

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

Thierry Magnac* Eric Maurin†

First version: February 2004

This revision: December 2006

Abstract

We investigate identification in semi-parametric binary regression models, $y = 1(x\beta + v + \epsilon > 0)$ when ϵ is assumed uncorrelated with a set of instruments z , ϵ is independent of v conditionally on x and z , and the support of $-(x\beta + \epsilon)$ is finite. We characterize the set of observationally equivalent parameters β when v is discrete or when interval data only are available on v . When there exist as many instruments z as variables x , the sets within which lie the scalar components β_k of parameter β can be expressed as simple linear moments. Also, in the case of interval data, it is shown that additional information on the distribution of v within intervals shrinks the identification set. Namely, the closer to uniformity the conditional distribution of v given z is, the smaller the identification set is. Point identification is achieved if and only if v is uniform within intervals.

*University of Toulouse (GREMAQ and IDEI), 21 Allée de Brienne, 31000 Toulouse, France, *Address:* magnac@cict.fr, <http://idei.fr/member.php?i=280>

†PSE, 48 Bld Jourdan, 75014, Paris, France, *Address:* eric.maurin@ens.fr

1 Introduction¹

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

We analyse in this paper partial identification in a binary regression model with discrete or interval-valued data, a case that is admittedly specific although surprisingly rich in terms of implications and facility of application. The identified set is bounded and convex and bounds are easy to characterise by simple moments of the data.

One of the practical interest of our model stems from the fact that the data on covariates that researchers have access to, are very often discrete or interval-valued. Such covariates tend to render point identification very problematic (Manski, 1988). When all covariates (denoted x) are discrete, Bierens and Hartog (1988) showed that there exists 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 $E(y | x)$ 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 $E(y | x)$ is assumed monotonic with respect to at least one particular regressor, say v . As a special case, they study the identification of the parameter of the familiar semiparametric binary regression model $y = 1(x\beta + v + \epsilon > 0)$, when interval-data only are available on v . They analyse 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 the estimation of the identified region in this setting only requires usual regression

¹We thank Arthur Lewbel and Francesca Molinari for helpful discussions, the editor 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.

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., $F_\varepsilon(\varepsilon | x, v) = F_\varepsilon(\varepsilon | x)$) and a finite support assumption ($Supp(-x\beta - \varepsilon) \subset [v_l, v_u]$, where v_l and v_u are finite) restricts the parameters of the semiparametric binary regression model to a non-empty bounded and convex set. We characterise this set and show that the bounds of the intervals in which lie any scalar linear combination of parameter β can be estimated through simple linear regression methods. 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 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 – that are not necessarily observed in the data – the conditional probability of success varies from zero to one. 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 last interesting feature of the set-up analysed in this paper is that, in the case where v is censored by interval, 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 birthdate) into interval data (e.g. month of birth). In order to minimize the size of the identified set, one should censor the data in a way such that the distribution of the censored variable is as close as possible to a uniform distribution within the resulting intervals.

As for references, 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 (1952). A very general setting of partial identification is also provided by Galichon and Henry (2006). Our results on bounds on parameters in binary regressions can be seen as generalizations of bounds on averages, derived in the paper by Green, Jacowitz, Kahneman & McFadden (1998). They are also reminiscent of the results presented by Leamer (1987). He considers a system of equations where covariates are mismeasured and shows that the vector of parameters of interest lies in an ellipsoid. Also, there exist striking similarities between our identification results and those of Chesher (2005), even though the topic is quite different. Chesher estimates the local effect of an endogenous discrete variable in non-separable models and shows that discrete variation of this endogenous variable as opposed to continuous variation gives rise to partial identification.

This paper focuses on identification issues, not on inference problems. As a matter of fact, the inference issues are addressed by several recent works. Chernozhukov, Hong and Tamer (2004) study inference in multivariate cases under more general conditions of set-identification than ours and we show how their findings can be applied to our results. It is also possible to follow Horowitz and Manski (2000) who study inference about intervals *i.e.* the lower and upper bounds of identified intervals. An alternative route is proposed by Imbens and Manski (2004) who changed focus by considering inference about the true value of the parameter (within an identified interval) and not about the interval in itself. Finally, a recent contribution by Beresteanu and Molinari (2006) develops inference procedures for partially identified population features and discusses how these procedures could be adapted to the monotone binary model using interval data analyzed in this paper.

The paper is organized as follows : The first section sets up notations and models. The second section examines the discrete case, the third section analyzes the case of interval data, the fourth section briefly reports Monte Carlo experiments and the last section concludes. Since the special case where variables x are exogenous is not simpler, we will consider right from the start the endogenous case where ϵ , though potentially correlated with the variables x , is uncorrelated with a set of instruments z . Some results will be specialized 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 variable²:

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

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 & 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.³ 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 finite interval $\Omega_v \subset [v_1, v_K]$ almost everywhere- $F_{x,z}$ (a.e. $F_{x,z}$). This conditional distribution, denoted $F_v(\cdot | x, z)$, 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, $(F_v(\cdot | x, z).F_{x,z})$, is equal to 1.

Assumptions *R.i* and *R.ii* defines 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 & 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.

²We only consider random samples and we do not subscript individual observations by i .

³With 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 $x, x^2, \log(x), \dots$). 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).

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 & 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 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 :

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

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

Assumption L(aten)

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

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

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

(L.2) (*Support*) There exist two finite real numbers v_l and v_u such that the support of $-x\beta - \epsilon$ 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 semiparametric models, i.e. 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 (L1) and a more restrictive support assumption ($\text{Supp}(-x\beta - \varepsilon) \subset \text{Supp}(v)$) as identifying restrictions.⁴

As defined by (L.1) and (L.2), conditional independence and support assumptions restrict the class of statistical models that can actually be analyzed. 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 satisfies necessarily,

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

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 $\Pr(y_i = 1 \mid v = v_u, x, z) = 1$.

