Université de Montréal

Essays in partial identification and applications to treatment effects and policy evaluation.

par
Ismael Yacoub Mourifié

Département de sciences économiques
Faculté des arts et sciences

Thèse présentée à la Faculté des arts et des sciences
en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)
en sciences économiques

May, 2013

Université de Montréal
Faculté des arts et des sciences

Cette thèse intitulée :
Essays in partial identification and applications to treatment effects and policy evaluation.

présentée par :
Ismael Y. Mourifié

a été évaluée par un jury composé des personnes suivantes :

Silvia Gonçalves,	président-rapporteur
Marc Henry,	directeur de recherche
William McCausland,	membre du jury
Victor Aguirregabaria,	examinateur externe
Peter Dietsch,	vice-doyen, représentant du doyen de la FAS

à mes parents, Ahmed Koffi Mourifié et Karidja Ouattara

Remerciements

Je tiens en premier à remercier Dieu et toutes les personnes qui, de près ou de loin, ont oeuvré à l'accomplissement de ce travail. Bien des pages ne suffiraient pour marquer ma gratitude à Dieu et à toutes ces personnes de bonne foi. Je me limite ici à quelques noms, mais mon coeur sera continuellement reconnaissant à tous.

Je voudrai avant tout remercier mon directeur de thèse Marc Henry pour sa présence continue, son amabilité, ses conseils de tous genres, pour m'avoir encouragé dans toutes les étapes de cette thèse et d'avoir surtout cru en moi. Il est sans doute la personne clé et la plus déterminante dans l'élaboration de cette thèse. Merci encore Marc.

Mes remerciements vont également à Yves Sprumont, Silvia Gonçalves et William Mccausland qui me suivent depuis la première présentation orale de cette thèse. J'aimerais tout aussi remercier Joris Pinkse, Quang Vuong et Isabelle Perrigne pour tout l'appui qu'ils m'ont apporté lors de mes deux dernières années de thèse.

Je tiens à remercier l'ensemble des étudiants de Doctorat en particulier Vincent, Nérée, Romuald, Ulrich, Idrissa, Désiré, Yourou, Juste, Guy, Mohamed, Brahim, William, Louis-Philippe, Hector et Vincent de Paul avec qui j'ai passé des moments inoubliables, particulièrement ces discussions dans les couloirs du département qui ont permis de raffermir mes connaissances en économie. Je ne pourrais omettre de citer Messan et Rachidi qui m'ont soutenu tout au long de ce parcours par leurs conseils et encouragements.

Ce parcours fut difficile d'un point de vue intellectuel, mais aussi moral, à des milliers de Kilomètres du beau soleil de la Côte d'ivoire ; je devais faire face à ce rude hiver de Montréal. Par la grâce de Dieu, j'ai trouvé sur place une grande "famille" qui m'a sans cesse soutenu dans toutes les épreuves et je ne pourrai terminer sans citer quelques noms.

Je remercie tout particulièrement mes très chers amis Chef Tolo, Bouba, Karim, Mohamed, Larissa, Fatheem à Montréal, Maï et Ibrahim à Dakar avec qui j'ai tout partagé et qui ont été d'un soutien incommensurable. Je tiens à remercier les familles Ano, Binaté, Fofana, Touré, Tamakholé, Traoré et particulièrement Richimond et Massa ainsi que leurs adorables enfants Latifa, Aemane, Imelle et Khyane pour toute la chaleur familiale qu'ils m'ont apporté. Je ne finirai jamais de les remercier.

Je remercie le Département de sciences économiques de l'Université de Montréal, le Centre Interuniversitaire de Recherche en économie Quantitative (CIREQ) et la Faculté des études supérieures et postdoctorales pour le soutien financier. Mes remerciements vont également à tout le corps professoral, au personnel administratif et particulièrement à Mélanie qui m'a toujours gratifié de sa bonne humeur et de son adorable sourire.

Je termine cette section en portant une attention particulière à ma famille : mes parents Ahmed et Karidja pour tout leur soutien et surtout cette éducation qu'ils m'ont inculqué, mes frères Mohammed et Issa qui m'ont toujours servi d'exemple, ma merveilleuse petite soeur Hadja Fanta qui m'a toujours gratifié de ses sages conseils et enfin mon adorable épouse Hajara qui a toujours été si indulgente avec moi et m'a toujours soutenu malgré tout le temps passé sur mon ordinateur à travailler sur cette thèse. Je ne peux que remercier Dieu de m'avoir gratifié d'une telle famille.

Résumé

Dans cette thèse, je me suis interessé à l'identification partielle des effets de traitements dans différents modèles de choix discrets avec traitements endogènes. Les modèles d'effets de traitement ont pour but de mesurer l'impact de certaines interventions sur certaines variables d'intérêt. Le type de traitement et la variable d'intérêt peuvent être défini de manière générale afin de pouvoir être appliqué à plusieurs différents contextes. Il y a plusieurs exemples de traitement en économie du travail, de la santé, de l'éducation, ou en organisation industrielle telle que les programmes de formation à l'emploi, les techniques médicales, l'investissement en recherche et développement, ou l'appartenance à un syndicat. La décision d'être traité ou pas n'est généralement pas aléatoire mais est basée sur des choix et des préférences individuelles. Dans un tel contexte, mesurer l'effet du traitement devient problématique car il faut tenir compte du biais de sélection.

Plusieurs versions paramétriques de ces modèles ont été largement étudiées dans la littérature, cependant dans les modèles à variation discrète, la paramétrisation est une source importante d'identification. Dans un tel contexte, il est donc difficile de savoir si les résultats empiriques obtenus sont guidés par les données ou par la paramétrisation imposée au modèle. Etant donné, que les formes paramétriques proposées pour ces types de modèles n'ont généralement pas de fondement économique, je propose dans cette thèse de regarder la version nonparamétrique de ces modèles. Ceci permettra donc de proposer des politiques économiques plus robustes.

La principale difficulté dans l'identification nonparamétrique de fonctions structurelles, est le fait que la structure suggérée ne permet pas d'identifier un unique processus générateur des données et ceci peut être du soit à la présence d'équilibres multiples ou soit à des contraintes sur les observables. Dans de telles situations, les méthodes d'identifications traditionnelles deviennent inapplicable d'où le récent développement de la littérature sur l'identification dans les modèles incomplets. Cette littérature porte une attention particuliere à l'identification de l'ensemble des fonctions structurelles d'intérêt qui sont compatibles avec la vraie distribution des données, cet ensemble est appelé : l'*ensemble identifié* .

Par conséquent, dans le premier chapitre de la thèse, je caractérise l'*ensemble identifié* pour les effets de traitements dans le modèle triangulaire binaire.

Dans le second chapitre, je considère le modèle de Roy discret. Je caractérise l'*ensemble identifié* pour les effets de traitements dans un modèle de choix de secteur lorsque la variable d'intérêt est discrète. Les hypothèses de sélection du secteur comprennent le choix de sélection simple, étendu et généralisé de Roy.

Dans le dernier chapitre, je considère un modèle à variable dépendante binaire avec plusieurs dimensions d'hétérogéneité, tels que les jeux d'entrées ou de participation. je caractérise l'*ensemble identifié* pour les fonctions de profits des firmes dans un jeux avec deux firmes et à information complète.

Dans tout les chapitres, l'*ensemble identifié* des fonctions d'intérêt sont écrites sous formes de bornes et assez simple pour être estimées à partir des méthodes d'inférence existantes.
Mots-clés : Effet de traitement, Evaluation de politique, Endogeneité, Modèle de sélection, Modèle incomplet, Ensemble identifié , Borne aigüe, Modèle nonparamétrique.

Abstract

In this thesis, I have been interested in the nonparametric (partial) identification of structural potential outcome functions and Average Treatment Effect (ATE) in various discrete models with endogenous selection and treatment. This topic of treatment effect concerns measuring the impact of an intervention on an outcome of interest. The type of treatments and outcomes may be broadly defined in order to be applied in many different contexts. There are many examples of treatment in economics (Labor, health, education, trade, industrial organization) such that Job training programs, surgical procedures, higher education level, research and development investment, being a member of a trade union etc. The decision to be treated or not, is usually not random but is based on individual choices or preferences. In such a context, determining the impact of the treatment becomes an important issue since we have to take into account the selectivity bias.

The parametric version of such models has been widely studied in the literature, however in models with discrete variation, the parametrization is a strong source of identification. Then, we don't know if the empirical results we obtain, are driven by the data or by the parametrization imposed on the model. I propose to look at a fully nonparametric version of those models, in order, to have more robust policy recommendations.

The central challenge in this nonparametric structural identification is that the hypothesized structure fails to identify a single generating process for the data, either because of multiple equilibria or data observability constraints. In such cases, many traditional identification techniques become inapplicable and a framework for identification in incomplete models is developing, with an initial focus on identification of the set of structural functions of interest compatible with the true data distribution (hereafter *identified set*).

Therefore, in the first chapter, I provide a full characterization of the

identified set for the ATE in a binary triangular system.

In the second chapter, I consider a model with sector specific unobserved heterogeneity. I provide the full characterization of the *identified set* for the structural potential outcome functions of an instrumental variables model of sectoral choice with discrete outcomes. Assumptions on selection include the simple, extended and generalized Roy models.

In the last chapter, I consider a binary model with several unobserved heterogeneity dimensions, such as entry and participation games. I provide the full characterization of the *identified set* for the payoffs in 2×2 games with perfect information, including duopoly entry and coordination games.

In all chapters, the *identified set* of the functions of interest are nonparametric intersection bounds and are simple enough to lend themselves to existing inference methods.

Keywords : Potential outcome, Average Treatment Effect, Policy evaluation, Endogeneity, Selection model, Incomplete model, Partial identification, Identified set, Sharp bound, Nonparametric model.

Table des matières

Table des figures

Liste des tableaux

Chapitre 1

Sharp bounds on treatment effects in binary triangular system.

1.1 Introduction

This paper considers the evaluation of the average treatment effect (ATE) of a binary endogenous regressor on a binary outcome when I impose a threshold crossing model on both the endogenous regressor and the outcome. This model encompasses many important applications in different areas of economics including labor economics as in Battistin and Rettore (2002), education, as in Canton and Bloom (2004), Beffy, Fougère and Maurel (2010), health economics as in Bhattacharya, Shaikh, and Vytlacil (2008), Carpenter and Dobkin (2009), political economy as in Lee (2008) among many others.

The joint threshold crossing model was recently investigated by Shaikh and Vytlacil (2011), but their proposed bounds are sharp only under a criti-

I am grateful to Louis-Philippe Béland, Ivan Canay, Marc Henry, Sung Jae Jun, Joris Pinkse, Christoph Rothe, Alexander Torgovitsky, Bernard Salanié for helpful discussions and comments from participants to Second CIREQ-CeMMAP Conference on Incomplete models and from seminar audiences in Cambridge, Chicago, Columbia, Exeter Business School, PennState, University of Toronto and Warwick economics departments. Parts of this chapter were written while I was visiting Penn State, I thank my hosts for their hospitality and support.

cal restriction imposed on the support of the covariates and the instruments. The support condition required is very strong and fails when we have deterministic treatment or partially deterministic treatment. Even without deterministic treatment, SV's support condition is likely to fail for a wide range of models. Basically, the SV critical support condition is more likely to hold in the rare case when there is no common covariates between the outcome and the treatment equations. Indeed, SV takes advantage of the threshold crossing condition imposed on the endogenous regressor, to refine known bounds on the ATE in the model with unrestricted endogenous regressor. However, when the support condition fails they do not take full advantage of the threshold crossing condition imposed on the endogenous regressor. In some cases, their bounds do not have any empirical content beyond the model with unrestricted endogenous regressor. I show throughout this paper, how it is possible to fully exploit the second threshold crossing restriction imposed on the endogenous regressor without imposing any support restrictions. More specifically, I show under the joint threshold crossing model, that the sign of the marginal average effect may be identified, and observable bounds of the marginal average effect can be derived. I take advantage of that to construct sharp bounds on the ATE.

Therefore, this paper complements SV's work by providing a methodology which allows to construct sharp bounds on the ATE by efficiently using variation on covariates and which does not need to impose any support restrictions. Our methodology requires only mild regularity conditions on the distribution of unobservable variables and an usual exogeneity assumption between the covariates (except the binary endogenous regressor) and the unobservable variables. The proof of sharpness of our proposed bounds is based on copula theory and a characterization theorem proposed by Chiburis (2010). A similar objective is pursued by Chiburis (2010), but his approach relies on an algorithm to determine existence of a copula, which is computationally infeasible in many cases of interest. I provide a methodology to reduce the computational burden of the Chiburis (2010) technique. However, the method proposed in this paper remains much simpler to apply. In addition, this method can be easily extended to a triangular system with nonbinary-valued discrete endogenous regressors and continuous outcome.

This joint threshold crossing model is a special case of nonparametric triangular systems. Imbens and Newey (2009) and Kasy (2011) develop nonparametric identifications results in triangular systems by using the "control function" approach. Their results hold when dependent and endogenous variables are continuous, but do not extend to the present context. This model is also a particular case of Chesher (2005), but his analysis requires a strong rank condition and an additional assumption on the joint distribution of the unobservable variables. His rank condition rules out the case where the endogenous regressor is binary. Jun, Pinkse and Xu (2010, JPX) relax this rank condition but still maintain an additional assumption on the joint distribution of the unobservable variables. Moreover, identification of the ATE in this model was previously considered by Vytlacil and Yildiz (2007, VY). They showed that under a strong support condition it is possible to point identify the ATE. Jun, Pinkse and Xu (2011) weaken the VY support condition by using the identification method proposed in JPX, but still maintain an additional assumption on the joint distribution of the unobservable variables. Here, I do not impose such restrictions, which might be very restrictive.

The rest of the paper is organized as follows. The next section revisits single threshold crossing models, when no structural form is assumed for the binary endogenous regressor. The following section considers joint threshold crossing models, explains why SV's bounds fails to be sharp without their support condition and proposes a methodology to sharpen their bounds in the case, where the latter fails to hold. The third section presents a numerical illustration and the fourth section presents an application of our methodology to the assessment of the effect of migration decisions on the standard of living of the family left behind in Cameroon. The last section concludes and proofs are collected in the appendix.

The last section concludes and proofs are collected in the appendix.

1.2 Threshold crossing model with unrestricted binary treatment

I adopt in this section the framework of the potential outcomes model $Y = Y_1 D + Y_0(1 - D)$, where Y is an observed outcome, D denotes the observed binary endogenous regressor and Y_1, Y_0 are unobserved potential outcomes. Potential outcomes are as follows :

$$Y_d = 1\{F(d, x, u) > 0\}, d = 1, 0 \tag{1.1}$$

where $1\{.\}$ denotes the indicator function and F is an unknown function of a vector of exogenous regressors X, and unobserved random variable u. The formal assumptions I use in this section may be expressed as follows :

Assumption 1. *The functions $F(d, x, u)$, d=1,0, both have weakly separable errors. As shown in Vytlacil (2002) and Vytlacil and Yildiz (2007), potential outcomes can then be written $Y_d = 1\{\nu(d, x) > u\}$ without loss of generality.*

Assumption 2. *(X, Z) and u are statistically independent, where Z is an available instrument.*

Assumption 3. *The distribution of u has positive density w.r.t Lebesgue measure on $\mathbb{R}$.*

According to equation (1.1) we have $\mathbb{E}[Y_d|X, Z] = \mathbb{E}[Y_d|X]$. It follows from assumptions 1 and 3 that we may impose, without loss of generality, the normalization that u is uniformly distributed on [0,1] ($u \sim$ Uniform [0,1]). This normalization is very convenient, since it implies $\mathbb{E}[Y_d \mid X = x] = P(Y_d = 1 \mid X = x)$ and bounds on treatment effects parameters can be derived from bounds on the structural parameters $\nu(1, x)$ and $\nu(0, x)$. Then we may define the average structural function (ASF) and the average treatment effect (ATE), respectively, as follows :

$$\begin{aligned} \nu(d, x) &= P(Y_d = 1 \mid X = x) \\ \Delta\nu(x) &= P(Y_1 = 1 \mid X = x) - P(Y_0 = 1 \mid X = x). \end{aligned}$$

In all this section , I shall use the notation $P(i, j|x, z) = P(Y = i, D = j|X = x, Z = z)$. Now, I will provide a proposition which recalls known result on sharp bounds on the ASF when the binary endogenous regressor is unrestricted.

Proposition 1 (Chiburis (2010))**.** *Suppose Y, D determined by model (1.1). Let $Dom(X)$ and $Dom(Z)$ denote the respective domains of the random variables X and Z. Under assumptions 1, 2 and 3, the following bounds are sharp for the average structural function (ASF) in the model (1.1) :*
For each x in $Dom(X)$,

— *if* $\nu(0, x) \leq \nu(1, x)$

$$\begin{aligned} &\sup_z\{P(1, 0|x, z)\} \leq \nu(0, x) \leq \inf_z\{P(Y = 1 \mid x, z)\} \\ &\sup_z\{P(Y = 1|x, z)\} \leq \nu(1, x) \leq \inf_z\{P(1, 1 \mid x, z) + P(D = 0 \mid x, z)\}, \end{aligned}$$

— *if* $\nu(0,x) \geq \nu(1,x)$

$$\sup_z\{P(Y=1|x,z)\} \leq \nu(0,x) \leq \inf_z\{P(1,0|x,z)+P(D=1|x,z)\}$$
$$\sup_z\{P(1,1|x,z)\} \leq \nu(1,x) \leq \inf_z\{P(Y=1|x,z)\},$$

where the supremum and the infimum are taken over the domain of the random variable Z ($Dom(Z)$).

One proof of this result is given by Chiburis (2010), but I give an alternative proof which considers this model as a particular case of discrete outcome models with multiple equilibria. (See Appendix .1). The proof I propose is convenient since it allows to derive sharp bounds for model with specific sector heterogeneity i.e $Y_d = 1\{F(d,x,u_d) > 0\}$, as it has been done in subsequent paper, see Henry and Mourifié (2012). It, also, allows to derive sharp bounds in a case of non-binary discrete endogenous regressor. In a case where the bounds cross for one of the two cases the sign of the ATE is identified. For instance, $\sup_z\{P(1,0|x,z)\} > \inf_z\{P(Y=1 \mid x,z)\}$ but $\sup_z\{P(Y=1|x,z)\} \leq \inf_z\{P(1,0|x,z)+P(D=1|x,z)\}$ and $\sup_z\{P(1,1|x,z)\} \leq \inf_z\{P(Y=1|x,z)\}$ then $\Delta\nu(x) < 0$. If the bounds cross in both cases, the joint assumption of weakly separable errors and the presence of a valid instrument Z is rejected. When no instrument is available, the previous sharp bounds of proposition 1 become :

— if $\nu(0,x) \leq \nu(1,x)$

$$P(1,0|x) \leq \nu(0,x) \leq P(Y=1|x)$$
$$P(Y=1 \mid x) \leq \nu(1,x) \leq P(1,1|x)+P(D=0|x),$$

— if $\nu(0,x) \geq \nu(1,x)$

$$P(Y=1|x) \leq \nu(0,x) \leq P(1,0|x)+P(D=1|x)$$
$$P(1,1|x) \leq \nu(1,x) \leq P(Y=1|x),$$

which are the same as Manski and Pepper's (2000) bounds on the ASF. In such a context, it is immediately apparent that the sign of the ATE is not identified and the weakly separable errors assumption cannot be falsified.

1.3 Joint threshold crossing model

In this section, I put more structure on the previous model by assuming that the binary endogenous regressor has the following structure $D = 1\{G(x,z,v) > 0\}$, where G is an unknown function weakly separable in v. Then I propose the following model :

$$\begin{aligned} Y_d &= 1\{F(d,x,u) > 0\}, d = 0,1 \\ D &= 1\{G(x,z,v) > 0\}. \end{aligned} \tag{1.2}$$

This model can be summarized without loss of generality (see Vytlacil (2002) and Vytlacil and Yildiz (2007)) as follows :

$$\begin{aligned} Y_d &= 1\{\nu(d,x) > u\}, d = 0,1 \\ D &= 1\{p(x,z) > v\}, \end{aligned} \tag{1.3}$$

where $Y = Y_1 D + Y_0(1-D)$ denotes the observed binary outcome of interest, D denotes the observed binary endogenous regressor, (X,Z) is a vector of exogenous regressors, (u,v) are unobserved random variables. The formal assumptions I use in this section may be expressed as follows :

Assumption 4. (X,Z) *and* (u,v) *are statistically independent.*

Assumption 5. *The distribution of* (u,v) *has positive density w.r.t Lebesgue measure on* $\mathbb{R}^2$.

It follows from assumptions 4 and 5, that we may impose, without loss of generality, the normalization that $u,v \sim U[0,1]$, $\nu(d,x) = \mathbb{E}[Y_d|X=x] = P(Y_d = 1 \mid X = x)$, $P(X,Z) = P(D = 1 \mid X,Z)$ and $\nu(d,x) = \mathbb{E}[Y_d|X = x, P(X,Z) = p]$. Then, the ASF is $\nu(d,x) = P(Y_d = 1 \mid X = x)$ and the ATE is $\Delta\nu(x) = P(Y_1 = 1 \mid X = x) - P(Y_0 = 1 \mid X = x)$. In all this section, I shall use the notation $P(i,j|x,p) = P(Y = i, D = j|X = x, P(X,Z) = p)$. Similar analysis has been carried out, previously, by SV. Their work provides bounds on the ASF, which are based on observable quantities, and which exploit covariates variation. SV used the joint threshold crossing equations determining Y and D and additional assumptions to determine the sign of $[\nu(1,x') - \nu(0,x)]$ from the distribution of observed data, and then take advantage of this information to construct bounds on ASF which exploit variation on covariates. Denote by $Supp(P(X,Z) \mid X)$ the support of $P(X,Z)$

conditional on X and write $P(X, Z) = P$ in the rest of the paper. Before going into details, let's provide a simple intuition of the main idea of this paper. We have

$$\nu(0, x) = P(u \leq \nu(0, x), v \geq p(x, z)) + P(u \leq \nu(0, x), v \leq p(x, z)),$$

where $P(u \leq \nu(0, x), v \geq p(x, z)) = P(1, 0|x, p)$, but the second term $P(u \leq \nu(0, x), v \leq p(x, z)) = P(Y_0 = 1, D = 1|X = x)$ is the unobserved counterfactual. SV proposed to bound this counterfactual by exploiting variation on covariates. Indeed, SV's idea suggests that, we may bound the unobserved counterfactual for untreated individuals ($D = 0$) with characteristic x by using information on treated individuals ($D = 1$) with different characteristics x' whenever they have exactly the same probability to be treated. In fact, if we have a treated individual with characteristic x' belonging to the following set $\Delta(x) = \{x' : \nu(0, x) \leq \nu(1, x')\} \cap \{x' : \exists z' \in Dom(Z),\ p(x, z) = p(x', z')\}$, the proposed bounds of SV for the unobserved counterfactual can be summarized as follows

$$P(u \leq \nu(0, x), v \leq p(x, z)) \leq \begin{cases} P(u \leq \nu(1, x'), v \leq p(x', z')) & \text{if } x' \in \Delta(x) \\ p(x, z) & \text{if } \Delta(x) = \emptyset \end{cases}$$

where $P(u \leq \nu(1, x'), v \leq p(x', z')) = P(1, 1|x', p')$. Their idea is quite interesting, but not sufficient to provide the sharp bounds. Our argument relies on the fact that, under the threshold crossing model assumption imposes on the treatment (D), we may bound the unobserved counterfactual $\mathbb{P}(Y_0 = 1, D = 1|x, z)$ by using information on treated individuals with different characteristics x' even if they have different probabilities to be treated. In fact, if we have a treated individual with characteristic x' belonging to the following subset $\tilde{\Delta}(x) = \{x' : \nu(0, x) \leq \nu(1, x')\} \cap \{x' : \exists z' \in Dom(Z),\ p(x, z) \leq p(x', z')\}$, the unobserved counterfactual may be bounded as follows

$$P(u \leq \nu(0, x), v \leq p(x, z)) \leq \begin{cases} P(1, 1|x', p') & \text{if } x' \in \tilde{\Delta}(x) \\ p(x, z) & \text{if } \tilde{\Delta}(x) = \emptyset. \end{cases}$$

Then, it is immediately apparent that, the subset $\Delta(x)$ is necessary but not sufficient to construct the sharp bound for $\nu(0, x)$. Instead of using $\Delta(x)$, I propose to visit $\tilde{\Delta}(x)$. Since $\Delta(x) \subseteq \tilde{\Delta}(x)$, it is easy to see that we may get an improvement over SV's bounds by using $\tilde{\Delta}(x)$ instead of $\Delta(x)$, especially when $\Delta(x)$ is empty or $\Delta(x)$=$\{x\}$. When $Supp(P|X = x) = Supp(P|X = x')$

we have $\tilde{\Delta}(x) = \Delta(x)$, this fact explains why the SV bounds would be sharp when $Supp(P|X = x) = Supp(P|X = x')$. This first idea is not sufficient to fully characterize all the empirical content of the model. In this first idea we show that, to bound the ASF for an untreated individual ($D = 0$) with characteristic x, we may use information on a treated individual ($D = 1$) with different characteristic x'. Our second idea relies on the fact that, under the threshold crossing model assumption imposes on the treatment (D), to bound the ASF for an untreated individual ($D = 0$) with characteristic x, we may, also, use information on other untreated individual ($D = 0$) with different characteristic x'. In fact, if we have two untreated individuals with characteristics x and x' such that $\nu(0,x) < \nu(0,x')$, we have the following

$$\begin{aligned} P(\nu(0,x) \leq u \leq \nu(0,x')) &= P(\nu(0,x) \leq u \leq \nu(0,x'), v \geq p(x,z)) \\ &\qquad + P(\nu(0,x) \leq u \leq \nu(0,x'), v \leq p(x,z)) \\ &\geq P(\nu(0,x) \leq u \leq \nu(0,x'), v \geq p(x,z)) \\ &\geq P(u \leq \nu(0,x'), v \geq p(x,z)) \\ &\qquad -P(u \leq \nu(0,x), v \geq p(x,z)), \end{aligned}$$

then for all $p(x,z) \leq p(x',z')$ we have

$$\begin{aligned} \nu(0,x') - \nu(0,x) &\geq P(u \leq \nu(0,x'), v \geq p(x',z')) \\ &\qquad -P(u \leq \nu(0,x), v \geq p(x,z)) \\ &\geq P(1,0|x',p') - P(1,0|x,p). \end{aligned}$$

Thus for all x' belonging to $\{x' : \nu(0,x) \leq \nu(0,x')\} \cap \{x' : \exists z' \in Dom(Z), p(x,z) \leq p(x',z')\}$, we shall have

$$\nu(0,x') - \nu(0,x) \geq \max[P(1,0|x',p') - P(1,0|x,p), 0]. \qquad (1.4)$$

We can easily see that SV's bounds do not recover this feature of the model. For instance, in a case where $Supp(P|X = x) \cap Supp(P|X = x') = \emptyset$, SV's bounds become the bounds derived in proposition 1, and then, if $\nu(1,x') > \nu(0,x') > \nu(0,x) > \nu(1,x)$ we have $\nu(0,x') - \nu(0,x) \geq P(1,0|x',p') - P(1,0|x,p) - P(D = 1|x,p)$ which is wider than the bound proposed in (1.4). To the best of our knowledge, this paper is the first to construct bounds on the ATE which respect such feature of the model. As we can remark throughout the above discussion the signs of the following quantities $[\nu(1,x') - \nu(0,x)]$, $[\nu(d,x') - \nu(d,x)]$, for $d = 1,0$ are very important in our analysis.

Model (1.3) has the additional nice feature that it allows identification of the sign of the following marginal average effect

$$\mathbb{E}[Y_1|X=x'] - \mathbb{E}[Y_0|X=x] = [\nu(1,x') - \nu(0,x)] \tag{1.5}$$

and

$$\mathbb{E}[Y_d|X=x'] - \mathbb{E}[Y_d|X=x] = [\nu(d,x') - \nu(d,x)],\ \text{d=1,0} \tag{1.6}$$

under very mild assumptions. SV showed that $[\nu(1,x') - \nu(0,x)]$ share the same sign as the following observable function $h(x,x',p,p') = (P(1,1|x',p) - P(1,1|x',p')) - (P(1,0|x,p') - P(1,0|x,p))$, when P is not degenerate, and both p and p' belong to $Supp(P \mid X=x) \cap Supp(P \mid X=x')$ such that $p' < p$. As we can remark, the SV idea cannot identified the sign of $[\nu(1,x') - \nu(0,x)]$ when $Supp(P \mid X=x) \cap Supp(P \mid X=x')$ is empty or a singleton. However, the sign of $[\nu(1,x') - \nu(0,x)]$ would still be identified. Indeed, if there are $p'_1 < p'_2 \in Supp(P \mid X=x')$ and $p_1 < p_2 \in Supp(P \mid X=x)$ such that $[p'_1, p'_2] \subseteq [p_1, p_2]$ the sign of $[\nu(1,x') - \nu(0,x)]$ would be identified using the following observable function $[P(1,1|x',p'_2) - P(1,1|x',p'_1)) - (P(1,0|x,p_1) - P(1,0|x,p_2)]$. In the Lemma 2 in Appendix .1, I show how the sign of $[\nu(1,x') - \nu(0,x)]$ may be identified under weaker assumptions. Moreover, I show that the sign of $[\nu(d,x') - \nu(d,x)]$ $d = 1,0$ may also be identified under very mild assumptions.

This interesting feature of the model, will help to reduce considerably the computation burden of our proposed bounds, as we will see later.

Now, let recall the SV bounds. SV take advantage of the knowledge of the sign of $[\nu(1,x') - \nu(0,x)]$, to construct an upper bound for $\nu(0,x)$, when $Supp(P \mid X=x) \cap Supp(P \mid X=x')$ is not empty. Therefore, if $p \in Supp(P \mid X=x) \cap Supp(P \mid X=x')$, and $h(x,x',p,p') \geq 0$ we have

$$\begin{aligned} \nu(0,x) &= P(u \leq \nu(0,x), v \geq p) + P(u \leq \nu(0,x), v \leq p) \\ &\leq P(u \leq \nu(0,x), v \geq p) + P(u \leq \nu(1,x'), v \leq p), \end{aligned}$$

for all $x' \in \mathbf{X}(x) = \{x' : h(x,x',p,p') \geq 0\}$. Hence,

$$\nu(0,x) \leq \inf_p \{P(1,1|x,p) + p \inf_{x' \in \mathbf{X}(x)} P(1|1,x',p')\}.$$

Similarly, we can derive the lower bound for $\nu(0,x)$, and also the lower and upper bounds for $\nu(1,x)$. Hereafter, I adopt the convention that the supremum over empty set is zero and the infimum over the empty set is one.

Remark 1. *SV take advantage of the knowledge of the sign of* $[\nu(1,x') - \nu(0,x)]$*, to construct an upper bound for* $\nu(0,x)$ *only whenever there exist* $p' < p$ *belong to* $Supp(P \mid X = x) \cap Supp(P \mid X = x')$*. Lemma 2 showed that we may take advantage of the knowledge of the sign of* $[\nu(1,x') - \nu(0,x)]$*, to construct an upper bound for* $\nu(0,x)$ *even if* $Supp(P \mid X = x) \cap Supp(P \mid X = x')$ *is empty or a singleton.*

SV showed that these bounds are sharp under further assumptions, which can be expressed as follows :

Assumption 6. *The functions* $\nu(0,.)$*,* $\nu(1,.)$ *and* $p(.)$ *are continuous.*

Assumption 7. *The support of the distribution of* (X,Z)*,* $Supp(X,Z)$*, is compact.*

Assumption 8. *(critical support condition)* $Supp(X,P) = Supp(X) \times Supp(P)$*.*

Their result remains restrictive due to the strong restriction imposed by the "critical support condition".

1.3.1 Plausibility of the "critical support condition"

Assumption 8 implies that $Supp(P \mid X = x) = Supp(P \mid X = x')$ for all $(x,x') \in Dom(X) \times Dom(X)$, in others terms for all $(x,x') \in Dom(X) \times Dom(X)$ and $z \in Dom(Z)$, there exist $z' \in Z$ such that $p(x,z) = p(x',z')$. This type of "perfect matching restriction" may be difficult to achieve in many applications.

partially deterministic treatment

In the treatment effect setting, there are many applications where we have additional information about the treatment assignment : it's known that the treatment assignment mechanism depends (at least in part) on the value of observed variables. Then, the treatment may be deterministic for some characteristics (x',z') such that the treated probability is degenerate in some points i.e ($p(x',z') = 1$ or $p(x',z') = 0$). One well suited example for this type of deterministic treatment is the financial aid selection rule. Van der Klaauw (2002) studied financial aid selection rule and showed that USA's colleges use generally SRT (Standard Reasoning Test) and GPA (Grade Point Average) to rank students into a small number of categories and then decide on a particular selection rule. Let

$$G_l = \begin{cases} 1 \text{ if } 0 \leq GPA \leq C_1, \\ 2 \text{ if } C_1 \leq GPA \leq C_2, \\ . \\ . \\ . \\ L \text{ if } C_{L-1} \leq GPA \leq C_L, \end{cases}$$

denote the financial aid group. Denote by Z the vector of others characteristics such as special awards, recommendation letters or/and extracurricular activities. The selection rule is usually the following : the higher ranked category is directly selected i.e $(p(C_{L-1} \leq GPA \leq C_L, z) = 1$ for all $z \in Dom(Z)$, the lower one is directly excluded i.e $(p(0 \leq GPA \leq C_1, z) = 0$ for all $z \in Dom(Z))$ and for the others, we look at others characteristics such as special awards, recommendation letters or/and extracurricular activities. i.e $(p(C_1 \leq GPA \leq C_{L-1}, z) \in (0,1))$. In this generic example, there does not exist $(z, z') \in Dom(Z) \times Dom(Z)$ such that $P(G_1, z) = P(G_l, z')$ for $1 < l \leq L$, then SV's "critical support condition" fails. We have showed in the latter example, that SV's "critical support condition" fails when we have deterministic treatment or partially deterministic treatment. Furthermore, our numerical illustration shows that even without deterministic treatment SV's "critical support condition" is likely to fail.

