THE REVIEW OF ECONOMIC STUDIES

Review of Economic Studies (2008) **75**, 835–864 0034-6527/08/00340835$02.00

Partial Identification in Monotone Binary Models: Discrete Regressors and Interval Data

THIERRY MAGNAC
Toulouse School of Economics (*GREMAQ and IDEI*)

and

ERIC MAURIN
Paris School of Economics

First version received February 2004; *final version accepted October* 2007 (*Eds.*)

We investigate identification in semi-parametric binary regression models, $y = 1(x\beta + \upsilon + \varepsilon > 0)$ when υ is either discrete or measured within intervals. The error term ε is assumed to be uncorrelated with a set of instruments z, ε is independent of υ conditionally on x and z, and the support of $-(x\beta + \varepsilon)$ is finite. We provide a sharp characterization of the set of observationally equivalent parameters β. When there are as many instruments z as variables x, the bounds of the identified intervals of the different scalar components β_k of parameter β can be expressed as simple moments of the data. Also, in the case of interval data, we show that additional information on the distribution of υ within intervals shrinks the identified set. Specifically, the closer the conditional distribution of υ given z is to uniformity, the smaller is the identified set. Point identified is achieved if and only if υ is uniform within intervals.

1. INTRODUCTION

In empirical research, point identification of parameters often requires assumptions that are difficult to justify. The recourse to more credible and weaker restrictions leading to partial identification remains rare. Yet, in the case where the identified region is convex and bounded, the context of partial identification is not conceptually different from the familiar context of confidence intervals.

In this paper, we analyse partial identification in a binary regression model with discrete or interval-valued data. Although admittedly specific, this is a case which is surprisingly rich in terms of implications and facility of application. The identified set is bounded and convex. Bounds are easy to characterize by simple moments of the data.

One of the practical advantages of our model stems from the fact that available data on covariates are very often discrete or interval valued. Such covariates tend to render point identification very problematic (Manski, 1988; Horowitz, 1998). When all covariates (denoted x) are discrete, Bierens and Hartog (1988) show that there is an infinite number of single-index representations for the mean regression of a dependent variable, y. Under weak conditions, for almost any parameter θ, there exists a measurable real function φ_θ such that $E(y \mid x)$ can be written as $\varphi_\theta(x\theta)$.

The contribution by Manski and Tamer (2002) considers a reasonably more specific framework where the non-parametric mean regression $E(y \mid x)$ is assumed monotonic with respect to at least one particular regressor, say υ. As a special case, they study the identification of the parameter of the familiar semi-parametric binary regression model $y = 1(x\beta + \upsilon + \varepsilon > 0)$, when only interval data are available on v. They analyse the identification of β under a quantile independence assumption and show that parameters belong to a non-empty convex set of observationally equivalent values.

The first message of our paper is that set identification of parameter β in the binary model can be obtained through a different set of weak restrictions (*i.e.* Lewbel, 2000) and that estimation of the identified region in this setting only requires usual regression tools. Specifically, it is shown that the combination of an uncorrelated error assumption (*i.e.* $E(x'\varepsilon) = 0$) with a conditional independence assumption (*i.e.* $F_\varepsilon(\varepsilon \mid x, v) = F_\varepsilon(\varepsilon \mid x)$) and a finite support assumption ($\text{Supp}(-x\beta - \varepsilon) \subset [v_l, v_u]$, where v_l and v_u are finite) restricts the parameters of the semi-parametric binary regression model to a non-empty bounded and convex set. We characterize the identified intervals of any scalar linear combination of parameter β and show that the bounds of these intervals can be estimated using simple linear regression methods. It yields a simple analytical expression of the support function, which completely characterizes the identified set B. These findings are probably the most important results of this paper for practitioners because they enable researchers to estimate the identified set very easily.

Conditional independence in the latent model implies that the binary outcome is monotone in v. Interestingly, the support condition does not impose supplementary restrictions on the binary outcome that can be analysed, although it requires one to be careful when it comes to application. The support assumption implies that when v is varying between the extreme points v_l to v_u—which are not necessarily observed in the data—the conditional probability of success varies from 0 to 1. As discussed below, there are many potential applications of this set-up, including contingent valuation studies, optimal schooling models, failure time experiments, and any models where $y^* = x\beta + \varepsilon$ represents a subject's latent ability, $-v$ an exogenous threshold, and where we observe $y = 1(y^* > -v)$ only.

A final interesting feature of the set-up analysed in this paper is for the case where v is censored by interval. In this case, additional information on the distribution of v within intervals might reduce the size of the identified region. Specifically, the size of the identified region diminishes (in a sense made precise below) as the conditional distribution of the special regressor within intervals becomes closer to uniformity. The identified set is a singleton, and the parameter of interest β is exactly identified if and only if v is uniformly distributed within intervals conditional on covariates. This property is particularly interesting when one has control over the process of censoring the continuous data on v (*e.g.* the birth date) into interval data (*e.g.* month of birth). In order to minimize the size of the identified set, one should censor the data such that the distribution of the censored variable is as close as possible to a uniform distribution within the resulting intervals.

This paper belongs to the small, but growing literature on partial identification as pioneered by Manski (2003, and references therein) and derived from seminal papers such as Marschak and Andrews (1944) and Fréchet (1951). Our results on bounds of parameters in binary regressions can be seen as generalizations of bounds on averages, derived by Green, Jacowitz, Kahneman and McFadden (1998). They are also reminiscent of the results presented by Leamer (1987) and Chesher (2003). Several recent papers address the inference issues about sets (Horowitz and Manski, 2000), or change focus by considering the true values of the parameter (Imbens and Manski, 2004; Andrews and Guggenberger, 2007) following Anderson and Rubin's (1949) seminal contribution. Specifically, Chernozhukov, Hong and Tamer (2007) study inference about true parameters under more general conditions of set identification than ours. We show how their findings can be applied to our results. Finally, Beresteanu and Molinari (2006) develop inference procedures that could be adapted to the monotone binary model considered here.

The paper is organized as follows: the first section sets up notations and models; the second section examines the discrete case; the third section analyses the case of interval data; the fourth section briefly reports Monte Carlo experiments and the last section concludes. Straight from the start, we will consider the endogenous case where ε, though potentially correlated with the variables x, is uncorrelated with a set of instruments z (the special case where variables x are exogenous is not simpler). Some results will be specific to the case where the number of instruments is equal to the number of explanatory variables. All proofs are in appendices.

2. THE SET-UP

Let the "data" be given by the distribution of the following random variable:[1]

$$\omega = (y, v, x, z),$$

where y is a binary outcome, while v , x, and z are covariates and instrumental variables whose role and properties are specified below. We first introduce some regularity conditions on the distribution of ω. They will be assumed valid in the rest of the text.

Assumption R (egularity):

$(R.i.)$ (*Binary model*) The support of the distribution of y is $\{0, 1\}$

$(R.ii.)$ (*Covariates and Instruments*) The support of the distribution, $F_{x,z}$ of (x, z) is $S_{x,z} \subset \mathbb{R}^p \times \mathbb{R}^q$. The dimension of the set $S_{x,z}$ is $r \leq p+q$ where $p+q-r$ are the potential overlaps and functional dependencies.[2] The condition of full rank, $\text{rank}(E(z'x)) = p$, holds.

$(R.iii.)$ (*Discrete or Interval-Valued Regressor*) The support of the distribution of v conditional on (x, z) is a set $\Omega_v \subset [v_1, v_K]$, a finite interval, almost everywhere-$F_{x,z}$ (a.e. $F_{x,z}$). This conditional distribution, denoted $F_v(\cdot \mid x, z)$, is defined a.e. $F_{x,z}$.

$(R.iv.)$ (*Functional Independence*) There is no subspace of $\Omega_v \times S_{x,z}$ of dimension strictly less than $r+1$ whose probability measure, $(F_v(\cdot \mid x, z).F_{x,z})$, is equal to 1.

Assumptions $(R.i)$ and $(R.ii)$ define a binary model where there are p explanatory variables and q instrumental variables. In assumption $(R.iii,)$ the support of v is assumed to be independent of variables (x, z). If this support is an interval in $\mathbb{R}$ (including $\mathbb{R}$ itself), we are back to the case studied by Lewbel (2000) and Magnac and Maurin (2007). In the next section (Section 3), this support is assumed to be discrete, $\Omega_v = \{v_1, \dots, v_K\}$ so that the special regressor is said to be discrete. In Section 4, the support is assumed continuous, $\Omega_v = [v_1, v_K)$, but v is observed imperfectly because it is censored. In such a case, the special regressor is said to be interval valued. In all cases, Assumption $(R.iv)$ avoids the degenerate case where v and (x, z) are functionally dependent.

There are many examples of discrete covariates in applied econometrics. Variables, such as gender, levels of education, occupational status, or household size of survey respondents are genuinely discrete. In contingent valuation studies, prices are set by the experimenter, and they are in general discrete, by steps of 0·10, 1, or 10 euros. There are also many examples of interval-valued data. They are common in surveys where, in case of non-response to an item, follow-up questions are asked. Manski and Tamer (2002) describe the example of the Health and Retirement Study. If a respondent does not want to reveal his wealth, he is then asked whether it falls in a sequence of intervals (unfolding brackets). Another important reason for interval data is anonymity. Age is a continuous covariate, which could, in theory, be used as a source of continuous exogenous

1. We only consider random samples and we do not subscript individual observations by i.

2. With no loss of generality, the p explanatory variables x can partially overlap with the $q \geq p$ instrumental variables z. Variables (x, z) may also be functionally dependent (for instance, x, x^2, $\log(x)$,...). A collection $(x_1, \dots, x_K)$ of real random variables is functionally independent if its support is of dimension K (*i.e.* there is no set of dimension strictly lower than K whose probability measure is equal to 1).

variation in many settings. For confidentiality reasons, however, statisticians often censor this information in the public versions of household surveys by transforming dates of birth into months (or years) of birth only. They are afraid that the exact date of birth along with other individual and household characteristics might reveal the identity of households responding to the survey.

2.1. *The latent model*

Assuming that the data satisfy $((R.i)-(R.iv))$, the question addressed in this paper is how they can be generated by the following semi-parametric latent variable index structure:

$$y = \mathbf{1}\{x\beta + v + \varepsilon > 0\}, \tag{LV}$$

where $\mathbf{1}\{A\}$ is the indicator function that equals 1 if A is true and 0 otherwise and where the random shock ε satisfies the following properties,

Assumption L (atent):

$(L.1)$ (*Conditional independence*) ε and v are independent conditionally on covariates x and variables z.

$$F_\varepsilon(\cdot \mid v, x, z) = F_\varepsilon(\cdot \mid x, z).$$

The support of ε is denoted $\Omega_\varepsilon(x, z)$.

$(L.2)$ (*Support*) There exist two finite real numbers v_l and v_u such that the support of $-x\beta - \varepsilon$ is included in $[v_l, v_u)$ and such that $v_l \leq v_1$ and $v_u \geq v_K$ (see $(R.iii)$).

$(L.3)$ (*Moment condition*) ε is uncorrelated with variables z:

$$E(z'\varepsilon) = 0.$$

Powell (1994) discusses conditional independence assumptions (calling them exclusion restrictions) in the context of other semi-parametric models, that is, without combining them with $(L.2)$ or $(L.3)$. More recently, Lewbel (2000) and Honoré and Lewbel (2002) provide an analysis of model (LV) using $(L.1)$ and a more restrictive support assumption $(\text{Supp}(-x\beta - \varepsilon) \subset \text{Supp}(v))$ as identifying restrictions.[3]

As defined by $(L.1)$ and $(L.2)$, conditional independence and support assumptions restrict the class of statistical models that can actually be analysed. If a binary reduced-form $\Pr(y = 1 \mid v, x, z)$ is generated through (LV) by a latent model satisfying$(L.1-L.3)$, it necessarily satisfies,

$$\Pr(y = 1 \mid v, x, z) = \Pr(\varepsilon > -x\beta - v \mid x, z) = 1 - F_\varepsilon(-x\beta - v \mid x, z),$$

which implies that $\Pr(y = 1 \mid v, x, z)$ is non-decreasing in v. Second, as the support of $-x\beta - \varepsilon$ is included in $[v_l, v_u)$, we have necessarily $\Pr(y = 1 \mid v_l, x, z) = 0$ and $\Pr(y = 1 \mid v_u, x, z) = 1$. To sum up, we have

$(NP.1)$ (*Monotonicity*) The conditional probability $\Pr(y_i = 1 \mid v, x, z)$ is non-decreasing in v (a.e. $F_{x,z}$).

$(NP.2)$ (*Complete Variation*) There exist two finite real numbers v_l and v_u such that $\Pr(y_i = 1 \mid v = v_l, x, z) = 0$ and $\Pr(y_i = 1 \mid v = v_u, x, z) = 1$.

3. There is another minor difference between assumptions L and the set-up introduced by Lewbel (2000), namely the distribution function F_ε can have mass points. When the special regressor is discrete or interval valued, it is much easier than in the continuous case to allow for such discrete distributions of the unobserved factor. If all distribution functions are CADLAG (*i.e.* continuous on the right, limits on left), the large support assumption $(L.2)$ has to be slightly rephrased, however, in order to exclude a mass point at $-x\beta - v_u$.

In the following, we focus on the class of statistical models satisfying ($NP.1-NP.2$) and analyse the conditions under which they can be generated through (LV) by a latent model satisfying ($L.1-L.3$). To better understand what the restriction ($NP.2$) implies for applied research, it is worth distinguishing two cases. First, when $\Pr(y=1 \mid v, x, z)$ is actually observed increasing from 0 when $v = v_1$ to 1 when $v = v_K$, then v_l can be set equal to v_1 and v_u equal to v_K. In such a case, ($NP.2$) is unambiguously satisfied. Second, when either $\Pr(y=1 \mid v_1, x, z) > 0$ or $\Pr(y=1 \mid v_K, x, z) < 1$, then either v_l or v_u has to be set outside the observed support of v. In such a case, $\Pr(y=1 \mid v, x, z)$ satisfies ($NP.2$) only if there are plausible values of v_l and v_u outside the observed support of v such that $\Pr(y=1 \mid v_u, x, z) = 0$ or $\Pr(y=1 \mid v_l, x, z) = 1$. By construction, this assumption is not testable. As we now discuss, there are many examples where this assumption is plausible. However, the case should be argued in each specific application.