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

⁴There 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 $-x\beta - v_u$.

set outside the observed support of v . In such a case, $Pr(y = 1 | v, x, z)$ satisfies (NP.2) only if there are plausible values of v_l and v_u outside the observed support of v such that $Pr(y = 1 | v_u, x, z) = 0$ or $Pr(y = 1 | v_l, x, z) = 1$. By construction, this assumption is not testable. As discussed next, there are many examples where this assumption is plausible, but the case should be argued in each specific application.

2.2 Examples

Potential applications of Assumption L include controlled experiments where $y^* = x\beta + \epsilon$ represent a latent failure time and where individuals (or animals, or equipments) 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 $y = 1(x\beta + \epsilon > -v)$ 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 it can be assumed that the probability of "failure" varies from zero at the beginning of the experiment to one after a sufficient long (or strong) exposition to the treatment.

A second type of applications are optimal investment models. For example, in optimal schooling models, $y^* = x\beta + \epsilon$ represents the number of years of post-compulsory education which maximises discounted lifetime wealth⁵ 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 boils down to assuming 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 $y^* = x\beta + \epsilon$ for a good or a resource, see e.g. Lewbel, Linton and McFadden (2006). Individuals are asked whether their willingness to pay exceeds a bid $-v$ chosen by experimental design. Again, we observe $y = 1(x\beta + \epsilon > -v)$ 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

⁵See for example Cameron and Heckman (1998).

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 the ability exceeds the 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 was possible to observe sufficiently young children at the entry into elementary school they would all have to repeat a grade, whereas sufficiently mature children would all be able to avoid grade repetition. 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 v .

2.3 Identifying Restrictions and Parameter of Interest

The relationship between our set-up and the one in Manski and Tamer (2002) are similar to the relationship between quantile independence (Manski, 1988) and Lewbel (2000) identifying restrictions. 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 that 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 v is continuously distributed and has large support (i.e. $\text{Supp}(-x\beta - \varepsilon) \subset [v_l, v_u]$), the latter paper shows that Assumption L is sufficient for exact identification of both β and $F_\varepsilon(\cdot | x, z)$.

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) $(v, x' = x + \delta_x, z, \varepsilon)$ is assigned to everyone of characteristics (v, x, z, ε) , namely a exogenous change δ_x in covariates holding the unobserved heterogeneity term ε constant. The counterfactual probability of success conditional on (v, x, z) is

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

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

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): $\Omega_v = \{v_1, \dots, v_K\}$, $v_k < v_{k+1}$ for any $k = 1, \dots, K - 1$.

We consider a binary reduced-form $\Pr(y = 1 \mid v, x, z)$ satisfying $(NP.1 - NP.2)$ and we ask whether there is a latent model $(\beta, F_\varepsilon(\cdot \mid x, z))$ satisfying assumptions $(L1 - L3)$ and generating through (LV) the values $\Pr(y = 1 \mid v = v_k, x, z)$, $k = 1, \dots, K$, taken by $\Pr(y = 1 \mid v, x, z)$ on the observed support of v . 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:

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

Using these notations, the transformation of the binary response variable which will be used to characterize the identified set is defined as:⁶

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

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 $k=1, \dots, K$, 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 $G_k(x, z)$, $k=1, \dots, K$, 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,

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

⁶For almost all (v, x, z) in its support, which justifies that we divide by $p_k(x, z)$. Division by zero is a null-probability event.

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 $F_\varepsilon(\cdot | x, z)$. Figure 1 provides an illustration of the results stated in Theorem 1. Given some (x, z) , the nodes represent the conditional probability distribution $G(v, x, z)$ as a function of v . $G(v, x, z)$ satisfies (NP.2) since it is equal to 0 at the lower bound ($v = -1$) and equal to 1 at the upper bound ($v = 1$). The other observed values of G are at $v = -.5, 0, 0.5$. By construction, a latent model (β, F_ε) generates G through (LV) if it satisfies $1 - F_\varepsilon(-x\beta - v | x, z) = G(v, x, z)$. Hence, the only compatible distribution functions of the shock ε are such that $1 - F_\varepsilon(-x\beta - v | x, z)$ is passing through the nodes at $v = -.5, 0, 0.5$. The only other restriction is that these distribution functions are non-decreasing within the rectangles between the nodes. An example is reported in the graph, but it is only one among many other possibilities. The total surface of the rectangles is given by function $2\Delta(x, z)$ and it measures the degree of our ignorance on the distribution of ε .

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 than 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 $u(x, z) = 0$,

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

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

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

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 $\Delta(x, z) \leq \Delta_M$, we have,

$$(\beta - \beta^*)'W(\beta - \beta^*) \leq E(u^2(x, z)) \leq E(\Delta(x, z)^2) \leq \Delta_M^2,$$

that 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 center 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 $(\beta, F_\varepsilon(\cdot | x, z))$, Proposition 2 proves that, for a sequence of support Ω_v indexed by Δ_M , the distance $d(\beta^*, B) = \inf_{b \in B} \|\beta^* - b\|$ between β^* and B converges to zero as $\Delta_M \mapsto 0$:

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

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

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

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 centered at β_p^* , the p -th component of β^* . Specifically, we have,

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

where \tilde{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 $E(|\tilde{x}_p| \Delta(x, z))$, it is worth emphasizing that $\Delta(x, z)$ can be rewritten $E(\tilde{y}_\Delta \mid x, z)$ where \tilde{y}_Δ , as \tilde{y} , is an affine function of y whose definition is given at the end of the proof of Proposition 3. Furthermore, by definition, $\beta_p^* = \frac{E(|\tilde{x}_p| \tilde{y})}{E(\tilde{x}_p^2)}$. 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 \tilde{x}_p and [3] the linear regression of $\tilde{y} + \tilde{y}_\Delta$, $\tilde{y} - \tilde{y}_\Delta$ on $|\tilde{x}_p|$. Estimation follows accordingly.

