* **113**, 703–741.

Baker, S. G. & Laird, N. M. (1988), 'Regression analysis for categorical variables with outcome subject to nonignorable nonresponse', *Journal of the American Statistical Association* **83**, 62–69.

Banerjee, A., Gertler, P. & Ghatak, M. (2002), 'Empowerment and efficiency : Tenancy reform in west bengal', *Journal of Political Economy* **110**, 239–280.

Baron, D. P. & Myerson, R. B. (1982), 'Regulating a monopolist with unknown costs', *Econometrica* **50**, 911–930.

Bayer, P. & Timmins, C. (2007), 'Estimating equilibrium models of sorting across location', *Economic Journal* **117**, 353–374.

Beron, K. J. & Vijverberg, W. P. M. (2004), Probit in a spatial context : A monte carlo analysis, *in* L. Anselin, R. J. G. M. Florax & S. J. Rey, eds, 'Advances in Spatial Econometrics : Methodology, Tools and Applications', Berlin : Springer-Verlag.

Bierens, H. J. (1982), 'Consistent model specification tests', *Journal of Econometrics* **20**, 105–134.

Bissantz, N., Hohage, T. & Munk, A. (2004), 'Consistency and rates of convergence of nonlinear tikhonov regularization with random noise', *Inverse Problems* **20**, 1773–1789.

Blume, L. & Durlauf, S. (2005), Identifying social interactions : A review. Mimeo, University of Wisonsin.

Blundell, R., Chen, X. & Kristensen, D. (2007), 'Nonparametric IV estimation of shape-invariant engel curves', *Econometrica* **75**, 1613–1669.

Bramoullé, Y., Djebbari, H. & Fortin, B. (2009), 'Identification of peer effects through social networks', *Journal of econometrics* **Forthcoming**.

Brock, W. A. & Durlauf, S. M. (2001), 'Discrete choice with social interactions', *Review of Economic Studies* **68**, 235–260.

Brock, W. A. & Durlauf, S. M. (2007), 'Identification of binary choice models with social interactions', *Journal of Econometrics* **140**, 52–75.

Calvó-Armengol, J., Pattacchini, E. & Zenou, Y. (2008), 'Peer effects and social networks in education', *Review of Economic Studies* **Forthcoming**.

Card, D. (2001), 'Estimating the return to schooling : Progress on some persistent econometric problems', *Econometrica* **69**, 1127–1160.

Carrasco, M., Florens, J. P. & Renault, E. (2006), Linear inverse problems and structural econometrics : Estimation based on spectral decomposition and regularization, *in* J. J. Heckman & E. E. Leamer, eds, 'Handbook of Econometrics', Vol. 6, North Holland.

Case, A. (1992), 'Neighborhood influence and technological change', *Regional Science and Urban Economics* **22**, 491–508.

Chamberlain, G. (1986), 'Asymptotic efficiency in semiparametric model with censoring', *Journal of Econometrics* **32**, 189–218.

Chen, K. (2001), 'Parametric models for response-biased sampling', *Journal of the Royal Statistical Society, Series B* **63**, 775–789.

Chen, X. & Hu, Y. (2006), Identification and inference of nonlinear models using two samples with arbitrary measurement errors. Cowles foundation discussion paper no. 1590.

Chernozhukov, V. & Hansen, C. (2005), 'An iv model of quantile treatment effects', *Econometrica* **73**, 245–261.

Chernozhukov, V., Imbens, G. W. & Newey, W. K. (2007), 'Nonparametric identification and estimation of non-separable models', *Journal of Econometrics* pp. –.

Chesher, A. (2003), 'Identification in nonseparable models', *Econometrica* **71**, 1405–1441.

Chesher, A. (2007), Identification of non-additive structural functions, *in* W. K. Newey, R. Blundell & T. Persson, eds, 'Advances in Economics and Econometrics : Theory and Applications, Ninth World Congress', Vol. 3, New York : Cambridge University Press.