Semiparametric bounds

Depending on the economic model you have, it is possible to assume some functional forms or parametric forms for either $\nu(d, x)$ or $p(x, z)$ or both of them. For instance, we may have $\nu(d, x) = F(x'\beta + d\alpha)$ or $p(x, z) = \Phi(x'\gamma + z'\eta)$. Manski (1988) discussed the identification of the single index model where F or Φ is unknown and showed that this model fits a wide range of model. Here, we will assume a linear index model for D, with Φ an unknown, strictly increasing, function. Then, we have the following semiparametric model :

$$\begin{aligned} Y &= 1\{\nu(D, x) > u\} \\ D &= 1\{\Phi(x\gamma + z\eta) > v\}, \end{aligned} \tag{1.7}$$

with $Dom(X) \subset \mathbb{R}$, $Dom(Z) = \{0, 1\}$ and $\Phi(.) \in [0, 1]$. We can easily see that the SV support condition fails to hold for this specification which fits a wide range of models. Indeed, since $\Phi(.)$ is strictly increasing, $Supp(P \mid X = 0) = \{\Phi(0), \Phi(\eta)\} \neq \{\Phi(\gamma), \Phi(\gamma + \eta)\} = Supp(P \mid X = 1)$ for all $(\gamma, \eta) \neq (0, 0)$. Moreover, since in this case $Supp(P \mid X = x) \cap Supp(P \mid X = x')$ is empty, SV's bounds fail to improve bounds that we found in proposition 1. It means that in this context their bounds do not take advantage of the linear index structure imposed on D. In a case where, $Dom(Z)$ is discrete non-binary the SV critical support condition still fails to hold. However, when the instrument Z has a continuous and a large support such that $\lim_{z \to +\infty} \Phi(x\gamma + z\eta) = 1$, and $\lim_{z \to -\infty} \Phi(x\gamma + z\eta) = 0$ for all $x \in Dom(X)$ and $\eta > 0$, the SV critical support condition would hold. But, in such a context, the partial identification approach entertains by SV and this present paper is less relevant since we have identification at infinity of our object interest $\nu(d, x)$ for all $x \in Dom(X)$.

Basically, the SV critical support condition is more likely to hold only when $p(x, z)$ does not depend on x, which is only true in the rare case of a complete dichotomy between variables in the outcome equation and variables in the treatment equation.

1.3.2 Failure of sharpness of SV's bounds without the critical support condition

SV's bounds take advantage of the observability of the sign of $[\nu(1, x') - \nu(0, x)]$ when $Supp(P \mid X = x) \cap Supp(P \mid X = x')$ is not empty and not reduced to a singleton. Whenever $Supp(P \mid X = x) \cap Supp(P \mid X = x')$ is empty or reduced to a singleton, SV's bounds do not take advantage of the additional weak separability restriction that we impose on the equation determining D. I will now show, how it is possible to sharpen bounds on the ASF and the ATE, without imposing the "critical support condition"(assumption 8). Before formalizing our idea, I will define some subsets summarized in Table 1.1.

1.3.3 Sharpening the bounds

I will show in a first step, that it is still possible to improve proposition 1's bounds when $p(x, z) \notin Supp(P \mid X = x) \cap Supp(P \mid X = x')$,

TABLE 1.1 – Collection of sets

$$\mathbf{P}^{+}(x',p) = \{p(x',z') = p' \in Supp(P|X = x') : p \leq p'\}$$

$$\mathbf{P}^{-}(x',p) = \{p(x',z') = p' \in Supp(P|X = x') : p \geq p'\}$$

$$\mathbf{\Omega}^{+}_{\mathbf{d_1 d_2}}(x) = \{x' : \nu(d_1,x) \leq \nu(d_2,x')\}$$

$$\mathbf{\Omega}^{-}_{\mathbf{d_1 d_2}}(x) = \{x' : \nu(d_1,x) \geq \nu(d_2,x')\}$$

by taking advantage simultaneously of the sign of $[\nu(1,x') - \nu(0,x)]$ and the range $\mathbf{P}^{+}(x',p)$. In a second step I will show how it is possible to narrow SV's bounds by also using the sign of $[\nu(0,x') - \nu(0,x)]$ and the range $\mathbf{P}^{+}(x',p)$.

First step :
When $p(x,z) \notin Supp(P \mid X = x) \cap Supp(P \mid X = x')$, we cannot identify $P(u \leq \nu(1,x'), v \leq p(x,z))$ from the data. Then SV proposed in this case to bound $P(u \leq \nu(1,x'), v \leq p(x,z))$ from above by $P(v \leq p(x,z)) = p(x,z)$. But whenever it is possible to find $p(x',z')$ in $\mathbf{P}^{+}(x',p)$, I propose to bound $P(u \leq \nu(1,x'), v \leq p(x,z))$ from above by $P(u \leq \nu(1,x'), v \leq p(x',z')) = P(1,1|x',p')$, which may be lower than $P(v \leq p(x,z)) = p(x,z)$ in some cases. The upper bound for $\nu(0,x)$ that we can build with this strategy is lower than the one proposed by SV. Indeed, for all $x' \in \mathbf{\Omega}^{+}_{\mathbf{01}}(x)$ and $p(x',z') \in \mathbf{P}^{+}(x',p)$ we have :

$$\begin{aligned}
\nu(0,x) &= P(u \leq \nu(0,x), v \geq p(x,z)) + P(u \leq \nu(0,x), v \leq p(x,z)) \\
&\leq P(u \leq \nu(0,x), v \geq p(x,z)) + P(u \leq \nu(1,x'), v \leq p(x,z)) \\
&\leq P(u \leq \nu(0,x), v \geq p(x,z)) + \min[P(u \leq \nu(1,x'), v \leq p(x',z')), p(x,z)] \\
&\leq P(1,0|x,p) + \min[\inf_{\mathbf{\Omega}^{+}_{\mathbf{01}}(x)} \inf_{\mathbf{P}^{+}(x',p)} P(1,1|x',p'), p].
\end{aligned}$$

The upper bound for $\nu(0,x)$ we just built is lower than SV's bounds, but may not be sharp, since it is also possible to take advantage of the knowledge of the sign of $[\nu(0,x')-\nu(0,x)]$ and the range of $\mathbf{P}^+(x,z)$.

Second step :
Let's assume that there exists (x',z^*) such that $p(x',z^*)=0$, then

$$\begin{aligned}
\nu(0,x') &= P(u\le\nu(0,x'))\\
&= P(u\le\nu(0,x'),v\ge p(x',z^*))+P(u\le\nu(0,x'),v\le p(x',z^*))\\
&= P(u\le\nu(0,x'),v\ge p(x',z^*))\\
&= P(1,0|x',p^*).
\end{aligned}$$

Thus $\nu(0,x')$ is point identified. As it was pointed out by Vytlacil and Yildiz (2007), we also have point identification if there exists (x'',z'') such that $\nu(0,x')=\nu(1,x'')$ and $p(x',z')=p(x'',z'')$. Indeed, we have :

$$\begin{aligned}
\nu(0,x') &= P(u\le\nu(0,x'),v\ge p')+P(u\le\nu(0,x'),v\le p')\\
&= P(u\le\nu(0,x'),v\ge p')+P(u\le\nu(1,x''),v\le p'')\\
&= P(1,0|x',p')+P(1,1|x'',p'').
\end{aligned}$$

Moreover, we may have identification under weaker assumptions as showed in Jun, Pinkse and Xu (2011). Previously we bounded $P(u\le\nu(0,x),v\le p(x,z))$ by $P(u\le\nu(1,x'),v\le p(x',z'))$ because the first term is non identifiable from the data while the second term may be identifiable. If $\nu(0,x')$ is point identified we can identify $P(u\le\nu(0,x'),v\le p(x',z'))$ from the data. Indeed, since $P(u\le\nu(0,x'),v\le p(x',z'))=\nu(0,x')-P(u\le\nu(0,x'),v\ge p(x',z'))$, we have $P(u\le\nu(0,x'),v\le p(x',z'))=P(1,0|x',p^*)-P(1,0|x',p')$ or $P(u\le\nu(0,x),v\le p(x,z))=P(1,1|x'',p'')$ depending on the source of identification. Then, depending on the sign of $[\nu(0,x')-\nu(0,x)]$ we will be able to bound $\nu(0,x)$ from above by terms other than $P(1,0|x,p)+\min[\inf_{\mathbf{\Omega}^+_{\mathbf{01}}(x)}\inf_{\mathbf{P}^+(x',p)}P(1,1|x',p'),p]$. Indeed, for all $x'\in\mathbf{\Omega}^+_{\mathbf{00}}(x)$ and $p'\in\mathbf{P}^+(x',p)$, such that $\nu(0,x')$ is point identified we have the following

$$\nu(0,x)\ \le\ P(u\le\nu(0,x),v\ge p(x,z))+P(u\le\nu(0,x'),v\le p(x',z')).$$

This means that we may take advantage from the identification of a given ASF (i.e $\nu(0,x')$) to refine bounds for others ASF (i.e $\nu(0,x)$). Moreover, our bounds should respect the necessary condition derived in (1.4). Therefore, I proposed the following strategy to take into account those features of the model. For all $x' \in \mathbf{\Omega^+_{00}}(x)$ and $p' \in \mathbf{P}^+(x', p)$ we have :

$$\begin{aligned} \nu(0, x') - \nu(0, x) &\geq \max[P(1, 0|x', p') - P(1, 0|x, p), 0] \\ &\geq \sup_p \Big\{ \sup_{\mathbf{P^+(x',p)}} \max[P(1, 0|x', p') - P(1, 0|x, p), 0] \Big\}. \end{aligned}$$

Similarly, we can derive the lower bound for $\nu(0, x)$, and also the lower and upper bounds for $\nu(1, x)$. I have just shown that $\nu(0, x)$ and $\nu(1, x)$ should respect some necessary conditions. The following theorem proves that these necessary conditions are sufficient to fully characterize the empirical content of the model. The proof is quite involved and it relies on copula theory and a characterization theorem in Chiburis (2010).

Theorem 1. *Suppose Y, D determined by model (1.3) . Under assumptions 4 and 5 , the characterization of the identified set for* $\nu(0, .), \nu(1, .)$ *is the following*

$$\begin{aligned} & \sup_p \Big\{ P(1, 0|x, p) + \sup_{\mathbf{\Omega^-_{01}}(x)} \sup_{\mathbf{P}^-(x',p)} P(1, 1|x', p') \Big\} \\ \leq\ & \nu(0, x) \\ \leq\ & \inf_p \Big\{ P(1, 0|x, p) + \min(\inf_{\mathbf{\Omega^+_{01}}(x)} \inf_{\mathbf{P}^+(x',p)} P(1, 1|x', p'), p) \Big\}, \end{aligned} \tag{1.8}$$

$$\begin{aligned} & \sup_p \Big\{ P(1, 1|x, p) + \sup_{\mathbf{\Omega^-_{10}}(x)} \sup_{\mathbf{P}^+(x',p)} P(1, 0 \mid x', p') \Big\} \\ \leq\ & \nu(1, x) \\ \leq\ & \inf_p \Big\{ P(1, 1|x, p) + \min(\inf_{\mathbf{\Omega^+_{10}}(x)} \inf_{\mathbf{P}^-(x',p)} P(1, 0|x', p'), (1 - p)) \Big\} \end{aligned} \tag{1.9}$$

for all $x' \in \mathbf{\Omega^+_{00}}(x)$

$$\nu(0, x') - \nu(0, x) \geq \sup_p \Big\{ \sup_{\mathbf{P^+(x',p)}} \max[P(1, 0|x', p') - P(1, 0|x, p), 0] \Big\}$$

and for all $x' \in \mathbf{\Omega^+_{11}}(x)$

$$\nu(1, x') - \nu(1, x) \geq \sup_p \Big\{ \sup_{\mathbf{P^-(x',p)}} \max[P(1, 1|x', p') - P(1, 1|x, p), 0] \Big\}.$$

We can remark in this characterization of the identified set that there exists a dependence between the sharp bounds for $\nu(d,x)$ and $\nu(d,x')$. An equivalent characterization of the identified set which provides more intuition on this dependence may be derived. Let's assume that we know the sharp bounds $[SL_0(x'), SU_0(x')]$ for $\nu(0,x')$, (1.10) may be rewritten equivalently as follows : for all $x' \in \mathbf{\Omega^+_{00}}(x)$ and $p' \in \mathbf{P}^+(x',p)$

$$\begin{aligned}\nu(0,x) - P(u \leq \nu(0,x), v \geq p) &\leq \nu(0,x') - P(u \leq \nu(0,x'), v \geq p') \\ &\leq SU_0(x') - P(u \leq \nu(0,x'), v \geq p').\end{aligned}$$

Hence,

$$\begin{aligned}\nu(0,x) &\leq P(u \leq \nu(0,x), v \geq p) + SU_0(x') - P(u \leq \nu(0,x'), v \geq p') \\ &\leq P(1,0|x,p) + (SU_0(x') - P(1,0|x',p')).\end{aligned}$$

Therefore,

$$\nu(0,x) \leq P(1,0|x,p) + \inf_{\mathbf{\Omega^+_{00}(x)}} \inf_{\mathbf{P^+(x',p)}} (SU_0(x') - P(1,0|x',p')). \quad (1.10)$$

By combining the upper bounds for $\nu(0,x)$ derived in (1.8) and (1.10). We may propose the following upper bound for $\nu(0,x)$.

$$\nu(0,x) \leq \inf_p \Big\{ P(1,0|x,p) + \min\Big[\min(\inf_{\mathbf{\Omega^+_{01}}(x)} \inf_{\mathbf{P}^+(x',p)} P(1,1|x',p'), p), \\ \inf_{\mathbf{\Omega^+_{00}(x)}} \inf_{\mathbf{P^+(x',p)}} (SU_0(x') - P(1,0|x',p'))\Big]\Big\}.$$

Similarly, we can derive the lower bound for $\nu(0,x)$, and also the lower and upper bounds for $\nu(1,x)$. Then, we have the following equivalent characterization of the identified set for $\nu(0,.)$ and $\nu(1,.)$. Hereafter, I shall denote by $SL_d(x)$ and $SU_d(x)$ respectively the sharp lower bound and sharp upper bound for $\nu(d,x)$ d=0,1.

Corollary 1. *Suppose Y, D determined by model (1.3). Under assumptions 4 and 5 , the characterization of the identified set for* $\nu(0,.), \nu(1,.)$ *is the fol-*

lowing

$$\sup_p \Big\{ P(1,0|x,p) + \max\Big[\sup_{\mathbf{\Omega}^-_{\mathbf{01}}(x)} \sup_{\mathbf{P}^-(x',p)} P(1,1|x',p'), \sup_{\mathbf{\Omega}^-_{\mathbf{00}}(x)} \sup_{\mathbf{P}^-(x',p)} (SL_0(x') - P(1,0|x',p'))\Big]\Big\}$$

$$\leq \nu(0,x)$$

$$\leq \inf_p \Big\{ P(1,0|x,p) + \min\Big[\min(\inf_{\mathbf{\Omega}^+_{\mathbf{01}}(x)} \inf_{\mathbf{P}^+(x',p)} P(1,1|x',p'),p), \inf_{\mathbf{\Omega}^+_{\mathbf{00}}(\mathbf{x})} \inf_{\mathbf{P}^+(\mathbf{x'},\mathbf{p})} (SU_0(x') - P(1,0|x',p'))\Big]\Big\}$$

and

$$\sup_p \Big\{ P(1,1|x,p) + \max\Big[\sup_{\mathbf{\Omega}^-_{\mathbf{10}}(x)} \sup_{\mathbf{P}^+(x',p)} P(1,0 \mid x',p'), \sup_{\mathbf{\Omega}^-_{\mathbf{11}}(x)} \sup_{\mathbf{P}^+(x',p)} (SL_1(x') - P(1,1|x',p'))\Big]\Big\}$$

$$\leq \nu(1,x)$$

$$\leq \inf_p \Big\{ P(1,1|x,p) + \min\Big[\min(\inf_{\mathbf{\Omega}^+_{\mathbf{10}}(x)} \inf_{\mathbf{P}^-(x',p)} P(1,0|x',p'),(1-p)), \inf_{\mathbf{\Omega}^+_{\mathbf{11}}(x)} \inf_{\mathbf{P}^-(x',p)} (SU_1(x') - P(1,1|x',p'))\Big]\Big\}.$$

The intuition which allows us to derive the sharp bounds for $\nu(0,x)$ by using variation in covariates in corollary 1 is the following : the width of the bounds on $\nu(0,x')$ depends on $Supp(P|X=x')$. Then, when $supp(P|X=x')$ is large the bounds for $\nu(0,x')$ become narrower. Using the relation that our model imposes between $\nu(0,x')$ and $\nu(0,x)$, we may refine the bounds for $\nu(0,x)$ using narrower bounds for $\nu(0,x')$. This fact explains why the sharp bounds for $\nu(0,x)$ depend on $SL_0(x')$ and $SU_0(x')$, which are the sharp bounds for $\nu(0,x')$. This dependence vanishes for the lower bound $SL_0(x)$ when $x' \notin \mathbf{\Omega}^-_{\mathbf{00}}(x)$ and for the upper bound when $x' \notin \mathbf{\Omega}^+_{\mathbf{00}}(x)$. Therefore, when $\mathbf{\Omega}^+_{\mathbf{00}}(x) = \emptyset$, the upper bound for $\nu(0,x)$ becomes $SU_0(x) = \inf_p \Big\{ P(1,0|x,p) + \min(\inf_{\mathbf{\Omega}^+_{\mathbf{01}}(x)} \inf_{\mathbf{P}^+(x',p)} P(1,1|x',p'),p)\Big\}$ which no longer depends on others ASF sharp bounds.

1.3.4 Computation of the bounds

For the construction of our bounds we need to know the ordering of the elements of the set $S_1 = \{\nu(d,x) : x \in Dom(X) \text{ and } d \in \{0,1\}\}$, in order to compute the collections of subsets defined in Table 1.1. Without restrictions on the true ordering on S_1, one may go over all possible orderings of S_1, and keep only orderings for which the bounds derived in Theorem 1 do not cross. Unfortunately, this method may be very costly, even under parametric restrictions for ν. Indeed, to derive sharp bounds on ATE for the model (1.3), Chiburis (2010) proposes to visit all possible orderings of S_1. In his empirical example, he assumes parametric restrictions for ν to reduce the number of orderings to be checked, but still fails to determine the bounds on the ATE in some cases due to computational intractability, even for a very limited $Dom(X)$. Even though our proposed bounds are easier to derive than bounds in Chiburis (2010), it is valuable to find a methodology to reduce the number of orderings to be checked. The properties of the function $h(x,x',p,p')$ and the functions $\tilde{h}_d(x,x',p,p')$, $h(d,x,x',p_1,p_2,p'_1,p'_2)$, $\tilde{h}_d(x,x',p_1,p_2,p'_1,p'_2)$ defined in Lemma 2 help to identify a partial ordering on S_1, which can dramatically reduce the number of orderings to be checked, particularly when $Supp(P \mid X = x) \cap Supp(P \mid X = x')$ is large. This is illustrated in this following example :

Example 1. *Denoting $P(Y = 1, D = d|X = x, Z = z) = p(1,d|x,z)$, consider $X = \{0,1\}$, $Z = \{0,1\}$, and the following observables :*

$$\begin{array}{ll} p(1,1|0,0) = \beta_1 & p(1,0|0,0) = \gamma_1 \\ p(1,1|0,1) = \beta_2 & p(1,0|0,1) = \gamma_2 \\ p(1,1|1,0) = \beta_3 & p(1,0|1,0) = \gamma_3 \\ p(1,1|1,1) = \beta_4 & p(1,0|1,1) = \gamma_4, \end{array}$$

with $0 < \beta_i, \gamma_i < 1$ for $i = 1...4$; and $p(x,z)$ such that

1. $\beta_1 - \beta_2 > \gamma_2 - \gamma_1 > \gamma_2 - \gamma_1 > \beta_3 - \beta_4 > \gamma_4 - \gamma_3$
2. $p(0,1) = p(1,1) < p(0,0) < p(1,0)$

Those conditions are sufficient to determine the sign of some marginal average effects. We can describe the set of all possible orderings on S_1 as follows :

$$\nu(d_1,d'_1) < \nu(d_2,d'_2) < \nu(d_3,d'_3) < \nu(d_4,d'_4),$$

where $\{d_i, d_i'\} \neq \{d_j, d_j'\}$ *for* $i \neq j$ *and* $d_i, d_i' \in \{0,1\}$*, hence there are 24 possible orderings.*

We have $Supp(P|X = 0) = \{p(0,1), p(0,0)\}$ *and* $Supp(P|X = 1) = \{p(1,1), p(1,0)\}$*. Thus,* $Supp(P \mid X = 0) \neq Supp(P \mid X = 1)$*. The approach of SV can allow to identify only the sign of* $[\nu(1,1) - \nu(0,1)]$ *and* $[\nu(1,0) - \nu(0,0)]$ *from the observable function h(x,x',p,p').*

$$\begin{aligned} sign[\nu(1,1) - \nu(0,1)] &= sign[h(1,1,p(1,1),p(1,0))] = + \\ sign[\nu(1,0) - \nu(0,0)] &= sign[h(0,0,p(0,0),p(0,1))] = +. \end{aligned}$$

Then, we have the following partial ordering $\nu(0,1) \leq \nu(1,1)$ *and* $\nu(0,0) \leq \nu(1,0)$ *on* S_1*. Among the 24 possible orderings only 6 are compatible with restrictions imposed by this partial ordering. We can see that even in a worst case when* $Supp(P \mid X = x) \cap Supp(P \mid X = x')$ *is empty or a singleton, the number of orderings to be checked drops from 24 to 6 (i.e when* $Supp(P \mid X = x)$ *is not a singleton we can always identified the sign of* $sign[\nu(1,x) - \nu(0,x)]$*). Moreover, we can see that the sign of* $[\nu(0,1) - \nu(1,0)]$ *may be identified using the function* $h(0,0,1,p(1,1),p(1,0),p(0,1),p(0,0))$ *defined in Lemma 2. Indeed,* $sign[\nu(0,1) - \nu(1,0)] = +$*. Then, we have only one ordering compatible with the data in this generic case :* $\nu(0,0) < \nu(1,0) < \nu(0,1) < \nu(1,1)$*.*

Since the number of orderings to be checked is reduced, it is now valuable to propose an efficient procedure to construct our bounds. The following iterative procedure can be used. The idea is to use the last information we obtained on previous sharp bounds, to sharpen remaining ASF bounds. For instance, first, construct $SU_0(x)$ such that $\mathbf{\Omega^+_{00}}(x) = \emptyset$, then construct $SU_0(x')$ such that $x = \mathbf{\Omega^+_{00}}(x')$, then $SU_0(x'')$ such that $\{x, x'\} = \mathbf{\Omega^+_{00}}(x'')$ and then iterate the strategy. The same iterative procedure holds to derive the lower sharp bounds for every element of the set $\{\nu(0,x) : x \in Dom(X)\}$, and also the lower and upper sharp bounds for every element of the set $\{\nu(1,x) : x \in Dom(X)\}$. For instance, in the latter example, one plausible ordering to be checked is the following :

$$\nu(0,0) < \nu(1,0) < \nu(0,1) < \nu(1,1).$$

Our procedure proposes to construct the lower bounds in the following ordering :

$$SL_0(0), \quad SL_1(0), \quad SL_0(1), \quad SL_1(1),$$

and the upper bounds in the following ordering :

$$SU_1(1), \quad SU_0(1), \quad SU_1(0), \quad SU_0(0).$$

1.4 Numerical illustration

Now, I provide a numerical illustration of the bounds on ATE using Theorem 1. In addition to the bounds proposed in Theorem 1, I will compute the SV bounds. Consider the following special case of the model :

$$\begin{aligned} Y &= 1\{\alpha D + x\beta > \epsilon_1\} \\ D &= 1\{x\gamma + z\eta > \epsilon_2\}, \end{aligned} \tag{1.11}$$

with $Dom(X) = [-2, 2]$ and $Dom(Z) = \{0, 1\}$ and $(\epsilon_1, \epsilon_2) \backsim N(0, \sum)$ where

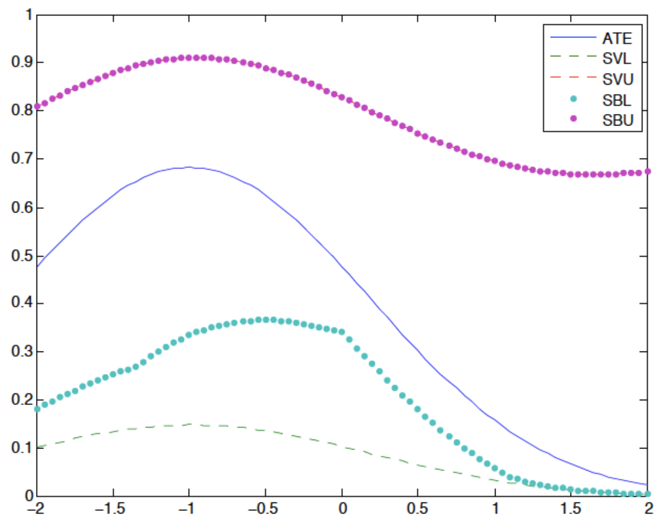
$$\sum = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

We can easily see that the SV support condition fails to hold for $\gamma \neq 0$. Indeed, $Supp(P \mid X = x) \cap Supp(P \mid X = x')$ is either empty or reduces to a singleton, for all $x \in [-2, 2]$. I will construct the bounds by fixing $\alpha = 2$, $\rho = \frac{1}{2}$ while varying other parameters. I consider all the ordering induced by the parametric form $\alpha' D + x\beta'$. In fact, every couple of parameter (α', β') induces one ordering. Before constructing the bounds I will apply Lemma 2 to reduce the number of ordering to be checked. For example, the function $h(x, x, p, p')$ allows to identify the sign of $[\nu(1, x) - \nu(0, x)]$ for all x, which implies that $\alpha' > 0$. Now, we may visit only the ordering induced by $(\alpha' > 0, \beta')$. In this numerical illustration, I consider all the ordering induced by $\alpha' \in (0, 5)$ and $\beta' \in [\beta - 2.5, \beta + 2.5]$. It is possible to consider a lager space, but within the simulation we note that most of the orderings induced by (α', β') are rejected whenever those orderings deviate slightly from the true ordering induced by the true parameters (α, β).

All the figures show the $ATE(x)$ while x varying from $[-2, 2]$ for different values of the parameters. Here, I will describe four different facts :

1. The Figure 1.1 represents the case where $(\beta = 1, \gamma = \frac{1}{5}, \eta = \frac{1}{2})$. We can see that our lower bound improves significantly on the SV lower bound, while the upper bound is exactly the same. Indeed, since

FIGURE 1.1 – Sharp bounds on ATE when ($\beta = 1, \gamma = \frac{1}{5}, \eta = \frac{1}{2}$).

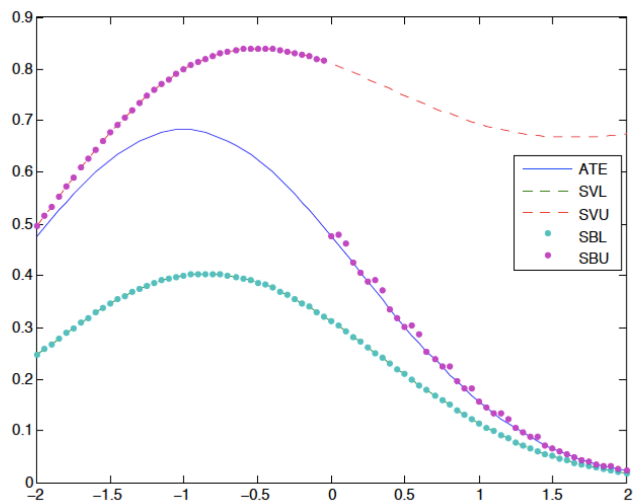


$\nu(0, x') \leq \nu(1, x)$ for many values of (x, x') our bounds refine the lower bound of $\nu(1, x)$, similarly for the upper bound of $\nu(0, x')$. However, the bounds do not refine the SV upper bound for two main raisons. There are only few values of (x, x') such that $\nu(1, x) \leq \nu(0, x')$ and in the case where it holds it is likely to have $p(x', z') \geq p(x, z)$, while we need to have $p(x', z') \leq p(x, z)$ if we want to refine $\nu(1, x)$ using $\nu(0, x')$.

2. In figure 1.2, I increase the strength of the effect of the instrument ($\beta = 1, \gamma = \frac{1}{5}, \eta = 4$). I note two important facts : First, I am now able to refine the upper bound of $\nu(1, x)$ using $\nu(0, x')$ since it is likely that $p(x', 0) \leq p(x, 1)$. The discontinuity denotes the point where both following conditions start to hold simultaneously : $\nu(1, x) \leq \nu(0, x')$ and $p(x', 0) \leq p(x, 1)$. Second, when the strength of the instrument increases we tend to have identification. This phenomenon is an example of identification at infinity as in Heckman (1990). We can see that the SV bounds do not respect this feature when their support condition fails.

3. In Figure 1.3, I decrease the strength of the covariate in the selection equation, to see how the bounds behave when the SV support condi-

FIGURE 1.2 – Sharp bounds on ATE when ($\beta = 1, \gamma = \frac{1}{5}, \eta = 4$).



tion almost holds ($\beta = 1, \gamma = \frac{1}{100}, \eta = 1$). We see that our bounds improve on the SV bounds. Indeed, when we are getting closer to the "perfect matching restriction" our bounds become better in a continuous way, but the SV bounds are not sensitive to that and jump directly to the tightest bounds when $\gamma = 0$.

4. In Figure 1.4, I reduce the strength of the covariate in the outcome equation, ($\beta = 1, \gamma = \frac{1}{5}, \eta = \frac{1}{2}$), I note that both type of bounds are very wide, and I get only a small improvement over the SV bounds. Indeed, the strength of this present analysis is based on the variation of the covariates. Then, when $Dom(X)$ is small, our improvement over the SV might be small. This fact explains why Chiburis (2010) found only a small improvement over the SV bounds within his empirical example where the domain of X was $\{0, 1\}^2$.

1.5 Empirical illustration

This empirical illustration examines the effect of migration decisions of cameroonian students on the standard of living of the family left behind. Modelling such effects is a difficult problem due to the endogeneity of the migration decision. As it was pointed out by Nakosteen and Zimmer (1980)

FIGURE 1.3 – Sharp bounds on ATE when $(\beta = 1, \gamma = \frac{1}{100}, \eta = 1)$.

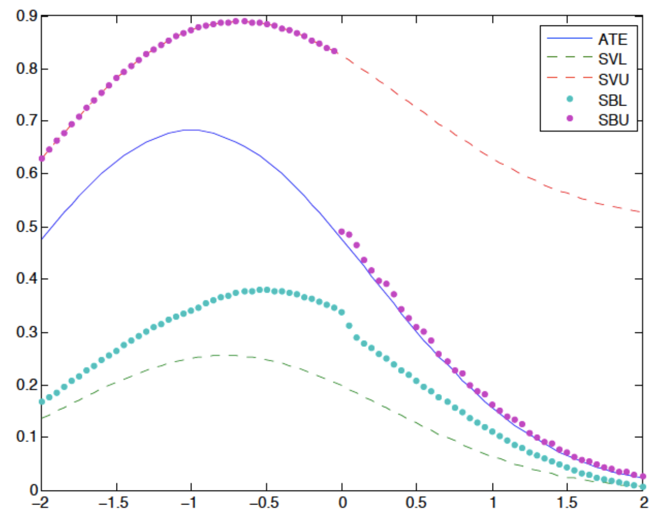
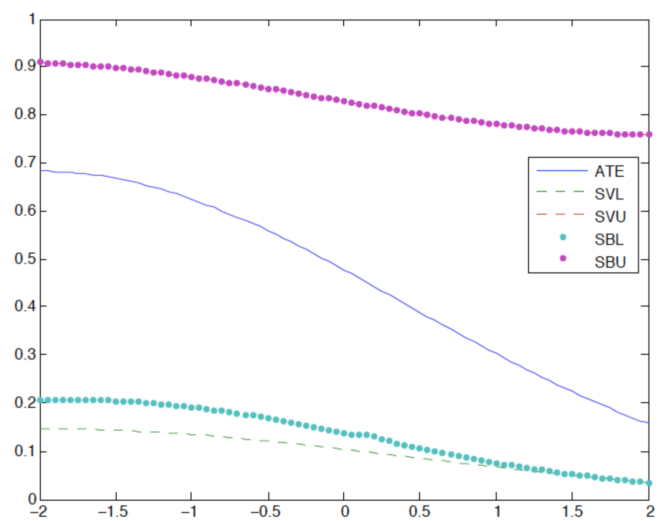


FIGURE 1.4 – Sharp bounds on ATE when $(\beta = 1, \gamma = \frac{1}{5}, \eta = \frac{1}{2})$.



the migration decision is not random because families make the decision to send their children to other countries because they have some basis for perceiving a more favorable future outcome from this decision. This rationality implies that individuals tend to self-select. Even if we were able to control the selection on observables, we would not be sure that all factors that may affect migration decisions and the standard of living of the family left behind were taken into account. We may have selection on unobservables, for instance, the income of the parents, child motivations and the degree of parental affection. Therefore, the selection issues are still present. Lucas (1997) and Adams (1993) suggest a non linear relationship between income or wealth and the migration decision and argue that there is a certain threshold such that for income or wealth over this threshold, migration decisions are not considered necessary. I, therefore, model the effect of the migration decision on the standard of living of the family left behind in Cameroon with the joint threshold crossing model considered throughout this paper.