3.2.3 Characterisation and Construction of the Identified Region

We begin by characterizing the identified region 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 to the p -th explanatory variable, $s_q = xq$. Let B_q the identified interval associated to this explanatory variable.

Denoting \tilde{s}_q the remainder of the projection of s_q onto xQ , we can apply Proposition 3 and write,

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

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 $q' \cdot \beta$.⁷ 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 $q' \cdot \beta$ when β lies in B . This function is known as the support function of B at q . It is denoted $\delta^*(q | B)$ and it is equal to:

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

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 halfspaces (Rockafellar, 1970),

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

Interestingly enough, we have an analytical definition of the support function of B . It makes it possible to construct B very easily by simulation. Randomly draw S vectors q_s , $\|q_s\| = 1$, construct the half-spaces $\{q'_s \beta \leq \delta^*(q_s | B)\}$ 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

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

which do not depend 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 $v_l \rightarrow -\infty$ or $v_u \rightarrow +\infty$. This result was obtained by Magnac and Maurin (2007) when v is continuous.

⁷It 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.

Regarding inference, Horowitz and Manski (1998), Imbens and Manski (2004) or Chernozhukov et al. (2004) provide tools that can be applied to estimate either the intervals of interest using the former two articles and the whole set B using the latter. In particular, Chernozhukov et al. (2004) provide confidence regions for sets of parameters that correspond to the zeroes of a non-negative continuous econometric criterion functions $Q(\beta)$. In our case,

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

where $d\mu(q)$ can be any strictly positive finite measure on the unit sphere \mathbb{S} . The sample analog Q_n of Q is not difficult to construct since $\delta^*(q | B)$ can be estimated through simple linear regressions. In such a context, assuming various regularity conditions, Chernozhukov et al. (2004) show how to construct confidence sets C_n such that $\lim_{n \rightarrow \infty} P(B \subset C_n) = \alpha$ or $\lim_{n \rightarrow \infty} \sup_{\beta \in B} P(\beta \subset C_n) = \alpha$ for a prespecified confidence level $\alpha \in (0, 1)$.

3.3 Supplementary 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 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:

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

As previously, the set on the RHS 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 work in progress (Bontemps, Magnac and Maurin, 2006) 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 set of parameters observationally equivalent to the true structural parameter has a similar structure to 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 $F_v(\cdot | v^*, x, z)$ and is assumed to be absolutely continuous. Its density function denoted $f_v(\cdot | v^*, x, z)$ 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 $-x\beta - \epsilon$ is included in the support of v), the moment condition (L.3) (i.e., $E(z'\epsilon) = 0$) and the following extension of the partial independence hypothesis,

$$F_\epsilon(\cdot | v, v^*, x, z) = F_\epsilon(\cdot | x, z) \quad (L.1^*)$$

The conditional probability distributions $Pr(y = 1 | v^*, x, z)$ 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. We thus restrict ourselves to this set of assumptions which is, not only in this respect, 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 exactly the monotonicity restriction imposed on $Pr(y = 1 | v^*, x, z)$ by our latent model. Also their Mean Independence assumption (MI) is a consequence of our assumption (L.1*), ours being slightly stronger. We however 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.

In analogy with 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 :

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

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

With these notations, the following theorem gives an exact 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 $Pr(y = 1 | v^*, x, z)$ (denoted $G_{v^*}(x, z)$) a conditional distribution function which is non decreasing in v^* . The two following statements are equivalent,

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

a. $(\beta, F_\varepsilon(. | x, z))$ satisfies (L.1*, L.2, L.3)

b. $G_{v^*}(x, z)$ is the image of $(\beta, F_\varepsilon(. | x, z))$ 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, $G_0(x, z) = 0$, $G_K(x, z) = 1$),

$$\begin{aligned}\overline{\Delta}^*(x, z) &= \sum_{k=1, \dots, K-1} (G_{k+1}(x, z) - G_k(x, z))(v_{k+1} - v_k), \\ \underline{\Delta}^*(x, z) &= - \sum_{k=1, \dots, K-1} (G_k(x, z) - G_{k-1}(x, z))(v_{k+1} - v_k),\end{aligned}$$

and such that,

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

Proof. See Appendix B ■

Theorem 4 provides a characterization of the set of latent models satisfying (L1* – L3) and generating $G_k(x, z)$, $k = 1, \dots, K - 1$, through (LV), when $[v_l, v_u)$ coincides with the observed support of v , $[v_1, v_K)$. If the observed support were $[v_2, v_{K-1})$ rather than $[v_l, v_u)$, the characterization of the identified set would follow similar lines.

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 .⁸ 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 $\Phi(v | v^*, x, z)$.

The first unsurprising result is that additional knowledge on $\Phi(v | x, z, v^*)$ actually helps to shrink the identified set. Secondly, knowing how identification is related to the conditional distribution $\Phi(v | x, z, v^*)$ 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 piece-wise uniform.

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

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

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

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

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 $\frac{\Phi - U}{\Phi}$ (resp. $\frac{\Phi - U}{1 - \Phi}$) is continuous and equal to zero at v_{k+1} (resp. v_k), the supremum of this function in the neighborhood of v_{k+1} (resp. v_k) is clearly non negative (resp. non

⁸Moffitt and Ridder (2006) provides a survey of two-sample techniques for such data design.