2.2. *Examples*

Potential applications of Assumption L include controlled experiments where $y^* = x\beta + \varepsilon$ represents a latent failure time and where individuals (or animals, or equipment) are observed at discrete, exogenously set, points in time or after having been exposed to discrete exogenously set doses of treatment (denoted $-v$). We observe $y = 1(x\beta + \varepsilon > -v)$ and we seek to identify β. By construction, this model satisfies the conditional independence and the uncorrelated error assumptions. It satisfies the support assumption provided that one can assume that the probability of "failure" goes from 0 at the beginning of the experiment to 1 after a sufficiently long (or strong) exposition to the treatment.

A second type of application is optimal investment models. For example, in optimal schooling models, $y^* = x\beta + \varepsilon$ represents the number of years of post-compulsory education which maximizes discounted lifetime wealth[4] and $-v$ is the respondent's age minus the minimum school-leaving age. The effects of family background x on y^* is the parameter of interest. Surveys provide us with information about $y = 1(y^* > -v)$ only, namely an indicator that the respondent still attends school at $-v$. In these models, the support assumption is essentially that there is an upper bound for the number of years that can be spent in the higher education system.

Other interesting empirical applications come from contingent valuation studies where we evaluate the impact of covariates x on the willingness to pay $y^* = x\beta + \varepsilon$ for a good or a resource, see, for example, Lewbel, Linton and McFadden (2006). Individuals are asked whether their willingness to pay exceeds a bid $-v$ chosen by experimental design. Again, we observe $y = 1(x\beta + \varepsilon > -v)$ and we seek to identify β. Bids are typically drawn from a discrete distribution. Given the experimental design, they may be constructed in order to satisfy the exclusion restriction ($L.1$). The model satisfies the support assumption provided that it can be assumed that nobody would answer "yes" for sufficiently high bids and nobody would answer "no" for sufficiently low bids.

Finally, applications of the conditional independence set-up are also provided by cases where y^* is a latent ability, v an exogenous ability threshold and where y indicates whether ability exceeds a given threshold. For example Maurin (2002) estimates a model on French data where y is grade repetition in primary schools, y^* is pupils' latent schooling ability, x is parental income, v is date of birth within the year. The date of birth within the year determines pupils' age at entry into elementary school and, as such, represents an important determinant of early performance at school. In this model, the support assumption means that if it were possible to observe sufficiently young children at entry into elementary school, they would all have to repeat a grade. In contrast, sufficiently mature children would all be able to avoid grade repetition.

4. See, for example, Cameron and Heckman (1998).

A related example is Lewbel (2006) who studies the ability to obtain a university degree using the cost of attending a local public college (relative to local unskilled wages) as the exogenous regressor υ.

2.3. *Identifying restrictions and parameter of interest*

The relationship between our set-up and the one in Manski and Tamer (2002) is similar to the relationship between quantile independence (Manski, 1988) and the identifying restrictions of Lewbel (2000). The quantile independence set-up assumes that one quantile of ε is independent of all covariates, whereas the conditional independence assumption used in this paper is equivalent to assuming that all quantiles of ε are independent of one covariate. In this crude sense, both assumptions are comparably restrictive. Another difference is that the conditional independence hypothesis makes it possible to characterize the domain of observationally equivalent distribution functions of the unobserved residuals. The price to pay is that conditional independence requires additional conditions on the support of the covariates and these are stronger than the conditions required under quantile independence. Assumption L and other examples are commented in Lewbel (2000) or Magnac and Maurin (2007). Once υ is continuously distributed and has large support (*i.e.* $\text{Supp}(-x\beta - \varepsilon) \subset [\upsilon_l, \upsilon_u)$), the latter paper shows that Assumption L is sufficient for exact identification of both β and $F_\varepsilon(\cdot \mid x, z)$.

Before moving on to the issue of identification of β, it is important to understand the relationship between this parameter and the effect of changes in covariates on the choice probability. Consider an experiment where (say) $(\upsilon, x' = x + \delta_x, z)$ is assigned to everyone of characteristics (υ, x, z), namely a exogenous change δ_x in covariates holding the unobserved heterogeneity term ε constant. The counterfactual probability of success conditional on (υ, x, z) is

$$\begin{aligned}\mathbf{E}(\mathbf{1}\{\upsilon + (x+\delta_x)\beta + \varepsilon > 0\} \mid \upsilon, x, z) &= \mathbf{E}(\mathbf{1}\{(\upsilon + \delta_x\beta) + x\beta + \varepsilon > 0\} \mid x, z)\\ &= \Pr(y = 1 \mid x, \upsilon + \delta_x\beta, z).\end{aligned}$$

In other words, the probability of success when $(\upsilon, x' = x + \delta_x, z)$ is assigned to everyone of characteristics (υ, x, z) is equal to the probability of success actually observed in the data conditional on $(\upsilon + \delta_x\beta, x, z)$. The parameter of interest β defines the shifts in υ whose effects on y are equivalent to exogenous shifts in x when we hold υ and z constant.

In the following, any $(\beta, F_\varepsilon(\cdot \mid x, z))$ satisfying Assumption L is called a latent model. The index parameter $\beta \in \mathbb{R}^p$ is the unknown parameter of interest. The distribution function of the error term, ε, is also unknown and may be considered as a nuisance parameter. Identification is studied in the set of all such parameters $(\beta, F_\varepsilon(\cdot \mid x, z))$.

3. THE DISCRETE CASE

In this section, the support of the special regressor is supposed to be a discrete set given by

Assumption D (iscrete): $\Omega_\upsilon = \{\upsilon_1, \ldots, \upsilon_K\}, \upsilon_k < \upsilon_{k+1}$ for any $k = 1, \ldots, K-1$.

We consider a binary reduced-form $\Pr(y = 1 \mid \upsilon, x, z)$ satisfying ($NP.1$–$NP.2$) and we ask whether there is a latent model $(\beta, F_\varepsilon(\cdot \mid x, z))$ satisfying assumptions ($L.1 - L.3$) and generating through (LV) the values $\Pr(y = 1 \mid \upsilon = \upsilon_k, x, z)$, $k = 1, \ldots, K$, taken by $\Pr(y = 1 \mid \upsilon, x, z)$ on the observed support of υ. The answer is positive though the admissible latent model is not unique. There are many possible latent models whose parameters are observationally equivalent. We start this section by proving that the identified set is given by a set of incomplete moment restrictions. We continue by showing that this set is non-empty, bounded and convex. We then give a sharp characterization of the identified set.

3.1. *Incomplete moment restrictions*

We begin with a one-to-one change in variables, which will allow us to characterize the set of observationally equivalent parameters through simple linear moment conditions. Denote

$$\delta_k = (v_{k+1} - v_{k-1})/2 \text{ for } k \in \{2, \dots, K-1\}$$
$$\delta_1 = (v_2 - v_l)/2,\ \delta_K = (v_u - v_{K-1})/2$$
$$p_k(x,z) = \Pr(v = v_k \mid x, z).$$

Using these notations, the transformation of the binary response variable, which will be used to characterize the identified set is defined as[5]

$$\tilde{y} = \frac{\delta_k . y}{p_k(x,z)} - \frac{v_u + v_K}{2} \quad \text{if } v = v_k, \quad \text{for } k \in \{1, \dots, K\}. \tag{1}$$

In contrast to the large-support, continuous case studied by Lewbel (2000) or Magnac and Maurin (2007), the identification of β when v is discrete is not exact anymore. The following theorem shows that β satisfies a set of moment conditions that are incomplete.

Theorem 1. *Let us consider β a vector of parameter and,* $\Pr(y = 1 \mid v = v_k, x, z)$ *(denoted $G_k(x,z)$) for k = 1,...,K, a conditional probability function which is non-decreasing in v. The two following statements are equivalent,*

(i) *there exists a latent random variable ε such that the latent model $(\beta, F_\varepsilon(\cdot \mid x,z))$ satisfies Assumption L and such that $G_k(x,z)$, k = 1,...,K, is the image of $(\beta, F_\varepsilon(\cdot \mid x,z))$ through the transformation (LV),*
(ii) *there exists a measurable function $u(x,z)$ from $S_{x,z}$ to $\mathbb{R}$ which takes its values in the interval (a.e.$F_{x,z}$) $I(x,z) = (-\Delta(x,z), \Delta(x,z)]$, where $\Delta(x,z)$ is positive and defined by,*

$$\Delta(x,z) = \frac{(v_1 - v_l)}{2} G_1(x,z) + \sum_{k=2}^{K} \left[\frac{(v_k - v_{k-1})}{2} (G_k(x,z) - G_{k-1}(x,z)) \right] + \frac{(v_u - v_K)}{2} (1 - G_K(x,z)),$$

and such that,

$$E(z'(x\beta - \tilde{y}) = E(z'u(x,z)). \tag{2}$$

Proof. See Appendix A. ||

Theorem 1 characterizes the set denoted B of all observationally equivalent values of parameter β. It shows that $\beta \in B$ if and only if it satisfies equation (2), which is an incomplete moment restriction since function $u(x,z)$ is not completely known. Green *et al.* (1998) proves a special case of this theorem when neither regressors x nor instruments z are present. It allows them to provide bounds for the average willingness to pay in a contingent valuation experiment. As discussed in a remark in Appendix A, the proof of Theorem 1 also leads to a characterization of the set of observationally equivalent distribution functions $F_\varepsilon(\cdot \mid x,z)$.

5. For almost all (v,x,z) in its support, which justifies that we divide by $p_k(x,z)$. Division by 0 is a null-probability event.

3.2. *Sharp bounds on structural parameters*

This section builds on Theorem 1 to provide a detailed description of B, the set of observationally equivalent parameters. We focus on the case where the number of instruments z is equal to the number of variables x (the exogenous case $z = x$ being the leading example). At the end of the section, we briefly discuss how the results could be extended to the case where the number of instruments z is larger that the number of explanatory variables, x.

3.2.1. General properties of the identified set. When the number of instruments is equal to the number of variables, the assumption that $E(z'x)$ is full rank (*R.ii*) implies that equation (2) has one and only one solution in β for any function $u(x,z)$. Given that $u(x,z) = 0$ is admissible, B is non-empty. It contains β^* the focal value associated with $u(x,z) = 0$,

$$E(z'(x\beta^* - \widetilde{y})) = 0.$$

Second, set B is convex because the set of admissible $u(x,z)$ is convex and equation (2) is linear in β. Lastly, since

$$\Delta(x,z) \leq \Delta_{\mathrm{M}} = \max(v_1 - v_l, \ldots, v_k - v_{k-1}, \ldots, v_u - v_K)/2$$

the admissible $u(x,z)$ is bounded by Δ_{M} so that B is bounded. Specifically, using the definition of β^* and rephrasing Theorem 1, β lies in B if and only if there is $u(x,z)$ taking its value in $I(x,z)$ such that

$$E(z'x)(\beta - \beta^*) = E(z'u(x,z)).$$

Denoting $W = E(z'z)^{-1/2}E(z'x)$ and using the generalized Cauchy–Schwarz inequality, we have,

$$(\beta - \beta^*)'W(\beta - \beta^*) = E(u'(x,z)z)E(z'z)^{-1}E(z'u(x,z)) \leq E(u^2(x,z)).$$

As $\Delta(x,z) \leq \Delta_{\mathrm{M}}$, we have,

$$(\beta - \beta^*)'W(\beta - \beta^*) \leq E(u^2(x,z)) \leq E(\Delta(x,z)^2) \leq \Delta_{\mathrm{M}}^2,$$

which shows that B is included in a sphere in the metric W. Previous developments are summarized in the following proposition,

Proposition 2. *The identified set B is non-empty, convex, and bounded. It contains the focal value $\beta^* = E(z'x)^{-1}E(z'\tilde{y})$. In the metric W, B is included in a sphere whose centre is β^* and whose radius is Δ_{M}.*

The maximum-length index, Δ_{M}, can be taken as a measure of distance to continuity of the distribution function of v (or between its support Ω_v and $[v_u, v_l]$). For a latent model (β, $F_\varepsilon(\cdot \mid x,z)$), Proposition 2 proves that, for a sequence of supports Ω_v indexed by Δ_{M}, the distance $d(\beta^*, B) = \inf_{b \in B} \|\beta^* - b\|$ between β^* and B converges to 0 as $\Delta_{\mathrm{M}} \mapsto 0$:

$$\lim_{\Delta_{\mathrm{M}} \mapsto 0} d(\beta^*, B) = 0,$$

and point identification is restored. We now give sharp bounds—first for single coefficients, second for linear combinations of coefficients—and show that it yields a sharp characterization of set B.

3.2.2. Interval identification in the coordinate dimensions. Let

$$B_p = \left\{\beta_p \in \mathbb{R} \mid \exists(\beta_1, \ldots, \beta_{p-1}) \in \mathbb{R}^{p-1}, (\beta_1, \ldots, \beta_{p-1}, \beta_p) \in B\right\}$$

represent the identified interval of the last coefficient (say). All scalars belonging to this interval, are observationally equivalent to the p-th component of the true parameter.

Proposition 3. *B_p is an interval centred at β_p^*, the p-th component of β^*. Specifically, we have,*

$$B_p = \left(\beta_p^* - \frac{E(|\widetilde{x_p}|\Delta(x,z))}{E(\widetilde{x_p}^2)}; \beta_p^* + \frac{E(|\widetilde{x_p}|\Delta(x,z))}{E(\widetilde{x_p}^2)}\right],$$

where $\widetilde{x_p}$ is the remainder of the IV-projection of x_p onto the other components of x using instruments z (as formally defined in the proof).