In the following, I first present the database, the explanatory variables I use, and some summary statistics, before giving a short discussion on the possible instruments. Then, I construct the empirical analogues of SV's bounds for the ASF, followed by the construction of the empirical analogues of our proposed bounds and discuss the results.

1.5.1 Data

For illustration I use a sample of 307 Cameroonians who are all high school graduates. This sample is taken from a survey conducted by Romuald Méango from the University of Montreal. He conducted an online survey from March 27, 2011 to May 8, 2011. The title of his survey was "Migration des jeunes Camerounais apres le BAC" (Migration of young Cameroonians after high-School). For each individual from our database we know her/his current migration status, her/his father's education at the moment of the migration decision, the migration status of her/his siblings at the moment of the migration decision, and if she/he graduated from high-school with honors. To measure the improvement in the standard of living of the family, I use available information we have on capital ownership (fridge, car and land) at the moment of the migration decision and at the time where the survey was made. Table 1.2 presents the list of the variables used in the analysis. More

http ://migration-cameroun.com/website/index.php.

summary statistics can be found in Appendix .2.

Table 1.2 – List of variables

	Variables	Equal to 1 if :	Percentage of sample
Edp	Father's educ level	univ graduate and more	34.52
Imp	Previous migration	1 and more	26.38
Hon	Graduate h-school with honors	yes	25.08
D	Migration decision	yes	60.91
Y	Improve in the standard of living	add unit on capital ownership	18.24

The following equation describes the relationship between the migration decision (D) and the improvement in the standard of living of the family left behind (Y) taking into consideration the available variables that are likely to influence D and Y.

$$\begin{aligned} Y &= 1\{\nu(D,\ Edp,\ Imp) > u\} \\ D &= 1\{P(Edp,\ Imp,\ Hon) > v\}. \end{aligned}$$

I consider in this model the variable Hon as an instrument. The validity of our proposed instrument remains debatable but I think that it is still the most appropriate one among all available candidates from our database. Indeed, graduating high-school with honors increases significantly the probability with which the student will get a scholarship from another country without appearing correlated with the income or the wealth of the family.

1.5.2 Empirical results

Shaikh and Vytlacil's (2011) bounds

I begin by deriving the empirical analogues of SV's bounds for the different ASF. I get the following result :

$$\begin{aligned}
0.0667 &\leq \nu(0,1,0) \leq 0.1053 \\
0.2000 &\leq \nu(1,1,0) \leq 0.4000 \\
0.2655 &\leq \nu(0,0,0) \leq 0.5840 \\
0.1327 &\leq \nu(1,0,0) \leq 0.1333 \\
0.1428 &\leq \nu(0,0,1) \leq 0.6250 \\
0.1250 &\leq \nu(1,0,1) \leq 0.1250.
\end{aligned}$$

I provide all ASF bounds except for $\nu(D, Edp = 1, Imp = 1)$. Indeed, for this ASF I find some inconsistency in the bounds, in the sense that the lower bound is greater than the upper bound. This inconsistency may have occurred due either to the small sample problem or the violation of the joint assumption of threshold crossing restriction imposed on (D) and the presence of a valid instrument Z.

To deal with this inconsistency I propose to remove the equation $(D = 1\{P(Edp, Imp, Hon) > v\})$ and to construct the bounds for the threshold crossing model with unrestricted binary treatment. Then, we may apply the result of proposition 1 when no instrument is available. I get the following result :

— if $\nu(0,1,1) \leq \nu(1,1,1)$

$$\begin{aligned}
0 &\leq \nu(0,1,1) \leq 0.1052 \\
0.1052 &\leq \nu(1,1,1) \leq 0.2368,
\end{aligned}$$

— if $\nu(0,1,1) \geq \nu(1,1,1)$

$$\begin{aligned}
0.1052 &\leq \nu(0,1,1) \leq 0.8684 \\
0.1052 &\leq \nu(1,1,1) \leq 0.1052.
\end{aligned}$$

These bounds give no improvement over Manski and Pepper's (2000) bounds on the ASF since I do not use an instrument to determine the sign of $[\nu(1,1,1) - \nu(0,1,1)]$ as it is possible with the others ASF.

Improvement over Shaikh and Vytlacil's (2011) bounds

The SV bounds would be sharp if we had $\{p(Edp, Imp, 1), p(Edp, Imp, 0)\} = \{p(Edp', Imp', 1), p(Edp', Imp', 0)\}$ for all $(Edp, Imp) \neq (Edp', Imp')$. As we can remark in Table 1.3, this "perfect matching restriction" doesn't hold.

TABLE 1.3 – Migrant's propensity scores

	Migrants	Non-migrants	Migrant's propensity scores
#{Edp=1,Imp=1,Hon=1}	17	2	0.8947
#{Edp=1,Imp=1,Hon=0}	16	3	0.8421
#{Edp=1,Imp=0,Hon=1}	22	8	0.7333
#{Edp=1,Imp=0,Hon=0}	20	18	0.5263
#{Edp=0,Imp=0,Hon=1}	33	12	0.7333
#{Edp=0,Imp=0,Hon=0}	51	62	0.4513
#{Edp=0,Imp=1,Hon=1}	5	3	0.6250
#{Edp=0,Imp=1,Hon=0}	23	12	0.6571

Then, let's construct the empirical analogues of our proposed bounds. In fact, according to the SV bounds we have :

$$\begin{aligned} \nu(0,1,0) &\leq \nu(1,0,0) \\ \nu(1,0,0) &\leq \nu(0,0,1). \end{aligned}$$

By using the ranges $\mathbf{P}^+(Edp = 1, Imp = 0, Hon = 0)$ and $\mathbf{P}^-(Edp = 0, Imp = 1, Hon = 0)$ I improve the upper bound for $\nu(0,1,0)$ and the lower bound for $\nu(0,0,1)$. Indeed, we have :

$$\begin{aligned} \nu(0,1,0) &\leq P(u \leq \nu(0,1,0), v \geq p(1,0,0)) + P(u \leq \nu(1,0,0), v \leq p(0,0,1)) \\ &\leq P(u \leq \nu(0,1,0), v \geq 0.5263) + P(u \leq \nu(1,0,0), v \leq 0.7333) \\ &= 0.0930 \\ &= 0.1053 - 0.0123, \end{aligned}$$

and

$$\begin{aligned}\nu(0,0,1) &\geq P(u \leq \nu(0,0,1), v \geq p(0,1,0)) + P(u \leq \nu(1,0,0), v \leq p(0,0,0)) \\ &\geq P(u \leq \nu(0,0,1), v \geq 0.6571) + P(u \leq \nu(1,0,0), v \leq 0.4513) \\ &= 0.1613 \\ &= 0.1428 + 0.0185.\end{aligned}$$

Then, our proposed bounds are significant improvement over the upper bound for $\nu(0,1,0)$ and the lower bound for $\nu(0,0,1)$.

Discussion on the empirical results

This illustration is flawed in two aspects. First, I used estimated rather than true probabilities without applying any inference procedure. Second, the sample is non-random. However, our aim was just to provide one empirical situation where the "critical support condition"of SV fails and show how it is possible to narrow their bounds in such a case. Nevertheless, our result give some insights on the effect of migration decisions on the standard of living of the family left behind in Cameroon for this particular sample. Indeed, the sample suggests that :

— The probability to improve the standard of living of the family in Cameroon is greater for the migrants with well-educated father rather than for the non-migrants with well-educated father i.e ($\nu(1,1,0) \geq \nu(0,1,0)$).

— The probability to improve the standard of living of the family left in Cameroon is greater for the first migrant of the family i.e ($\nu(1,0,1) \leq \nu(1,0,0)$).

1.6 Conclusion

I have considered the special case of joint threshold crossing model, where no parametric form or distributional assumptions are imposed. I provided sharp bounds on the Average Treatment effect (ATE) when I imposed only mild regularity conditions on the distribution of unobservable variables. I presented a methodology which allows to construct sharp bounds on the ATE by efficiently using variation on covariates without imposing any support restrictions. A numerical illustration showed our proposed bounds may have significant improvement over the Shaikh and Vytlacil (2011) bounds, which where until now, the tightest feasible bounds proposed in the literature for this model. Also, an illustration of our bounds was carried out on the analysis of the effect of migration decisions. There are several natural extensions of this work. First, this methodology may be easily used to provide sharp bounds for others functionals of treatment, not just the average. Second, this methodology efficiently exploits variation on covariates to sharpen bounds and it may be extended to narrow cross-sectional bounds using time variation in panel data. Finally, it has been extended to provide sharp bounds on the ATE in the generalized discrete Roy model.

Chapitre 2

Sharp bounds in the binary Roy model

2.1 Introduction

A large literature has developed since Heckman and Honoré (1990) on the empirical content of the Roy model of sectorial choice with sector specific unobserved heterogeneity. Most of this literature, however, concerns the case of continuous outcomes and many applications, where outcomes are discrete, fall outside its scope. They include analysis of the effects of different training programs on the probability of renewed employment, of competing medical treatments or surgical procedures on the probability of survival, of higher education on the probability of migration and of competing policies on schooling decisions in developing countries among numerous others. The Roy model is still highly relevant to those applications, but very little is known of its empirical content in such cases. Sharp bounds are derived in binary outcome models with a binary endogenous regressor in Chesher (2010),

---

This chapter is a joint work with Marc Henry. This chapter was conducted in part, while Marc Henry was visiting the University of Tokyo and I was visiting Penn State. We thank our respective hosts for their hospitality and support. Helpful discussions with Laurent Davezies, James Heckman, Hidehiko Ichimura, Koen Jochmans, Aureo de Paula and comments from seminar audiences in Cambridge, Ecole polytechnique, Princeton, UCL, UPenn and participants at the Vanderbilt conference on identification and inference in microeconometrics are also gratefully acknowledged.

Shaikh and Vytlacil (2011), Chiburis (2010), Jun, Pinkse, and Xu (2010) and Mourifié (2011)under a variety of assumptions, which all rule out sector specific unobserved heterogeneity. Finally, Heckman and Vytlacil (1999) derive identification conditions in a parametric version of the binary Roy model.

We consider three distinct versions of the binary Roy model : the original model, where selection is based solely on the probability of success ; the extended Roy model (in the terminology of Heckman and Vytlacil (1999)), where selection depends on the probability of success and a function of observable variables (sometimes called "nonpecuniary component") ; and the generalized Roy model (in the terminology of Heckman and Honoré (1990)), with selection specific unobservable heterogeneity. When considering the generalized Roy model, we further distinguish restrictions on the selection equation and restrictions on the joint distribution of sector specific unobserved heterogeneity. We specifically consider the case, where selection variables are independent of sector specific unobserved heterogeneity and the case, where sector specific unobserved heterogeneity follows a factor structure proposed in Aakvik, Heckman, and Vytlacil (2005).

Following Heckman, Smith, and Clements (1997), we apply results from optimal transportation theory to derive sharp bounds on the structural parameters, from which a range of treatment parameters can be derived. More specifically, we apply Theorem 1 of Galichon and Henry (2011) (equivalently Theorem 3.2 of Beresteanu, Molchanov, and Molinari (2011)) to derive bounds for the generalized binary Roy model. The latter Theorem was recently applied in a similar context by Chesher, Rosen, and Smolinski (2011) to derive sharp bounds for instrumental variable models of discrete choice. We spell out the point identification implications of the bounds under certain exclusion restrictions. The bounds are simple enough to lend themselves to existing inferential methods, specifically Chernozhukov, Lee, and Rosen (2009) and Andrews and Shi (2011) in the instrumental variables case.

The remainder of the paper is organized as follows. Section 3.2 clarifies the analytical framework and the objectives. In Section 2.3, sharp bounds are derived for the binary Roy model, when selection depends only on the probability of success and possibly on observable variables. Identification implications are spelled out under exclusion restrictions. Section 2.4 considers the generalized binary Roy model and the last section concludes.

2.2 Analytical framework

We adopt the framework of the potential outcomes model $Y = Y_1 D + Y_0(1-D)$, where Y is an observed outcome, D is an observed selection indicator and Y_1, Y_0 are unobserved potential outcomes. Heckman, and Vytlacil (2009) trace the genealogy of this model and we refer to them for terminology and attribution. Potential outcomes are as follows :

$$Y_d = 1\{Y_d^* > 0\} = 1\{F(d, X_d, u_d) > 0\},\ d = 1, 0, \tag{2.1}$$

where $1\{.\}$ denotes the indicator function and F is an unknown function of the vector of observable random variables X_d and unobserved random variable u_d. We make the following assumptions throughout Sections 2.3 and 2.4.

Assumption 9 (Weak separability)**.** *Potential outcomes can then be written $Y_d = 1\{f_d(X_d) > u_d\}$ for some unknown (measurable) functions f_d, $d = 0, 1$. As shown in Vytlacil (2002), the latter is implied by weak separability of the functions $F(d, X_d, u_d)$, d=1,0.*

Assumption 10 (Regularity)**.** *The sector specific unobserved variables u_d, $d = 1, 0$, are uniformly continuous with respect to Lebesgue measure, so that they may be assumed without loss of generality to be distributed uniformly on $[0, 1]$.*

The normalization of Assumption 10 is very convenient, since it implies $f_d(x_d) = \mathbb{E}(Y_d|x_d, z)$ and bounds on treatment effects parameters can be derived from bounds on the structural parameters f_1 and f_0.

Assumption 11 (Instruments)**.** *Observable variables X_d, $d = 1, 0$, and instruments Z are independent of (u_1, u_0). Common components of X_1 and X_0 will be dropped from the notation in the remainder of the paper and by slight abuse of notation, X_d will refer only to the variables that are excluded from the equation for Y_{1-d} and Z to variables that are excluded from both outcome equations (when the case arises).*

In all the paper, we shall use the notation $\mathbb{P}(i, j|X)$ for $\mathbb{P}(Y = i, D = j|X)$ and $W = (Z, X_1, X_0)$, $\omega = (z, x_1, x_0)$.

Our objective is to characterize all the information that can be gathered from the distribution of observed variables (Y, W) about the unknown elements of the model, namely the functions f_1 and f_0 and the joint distribution (or copula, since the marginals are normalized) of the sector specific heteroneity vector (u_1, u_0). We shall call this characterizing the empirical content of the model. The empirical content of the model, relative to the unknown functions f_1 and f_2 will be of primary interest and will take the form of sharp bounds such as :

$$\underline{G}(\omega) \leq f_d(x_d) \leq \bar{G}(\omega), \tag{2.2}$$

in which case, exhibiting the functions $\underline{G}$ and $\bar{G}$ will be the object of the analysis. In the case of a linear specification of the binary Roy model $f_d(x_d) = \tilde{f}_d(\beta_d' x_d)$, where $\tilde{f}_d : \mathbb{R} \to [0,1]$ is a known invertible function of the single index $\beta_d' x_d$ and the unknown parameter vector β_d is the object of analysis, we can derive sharp bounds on β_d from (2.2) straighforwardly. From $\underline{G}(\omega) \leq \tilde{f}_d(\beta_d' x_d) \leq \bar{G}(\omega)$, we derive $\tilde{f}_d^{-1}\underline{G}(\omega) \leq \beta_d' x_d \leq \tilde{f}_d^{-1}\bar{G}(\omega)$. Hence, the bounds on the parameter vector β_d will be given by the projections of $\tilde{f}_d^{-1}\underline{G}(\omega)$ and $\tilde{f}_d^{-1}\bar{G}(\omega)$ on x_d.

2.3 Sharp bounds for the binary Roy and extended Roy models

2.3.1 Simple binary Roy model

In the original model proposed by Roy (1951), the sector yielding the highest outcome is selected, i.e., $D = 1\{Y_1^* > Y_0^*\}$. In the binary case, this is equivalent to selecting the sector with the highest probability of success. The empirical content of the model under this selection rule is characterized in Figures 2.1 and 2.2.

Bounds on the structural functions

For each value of the exogenous observable variables and each value of the pair (u_1, u_0), the outcome is uniquely determined. If the joint distribution were known, the likelihood of each of the potential outcomes $(Y = 1, D = 1)$, $(Y = 1, D = 0)$, $(Y = 0, D = 1)$ and $(Y = 0, D = 0)$ would be determined. However, only the marginal distributions of u_1 and u_0 are fixed, not the copula, so that only the probability of vertical and horizontal bands in Figures 2.1 and 2.2 are uniquely determined. Thus we see for instance that $f_1 = \mathbb{P}(Y = 1, D = 1)$ is identified when $f_0 = 0$ (as in Figure 2.2)

FIGURE 2.1 – Characterization of the empirical content of the simple binary Roy model in the unit square of the (u_1, u_0) space.

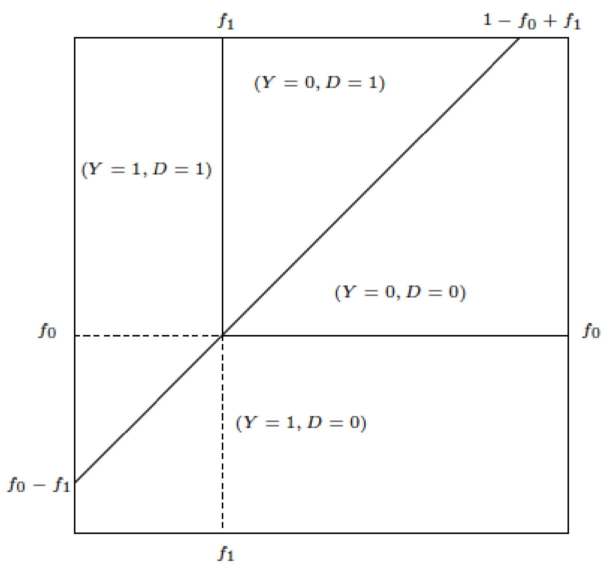
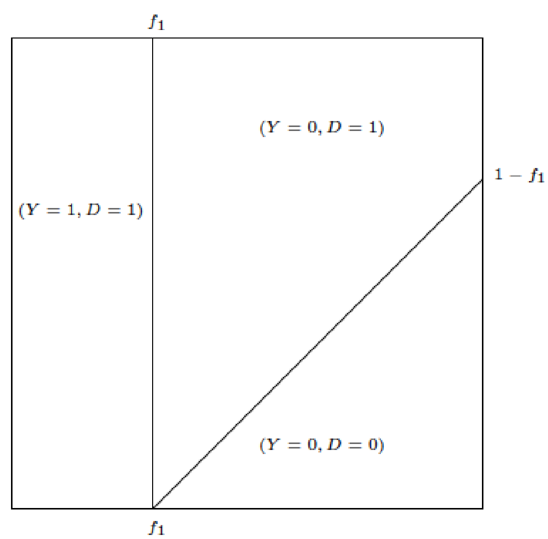


FIGURE 2.2 – Characterization of the empirical content of the simple binary Roy model in the unit square of the (u_1, u_0) space in case $f_0 = 0$.



and $f_0 = \mathbb{P}(Y = 1, D = 0)$ is identified when $f_1 = 0$ in a way that is akin to identification at infinity, as in Heckman (1990), when $f_i(x)$ follows a single index restriction. But in other cases (as in Figure 2.1), we only know $\mathbb{P}(Y = 1, D = 1) \leq f_1 \leq \mathbb{P}(Y = 1)$ and $\mathbb{P}(Y = 1, D = 0) \leq f_0 \leq \mathbb{P}(Y = 1)$. The following proposition, proved in the Appendix, shows that these bounds are jointly sharp.

Proposition 2 (Roy model)**.** *Under Assumptions 9-11, the following inequalities characterize the identified set for* (f_1, f_0) *under model (2.1) with* $D = 1\{Y_1^* > Y_0^*\}$.

$$\sup_{x_0} \mathbb{P}(1,1|x_1,x_0) \leq \quad f_1(x_1) \quad \leq \inf_{x_0} \Big[\mathbb{P}(1,1|x_1,x_0) + \mathbb{P}(1,0|x_1,x_0)1\{f_0(x_0) > 0\}\Big] \tag{2.3}$$

$$\sup_{x_1} \mathbb{P}(1,0|x_1,x_0) \leq \quad f_0(x_0) \quad \leq \inf_{x_1} \Big[\mathbb{P}(1,0|x_1,x_0) + \mathbb{P}(1,1|x_1,x_0)1\{f_1(x_1) > 0\}\Big] \tag{2.4}$$

where the infima and suprema are taken over the domains of the excluded variables X_1 *or* X_0 *as indicated and when they exist.*

The validity of the bounds was shown above. To prove sharpness, we show in Appendix .3 that we can construct joint distributions for (u_1, u_0) such that each of the extreme points of the identified region for $(f_1(x_1), f_0(x_0))$ defined by (2.3) and (2.4) are attained. Since the bounds in Proposition 2 are obtained as intersections over the domains of the excluded variables, they are called "intersection bounds". They are also semiparametric in the non excluded variables. Inference on such bounds can be conducted with existing methods described in Chernozhukov, Lee, and Rosen (2009) or in Andrews and Shi (2011).

A simple implication of selection equation $D = 1\{Y_1^* > Y_0^*\}$ is that actual success is more likely than counterfactual success.

Assumption 12 (Roy model)**.** $\mathbb{E}(Y_d|D = d, X_1, X_0) \geq \mathbb{E}(Y_{1-d}|D = d, X_1, X_0)$ *for* $d = 1, 0$.

Under Assumption 12, omitting conditioning variables for ease of notation,

$$\begin{aligned} f_d &= \mathbb{E}[Y_d] \\ &= \mathbb{E}[Y_d|D = d]\mathbb{P}(D = d) + \mathbb{E}[Y_d|D = 1-d]\mathbb{P}(D = 1-d) \\ &\leq \mathbb{P}[Y = 1, D = d] + \mathbb{E}[Y_{1-d}|D = 1-d]\mathbb{P}(D = 1-d) \\ &= \mathbb{P}(Y = 1, D = d) + \mathbb{P}(Y = 1, D = 1-d). \end{aligned}$$

Moreover, if $f_d > 0$ and $f_{1-d} = 0$, $\mathbb{P}(Y = 1, D = 1 - d) = 0$. This implies that

$$\mathbb{P}(1, d|x_1, x_0) \leq \mathbb{E}[Y_d|x_1, x_0] \leq \mathbb{P}(1, d|x_1, x_0) + \mathbb{P}(1, 1-d|x_1, x_0) 1\{\mathbb{E}[Y_{1-d}|x_1, x_0] > 0\}$$

characterizes the empirical content of the potential outcomes model $Y = Y_1 D + Y_0(1 - D)$ in all generality (i.e., without weak separability and without assumptions on the dimension of unobservable heterogeneity). It also shows that the simple binary Roy model has no empirical content relative to (f_1, f_0) beyond Assumption 4. More precisely, the identified set for (f_1, f_0) under Assumptions 9-12 is the same as under Assumption 9-11 with Roy selection $D = 1\{Y_1^* > Y_0^*\}$. Indeed, bounds (2.3) and (2.3) still hold under Assumptions 9-12. They are also sharp, since $D = 1\{Y_1^* > Y_0^*\}$ implies Assumption 12.

Corollary 2. *The identified set for* (f_1, f_0) *under Assumptions 9-12 is characterized by inequalities (2.3) and (2.4).*

In case of exclusion restrictions, an immediate corollary to Proposition 2 gives conditions for identification of the outcome equations. This identification result is related to Heckman (1990)'s identification at infinity in the following sense : in the special case of a single index model, where $f_0(x_0) = \phi(x_0\beta)$, where ϕ is a distribution function and β is a conformable vector of parameters, if $x_{0j} \to -\infty$ for some element x_{0j} of x_0 such that $\beta_j > 0$, then $f_0(x_0) \to 0$ as required.

Corollary 3 (Identification)**.** *Under Assumptions 9-12, the following hold (writing* $\omega = (z, x_1, x_0)$ *as before).*

a. If there is $x_0 \in Dom(X_0)$ *such that* $f_0(x_0) = 0$*, then* f_1 *is identified over* $Dom(X_1)$*.*

b. If there is $x_1 \in Dom(X_1)$ *such that* $f_1(x_1) = 0$*, then* f_0 *is identified over* $Dom(X_0)$*.*

a'. Take $x_1 \in Dom(X_1)$*. If there is* $x_0 \in Dom(X_0)$ *such that* $\mathbb{P}(1, 0|x_1, x_0) = 0$*, then* $f_1(x_1)$ *is identified.*

b'. Take $x_0 \in Dom(X_0)$*. If there is* $x_1 \in Dom(X_1)$ *such that* $\mathbb{P}(1, 1|x_1, x_0) = 0$*, then* $f_0(x_0)$ *is identified.*

The existence of valid instruments or exclusion restrictions is often problematic in applications of discrete choice models. However, in the Roy model of sectorial choice with sector specific unobserved heterogeneity, it is natural to expect some sector specific observed heterogeneity as well. Such sector specific observed heterogeneity would provide exclusion restrictions in the form of variables affecting outcome equation for Y_d without affecting outcome equation for Y_{1-d}. Such exclusion restrictions would yield intersection bounds in Proposition 2. Of course, even if it exists, sector specific observed heterogeneity may not satisfy a. or b. of Corollary 3. However, the availability of an exclusion restriction as in a. or b. of Corollary 3 is consistent with the spirit of a model of sector specific heterogeneity.

Bounds on the joint distribution of sector specific heterogeneity

The bounds proposed in Proposition 2 are joint sharp bounds on the structural functions, hence on treatment effects. To derive them, we treated the joint distribution of sector specific heterogeneity as a nuisance parameter. One may also be interested in the empirical content of the model relative to sector specific heterogeneity. Since the distributions of u_1 and u_0 are both normalized and assumed uniform on $[0,1]$, the joint distribution satisfies Fréchet bounds :

$$\begin{aligned}\max(f_1(x_1)+f_0(x_0)-1,0) \leq \mathbb{P}(u_1 \leq f_1(x_1), u_0 \leq f_0(x_0)|x_1,x_0) \\ \leq \min(f_1(x_1), f_0(x_0)).\end{aligned}$$

The relevant question, therefore, is whether the Roy model assumption on selection $D = 1\{Y_1^* > Y_0^*\}$ holds empirical content relative to the distribution of unobserved sector specific heterogeneity beyond Fréchet bounds. On Figure 2.1, $\mathbb{P}(Y = 1)$ is equal to the L-shaped region on the left side of the graph. The area of the left vertical band is f_1 and the area of the lower horizontal band is f_0. These two bands overlap on the lower left rectangle, whose area is equal to $\mathbb{P}(u_1 \leq f_1, u_0 \leq f_0)$. Hence $f_1 + f_0 = \mathbb{P}(Y = 1) + \mathbb{P}(u_1 \leq f_1, u_0 \leq f_0)$. Adding conditioning variables, we have the following bounds on the joint distribution of sector specific heterogeneity :

$$\mathbb{P}(u_1 \leq f_1(x_1), u_0 \leq f_0(x_0)|x_1,x_0) = f_1(x_1) + f_0(x_0) - \mathbb{P}(Y = 1|x_1,x_0). \quad (2.5)$$

This yields a sharper lower bound than the Fréchet bounds whenever $\mathbb{P}(Y = 1|x_1,x_0) < 1$. Note however, that the above constraint no longer holds when we replace the Roy selection hypothesis $D = \{Y_1^* > Y_0^*\}$ by Assumption 12. Hence the conclusion that the latter two assumptions hold the same empirical content is valid when considering empirical content relative to the structural functions and the treatment effects, but not when considering empirical content relative to the distribution of unobserved heterogeneity.

2.3.2 Extended binary Roy model

Extended selection assumption

Assumption 12 is very restrictive and recent research by Haultfoeuille and Maurel (2011) and Bayer, Khan, and Timmins (2011) on the Roy model with continuous outcomes has focused on an extended version according to the terminology of Heckman and Vytlacil (1999), where selection depends on $Y_1^* - Y_0^*$ and a function of observable variables $g(Z, X_1, X_0)$ sometimes called "non pecuniary component". We now investigate the implications of this selection assumption in the binary case.

Assumption 13 (Observable heterogeneity in selection)**.** $D = 1\{Y_1^* - Y_0^* > g(Z, X_1, X_0)\}$ *for some unknown function* g *of the vector of the observable variables* Z*,* X_1 *and* X_0*.*

Assumption 13 includes separability of the structural selection function in $Y_1^* - Y_0^*$ and $g(Z, X_1, X_0)$. The more general case without separability of the selection function is considered in Section 2.3.2. Under Assumptions 9-11 and 13, we may still characterize the empirical content of the model graphically, in Figures 2.3 and 2.4.

We drop Z, X_1 and X_0 from the notation in the discussion below. For each value of (u_1, u_0), the outcome is uniquely determined by f_1, f_0 and g. Again, the missing piece to compute the likelihood of outcomes $\mathbb{P}(i, j)$, $i, j = 1, 0$, is the copula for (u_1, u_0). From the knowledge of the probabilities of horizontal and vertical bands in the (u_1, u_0) space, we can derive the sharp bounds on structural parameters f_1, f_0 and g. Four cases are considered below to explain the bounds, which are derived formally in Proposition 3.

a. Case where $g \geq f_1$ on Figure 2.4. The probability of outcome $(Y = 1, D = 0)$ is seen to be exactly equal to the area of the lower horizontal band. Hence $f_0 = \mathbb{P}(1, 0)$ is identified. Moreover, the area of the horizontal band $(f_0, f_0 - f_1 + g)$ is smaller than the probability of outcome $(Y = 0, D = 0)$. Hence $g \leq f_1 + \mathbb{P}(0, 0)$. Similar reasoning yields $\mathbb{P}(1, 1) \leq f_1 \leq \mathbb{P}(Y = 1) + \mathbb{P}(0, 0)$.

FIGURE 2.3 – Characterization of the empirical content of the extended binary Roy model in the unit square of the (u_1, u_0) space in case $0 \leq g < f_1$.

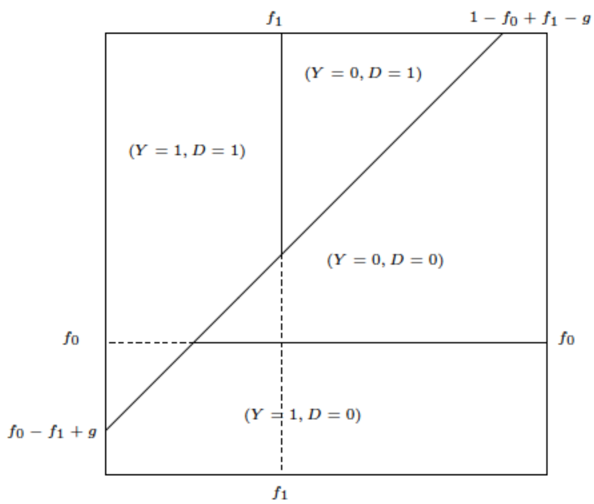
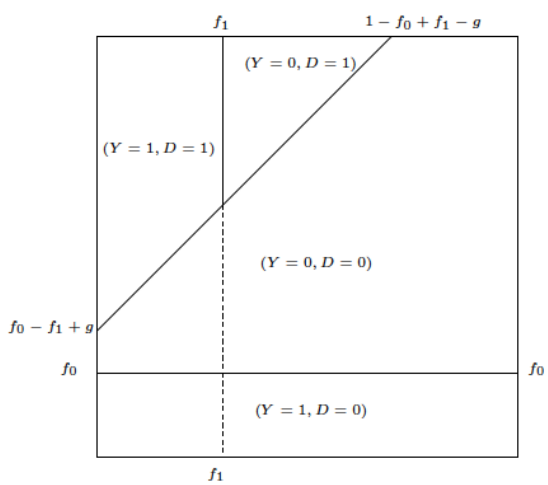


FIGURE 2.4 – Characterization of the empirical content of the extended binary Roy model in the unit square of the (u_1, u_0) space in case $g \geq f_1$.



b. Case where $0 \leq g < f_1$ on Figure 2.3. The area of the lower horizontal band $(0, f_0 - f_1 + g)$ is smaller than the probability of outcome $(Y = 1, D = 0)$. Hence $g \leq f_1 - f_0 + \mathbb{P}(1,0)$. Moreover, the area of the horizontal band $(0, f_0)$ is larger than the probability of outcome $(Y = 1, D = 0)$ and smaller than the probability of outcome $(Y = 1)$. Hence $\mathbb{P}(1,0) \leq f_0 \leq \mathbb{P}(Y = 1)$. Finally, $\mathbb{P}(1,1) \leq f_1 \leq \mathbb{P}(Y = 1) + \mathbb{P}(0,0)$ still holds.

c. Case where $-f_0 < g \leq 0$. Similarly to Case b., we obtain bounds $g \geq f_1 - f_0 + \mathbb{P}(1,1)$, $\mathbb{P}(1,0) \leq f_0 \leq \mathbb{P}(Y = 1) + \mathbb{P}(0,1)$ and $\mathbb{P}(1,1) \leq f_1 \leq \mathbb{P}(Y = 1)$.

d. Case where $g \leq -f_0$. Similarly to Case a., $f_1 = \mathbb{P}(1,1)$ is identified and $\mathbb{P}(1,0) \leq f_0 \leq \mathbb{P}(Y = 1) + \mathbb{P}(0,1)$ and $g \geq -f_0 - \mathbb{P}(0,1)$.