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 zero when Φ is equal to U . Using additional information, Theorem 4 is translated into:

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:*

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

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

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 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 piece-wise 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, i.e.:*

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

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 page 524). Their conditions (c) and (d) for point identification imply that the probability that the interval of observation of v (denoted v_0, v_1 in their notations) is as small as we want, is positive. 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 out of the scope of this paper.

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

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 $k \in \{1, \dots, K - 1\}$:

$$\begin{aligned} \xi_{k,1}^L(x, z) &\leq \xi_{k,2}^L(x, z) \\ \xi_{k,1}^U(x, z) &\geq \xi_{k,2}^U(x, z). \end{aligned}$$

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 the associated region of identification for β . Then:

$$\Phi_1 \succeq \Phi_2 \implies B_{\Phi_2} \subseteq B_{\Phi_1}$$

Proof. Straightforward using Theorem 5. ■

Assuming that we have some control on the construction on v^* (i.e., on the information on v that are made available to researchers), this result shows that this variable should

be constructed in 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 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 exist 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, so that it is uniformly distributed within months in most countries. In such a case, our results show that we only have to made available the month of birth of respondents (and not necessarily their exact date-of-birth) to achieve exact identification of 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 be easily 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.

Let:

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

be the projected set corresponding to the last (say) coefficient. All scalars belonging to this interval, are observationally equivalent to the p th component of the true parameter. 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 center 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 :

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

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

with \tilde{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 become 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 to 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.

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 $\bar{\sigma}^2$ the average sampling error defined as:

$$\bar{\sigma}^2 = (\sigma_u^2 + \sigma_b^2 + \sigma_{ub})/3$$

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 standard errors 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 $[\hat{\theta}_b, \hat{\theta}_u]$:

$$\begin{aligned} MSEI &= E \int_{\hat{\theta}_b}^{\hat{\theta}_u} (\theta - \theta_0)^2 \frac{d\theta}{\hat{\theta}_u - \hat{\theta}_b} = \frac{1}{3} E \left[\frac{(\hat{\theta}_u - \theta_0)^3 - (\hat{\theta}_b - \theta_0)^3}{\hat{\theta}_u - \hat{\theta}_b} \right] \\ &= \frac{1}{3} E \left[(\hat{\theta}_u - \theta_0)^2 + (\hat{\theta}_b - \theta_0)^2 + (\hat{\theta}_b - \theta_0)(\hat{\theta}_u - \theta_0) \right], \\ &= (\bar{\theta}_m - \theta_0)^2 + \frac{1}{3} ((\bar{\theta}_u - \bar{\theta}_b)/2)^2 \\ &\quad + \frac{1}{3} E \left[(\hat{\theta}_u - \bar{\theta}_u)^2 + (\hat{\theta}_b - \bar{\theta}_b)^2 + (\hat{\theta}_b - \bar{\theta}_b)(\hat{\theta}_u - \bar{\theta}_u) \right] \end{aligned}$$

The first term is the square of a “decentering” term (denoted *Dec*) which can be interpreted as a 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 standard errors (*ASE*) which can be interpreted as the effect of sample variability. 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 or 1000 observations. In all experiments, the number of Monte Carlo replications is equal to 1000. Additional replications do not affect any estimates (resp. standard errors) by more than a 1% margin of error (resp. 3%). Generally speaking, in all experiments, the true value of the parameter belongs to the 95% confidence interval built up around the estimates of the lower and upper bounds.

Table 1 presents the results of an experiment using interval data by reporting variations of *Dec*, *AL*, *ASE*, and $MSEI^{1/2}$ when both the size of the sample and the bandwidth used to compute the denominator of transformation (3) vary. It shows that the estimated intervals can be severely decentered for the intercept term, especially when the sample size

is small. Increasing the bandwidth decenters interval estimates for the coefficient of the variable towards the negative numbers though at a much lesser degree. The mean square error (*MSEI*) for the intercept decreases with the bandwidth, especially when the size of the sample is small. It has frequently a U-shape form for the coefficient of the variable. We have tried to look for a data-driven choice of the bandwidth by minimizing this quantity, but it was inconclusive. A larger bandwidth seems to be always preferred. Some further research is clearly needed on this issue. In the working paper version⁹, 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 v . It is shown that decentering can be quite severe when the degree of non-normality of the random shock or the degree of endogeneity of the covariate are large. Interval length is not affected by non-normality, but exhibits some non monotonic variations with the amount of endogeneity. Lastly, we find that interval length decreases with the number of points of the support of v , as predicted by the theory. This decrease is not much affected by sample sizes.

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 might 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 2 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 stems from the deterministic nature of one of the covariate (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 true set being around 75% of the area of the rectangle, with some variation according to the size of the sample.

⁹See <http://www.idei.fr/doc/wp/2005/magnac.pdf>

6 Conclusion

In this paper, we explored partial identification of coefficients of binary variable models when the very exogenous regressor is discrete or interval-valued. We derived 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 seem to be worthwhile considering. First, we do not provide proofs of consistency and asymptotic properties of the estimates of the bounds of the intervals because they would add little to the ones Lewbel (2000) presents. 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 some 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 object of work in progress.