Proof. See Appendix A. ||

Given that the estimation of B_p requires the estimation of $E(|\widetilde{x_p}|\Delta(x,z))$, it is worth emphasizing that $\Delta(x,z)$ can be rewritten $E(\tilde{y}_\Delta \mid x,z)$ where $\tilde{y}_\Delta$, as $\tilde{y}$, is an affine function of y whose definition is given in the appendix at the end of the proof of Proposition 3. Furthermore, by definition, $\beta_p^* = \frac{E(|\widetilde{x_p}|\tilde{y})}{E(\widetilde{x_p}^2)}$. Hence, the construction of the upper and lower bounds of B_p only requires (1) the construction of the transforms $\tilde{y}+\tilde{y}_\Delta$, $\tilde{y}-\tilde{y}_\Delta$, (2) the construction of the residual $\widetilde{x_p}$, and (3) the linear regression of $\tilde{y}+\tilde{y}_\Delta$, $\tilde{y}-\tilde{y}_\Delta$ on $|\widetilde{x_p}|$. Estimation follows accordingly.

To be as specific as possible, we now develop the estimation procedure in the case of contingent valuation studies, one of the examples examined in Section 2.2. The very exogenous regressor υ is the opposite of the discrete bid that is proposed to each interviewee. Variables x might include total expenditure, which is endogenous, and might be instrumented by exogenous income as in standard demand theory. The first step consists in constructing transformation (1) using the binary dependent variable, the discrete values of the support of the bids, υ_k, and the empirical frequency of these values in the sample, $\hat{p}_k$. Those frequencies do not depend on (x,z) since they are chosen by experimental design. The construction of the other transform $\tilde{y}_\Delta$ uses the same objects and is as simple as the first. In the second step, we regress the variable of interest on to the other explanatory variables using the instruments z. Finally, we perform the two regressions of the two transforms $\tilde{y}+\tilde{y}_\Delta$, $\tilde{y}-\tilde{y}_\Delta$ onto $|\widetilde{x_p}|$.

3.2.3. Characterization and construction of the identified region. We begin with characterizing the identified interval of any linear combination of the parameters.

Consider q a column vector of dimension $[p,1]$ such that $\| q \| = 1$. The issue is to characterize the identified interval of $\beta_q = q'\beta$. To begin with, we can always chose Q a matrix of dimension $[p, p-1]$ such that the matrix (Q,q) is an orthogonal matrix of dimension p. By construction, it satisfies $(Q,q)(Q,q)' = I$ so that

$$x\beta = x(Q,q)(Q,q)'\beta.$$

The p-th component of parameter $(Q,q)'\beta$ is $\beta_q = q'.\beta$. It is associated with the p-th explanatory variable, $s_q = xq$. Let B_q denote the identified interval associated with this explanatory variable. Denoting $\tilde{s}_q$ the remainder of the projection of s_q onto xQ, we can apply Proposition 3 and write,

$$B_q = \left(q'.\beta^* - \frac{E(|\tilde{s}_q|\Delta(x,z))}{E(\tilde{s}_q^2)}, q'.\beta^* + \frac{E(|\tilde{s}_q|\Delta(x,z))}{E(\tilde{s}_q^2)}\right].$$

For any normalized vector of weights q, this equation provides us with an analytical definition of the set of scalars that are observationally equivalent to the true $q'\beta$.[6] By construction, for any q in the unit sphere of $\mathbb{R}^p$ (denoted $\mathbb{S}$) the upper bound of B_q corresponds to the supremum of $q'.\beta$ when β lies in B. This function is known as the support function of B at q. It is denoted $\delta^*(q \mid B)$ and it is equal to

$$\delta^*(q \mid B) = \sup_{\beta \in B} q'.\beta = q'.\beta^* + \frac{E(|\tilde{s}_q|\Delta(x))}{E(\tilde{s}_q^2)}.$$

It leads to a sharp characterization of the identified set B. As B is bounded and convex, its closure cl(B) is indeed completely characterized by its support function and equal to the intersection of its supporting half spaces (Rockafellar, 1970),

$$\mathrm{cl}(B) = \left\{ \beta \text{ such that for any } q \in \mathbb{S},\ q'\beta \leq q'.\beta^* + \frac{E(|\tilde{s}_q|\Delta(x))}{E(\tilde{s}_q^2)} \right\}.$$

Interestingly enough, we have an analytical definition of the support function of B. This makes it possible to construct B very easily by simulation. Randomly draw S vectors q_s, $\| q_s \| = 1$, construct the half spaces $\{q_s'\beta \leq \delta^*(q_s \mid B)\}$ and their intersection. Then make S go to infinity.

The effect on set B of various auxiliary parameters can now be assessed. The impact of the limit points v_l, v_u are of particular interest in the case where they do not belong to the support of v ($v_l < v_1$ or $v_K < v_u$) and where $G_1(x,z) > 0$ or $G_K(x,z) < 1$. Using the definitions of β^* and $\Delta(x,z)$, it is not difficult to check that

$$\frac{\partial \delta^*(q \mid B)}{\partial v_l} = \frac{-E((|\tilde{s}_q| - \tilde{s}_q)G_1(x,z))}{E(\tilde{s}_q^2)} \text{ and } \frac{\partial \delta^*(q \mid B)}{\partial v_u} = \frac{E((|\tilde{s}_q| - \tilde{s}_q)(1 - G_K(x,z))}{E(\tilde{s}_q^2)},$$

which depend neither on v_l nor on v_u. As $\frac{\partial \delta^*}{\partial v_l} < 0$ or $\frac{\partial \delta^*}{\partial v_u} > 0$, the size of B decreases when v_l or v_u are getting closer to the actual limits of the support of v, $[v_1, v_K]$. Second, B becomes unbounded in all directions when either $v_l \to -\infty$ or $v_u \to +\infty$. This result was obtained by Magnac and Maurin (2007) when v is continuous.

Regarding inference, Chernozhukov *et al.* (2007) and Andrews and Guggenberger (2007) provide tools that can be applied to estimate either the intervals of interest or the whole of set B. Specifically, Chernozhukov *et al.* (2007) provide confidence regions for sets of parameters that correspond to the zeroes of a non-negative continuous function, $Q(\beta)$. In our case,

$$Q(\beta) = \int_{\mathbb{S}} (q'\beta - \delta^*(q \mid B))^2 1(q'\beta > \delta^*(q \mid B)) d\mu(q)$$

where $d\mu(q)$ can be any strictly positive finite measure on the unit sphere $\mathbb{S}$. The sample analogue Q_n of Q is not difficult to construct since $\delta^*(q \mid B)$ can be estimated through simple linear regressions. In such a context, assuming various regularity conditions, Chernozhukov *et al.* (2007) show how to construct confidence sets C_n such that $\lim_{n \mapsto \infty} P(B \subset C_n) = \alpha$ or $\lim_{n \mapsto \infty} \sup_{\beta \in B} P(\beta \subset C_n) = \alpha$ for a prespecified confidence level $\alpha \in (0, 1)$.

3.3. *Supernumerary restrictions*

A potentially interesting development of this framework is when the number of instruments is larger than the number of variables ($q > p$). In such a case, B is not necessarily non-empty, since

6. It is worth emphasizing that $\tilde{s}_q$ (and consequently, B_q) is independent of the choice of the matrix Q. It can be shown that it is invariant to the replacement of Q by QR, where R can be any orthogonal matrix.

condition (2) in Theorem 1 may have no solutions at all (*i.e.* some supernumerary restrictions may not be true).

Consider z_A, a random vector whose dimension is the same as random vector x, defined by

$$z_A = zA$$

and such that $E(z'_A x)$ is full rank. Define the set, $\mathcal{A}$, of such matrices A of dimension p, q. The previous analysis can then be repeated for any A in such a set. The identified set $B(A)$ is now indexed by A. Under the maintained assumption (L.3), the true parameter (or parameters) belongs to the intersection of all such sets when matrix A varies:

$$B \subset \bigcap_{A \in \mathcal{A}} B(A).$$

The set on the R.H.S. is convex because it is the intersection of convex sets. Also, we can always project this set onto its elementary dimensions. The intersection of the projections is the projection of the intersections. What changes is that it can be empty. As a subject of a companion paper (Bontemps, Magnac and Maurin, 2007) we are currently exploring ways of characterizing the support function of B and the possibility of constructing test procedures of supernumerary restrictions in such partial identification frameworks, when the number of instruments is greater than the number of covariates.

4. INTERVAL DATA

In this section, we deal with the case where v is continuous although it is observed by intervals only. We show that the identified set has a similar structure as in the discrete case. It is a convex set and, when there are no supernumerary restrictions ($p = q$), it is not empty. It contains the value corresponding to an IV regression of a transformation of y on x given instruments z. When some information is available on the conditional distribution function of regressor v within intervals, the identified set shrinks. Its size diminishes as the distribution function of the special regressor within intervals becomes closer to uniformity. When v is conditionally uniformly distributed within intervals, the identified set is a singleton and the parameter of interest, β, is exactly identified.

4.1. *Identified set: the general case*

The data are now characterized by a random variable (y, v, v^*, x, z) where v^* is the result of censoring v by interval:

$$v^* = k.1\{v \in [v_k, v_{k+1})\} \text{ for } k = 1, \dots, K-1.$$

Only realizations of (y, v^*, x, z) are observed and those of v are not. Variable v^* is discrete and defines the interval in which v lies. More specifically, assumption D is replaced by

Assumption ID:

(i) (*Interval Data*) The support of v^* conditional on (x, z) is $\{1, \dots, K-1\}$ almost everywhere $F_{x,z}$. The distribution function of v^* conditional on (x, z) is denoted $p_{v^*}(x, z)$. It is defined almost everywhere $F_{x,z}$.

(ii) *(Continuous Regressor)* The support of v conditional on $(x, z, v^* = k)$ is $[v_k, v_{k+1})$ (almost everywhere $F_{x,z}$). The overall support is $[v_1, v_K)$ where $v_1 = v_l$ and $v_K = v_u$. The distribution function of v conditional on x, z, v^* is denoted $F_v(\cdot \mid v^*, x, z)$ and is assumed to be absolutely continuous. Its density function denoted $f_v(\cdot \mid v^*, x, z)$ is strictly positive and bounded.

For the sake of simplicity, we focus on the case where $v_1 = v_l$ and $v_K = v_u$. Our results can readily be extended to cases where $v_1 > v_l$ or $v_K < v_u$ without additional insight. We consider latent models which satisfy the large support condition (L.2) (*i.e.* the support of $-x\beta - \varepsilon$ is included in the support of v), the moment condition (L.3) (*i.e.* $E(z'\varepsilon) = 0$) and the following extension of the partial independence hypothesis,

$$F_\varepsilon(\cdot \mid v, v^*, x, z) = F_\varepsilon(\cdot \mid x, z). \tag{$L.1^*$}$$

The conditional probability distributions $\Pr(y = 1 \mid v^*, x, z)$ generated through transformation (LV) by such latent models is necessarily non-decreasing in v^*.

Interval censorship that we consider does not cover cases where intervals are unbounded on the left and/or on the right. Using Magnac and Maurin (2007), we can indeed prove that parameter β does not belong to a bounded set in these cases. Generally speaking, the set of assumptions used in this section is quite similar to assumptions proposed by Manski and Tamer (2002). Their Interval (I) assumption is equivalent to Assumption (ID.ii) and their monotone assumption (M) is equivalent to the monotonicity restriction imposed on $\Pr(y = 1 \mid v^*, x, z)$ by our latent model. Also their Mean Independence (MI) assumption is a consequence of our assumption ($L.1^*$), ours being slightly stronger. However, we depart from the quantile restriction and exogeneity assumptions that they use in the binary case, since we assume that shocks are uncorrelated with some instruments z and that the bounds of the intervals of observation are not random.

Analogously to the discrete case, we begin with constructing a transformation of the dependent variable. If $\delta(v^*) = v_{v^*+1} - v_{v^*}$ denotes the length of the v^*-th interval, the transformation adapted to interval data is :

$$\bar{y} = \frac{\delta(v^*)}{p_{v^*}(x, z)} y - v_K. \tag{3}$$

It is slightly different from the transformation used in the previous section in equation (1) in terms of weights $\delta(v^*)$ and in reference to the end-points, yet the dependence on the random variable $y/p_{v^*}(x, z)$ remains the same.

Using these notations, the following theorem gives a sharp representation of the set of observationally equivalent parameters as solutions to incomplete linear moment conditions as in Theorem 1.

Theorem 4. *Consider β, a vector of parameter and* $\Pr(y = 1 \mid v^*, x, z)$*(denoted $G_{v^*}(x, z)$) a conditional distribution function, which is non-decreasing in v^*. The two following statements are equivalent,*

(i) *there exists a latent conditional distribution function of v, $F_v(\cdot \mid x, z, v^*)$, and a latent random variable ε defined by its conditional distribution function $F_\varepsilon(\cdot \mid x, z)$ such that*

  a. $(\beta, F_\varepsilon(\cdot \mid x, z))$ *satisfies* $(L.1^*, L.2, L.3)$
  b. $G_{v^*}(x, z)$ *is the image of* $(\beta, F_\varepsilon(\cdot \mid x, z))$ *through the transformation* (LV),

(ii) *there exists a function $u^*(x, z)$ taking its values in $I^*(x, z) = (\underline{\Delta}^*(x, z), \overline{\Delta}^*(x, z))$ where (by convention, $G_0(x, z) = 0$, $G_K(x, z) = 1$),*

$$\overline{\Delta}^*(x, z) = \sum_{k=1,\ldots,K-1} (G_{k+1}(x, z) - G_k(x, z))(v_{k+1} - v_k),$$

$$\underline{\Delta}^*(x, z) = -\sum_{k=1,\ldots,K-1} (G_k(x, z) - G_{k-1}(x, z))(v_{k+1} - v_k),$$

and such that,

$$E(z'(x\beta - \bar{y}) = E(z'u^*(x,z))). \tag{4}$$

Proof. See Appendix B. ‖

Theorem 4 provides a characterization of the set of latent models satisfying $(L.1^* - L.3)$ and generating $G_k(x,z)$, $k = 1, \dots, K-1$, through (LV). The identified set has the same general structure in the interval data case as in the discrete case. It is a bounded and convex set, which always contains the focal value defined by the moment condition $E(z'(x\beta^* - \bar{y}) = 0$. We now study how additional information helps to shrink the identified set.