In addition, in both cases a. and b., where $g > f_1 - f_0$, corresponding to Figures 2.4 and 2.3, the marginal constraint on u_1 fixes the probability mass in the thin right vertical band to $f_0 - f_1 + g$. Hence the maximum probability mass that can be shifted to the left of f_1 is $p_{11} + p_{10} + p_{00} - (f_0 - f_1 + g)$, so that we have the additional constraint $f_0 \leq p_{11} + p_{10} + p_{00} - g$. Symmetrically, in case $g < f_1 - f_0$, we have the constraint $f_1 \leq g + p_{11} + p_{10} + p_{00}$. Since $g > f_1 - f_0$ also implies $f_1 \leq g + p_{11} + p_{10} + p_{00}$ and $g < f_1 - f_0$ also implies $f_0 \leq p_{11} + p_{10} + p_{00} - g$, the two constraints $f_0 \leq p_{11} + p_{00} + p_{10} - g$ and $f_1 \leq g + p_{11} + p_{10} + p_{01}$ always hold. Proposition 3 shows validity of the bounds discussed above for the triplet $(f_1(x_1), f_0(x_0), g(\omega))$.

Proposition 3 (Bounds for the extended binary Roy model)**.** *Under Assumptions 9-11 and 13, the following bounds for (f_1, f_0, g) hold (writing $\omega = (z, x_1, x_0)$ as before).*

$$
\begin{aligned}
f_1(x_1) \;\in\; & \Big[\sup_{z,x_0} \mathbb{P}(1,1|\omega), \inf_{z,x_0} [\mathbb{P}(1,1|\omega) + \mathbb{P}(0,0|\omega)1\{g(\omega) > 0\} \\
& + \min[\min(\mathbb{P}(1,0|\omega), f_0(x_0) + g(\omega))1\{g(\omega) > -f_0(x_0)\}), \\
& g(\omega) + \mathbb{P}(1,0|\omega) + \mathbb{P}(0,1|\omega)]]\Big], \\
f_0(x_0) \;\in\; & \Big[\sup_{z,x_1} \mathbb{P}(1,0|\omega), \inf_{z,x_1} [\mathbb{P}(1,0|\omega) + \mathbb{P}(0,1|\omega)1\{g(\omega) < 0\} \\
& + \min[\min(\mathbb{P}(1,1|\omega), f_1(x_1) - g(\omega))1\{g(\omega) < f_1(x_1)\}), \\
& \mathbb{P}(1,1|\omega) + \mathbb{P}(0,0|\omega) - g(\omega)]]\Big]
\end{aligned}
\quad (2.6)
$$

and

$$g(\omega) \in \Big(\Big[-f_0(x_0) - \mathbb{P}(0,1|\omega), -f_0(x_0)\Big] \cup \Big[f_1(x_1) - f_0(x_0) - \mathbb{P}(1,1|\omega), f_1(x_1) - f_0(x_0) + \mathbb{P}(1,0|\omega)\Big] \cup \Big[f_1(x_1), f_1(x_1) + \mathbb{P}(0,0|\omega)\Big]\Big) \cap \Big[f_1(x_1) - \mathbb{P}(1,1|\omega) - \mathbb{P}(1,0|\omega) - \mathbb{P}(0,1|\omega), \mathbb{P}(1,1|\omega) + \mathbb{P}(1,0|\omega) + \mathbb{P}(0,0|\omega) - f_0(x_0)\Big] \tag{2.7}$$

where the infima and suprema are taken over the domain of Z, X_1 or X_0 as indicated and when they arise.

Identification implications of exclusion restrictions

Simple identification conditions can be derived for f_1 and f_0 from the bounds of Proposition 3 under exclusion restrictions. However, it can be seen immediately that exclusion restrictions cannot identify $g(\,)$, since it would require $\mathbb{P}(Y = 1, D = 1|\omega)$, $\mathbb{P}(Y = 0, D = 1|\omega)$, $\mathbb{P}(Y = 1, D = 0|\omega)$ and $\mathbb{P}(Y = 0, D = 0|\omega)$ to simultaneously equal zero.

Corollary 4 (Identification)**.** *Under Assumptions 9-11 and 13, the following hold (writing $\omega = (z, x_1, x_0)$ as before).*

a. If there is $z \in Dom(Z)$ and $x_0 \in Dom(X_0)$ such that $g(\omega) \leq -f_0(x_0)$, then $f_1(x_1) = \mathbb{P}(1,1|\omega)$ is identified.

b. If there is $z \in Dom(Z)$ and $x_1 \in Dom(X_1)$ such that $g(\omega) \geq f_1(x_1)$, then $f_0(x_0) = \mathbb{P}(1,0|\omega)$ is identified.

a'. Take $x_1 \in Dom(X_1)$. If there is $x_0 \in Dom(X_0)$ or $z \in Dom(Z)$ such that $\mathbb{P}(1,0|\omega) = \mathbb{P}(0,0|\omega) = 0$, then $f_1(x_1)$ is identified.

b'. Take $x_0 \in Dom(X_0)$. If there is $x_1 \in Dom(X_1)$ or $z \in Dom(Z)$ such that $\mathbb{P}(1,1|\omega) = \mathbb{P}(0,1|\omega) = 0$, then $f_0(x_0)$ is identified.

Sharp bounds in the extended Roy model

When the object of interest is treatment parameters only, the three dimensional identification region defined by the sharp bounds on (f_1, f_0, g) is projected on the two-dimensional space (f_1, f_0) as follows.

Proposition 4 (Sharp bounds for the extended Roy model)**.** *Under Assumptions 9-11 13, the identified set for (f_1, f_0) is characterized by the following bounds, where λ takes the values 1 or 0 and $\varepsilon > 0$ is arbitrarily small :*

$$\begin{array}{lllll} \sup_{z,x_0} \mathbb{P}(1,1|\omega) & \leq & f_1(x_1) & \leq & \inf_{z,x_0} \left[\mathbb{P}(1,1|\omega) + \mathbb{P}(1,0|\omega) + \lambda \max(0, \mathbb{P}(0,0|\omega) - \varepsilon)\right] \\ \sup_{z,x_1} \mathbb{P}(1,0|\omega) & \leq & f_0(x_0) & \leq & \inf_{z,x_1} \left[\mathbb{P}(1,0|\omega) + \mathbb{P}(1,1|\omega) + (1-\lambda) \max(0, \mathbb{P}(0,1|\omega) - \varepsilon)\right] \end{array} \quad (2.8)$$

The binary λ ensures joint sharpness of the bounds on f_1 and f_0. It reflects the fact that in Figure 2.3 the diagonal separating the $D = 1$ from the $D = 0$ regions is either on the right of the point (f_1, f_0), in which case the sharper bound on f_1 holds, or on the left of point (f_1, f_0), in which case the sharper bound on f_0 holds, but not both at the same time. The presence of $\varepsilon > 0$ in the bounds reflects the fact that if, as in Figure 2.3, the diagonal is on the right of the point (f_1, f_0), i.e., $\lambda = 1$, then f_1 could only attain the upper bound $p_{11} + p_{10} + p_{00}$ if all the mass corresponding to Region $(Y = 0, D = 0)$ was shifted into the triangle below the diagonal, above f_0 and left of f_1. However, this is impossible since the vertical band on the right has non zero mass by the uniform marginal constraint on u_1. Hence, the presence of ε in the bounds is due to the linearity of the boundary between Regions $D = 0$ and $D = 1$. This explains why it disappears in the nonseparable case of the next section.

If the object of interest is the non pecuniary component g, the three dimensional identification region is projected on the one-dimensional space for g into the single interval $[-\mathbb{P}(1,1|\omega) - \mathbb{P}(1,0|\omega), \mathbb{P}(1,1|\omega) + \mathbb{P}(1,0|\omega)]$, since the bounds in (2.6) cross at those values. In the presence of instruments (or exclusion restrictions), the projections on (f_1, f_0) and on g can be much tighter and the projection on (f_1, f_0) may even be reduced to a point, as in Corollary 4.

Testing the Roy selection assumption

As we have just seen, in the absence of exclusion restrictions, the identified region always contains the hyperplane $g = 0$, so that it is impossible to test the classical Roy selection hypothesis. However, in the presence of exclusion restrictions, the hypothesis $g(\omega) = 0$ may become testable. There is a non zero non pecuniary component in the selection equation if the hyperplane $g(\omega) = 0$ does not intersect the three dimensional identification region for $(f_1(x_1), f_0(x_0), g(\omega))$ defined by the bounds in Proposition 3. This implies the crossing of the intersection bounds in Proposition 2, in the sense that

$$\sup_{x_0,z} \mathbb{P}(1,1|x_1,x_0,z) \quad > \quad \inf_{x_0,z} \Big[\mathbb{P}(1,1|\omega) + \mathbb{P}(1,0|\omega) 1\{f_0(x_0) > 0\}\Big]$$

or

$$\sup_{x_1,z} \mathbb{P}(1,0|\omega) \quad > \quad \inf_{x_1,z}\Big[\mathbb{P}(1,0|\omega) + \mathbb{P}(1,1|\omega)1\{f_1(x_1) > 0\}\Big]$$

so that by Proposition 2, the simple Roy model is rejected. In practice, the test for the existence of a non pecuniary component would be carried out by constructing a confidence region according to the methods proposed in Chernozhukov, Lee, and Rosen (2009) or Andrews and Shi (2011) and checking, whether the hyperplane $g(\omega) = 0$ intersects the confidence region. If it does, we fail to reject the hypothesis of existence of a non pecuniary component and if it doesn't, we reject the hypothesis at significance level equal to 1 minus the confidence level chosen for the confidence region. The hypotheses $g \geq 0$ or $g \leq 0$ may be tested in the same way.

Sharp bounds without separability of the selection function

The same arguments can be applied to derive the empirical content of the model where the selection equation generalizes Assumption 13 with the following.

Assumption 14 (Nonseparable selection function)**.** *Suppose the selection rule is $D = 1\{u_0 > h(u_1, W)\}$ and h strictly increasing in u_1, for all W.*

Assumption 13 is a special case of Assumption 14, where $h(u_1, W) = u_1 + f_0(X_0) - f_1(X_1) + g(W)$.

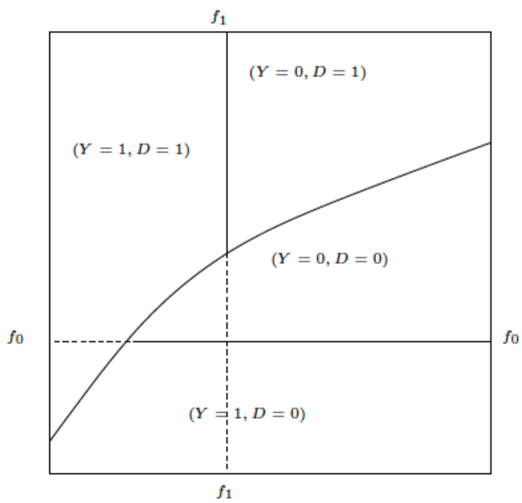
The identified region for the pair $(f_1(x_1), f_0(x_0))$ is obtained in the same way as the separable case except that f_1 attains $\mathbb{P}(1,1) + \mathbb{P}(1,0) + \mathbb{P}(0,0)$ and f_0 attains $\mathbb{P}(1,1)+\mathbb{P}(1,0)+\mathbb{P}(0,0)$. This occurs because the nonlinearity of the curve separating region $D = 1$ from region $D = 0$ allows all the mass corresponding to $\mathbb{P}(0,0)$ to be shifted on the left of f_1, as in Figure 2.5.

Proposition 5 (Sharp bounds for the extended Roy model without separability)**.** *Under Assumptions 9-11 and 14, the identified set for (f_1, f_0) is characterized by the following inequalities, where λ takes the values* 1 *or* 0 *:*

$$\begin{aligned} \sup\nolimits_{z,x_0} \mathbb{P}(1,1|\omega) \quad &\leq \quad f_1(x_1) \quad \leq \quad \inf\nolimits_{z,x_0} [\mathbb{P}(1,1|\omega) + \mathbb{P}(1,0|\omega) + \lambda\mathbb{P}(0,0|\omega)] \\ \sup\nolimits_{z,x_1} \mathbb{P}(1,0|\omega) \quad &\leq \quad f_0(x_0) \quad \leq \quad \inf\nolimits_{z,x_1} [\mathbb{P}(1,0|\omega) + \mathbb{P}(1,1|\omega) + (1-\lambda)\mathbb{P}(0,1|\omega)] \end{aligned} \tag{2.9}$$

In this context, however, the Roy selection assumption $D = 1\{Y_1^* > Y_0^*\}$ may not be tested with the strategy developed above.

FIGURE 2.5 – Characterization of the empirical content of the binary Roy model in the unit square of the (u_1, u_0) space without separability of the selection function.



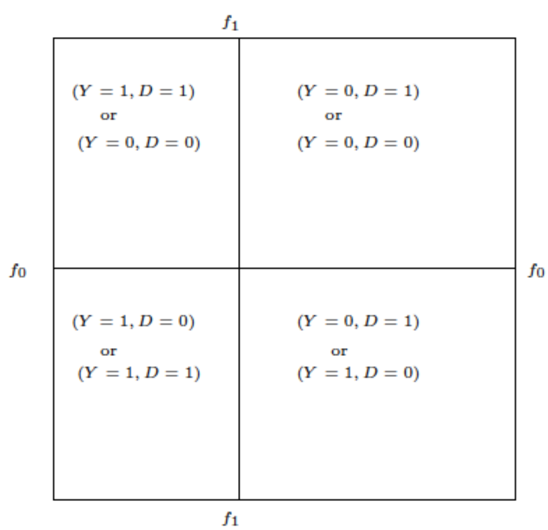
2.4 Sharp bounds for the generalized binary Roy model

So far, we have assumed that selection occurs on the basis of success probability and other observable variables. We now turn to the general case, where unobservable heterogeneity, beyond $u_0 - u_1$, may play a role in sectorial selection. Knowledge of (u_1, u_0) now no longer uniquely determines the outcome $(Y = i, D = j)$ as seen on Figure 2.6.

Multiplicity of equilibria and lack of coherence of the model can be dealt with, however, with the optimal transportation approach of Galichon and Henry (2011) (or equivalently with the random set approach of Beresteanu, Molchanov, and Molinari (2011) as in Chesher, Rosen, and Smolinski (2011)), as shown in the proof of Theorem 2 below.

Theorem 2 (Sharp bounds for the generalized Roy model)**.** *Under Assumption 9-11, the empirical content of the model is characterized by inequali-*

FIGURE 2.6 – Characterization of the empirical content of the generalized binary Roy model in the unit square of the (u_1, u_0) space.



ties (2.10)-(2.12) below (writing $\omega = (z, x_1, x_0)$ as before).

$$\sup_{z,x_0} \mathbb{P}(1,1|\omega) \leq f_1(x_1) \leq 1 - \sup_{z,x_0} \mathbb{P}(0,1|\omega), \tag{2.10}$$

$$\sup_{z,x_1} \mathbb{P}(1,0|\omega) \leq f_0(x_0) \leq 1 - \sup_{z,x_1} \mathbb{P}(0,0|\omega) \tag{2.11}$$

and

$$\begin{aligned} & \sup_z \max\Big(0, f_0(x_0) - \mathbb{P}(1,0|\omega) - \mathbb{P}(0,1|\omega), f_1(x_1) - \mathbb{P}(1,1|\omega) - \mathbb{P}(0,0|\omega)\Big) \\ \leq\ & \mathbb{P}(u_1 \leq f_1(x_1), u_0 \leq f_0(x_0)|x_1, x_0) \qquad (2.12) \\ \leq\ & \inf_z \min\Big(\mathbb{P}(Y=1|\omega), f_1(x_1) + f_0(x_0) - \mathbb{P}(Y=1|\omega)\Big). \end{aligned}$$

Theorem 2 is not an operational characterization of the empirical content of the model since the sharp bounds involve the unknown quantity $\mathbb{P}(u_1 \leq f_1(x_1), u_0 \leq f_0(x_0)|x_1, x_0)$, which, by the normalization of Assumption 10, is exactly the copula of (u_1, u_0). In the case of total ignorance about the copula of (u_1, u_0), after plugging Fréchet bounds $\max(f_1(x_1)+f_0(x_0)-1, 0) \leq \mathbb{P}(u_1 \leq f_1(x_1), u_0 \leq f_0(x_0)|x_1, x_0) \leq \min(f_1(x_1), f_0(x_0))$, inequalities (2.12) are shown to be redundant. Hence we have the following.

Corollary 5. *The identified set for (f_1, f_0) under Assumption 9-11 is characterized by inequalities (2.10) and (2.11).*

In order to sharpen those bounds, we may consider restrictions on the copula for (u_1, u_0) or restrictions on the selection equation. We consider both strategies in turn.

2.4.1 Restrictions on selection

Consider the following selection model, where selection depends on $Y_1^* - Y_0^*$ and $g(Z, X_1, X_0)$ and selection specific unobserved heterogeneity v, which is separable and which is independent of (resp. dependent on) sector specific unobserved heterogeneity (u_1, u_0) under Assumption 15 (resp. Assumption 16). As before, write $W = (Z, X_1, X_0)$.

Assumption 15. $D = 1\{Y_1^* - Y_0^* > g(W) + v\}$*, with* $v \perp\!\!\!\perp (u_1, u_0, W)$ *and* $\mathbb{E}v = 0$ *(without loss of generality).*

With $v \perp\!\!\!\perp (u_1, u_0)$, we have $\mathbb{P}(u_d \leq g(z,x_1,x_0)+v+f_1(x_1)-f_0(x_0)|z,x_1,x_0) = \mathbb{E}_v\mathbb{E}[1\{u_d \leq g(z,x_1,x_0)+v-f_1(x_1)+f_0(x_0)\}|z,x_1,x_0,v] = \max(0, g(z,x_1,x_0)-f_1(x_1)+f_0(x_0))$ and it is shown in Corollary 6 that the bounds on $g(\,)$ derived in Section 2.3 remain valid.

Corollary 6. *Under assumptions 9-11 and 15, (2.7) holds.*

As for the bounds on (f_1, f_0), (2.6) remain valid under specific domain restrictions for v.

Assumption 16. $D = 1\{Y_1^* - Y_0^* > g(W) + v\}$, *with* $v \perp\!\!\!\perp W$, $\mathbb{E}v = 0$ *(without loss of generality).*

Note that Assumption 16 is equivalent to assuming the selection equation $D = 1\{h(W) > \eta\}$ with η arbitrarily dependant on (u_1, u_0). Indeed, one can take $h(W) = f_1(X_1) - f_0(X_0) - g(W)$ and $\eta = v + u_1 - u_0$.

Corollary 7. *Under Assumption 9-11 and 16, (2.10) and (2.11) are sharp bounds for the pair* (f_1, f_2).

From Corollary 7, we conclude that the separability of the selection specific unobserved heterogeneity term has no empirical content, in the sense that the identified set for (f_1, f_0) is identical to the case, where there is no information on selection. This is related to the lack of empirical content of LATE in Kitagawa (2009) and it is in sharp contrast with the case of no sector specific heterogeneity in Shaikh and Vytlacil (2011), Jun, Pinkse, and Xu (2010) and Mourifié (2011), where the ordering between f_1 and f_0 can be used as identifying information. Indeed, if $f_1 \leq f_0$, we have $f_1 = \mathbb{P}(Y = 1, D = 1) + \mathbb{P}(u_1 \leq f_1, D = 0) \leq \mathbb{P}(Y = 1, D = 1) + \mathbb{P}(u_1 \leq f_0, D = 0)$. The last term is equal to $\mathbb{P}(Y = 1, D = 0)$ if $u_1 = u_0$ but is not identified in the case with sector specific unobserved heterogeneity.

2.4.2 Restrictions on the joint distribution of sector specific heterogeneity

Parametric restrictions on the copula

In case the copula for (u_1, u_0) is parameterized with parameter vector θ, sharp bounds are obtained straightforwardly by replacing $\mathbb{P}(u_1 \leq f_1(x_1), u_0 \leq f_0(x_0)|x_1, x_0)$ with the parametric version $F(f_1(x_1), f_0(x_0); \theta)$ in (2.12).

Perfect correlation

In the case of perfect correlation between the two sector specific unobserved heterogeneity variables, $\mathbb{P}(u_1 \leq f_1(x_0), u_0 \leq f_0(x_0)) = \min(f_1(x_1), f_0(x_0))$ so that the sharp bounds of Theorem 2 specialize to (2.10), (2.11), $\min(f_1(x_1), f_0(x_0)) \leq \inf_z \mathbb{P}(Y = 1|z, x_1, x_0)$ and $\sup_z \mathbb{P}(Y = 1|z, x_1, x_0) \leq \max(f_1(x_1), f_0(x_0))$, which are the bounds derived in Chiburis (2010).

Independence

In the special case, where the two sector specific errors are independent of each other $u_1 \perp\!\!\!\perp u_0$, sharp bounds can be derived from Theorem 2 and $\mathbb{P}(u_1 \leq f_1(x_0), u_0 \leq f_0(x_0)) = \mathbb{P}(u_1 \leq f_1(x_1))\mathbb{P}(u_0 \leq f_0(x_0)) = f_1(x_1)f_0(x_0)$. The sharp bounds obtained allow formal tests of the hypothesis of independence of the two unobserved heterogeneity components. This would not be achievable based only on Fréchet bounds (as noted by Tsiatis (1975) in the case of competing risks), as we always have $f_0 + f_1 - 1 \leq f_0 f_1 \leq \min(f_1, f_0)$ when $0 \leq f_1, f_0 \leq 1$.

Factor structure

Theorem 2 also allows us to characterize the empirical content of the factor model for sector specific unobserved heterogeneity proposed in Aakvik, Heckman, and Vytlacil (2005).

Assumption 17 (Factor model)**.** *Sector specific unobserved heterogeneity has factor structure $u_d = \alpha_d u + \eta_d$, $d = 1, 0$, with $\mathbb{E}u = 0$, $\mathbb{E}u^2 = 1$ (without loss of generality) and $\eta_1 \perp\!\!\!\perp \eta_0 | u$. η_d is uniformly distributed on $[0, 1]$ for $d = 1, 0$, conditionally on u.*

This factor specification for sector specific unobserved heterogeneity is particularly appealing in applications to the effects of employment programs. Success in securing a job depends on common unobservable heterogeneity in talent and motivation and sector specific noise. Under Assumptions 9, 11 and 17, we still have $\mathbb{E}[Y_d|z, x_1, x_0] = f_d(x_d)$ and

$$\begin{aligned}
\mathbb{P}(u_1 \leq f_1(x_1), u_0 \leq f_0(x_0)|x_1, x_0) &= \mathbb{E}_u \mathbb{P}(\eta_1 \leq f_1(x_1) - \alpha_1 u, \\
&\qquad \eta_0 \leq f_0(x_0) - \alpha_0 u | x_1, x_0, u) \\
&= \mathbb{E}_u \mathbb{P}(\eta_1 \leq f_1(x_1) - \alpha_1 u | x_1, u) \\
&\qquad \mathbb{P}(\eta_0 \leq f_0(x_0) - \alpha_0 u | x_1, x_0, u) \\
&= f_1(x_1) f_0(x_0) + \alpha_1 \alpha_0.
\end{aligned}$$

Hence we can obtain sharp bounds on parameters f_1, f_0, α_1 and α_0 as follows.

Corollary 8 (Sharp bounds for the factor model)**.** *Under Assumptions 9, 11 and 17, the empirical content of the model is characterized by (2.10), (2.11) and (writing $\omega = (z, x_1, x_0)$ as before)*

$$\begin{aligned} & \sup_z \max\Big(0, f_0(x_0) - \mathbb{P}(1,0|\omega) - \mathbb{P}(0,1|\omega), f_1(x_1) - \mathbb{P}(1,1|\omega) - \mathbb{P}(0,0|\omega)\Big) \\ \leq \;& f_1(x_1) f_0(x_0) + \alpha_1\alpha_0 \\ \leq \;& \inf_z \min\Big(\mathbb{P}(Y=1|\omega), f_1(x_1) + f_0(x_0) - \mathbb{P}(Y=1|\omega)\Big) \end{aligned}$$

We recover the case of independent sector specific heterogeneity variables, when $\alpha_1 = \alpha_0 = 0$.

2.5 Conclusion

We have derived sharp bounds in the simple, extended and generalized binary Roy models, including a factor specification proposed by Aakvik, Heckman, and Vytlacil (2005). The bounds are simple enough to lend themselves to existing inference methods for intersection bounds as in Chernozhukov, Lee, and Rosen (2009) and Andrews and Shi (2011). The methods introduced here can be applied to the derivation of nonparametric sharp bounds for the Tobit version of the Roy model as well as in other binary models with several unobserved heterogeneity dimensions, such as entry and participation games.

Chapitre 3

Nonparametric sharp bounds for payoffs in 2×2 games.

3.1 Introduction

The empirical analysis of full information game theoretic models has emerged as a leading way to learn about strategic interactions between economic agents and to estimate, for example, the extent of monopoly advantage in imperfect competitive environments or free riding incentives in cooperation settings. Beside the numerous applications in industrial organization, as evidenced by the recent survey in Bajari, Hong, and Nekipelov (2012), areas of impact include labor economics, as in Bjorn and Vuong (1984) and Kooreman (1994), social interactions, as in Soetevent and Kooreman (2007), family economics, as in Engers and Stern (2002), or development economics, as in Méango (2012). The empirical approach to models of multiperson simultaneous decisions goes back at least to Bjorn and Vuong (1984) and was popularized in the field of industrial organization by Bresnahan and Reiss (1990, 1991) and Berry (1992) among others. In those cases, attention was restricted to specific parametric utilities or profits and unobserved heterogeneity types. Coherency of the model, in the sense of Heckman (1978) and Gourié-

This chapter is a joint work with Marc Henry. This chapter was conducted in part, while Marc Henry was visiting Sciences-Po and Polytechnique and I was visiting Penn State. We thank our respective hosts for their hospitality and support.

roux, Laffont, and Monfort (1980), was obtained by removing multiplicity of predicted outcomes in the game Bjorn and Vuong (1984) assume an ad-hoc uniform equilibrium selection device, whereas Bresnahan and Reiss (1991) coarsen the outcome space). The multiplicity issue was addressed head-on by Jovanovic (1989) and Tamer (2003) and both Galichon and Henry (2006, 2011) and Beresteanu, Molchanov, and Molinari (2008, 2011) propose characterizations of the empirical content of Nash equilibrium play in models with simultaneous decisions by multiple agents, while retaining the parametric framework for payoffs and unobserved types. Much of the empirical content in the latter characterizations, however, rests on the specific parametric assumptions maintained, some of which may be structurally motivated, but others, especially parametric assumptions on unobserved type distributions, are entirely ad-hoc. Kline and Tamer (2012) seem to be the first to remove parametric assumptions and consider sharp bounds in full information games, but their focus, however, is best response functions, which may be of interest in their own right, but which are not the focus of the literature, generally interested in recovering payoff functions (utilities and profits). Aradillas-Lopez (2011) considers nonparametric bounds on predicted probabilities of strategy profiles under asymmetric information. Neither considers nonparametric sharp bounds on payoff functions in full information games as we do here.

Within the class of two person games with binary strategies in full information, we consider the identification problem, where the distribution of realized decisions is known by the analyst, who assumes that such realizations emerge from Nash equilibrium play (in pure or mixed strategies) in the game. Hence we adopt a pure revealed preference approach to the model of interaction and analyze the empirical content of maximizing behavior as in Henry and Mourifié (2012), with the additional complication that the dummy endogenous variable is the result of a simultaneous decision by a second agent. Based on the characterization of the empirical content of games with Shapley regular core in Galichon and Henry (2011), we derive sharp nonparametric bounds on payoffs and unobserved heterogeneity distributions. Additional constraints on the order of payoffs, to consider games of complements or games of susbstitutes, and on type distributions, to evaluate shape and other distributional restrictions, can be easily added to see how they shrink the identified region. One of the main arguments for allowing agents to randomize in the empirical analysis of games, as in Bajari, Hong, and Ryan (2010), Beresteanu, Molchanov, and Molinari (2008, 2011), Bajari, Hahn, Hong, and Ridder (2011) and Galichon and Henry (2011), is almost sure existence of equilibrium in mixed strategies, whereas existence of equilibrium in pure strategies only is not garanteed. This argument is only relevant in case of parametric assumptions on the unobserved heterogeneity (or type) distribution, but fails to sway in the framework entertained here, as regions of the type space may well have zero probability. We therefore analyze the implications of restricting play to pure strategies and derive sharp bounds on payoffs and type distributions in that case too. Considering type distributions as nuisance infinite dimensional parameters and projecting the identified region allows us then to derive sharp nonparametric bounds on the payoff functions themselves. We find that the hypothesis of Nash equilibrium play is not falsifiable in this framework, as the identified region is never empty. Rejection of the model becomes possible under the assumption of an exclusion restriction, namely variation in the payoff of a player that leaves the opponent's profit unchanged. In the latter case, the bounds become intersection bounds, as in Chernozhukov, Lee, and Rosen (2009) and they can cross. We also find that, without additional prior information, we cannot identify, whether the game is of complements or substitutes. However, we obtain non trivial sharp bounds on monopoly advantage and free-riding incentives, when they arise.

The remainder of the paper is organized as follows. Section 3.2 derives the analytical framework, the games analyzed, their equilibrium correspondences and the objects of interest. Section 3.3 derives joint sharp bounds for payoff functions and type distributions, treating the equilibrium selection mechanism as a nuisance parameter. Section 3.4 considers implications of pure strategy play and derives the projection of the identified set to obtain sharp bounds for the payoff functions. Sharp bounds are also given for monopoly advantage and free riding incentives. The last section concludes.

3.2 Analytical framework

We shall be concerned with the following econometric model.

$$Y_i = 1\{\Pi_i(Y_{3-i}, X_i) > \varepsilon_i\} \text{ and } \varepsilon_i \sim U[0,1],\ i = 1,2, \tag{3.1}$$

where $1\{A\} = 1$ if A is true and zero otherwise, $Y = (Y_1, Y_2)$ is a pair of observed binary outcome variables, $\Pi = (\Pi_1, \Pi_2)$ are unknown functions of Y_{3-i}, observable random vectors $X = (X_1, X_2)$ and unobservable random variables $\varepsilon = (\varepsilon_1, \varepsilon_2)$. We assume that the only source of endogeneity in the econometric model is the interaction between players and the simultaneous choice. Hence, we assume that observable heterogeneity variables are exogenous.

Assumption 18 (Exogeneity)**.** *The following exogeneity assumption holds :* $(X_1, X_2) \perp\!\!\!\perp (\varepsilon_1, \varepsilon_2)$ *and for ease of notation, we shall drop all components that are common to* X_1 *and* X_2 *and relabel* X_i *as the vector of observable heterogeneity variables (if they arise) that affect* Π_i *but are excluded from* Π_{3-i}.

We give two structural interpretation of this model within the range of noncooperative games of perfect information with 2 players and 2 strategies each.

3.2.1 General 2×2 games

In a first structural interpretation of Model (3.1) we consider general 2×2 games of perfect information with payoff structure given in Table 3.1, which is common knowledge to the two players. Working under assumptions that rule out ties, the best response of Player 1 to $Y_2 = 1$ is $Y_1 = 1$ if $\tilde{\Pi}_1(1, 1, X_1) - \tilde{\Pi}_1(0, 1, X_1) > [\tilde{\varepsilon}_1(0, 1) - \tilde{\varepsilon}_1(1, 1)]$ and zero otherwise, whereas the best response to $Y_2 = 0$ is $Y_1 = 1$ if $\tilde{\Pi}_1(1, 0, X_1) - \tilde{\Pi}_1(0, 0, X_1) > [\tilde{\varepsilon}_1(0, 0) - \tilde{\varepsilon}_1(1, 0)]$ and zero otherwise. Best responses for Player 2 are obtained symmetrically. Assuming that the unobserved heterogeneity differences $\tilde{\varepsilon}_1(1, Y_2) - \tilde{\varepsilon}_1(0, Y_2)$ and $\tilde{\varepsilon}_2(Y_1, 1) - \tilde{\varepsilon}_2(Y_1, 0)$ are independent of the opponent's action and are absolutely continuous with respect to Lebesgue measure and setting $\Pi_1(Y_2, X_1) = \tilde{\Pi}_1(1, Y_2, X_1) - \tilde{\Pi}_1(0, Y_2, X_1)$, $\Pi_2(Y_1, X_2) = \tilde{\Pi}_2(Y_1, 1, X_2) - \tilde{\Pi}_2(Y_1, 0, X_2)$, $\varepsilon_1 = -\tilde{\varepsilon}_1(1, Y_2) + \varepsilon_1(0, Y_2)$ and $\varepsilon_2 = -\tilde{\varepsilon}_2(Y_1, 1) + \varepsilon_2(Y_1, 0)$, we obtain Model (3.1), where $\varepsilon_i \sim U[0, 1]$ is without loss of generality.

3.2.2 Participation games

In a second structural interpretation of Model (3.1), we consider the special case of 2×2 participation games, where a player's payoff when she chooses not to participate is independent of the opponent's behavior and can therefore be normalized to zero. Each player has 2 strategies and 3 different payoffs. For each player, the 3 different payoffs can be ranked in 3! distinct

TABLE 3.1 – Payoff structure of 2×2 games.