Chesher, A. (2008), Instrumental variable models for discrete outcomes. cemmap working paper CWP30/08.

Chiappori, P. A., Durand, F. & Geoffard, P. (1998), 'Moral hazard and the demand for physician services : First lessons from a french natural experiment', *European Economic Review* **42**, 499–511.

Chiappori, P. A., Geoffard, P. & Kyriadizou, E. (1998), Cost of time, moral hazard, and the demand for physician services. Mimeo, University of Chicago.

Chiappori, P. A., Jullien, B., Salanié, B. & Salanié, F. (2006), 'Asymmetric information in insurance : General testable implications', *The RAND Journal of Economics* pp. –.

Chiappori, P. A. & Salanié, B. (2000), 'Testing for asymmetric information in insurance markets', *Journal of Political Economy* **108**, 56–78.

Chiappori, P. A. & Salanié, B. (2002), Testing contract theory : A survey of some recent work, *in* L. H. e. S. T. M. Dewatripont, ed., 'Advances in Economics and Econometrics', Vol. 1, New York : Cambridge University Press.

Ciliberto, F. & Tamer, E. (2006), Market structure and multiple equilibria in airline markets. Mimeo.

Cooley, J. (2007), Desegregation and the achievement gap : Do diverse peers help ? Working Paper, Duke University.

Cosnefroy, O. & Rocher, T. (2004), 'Le redoublement au cours de la scolarité obligatoire : nouvelles analyses, mÃªmes constats', *Education et Formation* **70**, 73–82.

Crahaye, M. (1996), *Peut-on lutter contre l'échec scolaire ?*, De Boeck.

Cross & Manski, C. F. (2002), 'Regression, short and long', *Econometrica* **70**, –.

Darolles, S., Florens, J. P. & Renault, E. (2007), Nonparametric instrumental regression. Working paper.

Das, M. (2005), 'Instrumental variables estimators of nonparametric models with discrete endogenous regressors', *Journal of Econometrics* **124**, 335–361.

Deville, J. C. (2001), La correction de la non-réponse par calage généralisé. mimeo INSEE UMS-DSDS.

Devroye, L. (1989), 'Consistent deconvolution in density estimation', *Canadian Journal of Statistics* **17**, 235–239.

Diamond, P. (1998), 'Optimal income taxation : An example with a u-shaped pattern of optimal marginal tax rates', *The American Economic Review* **88**, 83–95.

Dionne, G. & Vanasse, C. (1996), 'Une évaluation empirique de la nouvelle tarification de l'assurance automobile au québec', *Actualité économique* **73**, 47–80.

Donald, S. & Paarsch, H. (1996), 'Identification, estimation and testing in parametric empirical models of auctions within the independent private values paradigm', *Econometric Theory* **12**, 517–567.

Duflo0, E. & Saez, E. (2003), 'The role of information and social interactions in retirement plan decisions : Evidence from a randomized experiment', *Quarterly Journal of Economics* **118**, 815–842.

Dufour, J. M. (2003), 'Identification, weak instruments and statistical inference in econometrics. presidential address to the canadian economics association', *Canadian Journal of Economics* **36**, 767–808.

Elyakime, B., Laffont, J.-J., Loisel, P. & Vuong, Q. (1994), 'First-price sealed-bid auctions with secret reservation prices', *Annales d'Economie et de Statistique* **34**, 115–141.

Elyakime, B., Laffont, J.-J., Loisel, P. & Vuong, Q. (1997), 'Auctioning and bargaining : An econometric study of timber auctions with secret reservation prices', *Journal of Business and Economic Statistics* **15**, 209–220.

Fan, J. & Truong, Y. K. (1993), 'Nonparametric regression with errors in variables', *The Annals of Statistics* **21**, 1900–1925.

Fay, R. E. (1986), 'Causal models for patterns of nonresponse', *Journal of the American Statistical Association* **81**, 354–365.

Ferrall, C. & Shearer, B. (1999), 'Incentives and transaction costs within the firm : Estimating an agency model using payroll records', *Review of Economic Studies* **99**, 309–338.