Generally speaking, the identification results obtained in this paper when data are not continuous may be used to enhance identification power when the data are actually continuous. Specifically, if the support of the continuous very exogenous regressor is not large enough, one could use additional measurements or structural priors at discrete points at 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 analyzed by using simultaneously the proofs used in the discrete, interval or continuous settings. An interesting situation corresponds to a binary variable whose probability of occurrence is known to be monotone in some regressor v and varies between 0 and 1 in a known interval. 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 the very exogenous regressor is complete, discrete, by interval or continuous. Two extreme cases lead to exact identification, i.e. complete and continuous observation in the interval on the one hand, and, on the other hand, complete & interval-data observation when the distribution of the very exogenous regressor is uniform within intervals. Other

cases are nevertheless 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.*, 2006).

REFERENCES

- Bierens, H.J., and J., Hartog**, 1988, "Nonlinear Regression with Discrete Explanatory Variables with an Application to the Earnings Function", *Journal of Econometrics*, 38:269-299.
- Beresteanu, A. and F., Molinari**, 2006, "Asymptotic Properties for a Class of Partially Identified Models", working paper.
- Bontemps, C., T. Magnac and E. Maurin**, 2006, "Set Identified Linear Models", unpublished manuscript.
- Chernozhukov, V., H. Hong, E. Tamer**, 2004, "Inference on Parameter Sets in Econometric Models", unpublished manuscript.
- Chesher, A.**, 2003, "Non Parametric Identification under Discrete Variation", *Econometrica*, 73:1525-1550.
- Galichon A., and M., Henry**, 2006, "Inference in Incomplete Models", Unpublished Manuscript.
- Green, D., K.E. Jacowitz, D. Kahneman, D. McFadden**, 1998, "Referendum Contingent Valuation, Anchoring and Willingness to Pay for Public Goods", *Resource and Energy Economics*, 20:85-116.
- Honoré, B., and A., Lewbel**, 2002, "Semiparametric Binary Choice Panel Data Models without Strict Exogeneity", *Econometrica*, 70:2053-2063.
- Horowitz, J.**, 1998, *Semiparametric methods in Econometrics*, Springer: Berlin.
- Horowitz, J., and C.F., Manski**, 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 C.F., Manski**, 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*, 97:145-77.
- Lewbel, A.**, 2006, "Endogenous Selection or Treatment Model Estimation," forthcoming *Journal of Econometrics*.
- Lewbel, A., O., Linton and D., McFadden**, 2006, "Estimating Features of a Distribution From Binomial Data", unpublished manuscript.
- Magnac, T., and E., Maurin**, 2007, "Identification & Information in Monotone Binary Models", forthcoming *Journal of Econometrics*.
- 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*, Springer-Verlag: Berlin.
- Manski, C.F., and E. Tamer**, 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 W.H. Andrews**, 1944, "Random Simultaneous Equations and the theory of Production", *Econometrica*, 12:143-205.

Moffitt, R. and G. E. Ridder, (2006), "The econometrics of data combination", in J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6, North-Holland, Amsterdam.

Powell, J., 1994, "Estimation of Semiparametric Models", in eds. R. Engle and D. McFadden, *Handbook of Econometrics*, 4:2444-2521.

Rockafellar, R.T., 1970, *Convex Analysis*, Princeton University Press: Princeton.

A Proofs in Section 3

A.1 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 $(\beta, F_\varepsilon(\cdot | x, z))$ 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,

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

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

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

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

As the support of $w = -(x\beta + \varepsilon)$ is included in $[v_0, v_{K+1}]$, we can also 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 $\sum_{k=1}^K (s_{k+1} - s_k) \mathbf{1}\{v_k > w\} = (s_{K+1} - s_{j(w)+1})$. Hence, we have :

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

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

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

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 equation, we use 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 (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 (v, x, z) is independent of v (a.e. $F_{x,z}$) and is such that:

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

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 (v, x, z) is independent of v and is given by:

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

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:

$$\varepsilon(\lambda, k) = -x\beta - \lambda v_{k-1} - (1 - \lambda)v_k$$

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

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

whose support is $(-x\beta - v_K, -x\beta - v_1]$ and which is independent of v (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) ,

$$\int_{\Omega(\varepsilon \mid x, z)} (s_{j(-x\beta - \varepsilon)+1} + x\beta + \varepsilon) f(\varepsilon \mid x, z) d\varepsilon =$$

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

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:

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

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

Remark: It is worth emphasizing that this proof also provides a characterization of the domain of observationally equivalent distribution functions F_ε , i.e. 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 κ is $\{2, \dots, K\}$, equation (A.4) holds and such that:*

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

where β verifies:

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

A.2 Proof of Corollary 3

For the sake of clarity, we start with the exogeneous 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 $\beta^* = (E(x'x))^{-1} E(x'\tilde{y})$. There exists a function $u(x)$ in $]-\Delta(x), \Delta(x)]$ such that $\beta - \beta^* = (E(x'x))^{-1} E(x'u(x))$ 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 \tilde{x}_p :

$$\tilde{x}_p = x_p - x_{-p} (E(x'_{-p}x_{-p}))^{-1} E(x'_{-p}x_p)$$

Applying the principle of Frish-Waugh, we have

$$\beta_p - \beta_p^* = (E((\tilde{x}_p)^2))^{-1} E(\tilde{x}_p u(x)),$$

since \tilde{x}_p is a scalar. The maximum (resp. minimum) of $E(\tilde{x}_p u(x))$ when $u(x, z)$ varies in $(-\Delta(x), \Delta(x)]$ is obtained by setting $u(x) = \Delta(x)\mathbf{1}\{\tilde{x}_p > 0\} - \Delta(x)\mathbf{1}\{\tilde{x}_p \leq 0\}$ (resp.