	1	0
1	$\tilde{\Pi}_1(1,1,X_1)+\tilde{\varepsilon}_1(1,1),\tilde{\Pi}_2(1,1,X_2)+\tilde{\varepsilon}_2(1,1)$	$\tilde{\Pi}_1(1,0,X_1)+\tilde{\varepsilon}_1(1,0),\tilde{\Pi}_2(1,0,X_2)+\tilde{\varepsilon}_2(1,0)$
0	$\tilde{\Pi}_1(0,1,X_1)+\tilde{\varepsilon}_1(0,1),\tilde{\Pi}_2(0,1,X_2)+\tilde{\varepsilon}_2(0,1)$	$\tilde{\Pi}_1(0,0,X_1)+\tilde{\varepsilon}_1(0,0),\tilde{\Pi}_2(0,0,X_2)+\tilde{\varepsilon}_2(0,0)$

TABLE 3.2 – Payoff structure of 2×2 participation games.

	1	0
1	$\Pi_1(1,X_1,\varepsilon_1),\Pi_2(1,X_2,\varepsilon_2)$	$\Pi_1(0,X_1,\varepsilon_1),0$
0	$0,\Pi_2(0,X_2,\varepsilon_2)$	$0,0$

ways. Hence there are 36 classes of ordinally equivalent such 2×2 participation games (but only 7 strategically distinct classes of games as we shall see). The payoff structure as in Table 3.2, which is common knowledge to the two players.

Assuming that the profit functions are weakly separable in ε_i, $i = 1, 2$, and the latter are absolutely continuous with respect to Lebesgue measure, the game can be summarized by Model (3.1) without loss of generality (see Vytlacil (2002)).

3.2.3 Implications of each structural interpretation

Depending on the chosen structural interpretation, the analyst will be able to answer different empirical questions. Two questions of particular relevance in 2×2 game theoretic modeling of economic interactions are the price of competition and the extent of free riding incentives. 2×2 games are applied to the empirical analysis of imperfect competition since at least

Bresnahan and Reiss (1990) and Berry (1992). Two questions of particular interest arise : whether the two players (firms) are complements or substitutes and the extent of the monopoly advantage if they are substitutes. Both questions can be answered (partially) if the quantities

$$\tilde{\Pi}_1(1,0,X_1) + \tilde{\varepsilon}_1(1,0) - [\tilde{\Pi}_1(1,1,X_1) + \tilde{\varepsilon}_1(1,1)]$$

$$\tilde{\Pi}_2(0,1,X_2) + \tilde{\varepsilon}_2(0,1) - [\tilde{\Pi}_2(1,1,X_2) + \tilde{\varepsilon}_2(1,1)]$$

are (partially) identified. Now, with the structural interpretation of participation games in Section 3.2.2, we have

$$\begin{aligned}
\tilde{\Pi}_1(1,0,X_1) + \tilde{\varepsilon}_1(1,0) - [\tilde{\Pi}_1(1,1,X_1) + \tilde{\varepsilon}_1(1,1)] &= \Pi_1(0,X_1) - \Pi_1(1,X_1) \\
\tilde{\Pi}_2(0,1,X_2) + \tilde{\varepsilon}_2(0,1) - [\tilde{\Pi}_2(1,1,X_2) + \tilde{\varepsilon}_2(1,1)] &= \Pi_2(0,X_2) - \Pi_2(1,X_2)
\end{aligned}$$

and we shall derive sharp bounds on $\Pi = (\Pi_1(1,X_1), \Pi_1(0,X_1), \Pi_2(1,X_2), \Pi_2(0,X_2)$. 2×2 games are also used to model the provision of public goods. In that context, the extent of free riding incentives is of particular empirical relevance and it is measured by the following quantities.

$$\tilde{\Pi}_1(0,1,X_1) + \tilde{\varepsilon}_1(0,1) - [\tilde{\Pi}_1(1,1,X_1) + \tilde{\varepsilon}_1(1,1)]$$

$$\tilde{\Pi}_2(1,0,X_2) + \tilde{\varepsilon}_2(1,0) - [\tilde{\Pi}_2(1,1,X_2) + \tilde{\varepsilon}_2(1,1)].$$

Under both the structural interpretations of Sections 3.2.1 and 3.2.2, we have the following.

$$\begin{aligned}
\tilde{\Pi}_1(0,1,X_1) + \tilde{\varepsilon}_1(0,1) - [\tilde{\Pi}_1(1,1,X_1) + \tilde{\varepsilon}_1(1,1)] &= \varepsilon_1 - \Pi_1(1) \\
\tilde{\Pi}_2(1,0,X_2) + \tilde{\varepsilon}_2(1,0) - [\tilde{\Pi}_2(1,1,X_2) + \tilde{\varepsilon}_2(1,1)] &= \varepsilon_2 - \Pi_2(1)
\end{aligned}$$

and we shall derive sharp bounds on $\Pi_1(1)$ and $\Pi_2(1)$.

3.2.4 Equilibrium

We assume, as is customary, that players choose the strategy that maximizes their payoff in pure or mixed strategy Nash equilibrium (see Aradillas-Lopez and Tamer (2008) for some discussion of the empirical content of other notions of rationality in games). We distinguish four cases, according to the ordering between $\Pi_i(1,X_i)$ and $\Pi_i(0,X_i)$.

1. Duopoly entry game : $\Pi_i(1, X_i) \leq \Pi_i(0, X_i)$, $i = 1, 2$.
2. Coordination game : $\Pi_i(0, X_i) \leq \Pi_i(1, X_i)$, $i = 1, 2$.
3. Asymmetric game 1 : $\Pi_1(0, X_1) \leq \Pi_1(1, X_1)$ and $\Pi_2(1, X_2) \leq \Pi_2(0, X_2)$.
4. Asymmetric game 2 : $\Pi_1(1, X_1) \leq \Pi_1(0, X_1)$ and $\Pi_2(0, X_2) \leq \Pi_2(1, X_2)$.

In each case, the equilibrium correspondence is represented on the unit square as a function of the pair $(\varepsilon_1, \varepsilon_2)$ in Figures 3.1-3.3 and Case (4) can be obtained from Case (3) by permuting the two players.

Definition 1 (Equilibrium correspondence)**.** *The equilibrium correspondence, denoted $G(\varepsilon, X, \Pi)$, is the set of equilibria of the game for a given values of (ε, X, Π). It is a subset of the simplex on $\{(1,1),(1,0),(0,1),(0,0)\}$ and its elements are non degenerate probabilities in case the equilibrium is in mixed strategies and degenerate probabilities in case the equilibrium is in pure strategies.*

The equilibrium has similar features in the duopoly, coordination and asymmetric games. When

$$\varepsilon \notin [\min(\Pi_1(1, X_1), \Pi_1(0, X_1)), \max(\Pi_1(1, X_1), \Pi_1(0, X_1))]$$

$$\times[\min(\Pi_2(1, X_2), \Pi_2(0, X_2)), \max(\Pi_2(1, X_2), \Pi_2(0, X_2))],$$

there is a unique equilibrium in pure strategies. For instance, when $\varepsilon_i > \max(\Pi_i(1, X_i), \Pi_i(0, X_i))$, $i = 1, 2$, the game is a Prisoner's Dilemma. When, on the other hand,

$$\varepsilon \in [\min(\Pi_1(1, X_1), \Pi_1(0, X_1)), \max(\Pi_1(1, X_1), \Pi_1(0, X_1))]$$

$$\times[\min(\Pi_2(1, X_2), \Pi_2(0, X_2)), \max(\Pi_2(1, X_2), \Pi_2(0, X_2))],$$

there is always one equilibrium in mixed strategies. There is also two equilibria in pure strategies in the case of duopoly entry and coordination. For instance, when $\Pi_i(1) < \varepsilon_i < \Pi_i(0)$, $i = 1, 2$, we have a game of Chicken (or public good provision) and when $\Pi_i(0) < \varepsilon_i < \Pi_i(1)$, $i = 1, 2$, we have a Battle of the Sexes.

3.2.5 Object of inference

The analyst observes the realized strategy profile and realized values of the heterogeneity variables X_1 and X_2. However, realized values of heterogeneity

FIGURE 3.1 – Equilibrium correspondence in the duopoly entry case. For each value of the pair $(\varepsilon_1, \varepsilon_2)$, the predicted equilibria are given. In the central rectangle, corresponding to values of unobserved heterogeneity such that $\Pi_i(1) \leq \varepsilon_i \leq \Pi_i(0)$, for $i = 1, 2$, three equilibria are predicted, including two in pure strategies, $(Y_1 = 1, Y_2 = 0)$ and $(Y_1 = 0, Y_2 = 1)$ and one in mixed strategies, with Player i participating with probability $\sigma_i(\varepsilon_{3-i}) = (\Pi_{3-i}(0) - \Pi_{3-i}(1))^{-1}(\Pi_{3-i}(0) - \varepsilon_{3-i})$. In the rest of the $(\varepsilon_1, \varepsilon_2)$ space, single pure strategy Nash equilibria are predicted for each value of the unobserved heterogeneity pair $(\varepsilon_1, \varepsilon_2)$.

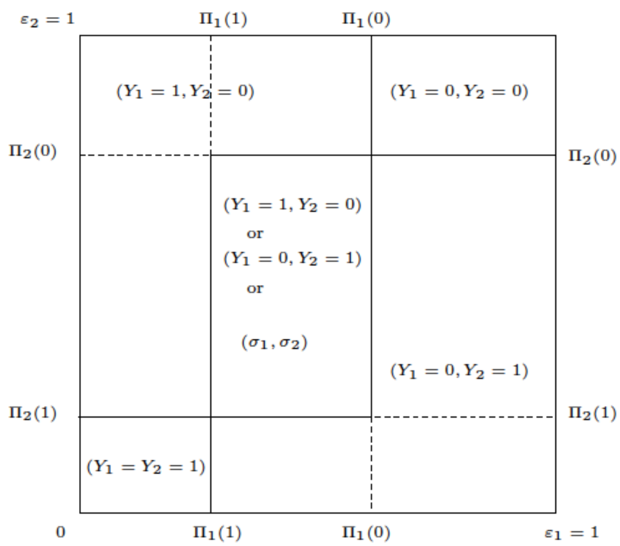


FIGURE 3.2 – Equilibrium correspondence in the coordination case. For each value of the pair $(\varepsilon_1, \varepsilon_2)$, the predicted equilibria are given. In the central rectangle, corresponding to values of unobserved heterogeneity such that $\Pi_i(0) \leq \varepsilon_i \leq \Pi_i(1)$, for $i = 1, 2$, three equilibria are predicted, including two in pure strategies, $(Y_1 = 1, Y_2 = 1)$ and $(Y_1 = 0, Y_2 = 0)$ and one in mixed strategies, with Player i participating with probability $\sigma_i(\varepsilon_{3-i}) = (\Pi_{3-i}(1) - \Pi_{3-i}(0))^{-1}(\varepsilon_{3-i} - \Pi_{3-i}(0))$. In the rest of the $(\varepsilon_1, \varepsilon_2)$ space, single pure strategy Nash equilibria are predicted for each value of the unobserved heterogeneity pair $(\varepsilon_1, \varepsilon_2)$.

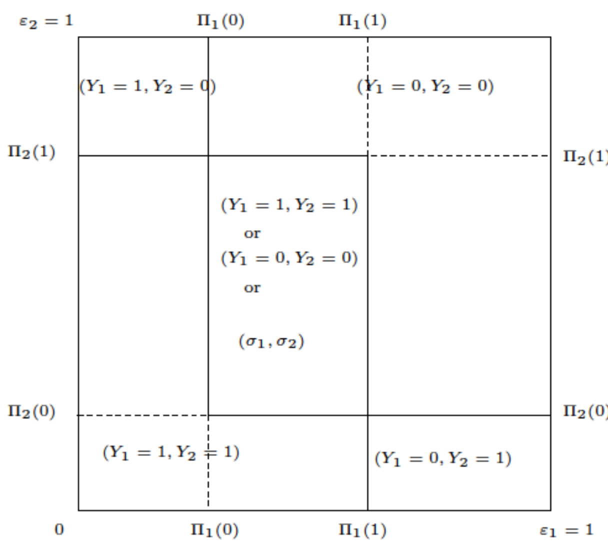
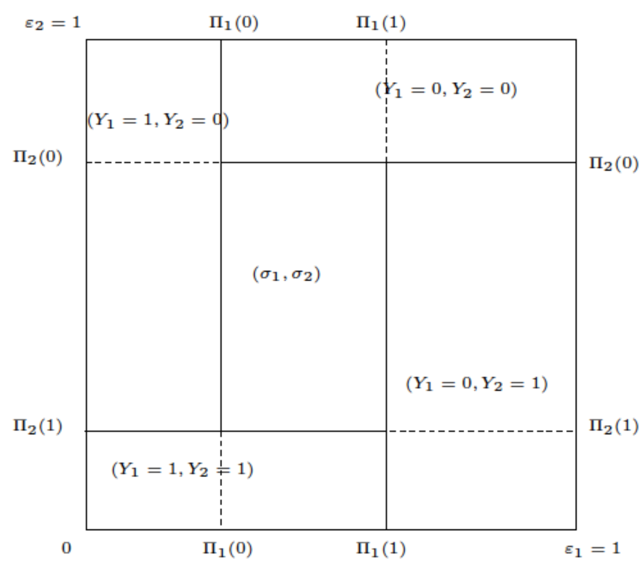


FIGURE 3.3 – Equilibrium correspondence in the asymmetric case. For each value of the pair $(\varepsilon_1, \varepsilon_2)$, the predicted equilibria are given. In the central rectangle, corresponding to values of unobserved heterogeneity such that $\Pi_1(0) \leq \varepsilon_1 \leq \Pi_1(1)$ and $\Pi_2(1) \leq \varepsilon_2 \leq \Pi_2(0)$, a single equilibrium in mixed strategies is predicted, with Player i participating with probability $\sigma_i(\varepsilon_{3-i}) = (\Pi_{3-i}(0) - \Pi_{3-i}(1))^{-1}(\Pi_{3-i}(0) - \varepsilon_{3-i})$. In the rest of the $(\varepsilon_1, \varepsilon_2)$ space, single pure strategy Nash equilibria are predicted for each value of the unobserved heterogeneity pair $(\varepsilon_1, \varepsilon_2)$.



variables ε_1 and ε_2 are not observed and the payoff functions Π_1 and Π_2 are unknown and are the object of inference. The model is incomplete in two respects :

1. The marginal distributions of the unobserved heterogeneity variables ε_1 and ε_2 are normalized. However, the joint distribution of $(\varepsilon_1, \varepsilon_2)$, which we shall denote $C(\varepsilon_1, \varepsilon_2)$ (since it is equal to the copula, given the uniform normalization) is unknown. This implies that although the probability of any horizontal or any vertical band in Figures 3.1-3.3 is predicted by the model, the probability of other rectangles are not. This means in particular that the likelihood of observing, say, $(Y_1 = 1, Y_2 = 1)$ in the duopoly entry case of Figure 3.1 is not pinned down by the model.
2. In each of the three Figures 3.1-3.3, multiple equilibria arise in the central region of the $(\varepsilon_1, \varepsilon_2)$ space. This implies that, short of additional information about the equilibrium selection mechanism, the model delivers multiple predictions for the strategy profile, only one of which is actually realized.

Model incompleteness results here, as we shall see, in partial identification of the payoff functions, the joint distribution of unobserved heterogeneity and the equilibrium selection mechanism. Throughout the paper, we shall treat the equilibrium selection mechanism as a nuisance parameter and concentrate on the derivation of the empirical content of the model, when no additional assumption is maintained about equilibrium selection. We shall proceed in two steps.

1. First we define and characterize the identified set for the distribution of unobserved heterogeneity and for the payoff functions (jointly). This will be achieved in Section 3.3 with an application of the characterization of the identified set for Shapley regular games in Galichon and Henry (2011).
2. Second we treat both the equilibrium selection mechanism and the distribution of unobserved heterogeneity as nuisance parameters and we derive in Section 3.4 the identified set for the payoff functions as the projection of the joint identified set obtained in Point (1).

3.3 Identified set for payoffs and heterogeneity distribution

In order to define and characterize the empirical content of Nash equilibrium play in 2×2 games of perfect information, we first clarify the observability structure and the structural elements to be identified.

Definition 2 (True frequencies)**.** *The probabilities of each of the four strategy profiles* $(Y_1 = j_1, Y_2 = j_2)$*, for* $j_1, j_2 = 1, 0$*, (as would be obtained from an infinite sample of i.i.d. replications of the game) are called* true frequencies *and denoted* $P(Y_1 = j_1, Y_2 = j_2 | X_1, X_2)$ *or* $P((j_1, j_2)|X_1, X_2)$ *for* $j_1, j_2 = 1, 0$*. We shall assume throughout this (partial) identification analysis that the true frequencies are known.*

Knowing the true frequencies of strategy profiles, we seek to characterize all the informational content of Nash equilibrium play with a finite collection of inequalities involving payoff functions $\Pi_i(j, X_i)$, $i = 1, 2$ and $j = 1, 0$ and the joint distribution of unobserved heterogeneity denoted :

$$C(u_1, u_2) = P(\varepsilon_1 \leq u_1, \varepsilon_2 \leq u_2),\ \forall (u_1, u_2) \in [0, 1]^2. \tag{3.2}$$

The notation $C(u_1, u_2)$ is chosen in reference to the fact that, given the uniform normalization of the marginals, C is also the copula of the pair $(\varepsilon_1, \varepsilon_2)$.

The inequalities characterizing the empirical content of the model will be sharp in the sense that all (C, Π_1, Π_2) that satisfy them are compatible with Nash equilibrium play in the 2×2 perfect information game specification. We define the identified set as in Beresteanu, Molchanov, and Molinari (2011).

Definition 3 (Identified set)**.** *The identified set for payoff functions and unobserved heterogeneity distribution is the collection of values of* (C, Π_1, Π_2) *such that there exists a probability* $\sigma \mapsto \mu(\sigma|\varepsilon, X, \Pi)$ *(an equilibrium selection mechanism) on the equilibrium correspondence* $G(\varepsilon, X, \Pi)$ *satisfying for each strategy profile* $(Y_1 = j_1, Y_2 = j_2)$*,* $j_1, j_2 = 1, 0$*,*

$$P((j_1, j_2)|X) = \int_{[0,1]^2} \left\{ \int_{G(\varepsilon, X, \Pi)} \sigma((j_1, j_2), \varepsilon, \Pi) \mu(\sigma|\varepsilon, X, \Pi) \right\} dC(\varepsilon_1, \varepsilon_2),$$

where $P((j_1, j_2)|X)$ *is the true frequency of* $(Y_1 = j_1, Y_2 = j_2)$*.*

Definition 3 is a rephrasing of the fact that there is a way to complete the model so that predicted probabilities are equal to true frequencies. Applying Theorem 5 of Galichon and Henry (2011) for Shapley regular games and removing redundant inequalities yields the characterization of the identified set given in Theorem 3 (see Appendix .3 for the proof). First, we need some additional notation relative to the probabilities of each strategy profile under mixed strategies.

Lemma 1 (Profile probabilities under mixed strategies)**.** *The probability that Player i participates in case the mixed strategy equilibrium is selected is* $\sigma_i(\varepsilon_{3-i}, \Pi_{3-i}) = (\Pi_{3-i}(0) - \Pi_{3-i}(1))^{-1}(\Pi_{3-i}(0) - \varepsilon_{3-i})$ *and the predicted probability of strategy profile* (j_1, j_2) *is* $\Sigma_{j_1,j_2}(C, \Pi)$ *with :*

$$\begin{aligned}
\Sigma_{11}(C,\Pi) &= \left|\int_{\Pi_1(1)}^{\Pi_1(0)}\int_{\Pi_2(1)}^{\Pi_2(0)} \sigma_1(\varepsilon_2,\Pi_2)\sigma_2(\varepsilon_1,\Pi_1)dC(\varepsilon_1,\varepsilon_2)\right|, \\
\Sigma_{00}(C,\Pi) &= \left|\int_{\Pi_1(1)}^{\Pi_1(0)}\int_{\Pi_2(1)}^{\Pi_2(0)} (1-\sigma_1(\varepsilon_2,\Pi_2))\,(1-\sigma_2(\varepsilon_1,\Pi_1))\,dC(\varepsilon_1,\varepsilon_2)\right|, \\
\Sigma_{10}(C,\Pi) &= \left|\int_{\Pi_1(1)}^{\Pi_1(0)}\int_{\Pi_2(1)}^{\Pi_2(0)} \sigma_1(\varepsilon_2,\Pi_2)\,(1-\sigma_2(\varepsilon_1,\Pi_1))\,dC(\varepsilon_1,\varepsilon_2)\right|, \\
\Sigma_{01}(C,\Pi) &= \left|\int_{\Pi_1(1)}^{\Pi_1(0)}\int_{\Pi_2(1)}^{\Pi_2(0)} (1-\sigma_1(\varepsilon_2,\Pi_2))\,\sigma_2(\varepsilon_1,\Pi_1)dC(\varepsilon_1,\varepsilon_2)\right|.
\end{aligned} \tag{3.3}$$

With this notation, we can state the characterization of the identified set.

Theorem 3 (Identified set)**.** (C, Π) *belongs to the identified set if and only if one of the following holds for almost all values of* X*. For ease of exposition, we denote* $P(i, j) = P(Y_1 = i, Y_2 = j|X_1, X_2)$ *and* $\Pi_i(j, X_i) = \Pi_i(j)$*,* $i = 1, 2$*, and* $j = 1, 0$*.*

1. *(Duopoly entry)* $\Pi_i(1) \leq \Pi_i(0)$*,* $i = 1, 2$*, and*

$$\begin{aligned}
&C(\Pi_1(1),\Pi_2(1)) \leq P(1,1) \leq C(\Pi_1(1),\Pi_2(1)) + \Sigma_{11}(C,\Pi) \\
&1-\Pi_1(0)-\Pi_2(0)+C(\Pi_1(0),\Pi_2(0)) \leq P(0,0) \\
&\leq 1-\Pi_1(0)-\Pi_2(0)+C(\Pi_1(0),\Pi_2(0))+\Sigma_{00}(C,\Pi), \\
&\Pi_2(0)+[C(\Pi_1(0),\Pi_2(1))-C(\Pi_1(1),\Pi_2(1))]-C(\Pi_1(0),\Pi_2(0)) \leq P(0,1) \\
&\leq \Pi_2(0)-C(\Pi_1(1),\Pi_2(0)), \\
&\Pi_1(0)+[C(\Pi_1(1),\Pi_2(0))-C(\Pi_1(1),\Pi_2(1))]-C(\Pi_1(0),\Pi_2(0)) \leq P(1,0) \\
&\leq \Pi_1(0)-C(\Pi_1(0),\Pi_2(1)).
\end{aligned} \tag{3.4}$$

2. *(Coordination game)* $\Pi_i(1) \geq \Pi_i(0)$, $i = 1, 2$, *and*

$$\begin{aligned}
&\Pi_2(0) - C(\Pi_1(1), \Pi_2(0)) \leq P(0,1) \\
&\leq \Pi_2(0) - C(\Pi_1(1), \Pi_2(0)) + \Sigma_{01}(C, \Pi), \\
&\Pi_1(0) - C(\Pi_1(0), \Pi_2(1)) \leq P(1,0) \\
&\leq \Pi_1(0) - C(\Pi_1(0), \Pi_2(1)) + \Sigma_{10}(C, \Pi), \\
&C(\Pi_1(0), \Pi_2(1)) + [C(\Pi_1(1), \Pi_2(0)) - C(\Pi_1(0), \Pi_2(0))] \leq P(1,1) \\
&\leq C(\Pi_1(1), \Pi_2(1)), \\
&1 - \Pi_1(0) - \Pi_2(0) + [C(\Pi_1(0), \Pi_2(1)) - C(\Pi_1(1), \Pi_2(1))] + C(\Pi_1(1), \Pi_2(0)) \\
&\leq P(0,0) \leq 1 - \Pi_1(0) - \Pi_2(0) + C(\Pi_1(0), \Pi_2(0)).
\end{aligned} \tag{3.5}$$

3. *(Asymmetric case 1)* $\Pi_1(1) \geq \Pi_1(0)$, $\Pi_2(1) \leq \Pi_2(0)$ *and*

$$\begin{aligned}
P(1,1) &= C(\Pi_1(1), \Pi_2(1)) + \Sigma_{11}(C, \Pi), \\
P(0,0) &= 1 - \Pi_1(0) - \Pi_2(0) + C(\Pi_1(0), \Pi_2(0)) + \Sigma_{00}(C, \Pi), \\
P(0,1) &= \Pi_2(0) - C(\Pi_1(1), \Pi_2(0)) + \Sigma_{01}(C, \Pi), \\
P(1,0) &= \Pi_1(0) - C(\Pi_1(0), \Pi_2(1)) + \Sigma_{10}(C, \Pi)..
\end{aligned} \tag{3.6}$$

4. *(Asymmetric case 2) The constraints of Case (3) hold after permutation of the two players.*

Consider the duopoly entry case. All other cases are derived in the same way. The equilibrium correspondence of the game is represented in Figure 3.1. The observation of strategy profile $(Y_1 = 1, Y_2 = 1)$ is rationalizable as the result of a pure strategy equilibrium in region $\varepsilon \in [0, \Pi_1(1)] \times [0, \Pi_2(1)]$ with probability $C(\Pi_1(1), \Pi_2(1))$ or as the result of a mixed strategy equilibrium in region $\varepsilon \in [\Pi_1(1), \Pi_1(0)] \times [\Pi_2(1), \Pi_2(0)]$ with probability Σ_{11} if the equilibrium in mixed strategies is selected. Hence the true frequency $P(1,1)$ is at least equal to $C(\Pi_1(1), \Pi_2(1))$ if the equilibrium in mixed strategies is never selected and at most equal to $C(\Pi_1(1), \Pi_2(1)) + \Sigma_{11}$ if the equilibrium in mixed strategies is always selected. Hence we recover the bounds on the first line of (3.4). The same reasoning applies to strategy profile $(Y_1 = 0, Y_2 = 0)$ to yield the second line of (3.4).

The observation of strategy profile $(Y_1 = 0, Y_2 = 1)$ can be rationalized as the result of a pure strategy equilibrium in the lower right L-shaped region or as the result of a pure strategy equilibrium or a mixed strategy equilibrium in region $\varepsilon \in [\Pi_1(1), \Pi_1(0)] \times [\Pi_2(1), \Pi_2(0)]$. The maximum rationalizable true frequency $P(0,1)$ is therefore obtained when the pure strategy equilibrium $(Y_1 = 0, Y_2 = 1)$ is always selected in region $\varepsilon \in [\Pi_1(1), \Pi_1(0)] \times [\Pi_2(1), \Pi_2(0)]$. The resulting upper bound is equal to $P(\varepsilon_1 \geq \Pi_1(1), \varepsilon_2 \leq \Pi_2(0))$, which is equal to the right-hand side on the third line of (3.4). The minimum rationalizable true frequency $P(0,1)$ is obtained when the pure strategy equilibrium $(Y_1 = 1, Y_2 = 0)$ is always selected so that $(Y_1 = 0, Y_2 = 1)$ never occurs in region $\varepsilon \in [\Pi_1(1), \Pi_1(0)] \times [\Pi_2(1), \Pi_2(0)]$. The resulting lower bound is the probability of the lower left L-shaped region, whose probability is equal to the left-hand side of Line 3 of (3.4). The same reasoning applies to true frequency $P(1,0)$ and Line 4 of (3.4).

Note that additional constraints can be derived from the analysis of the game. In particular, the maximum rationalizable frequency $P(0,1)$ is obtained when the pure strategy equilibrium $(Y_1 = 0, Y_2 = 1)$ is always selected in the region with multiple equilibria. This implies of course that the other equilibria are never selected, which constrains the rationalizable frequency $P(Y_1 = 1, Y_2 = 0)$. Hence $P(0,1) + P(1,0)$ is bounded above by $1 - \Pi_1(0) - \Pi_2(0) + C(\Pi_1(1), \Pi_2(1)) - C(\Pi_1(0), \Pi_2(0))$. However, the latter constraint on (C, Π) is redundant, as it is implied by the combination of $P(1,1) \geq C(\Pi_1(1), \Pi_2(1))$ and $P(0,0) \geq 1 - \Pi_1(0) - \Pi_2(0) + C(\Pi_1(0), \Pi_2(0))$.

This shows that true frequencies that are rationalizable as Nash equilibrium strategy profiles of the 2×2 game necessarily satisfy inequalities in (3.4-3.6). The proof of Theorem 3 in Appendix .3 shows the converse, namely that true frequencies that satisfy inequalities (3.4-3.6) are rationalizable as Nash equilibrium strategy profiles of the 2×2 game. Hence, the bounds of Theorem 3 are sharp.

3.4 Empirical content of equilibrium in pure strategies

When equilibria in mixed strategies are ruled out, $\Sigma_{j_1 j_2} = 0$ for $j_1, j_2 = 1, 0$ and the lower bounds in each of the inequalities in (3.4)-(3.6) are redundant. Hence we have the following result.

Theorem 4 (Identified set for (C, Π) with only pure strategies)**.** (C, Π) *belongs to the identified set if and only if one of the following holds for almost all values of X. For ease of exposition, we denote $P(i,j) = P(Y_1 = i, Y_2 = j | X_1, X_2)$ and $\Pi_i(j, X_i) = \Pi_i(j)$, $i = 1, 2$, and $j = 1, 0$.*

1. *(Duopoly entry)* $\Pi_i(1) \leq \Pi_i(0)$, $i = 1, 2$, *and*

$$\begin{aligned} P(1,1) &= C(\Pi_1(1), \Pi_2(1)) \\ P(0,0) &= 1 - \Pi_1(0) - \Pi_2(0) + C(\Pi_1(0), \Pi_2(0)), \\ P(0,1) &\leq \Pi_2(0) - C(\Pi_1(1), \Pi_2(0)), \\ P(1,0) &\leq \Pi_1(0) - C(\Pi_1(0), \Pi_2(1)). \end{aligned} \tag{3.7}$$

2. *(Coordination game)* $\Pi_i(1) \geq \Pi_i(0)$, $i = 1, 2$, *and*

$$\begin{aligned} P(0,1) &= \Pi_2(0) - C(\Pi_1(1), \Pi_2(0)), \\ P(1,0) &= \Pi_1(0) - C(\Pi_1(0), \Pi_2(1)), \\ P(1,1) &\leq C(\Pi_1(1), \Pi_2(1)), \\ P(0,0) &\leq 1 - \Pi_1(0) - \Pi_2(0) + C(\Pi_1(0), \Pi_2(0)). \end{aligned} \tag{3.8}$$

3. *(Asymmetric case 1)* $\Pi_1(1) \geq \Pi_1(0)$, $\Pi_2(1) \leq \Pi_2(0)$ *and*

$$\begin{aligned} P(1,1) &= C(\Pi_1(1), \Pi_2(1)) \\ P(0,0) &= 1 - \Pi_1(0) - \Pi_2(0) + C(\Pi_1(0), \Pi_2(0)) \\ P(0,1) &= \Pi_2(0) - C(\Pi_1(1), \Pi_2(0)), \\ P(1,0) &= \Pi_1(0) - C(\Pi_1(0), \Pi_2(1)). \end{aligned} \tag{3.9}$$

4. *(Asymmetric case 2) The constraints of Case (3) hold after permutation of the two players.*

The results of Theorem 4 can be applied in several ways. We describe two polar cases. On the one hand, we may add assumptions on the joint distribution of firm specific unobserved heterogeneity $(\varepsilon_1, \varepsilon_2)$, positing (1) a parametric copula, (2) perfect correlation of $(\varepsilon_1, \varepsilon_2)$, as in the case of an industry-wide shock or (3) independence of ε_1 and ε_2 as in the case of purely idiosyncratic shocks. A combination of the latter two cases can also be entertained in the form of (4) a factor model. These implications are detailed in Section 3.4.1. On the other hand, we may acknowledge total ignorance of the joint distribution of firm specific unobserved heterogeneity and project the identified region of Theorem 5 to obtain nonparametric sharp bounds on the payoff functions only. We describe this in Section 3.4.2.

3.4.1 Restrictions on the joint distribution of firm specific heterogeneity

We consider first refinements of the bounds of Theorem 4 based on a variety of assumptions on the joint distribution of firm specific unobserved heterogeneity.

Parametric restrictions on the copula

In the case where the copula for $(\varepsilon_1, \varepsilon_2)$ is parameterized with parameter vector θ, sharp bounds are obtained straightforwardly by replacing $C(\varepsilon_1, \varepsilon_2)$ with the parametric version $C(\varepsilon_1, \varepsilon_2, \theta)$ in Lemma 1 and Theorems 3 and 4. Parameterizing the copula $C(\varepsilon_1, \varepsilon_2)$ while leaving the marginal distributions of ε_1 and ε_2 unrestricted yields nonparametric bounds, akin to those derived by Aradillas-Lopez (2010) in the case of incomplete information games.