4.2. *Inference using additional information on the distribution function of the special regressor*

There are many instances where additional information on the conditional distribution function of v within intervals is available. Variable v could be observed at the initial stage of a survey or a census and for confidentiality reasons, dropped from the files that are provided to researchers. Only information about interval data and the conditional distribution function of v remains. Another instance is when the conditional distribution function of v is available in one database that does not contain information on y, while the information on y is available in another database, which contains only interval information on v.[7] To analyse these situations, we complete the statistical model by assuming that we have full information on the conditional distribution of v which is denoted $\Phi(v \mid v^*, x, z)$.

The first unsurprising result is that additional knowledge of $\Phi(v \mid x, z, v^*)$ actually helps to shrink the identified set. Secondly, knowing how identification is related to the conditional distribution $\Phi(v \mid x, z, v^*)$ may provide interesting guidelines to control censorship and choose intervals for defining v^* in an optimal way. It is thus quite surprising to find that point identification is restored provided that the conditional distribution function of the censored variable v is piecewise uniform.

To state these two results, we are going to use indices measuring the distance of a distribution function $\Phi(v \mid v^* = k, x, z)$ to uniformity. Specifically, we denote

$$U(v \mid v^* = k) = \frac{v - v_k}{v_{k+1} - v_k},$$

the uniform c.d.f., and we consider the two following indexes,

$$\xi_k^U(x,z) = \sup_{v \in (v_k, v_{k+1})} \left[\frac{\Phi - U}{\Phi}\right], \qquad \xi_k^L(x,z) = \inf_{v \in (v_k, v_{k+1})} \left[\frac{\Phi - U}{1 - \Phi}\right],$$

where the arguments of Φ and U are made implicit for expositional simplicity.

Given that Φ is absolutely continuous and its density is positive everywhere (ID.ii), $\frac{\Phi - U}{\Phi}$ and $\frac{\Phi - U}{1-\Phi}$ are well defined on (v_k, v_{k+1}) and satisfy $\frac{\Phi-U}{\Phi} < 1$ and $\frac{\Phi-U}{1-\Phi} > -1$. Furthermore, given that $\frac{\Phi-U}{\Phi}$ (respectively $\frac{\Phi-U}{1-\Phi}$) is continuous and equal to 0 at v_{k+1} (respectively v_k), the supremum of this function in the neighbourhood of v_{k+1}(respectively v_k) is non-negative (respectively non-positive). Hence, we have $\xi_k^L(x,z) \in (-1, 0]$ and $\xi_k^U(x,z) \in [0, 1)$, the two indices being equal to 0 when Φ is equal to U. Using additional information, Theorem 4 is translated into

7. Moffitt and Ridder (2006) provides a survey of two-sample techniques for such data design.

Theorem 5. *Consider β, a vector of parameters,* $\Pr(y=1 \mid v^*, x, z)$ *(denoted $G_{v^*}(x,z)$) a conditional distribution function, which is non-decreasing in v^* and $\Phi(v \mid v^*, x, z)$ a conditional distribution function. The two following statements are equivalent,*

(i) *there exists a latent random variable ε defined by its conditional distribution function $F_\varepsilon(\cdot \mid x, z)$ such that*

a. $(\beta, F_\varepsilon(\cdot \mid x, z))$ *satisfies* $(L.1^*, L.2, L.3)$
b. $G_{v^*}(x,z)$ *is the image of* $(\beta, F_\varepsilon(\cdot \mid x, z))$ *through the transformation* (LV),

(ii) *there exists a function $u^*(x,z)$ taking its values in $[\underline{\Delta}^*_\Phi(x,z), \overline{\Delta}^*_\Phi(x,z)]$ where*

$$\underline{\Delta}^*_\Phi(x,z) = \sum_{k=1,\ldots,K-1} (v_{k+1} - v_k)\xi_k^L(x,z)(G_k(x,z) - G_{k-1}(x,z))$$

$$\overline{\Delta}^*_\Phi(x,z) = \sum_{k=1,\ldots,K-1} (v_{k+1} - v_k)\xi_k^U(x,z)(G_{k+1}(x,z) - G_k(x,z))$$

and such that

$$E(z'(x\beta - \bar{y}) = E(z'u^*(x,z))).$$

Proof. See Appendix B. ||

Given that $\xi_k^L(x,z) \in (-1, 0]$ and $\xi_k^U(x,z) \in [0,1)$, the identified set characterized by Theorem 5 is smaller than the identified set characterized by Theorem 4 when no information is available on v (see below). Also, Theorem 5 makes clear that the size of the identified set diminishes with respect to the distance between the conditional distribution of v and the uniform distribution, as measured by $\xi_k^L(x,z)$ and $\xi_k^U(x,z)$. When v is piecewise uniform (conditionally), the identified set boils down to a singleton.

Corollary 6. *The identified set is a singleton if and only if the conditional distribution, $\Phi(v \mid x, z, v^*)$, for all $v^* = k$, and a.e. $F_{x,z}$, is uniform, that is,*

$$\Phi(v \mid v^* = k, x, z) = \frac{v - v_k}{v_{k+1} - v_k}.$$

Proof. See Appendix B. ||

Corollary 6 provides a necessary and sufficient condition for identification, which is very different from the sufficient conditions given in Manski and Tamer (2002, corollary, p. 524). Their conditions (c) and (d) for point identification imply a positive probability that the interval of observation of v (denoted v_0, v_1 in their notation) is as small as we want. In our case, this length is fixed. As our condition is necessary and sufficient, a complete comparison with what can be obtained in the setting of Manski and Tamer (2002) is beyond the scope of this paper.

Assuming that the distribution of v is not piecewise uniform, the question remains as to whether it is possible to rank the potential distributions of v according to the corresponding degree of underidentification of β. The answer is that we can. Specifically, the closer the conditional distribution of v is to uniformity, the smaller is the identified set.

To state this result, we first need to rank distributions according to the magnitude of their deviations from the uniform distribution.

Definition 7. *$\Phi_2(v \mid x, z, v^*)$ is closer to uniformity than $\Phi_1(v \mid x, z, v^*)$, when a.e. $F_{x,z}$ and for any $k \in \{1, \ldots, K-1\}$*

$$\xi^L_{k,1}(x,z) \leq \xi^L_{k,2}(x,z)$$
$$\xi^U_{k,1}(x,z) \geq \xi^U_{k,2}(x,z).$$

The corresponding preorder is denoted $\Phi_1 \succeq \Phi_2$.

Using this definition

Corollary 8. *Let $\Phi(v \mid v^* = k, x, z)$ any conditional distribution. Let B denote the associated region of identification for β. Then*

$$\Phi_1 \succeq \Phi_2 \Longrightarrow B_{\Phi_2} \subseteq B_{\Phi_1}.$$

Proof. Straightforward using Theorem 5. ||

Assuming that we have some control over the construction of v^* (*i.e.* using the information about v that is made available to researchers), this result shows that the variable should be constructed in such a way that minimizes the distance between the uniform distribution and the distribution of v conditional on v^* (and other regressors). Such a choice minimizes the length of the identified interval. Consider, for instance, the date of birth. The frequency of this variable plausibly varies from one season to another, or even from one month to another, especially in countries where there are strong seasonal variations in economic activity. At the same time, it is likely that the frequency of this variable does not vary significantly within months. Therefore, it is uniformly distributed within months in most countries. In such a case, our results show that we need only make available the month of birth of respondents (and not necessarily their exact date of birth) to achieve exact identification of the structural parameters of binary models, which are monotone with respect to date of birth.

4.3. *Projections of the identified set*

Results concerning projections of the identified set in the discrete case can easily be extended to the case of interval data. As in the discrete case and for simplicity, we restrict our analysis to the leading case when the dimension of z and x are the same. The identified set B can be projected onto its elementary dimensions using the same usual rules as in Corollary 3.

Again let

$$B_p = \left\{\beta_p \in \mathbb{R} \mid \exists(\beta_1, \ldots, \beta_{p-1}) \in \mathbb{R}^{p-1}, (\beta_1, \ldots, \beta_{p-1}, \beta_p) \in B\right\}$$

be the projected set corresponding to the last coefficient. We denote β^* the solution of equation (4) when function $u^*(x,z) = 0$ (as $E(z'x)$ is a square invertible matrix):

$$\beta^* = E(z'x)^{-1} E(z'\bar{y}).$$

To begin with, we consider the case where no information is available on the distribution of v and state the corollary to Theorem 4.

Corollary 9. *B_p is an interval whose centre is β^*_p, where β^*_p represents the p-th component of β^*. Specifically, we have,*

$$B_p = (\beta^*_p + \varsigma_{L,p}; \beta^*_p + \varsigma_{U,p}]$$

where

$$\varsigma_{L,p} = [E(\widetilde{x_p}^2)]^{-1} E(\widetilde{x_p}(\mathbf{1}\{\widetilde{x_p} > 0\}\underline{\Delta}^*(x,z) + \mathbf{1}\{\widetilde{x_p} \leq 0\}\overline{\Delta}^*(x,z))$$

$$\varsigma_{U,p} = [E(\widetilde{x_p}^2)]^{-1} E(\widetilde{x_p}(\mathbf{1}\{\widetilde{x_p} \leq 0\}\underline{\Delta}^*(x,z) + \mathbf{1}\{\widetilde{x_p} > 0\}\overline{\Delta}^*(x,z))$$

with $\widetilde{x_p}$ is the residual of the projection of x_p onto the other components of x using instruments z.

Proof. See Appendix B. ||

The corresponding corollary to Theorem 5 replaces $\underline{\Delta}^*$ and $\overline{\Delta}^*$ by $\underline{\Delta}^*_\Phi$ and $\overline{\Delta}^*_\Phi$. In the proof, we also show how to construct $\varsigma_{L,p}$ and $\varsigma_{U,p}$ as functions of moments of observable variables as in the previous section. Analogously, we can define the support function of set B for any vector q of the unit sphere and characterize exactly set B and a criterium function $Q(\beta)$ whose zeroes define B.

5. MONTE CARLO EXPERIMENTS

Deriving empirical estimates for the upper and lower bounds of intervals of interest is straightforward since these bounds can be expressed as moments. When the number of observations becomes large, the properties of interval estimates conform with theoretical properties that have just been derived. It remains to be seen how these estimators of the bounds behave in small and medium-sized samples (*i.e.* 100–1000 observations). This is why we briefly present Monte Carlo experiments in this section. The simulated model is $y = \mathbf{1}\{1 + v + x_2 + \varepsilon > 0\}$. For the sake of clarity, the set-up is chosen to be as close as possible to the set-up originally used by Lewbel (2000). We adapt this original setting to the case where regressor v is discrete or interval valued. The design is described in Appendix B.

5.1. *Statistical summaries*

Before moving on to the results, we introduce simple statistics to describe the small sample properties of our estimates. Let $\hat{\theta}_i$, $i = u, b$, be the estimators of the lower and upper bounds of the estimated interval and $\bar{\theta}_i = E(\hat{\theta}_i)$, $i = u, b$, the expected values of these estimators. Using these notations, let us consider $\bar{\theta}_m = (\bar{\theta}_u + \bar{\theta}_b)/2$ the expected average of the estimated lower and upper bounds, $(\bar{\theta}_u - \bar{\theta}_b)/2\sqrt{3}$ the expected adjusted length of the estimated interval, and $\overline{\sigma}^2$ the average sampling error defined as

$$\overline{\sigma}^2 = \frac{(\sigma_u^2 + \sigma_b^2 + \sigma_{ub})}{3},$$

where $\sigma_u^2 = E(\hat{\theta}_u - \bar{\theta}_u)^2$ and $\sigma_b^2 = E(\hat{\theta}_b - \bar{\theta}_b)^2$ are the estimated S.E. of the estimated lower and upper bounds whereas $\sigma_{ub} = E[(\hat{\theta}_b - \bar{\theta}_b)(\hat{\theta}_u - \bar{\theta}_u)]$ is their estimated covariance. Interestingly enough, these three statistics provide a way to decompose the mean square error uniformly integrated over the interval $[\hat{\theta}_b, \hat{\theta}_u]$:

$$\text{MSEI} = E\int_{\hat{\theta}_b}^{\hat{\theta}_u} (\theta - \theta_0)^2 \frac{d\theta}{\hat{\theta}_u - \hat{\theta}_b} = \frac{1}{3} E\left[\frac{(\hat{\theta}_u - \theta_0)^3 - (\hat{\theta}_b - \theta_0)^3}{\hat{\theta}_u - \hat{\theta}_b}\right]$$

$$= \frac{1}{3} E[(\hat{\theta}_u - \theta_0)^2 + (\hat{\theta}_b - \theta_0)^2 + (\hat{\theta}_b - \theta_0)(\hat{\theta}_u - \theta_0)],$$

$$= (\bar{\theta}_m - \theta_0)^2 + \frac{1}{3}((\bar{\theta}_u - \bar{\theta}_b)/2)^2$$

$$+ \frac{1}{3}E[(\hat{\theta}_u - \bar{\theta}_u)^2 + (\hat{\theta}_b - \bar{\theta}_b)^2 + (\hat{\theta}_b - \bar{\theta}_b)(\hat{\theta}_u - \bar{\theta}_u)].$$

The first term is the square of a "decentring" term (denoted Dec), which can be interpreted as the familiar bias term. The second term is the square of the "adjusted" length (AL), which can be interpreted as the specific "uncertainty" due to partial identification instead of point identification. The third term is an average of S.E. (ASE), which can be interpreted as the familiar variance term. This decomposition is an adaptation of the classical decomposition of mean square error to the case where identification is partial.

5.2. *Results*

We have performed several Monte Carlo experiments using discrete and interval data where the sample size varies between 100, 200, 500, and 1000 observations. In all experiments, the number of Monte Carlo replications is equal to 1000. Additional replications do not affect any estimates (respectively S.E.) by more than a 1% margin of error (respectively 3%). Generally speaking, in all experiments, the true value of the parameter is within the 95% confidence interval built up around the lower and upper bounds of the estimates.