Février, P. (2003). Thèse de doctorat, Université Paris 1.

Février, P. (2007), Nonparametric identification and estimation of a common value auction model. Crest working paper.

Florens, J., Heckman, J. J., Meghir, C. & Vytlacil, E. (2008), 'Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects', *Econometrica* **76**, 1191–1206.

Florens, J. P., Heckman, J. J., Meghir, C. & Vytlacil, E. (2003), Instrumental variables, local instrumental variables and control functions. IDEI Working Paper 249.

Florens, J. P., Mouchart, M. & Rolin, J. M. (1990), *Elements of Bayesian statistics*, Marcel Dekker, New York - Basel.

Freixas, X. & Laffont, J. J. (1990), Optimal banking contracts, *in* 'Essays in Honor of Edmond Malinvaud, vol. 2 : Macroeconomics', MIT Press, pp. 33–61.

Freixas, X. & Rochet, J. C. (1997), *The Microeconomics of Banking*, MIT Press.

Gagliardini, P. & Scaillet, O. (2006), Tikhonov regularization for functional minimum distance estimators. Working Paper.

Gagnepain, P. & Ivaldi, M. (2002), 'Incentive regulatory policies : The case of public transit systems in france', *The RAND Journal of Economics* **33**, 605–629.

Gagnepain, P. & Ivaldi, M. (2007), Contract choice, incentives, and political capture in public transit. Mimeo.

Gence-Creux, C. (2000), Identification of non-additive structural functions, *in* 'Econometric Society World Congress 2000 Contributed Papers', New York : Cambridge University Press.

Geweke, J. (1989), 'Bayesian inference in econometric models using monte carlo integration', *Econometrica* **57**, 1317–1339.

Ghosh, J. K. & Singh, R. (1966), 'Unbiased estimation of location and scale parameters', *The Annals of Statistics* **37**, 1671–1675.

Glaeser, E., Sacerdote, B. & Scheinkman, J. (1996), 'Crime and social interaction', *Quarterly Journal of Economics* **111**, 507–548.

Gouriéroux, C. & Monfort, A. (1995), *Statistics and Econommetric Models*, Cambridge University Press.

Graham, B. S. (2008), 'Identifying social interactions through excess variance contrasts', **76**. Econometrica.

Graham, B. S. & Hahn, J. (2005), 'Identification and estimation of the linear-in-means model of social interactions', *Economics Letters* **88**, 1–6.

Grogger, J. T. & Carson, R. T. (1991), 'Models for truncated counts', *Journal of Applied Econometrics* **6**, 225–238.

Guerre, E., Perrigne, I. & Vuong, Q. (2000), 'Optimal nonparametric estimation of first-price auctions', *Econometrica* **68**, 525–574.

Guerre, E., Perrigne, I. & Vuong, Q. (2008), Nonparametric identification of risk aversion in frist-price auctions under exclusion restrictions. Mimeo.

Gyorfi, L., Kohler, M., Kryzak, A. & Walk, H. (2002), *A Distribution-Free Theory of Nonparametric Regression*, New York : Springer.

Haile, P. A. & Tamer, E. (2003), 'Inference with an incomplete model of english auctions', *Journal of Political Economy* **111**, 1–51.

Hajivassiliou, V. A., McFadden, D. & Ruud, P. (1996), 'Simulation of multivariate normal rectangle probabilities and their derivatives : theoretical and computational results', *Journal of Econometrics* **72**, 85–134.

Hall, P. & Horowitz, J. L. (2005), 'Nonparametric methods for inference in the presence of instrumental variables', *Annals of Statistics* **33**, 2904–2929.

Haurin, D. R. & Sridhar, K. S. (2003), 'The impact of local unemployment rates on reservation wages and the duration of search for a job', *Applied Economics* **35**, 1469–1475.

Heckman, J. J. (1974), 'Shadow prices, market wages, and labor supply', *Econometrica* **42**, 679–694.