Perfect correlation

The case of perfect correlation between the two firm specific unobserved heterogeneity components is also of interest, as it corresponds to an industry-wide productivity shock in industrial organization applications. In that case, the copula attains its Fréchet upper bounds $C(\varepsilon_1, \varepsilon_2) = \min(\varepsilon_1, \varepsilon_2)$ so that the sharp bounds of Theorem 4 in case of duopoly entry yield $P(1,1) = \min(\Pi_1(1), \Pi_2(1))$, $P(0,0) = \min(1 - \Pi_1(0), 1 - \Pi_2(0))$, $P(0,1) \leq \max(\Pi_2(0) - \Pi_1(1), 0)$ and $P(1,0) \leq \max(\Pi_1(0) - \Pi_2(1), 0)$. Similar sharp bounds for the three others cases may be easily derived.

Independence

In the other polar case, where the two firms specific unobserved heterogeneity components are purely idiosyncratic shocks, $\varepsilon_1 \perp\!\!\!\perp \varepsilon_0$ and sharp bounds are derived from Theorem 2 by simply setting $C(\varepsilon_1, \varepsilon_2) = \varepsilon_1 \varepsilon_2$.

Factor structure

Intermediate cases between the two polar cases of industry-wide shock and idiosyncratic shocks can also be entertained with a simple factor model for the pair of unobserved heterogeneities $(\varepsilon_1, \varepsilon_2)$. Suppose unobserved heterogeneity has factor structure $\varepsilon_d = \alpha_d \varepsilon + \eta_d$, $d = 1, 2$, with $\mathbb{E}\varepsilon = 0$, $\mathbb{E}\varepsilon^2 = 1$ (without loss of generality) and $\eta_1 \perp\!\!\!\perp \eta_2 | \varepsilon$. η_d is uniformly distributed on $[0, 1]$ for $d = 1, 2$, conditionally on ε. This factor specification achieves a decomposition of unobserved heterogeneity components into an industry common shock ε and a purely idiosyncratic shock η_d, $d = 1, 2$. We recover the case of purely idiosyncratic firm specific unobserved heterogeneity, when $\alpha_1 = \alpha_0 = 0$. By iterated expectations, we find for each $i, j = 1, 0$:

$$\begin{aligned}
C(\Pi_1(i, x_1), \Pi_2(j, x_2)|x_1, x_2) &= \mathbb{P}(\varepsilon_1 \leq \Pi_1(i, x_1), \varepsilon_2 \leq \Pi_2(j, x_2)|x_1, x_2) \\
&= \mathbb{E}_\varepsilon \mathbb{P}(\eta_1 \leq \Pi_1(i, x_1) - \alpha_1 \varepsilon, \\
&\qquad\qquad \eta_2 \leq \Pi_2(j, x_2) - \alpha_2 \varepsilon | x_1, x_2, \varepsilon) \\
&= \mathbb{E}_\varepsilon \mathbb{P}(\eta_1 \leq \Pi_1(i, x_1) - \alpha_1 \varepsilon | x_1, \varepsilon) \\
&\qquad\qquad \mathbb{P}(\eta_2 \leq \Pi_2(j, x_2) - \alpha_2 \varepsilon | x_1, x_2, \varepsilon) \\
&= \Pi_1(i, x_1)\Pi_2(j, x_2) + \alpha_1 \alpha_2,
\end{aligned}$$

from which sharp bounds can be derived for the payoff functions and the pair (α_1, α_2).

3.4.2 Sharp bounds on the payoff functions

From the identified set for (Π, C) we can derive sharp bounds for the payoff functions alone using Fréchet bounds on C in each of the four cases. Consider the duopoly entry case for instance. Line 1 of (3.7) yields $P(1, 1) = C(\Pi_1(1), \Pi_2(1)) \leq \min(\Pi_1(1), \Pi_2(2))$ (Fréchet bound). Similarly, Line 2 of (3.7) yields $1 - P(0, 0) \geq \max(\Pi_1(0), \Pi_2(0))$. Since $\Pi_2(0) \geq \Pi_2(1)$, we have $C(\Pi_1(1), \Pi_2(0)) \geq C(\Pi_1(1), \Pi_2(1))$ and Lines 1 and 3 of (3.7) combined yield $P(1, 1)+P(0, 1) \leq \Pi_2(0)-[C(\Pi_1(1), \Pi_2(0))-C(\Pi_1(1), \Pi_2(1))] \leq \Pi_2(0)$. Similarly, Lines 1 and 4 yield $P(1, 1)+P(1, 0) \leq \Pi_1(0)$. Finally, $P(0, 1)+P(1, 1) = 1-P(1, 0)-P(0, 0) \geq \Pi_2(0)-[C(\Pi_1(0), \Pi_2(0))-C(\Pi_1(0), \Pi_2(1))] \geq \Pi_2(0)-[\Pi_2(0)-\Pi_2(1)] = \Pi_2(1)$ and similarly $P(1, 0)+P(1, 1) \geq \Pi_1(1)$. We therefore have the validity of the following bounds for the duopoly entry case :

$$\begin{aligned}
P(1, 1) &\leq \Pi_1(1) \leq P(1, 1) + P(1, 0) \leq \Pi_1(0) \leq 1 - P(0, 0), \\
P(1, 1) &\leq \Pi_2(1) \leq P(1, 1) + P(0, 1) \leq \Pi_2(0) \leq 1 - P(0, 0).
\end{aligned}$$

For the coordination case, the same method (see the proof of Theorem 5) yields :

$$\begin{array}{lllllllll} P(1,0) & \leq & \Pi_1(0) & \leq & P(1,1)+P(1,0) & \leq & \Pi_1(1) & \leq & 1-P(0,1), \\ P(0,1) & \leq & \Pi_2(0) & \leq & P(1,1)+P(0,1) & \leq & \Pi_2(1) & \leq & 1-P(1,0). \end{array}$$

and finally for the asymmetric cases :

$$\begin{array}{lllllllll} P(1,0) & \leq & \Pi_1(0) & \leq & P(1,1)+P(1,0) & \leq & \Pi_1(1) & \leq & 1-P(0,1), \\ P(1,1) & \leq & \Pi_2(1) & \leq & P(1,1)+P(0,1) & \leq & \Pi_2(0) & \leq & 1-P(0,0), \end{array}$$

and similarly after permutation of the two players. We can now formally characterize the joint sharp bounds on payoff functions when only pure strategies are entertained.

Theorem 5 (Sharp bounds for payoff functions)**.** Π *belongs to the identified set if and only if (3.10) and (3.11) below hold.*

$$\begin{array}{lll} \min\left(\Pi_1(1,x_1),\Pi_1(0,x_1)\right) & \leq & \inf_{x_2}\Big(P(1,1|x_1,x_2)+P(1,0|x_1,x_2)\Big) \\ \max\left(\Pi_1(1,x_1),\Pi_1(0,x_1)\right) & \geq & \sup_{x_2}\Big(P(1,1|x_1,x_2)+P(1,0|x_1,x_2)\Big) \\ \min\left(\Pi_2(1,x_2),\Pi_2(0,x_2)\right) & \leq & \inf_{x_1}\Big(P(1,1|x_1,x_2)+P(0,1|x_1,x_2)\Big) \\ \max\left(\Pi_2(1,x_2),\Pi_2(0,x_2)\right) & \geq & \sup_{x_1}\Big(P(1,1|x_1,x_2)+P(0,1|x_1,x_2)\Big) \end{array} \tag{3.10}$$

and

$$\begin{array}{lllll} \sup_{x_2} P(1,1|x_1,x_2) & \leq & \Pi_1(1,x_1) & \leq & \inf_{x_2}\Big(1-P(0,1|x_1,x_2)\Big) \\ \sup_{x_2} P(1,0|x_1,x_2) & \leq & \Pi_1(0,x_1) & \leq & \inf_{x_2}\Big(1-P(0,0|x_1,x_2)\Big) \\ \sup_{x_1} P(1,1|x_1,x_2) & \leq & \Pi_2(1,x_2) & \leq & \inf_{x_1}\Big(1-P(1,0|x_1,x_2)\Big) \\ \sup_{x_1} P(0,1|x_1,x_2) & \leq & \Pi_2(0,x_2) & \leq & \inf_{x_1}\Big(1-P(0,0|x_1,x_2)\Big). \end{array} \tag{3.11}$$

In the case without excluded variables, it is immediately apparent from the bounds of Theorem 5 that the sign of $\Pi_i(1)-\Pi_i(0)$ is not identified, hence we cannot determine from the data only, whether the game is a duopoly entry game, a game of cooperation or an asymmetric game. With exclusion restrictions, however, it becomes possible to identify the class of games if bounds cross in all cases except one. An example is the case when

$$\sup_{x_2}\left(P(1,0|x_1,x_2)+P(1,1|x_1,x_2)\right) > \inf_{x_2}\left(1-P(0,1|x_1,x_2)\right)$$

and

$$\sup_{x_1}\left(P(0,1|x_1,x_2)+P(1,1|x_1,x_2)\right) > \inf_{x_1}\left(1-P(1,0|x_1,x_2)\right),$$

so that cooperation and asymmetric games are rejected, whereas

$$\sup_{x_2}\left(P(1,0|x_1,x_2)+P(1,1|x_1,x_2)\right) \leq \inf_{x_2}\left(1-P(0,0|x_1,x_2)\right)$$

and

$$\sup_{x_1}\left(P(0,1|x_1,x_2)+P(1,1|x_1,x_2)\right) \leq \inf_{x_1}\left(1-P(0,0|x_1,x_2)\right),$$

so that the duopoly entry game is not rejected.

In the case without excluded variable, the bounds on the payoff functions $\Pi_i(1)$ and $\Pi_i(0)$ can be reduced to a point, but may never cross, so that the hypothesis of Nash equilibrium play is not falsifiable. If, on the other hand, there is an exclusion restriction, hence variation in the payoff of one player that leaves the other player's payoff unchanged, the bounds may cross and the joint assumption of Nash equilibrium play and the exclusion restriction may be rejected. For instance, if $\inf_{x_2}\left(P(1,0|x_1,x_2)+P(1,1|x_1,x_2)\right) < \min(\sup_{x_2}\left(P(1,1|x_1,x_2)\right), \sup_{x_2}\left(P(1,0|x_1,x_2)\right))$ then the bounds cross in all cases and the model is rejected.

Sharp bounds on monopoly advantage can also be easily derived from the bounds of Theorem 5. Indeed, considering Player 1 only for simplicity, monopoly advantage is

$$|\Pi_1(0,x_1)-\Pi_1(1,x_1)| \leq 1-\min\left(\sup_{x_2}P(0,0|x_1,x_2), \sup_{x_2}P(0,1|x_1,x_2)\right) - \sup_{x_2}P(1,1|x_1,x_2).$$

If we assume a priori that the game is duopoly entry, then the bounds on monopoly advantage simplify to

$$\Pi_1(0, x_1) - \Pi_1(1, x_1) \leq 1 - \sup_{x_2} P(0, 0|x_1, x_2) - \sup_{x_2} P(1, 1|x_1, x_2)$$

. Bounding free-riding incentives $\varepsilon_1 - \Pi_1(1)$ (free riding incentives of Player 1) is a little more involved, since they involve the unobserved heterogeneity component ε. We may however apply the bounds of Theorem 4 to derive joint sharp bounds on the distribution of the pair $(\varepsilon_1 - \Pi_1(1), \varepsilon_2 - \Pi_2(1))$.

Conclusion

This paper contributed to the literature on the empirical analysis of game theoretic models of economic interactions by providing sharp bounds on nonparametrically specified payoff functions and type distributions. This complements results of Kline and Tamer (2012) who derive sharp bounds on best response functions. The bounds obtained lend themselves to standard partial identification inference methods, and therefore allow nonparametric inference on utility functions, profit functions, unobserved heterogeneity distributions and more specific quantities such as the extent of monopoly advantage in duopoly entry games and free riding incentives in cooperation games. The method employed to derive sharp bounds on payoff functions only as a projection of the joint identified region for payoff functions and type distributions could be applied to higher dimensions to extend the present results to multiperson games with more complex strategy spaces. Other equilibrium concepts (Stackelberg, correlated strategies etc...) could also be entertained in future work.

Bibliographie

Arild Aakvik, James Heckman, and Edward Vytlacil. Estimating treatment effects for discrete outcomes when responses to treatment vary : an application to norwegian vocational rehabilitation programs. *Journal of Econometrics*, 125 :15–51, 2005.

R.H Adams. The economic and demographic determinants of international migration in rural egypt. *Journal of Development Studies*, 30(1) :146–167, 1993.

Donald Andrews and Xiaoxia Shi. Nonparametric inference based on conditional moment inequalities. Cowles Foundation Discussion Paper 1840, 2011.

Andres Aradillas-Lopez. Semiparametric estimation of simultaneous games with incomplete information. *Journal of Econometrics*, 157 :409–431, 2010.

Andres Aradillas-Lopez. Nonparametric probability bounds for nash equilibrium actions in a simultaneous discrete game. *Quantitative Economics*, 2 :135–171, 2011.

Andres Aradillas-Lopez and Elie Tamer. The identification power of equilibrium in simple games. *Journal of Business and Economic Statistics*, 26 : 261–283, 2008.

Pat Bajari, Han Hong, and Steve Ryan. Identification and estimation of a discrete game of complete information. *Econometrica*, 78 :1529–1568, 2010.

Pat Bajari, Jin Hahn, Han Hong, and Geert Ridder. A note on semiparametric estimation of finite mixtures of discrete choice models with applications to game theoretic models. *International Economic Review*, 52 :807–824, 2011.

Pat Bajari, Han Hong, and Denis Nekipelov. Game theory and econometrics : a survey of some recent results. unpublished manuscript, 2012.

E. Battistin and E. Rettore. Testing for programme effects in a regression discontinuity design with imperfect compliance. *Journal of the Royal Statistical Society*, 165(1) :39–57, 2002.

Patrick Bayer, Shakeeb Khan, and Christopher Timmins. Nonparametric identification and estimation in a roy model with common nonpecuniary returns. *Journal of Business and Economic Statistics*, 29, 2011. 201–215.

M. Beffy, D. Fougère, and A. Maurel. The effect of part-time work on post-secondary educational attainment : New evidence from french data. unpublished manuscript, 2010.

Ari Beresteanu, Ilya Molchanov, and Francesca Molinari. Sharp identification regions in games. cemmap working paper CWP15/08, 2008.

Ari Beresteanu, Ilya Molchanov, and Francesca Molinari. Sharp identification regions in models with convex predictions. *Econometrica*, 79 :1785–1821, 2011.

Steve Berry. Estimation of a model of entry in the airline industry. *Econometrica*, 60 :889–917, 1992.

J. Bhattacharya, A. Shaikh, and E. Vytlacil. Treatment effect bounds under monotonicity conditions : An application to swan-ganz catherization. *American Economic Review*, 98 :351–356, 2008.

P Bjorn and Quang Vuong. Simultaneous equations models for dummy endogenous variables. Caltech Working Paper 537, 1984.

R. Blundell, A. Gosling, H. Ichimura, and C. Meghir. Changes in the distribution of male and female wages accounting for employement composition using bounds. *Econometrica*, 75 :323–363, 2007.

Tim Bresnahan and P. Reiss. Entry in monopoly markets. *Review of Economic Studies*, 57 :531–553, 1990.

Tim Bresnahan and P. Reiss. Empirical models of discrete games. *Journal of Econometrics*, 48 :57–81, 1991.

E. Canton and A. Blom. Can student loans improve accessibility to higher education and student performance ? an impact study of the case of sofes, mexico. unpublished manuscript, 2004.

C. Carpenter and C. Dobkin. The effect of alcohol consumption on mortality : Regression discontinuity evidence from the minimum drinking age. *American Economic Journal : Applied Economics*, 1 :164—182, january 2009.

Victor Chernozhukov, Simon Lee, and Adam Rosen. Inference on intersection bounds. CEMMAP Working Paper CWP19/09, 2009.

A. Chesher. Identification in nonseparable models. *Econometrica*, 71 :1405–1441, 2003.

A. Chesher. Nonparametric identification under discrete variation. *Econometrica*, 73 :1525–1550, 2005.

Andrew Chesher. Instrumental variable models for discrete outcomes. *Econometrica*, 78 :575–601, 2010.

Andrew Chesher and Konrad Smolinski. An instrumental variable model of ordered choice. CEMMAP Working Paper CWP37/09, 2009.

Andrew Chesher, Adam Rosen, and Konrad Smolinski. An instrumental variable model of multiple discrete choice. CEMMAP Working Paper CWP06/11, 2011.

R. C. Chiburis. Semiparametric bounds on treatment effect. *Journal of Econometrics*, 159 :267–275, 2010.

M Engers and S Stern. Family bargaining and long term care. *International Economic Review*, 43 :73–114, 2002.

Y. Fan and S. S. Park. Sharp bounds on the distribution of treatment effects and their statistical inference. *Econometric Theory*, 26 :931–951, 2010.

Yanqin Fan and Jisong Wu. Partial identification of the distribution of treatment effects in switching regime models and its confidence sets. *Review of Economic Studies*, 77, 2010. 1002–1041.

Alfred Galichon and Marc Henry. Inference in incomplete models. unpublished manuscript, 2006.

Alfred Galichon and Marc Henry. Set identification in models with multiple equilibria. *Review of Economic Studies*, 78(4) :1264–1298, 2011.

Christian Gouriéroux, Jean-Jacques Laffont, and Alain Monfort. Coherency conditions in simultaneous linear equations models with endogenous switching regimes. *Econometrica*, 48 :675–695, 1980.

Xavier Haultfoeuille and Arnaud Maurel. Inference on an extended roy model, with an application to schooling decisions in france. unpublished manuscript, 2011.

James Heckman. Dummy endogenous variables in a simultaneous equations system. *Econometrica*, 46 :931–960, 1978.

James Heckman. Varieties of selection bias. *American Economic Review*, 80 : 313–318, 1990.

James Heckman and Bo Honoré. The empirical content of the roy model. *Econometrica*, 58 :1121–1149, 1990.

James Heckman and Edward Vytlacil. Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences*, 96 :4730–4734, 1999.

James Heckman and Edward Vytlacil. Instrumental variables, selection models and tight bounds on the average treatment effect. In *Econometric Evaluations of Labour Market Policies, Lechner, M., and F. Pfeiffer, eds.*, pages 1–16. Heidelberg : Springer-Verlag, 2001.

James Heckman, Jeffrey Smith, and Nancy Clements. Making the most out of programme evaluation and social experiments : accounting for heterogeneity in programme impacts. *Review of Economic Studies*, 64 :487–535, 1997.

M. Hinich and M. Munger. *Analytical politics.* Cambridge University Press : Cambridge, 1997.

Bo Honoré and Adriana Lleras-Muney. Bounds in competing risks models and the war on cancer. *Econometrica*, 74 :1675–1698, 2006.

G. Imbens and T. Lemieux. Regression discontinuity designs : A guide to practice. *Journal of Econometrics*, 142 :615–635, 2008.

G. Imbens and W. Newey. Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, 77 :1481–1512, 2009.

Boyan Jovanovic. Observable implications of models with multiple equilibria. *Econometrica*, 57 :1431–1437, 1989.

S. Jun, J. Pinkse, and H. Xu. Discrete endogenous variables in weakly separable models. unpublished manuscript, 2011.

Sun Jun, Joris Pinkse, and Haiqing Xu. Tighter bounds in triangular systems. *Journal of Econometrics*, 161 :122–128, 2010.

M. Kasy. Identification in triangular systems using control functions. forthcoming in *Econometric Theory*, 2011.

Toru Kitagawa. Identification region of the potential outcome distributions under instrument independence. CEMMAP Working Paper CWP30/09, 2009.

W. Van Der Klaauw. Estimating the effect of financial aid offers on college enrollement : a regression-discontinuity approach. *International Economic Review*, 43 :1249–1287, 2002.

Brendan Kline and Elie Tamer. Bounds for best response functions in binary games. *Journal of Econometrics*, 166 :92–105, 2012.

P Kooreman. Estimation of some econometric models of some discrete games. *Journal of Applied Econometrics*, 9 :255–268, 1994.

D. Lee. Randomized experiments from non-random selection in u.s. house elections. *Journal of Econometrics*, 142 :675–697, 2008.

D. Lee and T. Lemieux. Regression discontinuity designs in economics. unpublished manuscript, 2009.

R.E Lucas. *Internal Migration in Developing Countries*, volume 1B of *Handbook of Population and Family Economics.* Amsterdam : Elsevier Science Publishing, 1997.

Thierry Magnac and Eric Maurin. Partial identification in monotone binary models : discrete regressors and interval data. *Review of Economic Studies*, 75 :835–864, 2008.

C. Manski. Identification of binary response models. *Journal of the American Statistical Association*, 83(403) :729–738, Sep 1988.

C. Manski. Anatomy of the selection problem. *Journal of Human ressources*, 24 :243–360, 1989.

C. Manski and J. Pepper. Monotone instrumental variables with an application to the returns to schooling. *Econometrica*, 68 :997–1010, 2000.

Charles Manski. Nonparametric bounds on treatment effects. *American Economic Review, Papers and Proceedings*, 80 :319–323, 2000.

Romuald Méango. International student migration. who goes ? who pays ? who gains ? unpublished manuscript, 2012.

R.A. Nakosteen and M. Zimmer. Migration and income : The question of self-selection. *Southern Economic Journal*, 46(3) :840–851, 1980.

Roger. B. Nelsen. *An introduction to copulas*. Springer, New york, 2006.

Arthur Peterson. Bounds on a joint distribution function with fixed subdistribution functions : application to competing risks. *Proceedings of the National Academy of Sciences*, 73 :11–13, 1976.

Anatol Rapoport and Melvin Guyer. A taxonomy of 2×2 games. *General Systems*, 11 :203–14, 1966.

Albert Roy. Some thoughts on the distribution of earnings. *Oxford Economic Papers*, 3 :135–146, 1951.

R. Sabates-Wheeler, R. Sabates, and A. Castaldo. Tackling poverty-migration linkages : Evidence from ghana and egypt. unpublished manuscript, 2005.

Azeem Shaikh and Edward Vytlacil. Partial identification in triangular systems of equations with binary dependent variables. *Econometrica*, 79 : 949–955, 2011.

A Soetevent and Kooreman. A discrete choice model with social interactions : with an application to high school teen behaviour. *Journal of Applied Econometrics*, 22 :599–624, 2007.

Elie Tamer. Incomplete simultaneous discrete response model with multiple equilibria. *Review of Economic Studies*, 70 :147–165, 2003.

A Tsiatis. A nonidentifiability aspect of the problem of competing risks. *Proceedings of the Natinal Academy of Sciences*, 72 :20–22, 1975.

E. Vytlacil and N. Yildiz. Dummy endogenous variables in weakly separable models. *Econometrica*, 75 :757–779, 2007.

Edward Vytlacil. Independence, monotonicity and latent index models : an equivalence result. *Econometrica*, 70 :331–341, 2002.

Annexes

.1 Proof of main results in Chapter 1

Lemma 2. *Suppose $Supp(P|X = x)$ is not a singleton for all $x \in Dom(X)$.*

1. *(Lemma 1, SV) Let p and p' belong to $Supp(P \mid X = x) \cap Supp(P \mid X = x')$ such that $p' < p$, then*

$$sign([\nu(1, x') - \nu(0, x)]) = sign(h(x, x', p, p'))$$

where $h(x, x', p, p') = (P(1, 1|x', p) - P(1, 1|x', p')) - (P(1, 0|x, p') - P(1, 0|x, p))$.

2. *Let $p'_1 < p'_2 \in Supp(P \mid X = x')$ and $p_1 < p_2 \in Supp(P \mid X = x)$ such that $[p'_1, p'_2] \subseteq [p_1, p_2]$ then,*

$$h(1, x, x', p_1, p_2, p'_1, p'_2) \geq 0 \Rightarrow [\nu(1, x') - \nu(0, x)] \geq 0$$

where $h(1, x, x', p_1, p_2, p'_1, p'_2) = (P(1, 1|x', p'_2) - P(1, 1|x', p'_1)) - (P(1, 0|x, p_1) - P(1, 0|x, p_2))$.

3. *Let $p'_1 < p'_2 \in Supp(P \mid X = x')$ and $p_1 < p_2 \in Supp(P \mid X = x)$ such that $[p'_1, p'_2] \subseteq [p_1, p_2]$ then,*

$$h(0, x, x', p_1, p_2, p'_1, p'_2) \geq 0 \Rightarrow [\nu(0, x') - \nu(1, x)] \geq 0$$

where $h(0, x, x', p_1, p_2, p'_1, p'_2) = (P(1, 0|x', p'_1) - P(1, 0|x', p'_2)) - (P(1, 1|x, p_2) - P(1, 1|x, p_1))$.

4. *Let p and p' belong to $Supp(P \mid X = x) \cap Supp(P \mid X = x')$ such that $p' < p$, then*

$$sign([\nu(1, x') - \nu(1, x)]) = sign(\tilde{h}_1(x, x', p, p'))$$

where $\tilde{h}_1(x, x', p, p') = (P(1, 1|x', p) - P(1, 1|x', p')) - (P(1, 1|x, p) - P(1, 1|x, p'))$.

5. *Let p and p' belong to $Supp(P \mid X = x) \cap Supp(P \mid X = x')$ such that $p' < p$, then*

$$sign([\nu(0, x') - \nu(0, x)]) = sign(\tilde{h}_0(x, x', p, p'))$$

where $\tilde{h}_0(x, x', p, p') = (P(1, 0|x', p') - P(1, 0|x', p)) - (P(1, 0|x, p') - P(1, 1|x, p))$.

6. *Let $p'_1 < p'_2 \in Supp(P \mid X = x')$ and $p_1 < p_2 \in Supp(P \mid X = x)$ such that $[p'_1, p'_2] \subseteq [p_1, p_2]$ then,*

$$\tilde{h}_1(x, x', p_1, p_2, p'_1, p'_2) \geq 0 \Rightarrow [\nu(1, x') - \nu(1, x)] \geq 0$$

where $\tilde{h}_1(x, x', p_1, p_2, p'_1, p'_2) = (P(1, 1|x', p'_2) - P(1, 1|x', p'_1)) - (P(1, 1|x, p_2) - P(1, 1|x, p_1))$.

7. *Let $p'_1 < p'_2 \in Supp(P \mid X = x')$ and $p_1 < p_2 \in Supp(P \mid X = x)$ such that $[p'_1, p'_2] \subseteq [p_1, p_2]$ then,*

$$\tilde{h}_0(x, x', p_1, p_2, p'_1, p'_2) \geq 0 \Rightarrow [\nu(0, x') - \nu(0, x)] \geq 0$$

where $\tilde{h}_1(x, x', p_1, p_2, p'_1, p'_2) = (P(1, 0|x', p'_1) - P(1, 0|x', p'_2)) - (P(1, 0|x, p_1) - P(1, 1|x, p_2))$.

Proof of lemma 2. I will prove cases (2), (4) and (6). The cases (3), (5), and (7) can be similarly proved.

— Case (2) Let $p'_1 < p'_2 \in Supp(P \mid X = x')$ and $p_1 < p_2 \in Supp(P \mid X = x)$ such that $[p'_1, p'_2] \subseteq [p_1, p_2]$.

$$\begin{aligned} h(1, x, x', p_1, p_2, p'_1, p'_2) &= (P(1, 1|x', p'_2) - P(1, 0|x', p'_1)) \\ &\qquad -(P(1, 0|x, p_1) - P(1, 1|x, p_2)) \\ &= P(u \leq \nu(1, x'), p'_1 < v < p'_2) \\ &\qquad -P(u \leq \nu(0, x), p_1 < v < p_2) \\ &\leq P(u \leq \nu(1, x'), p_1 < v < p_2) \\ &\qquad -P(u \leq \nu(0, x), p_1 < v < p_2). \end{aligned}$$

The last inequality holds since $[p'_1, p'_2] \subseteq [p_1, p_2]$. Therefore, if $h(1, x, x', p_1, p_2, p'_1, p'_2) \geq 0$ then

$$P(u \leq \nu(1, x'), p_1 < v < p_2) - P(u \leq \nu(0, x), p_1 < v < p_2) \geq 0,$$

which implies that $[\nu(1,x') - \nu(0,x)] \geq 0$ since we have the following :

$$\Big[P(u \leq \nu(1,x'), p_1 < v < p_2) - P(u \leq \nu(0,x), p_1 < v < p_2)\Big] =$$

$$\begin{cases} P(\nu(0,x) \leq u \leq \nu(1,x'), p_1 < v < p_2) \text{ if } \nu(1,x') > \nu(0,x) \\ 0 \text{ if } \nu(1,x') = \nu(0,x) \\ -P(\nu(1,x') \leq u \leq \nu(0,x), p_1 < v < p_2) \text{ if } \nu(1,x') < \nu(0,x). \end{cases}$$

— Case (4) Let p and p' belong to $Supp(P \mid X = x) \cap Supp(P \mid X = x')$ such that $p' < p$.

$$\begin{aligned} \tilde{h}_1(x,x',p,p') &= (P(1,1|x',p) - P(1,d|x',p')) - (P(1,1|x,p) - P(1,1|x,p')) \\ &= P(u \leq \nu(d,x'), p' < v < p) - P(u \leq \nu(d,x), p' < v < p), \end{aligned}$$

then

$$\tilde{h}_d(x,x',p,p') = \begin{cases} P(\nu(d,x) \leq u \leq \nu(d,x'), p' < v < p) \text{ if } \nu(d,x') > \nu(d,x) \\ 0 \text{ if } \nu(d,x') = \nu(d,x) \\ -P(\nu(d,x') \leq u \leq \nu(d,x), p' < v < p) \text{ if } \nu(d,x') < \nu(d,x). \end{cases}$$

— Case (6) Let $p'_1 < p'_2 \in Supp(P \mid X = x')$ and $p_1 < p_2 \in Supp(P \mid X = x)$ such that $[p'_1, p'_2] \subseteq [p_1, p_2]$.

$$\begin{aligned} \tilde{h}_1(x,x',p_1,p_2,p'_1,p'_2) &= (P(1,1|x',p'_2) - P(1,0|x',p'_1)) \\ &\qquad -(P(1,1|x,p_2) - P(1,1|x,p_1)) \\ &= P(u \leq \nu(1,x'), p'_1 < v < p'_2) \\ &\qquad -P(u \leq \nu(1,x), p_1 < v < p_2) \\ &\leq P(u \leq \nu(1,x'), p_1 < v < p_2) \\ &\qquad -P(u \leq \nu(1,x), p_1 < v < p_2). \end{aligned}$$

The last inequality holds since $[p'_1, p'_2] \subseteq [p_1, p_2]$. Therefore, if $\tilde{h}_1(x,x',p_1,p_2,p'_1,p'_2) \geq 0$ then

$$P(u \leq \nu(1,x'), p_1 < v < p_2) - P(u \leq \nu(1,x), p_1 < v < p_2) \geq 0,$$

which implies that $[\nu(1,x') - \nu(1,x)] \geq 0$ since we have the following :

$$\Big[P(u \leq \nu(1,x'), p_1 < v < p_2) - P(u \leq \nu(1,x), p_1 < v < p_2)\Big] =$$

$$\begin{cases} P(\nu(1,x) \leq u \leq \nu(1,x'), p_1 < v < p_2) \text{ if } \nu(1,x') > \nu(1,x) \\ 0 \text{ if } \nu(1,x') = \nu(1,x) \\ -P(\nu(1,x') \leq u \leq \nu(1,x), p_1 < v < p_2) \text{ if } \nu(1,x') < \nu(1,x). \end{cases}$$

This completes our proof. □

Proof of proposition 1. Under assumption 1, if $\nu(0,X) < \nu(1,X)$ the model (1.1) can be written in the form of a multi-valued mapping $G_\nu(.,X)$ from unobservable u to observable (Y,D,X,Z) in the following way :

$$\begin{aligned} G_\nu : u &\longmapsto (y,d,x,z) \\ [0,\nu(0,x)] &\longmapsto \{(1,0,x,z);(1,1,x,z)\} \\ [\nu(0,x),\nu(1,x)] &\longmapsto \{(0,0,x,z);(1,1,x,z)\} \\ [\nu(1,x),1] &\longmapsto \{(0,1,x,z);(0,0,x,z)\} \end{aligned}$$

When $\nu(0,X) > \nu(1,X)$ we can derived an analogous multi-valued mapping :

$$\begin{aligned} G_\nu : u &\longmapsto (y,d,x,z) \\ [0,\nu(1,x)] &\longmapsto \{(1,0,x,z);(1,1,x,z)\} \\ [\nu(1,x),\nu(0,x)] &\longmapsto \{(1,0,x,z);(0,1,x,z)\} \\ [\nu(0,x),1] &\longmapsto \{(0,1,x,z);(0,0,x,z)\} \end{aligned}$$

Let call P the distribution of observable variables $(Y,D) \in \{(0,0),(0,1),(1,0),(1,1)\} = (\mathcal{Y},\mathcal{D})$ which can be estimated from the data and U the uniform distribution on $[0,1]$, of the unobservable variable u. The model relating outcomes variables (Y,D) and unobservable variable u is given by :

$$P(\{(Y,D) \in G_\nu(u \mid X,Z)\}) = 1 \text{ X,Z-as for some } \nu$$

such as ν is generally non-unique, which prompts the following definition.