$u(x) = -\Delta(x)\mathbf{1}\{\tilde{x}_p > 0\} + \Delta(x)\mathbf{1}\{\tilde{x}_p \leq 0\}$). Hence $E(\tilde{x}_p u(x))$ lies between $-E(|\tilde{x}_p| \Delta(x))$ and $E(|\tilde{x}_p| \Delta(x))$ and the difference $\beta_p - \beta_p^*$ varies in:

$$\left(-\frac{E(|\tilde{x}_p| \Delta(x))}{E(\tilde{x}_p^2)}, \frac{E(|\tilde{x}_p| \Delta(x))}{E(\tilde{x}_p^2)} \right).$$

To show the reciprocal, consider any β_p in

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

Denote

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

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

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

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:

$$\begin{aligned} \tilde{x}_p &= z(E(z'z))^{-1}E(z'x_p) - \\ & z(E(z'z))^{-1}E(z'x_{-p}) [E(x'_{-p}z)(E(z'z))^{-1}E(z'x_{-p})]^{-1} E(x'_{-p}z)(E(z'z))^{-1}E(z'x_p) \end{aligned}$$

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

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

Specifically,

$$\begin{aligned} \Delta(x, z) &= \frac{(v_1 - v_0)}{2} G_1(x, z) + \sum_{k=2}^K \left[\frac{(v_k - v_{k-1})}{2} (G_k(x, z) - G_{k-1}(x, z)) \right] \\ &+ \frac{(v_{K+1} - v_K)}{2} (1 - G_K(x, z)) \\ &= \sum_{k=1}^K \frac{(v_k - v_{k-1} - (v_{k+1} - v_k))}{2} G_k(x, z) + \frac{v_{K+1} - v_K}{2} \\ &= E\left(\left[\sum_{k=1}^K \mathbf{1}(v = v_k) \frac{(v_k - v_{k-1} - (v_{k+1} - v_k))}{2p_k(x, z)} \right] y + \frac{v_{K+1} - v_K}{2} \mid x, z \right) = E(\tilde{y}_\Delta \mid x, z) \end{aligned}$$

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

B Proofs in Section 4

B.1 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 $R(vi)$, $F_v(\cdot \mid x, z, v^*)$ is absolutely continuous and its density function is denoted f_v . By assumption (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 $(L2)$. 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:

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

Note that this condition implies:

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

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

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

Using the definition of \bar{y} , the term $p_{v^*}(x, z)$ cancels out and using condition $(L1^*)$, the integral over dw on the one hand, and the sum and other integral on the other hand, can be permuted:

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

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 $v_{j(w)} \leq w < v_{j(w)+1}$. Distinguish three cases. First, when $v^* < j(w)$, the whole conditional support of v lies below w and,

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

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

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

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

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

Summing over values of v^* ,

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

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

$$E(\bar{y} | x, z) = -E(w | x, z) - u^*(x, z) = x\beta + E(\epsilon | x, z) - u^*(x, z). \quad (\text{B.4})$$

where

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

Integrating (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. Let write,

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

where:

$$\phi_k(x, z) = \int_{v_k}^{v_{k+1}} (F_v(w | k, x, z) + \frac{v_k - w}{v_{k+1} - v_k}) dF_w(w | x, z). \quad (\text{B.6})$$

By integration by parts, the first term is:

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

Therefore, using the convolution equation (B.1),

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

Using (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 $F_w(w = -(x\beta + v_{k+1}) | x, z) - F_w(w = -(x\beta + v_k) | x, z) > 0$. 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 $(\beta, F_\varepsilon(\cdot | x, z))$ satisfies (L.1*, L.2, L.3) and $G_{v^*}(x, z)$ is the image of $(\beta, F_\varepsilon(\cdot | x, z))$ through the transformation (LV).

To begin with, we are going to construct w . We proceed in three steps.

First, we choose a sequence of functions $H_k(x, z)$ such that $H_1 = 0$, $H_K = 1$, and such that:

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

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 for instance

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

for instance, $\theta(x, z) = \frac{\overline{\Delta}^*(x, z) - u^*(x, z)}{\overline{\Delta}^*(x, z) - \underline{\Delta}^*(x, z)}$. 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 .

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

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

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

$$E(\lambda | x, z) = \frac{\sum_{k=1}^{K-1} (v_{k+1} - v_k)(H_{k+1}(x, z) - G_k(x, z)) - u^*(x, z)}{\sum_{k=1}^{K-1} (v_{k+1} - v_k)(H_{k+1}(x, z) - H_k(x, z))} \quad (\text{B.9})$$

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

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

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:

$$w = (1 - \lambda)v_k + \lambda v_{k+1}$$

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

To construct v , 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:

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

where $F_\lambda(\cdot | x, z)$ 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 chose η such that

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

Within this framework, we define v by the following expression:

$$v = v_k + (v_{k+1} - v_k)\eta$$

Having defined w and v , we are now going to prove that the image of $(\beta, F_w(\cdot | x, z))$ through (LV) is $G_{v^*}(x, z)$ because it satisfies equation (B.1):

$$\begin{aligned} & \int_{v_k}^{v_{k+1}} F_w(v | x, z) \cdot f_v(v | k, x, z) dv = H_k(x, z) + \\ & + (H_{k+1}(x, z) - H_k(x, z)) \int_{v_k}^{v_{k+1}} \Pr(w = (1 - \lambda)v_k + \lambda v_{k+1} \leq v | x, z) \cdot f_v(v | k, x, z) dv = \\ & H_k(x, z) + (H_{k+1}(x, z) - H_k(x, z)) \int_0^1 \Pr(\lambda \leq \eta | x, z) \cdot f_\eta(\eta | k, x, z) d\eta = G_k(x, z) \end{aligned}$$