Heckman, J. J. & Vytlacil, E. (2005), 'Structural equations, treatment effects, and econometric policy evaluation', *Econometrica* **73**, 669–738.

Hellerstein, J. K. & Imbens, G. W. (1999), 'Imposing moment restrictions from auxiliary data by weighting', *The Review of Economics and Statistics* **81**, 1–14.

Hemvanich, S. (2004), The general missingness problems and estimation in discrete choice models. Working Paper.

Hirano, K., Imbens, G. W., Ridder, G. & Rubin, D. B. (2001), 'Combining panel data sets with attrition and refreshment samples', *Econometrica* **69**, 1645–1659.

Hoeffding, W. (1977), 'Some incomplete and boundedly complete families of distributions', *The Annals of Statistics* **5**, 278–291.

Holmes, T. (1989), Grade level retention effects : A meta-analysis of research studies, *in* L. A. Sheppard & M. L. Smith, eds, 'Flunking Grades. Research and Policies on Retention', New York, The Falmer Press, pp. 16–33.

Horowitz, J. L. & Lee, S. (2007), 'Nonparametric instrumental variables estimation of a quantile regression model', *Econometrica* **75**, 1191–1208.

Horowitz, J. L. & Manski, C. F. (1995), 'Identification and robustness with contaminated and corrupted data', *Econometrica* **63**, 281–302.

Horvitz, D. G. & Thompson, D. J. (1952), 'A generalization of sampling without replacement from a finite universe', *Journal of the American Statistical Association* **47**, 663–685.

Hoxby, C. (2000), Peer effects in the classroom : Learning from gender and race variation. NBER Working Paper No 7867.

Hu, Y. & Schennach, S. M. (2008), 'Identification and estimation of nonclassical errors-in-variable models with continuous distributions using instruments', *Econometrica* **76**, 195–216.

Imbens, G. (2004), 'Nonparametric estimation of average treatment effects under exogeneity : a review', *The Review of Economics and Statistics* **86**, 4–29.

Imbens, G. W. & Lancaster, T. (1994), 'Combining micro and macro data in microeconometric models', *Review of Economic Studies* **61**, 655–680.

Imbens, G. W. & Newey, W. K. (2008), Identification and estimation of triangular simultaneous equations models without additivity. Working Paper.

Ivaldi, M. & Martimort, D. (1994), 'Competition under nonlinear pricing', *Annales d'Economie et de Statistique* **34**, 71–114.

Jacob, B. A. (2005), 'Accountability, incentives and behavior : Evidence from school reform in chicago', *Journal of Public Economics* **89**, 761–796.

Jacob, B. A. & Lefgren, L. (2004), 'Remedial education and student achievement : A regression-discontinuity analysis', *Review of Economics and Statistics* **86**, 226–244.

Jimerson, S. (2001), 'Meta-analysis of grade retention research : Implications for practice in the 21st century', *School Psychology Review* **30**, 420–437.

Jimerson, S., Anderson, G. E. & Whipple, A. D. (2002), 'Winning the battle and losing the war : Examining the relationship between grade retention and dropping out of high school', *Psychology in the Schools* **39**, 441–457.

Jocanovic, B. (1989), 'Observable implications of models with multiple equilibria', *Econometrica* **57**, 1431–1437.

Katz, L. F., Kling, J. R. & Liebman, J. B. (2001), 'Moving to opportunity in boston : Early results of a randomized mobility experiment', *Quaterly Journal of Economics* **116**, 607–654.

Keane, M. (1994), 'A computationally practical simulation estimator for panel data', *Econometrica* **62**, 95–116.

Klier, T. & McMillen, D. P. (2008), 'Clustering of auto supplier plants in the united states : Generalized method of moments spatial logit for large samples', *Journal of Business and Economic Statistics* **26**, 460–471.

Koopmans, T. C. & Reiersøl, O. (1950), 'The identification of structural characteristics', *Annals of Mathematical Statistics* **21**, 165–181.