Definition 4. $\nu(d,x) = P(Y_d = 1 \mid X = x)$ *belongs to the identified set if and only if* $P(\{(Y,D) \in G_\nu(u \mid X,Z)\}) = 1$ X,Z *a.s*

According to Theorem 1 of Galichon and Henry (2011), the identified set is equal to the set ν such that the following inequalities hold :

$$\mathbb{P}(A \mid X,Z) \leq U(\{u \mid G_\nu(u \mid X,Z) \cap A \neq \emptyset\}), \forall A \in 2^{(\mathcal{Y},\mathcal{D})} \tag{12}$$

We find exactly 6 non redundant inequalities which can be rewritten as follows :

$$\begin{array}{rcl} P(Y=1,D=0 \mid X,Z) & \leq & \nu(0,X) \\ P(Y=0,D=0 \mid X,Z) & \leq & 1-\nu(0,X) \\ P(Y=1,D=1 \mid X,Z) & \leq & \nu(1,X) \\ P(Y=0,D=1 \mid X,Z) & \leq & 1-\nu(1,X) \\ P(Y=0 \mid X,Z) & \leq & max(1-\nu(0,X),1-\nu(1,X)) \\ P(Y=1 \mid X,Z) & \leq & max(\nu(0,X),\nu(1,X)). \end{array}$$

□

This completes our proof.

Proof of theorem 1.

Definition 5. *(Nelsen 2006) A two-dimentional subcopula (or briefly subcopula) is a function C with the following properties :*

1. *Domain(C)=$D_1 \times D_2$ where D_1 and D_2 are subsets of $[0,1]$ containing 0 and 1.*
2. *$C(u_1,v_1)-C(u_1,v_2)-C(u_2,v_1)+C(u_2,v_2) \geq 0$, for all $u_1,u_2 \in D_1$ and $v_1,v_2 \in D_2$ such that $u_1 \geq u_2$ and $v_1 \geq v_2$.*
3. *$C(u,1)=u$ and $C(1,v)=v$ for all $u \in D_1$ and for all v in D_2.*

Claim 1. *Under assumptions 4 and 5, $\nu(d,x) : \{0,1\} \times Dom(X) \longrightarrow [0,1]$ is in the identified set if and only if there exists a subcopula C whose domain is $S_1 \cup \{0,1\} \times S_2 \cup \{0,1\}$ such that :*

1. *$C(u,0)=C(0,v)=0$, for all $u \in S_1 \cup \{0,1\}$ and for all v in $S_2 \cup \{0,1\}$.*
2. *$C(\nu(1,x),p)=P(1,1|x,p)$ and $C(\nu(0,x),p)=\nu(0,x)-P(1,0|x,p)$ for all $(x,p) \in Dom(X,supp(P|X))$.*

where $S_2=\{p(x,z):(x,z) \in Dom(x,z)\}$.

The proof of this claim is given in Chiburis (2010).

In the main text, we prove that the characterization of the identified set derived in the theorem (1) is equivalent to the characterization derived in the corollary (1). Then to prove the theorem (1) I will prove that the bounds proposed in the corollary (1) are sharp. For the sake of simplicity, I shall use in this section the following notation, $L_0(x,p)=\sup_{\mathbf{\Omega}_{\mathbf{01}}^{-}(x)} \sup_{\mathbf{P}^{-}(x',p)} P(1,1|x',p')$ and $M_0(x,p)=\sup_{\mathbf{\Omega}_{\mathbf{00}}^{-}(x)} \sup_{\mathbf{P}^{-}(x',p)} (SL_0(x')-P(1,0|x',p'))$. Throughout the main text, we showed that $\nu(0,x)$ lies inside the following interval $[SL_0(x), SU_0(x)]$. To show that these bounds are sharp, it is sufficient to construct a subcopula which respects conditions cited in claim 1 when $\nu(0,x)$ equals $SL_0(x)$ or $SU_0(x)$. Then, assume that,

$$\begin{aligned}\nu(0,x) &= SL_0(x)\\ &= \sup_p\{P(1,0|x,p) + \max[L_0(x,p), M_0(x,p)]\},\end{aligned}$$

where the supremum is taken over $Supp(P \mid X)$. I will now show that the following function is a subcopula on domain $S_1 \cup \{0,1\} \times S_2 \cup \{0,1\}$:

$$\begin{aligned}C(\nu(1,x),p) &= P(u \leq \nu(1,x), v \leq p)\\ C(\nu(0,x),p) &= -P(u \leq \nu(0,x), v \geq p) + \sup_p\{P(1,0|x,p)\\ &\qquad + \max[L_0(x,p), M_0(x,p)]\}.\end{aligned}$$

By construction, our function verifies properties (1) and (3) of definition 5, it remains to verify property (2). When $Supp(P \times X) = Supp(P) \times Supp(X)$, property (2) imposes restrictions on $\nu(0,x)$ for all $p, p' \in Supp(P) = Supp(P \mid X) = Supp(P \mid X')$. It's no longer the case when we have $Supp(P \times X) \neq Supp(P) \times Supp(X)$ because of additional data observability constraints. Indeed, $C(\nu(1,x), p(x',z')) = P(u \leq \nu(1,x), v \leq p(x',z'))$ cannot be identified from the data when $p(x',z') \notin Supp(P \mid X)$. So, property (2) doesn't always impose additional testable constraints. To clarify this point, consider the two following situations :

1. $Supp(P \mid X) \cap Supp(P \mid X') = \emptyset$, $u_1 = \nu(0,x)$, $u_2 = \nu(1,x')$, $v_1 = p(x',z')$ and $v_2 = p(x,z)$. Then property 2 doesn't impose additional restrictions on $\nu(0,x)$ since we cannot identify $C(\nu(1,x), p(x',z')) = P(u \leq \nu(1,x), v \leq p(x',z'))$.
2. $Supp(P \mid X) \cap Supp(P \mid X') = \emptyset$, $u_1 = \nu(0,x)$, $u_2 = \nu(1,x')$, $v_1 = p(x,z)$ and $v_2 = p(x',z')$.

The only constraint from property 2 is : $C(\nu(1,x), p(x,z)) \geq C(\nu(1,x'), p(x',z'))$. I now prove in 2 steps, that our proposed function verifies property (2). Before going over these steps, we need a technical result :

Claim 2. *For all* $(u_1, u_2, v_1, v_2) \in [0,1]^4$ *such that* $u_1 \geq u_2$ *and* $v_1 \geq v_2$*, we* *have :*

$$\begin{aligned} P(u_2 \le u \le u_1, v_2 \le v \le v_1) &= (P(u \le u_1, v \le v_1) + P(u \le u_2, v \ge v_1)) \\ &\quad -(P(u \le u_1, v \le v_2) + P(u \le u_2, v \ge v_2)) \\ &\ge 0. \end{aligned}$$

First step : Let $p \in Supp(P \mid X) \cap Supp(P \mid X') \neq \emptyset$

1. Let (x, x') satisfy $\nu(0,x) \ge \nu(1,x')$.

$$\begin{aligned} & C(\nu(0,x),p) - C(\nu(1,x'),p) \\ = \; & -(P(u \le \nu(0,x), v \ge p) + P(u \le \nu(1,x'), v \le p)) \\ & + \sup_{p \in Supp(P|X)} \{P(u \le \nu(0,x), v \ge p) + \max[L_0(x,p), M_0(x,p)]\} \\ \ge \; & -(P(u \le \nu(0,x), v \ge p) + P(u \le \nu(1,x'), v \le p)) \\ & + \sup_{p \in Supp(P|X)} \{P(u \le \nu(0,x), v \ge p) + L_0(x,p)\} \\ \ge \; & -(P(u \le \nu(0,x), v \ge p) + P(u \le \nu(1,x'), v \le p)) \\ & + \sup_{p \in Supp(P|X)} \{P(u \le \nu(0,x), v \ge p) \\ & + \sup_{\Omega_{01}^-(x)} \sup_{P^-(x^*,p)} P(u \le \nu(1,x^*), v \le p^*)\} \\ \ge \; & 0. \end{aligned}$$

The last inequality holds because $p \in Supp(P \mid X)$, $x' \in \Omega_{01}^-(x)$ and $p \in P^-(x', p)$. Also, $C(\nu(0,x),p) - C(\nu(1,x'),p)$ is increasing in p by the first equality. Indeed, by claim 2 $(P(u \le \nu(1,x'), v \le p) + P(u \le \nu(0,x), v \ge p))$ is decreasing in p since $\nu(0,x) \ge \nu(1,x')$. Then, for all $p' < p \in Supp(P \mid X) \cap Supp(P \mid X')$ we have $C(\nu(0,x),p) - C(\nu(1,x'),p) \ge C(\nu(0,x),p') - C(\nu(1,x'),p')$. Thus, $C(\nu(0,x),p) - C(\nu(1,x'),p) - C(\nu(0,x),p') + C(\nu(1,x'),p') \ge 0$. So, the property 2 is verified.

2. Let (x, x') satisfy $\nu(0, x) \leq \nu(1, x')$.

$$\begin{aligned}
& C(\nu(0,x),p) - C(\nu(1,x'),p) \\
= \quad & -(P(u \leq \nu(0,x), v \geq p) + P(u \leq \nu(1,x'), v \leq p)) \\
& + \sup_{p \in Supp(P|X)} \{P(u \leq \nu(0,x), v \geq p) + \max[L_0(x,p), M_0(x,p)]\} \\
\leq \quad & -(P(u \leq \nu(0,x), v \geq p) + P(u \leq \nu(0,x), v \leq p)) \\
& + \sup_{p \in Supp(P|X)} \{P(u \leq \nu(0,x), v \geq p) + \max[L_0(x,p), M_0(x,p)]\} \\
\leq \quad & -P(u \leq \nu(0,x)) \\
& + \sup_{p \in Supp(P|X)} \{P(u \leq \nu(0,x), v \geq p) + \max[L_0(x,p), M_0(x,p)]\} \\
\leq \quad & -P(u \leq \nu(0,x)) + SL_0(x) \\
\leq \quad & 0.
\end{aligned}$$

The first inequality holds because $\nu(0, x) \leq \nu(1, x')$. Also, $C(\nu(0, x), p) - C(\nu(1, x'), p)$ is increasing in p by the first equality. Indeed, by claim 2 $(P(u \leq \nu(1, x'), v \leq p) + P(u \leq \nu(0, x), v \geq p))$ is increasing in p since $\nu(0, x) \leq \nu(1, x')$.

3. Let (x, x') satisfy $\nu(0,x) \geq \nu(0,x')$

$$
\begin{aligned}
& C(\nu(0,x),p) - C(\nu(0,x'),p) \\
= \; & -(P(u \leq \nu(0,x), v \geq p) - P(u \leq \nu(0,x'), v \geq p)) \\
& + \sup_{p \in Supp(P|X)} \{P(u \leq \nu(0,x), v \geq p) + \max[L_0(x,p), M_0(x,p)]\} \\
& - \sup_{p' \in Supp(P|X')} \{P(u \leq \nu(0,x'), v \geq p) + \max[L_0(x',p'), M_0(x',p')]\} \\
\geq \; & -(P(u \leq \nu(0,x), v \geq p) - P(u \leq \nu(0,x'), v \geq p)) \\
& + \sup_{p \in Supp(P|X)} \{P(u \leq \nu(0,x), v \geq p) + M_0(x,p)]\} - SL_0(x') \\
\geq \; & -(P(u \leq \nu(0,x), v \geq p) - P(u \leq \nu(0,x'), v \geq p)) \\
& + \sup_{p \in Supp(P|X)} \{P(u \leq \nu(0,x), v \geq p) \\
& + \sup_{\Omega^{-}_{00}(x)} \sup_{P^{-}(x^*,p)} (SL_0(x^*) - P(u \leq \nu(0,x^*), v \geq p^*))\} \\
& - SL_0(x') \\
\geq \; & -(P(u \leq \nu(0,x), v \geq p) - P(u \leq \nu(0,x'), v \geq p)) \\
& \sup_{p \in Supp(P|X)} \{P(u \leq \nu(0,x), v \geq p) - P(u \leq \nu(0,x'), v \geq p))\} \\
\geq \; & 0.
\end{aligned}
$$

The fourth inequality holds because $x' \in \mathbf{\Omega^{-}_{00}}(x)$ and $p \in \mathbf{P}^{-}(x',p)$. Also, $C(\nu(0,x),p) - C(\nu(0,x'),p)$ is increasing in p by the first equality.

4. Let (x, x') satisfy $\nu(0,x) \leq \nu(0,x')$
by interchanging x by x' in point (3) we obviously get that $C(\nu(0,x),p) - C(\nu(0,x'),p)$ is decreasing in p and greater than 0.

Second step : $Supp(P \mid X) \cap Supp(P \mid X') = \emptyset$

1. $\nu(0,x) \geq \nu(1,x')$ and $p(x,z) \geq p(x',z')$

$$
\begin{aligned}
C(\nu(0,x),p) - C(\nu(1,x'),p') & \geq C(\nu(0,x),p) - C(\nu(1,x'),p) \\
& \geq 0.
\end{aligned}
$$

The last inequality holds by point (1) of first step.

2. $\nu(0,x) \leq \nu(1,x')$ and $p(x,z) \leq p(x',z')$

$$
\begin{aligned}
C(\nu(0,x),p) - C(\nu(1,x'),p') & \leq C(\nu(0,x),p) - C(\nu(1,x'),p) \\
& \leq 0.
\end{aligned}
$$

The last inequality holds by point (2) of first step.

3. $\nu(0,x) \geq \nu(0,x')$ and $p(x,z) \geq p(x',z')$

$$\begin{aligned} C(\nu(0,x),p) - C(\nu(0,x'),p) &\geq C(\nu(0,x),p) - C(\nu(0,x'),p) \\ &\geq 0. \end{aligned}$$

The last inequality holds by point (3) of first step.

4. $\nu(0,x) \leq \nu(0,x')$

$$\begin{aligned} C(\nu(0,x),p) - C(\nu(0,x'),p) &\geq C(\nu(0,x),p) - C(\nu(0,x'),p) \\ &\geq 0. \end{aligned}$$

The last inequality holds by point (4) of first step.

Then the property (2) holds. We can proceed in the same way for $\nu(0,x) = SU_0(x)$. This completes our proof. $\square$

.2 Data summary statistics

TABLE 3 – Summary statistics

	Migrants	Non-migrants	$m_1(E,I,H)$	$m_0(E,I,H)$
{Y=1,Edp=1,Imp=1,Hon=1}	1	0	0.0588	0
{Y=1,Edp=1,Imp=1,Hon=0}	3	0	0.1875	0
{Y=1,Edp=1,Imp=0,Hon=1}	4	2	0.1818	0.2500
{Y=1,Edp=1,Imp=0,Hon=0}	3	1	0.1500	0.0555
{Y=1,Edp=0,Imp=0,Hon=1}	3	3	0.0909	0.2500
{Y=1,Edp=0,Imp=0,Hon=0}	15	15	0.2941	0.2419
{Y=1,Edp=0,Imp=1,Hon=1}	1	0	0.2000	0
{Y=1,Edp=0,Imp=1,Hon=0}	4	1	0.1739	0.0833

.3 Proof of main results in Chapter 2

In all the proofs, we use the notation $\omega = (z, x_1, x_0)$. When there is no ambiguity, we shall write $f_1 = f_1(x_1)$, $f_0 = f_0(x_0)$ and $g = g(\omega)$.

.3.1 Proof of Proposition 2

Validity of the bounds

See main text.

Sharpness of the bounds

To show the sharpness of the joint bounds for $f_1(x_1)$ and $f_0(x_0)$, it is sufficient to construct joint distributions for the unobserved heterogeneity vector (u_0^*, u_1^*) such that each point in the identified region is attained and which is compatible with the observed data in the following sense :

1. $P(u_0^* \leq f_0(x_0), u_1^* \geq u_0^* + f_1(x_1) - f_0(x_0)|x_1, x_0) = P(Y = 1, D = 0|\omega)$,
2. $P(u_1^* \leq f_1(x_1), u_1^* \leq u_0^* + f_1(x_1) - f_0(x_0)|x_1, x_0) = P(Y = 1, D = 1|\omega)$,
3. $P(u_0^* \geq f_0(x_0), u_1^* \geq u_0^* + f_1(x_1) - f_0(x_0)|x_1, x_0) = P(Y = 0, D = 0|\omega)$,
4. $P(u_1^* \geq f_1(x_1), u_1^* \leq u_0^* + f_1(x_1) - f_0(x_0)|x_1, x_0) = P(Y = 0, D = 1|\omega)$.

We construct a joint distribution for (u_0^*, u_1^*) such that $f_1(x_1) = \mathbb{P}(1, 1|\omega) + \alpha_1$ and $f_0(x_0) = \mathbb{P}(1, 0|\omega) + \alpha_0$, for any (α_1, α_0) satisfying $0 \leq \alpha_1 \leq \mathbb{P}(1, 0|\omega)$ and $0 \leq \alpha_0 \leq \mathbb{P}(1, 1\omega)$ and consider the case where $f_0 - f_1 \geq 0$. The case where $f_0 - f_1 \leq 0$ can be derived similarly. We propose the following candidate as a potential joint distribution. Denoting $\delta = \mathbb{P}(1, 0|\omega) + \alpha_0 - \mathbb{P}(1, 1|\omega) - \alpha_1$, let $\mathbb{P}(u_0^* \leq s, u_1^* \leq t) = 0$ if $0 \leq t \leq 1 - \delta, s \leq \delta$, $\mathbb{P}(u_0^* \geq s, u_1^* \geq t) = 0$ if

$t \geq 1-\delta, s \geq \delta$ and

$$\mathbb{P}(\mathbb{P}(1,0|\omega)+\alpha_0 \leq u_0^* \leq s, u_1^* \leq t) = \frac{a(t,s)}{a(\mathbb{P}(1,1|\omega)+\alpha_1,1)}(\mathbb{P}(1,1|\omega)-\alpha_0)$$
if $0 \leq t \leq \mathbb{P}(1,1|\omega)+\alpha_1, s \geq \mathbb{P}(1,0|\omega)+\alpha_0$,

$$\mathbb{P}(u_0^* \leq s, 1-\delta \leq u_1^* \leq t) = \frac{(t-(1-\delta))s}{\delta} \text{ if } t \geq 1-\delta, s \leq \delta,$$

$$\mathbb{P}(\delta \leq u_0^* \leq s, \mathbb{P}(1,1|\omega)+\alpha_1 \leq u_1^* \leq t) = \frac{b(t,s)}{b(1-\delta,\mathbb{P}(1,0|\omega)+\alpha_0))}(\mathbb{P}(1,1|\omega)-\alpha_0)$$
if $\mathbb{P}(1,1|\omega)+\alpha_1 \leq t \leq 1-\delta, \delta \leq s \leq \mathbb{P}(1,0|\omega)+\alpha_0$,

$$\mathbb{P}(\delta+u_1^* \leq u_0^* \leq s, u_1^* \leq t) = \frac{c(t,s)}{c(\mathbb{P}(1,1|\omega)+\alpha_1,\mathbb{P}(1,0|\omega)+\alpha_0))}\alpha_0$$
if $0 \leq t \leq \mathbb{P}(1,1|\omega)+\alpha_1, \delta+u_1^* \leq s \leq \mathbb{P}(1,0|\omega)+\alpha_0$,

$$\mathbb{P}(\delta \leq u_0^* \leq u_1^*+\delta, u_1^* \leq t) = t - \frac{c(t,\mathbb{P}(1,0|\omega)+\alpha_0)}{c(\mathbb{P}(1,1|\omega)+\alpha_1,\mathbb{P}(1,0|\omega)+\alpha_0))}\alpha_0$$
$$-\frac{a(t,1)}{a(\mathbb{P}(1,1|\omega)+\alpha_1,1)}(\mathbb{P}(1,1|\omega)-\alpha_0)$$
if $0 \leq t \leq \mathbb{P}(1,1|\omega)+\alpha_1, \delta+u_1^* \leq s \leq \mathbb{P}(1,0|\omega)+\alpha_0$

$$\mathbb{P}(\delta \leq u_0^* \leq s, s-\delta \leq u_1^* \leq t) = \frac{(s-\delta)t}{\mathbb{P}(1,1|\omega)+\alpha_1)} - \mathbb{P}(\delta \leq u_0^* \leq s, u_1^* \leq s-\delta)$$
$$-\frac{b(t,s)}{b(1-\delta,\mathbb{P}(1,0|\omega)+\alpha_0))}(\mathbb{P}(1,1|\omega)-\alpha_0),$$
if $s-\delta \leq t \leq \mathbb{P}(1,1|\omega)+\alpha_1, \delta \leq s \leq \mathbb{P}(1,0|\omega)+\alpha_0$,

$$\mathbb{P}(u_1^*+\delta \leq u_0^* \leq s, \mathbb{P}(1,1|\omega)+\alpha_1 \leq u_1^* \leq t) = \frac{d(t,s)}{d(1-\delta,1)}\mathbb{P}(0,1|\omega)$$
if $\mathbb{P}(1,1|\omega)+\alpha_1 \leq t \leq 1-\delta, u_1^*+\delta \leq s \leq 1$,

$$\mathbb{P}(\mathbb{P}(1,0|\omega)+\alpha_0 \leq u_0^* \leq u_1^*+\delta, \mathbb{P}(1,1|\omega)+\alpha_1 \leq u_1^* \leq t) = (t-\mathbb{P}(1,1|\omega)-\alpha_1)$$
$$-\frac{d(t,1)}{d(1-\delta,1)}\mathbb{P}(0,1|\omega) - \frac{b(t,\mathbb{P}(1,0|\omega)+\alpha_0)}{b(1-\delta,\mathbb{P}(1,0|\omega)+\alpha_0))}(\mathbb{P}(1,1|\omega)-\alpha_0)$$
if $\mathbb{P}(1,1|\omega)+\alpha_1 \leq t \leq 1-\delta$,

$$\mathbb{P}(\mathbb{P}(1,0|\omega)+\alpha_0 \leq u_0^* \leq s, s-\delta \leq u_1^* \leq t) = \frac{(s-\mathbb{P}(1,0|\omega)+\alpha_0)t}{1-\delta}$$
$$-\mathbb{P}(\mathbb{P}(1,0|\omega)+\alpha_0 \leq u_0^* \leq s, \mathbb{P}(1,0|\omega)+\alpha_0 \leq u_1^* \leq s-\delta)$$
$$-\frac{a(\mathbb{P}(1,1|\omega)+\alpha_1,s)}{a(\mathbb{P}(1,1|\omega)+\alpha_1,1)}(\mathbb{P}(1,1|\omega)-\alpha_0)$$
if $s-\delta \leq t \leq 1-\delta, \mathbb{P}(1,0|\omega)+\alpha_0 \leq s \leq 1$,

where

$$\begin{aligned} a(t,s) &= t(s-(\mathbb{P}(1,0|\omega)+\alpha_0)), \\ b(t,s) &= (t-(\mathbb{P}(1,1|\omega)+\alpha_1))(s-\delta), \\ c(t,s) &= \frac{1}{2}[s-(\delta)+(s-(t+(\delta)))]t, \\ d(t,s) &= \frac{1}{2}[(s-(\mathbb{P}(1,0|\omega)+\alpha_0)+(s-(t+(\delta)))](t-(\mathbb{P}(1,1|\omega)+\alpha_1)) \end{aligned}$$

The proof is complete upon verifying that this function is a joint distribution such as the marginals are uniform distribution over [0,1] and which is compatible with the observed data (i.e., respects Conditions 1 to 4) when $f_1(x_1)=\mathbb{P}(1,1|\omega)+\alpha_1$ and $f_0(x_0)=\mathbb{P}(1,0|\omega)+\alpha_0$, for any (α_1,α_0) satisfying $0\leq\alpha_1\leq\mathbb{P}(1,0|\omega)$ and $0\leq\alpha_0\leq\mathbb{P}(1,1\omega)$.

.3.2 Proof of Proposition 3

To show validity of the bounds, we drop all the conditioning variables $\omega=(z,x_1,x_0)$ from the notation. We have $D=1\Rightarrow Y_0^*+g\leq Y_1^*\Rightarrow 1\{Y_0^*+g\geq 0\}\leq 1\{Y_1^*\geq 0\}\Rightarrow 1\{Y_0^*+g\geq 0\}1\{D=1\}\leq 1\{Y_1^*\geq 0\}1\{D=1\}\Rightarrow \mathbb{E}[1\{Y_0^*+g\geq 0\}|D=1]\leq \mathbb{E}[1\{Y_1^*\geq 0\}|D=1]\Rightarrow \mathbb{E}[1\{Y_0^*+g\geq 0\}|D=1]\leq \mathbb{E}[Y_1|D=1]$. We can easily derive equivalent inequalities when $D=0$. Hence, if $D=1\{Y_1^*>Y_0^*+g\}$ then $\mathbb{E}[1\{Y_0^*+g\geq 0\}|D=1]\leq\mathbb{E}[Y_1|D=1]$ and $\mathbb{E}[Y_1|D=0]\leq\mathbb{E}[1\{Y_0^*+g\geq 0\}|D=0]$. Hence, when $g\geq 0$, $\mathbb{E}[Y_0|D=1]\leq \mathbb{E}[Y_1|D=1]$ and when $g\leq 0$, $\mathbb{E}[Y_1|D=0]\leq\mathbb{E}[Y_0|D=0]$. Finally, if $g=0$ we have $\mathbb{E}[Y_d|D=d]\geq\mathbb{E}[Y_d|D=1-d]$ where $d\in\{0,1\}$. Those inequalities allow us to construct the sharp bounds for f_1 and f_0 in the case where $D=1\{Y_1^*>Y_0^*+g\}$. Indeed, $f_1=\mathbb{E}[Y_1]=\mathbb{E}[Y_1,D=1]+\mathbb{E}[Y_1|D=0]P(D=0)$ and $f_0=\mathbb{E}[Y_0]=\mathbb{E}[Y_0,D=0]+\mathbb{E}[Y_0|D=1]P(D=1)$. Now, if $g\geq 0$, then $P(Y=1,D=1)\leq f_1\leq P(Y=1,D=1)+P(D=0)$ and $P(Y=1,D=0)\leq f_0\leq P(Y=1)$. On the other hand, if $g\leq 0$, $P(Y=1,D=1)\leq f_1\leq P(Y=1)$ and $P(Y=1,D=0)\leq f_0\leq P(Y=1,D=0)+P(D=1)$. Finally, $f_0=\mathbb{E}[1\{u_0\leq f_0\}1\{u_1\geq u_0+f_1-f_0-g\}]+\mathbb{E}[1\{u_0\leq f_0\}1\{u_1\leq u_0+f_1-f_0-g\}]$. Hence, if $g\geq f_1$, then $\{u_1\leq u_0+f_1-f_0-g\}\Rightarrow\{u_0\geq f_0\}$ and $f_0(X_0)\leq\mathbb{E}[1\{u_0\leq f_0\}1\{u_1\geq u_0+f_1-f_0-g\}]+\mathbb{E}[1\{u_0\leq f_0\}1\{u_0\geq f_0\}]\leq\mathbb{E}[1\{u_0\leq f_0\}1\{u_1\geq u_0+f_1-f_0-g\}]=P(Y=1,D=0)$.

Now the bounds for g can be obtained as follows.

— If $g+f_0-f_1\geq 0$ and $g\leq f_1$, then $\{u_0\leq g+f_0-f_1\}\Rightarrow\{u_0\leq u_1+g+f_0-f_1\}$ and $\{u_0\leq g+f_0-f_1\}\Rightarrow\{u_0\leq f_0\}$. So $\{u_0\leq g+f_0-f_1\}\Rightarrow$

$\{u_0 \leq u_1 + g + f_0 - f_1\} \cap \{u_0 \leq f_0\}$. Hence $g + f_0 - f_1 = P(u_0 \leq g+f_0-f_1) \leq P(\{u_0 \leq u_1+g+f_0-f_1\} \cap \{u_0 \leq f_0\}) = P(Y=1, D=0)$.

— If $g + f_0 - f_1 \geq 0$ and $g \geq f_1$, then $\{u_0 \leq g + f_0 - f_1\} \Rightarrow \{u_0 \leq u_1 + g + f_0 - f_1\}$, hence $g + f_0 - f_1 = P(u_0 \leq g + f_0 - f_1) \leq P(\{u_0 \leq u_1 + g + f_0 - f_1\}) = P(D=0)$. As $f_0 = P(Y=1, D=0)$ we have $g - f_1 \leq P(Y=0, D=0)$.

— If $g + f_0 - f_1 \leq 0$ and $g \geq -f_0$, then by similar arguments, we have $g + f_0 - f_1 \geq -P(Y=1, D=1)$.

— If $g + f_0 - f_1 \leq 0$ and $g \leq -f_0$, then $g + f_0 \geq -P(Y=0, D=1)$.

Finally, the validity of bounds $f_1(x_1) - \mathbb{P}(1,1|\omega) - \mathbb{P}(1,0|\omega) - \mathbb{P}(0,1|\omega) \leq g(\omega) \leq \mathbb{P}(1,1|\omega) + \mathbb{P}(1,0|\omega) + \mathbb{P}(0,0|\omega) - f_0(x_0)$ is shown formally in the main text. This completes the proof.

.3.3 Proof of Proposition 4

As previously, our method consists in constructing joint distributions for (u_1, u_0) such that all points of the identified set for (f_1, f_0) are attained. All points in the identified set of Proposition 2 can be attained as shown in the proof of Proposition 2.

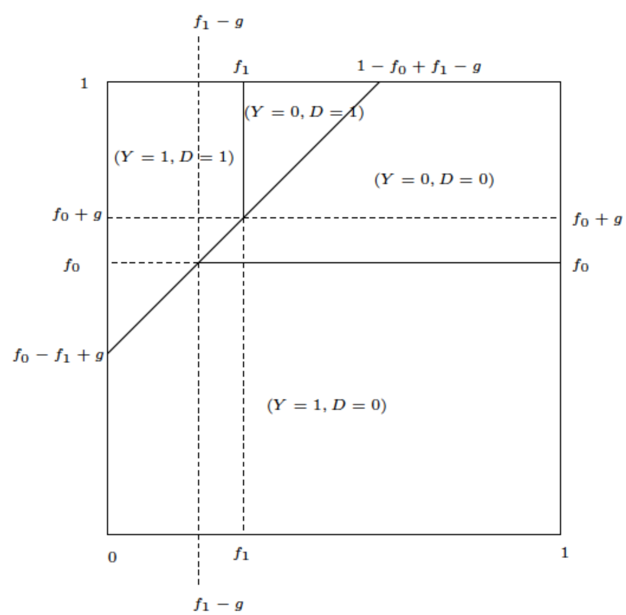
There remains to show that all points in the rectangle with corners $(\mathbb{P}(1,1|\omega)+\mathbb{P}(1,0|\omega)+\min(0, \mathbb{P}(0,0|\omega)-\varepsilon), \mathbb{P}(1,0|\omega))$, $(\mathbb{P}(1,1|\omega)+\mathbb{P}(1,0|\omega)+\min(0, \mathbb{P}(0,0|\omega) - \varepsilon), \mathbb{P}(1,0|\omega) + \mathbb{P}(1,1|\omega))$, $(\mathbb{P}(1,1|\omega) + \mathbb{P}(1,0|\omega), \mathbb{P}(1,0|\omega))$ and $(\mathbb{P}(1,1|\omega)+\mathbb{P}(1,0|\omega), \mathbb{P}(1,0|\omega)+\mathbb{P}(1,1|\omega))$ for $\varepsilon > 0$ arbitrarily small (and symmetrically all points in the rectangle with corners $(\mathbb{P}(1,1|\omega), \mathbb{P}(1,0|\omega) + \mathbb{P}(1,1|\omega)+\min(0, \mathbb{P}(0,1|\omega)-\varepsilon))$, $(\mathbb{P}(1,1|\omega)+\mathbb{P}(1,0|\omega), \mathbb{P}(1,0|\omega)+\mathbb{P}(1,1|\omega)+\min(0, \mathbb{P}(0,1|\omega)-\varepsilon))$, $(\mathbb{P}(1,1|\omega)+\mathbb{P}(1,0|\omega), \mathbb{P}(1,0|\omega)+\mathbb{P}(1,1|\omega))$ and $(\mathbb{P}(1,1|\omega)+\mathbb{P}(1,0|\omega), \mathbb{P}(1,0|\omega) + \mathbb{P}(1,1|\omega)))$.

Compatibility between the joint distribution and the observed data can be expressed as follows :

1. $P(u_0^* \leq f_0(x_0), u_1^* \geq u_0^* + f_1(x_1) - f_0(x_0) - g(\omega)|\omega) = P(Y=1, D=0|\omega)$,
2. $P(u_1^* \leq f_1(x_1), u_1^* \leq u_0^* + f_1(x_1) - f_0(x_0) - g(\omega)|\omega) = P(Y=1, D=1|\omega)$,
3. $P(u_0^* \geq f_0(x_0), u_1^* \geq u_0^* + f_1(x_1) - f_0(x_0) - g(\omega)|\omega) = P(Y=0, D=0|\omega)$,
4. $P(u_1^* \geq f_1(x_1), u_1^* \leq u_0^* + f_1(x_1) - f_0(x_0) - g(\omega)|\omega) = P(Y=0, D=1|\omega)$,

The method of proof is illustrated in Figure 4. Assume that $\mathbb{P}(0,0|\omega) > 0$ (otherwise the rectangle treated below collapses). We construct a joint distribution for (u_1, u_0) such that $f_1(x_1) = (\mathbb{P}(1,1|\omega)+\mathbb{P}(1,0|\omega)+\mathbb{P}(0,0|\omega) - \alpha_1$ and $f_0(x_0) = \mathbb{P}(1,0|\omega)+\alpha_0$, for any (α_1, α_0) satisfying $0 < \alpha_1 \leq \mathbb{P}(0,0|\omega)$ and $0 \leq \alpha_0 \leq \mathbb{P}(1, 1\omega)$.

FIGURE 4 – Characterization of the empirical content of the extended binary Roy model in the unit square of the (u_1, u_0) space in case $f_0(x_0) > f_1(x_1)$, $0 < g(\omega) < f_1(x_1)$ and $f_0(x_0) + g(\omega) < 1$.