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

$$\begin{aligned} \phi_k(x, z) &= -G_k(x, z) + \int_{v_k}^{v_{k+1}} \frac{F_w(w | x, z)}{v_{k+1} - v_k} dw \\ &= -G_k(x, z) + H_{k+1}(x, z) - (H_{k+1}(x, z) - H_k(x, z))E(\lambda | x, z). \end{aligned}$$

Therefore,

$$\begin{aligned} \sum_{k=1}^{K-1} (v_{k+1} - v_k) \phi_k(x, z) &= \sum_{k=1}^{K-1} (v_{k+1} - v_k) (H_{k+1}(x, z) - G_k(x, z)) \\ &= \sum_{k=1}^{K-1} (v_{k+1} - v_k) (H_{k+1}(x, z) - H_k(x, z)) E(\lambda | x, z) = u^*(x, z). \end{aligned}$$

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

B.2 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:

$$\phi_k(x, z) = \int_{v_k}^{v_{k+1}} \left(\Phi(v | k, x, z) - \frac{v - v_k}{v_{k+1} - v_k} \right) dF_w(v | x, z).$$

We then have first:

$$\begin{aligned} \phi_k(x, z) &= \int_{v_k}^{v_{k+1}} \left(1 - \frac{\frac{v - v_k}{v_{k+1} - v_k}}{\Phi(v | k, x, z)} \right) \Phi(v | k, x, z) dF_w(v | x, z) \\ &\leq \sup_{v \in (v_k, v_{k+1})} \left(1 - \frac{\frac{v - v_k}{v_{k+1} - v_k}}{\Phi(v | k, x, z)} \right) \int_{v_k}^{v_{k+1}} \Phi(v | k, x, z) dF_w(v | x, z) \\ &= \xi_k^U(x, z) \int_{v_k}^{v_{k+1}} \Phi(v | k, x, z) dF_w(v | x, z). \end{aligned}$$

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

$$\begin{aligned} &\int_{v_k}^{v_{k+1}} \Phi(v | k, x, z) dF_w(v | x, z). \\ &= \int_{v_k}^{v_{k+1}} d[\Phi(v | k, x, z) F_w(v | x, z)] - \int_{v_k}^{v_{k+1}} F_w(v | x, z) d\Phi(v | k, x, z) \\ &= \int_{v_k}^{v_{k+1}} d[\Phi(v | k, x, z) F_w(v | x, z)] - G_k(x, z). \end{aligned}$$

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

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

The derivation of the lower bound follows the same logic:

$$\begin{aligned} \phi_k(x, z) &\geq \inf_{v \in (v_k, v_{k+1})} \left(-1 - \frac{\frac{v - v_{k+1}}{v_{k+1} - v_k}}{1 - \Phi(v | k, x, z)} \right) \int_{v_k}^{v_{k+1}} (1 - \Phi(v | k, x, z)) dF_w(v | x, z) \\ &\geq \xi_k^L(x, z) \left[\int_{v_k}^{v_{k+1}} d[(1 - \Phi(v | k, x, z)) F_w(v | x, z)] + G_k(x, z) \right] \end{aligned}$$

where $\xi_k^L(x, z)$ is defined in the text. Hence, using $F_w(v_k | x, z) \geq G_k(x, z)$, 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 $u^*(x, z)$ (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

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

Under this assumption, we are going to prove that there exists a distribution function of the random term ε such that $(\beta, F_\varepsilon(\cdot | x, z))$ satisfies (L.1*, L.2, L.3) and $G_{v^*}(x, z)$ is the image of $(\beta, F_\varepsilon(\cdot | x, z))$ through the transformation (LV), when the distribution function of the special regressor v is $\Phi(v | k, x, z)$. 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 $\{1, \dots, K-1\}$ such as:

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

and such as:

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

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:

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

There are many decompositions of this type. Choose one.

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

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

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:

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

and such that:

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

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_k s and κ are. Hence, $\varepsilon = -(x\beta + w)$ satisfies (L.1) and (L.2).

Finish the proof as in Theorem 4.

B.3 Proof of Corollary 6

(*Necessity*) Let the conditional distribution of v , Φ_0 , be piece-wise uniform by intervals, $v^* = k$. Then, for any $k = 1, \dots, K - 1$, $\xi_k^U(x, z) = \xi_k^L(x, z) = 0$. 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 $k \in \{1, \dots, K - 1\}$, a measurable set A included in $[v_k, v_{k+1})$ with positive Lebesgue measure and a measurable set S of elements (x, z) with positive probability $F_{x,z}(S) > 0$ such that $\Phi(v | k, x, z)$ 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:

$$\forall v \in A; \forall (x, z) \in S; \Phi(v | k, x, z) - \frac{v - v_k}{v_{k+1} - v_k} > 0$$

Because $\xi_k^U(x, z) > 0$, we can always construct a function $u_1^*(x, z)$ which is strictly positive on S satisfying the conditions of Theorem 5. Thus $E(z'u_1^*(x, z)) \neq 0$ and the moment condition (4) can be used to construct parameter β_1 . It implies that the identified set B contains at least two different parameters β , i.e. 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).