Krauth, B. V. (2006), 'Simulation-based estimation of peer effects', *Journal of Econometrics* **133**, 243–271.

Laffont, J. J. & Martimort, D. (2002), *The Theory of Incentives : The Principal-Agent Model*, Princeton University Press.

Laffont, J.-J., Ossard, H. & Vuong, Q. (1995), 'Econometrics of first-price auctions', *Econometrica* **63**, 953–980.

Laffont, J. J. & Tirole, J. (1986), 'Using cost observation to regulate firms', *The Journal of Political Economy* **94**, 614–641.

Laffont, J. J. & Tirole, J. (1993), *A Theory of Incentives in Procurement and Regulation*, MIT Press.

Laffont, J. J. & Vuong, Q. (1996), 'Structural analysis of auction data', *American Economic Review* **86**, 414–420.

Lavergne, P. & Thomas, A. (2005), 'Semiparametric estimation and testing in a model of environmental regulation with adverse selection', *Empirical Economics* **30**, 171–192.

Lazear, E. (1990), 'Performance pay and productivity', *American Economic Review* **90**, 1346–1361.

LeCam, L. & Schwartz, L. (1960), 'A necessary and sufficient condition for the existence of consistent estimates', *Annals of Mathematical Statistics* **31**, 140–150.

Lee, L. F. (2007), 'Identification and estimation of econometric models with group interactions, contextual factors and fixed effects', *Journal of Econometrics* **140**, 333–374.

Lehmann, E. L. (1983), *Theory of Point Estimation*, Springer.

Lehmann, E. L. (1986), *Testing Statistical Hypothesis*, 2nd ed. Wiley : New-York.

Lehmann, E. L. & Scheffé, H. (1947), 'On the problem of similar region', *Proceedings of the National Academy of Science* **33**, 382–386.

Lewbel, A. (1997), 'Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and r & d', *Econometrica* **65**, 1201–1213.

Lewbel, A. (2000), 'Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables', *Journal of Econometrics* **97**, 145–177.

Lewbel, A. (2007), 'Endogenous selection or treatment model estimation', *Journal of Econometrics* **141**, 777–806.

Lewis, T. & Sapppington, D. (1995), 'Optimal capital structure in agency relationships', *The RAND Journal of Economics* **26**, 343–361.

Li, T., Perrigne, I. & Vuong, Q. (2000), 'Conditionally independent private information in ocs wildcat auctions', *Journal of Econometrics* **98**, 129–161.

Li, T. & Vuong, Q. (1998), 'Nonparametric estimation of the measurement error model using multiple indicators', *Journal of Multivariate Analysis* **65**, 139–165.

Little, R. & Rubin, D. B. (1987), *Statistical analysis with Missing Data*, John Wiley & Sons, New York.

Lorence, J. (2006), 'Retention and academic achievement research revisited from an united states perspective', *International Education Journal* **7**, 731–777.

Ludwig, J., Duncan, G. J. & Hierschfield, P. (2001), 'Urban poverty and juvenile crime : Evidence from a randomized housing-mobility experiment', *Quaterly Journal of Economics* **116**, 655–679.

Magnac, T. & Maurin, E. (2007), 'Identification and information in monotone binary models', *Journal of Econometrics* **139**, 76–104.

Manning, W., Newhouse, J., Duan, N., Keeler, E. & Leibowitz, A. (1987), 'Health insurance and the demand for medical care : Evidence from a randomized experiment', *American Economic Review* **77**, 251–277.

Manski, C. (1997), 'Monotone treatment response', *Econometrica* **65**, 1311–1334.

Manski, C. F. (1987), 'Semiparametric analysis of random effects linear models from binary panel data', *Econometrica* **55**, 357–362.

Manski, C. F. (1988), 'Identification of binary response models', *Journal of the American Statistical Association* **83**, 729–738.

Manski, C. F. (1993), 'Identification of endogenous social effects the reflection problem', *Review of Economic Studies* **60**, 531–542.