There remains to verify that this function is defines a joint distribution which is compatible with the observed data (i.e respects conditions 1 to 4) and such that $f_1(x_1) = P(u_1^* \leq f_1(x_1)|x_1) = P(Y = 1, D = 1|\omega) + P(D = 0|\omega)$ and $g(\omega) + f_0(x_0) - f_1(x_1) = P(u_0^* \leq g(\omega) + f_0(x_0) - f_1(x_1)|\omega) = P(Y = 1, D = 0|\omega)$, when $g(\omega)$ is set equal to $\alpha_1 - f_0(x_0) + f_1(x_1)$.

Symmetrically, we can show that any point in the rectangle with corners $(\mathbb{P}(1,1|\omega), \mathbb{P}(1,0|\omega) + \mathbb{P}(1,1|\omega) + \min(0, \mathbb{P}(0,1|\omega) - \varepsilon))$, $(\mathbb{P}(1,1|\omega) + \mathbb{P}(1,0|\omega), \mathbb{P}(1,0|\omega)+\mathbb{P}(1,1|\omega)+\min(0, \mathbb{P}(0,1|\omega)-\varepsilon))$, $(\mathbb{P}(1,1|\omega)+\mathbb{P}(1,0|\omega), \mathbb{P}(1,0|\omega)+\mathbb{P}(1,1|\omega))$ and $(\mathbb{P}(1,1|\omega)+\mathbb{P}(1,0|\omega), \mathbb{P}(1,0|\omega)+\mathbb{P}(1,1|\omega))$ can be attained and this completes the Proof.

.3.4 Proof of Proposition 5

The proof is exactly identical to the that of Proposition 4 except that $h(u_1, \omega)$ can be chosen as in Figure 2.5 so that all the mass $\mathbb{P}(0,0|\omega)$ can be shifted on the left of $f_1(x_1)$ and therefore we can no longer restrict α_1 to be strictly positive. The case $\alpha_1 = 0$ is also attained. The result follows immediately.

.3.5 Proof of Theorem 2

Under Assumptions 9-11, the model can be equivalently written $(Y, D) \in G((u_1, u_0)|W)$ almost surely conditionally on $W = (Z, X_1, X_0)$, where G is a multi-valued mapping, which to (u_1, u_0) associates $(y, d) = G((u_1, u_0)|W) = \{(1,1),(1,0)\}$ if $u_1 \leq f_1(x_1)$ and $u_0 \leq f_0(x_0)$, $\{(0,1),(1,0)\}$ if $u_1 > f_1(x_1)$ and $u_0 \leq f_0(x_0)$, $\{(1,1),(0,0)\}$ if $u_1 \leq f_1(x_1)$ and $u_0 > f_0(x_0)$ and $\{(0,1),(0,0)\}$ if $u_1 > f_1(x_1)$ and $u_0 > f_0(x_0)$. Hence Theorem 1 of Galichon and Henry (2011), applies and the empirical content of the model is characterized by the collection of inequalities $P(A|W) \leq P((u_1, u_0) : \; G((u_1, u_0)|W)$ hits $A|W)$ for each subset A of $\{(0,0),(0,1),(1,0),(1,1)\}$ (i.e., 16 inequalities). The only non redundant inequalities are $P(1,1|W) \leq f_1(X_1)$, $P(1,0|W) \leq f_0(X_0)$, $P(0,1|W) \leq 1 - f_1(X_1)$, $P(0,0|W) \leq 1 - f_0(X_0)$, $P(Y=0|W) \leq 1 - P(u_1 \leq f_1(X_1), u_0 \leq f_0(X_0)|X_1, X_0)$, $P(Y=1|W) \leq 1 - P(u_1 > f_1(X_1), u_0 > f_0(X_0)|X_1, X_0)$, $P(0,0|W) + P(1,1|W) \leq P(u_1 \leq f_1(X_1), u_0 \leq f_0(X_0)|X_1, X_0) + P(u_0 > f_0(X_0)|X_0)$ and $P(0,1|W) + P(1,0|W) \leq P(u_1 \leq f_1(X_1), u_0 \leq f_0(X_0)|X_1, X_0) + P(u_1 > f_1(X_1)|X_1)$. After some manipulation, the result follows.

.3.6 Proof of Corollary 6

We show that the bounds (2.7) for g remain valid. We drop conditioning variables from the notation throughout this section.

- If $g + v + f_0 - f_1 \geq 0$ and $g + v \leq f_1$, then $\{u_0 \leq g + v + f_0 - f_1\} \Rightarrow \{u_0 \leq u_1 + g + v + f_0 - f_1\}$ and $\{u_0 \leq g + v + f_0 - f_1\} \Rightarrow \{u_0 \leq f_0\}$. So $\{u_0 \leq g + v + f_0 - f_1\} \Rightarrow \{u_0 \leq u_1 + g + v + f_0 - f_1\} \cap \{u_0 \leq f_0\}$. Therefore $P(u_0 - v \leq g + f_0 - f_1) \leq P(\{u_0 \leq u_1 + g + v + f_0 - f_1\} \cap \{u_0 \leq f_0\}) = P(Y = 1, D = 0)$.
- If $g + v + f_0 - f_1 \geq 0$ and $g + v \geq f_1$, then $\{u_0 \leq g + v + f_0 - f_1\} \Rightarrow \{u_0 \leq u_1 + g + v + f_0 - f_1\}$. Therefore $P(u_0 - v \leq g + f_0 - f_1) \leq$

$P(\{u_0 \leq u_1 + g + v + f_0 - f_1\}) = P(D = 0)$.

— If $g+v+f_0-f_1 \leq 0$ and $g+v \geq -f_0$, then $\{u_1 \leq f_1 - f_0 - g - v\} \Rightarrow \{u_1 \leq u_0 + f_1 - f_0 - g - v\}$ and $\{u_1 \leq f_1 - f_0 - g - v\} \Rightarrow \{u_1 \leq f_1\}$. So $\{u_1 \leq f_1 - f_0 - g - v\} \Rightarrow \{u_1 \leq u_0 + f_1 - f_0 - g - v\} \cap \{u_1 \leq f_1\}$. Therefore $P(u_1+v \leq f_1-f_0-g) \leq P(\{u_1 \leq u_0+f_1-f_0-g-v\}\cap\{u_1 \leq f_1\}) = P(Y = 1, D = 1)$.

— If $g+v+f_0-f_1 \leq 0$ and $g+v \leq -f_0$, then $\{u_1 \leq f_1 - f_0 - g - v\} \Rightarrow \{u_1 \leq u_0 + f_1 - f_0 - g - v\}$. Hence $P(u_1 + v \leq f_1 - f_0 - g) \leq P(u_1 \leq u_0 + f_1 - f_0 - g - v) = P(D = 1)$.

Now, since $v \perp\!\!\!\perp (u_0, u_1)$, we have : $P(u_0 \leq g + v + f_0 - f_1) = \mathbb{E}_v[\mathbb{E}[1\{u_0 \leq g + v + f_0 - f_1\}|v]] = \mathbb{E}_v[g + v + f_0 - f_1] = g + f_0 - f_1$. Then, we get the following :

— If $g+v+f_0-f_1 \geq 0$ and $g+v \leq f_1$, then $g+f_0-f_1 \leq P(Y = 1, D = 0)$.

— If $g + v + f_0 - f_1 \geq 0$ and $g + v \geq f_1$, then $g - f_1 \leq P(Y = 0, D = 0)$.

— If $g + v + f_0 - f_1 \leq 0$ and $g + v \geq -f_0$, then $g + f_0 - f_1 \geq -P(Y = 1, D = 1)$.

— If $g+v+f_0-f_1 \leq 0$ and $g+v \leq -f_0$, then $g+f_0 \geq -P(Y = 0, D = 1)$.

which completes the proof.

.3.7 Proof of Corollary 7

Our goal here is to show that the following bounds are sharp for f_0 and f_1.

$$\begin{aligned} P(Y = 1, D = 1|\omega) &\leq f_1(x_1) \leq P(Y = 1, D = 1|\omega) + P(D = 0|\omega) \\ P(Y = 1, D = 0 \mid \omega) &\leq f_0(x_0) \leq P(Y = 1, D = 0|\omega) + P(D = 1|\omega). \end{aligned}$$

The previous results show that the lower bounds are sharp. Now, to show that these bounds are sharp for $f_1(x_1)$ it is sufficient to construct a joint distribution (u_0^*, u_1^*) such that $f_1(x_1)$ equals $P(Y = 1, D = 1|\omega)+P(D = 0|\omega)$ and $f(0, x) = P(Y = 1, D = 0|\omega) + P(D = 1|\omega)$ and which is compatible with the observed data in the following sense :

1. $P(u_0^* \leq f_0(x_0), u_1^* \geq u_0^* + f_1(x_1) - f_0(x_0) - g(\omega) - v|\omega) = P(Y = 1, D = 0|\omega)$
2. $P(u_1^* \leq f_1(x_1), u_1^* \leq u_0^* + f_1(x_1) - f_0(x_0) - g(\omega) - v|\omega) = P(Y = 1, D = 1|\omega)$
3. $P(u_0^* \geq f_0(x_0), u_1^* \geq u_0^* + f_1(x_1) - f_0(x_0) - g(\omega) - v|\omega) = P(Y = 0, D = 0|\omega)$

4. $P(u_1^* \geq f_1(x_1), u_1^* \leq u_0^* + f_1(x_1) - f_0(x_0) - g(\omega) - v|\omega) = P(Y = 0, D = 1|\omega)$

Define the following joint distribution (u_0^*, u_1^*, v^*) such that $u_0^* + u_1^* \leq f_1(x_1) + f_0(x_0)$ and $2v^* = 3u_0^* - 3u_1^* - 3f_0(x_0) + 3f_1(x_1) - 2g(\omega)$. Under the condition that $u_0^* + u_1^* \leq f_1(x_1) + f_0(x_0)$, we have $\{u_1^* \geq u_0^* + f_1(x_1) - f_0(x_0) - g(\omega) - v^*\} \Rightarrow \{u_1^* \leq f_1(x_1)\}$ and $\{u_1^* \leq u_0^* + f_1(x_1) - f_0(x_0) - g(\omega) - v^*\} \Rightarrow \{u_0^* \leq f_0(x_0)\}$. Hence,

$$
\begin{aligned}
f_1(x_1) &= P(u_1^* \leq f_1(x_1), u_1^* \leq u_0^* + f_1(x_1) - f_0(x_0) - g(\omega) - v|\omega) \\
&\qquad + P(u_1^* \leq f_1(x_1), u_1^* \geq u_0^* + f_1(x_1) - f_0(x_0) - g(\omega) - v|\omega) \\
&= P(u_1^* \leq f_1(x_1), u_1^* \leq u_0^* + f_1(x_1) - f_0(x_0) - g(\omega) - v|\omega) \\
&\qquad + P(u_1^* \geq u_0^* + f_1(x_1) - f_0(x_0) - g(\omega) - v|\omega) \\
&= P(Y = 1, D = 1|\omega) + P(D = 0|\omega).
\end{aligned}
$$

With the same strategy, we can also show $f_0(x_0) = P(Y = 1, D = 0|\omega) + P(D = 1|\omega)$.

.4 Proof of main results in Chapter 3

In all that follows, for ease of notation, we drop the conditioning variables and write $\Pi_i(Y_{3-i} = j, X_i) = \Pi_i(j)$ for $i = 1, 2$ and $j = 1, 0$ and $p_{j_1 j_2} = P(Y_1 = j_1, Y_2 = j_2|X_1, X_2)$. We shall also use the following symmetries in the game. All results concerning the second asymmetric game can be obtained from results concerning the first asymmetric game after permutation of the two players. The coordination game is obtained from the duopoly entry game by relabeling. Hence all results for the coordination game can be obtained from the results for the duopoly entry game with the following conversion table : $\Pi_1(0)$ in the duopoly entry case is replaced by $1 - \Pi_1(1)$ and vice-versa. $\Pi_2(1)$ is replaced by $\Pi_2(0)$ and vice-versa. Σ_{j_1, j_2} is replaced by Σ_{1-j_1, j_2}. $P(j_1, j_2)$ is replaced by $P(1 - j_1, j_2)$. Finally, σ_1 is replaced by $1 - \sigma_1$.

.4.1 Proof of Theorem 3

Dropping all explanatory variables from the notation, the equilibrium correspondence $\varepsilon = (\varepsilon_1, \varepsilon_2) \mapsto G(\varepsilon)$, namely the set of all Nash equilibria in mixed strategies, for a given value of $\varepsilon = (\varepsilon_1, \varepsilon_2)$, is represented in Figure 3.1 and formally defined by $G(\varepsilon) = \{(0, 0)\}$ if $\varepsilon_i > \Pi_i(0)$, $i = 1, 2$, $G(\varepsilon) = \{(1, 1)\}$

if $\varepsilon_i < \Pi_i(1)$, $i = 1, 2$, $G(\varepsilon) = \{(\sigma_1, \sigma_2)\}$ if $\Pi_i(1) < \varepsilon_i < \Pi_i(0)$, $i = 1, 2$, $G(\varepsilon) = \{(1, 0)\}$ if $\varepsilon_1 < \Pi_1(1)$ $\varepsilon_2 > \Pi_2(1)$ or $\varepsilon_1 < \Pi_1(0)$ $\varepsilon_2 > \Pi_2(0)$, with the convention that a degenerate mixed strategy is denoted as its realization.

For almost all values of ε, there is at most one equilibrium in non degenerate mixed strategies. Hence, by Lemma 2 of Galichon and Henry (2011),, the game has a Shapley regular core (see for instance Definition 9 of Galichon and Henry (2011),) and Theorem 5 of Galichon and Henry (2011), applies. The identified set for payoff functions and type distributions is therefore characterized by $P(B) \leq \int \left(\max_{\sigma \in G(\epsilon)} \sigma(B)\right) dC(\epsilon)$, for all subsets B of the set of realized decision profiles $\{(0, 1), (1, 0), (0, 0), (1, 1)\}$. This induces the following inequalities.

$$
\begin{aligned}
P(1,1) &\leq C(\Pi_1(1), \Pi_2(1)) + \int_\Delta \sigma_1(u_2)\sigma_2(u_1) dC(u_1, u_2), \qquad (13)\\
P(0,0) &\leq 1 - \Pi_1(0) - \Pi_2(0) + C(\Pi_1(0), \Pi_2(0))\\
&\quad + \int_\Delta (1 - \sigma_1(u_2))(1 - \sigma_2(u_1)) dC(u_1, u_2), \qquad (14)\\
P(0,1) &\leq \Pi_2(0) - C(\Pi_1(1), \Pi_2(0)),\\
P(1,0) &\leq \Pi_1(0) - C(\Pi_1(0), \Pi_2(1)),
\end{aligned}
$$

$$
\begin{aligned}
P(1,1) &\geq C(\Pi_1(1), \Pi_2(1)) \qquad (15)\\
P(0,0) &\geq 1 - \Pi_1(0) - \Pi_2(0) + C(\Pi_1(0), \Pi_2(0)) \qquad (16)\\
P(0,1) &\geq \Pi_2(0) - C(\Pi_1(1), \Pi_2(1)) - [C(\Pi_1(0), \Pi_2(0)) - C(\Pi_1(0), \Pi_2(1))], \qquad (17)\\
P(1,0) &\geq \Pi_1(0) - C(\Pi_1(1), \Pi_2(1)) - [C(\Pi_1(0), \Pi_2(0)) - C(\Pi_1(1), \Pi_2(0))], \qquad (18)
\end{aligned}
$$

and

$$\begin{aligned}
P(0,0)+P(0,1) &\leq 1-\Pi_1(0)+[C(\Pi_1(0),\Pi_2(0))-C(\Pi_1(1),\Pi_2(0))], && (19)\\
P(0,0)+P(1,0) &\leq 1-\Pi_2(0)+[C(\Pi_1(0),\Pi_2(0))-C(\Pi_1(0),\Pi_2(1))], && (20)\\
P(1,1)+P(0,1) &\leq \Pi_2(0)+[C(\Pi_1(1),\Pi_2(1))-C(\Pi_1(1),\Pi_2(0))], && (21)\\
P(1,1)+P(1,0) &\leq \Pi_1(0)+[C(\Pi_1(1),\Pi_2(1))-C(\Pi_1(0),\Pi_2(1))], && (22)\\
P(1,0)+P(0,1) &\leq \Pi_1(0)+\Pi_2(0)-C(\Pi_1(1),\Pi_2(1))-C(\Pi_1(0),\Pi_2(0)), && (23)\\
P(0,0)+P(1,1) &\leq C(\Pi_1(1),\Pi_2(1))+\int_\Delta \sigma_1(u_2)\sigma_2(u_1)dC(u_1,u_2)\\
&\quad +1-\Pi_1(0)-\Pi_2(0)+C(\Pi_1(0),\Pi_2(0))\\
&\quad +\int_\Delta (1-\sigma_1(u_2))(1-\sigma_2(u_1))dC(u_1,u_2). && (24)
\end{aligned}$$

Now, we will show that (19)-(24) are redundant. (15) and (18) jointly imply that $P(1,1)+P(1,0) \geq \Pi_1(0)-[C(\Pi_1(0),\Pi_2(0))-C(\Pi_1(1),\Pi_2(0))]$ so that $1-\Pi_1(0)+[C(\Pi_1(0),\Pi_2(0))-C(\Pi_1(1),\Pi_2(0))] \geq 1-P(1,1)-P(1,0)$, hence (19) holds. Similarly, (15) and (17) imply (20), (16) and (18) imply (21), (16) and (17) imply (22), (15) and (16) imply (23) and finally (13) and (14) imply (24). The result follows.

.4.2 Proof of Theorem 5

Duopoly entry case

Consider first the duopoly entry case, with $\Pi_i(0) \geq \Pi_i(1)$, $i=1,2$. The bounds are shown to hold in the main text as a corollary of Theorem 4. We show now that the bounds are jointly sharp. To do so, take any given true frequency profile $(p_{11},p_{10},p_{01},p_{00})$ and exhibit a joint distribution $C(\varepsilon_1,\varepsilon_2)$ and an equilibrium selection mechanism $\delta \in [0,1]$ (denoting the probability that $(Y_1=1,Y_2=0)$ is selected in the region of multiplicity) such that all Π can be rationalized.

Construction of the joint distribution

We construct the joint distribution in the following way. Assume $P(\varepsilon_1 \leq \Pi_1(1),\varepsilon_2 \leq \Pi_2(1)) = p_{11}$ and $P(\varepsilon_1 \geq \Pi_1(0),\varepsilon_2 \geq \Pi_2(0)) = p_{00}$. From the marginal constraints, $P(\varepsilon_1 \leq \Pi_1(1)) = \Pi_1(1)$ and $P(\varepsilon_1 \geq \Pi_1(0) = 1-\Pi_1(0)$.

Hence we can choose s and t in $[0,1]$ such that the following hold.

$$\begin{aligned}
&P(\varepsilon_1 \leq \Pi_1(1), \varepsilon_2 \geq \Pi_2(0)) = (1-s)(\Pi_1(1)-p_{11}),\\
&P(\varepsilon_1 \leq \Pi_1(1), \Pi_2(1) \leq \varepsilon_2 \leq \Pi_2(0)) = s(\Pi_1(1)-p_{11}),\\
&P(\varepsilon_1 \geq \Pi_1(0), \varepsilon_2 \leq \Pi_2(1)) = (1-t)(1-p_{00}-\Pi_1(0)),\\
&P(\varepsilon_1 \geq \Pi_1(0), \Pi_2(1) \leq \varepsilon_2 \leq \Pi_2(0)) = t(1-p_{00}-\Pi_1(0)).
\end{aligned}$$

The mass in the remaining regions is constrained accordingly. In particular, we have :

$$\begin{aligned}
P(\Pi_1(1) \leq \varepsilon_1 \leq \Pi_1(0), \Pi_2(1) \leq \varepsilon_2 \leq \Pi_2(0)) = \Pi_2(0) - \Pi_2(1) - s(\Pi_1(1)-p_{11})\\
-t(1-p_{00}-\Pi_1(0)).
\end{aligned}$$

This mass can be divided between $(Y_1=1, Y_2=0)$ and $(Y_1=0, Y_2=1)$ with an appropriate choice of equilibrium selection mechanism, in order to satisfy the following constraint.

$$\begin{aligned}
p_{10} &= 1-p_{00}-\Pi_2(0)+s(\Pi_1(1)-p_{11})\\
&\quad +\delta\Big(\Pi_2(0)-\Pi_2(1)-s(\Pi_1(1)-p_{11})-t(1-p_{00}-\Pi_1(0))\Big) \qquad (25)
\end{aligned}$$

with equilibrium selection parameter $\delta \in [0,1]$. There remains to show that equation (25) has a solution for $(s,t,\delta) \in [0,1]^3$.

Case $1-p_{00} = \Pi_1(0)$ **:** When $1-p_{00} = \Pi_1(0)$, equation (25) becomes

$$p_{01}+p_{11}-\Pi_2(0)+\delta(\Pi_2(0)-\Pi_2(1))+s(1-\delta)(\Pi_1(1)-p_{11}) = 0.$$

If $\Pi_1(1) = p_{11}$, then δ can be chosen equal to $(\Pi_2(0)-p_{01}-p_{11})/(\Pi_2(0)-\Pi_2(1))$ (or δ unrestricted in case $\Pi_2(0)=\Pi_2(1)$). If $(1-\delta)(\Pi_1(1)-p_{11}) > 0$, then

$$s = \frac{\Pi_2(0)-p_{01}-p_{11}-\delta(\Pi_2(0)-\Pi_2(1))}{(1-\delta)(\Pi_1(1)-p_{11})}$$

must be between 0 and 1. So we must have $\delta \leq (\Pi_2(0)-p_{01}-p_{11})/(\Pi_2(0)-\Pi_2(1))$ (no restriction if $\Pi_2(0)=\Pi_2(1)$) and

$$(\Pi_2(0)-p_{01}-p_{11})-(\Pi_1(1)-p_{11}) \leq \delta\Big((\Pi_2(0)-\Pi_2(1))-(\Pi_1(1)-p_{11})\Big). \qquad (26)$$

We denote the latter $A \leq \delta B$. Since $\Pi_2(1) \leq p_{01}+p_{11}$, only three cases need to be considered :

1. $0 < A \leq B$: the δ needs to be chosen larger than or equal to A/B. Combined with the above, it yields $0 < A/B \leq \delta \leq (A + \Pi_1(1) - p_{11})/(B + \Pi_1(1) - p_{11}) \leq 1$, which has solutions since $A \leq B$ and $\Pi_1(1) \geq p_{11}$.
2. $A < 0 \leq B$: then (26) is always satisfied for $\delta \geq 0$ since the left-hand-side is negative and the right-hand-side is non negative.
3. $A < B < 0$: then any $\delta \in [0, 1]$ satisfies (26) since $-A > -B$.

Case $1 - p_{00} - \Pi_1(0) > 0$ **:** When $\delta(1 - p_{00} - \Pi_1(0)) > 0$, equation (25) can be rewritten :

$$t = \frac{p_{01} + p_{11} - \Pi_2(0) + \delta(\Pi_2(0) - \Pi_2(1)) + s(1-\delta)(\Pi_1(1) - p_{11})}{\delta(1 - p_{00} - \Pi_1(0))}$$

so we need to show there exists $(s, \delta) \in [0, 1]^2$ such that

$$\begin{aligned} 0 \leq p_{01} + p_{11} - \Pi_2(0) + \delta(\Pi_2(0) - \Pi_2(1)) + s(1-\delta)(\Pi_1(1) - p_{11}) \\ \leq \delta(1 - p_{00} - \Pi_1(0)). \end{aligned} \tag{27}$$

Subcase $\Pi_1(1) = p_{11}$ **:** We need to show the existence of $\delta \in [0, 1]$ such that

$$0 \leq \delta(\Pi_2(0) - \Pi_2(1)) - (\Pi_2(0) - p_{01} - p_{11}) \leq \delta(1 - p_{00} - \Pi_1(0)).$$

The left inequality is satisfied for

$$\frac{\Pi_2(0) - p_{01} - p_{11}}{\Pi_2(0) - \Pi_2(1)} \leq \delta \leq 1, \tag{28}$$

since $\Pi_2(1) \leq p_{01} + p_{11} \leq \Pi_2(0)$. The right inequality is equivalent to

$$-\Big(\Pi_2(0) - p_{01} - p_{11}\Big) \leq \delta\Big(1 - p_{00} - \Pi_1(0) - (\Pi_2(0) - \Pi_2(1))\Big),$$

which is true for any $\delta \geq 0$ if $1 - p_{00} - \Pi_1(0) \geq \Pi_2(0) - \Pi_2(1)$ and for any

$$0 \leq \delta \leq \frac{\Pi_2(0) - p_{01} - p_{11}}{\Pi_2(0) - \Pi_2(1) - (1 - p_{00} - \Pi_1(0))} \tag{29}$$

otherwise. (28) and (29) are compatible since $1 - p_{00} \geq \Pi_1(0)$.

Subcase $\Pi_1(1) > p_{11}$ **:** (27) is equivalent to

$$\frac{-\Big(p_{01}+p_{11}-\Pi_2(0)+\delta(\Pi_2(0)-\Pi_2(1))\Big)}{(1-\delta)(\Pi_1(1)-p_{11})} \leq s$$
$$\leq \frac{\delta(1-p_{00}-\Pi_1(0))-\Big(p_{01}+p_{11}-\Pi_2(0)+\delta(\Pi_2(0)-\Pi_2(1))\Big)}{(1-\delta)(\Pi_1(1)-p_{11})}.$$

The latter admits a solution $s \in [0,1]$ if and only if

$$\delta(1-p_{00}-\Pi_1(0))-\Big(p_{01}+p_{11}-\Pi_2(0)+\delta(\Pi_2(0)-\Pi_2(1))\Big) \geq 0 \tag{30}$$
$$\text{and} \quad -\Big(p_{01}+p_{11}-\Pi_2(0)+\delta(\Pi_2(0)-\Pi_2(1))\Big) \leq (1-\delta)(\Pi_1(1)-p_{11}) \tag{31}$$

(31) is equivalent to

$$\delta\Big((\Pi_1(1)-p_{11})-(\Pi_2(0)-\Pi_2(1))\Big) \leq (\Pi_1(1)-p_{11})-(\Pi_2(0)-p_{01}-p_{11}), \tag{32}$$

which we denote $\delta B \leq A$. Since $\Pi_2(1) \leq p_{01}+p_{11}$, we have $A \geq B$ and we need only consider the following three cases :

1. If $0 < B \leq A$, (32) is satisfied for all $\delta \in [0,1]$.
2. If $B \leq 0 \leq A$, (32) is satisfied for all $\delta \geq 0$, since the left hand side is negative and the right-hand-side positive.
3. If $B \leq A < 0$: (32) is satisfied for a choice of $\delta \geq A/B$, namely

$$\frac{(\Pi_2(0)-p_{01}-p_{11})-(\Pi_1(1)-p_{11})}{(\Pi_2(0)-\Pi_2(1))-(\Pi_1(1)-p_{11})} \leq \delta \leq 1. \tag{33}$$

(30) is equivalent to

$$\delta\Big(1-p_{00}-\Pi_1(0)-(\Pi_2(0)-\Pi_2(1))\Big) \geq p_{01}+p_{11}-\Pi_2(0).$$

The right-hand-side is negative, so the statement is

1. true for all $\delta \in [0,1]$ when $1-p_{00}-\Pi_1(0)-(\Pi_2(0)-\Pi_2(1)) \geq 0$,
2. true for all

$$0 \leq \delta \leq \frac{\Pi_2(0)-p_{01}-p_{11}}{(\Pi_2(0)-\Pi_2(1))-(1-p_{00}-\Pi_1(0))} \tag{34}$$

when $1-p_{00}-\Pi_1(0)-(\Pi_2(0)-\Pi_2(1)) < 0$.

Note that there is a solution to both (33) and (34). Indeed, calling the left-hand-side of (33) $-A/(-B)$, with both numerator and denominator positive, we can write the right-hand-side of (34) as $[-A+(\Pi_1(1)-p_{11})]/(-B+(\Pi_1(1)-p_{11})-(1-p_{00}-\Pi_1(0)))$, which is larger than or equal to $-A/(-B)$. This completes the proof for the duopoly entry case.

Coordination case :

As shown above, results for the coordination case can be obtained from results pertaining to the duopoly entry case by relabeling of payoff functions.

Asymmetric cases :

From the identified set for (Π, C) we can derive sharp bounds for the payoff functions alone using Fréchet bounds on C. Line 1 of (3.9) yields $P(1,1) = C(\Pi_1(1), \Pi_2(1)) \leq \min(\Pi_1(1), \Pi_2(2)) \leq \Pi_2(1)$ (Fréchet bound). Similarly, Line 4 of (3.9) yields $P(1,0) = \Pi_1(0) - C(\Pi_1(0), \Pi_2(1)) \leq \Pi_1(0)$ and $1 - P(1,0) = 1 - \Pi_1(0) + C(\Pi_1(0), \Pi_2(1)) \geq \Pi_2(1)$ (Fréchet lower bound). Line 3 yields $1 - P(0,1) = 1 - \Pi_2(0) + C(\Pi_1(1), \Pi_2(0)) \geq \Pi_1(1)$. Since $\Pi_1(1) \geq \Pi_1(0)$, we have $C(\Pi_1(1), \Pi_2(1)) \geq C(\Pi_1(0), \Pi_2(1))$ and Lines 1 and 4 of (3.9) combined yield $P(1,1) + P(1,0) = \Pi_1(0) + [C(\Pi_1(1), \Pi_2(1)) - C(\Pi_1(0), \Pi_2(1))] \geq \Pi_1(0)$. Similarly, since $\Pi_2(1) \leq \Pi_2(0)$, we have $C(\Pi_1(1), \Pi_2(1)) \leq C(\Pi_1(1), \Pi_2(0))$ and Lines 1 and 3 yield $P(1,1)+P(0,1) = \Pi_2(0)+[C(\Pi_1(1), \Pi_2(1)) - C(\Pi_1(1), \Pi_2(0))] \leq \Pi_2(0)$. Finally, $P(0,1)+P(1,1) = \Pi_2(0)-[C(\Pi_1(1), \Pi_2(0)) - C(\Pi_1(1), \Pi_2(1))] \geq \Pi_2(0) - [\Pi_2(0) - \Pi_2(1)] = \Pi_2(1)$ and similarly $P(1,0) + P(1,1) \geq \Pi_1(0)$.

We show now that the bounds are jointly sharp. To do so, take any given true frequency profile $(p_{11}, p_{10}, p_{01}, p_{00})$ and exhibit a joint distribution $C(\varepsilon_1, \varepsilon_2)$ such that all Π can be rationalized.

Construction of the joint distribution

We construct the joint distribution in the following way. Let $(s, t, u, v) \in [0, 1]^4$ be such that the following hold.

$$\begin{aligned}
&P(\varepsilon_1 \leq \Pi_1(0), \varepsilon_2 \leq \Pi_2(1)) = (1-u)p_{11}, \\
&P(\Pi_1(0) \leq \varepsilon_1 \leq \Pi_1(1), \varepsilon_2 \leq \Pi_2(1)) = up_{11}, \\
&P(\varepsilon_1 \geq \Pi_1(1), \varepsilon_2 \leq \Pi_2(1)) = (1-t)p_{01}, \\
&P(\varepsilon_1 \geq \Pi_1(1), \Pi_2(1) \leq \varepsilon_2 \leq \Pi_2(0)) = tp_{01}, \\
&P(\varepsilon_1 \geq \Pi_1(1), \varepsilon_2 \geq \Pi_2(0)) = (1-s)p_{00}, \\
&P(\Pi_1(0) \leq \varepsilon_1 \leq \Pi_1(1), \varepsilon_2 \geq \Pi_2(0)) = sp_{00}, \\
&P(\varepsilon_1 \leq \Pi_1(0), \varepsilon_2 \geq \Pi_2(0)) = (1-v)p_{10}, \\
&P(\varepsilon_1 \leq \Pi_1(0), \Pi_2(1) \leq \varepsilon_2 \leq \Pi_2(0)) = vp_{10}.
\end{aligned}$$

Marginal constraints are given by $1 - \Pi_1(1) = p_{01} + (1-s)p_{00}$, $1 - \Pi_2(0) = p_{00} + (1-v)p_{10}$, $\Pi_1(0) = p_{10} + (1-u)p_{11}$ and $\Pi_2(1) = p_{11} + (1-t)p_{01}$.

and the solution for $(s, t, u, v) \in [0, 1]^4$ is the following.

$$(s, t, u, v) = \left(\frac{\Pi_1(1) - p_{11} - p_{10}}{p_{00}}, \frac{p_{11} + p_{01} - \Pi_2(1)}{p_{01}}, \frac{p_{11} + p_{10} - \Pi_1(0)}{p_{11}}, \frac{\Pi_2(0) - p_{11} - p_{01}}{p_{10}} \right).$$

Note that $\Pi_1(0) = 0$ and $\Pi_1(1) = 1$ can only be reached if $p_{10} = p_{01} = 0$, which in turns forces $\Pi_2(1) = \Pi_2(0) = p_{11} = 1 - p_{00}$. Similarly, $\Pi_2(0) = 1$ and $\Pi_2(1) = 0$ can only be reached if $p_{11} = p_{00} = 0$, which in turns forces $\Pi_1(1) = \Pi_1(0) = p_{10} = 1 - p_{01}$.

The bounds for the second asymmetric game are obtained by permuting the two players and the result follows.