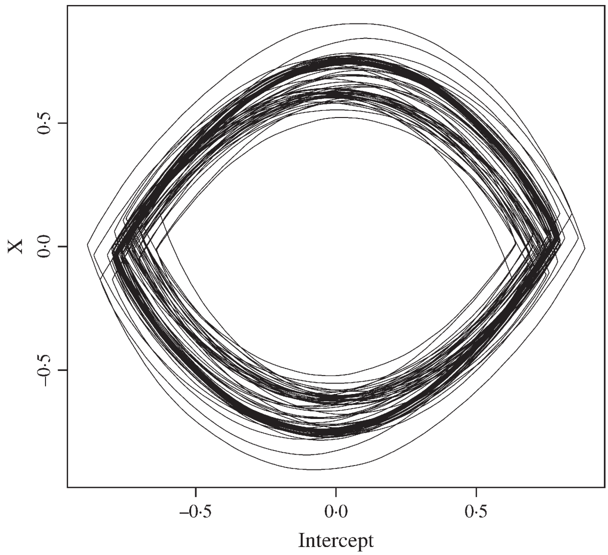
Table 1 presents the results of an experiment using interval data by reporting variations of Dec, AL, ASE, and Root MSEI when both the sample size and the bandwidth used to compute the denominator of transformation (3) vary. It shows that the estimated intervals can be severely decentred for the intercept term, especially when the sample size is small. Increasing the bandwidth

TABLE 1

Simple experiment, interval data: sensitivity to bandwidth. Error decomposition: decentring, adjusted length, and sampling error

		Intercept				Variable			
No. observations	Bandwidth	Dec	AL	ASE	RMSEI	Dec	AL	ASE	RMSEI
100	1·000	0·554	0·396	0·398	0·789	−0·041	0·353	0·491	0·606
	1·500	0·410	0·404	0·325	0·661	−0·068	0·352	0·411	0·545
	3·000	0·247	0·401	0·266	0·541	−0·079	0·337	0·327	0·476
	5·000	0·182	0·383	0·281	0·508	−0·090	0·321	0·322	0·464
200	1·000	0·358	0·428	0·204	0·594	0·007	0·375	0·267	0·461
	1·500	0·218	0·422	0·172	0·505	−0·039	0·367	0·225	0·432
	3·000	0·079	0·412	0·144	0·444	−0·101	0·350	0·191	0·412
	5·000	0·046	0·410	0·140	0·435	−0·126	0·347	0·202	0·420
500	1·000	0·096	0·411	0·094	0·433	−0·053	0·357	0·122	0·381
	1·500	0·025	0·406	0·085	0·415	−0·088	0·349	0·106	0·375
	3·000	−0·038	0·400	0·080	0·410	−0·129	0·340	0·104	0·378
	5·000	−0·050	0·399	0·080	0·410	−0·138	0·338	0·113	0·383
1000	1·000	−0·006	0·402	0·059	0·406	−0·094	0·345	0·073	0·364
	1·500	−0·044	0·399	0·056	0·405	−0·119	0·339	0·067	0·365
	3·000	−0·076	0·396	0·055	0·407	−0·146	0·334	0·071	0·372
	5·000	−0·082	0·396	0·055	0·408	−0·151	0·334	0·078	0·374

The number of interval values is equal to 10. The simple experiment refers to the case where $\alpha = \rho = 0$. All details are reported in the text. Experimental results are based on 1000 replications. Bandwidth refers to the bandwidth that is used. Dec stands for decentring of the mid-point of the interval. AL is the adjusted length of the interval. ASE is the sampling variability of bounds as defined in the text. The identity $\text{Dec}^2 + \text{AL}^2 + \text{ASE}^2 = \text{RMSEI}^2$ is shown in the text. RMSEI is the root mean square error integrated over the identified interval.



Note: Simple experiment. 1000 observations and 100 replicated estimates of the set

FIGURE 1

Estimated sets

decentres interval estimates for the coefficient towards negative numbers though to a much lesser extent. The mean square error (MSEI) for the intercept decreases with the bandwidth, especially when the sample size is small. It frequently has a U-shaped form for the coefficient the variable. We have tried to look for a data-driven choice of the bandwidth by minimizing this quantity, but this was inconclusive. A larger bandwidth seems to be always preferred. Some further research is clearly needed on this issue. In the working paper version,[8] we report results when other parameters vary: the degree of non-normality of ε, the degree of endogeneity of x_2 and the number of points in the support of υ. It is shown that decentring can be quite severe when the degree of non-normality of the random shock or the degree of endogeneity of the covariate is large. Interval length is not affected by non-normality, but exhibits some non-monotonic variations with the amount of endogeneity. Finally, we find that interval length decreases with the number of points of the support of υ, as predicted by the theory. This decrease is not sensitive to sample size.

The projections of the identified set onto one-dimensional intervals may provide a distorted picture of the bi-dimensional set. For example, when the identified set is stretched along the 45° line, the single-dimensional intervals might be very wide even when the total area of the bi-dimensional set is small. This is why it may be informative to construct the complete identified region on top of its one-dimensional projections. In our example, the complete bi-dimensional set can be computed quite easily. Figure 1 shows 100 replications of the complete bi-dimensional estimated sets. An interesting feature of these complete sets (*i.e.* a feature, which is not perceptible when working with one-dimensional projections) is their "ocular" shape. The kinks on both sides stem from the deterministic nature of one of the covariates (the intercept). If x_2 were discretely distributed, we would obtain polyhedral sets. Table 2 reports the magnitude of error when we proxy the complete estimated set by the rectangle given by the projections of the estimated set on both vertical and horizontal axes. In our specific case, the error is moderate, the surface of the

8. See http://www.idei.fr/doc/wp/2005/magnac.pdf.

TABLE 2

Approximating true sets by their projections on axes

Observations	True area	Square area	Proportion	
			Mean	S.D.
100	1·57	1·81	0·91	0·41
200	1·40	1·79	0·80	0·12
500	1·41	1·91	0·74	0·05
1000	1·51	2·12	0·71	0·02

Simple experiment. 100 replicated estimates of the set. "True area" is the mean of the estimated areas of the true sets. "Square area" is the area of the square given by the estimated intervals on the two axis. "Proportion" is the ratio of the former over the latter.

true set being around 75% of the area of the rectangle, with some variation according to the size of the sample.

6. CONCLUSION

In this paper, we explore partial identification of coefficients of binary variable models when exogenous regressors are discrete or interval valued. We derive bounds for the coefficients and show that they can be written as moments of the data generating process. We also show that in the case of interval data, additional information can shrink the identified set. When the unknown variable is distributed uniformly within intervals, these sets are reduced to one point.

Some additional points are worthy of consideration. First, we do not provide proofs of consistency and asymptotic properties of the estimates of the interval bounds because they would add little to those presented by Lewbel (2000). The asymptotic variance–covariance matrix of the bounds can also be derived along similar lines. Moreover, adapting the proofs of Magnac and Maurin (2007), these estimates are efficient in a semi-parametric sense under certain conditions. In contrast, constructing confidence sets for the identified set or the true value of the parameter is more involved (Beresteanu and Molinari, 2006) and is the subject of work in progress.

Generally speaking, the identification results obtained in this paper for the case where data on v are not continuous may also be used to enhance identification power when the data are actually continuous. For example, in the case where v is continuous but not "large-support", one could use additional measurements or structural priors at discrete points to the left and right of the actual support in order to achieve partial or point identification. Such additional information generates a case with mixed discrete and continuous support. It can be analysed by using (simultaneously) the proofs developed for the discrete, interval, and continuous settings. An interesting case arises for a binary variable whose probability of occurrence is known to be monotone in some regressor v and to vary between 0 and 1 in a known interval of v. School leaving (as a function of age) is such an example. In such a case, the coefficients of the binary latent model are partially identified regardless of whether the scheme of observation of v is complete, discrete, interval valued or continuous. Two extreme cases lead to exact identification: (1) complete and continuous observation in the interval; (2) complete and interval-data observation when the distribution of v is uniform within intervals. Nevertheless, other cases are still informative. Finally, a more complex research question is whether our results can be extended to settings where the moment condition (L.3) is replaced by stronger conditional mean independence or conditional independence assumptions. Such assumptions can be analysed as supernumerary moment conditions which are the object of a companion paper (Bontemps *et al.*, 2007).

APPENDIX A. PROOFS IN SECTION 3

Proof of Theorem 1. Let $\{G_k(x,z)\}_{k=1,\dots,K}$ satisfy monotonicity ($G_k < G_{k+1}$). It is an ordered set of functions such that $G_1 \geq 0$ and $G_K \leq 1$. Fix β. We first prove that (i) implies (ii).

(*Necessity*) Assume that there exists a latent random variable ε such that $(\beta, F_\varepsilon(\cdot \mid x,z))$ satisfies $(L.1 - L.3)$ and such that $\{G_k(x,z)\}_{k=1,\dots,K}$ is its image through transformation (LV). In the following, we denote $v_0 = v_l$ and $v_{K+1} = v_u$ and $G_0(x,z) \equiv 0$ and $G_{K+1}(x,z) \equiv 1$. By $(L.2)$, the conditional support of ε given (x,z), is included in $(-(v_{K+1} + x\beta), -(v_0 + x\beta)]$ and we can write,

$$\forall k \in \{1,\dots,K+1\},\ G_k(x,z) = \int_{-(v_k+x\beta)}^{-(v_0+x\beta)} f_\varepsilon(\varepsilon \mid x,z) d\varepsilon = 1 - F_\varepsilon(-(v_k + x\beta) \mid x,z). \tag{A.1}$$

Put differently, we necessarily have $F_\varepsilon(-(v_k + x\beta) \mid x,z) = 1 - G_k(x,z)$, for each k in $\{0,\dots,K+1\}$.

Denote $s_k = (v_k + v_{k-1})/2$ and $\delta_k = \dfrac{v_{k+1} - v_{k-1}}{2} = s_{k+1} - s_k$ for all $k = 1,\dots,K$. Set the transformed variable $\tilde{y}$ to $(\dfrac{\delta_k y}{p_k(x,z)} - s_{K+1})$ where $y = \mathbf{1}\{v > -(x\beta + \varepsilon)\}$. Integrate $\widetilde{y}$ with respect to v (of support $\{v_1,\dots,v_K\}$) and ε:

$$\begin{aligned} E(\widetilde{y} \mid x,z) &= \int_{\Omega_{(\varepsilon|x,z)}} \left[\sum_{k=1}^{K} \delta_k \mathbf{1}\{v_k > -(x\beta + \varepsilon)\}\right] f(\varepsilon \mid x,z) d\varepsilon - s_{K+1} \\ &= \int_{\Omega_{(\varepsilon|x,z)}} \left[\sum_{k=1}^{K} (s_{k+1} - s_k) \mathbf{1}\{v_k > -(x\beta + \varepsilon)\}\right] f(\varepsilon \mid x,z) d\varepsilon - s_{K+1}. \end{aligned}$$

As the support of $w = -(x\beta + \varepsilon)$ is included in $[v_0, v_{K+1}]$, define an integer function $j(w)$ in $\{0,\dots,K\}$, such that $v_{j(w)} \leq w < v_{j(w)+1}$. By construction, $v_k > w \Leftrightarrow k > j(w)$ and $\sum_{k=1}^{K} (s_{k+1} - s_k) 1\{v_k > w\} = (s_{K+1} - s_{j(w)+1})$. Hence, we have

$$\begin{aligned} E(\widetilde{y} \mid x,z) &= \int_{\Omega_{(\varepsilon|x,z)}} (s_{K+1} - s_{j(-(x\beta+\varepsilon))+1}) f(\varepsilon \mid x,z) d\varepsilon - s_{K+1} = -E[s_{j(-x\beta-\varepsilon)+1} \mid x,z] \\ &= x\beta + E(\varepsilon \mid x,z) - E[s_{j(-x\beta-\varepsilon)+1} + x\beta + \varepsilon \mid x,z] \\ &= x\beta + E(\varepsilon \mid x,z) - u(x,z) \end{aligned} \tag{A.2}$$

where (recall that $w \equiv -(x\beta + \varepsilon)$):

$$u(x,z) = E(s_{j(w)+1} - w \mid x,z).$$

Bounds on $u(x,z)$ can be obtained using the definition of $j(w)$. Given that $v_{j(w)} \leq w < v_{j(w)+1}$, we have :

$$-\frac{v_{j(w)+1} - v_{j(w)}}{2} < s_{j(w)+1} - w = \frac{v_{j(w)+1} + v_{j(w)}}{2} - w \leq \frac{v_{j(w)+1} - v_{j(w)}}{2}.$$

Hence, we can write using the upper bound and decomposing the support of ε into intervals,

$$\begin{aligned} E(s_{j(w)+1} - w \mid x,z) &= \sum_{k=1}^{K+1} \int_{-(v_k+x\beta)}^{-(v_{k-1}+x\beta)} (s_k + x\beta + \varepsilon) f(\varepsilon \mid x,z) d\varepsilon \\ &\leq \sum_{k=1}^{K+1} \int_{-(v_k+x\beta)}^{-(v_{k-1}+x\beta)} \frac{v_k - v_{k-1}}{2} f(\varepsilon \mid x,z) d\varepsilon \\ &= \sum_{k=1}^{K+1} \left[\frac{v_k - v_{k-1}}{2} (G_k(x,z) - G_{k-1}(x,z))\right] = \Delta(x,z) \end{aligned}$$

where in the last line, we used equation (A.1). For the lower bound, a similar proof yields:

$$-\Delta(x,z) < u(x,z) \leq \Delta(x,z).$$

Since $G_{K+1}(x,z) = 1$ and $G_0(x,z) = 0$, we have $\Delta(x,z) \geq \min_k(\frac{v_k - v_{k-1}}{2})$, meaning that $\Delta(x,z) > 0$ and that $I(x,z)$ is non-empty. It finishes the proof that statement (i) implies statement (ii) since equation (A.2) implies equation (2).

(*Sufficiency*) Conversely, let us prove that statement (ii) implies statement (i). We assume that there exists $u(x,z)$ in $I(x,z) = (-\Delta(x,z), \Delta(x,z)]$ such that equation (2) holds true and we construct a distribution function $F_\varepsilon(\cdot \mid x,z)$ satisfying $(L.1 - L.3)$ such that the image of $(\beta, F_\varepsilon(\cdot \mid x,z))$ through (LV) is $\{G_k(x,z)\}_{k=1,\dots,K}$.