Manski, C. F. (1994), The selection problem, *in* C. Sims, ed., 'Advances in Econometrics, Sixth World Congress', Cambridge University Press.

Manski, C. F. (2000), 'Economic analysis of social interactions', *Journal of Economic Perspectives* **14**, 115–136.

Manski, C. F. (2003), *Partial identification of probability distributions*, Springer.

Manski, C. F. & Pepper, J. V. (2000), 'Monotone instrumental variables : With an application to the returns to schooling', *Econometrica* **68**, 997–1010.

Maskin, E. & Riley, J. (1984), 'Monopoly with incomplete information', *The RAND Journal of Economics* **15**, 171–196.

Mattner, L. (1992), 'Completeness of location families, translated moments, and uniqueness of charges', *Probability Theory and Related Fields* **92**, 137–149.

Mattner, L. (1993), 'Some incomplete but boundedly complete location families', *The Annals of Statistics* **21**, 2158–2162.

McMillen, D. P. (1992), 'Probit with spatial autocorrelation', *Journal of Regional Science* **32**, 335–348.

Milgrom, P. & Weber, R. (1982), 'A theory of auctions and competitive bidding', *Econometrica* **50**, 1089–1122.

Miravette, E. J. (2002), 'Estimating demand for local telephone service with asymmetric information and optional calling plans', *Review of Economic Studies* **69**, 943–971.

Mundlak, Y. (1961), 'Empirical production function free of management bias', *Journal of Farm Economics* **43**, 44–56.

Mussa, M. & Rosen, S. (1978), 'Monopoly and product quality', *Journal of Economic Theory* **18**, 301–317.

Myerson, R. B. (1981), 'Optimal auction design', *Mathematics of Operation Research* **6**, 58–73.

Nevo, A. (2002), 'Using weights to adjust for sample selection when auxiliary information is available', *Journal of Business and Economics Statistics* **21**, 43–52.

Newey, W. K. & Powell, J. L. (2003), 'Instrumental variable estimation of nonparametric models', *Econometrica* **71**, 1565–1578.

Newey, W. K., Powell, J. L. & Vella, F. (1999), 'Nonparametric estimation of triangular simultaneous equations models', *Econometrica* **67**, 565–603.

Paarsch, L. (1992), 'Deciding between the common and private value paradigms in empirical models of auctions', *Journal of Econometrics* **51**, 191–215.

Paarsch, L. & Shearer, B. (2000), 'Piece rates, fixed wages and incentive effects : Statistical evidence from payroll records', *International Economic Review* **41**, 59–92.

Park, T. & Brown, M. B. (1994), 'Models for categorical data with nonignorable nonresponse', *Journal of the American Statistical Association* **89**, 44–52.

Perrigne, I. (2002), Incentive regulatory contracts in public transportation : An empirical study. Working Paper, Pennsylvania State University.

Perrigne, I. & Vuong, Q. (2004), Nonparametric identification of incentive regulation models. Mimeo.

Pinkse, J. & Slade, M. E. (1998), 'Contracting in space : An application of spatial statistics to discrete- choice models', *Journal of Econometrics* **85**, 125–154.

Popper, K. (1968), *The logic of scientific discovery*, Harper Torchbooks, New York, revised ed.

Prendergast, C. (1999), 'The provision of incentives in firms', *Journal of Economic Literature* **37**, 7–63.

Puelz, R. & Snow, A. (1994), 'Evidence on adverse selection : Equilibrium signalling and cross-subsidization in the insurance market', *Journal of Political Economy* **102**, 236–257.

Ramalho, E. A. & Smith, R. J. (2007), Discrete choice nonresponse. CEMMAP working paper.

Rees, D. I., Zax, J. S. & Herries, J. (2003), 'Interdependence in worker productivity', *Journal of Applied Econometrics* **18**, 585–604.

Robin, J. M. & Smith, R. J. (2000), 'Test of rank', *Econometric Theory* **16**, 151–175.