B.4 Proof of Corollary 9

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

$$u^*(x, z) = \mathbf{1}\{\tilde{x}_p \leq 0\} \underline{\Delta}^*(x, z) + \mathbf{1}\{\tilde{x}_p > 0\} \overline{\Delta}^*(x, z)$$

and the minimization of such an expression is obtained when:

$$u^*(x, z) = \mathbf{1}\{\tilde{x}_p > 0\} \underline{\Delta}^*(x, z) + \mathbf{1}\{\tilde{x}_p \leq 0\} \overline{\Delta}^*(x, z)$$

Furthermore, we have:

$$\begin{aligned} \overline{\Delta}^*(x, z) &= \sum_{k=1}^{K-1} [(v_{k+1} - v_k)(G_{k+1}(x, z) - G_k(x, z))] \\ &= -(v_2 - v_1)G_1(x, z) + \sum_{k=2}^{K-1} (v_k - v_{k-1} - (v_{k+1} - v_k))G_k(x, z) + v_K - v_{K-1} \\ &= \sum_{k=1}^{K-1} (v_k - v_{k-1} - (v_{k+1} - v_k))E(y | v = v_k, x, z) + v_K - v_{K-1} \\ &= E\left(\frac{\theta_{U,k} \cdot y}{p_k(x, z)} \mid x, z\right) + v_K - v_{K-1} = E(\bar{y}_U \mid x, z) \end{aligned}$$

where by convention $v_0 = v_1$. Similarly:

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

if the convention $v_{K+1} = v_K$ is adopted.

B.5 Design of Monte Carlo Experiments

The construction of the special regressor v , 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 $N(-.3, .91)$ using a weight of .75 and a normal variate $N(.9, .19)$ using a weight of .25. Using these notations, we define:

$$\begin{aligned}
\eta &= 2e_2 + \alpha e_4, & x_2 &= e_1 + e_4 \\
\varepsilon &= \rho(e_1 - .5) + e_3, & z &= e_4.
\end{aligned}$$

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 $\alpha = \rho = 0$ (resp. $\alpha = \rho = 1$) roughly corresponds to what Lewbel calls the simple (resp. 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:

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

Figure 1: A graphical argument for set-identification

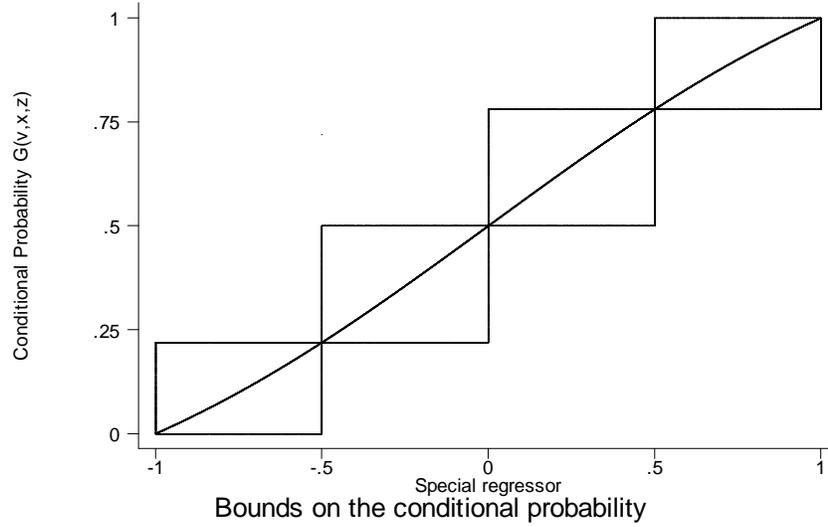
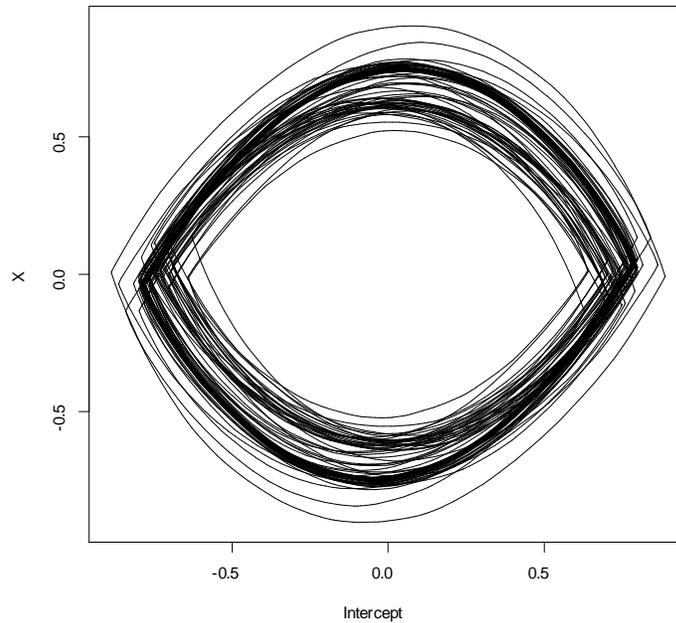


Figure 2:

Figure 2: Estimated sets



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

Table 1: Simple experiment, Interval Data: Sensitivity to Bandwidth

Lower and upper estimated bounds with standard errors

Error Decomposition: Decentering, Adjusted Length and Sampling Error

Nobs	Bwidth	Intercept				Variable			
		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

Notes: 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. **LB** and **UB** refer to the estimated lower and upper bounds of intervals with their standard errors (**SE**). **Bwidth** refers to the bandwidth that is used. **Dec** stands for decentering of the mid-point of the interval that is, $(\mathbf{UB} + \mathbf{LB})/2$. **AL** is the adjusted length of the interval, $(\mathbf{UB} - \mathbf{LB})/2\sqrt{3}$. **ASE** is the sampling variability of bounds as defined in the text. The identity $\mathbf{Dec}^2 + \mathbf{AL}^2 + \mathbf{ASE}^2 = \mathbf{RMSEI}^2$ is shown in the text. **RMSEI** is the root mean square error integrated over the identified interval.

Table 2: Approximating true sets by their projections on axes.

Observations	True Area	Square Area	Proportion	
			Mean	Standard Deviation
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

Notes: 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