First, let λ a random variable whose support is $(0,1]$, whose conditional density given (υ, x, z) is independent of υ (a.e. $F_{x,z}$) and is such that:

$$E(\lambda \mid x,z) = (u(x,z) + \Delta(x,z))/(2\Delta(x,z)). \tag{A.3}$$

As this expectation lies between 0 and 1, it is always possible to find such a random variable. Second, let κ a discrete random variable whose support is $\{1,\dots,K+1\}$ and whose conditional distribution given (υ,x,z) is independent of υ and is given using:

$$\Pr(\kappa = k \mid x,z) = G_k(x,z) - G_{k-1}(x,z) \tag{A.4}$$

where we keep on denoting $G_{K+1}(x,z) \equiv 1$ and $G_0(x,z) \equiv 0$. For any $k \in \{1,\dots,K+1\}$, consider K random variables, say $\varepsilon(\lambda,k)$ which are constructed from λ by

$$\varepsilon(\lambda,k) = -x\beta - \lambda\upsilon_{k-1} - (1-\lambda)\upsilon_k.$$

Given that $\lambda > 0$, the support of $\varepsilon(\lambda,k)$ is $(-x\beta - \upsilon_k, -x\beta - \upsilon_{k-1}]$. Finally, consider the random variable

$$\varepsilon = \varepsilon(\lambda,\kappa), \tag{A.5}$$

whose support is $(-x\beta - \upsilon_K, -x\beta - \upsilon_1]$ and which is independent of υ (because both λ and κ are). It therefore satisfies $(L.1)$ and $(L.2)$. Furthermore, because of (A.4), the image of $(\beta, F_\varepsilon(\cdot \mid x,z))$ through (LV) is $\{G_k(x,z)\}_{k=1,\dots,K}$ because these functions satisfy equation (A.1). The last condition to prove is $(L.3)$. Consider, for almost any (x,z),

$$\begin{aligned}
&\int_{\Omega_{(\varepsilon|x,z)}} (s_{j(-x\beta-\varepsilon)+1} + x\beta + \varepsilon) f(\varepsilon \mid x,z)\, d\varepsilon \\
&= \sum_{k=1}^{K+1} \left(\int_{-x\beta-\upsilon_k}^{-x\beta-\upsilon_{k-1}} \left(\frac{\upsilon_k + \upsilon_{k-1}}{2} + x\beta + \varepsilon \right) f(\varepsilon \mid x,z,\kappa = k)\, d\varepsilon \right) \\
&= \sum_{k=1}^{K+1} E\left(\frac{\upsilon_k + \upsilon_{k-1}}{2} - \lambda\upsilon_{k-1} - (1-\lambda)\upsilon_k \mid x,z \right) (G(\upsilon_k,x,z) - G(\upsilon_{k-1},x,z)) \\
&= \sum_{k=1}^{K+1} E(\lambda - 1/2 \mid x,z).(\upsilon_k - \upsilon_{k-1}).(G_k(x,z) - G_{k-1}(x,z)) \\
&= (u(x,z)/(2\Delta(x,z)))(2\Delta(x,z)) = u(x,z),
\end{aligned}$$

where the third line is the consequence of the definition of ε and the last line is using equation (A.3). Therefore, equation (A.2) holds and:

$$E(z'\widetilde{y}) = E(z'x)\beta + E(z'\varepsilon) - E(z'u(x,z)).$$

Equation (2) implies $E(z'\varepsilon) = 0$, that is $(L.3)$, which finishes the proof of Theorem 1. ‖

Remark A1. It is worth emphasizing that this proof also provides a characterization of the domain of observationally equivalent distribution functions F_ε, that is, the set of random variables ε such that there exists β with (β, F_ε) satisfying conditions $(L.1 - L.3)$ and generating $\{G_k(x,z)\}_{k=1,\dots,K}$. We have

The two following statements are equivalent,

(i) there exists a vector of parameter β such that the latent model $(\beta, F_\varepsilon(\cdot \mid x,z))$ verifies conditions L and such that $\{G_k(x,z)\}_{k=1,\dots,K}$ is its image through the transformation (LV),

(ii) there exist two independent random variables (λ,κ), conditional on (x,z) such that the support of λ is $(0,1]$, the support of $\kappa$is $\{2,\dots,K\}$, equation (A.4) holds and such that

$$\varepsilon = -x\beta - \lambda\upsilon_{\kappa-1} - (1-\lambda)\upsilon_\kappa$$

where β verifies

$$E(z'(x\beta - \tilde{y}) = E(z'\Delta(x,z)(2\lambda - 1)).$$

Proof of Corollary 3. For the sake of clarity, we start with the exogenous case where $z = x$. Denote x_p the last variable in x , x_{-p} all the other variables (*i.e.* $x = (x_{-p}, x_p)$). Consider any $\beta \in B$ and $\beta^* = \left(E(x'x)\right)^{-1} E(x'\widetilde{y})$. There exists a function $u(x)$ in $(-\Delta(x), \Delta(x)]$ such that $\beta - \beta^* = \left(E(x'x)\right)^{-1} E(x'u(x))$, which is also the result of the regression of $u(x)$ on x.

Denote the residual of the projection of x_p onto the other components x_{-p} as $\widetilde{x_p}$:

$$\widetilde{x_p} = x_p - x_{-p}\left(E(x'_{-p}x_{-p})\right)^{-1} E(x'_{-p}x_p).$$

Applying the principle of Frish–Waugh, we have

$$\beta_p - \beta_p^* = \left(E((\tilde{x}_p)^2)\right)^{-1} E(\widetilde{x}_p u(x)).$$

Since $\tilde{x}_p$ is a scalar, the maximum (respectively minimum) of $E(\widetilde{x}_p u(x))$ when $u(x,z)$ varies in $(-\Delta(x), \Delta(x)]$ is obtained by setting $u(x) = \Delta(x)\mathbf{1}\{\tilde{x}_p > 0\} - \Delta(x)\mathbf{1}\{\tilde{x}_p \leq 0\}$ (respectively $u(x) = -\Delta(x)\mathbf{1}\{\tilde{x}_p > 0\} + \Delta(x)\mathbf{1}\{\tilde{x}_p \leq 0\}$). Hence $E(\widetilde{x}_p u(x))$ lies between $-E(|\widetilde{x_p}|\Delta(x))$ and $E(|\widetilde{x_p}|\Delta(x))$ and the difference $\beta_p - \beta_p^*$ varies in

$$\left(-\frac{E(|\widetilde{x_p}|\Delta(x))}{E(\widetilde{x_p}^2)}, \frac{E(|\widetilde{x_p}|\Delta(x))}{E(\widetilde{x_p}^2)}\right].$$

To show the reciprocal, consider any β_p in

$$\left(\beta_p^* - \frac{E(|\widetilde{x_p}|\Delta(x))}{E(\widetilde{x_p}^2)}; \beta_p^* + \frac{E(|\widetilde{x_p}|\Delta(x))}{E(\widetilde{x_p}^2)}\right].$$

Denote

$$\lambda = \frac{E(\widetilde{x_p}^2)}{E(|\widetilde{x_p}|\Delta(x))}(\beta_p - \beta_p^*) \in (-1, 1].$$

Consider $u(x) = \lambda\Delta(x)$ when $\widetilde{x_p} > 0$ and $u(x) = -\lambda\Delta(x)$ otherwise, which means that

$$\frac{E(\widetilde{x}_p u(x))}{E(\widetilde{x_p}^2)} = (\beta_p - \beta_p^*).$$

Function $u(x)$ takes its values in $(-\Delta(x), \Delta(x)]$ and therefore satisfies point (ii) of Theorem 1. Thus, there exists $\beta \in B$ such that its last component is β_p.

The adaptation to the general IV case uses the generalized transformation

$$\widetilde{x_p} = z(E(z'z))^{-1}E(z'x_p) - z(E(z'z))^{-1}E(z'x_{-p})\left[E(x'_{-p}z)(E(z'z))^{-1}E(z'x_{-p})\right]^{-1}E(x'_{-p}z)(E(z'z))^{-1}E(z'x_p).$$

Generally speaking, the estimation of B_p requires the estimation of $E(|\widetilde{x_p}|\Delta(x,z))$. Given this fact, it is worth emphasizing that $\Delta(x,z)$ can be rewritten as $E(\tilde{y}_\Delta \mid x, z)$ where

$$\tilde{y}_\Delta = \sum_{K=1}^{K}\left[\frac{(v_k - v_{k-1} - (v_{k+1} - v_k))}{2p_k(x,z)}1(v = v_k)\right]y + \frac{v_{K+1} - v_K}{2}.$$

Specifically,

$$\Delta(x,z) = \frac{(v_1 - v_0)}{2}G_1(x,z) + \sum_{k=2}^{K}\left[\frac{(v_k - v_{k-1})}{2}(G_k(x,z) - G_{k-1}(x,z))\right] + \frac{(v_{K+1} - v_K)}{2}(1 - G_K(x,z))$$

$$= \sum_{k=1}^{K}\frac{(v_k - v_{k-1} - (v_{k+1} - v_k))}{2}G_k(x,z) + \frac{v_{K+1} - v_K}{2}$$

$$= E\left(\left[\sum_{k=1}^{K}1(v = v_k)\frac{(v_k - v_{k-1} - (v_{k+1} - v_k))}{2p_k(x,z)}\right]y + \frac{v_{K+1} - v_K}{2} \mid x, z\right) = E(\tilde{y}_\Delta \mid x, z).$$

Using these notations, $E(|\widetilde{x_p}|\Delta(x,z))$ can be rewritten $E(|\widetilde{x_p}|\tilde{y}_\Delta)$, which means that the estimation of the upper and lower bounds of B_p only requires (1) the construction of the transform $\tilde{y}_\Delta$, (2) an estimation of the residual $\widetilde{x_p}$ and (3) the linear regression of $\tilde{y}_\Delta$ on $|\widetilde{x_p}|$. ‖

APPENDIX B. PROOFS IN SECTION 4

Proof of Theorem 4. Consider a vector of parameters β and a conditional probability distribution $\Pr(y = 1 \mid v^*, x, z)$ (denoted $G_{v^*}(x,z)$) which is non-decreasing in v^*.

(*Necessity*) We prove that (i) implies (ii). Denote, $F_v(\cdot \mid x,z,v^*)$, and $F_\varepsilon(\cdot \mid x,z)$, two conditional distribution functions satisfying (i). By Assumption *ID(ii)*, $F_v(\cdot \mid x,z,v^*)$ is absolutely continuous and its density function is denoted f_v. By assumption Theorem 4 (i), $(\beta, F_\varepsilon(\cdot \mid x,z))$ satisfies condition $(L1^*)$, $(L2)$, and $(L3)$ and $\{G_k(x,z)\}_{k=1,\dots,K-1}$ is its image through transformation (LV).

For the sake of clarity, set $w = -(x\beta + \varepsilon)$ so that $y = \mathbf{1}\{v > w\}$ and the support of w is a subset of $[v_1, v_K)$ by $(L.2)$. The variable w is conditionally (on (x,z)) independent of v and v^* and the corresponding conditional distribution is

$$F_w(w \mid x,z) = 1 - F_\varepsilon(-(x\beta + w) \mid x,z).$$

The conditional probability of occurrence of $y = 1$ in the k-th interval ($v^* = k$ in $\{1,\dots,K-1\}$) is

$$G_k(x,z) = \int_{v_k}^{v_{k+1}} E(\mathbf{1}\{v > w \mid v, v^* = k, x, z) f_v(v \mid k, x, z) dv,$$

which yields the convolution equation:

$$G_k(x,z) = \int_{v_k}^{v_{k+1}} F_w(v \mid x,z) f_v(v \mid k,x,z) dv. \tag{B.1}$$

Note that this condition implies

$$F_w(v_k \mid x,z) \le G_k(x,z) \le F_w(v_{k+1} \mid x,z) \tag{B.2}$$

with a strict inequality on the right if $F_w(v_k \mid x,z) < F_w(v_{k+1} \mid x,z)$ because F_v is absolutely continuous and F_w is continuous on the right (CADLAG).