Rothkopf, M. H. (1969), 'A model of rational competitive bidding', *Management Science* **15**, 362–373.

Rotnitzky, A., Robins, J. M. & Scharfstein, D. O. (1998), 'Semiparametric regression for repeated outcomes with nonignorable nonresponse', *Journal of the American Statistical Association* **93**, 1321–1339.

Rudin, W. (1987), *Real and Complex Analysis, third edition*, McGraw-Hill.

Rudin, W. (1991), *Functional Analysis, second edition*, McGraw-Hill.

Sacerdote, B. (1996), 'Peer effects with random assignment : Results for dartmouth roommates', *Quaterly Journal of Economics* **116**, 681–704.

Salanié, B. (2005*a*), *The Economics of Contracts : a Primer, 2nd edition*, MIT Press.

Salanié, F. (2005*b*), On screening. Working paper.

Sándor, Z. & András, P. (2004), 'Alternative sampling methods for estimating multivariate normal probabilities', *Journal of Econometrics* **120**, 207–234.

Scharfstein, D. O., Rotnitzky, A. & Robins, J. M. (1999), 'Adjusting for nonignorable drop-out using semiparametric nonresponse models', *Journal of the American Statistical Association* **94**, 1096–1120.

Schennach, S. M. (2004), 'Estimation of nonlinear models with measurement error', *Econometrica* **72**, 33–75.

Schennach, S. M. (2007), 'Instrumental variables estimation of nonlinear errors-in-variables models', *Econometrica* **75**, 201–239.

Schwartz, L. (1973), *Théorie des distributions, deuxième édition*, Hemann.

Shearer, B. (2004), 'Piece rates, fixed wages and incentives : Evidence from a field experiment', *Review of Economic Studies* **71**, 513–534.

Shum, M. & Hu, Y. (2008), Estimating first-price auction models with unknown number of bidders : a misclassification approach. Mimeo.

Smith, R. J. & Ramalho, E. A. (2007), Discrete choice nonresponse. Mimeo.

Soetevent, A. (2006), 'Empirics of the identification of social interactions : An evaluation of the approaches and their results', *Journal of Economic Surveys* **20**, 193–228.

Spence, M. (1973), 'Job market signaling', *Quarterly Journal of Economics* **87**, 355–374.

Tamer, E. (2003), 'Incomplete simultaneous discrete response models with multiple equilibria', *Review of Economic Studies* **70**, 147–165.

Tang, G., Little, R. J. A. & Raghunathan, T. E. (2003), 'Analysis of multivariate missing data with nonignorable nonresponse', *Biometrika* **90**, 747–764.

Troncin, T. (2005), Le redoublement : radiographie d'une décision Ã la recherche de sa légitimité. PhD Thesis, available at http ://tel.archives-ouvertes.fr/docs/00/14/05/31/PDF/05076.pdf.

van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics.

Vickrey, W. (1961), 'Counterspeculation, auctions and competitive sealed tenders', *Journal of Finance* **16**, 8–37.

Wilson, R. (1969), 'Competitive bidding with disparate information', *Management Science* **15**, 446–448.

Wilson, R. (1977), 'A bidding model of perfect competition', *Review of Economic Studies* **44**, 511–518.

Wilson, R. (1993), *Nonlinear pricing*, Oxford University Press.

Wolak, F. (1994), 'An econometric analysis of the asymmetric information, regulator-utility interaction', *Annales d'Economie et de Statistique* **34**, 13–69.

Wooldridge, J. (2002), 'Inverse probability weighted m-estimation for sample selection, attrition, and stratification', *Portuguese Economic Journal* **1**, 117–139.

Wooldridge, J. (2005), Inverse probability weighted estimation for general missing data problems. Michigan State University, Working paper.

Yatchew, A. (1998), 'Nonparametric regression techniques in economics', *Journal of Economic Literature* **36**, 669–721.

Zinde-Walsh, V. (2007), Errors-in-variables models : a generalized functions approach. CIREQ - Cahier 14-2007.