To prove (4), write $E(\bar{y} \mid x,z)$ as

$$\sum_{v^*=1,..,K-1} \int_{\Omega(v|v^*,x,z)} \int_{\Omega(w|v^*,v,x,z)} [\bar{y}.p_{v^*}(x,z).f_v(v \mid v^*,x,z) dv dF_w(w \mid v^*,v,x,z)].$$

Using the definition of $\bar{y}$, the term $p_{v^*}(x,z)$ cancels out and using condition $(L.1^*)$, the integral over *dw* on the one hand, and the sum and other integral on the other hand, can be permuted:

$$\int_{\Omega(w|x,z)} \left[\sum_{v^* \in \{1,\dots,K-1\}} \delta(v^*) \int_{\Omega(v|v^*,x,z)} \mathbf{1}(v > w)) f_v(v \mid v^*,x,z) dv \right] dF_w(w \mid x,z) - v_K. \tag{B.3}$$

Evaluate first the inner integral with respect to v. As the support of w is included in $[v_1, v_K)$, we can define for any value of w in its support, an integer function $j(w)$ in $\{1,\dots,K-1\}$, such that $v_{j(w)} \le w < v_{j(w)+1}$. Distinguish three cases. First, when $v^* < j(w)$, the whole conditional support of v lies below w and

$$\int_{\Omega(v|v^*,x,z)} \mathbf{1}(v > w) f_v(v \mid v^*,x,z) dv = 0,$$

while when $v^* > j(w)$, the whole conditional support of v lies strictly above w and thus

$$\int_{\Omega(v|v^*,x,z)} \mathbf{1}(v > w) f_v(v \mid v^*,x,z) dv = 1.$$

Last, when $v^* = j(w)$,

$$\int_{\Omega(v|v^*,x,z)} \mathbf{1}(v > w) f_v(v \mid v^*,x,z) dv = 1 - F_v(w \mid v^*,x,z).$$

Summing over values of v^*,

$$\sum_{v^* \in \{1,\dots,K-1\}} \delta(v^*) \int_{\Omega(v|v^*,x,z)} \mathbf{1}(v > w) f_v(v \mid v^*,x,z) dv = -F_v(w \mid v_{j(w)},x,z)(v_{j(w)+1} - v_{j(w)}) + v_K - v_{j(w)}.$$

Replacing in equation (B.3) and integrating w.r.t. w, implies that

$$E(\bar{y} \mid x,z) = -E(w \mid x,z) - u^*(x,z) = x\beta + E(\varepsilon \mid x,z) - u^*(x,z). \tag{B.4}$$

where

$$u^*(x,z) = \int_{\Omega_{(w|x,z)}} (F_v(w \mid v_{j(w)}, x, z)(v_{j(w)+1} - v_{j(w)}) + v_{j(w)} - w) dF_w(w \mid x,z).$$

Integrating equation (B.4) with respect to x, z and using condition $(L.3)$ yields condition (4). To finish the proof, upper and lower bounds for $u^*(x,z)$ are now provided. Write,

$$u^*(x,z) = \sum_{k=1}^{K-1} (v_{k+1} - v_k)\phi_k(x,z) \tag{B.5}$$

where

$$\phi_k(x,z) = \int_{v_k}^{v_{k+1}} \left(F_v(w \mid k, x, z) + \frac{v_k - w}{v_{k+1} - v_k} \right) dF_w(w \mid x,z). \tag{B.6}$$

Using integration by parts, the first term is

$$\phi_k(x,z) = \int_{v_k}^{v_{k+1}} \left(\frac{1}{v_{k+1} - v_k} - f_v(w \mid k, x, z) \right) F_w(w \mid x,z) dw,$$

which yields, using the convolution equation (B.1),

$$\phi_k(x,z) = -G_k(x,z) + \int_{v_k}^{v_{k+1}} \frac{F_w(w \mid x,z)}{v_{k+1} - v_k} dw.$$

Furthermore, using equation (B.2), implies

$$G_{k-1}(x,z) - G_k(x,z) \leq \phi_k(x,z) \leq G_{k+1}(x,z) - G_k(x,z)$$

where at least one inequality on the right and one inequality on the left are strict, since there exists at least one k such that $F_w(w = -(x\beta + v_{k+1}) \mid x,z) - F_w(w = -(x\beta + v_k) \mid x,z) > 0$. Therefore

$$\underline{\Delta}^*(x,z) < u^*(x,z) < \overline{\Delta}^*(x,z)$$

where the definitions of $\overline{\Delta}^*(x,z)$ and $\underline{\Delta}^*(x,z)$ correspond to those given in the body of the theorem.

(*Sufficiency*) We now prove that (ii) implies (i). Denote $u^*(x,z)$ in $(\underline{\Delta}^*(x,z), \overline{\Delta}^*(x,z))$ such that

$$E(z'(x\beta - \bar{y})) = E(z'u^*(x,z)).$$

We are going to prove that there exists a distribution function of $w = -(x\beta + \varepsilon)$ and a distribution function of v such that $(\beta, F_\varepsilon(\cdot \mid x,z))$ satisfies $(L.1^*, L.2, L.3)$ and $G_{v^*}(x,z)$ is the image of $(\beta, F_\varepsilon(\cdot \mid x,z))$ through the transformation (LV).

To begin with, we proceed in three steps to construct w. First, we choose a sequence of functions $H_k(x,z)$ such that $H_1 = 0$, $H_K = 1$, and such that:

$$H_k(x,z) \leq G_k(x,z) \leq H_{k+1}(x,z), \quad \text{for } k \in \{1, \ldots, K-1\} \tag{B.7}$$

where at least one inequality on the right is strict and

$$\sum_{k=1}^{K-1} (v_{k+1} - v_k)(H_k(x,z) - G_k(x,z)) < u^*(x,z) < \sum_{k=1}^{K-1} (v_{k+1} - v_k)(H_{k+1}(x,z) - G_k(x,z)).$$

Consider

$$\theta(x,z) > \frac{u^*(x,z)}{\underline{\Delta}^*(x,z)}, \quad 1 - \theta(x,z) > \frac{u^*(x,z)}{\overline{\Delta}^*(x,z)},$$

where, for instance, $\theta(x,z) = \dfrac{\overline{\Delta(x,z)} - u^*(x,z)}{\overline{\Delta}(x,z) - \underline{\Delta}(x,z)}$. By construction $\theta(x,z) \in (0,1]$ and one checks that

$$H_k(x,z) = \theta(x,z)G_{k-1}(x,z) + (1-\theta(x,z))G_k(x,z)$$

satisfies the two previous conditions. Generally speaking, the closer $u^*(x,z)$ is from the lower bound $\underline{\Delta}^*(x,z)$, the closer is H_k to G_{k-1} and the closer $u^*(x,z)$ is from the upper bound $\overline{\Delta}^*(x,z)$, the closer is H_k to G_k.

Second, we consider κ, a discrete random variable, whose support is $\{1,\dots,K-1\}$, which is independent of υ^* (a.e. $F_{x,z}$) and whose conditional on (x,z) distribution is

$$\Pr(\kappa = k \mid x,z) = H_{k+1}(x,z) - H_k(x,z). \tag{B.8}$$

Third, we consider λ a random variable whose support is $(0,1)$, which is independent of υ^* (a.e. $F_{x,z}$), and whose conditional (on (x,z)) expectation is

$$E(\lambda \mid x,z) = \frac{\sum_{k=1}^{K-1}(\upsilon_{k+1}-\upsilon_k)(H_{k+1}(x,z)-G_k(x,z)) - u^*(x,z)}{\sum_{k=1}^{K-1}(\upsilon_{k+1}-\upsilon_k)(H_{k+1}(x,z)-H_k(x,z))}. \tag{B.9}$$

For instance, λ can be chosen discrete with a mass point on

$$\lambda_0(x,z) = \frac{\sum_{k=1}^{K-1}(\upsilon_{k+1}-\upsilon_k)(H_{k+1}(x,z)-G_k(x,z)) - u^*(x,z)}{\sum_{k=1}^{K-1}(\upsilon_{k+1}-\upsilon_k)(H_{k+1}(x,z)-H_k(x,z))}.$$

Given the constraints on the $H_k(x,z)$ and given that $u^*(x,z)$ is in in $(\underline{\Delta}^*(x,z), \overline{\Delta}^*(x,z))$, $\lambda_0(x,z)$ belongs to $(0,1)$. Within this framework, we can define w as

$$w = (1-\lambda)\upsilon_\kappa + \lambda\upsilon_{\kappa+1}$$

By construction, the support of w is $[\upsilon_1, \upsilon_K)$ and w is independent of υ^* conditionally on (x,z) because both λ and κ are. Hence, $\varepsilon = -(x\beta + w)$ satisfies $(L.1)$ and $(L.2)$.

To construct υ, we first introduce a random variable η whose support is $[0,1)$, which is absolutely continuous, which is defined conditionally on (k,x,z), which is independent of λ and such that

$$\int_0^1 F_\lambda(\eta \mid x,z).f_\eta(\eta \mid k,x,z)d\eta = \frac{G_k(x,z)-H_k(x,z)}{H_{k+1}(x,z)-H_k(x,z)} \in [0,1)$$

where $F_\lambda(\cdot \mid x,z)$ denotes the distribution of λ conditional on (x,z).

For instance, when λ is chosen discrete with a mass point on $\lambda_0(x,z)$, we simply have to choose η such that

$$F_\eta(\lambda_0(x,z) \mid x,z) = \frac{H_{k+1}(x,z)-G_k(x,z)}{H_{k+1}(x,z)-H_k(x,z)}.$$

Now, define υ as

$$\upsilon = \upsilon_k + (\upsilon_{k+1}-\upsilon_k)\eta.$$

Let us now prove that the image of $(\beta, F_w(\cdot \mid x,z))$ through (LV) is $G_{\upsilon^*}(x,z)$ because it satisfies equation (B.1):

$$\int_{\upsilon_k}^{\upsilon_{k+1}} F_w(\upsilon \mid x,z).f_\upsilon(\upsilon \mid k,x,z)d\upsilon = H_k(x,z)$$

$$+(H_{k+1}(x,z)-H_k(x,z)).\int_{\upsilon_k}^{\upsilon_{k+1}} \Pr(w = (1-\lambda)\upsilon_k + \lambda\upsilon_{k+1} \le \upsilon \mid x,z).f_\upsilon(\upsilon \mid k,x,z)d\upsilon = H_k(x,z)$$

$$+(H_{k+1}(x,z)-H_k(x,z)).\int_0^1 \Pr(\lambda \le \eta \mid x,z).f_\eta(\eta \mid k,x,z)d\eta = G_k(x,z).$$

The last condition to prove is $(L.3)$. Rewrite equation (B.6), for almost any (x,z),

$$\phi_k(x,z) = -G_k(x,z) + \int_{\upsilon_k}^{\upsilon_{k+1}} \frac{F_w(w \mid x,z)}{\upsilon_{k+1}-\upsilon_k} dw$$

$$= -G_k(x,z) + H_{k+1}(x,z) - (H_{k+1}(x,z) - H_k(x,z))E(\lambda \mid x,z).$$

Therefore,

$$\sum_{k=1}^{K-1} (\upsilon_{k+1}-\upsilon_k)\phi_k(x,z) = \sum_{k=1}^{K-1} (\upsilon_{k+1}-\upsilon_k)(H_{k+1}(x,z) - G_k(x,z))$$

$$-\sum_{k=1}^{K-1} (\upsilon_{k+1}-\upsilon_k)(H_{k+1}(x,z) - H_k(x,z))E(\lambda \mid x,z) = u^*(x,z)$$

using equation (B.9). Plugging equation (4) in equation (B.4) yields $E(z'\varepsilon) = 0$ that is condition $(L.3)$. ‖

Proof of Theorem 5. We use large parts of the proof of Theorem 4.

(*Necessity*) Same as the proof of Theorem 4 until equation (B.6) that we rewrite as

$$\phi_k(x,z) = \int_{\upsilon_k}^{\upsilon_{k+1}} (\Phi(\upsilon \mid k,x,z) - \frac{\upsilon-\upsilon_k}{\upsilon_{k+1}-\upsilon_k}) dF_w(\upsilon \mid x,z).$$

First

$$\phi_k(x,z) = \int_{\upsilon_k}^{\upsilon_{k+1}} (1 - \frac{\frac{\upsilon-\upsilon_k}{\upsilon_{k+1}-\upsilon_k}}{\Phi(\upsilon \mid k,x,z)})\Phi(\upsilon \mid k,x,z) dF_w(\upsilon \mid x,z)$$

$$\leq \sup_{\upsilon\in(\upsilon_k,\upsilon_{k+1})} (1 - \frac{\frac{\upsilon-\upsilon_k}{\upsilon_{k+1}-\upsilon_k}}{\Phi(\upsilon \mid k,x,z)}) \int_{\upsilon_k}^{\upsilon_{k+1}} \Phi(\upsilon \mid k,x,z) dF_w(\upsilon \mid x,z)$$

$$= \xi_k^U(x,z) \int_{\upsilon_k}^{\upsilon_{k+1}} \Phi(\upsilon \mid k,x,z) dF_w(\upsilon \mid x,z)$$

where $\xi_k^U(x,z)$ is defined in the text. Equation (B.1) delivers

$$\int_{\upsilon_k}^{\upsilon_{k+1}} \Phi(\upsilon \mid k,x,z) dF_w(\upsilon \mid x,z) = \int_{\upsilon_k}^{\upsilon_{k+1}} d[\Phi(\upsilon \mid k,x,z)F_w(\upsilon \mid x,z)] - \int_{\upsilon_k}^{\upsilon_{k+1}} F_w(\upsilon \mid x,z) d\Phi(\upsilon \mid k,x,z)$$

$$= \int_{\upsilon_k}^{\upsilon_{k+1}} d[\Phi(\upsilon \mid k,x,z)F_w(\upsilon \mid x,z)] - G_k(x,z).$$

Hence, using $F_w(\upsilon_{k+1} \mid x,z) \leq G_{k+1}(x,z)$, we have,

$$\phi_k(x,z) \leq \xi_k^U(x,z).(G_{k+1}(x,z) - G_k(x,z)).$$

The derivation of the lower bound follows the same logic:

$$\phi_k(x,z) \geq \inf_{\upsilon\in(\upsilon_k,\upsilon_{k+1})} \left(-1 - \frac{\frac{\upsilon-\upsilon_{k+1}}{\upsilon_{k+1}-\upsilon_k}}{1-\Phi(\upsilon \mid k,x,z)}\right) \int_{\upsilon_k}^{\upsilon_{k+1}} (1-\Phi(\upsilon \mid k,x,z)) dF_w(\upsilon \mid x,z)$$

$$\geq \xi_k^L(x,z) \left[\int_{\upsilon_k}^{\upsilon_{k+1}} d[(1-\Phi(\upsilon \mid k,x,z))F_w(\upsilon \mid x,z)] + G_k(x,z)\right],$$

where $\xi_k^L(x,z)$ is defined in the text. Hence, using $F_w(v_k \mid x,z) \geq G_k(x,z)$, we have

$$\phi_k(x,z) \geq \xi_k^L(x,z)(G_k(x,z) - G_{k-1}(x,z)).$$

Therefore, using the definition of $u^*(x,z)$ in equation (B.5), we have

$$\underline{\Delta}_\Phi^*(x,z) \leq u^*(x,z) \leq \overline{\Delta}_\Phi^*(x,z) \tag{B.10}$$

where $\underline{\Delta}_\Phi^*(x,z)$ and $\overline{\Delta}_\Phi^*(x,z)$ are defined in the text.

(*Sufficiency*) We now prove that (ii) implies (i). We assume that there exists $u^*(x,z)$ in $[\underline{\Delta}_\Phi^*(x,z), \overline{\Delta}_\Phi^*(x,z)]$ such that

$$E(z'(x\beta - \tilde{y})) = E(z'u^*(x,z)).$$

Under this assumption, we are going to prove that there exists a distribution function of the random term ε such that $(\beta, F_\varepsilon(\cdot \mid x,z))$ satisfies $(L.1^*, L.2, L.3)$ and $G_{v^*}(x,z)$ is the image of $(\beta, F_\varepsilon(\cdot \mid x,z))$ through the transformation (LV), when the distribution function of the special regressor v is $\Phi(v \mid k,x,z)$. As in the proof of Theorem 4, we proceed by construction in three steps.

First, choose a sequence of functions $H_k(x,z)$ such that $H_1 = 0$, $H_K = 1$, and for any k in $\{1,\ldots,K-1\}$ such as

$$H_k(x,z) \leq G_k(x,z) < H_{k+1}(x,z) \tag{B.11}$$

and such as

$$\sum_{k=1}^{K-1}(v_{k+1} - v_k)\xi_k^L(x,z)(G_k(x,z) - H_k(x,z)) \leq u^*(x,z) \leq \sum_{k=1}^{K-1}(v_{k+1} - v_k)\xi_k^U(x,z)(H_{k+1}(x,z) - G_k(x,z)).$$

If $\xi_k^L(x,z) < 0$ and $\xi_k^U(x,z) > 0$, the closer $u^*(x,z)$ is from the lower bound $\underline{\Delta}_\Phi^*(x,z)$, the closer is H_k to G_{k-1}, and the closer $u^*(x,z)$ is from the upper bound $\overline{\Delta}_\Phi^*(x,z)$, the closer is H_k to G_k.

Decompose now $u^*(x,z)$ into $\phi_k^*(x,z)$ such that

$$u^*(x,z) = \sum_{k=1}^{K-1}(v_{k+1} - v_k)\phi_k^*(x,z)$$

and such that the bounds on u^* can be translated into

$$\xi_k^L(x,z)(G_k(x,z) - H_k(x,z)) \leq \phi_k^*(x,z) \leq \xi_k^U(x,z)(H_{k+1}(x,z) - G_k(x,z)). \tag{B.12}$$

There are many decompositions of this type. Choose one.

Second, consider κ a discrete random variable whose support is $\{1,\ldots,K-1\}$, which is independent of v^* (a.e. $F_{x,z}$) and whose conditional on (x,z) distribution is

$$\Pr(\kappa = k \mid x,z) = H_{k+1}(x,z) - H_k(x,z). \tag{B.13}$$

Consider also $K-1$ random variable λ_k whose support is $(0,1)$, which are independent of v^* (a.e. $F_{x,z}$) and whose conditional (on (x,z)) expectation is

$$E(\lambda_k \mid x,z) = \frac{H_{k+1}(x,z) - G_k(x,z) - \phi_k^*(x,z)}{H_{k+1}(x,z) - H_k(x,z)} \tag{B.14}$$

and such that

$$\int_0^1 (\Phi_v(\lambda v_k + (1-\lambda)v_{k+1} \mid k,x,z) - \frac{v - v_k}{v_{k+1} - v_k})dF_{\lambda_k}(\lambda \mid x,z) = \frac{\phi_k^*(x,z)}{H_{k+1}(x,z) - H_k(x,z)}.$$

Given constraints (B.11) and (B.12), it is always possible to construct such a random variable.

Finally, define the random variable

$$w = (1-\lambda)v_\kappa + \lambda v_{\kappa+1}.$$

By construction, the support of w is $[v_1, v_K)$ and w is independent of v^* conditionally on (x,z) because all λ_ks and κ are. Hence, $\varepsilon = -(x\beta + w)$ satisfies $(L.1)$ and $(L.2)$.

Finish the proof as in Theorem 4. ‖

Proof of Corollary 6. (*Necessity*) Let the conditional distribution of υ, Φ_0, be piecewise uniform by intervals, $\upsilon^* = k$. Then, for any $k = 1, \ldots, K-1$, $\xi_k^U(x,z) = \xi_k^L(x,z) = 0$. Using Theorem 5 yields that $\underline{\Delta}^*_\Phi(x,z) = \overline{\Delta}^*_\Phi(x,z) = 0$ and therefore $u^*(x,z) = 0$. Identification of β is exact and its value is given by the moment condition (4).

(*Sufficiency*) By contraposition assume that there exists $k \in \{1, \ldots, K-1\}$, a measurable set A included in $[\upsilon_k, \upsilon_{k+1})$ with positive Lebesgue measure and a measurable set S of elements (x,z) with positive probability $F_{x,z}(S) > 0$ such that $\Phi(\upsilon \mid k, x, z)$ is different from a uniform distribution function on A for any (x,z) in S. Because Φ is absolutely continuous (ID.ii), and for the sake of simplicity, assume that

$$\forall \upsilon \in A; \forall (x,z) \in S; \Phi(\upsilon \mid k, x, z) - \frac{\upsilon - \upsilon_k}{\upsilon_{k+1} - \upsilon_k} > 0.$$

Because $\xi_k^U(x,z) > 0$, we can always construct a function $u_1^*(x,z)$, which is such that $E(z' u_1^*(x,z))$ is strictly positive on S and satisfies the conditions of Theorem 5. The moment condition (4) is then used to construct parameter β_1. It implies that the identified set B contains at least two different parameters β, that is, the one corresponding to $u^*(x,z) = 0$ and the one corresponding to $u_1^*(x,z)$ (and in fact the whole real line between them as B is convex). ‖

Proof of Corollary 9.

Same as Corollary 2 except that the maximization of $E(\tilde{x}_p u^*(x,z))$ is obtained when

$$u^*(x,z) = \mathbf{1}\{\widetilde{x_p} \le 0\}\underline{\Delta}^*(x,z) + \mathbf{1}\{\widetilde{x_p} > 0\}\overline{\Delta}^*(x,z)$$

and the minimization of such an expression is obtained when

$$u^*(x,z) = \mathbf{1}\{\widetilde{x_p} > 0\}\underline{\Delta}^*(x,z) + \mathbf{1}\{\widetilde{x_p} \le 0\}\overline{\Delta}^*(x,z).$$

Furthermore, we have

$$\begin{aligned}
\bar{\Delta}^*(x,z) &= \sum_{k=1}^{K-1} \left[(\upsilon_{k+1} - \upsilon_k)(G_{k+1}(x,z) - G_k(x,z))\right] \\
&= -(\upsilon_2 - \upsilon_1)G_1(x,z) + \sum_{k=2}^{K-1} (\upsilon_k - \upsilon_{k-1} - (\upsilon_{k+1} - \upsilon_k))G_k(x,z) + \upsilon_K - \upsilon_{K-1} \\
&= \sum_{k=1}^{K-1} (\upsilon_k - \upsilon_{k-1} - (\upsilon_{k+1} - \upsilon_k))E(y \mid \upsilon = \upsilon_k, x, z) + \upsilon_K - \upsilon_{K-1} \\
&= E\left(\frac{\theta_{U,k} \cdot y}{p_k(x,z)} \mid x, z\right) + \upsilon_K - \upsilon_{K-1} = E(\bar{y}_U \mid x, z)
\end{aligned}$$

where by convention $\upsilon_0 = \upsilon_1$. Similarly

$$\begin{aligned}
\underline{\Delta}^*(x,z) &= \sum_{k=1}^{K-1} \left[(\upsilon_{k+1} - \upsilon_k)(G_{k-1}(x,z) - G_k(x,z))\right] \\
&= \sum_{k=1}^{K-2} (\upsilon_{k+2} - \upsilon_{k+1} - (\upsilon_{k+1} - \upsilon_k))G_k(x,z) - (\upsilon_K - \upsilon_{K-1})G_{K-1}(x,z) \\
&= \sum_{k=1}^{K-1} (\upsilon_{k+2} - \upsilon_{k+1} - (\upsilon_{k+1} - \upsilon_k))E(y \mid \upsilon = \upsilon_k, x, z) \\
&= E\left(\frac{\theta_{L,k} \cdot y}{p_k(x,z)} \mid x, z\right) = E(\bar{y}_L \mid x, z)
\end{aligned}$$

if the convention $\upsilon_{K+1} = \upsilon_K$ is adopted. ‖

B.1. *Design of Monte Carlo experiments*

The construction of the special regressor υ, the covariate x_2, the instrument z and the random shock ε proceeds in several steps. To begin with, we consider four random variables : e_1 is uniform on [0, 1], e_2 and e_3 are zero mean unit variance normal variates, and e_4 is a mixture of a normal variate $N(-0{\cdot}3, 0{\cdot}91)$ using a weight of 0·75 and a normal variate $N(0{\cdot}9, 0{\cdot}19)$ using a weight of 0·25. Using these notations, we define

$$\eta = 2e_2 + \alpha e_4, \qquad x_2 = e_1 + e_4$$
$$\varepsilon = \rho(e_1 - 0{\cdot}5) + e_3, \qquad z = e_4.$$

where α is a parameter that makes the random shock a non-normal variate and ρ is a parameter that renders x_2 endogenous. The case where $\alpha = \rho = 0$ (respectively $\alpha = \rho = 1$) roughly corresponds to what Lewbel calls the simple (respectively messy) design.

Regressor v is defined by truncating η to the 95% symmetric interval around 0, denoted $[v_1, v_K]$. To comply with assumption $L.2$, we then truncate $x_2 + \varepsilon$ so that $1 + x_2 + \varepsilon + v_1 < 0$ and $1 + x_2 + \varepsilon + v_K > 0$. We construct the censored $K - 1$ intervals using

$$v^* = k \quad \text{if } v \in [v_k, v_{k+1}).$$

Acknowledgements. We thank Arthur Lewbel and Francesca Molinari for helpful discussions, the editors and two referees for their constructive comments, and participants at seminars at LSE, CREST, CEMFI, Toulouse, Montréal and Rochester, and at conferences (ESRC Econometrics Study Group '04 in Bristol, ESWC'05, Northwestern U.'05, Econometrics in Rio '06) for helpful comments. The usual disclaimer applies.

REFERENCES

ANDERSON, T. W. and RUBIN, H. (1949), "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations", *Annals of Mathematical Statistics*, **20**, 46–63.

ANDREWS, D. W. K. and GUGGENBERGER, P. (2007), "Validity of Subsampling and Plug-in Asymptotic Inference for Parameters Defined by Moment Inequalities" (Unpublished manuscript, Cowles Foundation, Yale University).

BIERENS, H. J. and HARTOG, J. (1988), "Nonlinear Regression with Discrete Explanatory Variables with an Application to the Earnings Function", *Journal of Econometrics*, **38**, 269–299.

BERESTEANU, A. and MOLINARI, F. (2006), "Asymptotic Properties for a Class of Partially Identified Models" (Working Paper, Cornell University).

BONTEMPS, C., MAGNAC, T. and MAURIN, E. (2007), "Set Identified Linear Models" (Unpublished Manuscript, Toulouse School of Economics).

CAMERON, S. and HECKMAN, J. (1998), "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts Of American Males", *Journal of Political Economy*, **106** (2), 262–333.

CHERNOZHUKOV, V., HONG, H. and TAMER, E. (2007), "Inference on Parameter Sets in Econometric Models", *Econometrica*, **75**, 1243–1284.

CHESHER, A. (2003), "Non Parametric Identification under Discrete Variation", *Econometrica*, **73**, 1525–1550.

FRÉCHET, M. (1951), "Sur les tableaux de corrélation dout les marges sont données", *Annales de l'Université de Lyon*, 111° Série Sci., A, 53–77.

GALICHON A. and HENRY M. (2006), "Inference in Incomplete Models" (Unpublished Manuscript, Columbia University).

GREEN, D., JACOWITZ, K. E., KAHNEMAN, D. and MCFADDEN, D. (1998), "Referendum Contingent Valuation, Anchoring and Willingness to Pay for Public Goods", *Resource and Energy Economics*, **20**, 85–116.

HONORÉ, B. and LEWBEL, A. (2002), "Semiparametric Binary Choice Panel Data Models without Strict Exogeneity", *Econometrica*, **70**, 2053–2063.

HOROWITZ, J. (1998), *Semiparametric Methods in Econometrics* (Berlin: Springer).

HOROWITZ, J. and MANSKI, C. F. (2000), "Non Parametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data", *Journal of the American Statistical Association*, **95**, 77–84.

IMBENS, G. and MANSKI, C. F. (2004), "Confidence Intervals for Partially Identified Parameters", *Econometrica*, **72**, 1845–1859.

LEAMER, E. E. (1987), "Errors in Variables in Linear Systems", *Econometrica*, **55** (4), 893–909.

LEWBEL, A. (2000), "Semiparametric Qualitative Response Model Estimation With Unknown Heteroskedasticity or Instrumental Variables", *Journal of Econometrics*, **97**, 145–177.

LEWBEL, A. (2006), "Endogenous Selection or Treatment Model Estimation", *Journal of Econometrics* (forthcoming).

LEWBEL, A., LINTON, O. and MCFADDEN, D. (2006), "Estimating Features of a Distribution from Binomial Data" (Unpublished Manuscript, Boston College).

MAGNAC, T. and MAURIN, E. (2007), "Identification & Information in Monotone Binary Models", *Journal of Econometrics*, **139**, 76–104.

MANSKI, C. F. (1988), "Identification of Binary Response Models", *Journal of the American Statistical Association*, **83**, 729–738.

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

MANSKI, C. F. and TAMER, E. (2002), "Inference on Regressions with Interval Data on a Regressor or Outcome", *Econometrica*, **70**, 519–546.

MAURIN, E. (2002), "The Impact of Parental Income on Early Schooling Transitions: A Re-examination Using Data Over Three Generations", *Journal of Public Economics*, **85**, 301–332.

MARSCHAK, J. and ANDREWS, W. H. (1944), "Random Simultaneous Equations and the Theory of Production", *Econometrica*, **12**, 143–205.

MOFFITT, R. and RIDDER, G. E. (2006), "The Econometrics of Data Combination", in J. J. Heckman and E. E. Leamer (eds.) *Handbook of Econometrics*, Vol. 6 (Amsterdam: North-Holland).

POWELL, J. (1994),"Estimation of Semiparametric Models" , in R. Engle and D. McFadden (eds.) *Handbook of Econometrics*, Vol. 4 (Amsterdam: North-Holland) 2444–2521.

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