**INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES**
**Série des Documents de Travail du CREST**
**(Centre de Recherche en Economie et Statistique)**

# n° 2004-11

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

**T. MAGNAC**[1]
**E. MAURIN**[2]



[1] University of Toulouse (INRA and IDEI) and CREST-INSEE. Email : tmagnac@ens.fr
*Address : 21 Allée de Brienne, 31000 Toulouse, France*. http://www.inra.fr/ESR/UR/lea/magnac.htm
[2] CREST-INSEE, 15 Boulevard Gabriel Péri, 92245 Malakoff Cedex, France. maurin@ensae.fr

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

Thierry Magnac[*] Eric Maurin[†]

First version: February 2004
This version: May 2004


## Abstract

We investigate inference in semi-parametric binary regression models, $y = 1(x\beta + v + \varepsilon > 0)$ when $\varepsilon$ is assumed uncorrelated with a set of instruments $z$, $\varepsilon$ is independent of $v$ conditionally on $x$ and $z$, and the conditional support of $\varepsilon$ is sufficiently small relative to the support of $v$. We characterize the set of observationally equivalent parameters $\beta$ when interval data only are available on $v$ or when $v$ is discrete. When there exist as many instruments $z$ as variables $x$, the sets within which lie the scalar components $\beta_k$ of parameter $\beta$ can be estimated by simple linear regressions. 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 distribution of $v$ is, the smaller the identification set is. Point identification is achieved if and only if $v$ is uniform within intervals.


[*] University of Toulouse (INRA and IDEI) and CREST-INSEE, 21 Allée de Brienne, 31000 Toulouse, France. tmagnac@ens.fr, http://www.inra.fr/ESR/UR/lea/magnac.htm

[†] CREST-INSEE, 15, Boulevard Gabriel Péri, 92245 Malakoff Cedex, maurin@ensae.fr

# 1 Introduction[1]

Data on covariates that researchers have access to, are very often discrete or interval-valued. There are many such examples in applied econometrics. Variables such as gender, levels of education, occupation, employment status or household size of survey respondents typically take a discrete number of values. In contingent valuation studies, prices are set by the experimenter and they are in general discrete, 1, 10, 100 or 1000 euros. It would sound funny to ask a person whether she wants to buy a salmon-fishing permit for 15 euros and 24 cents. 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 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, the French National Statistical Office, for instance, censors this information in the public versions of household surveys, by transforming dates of birth into months (or years) of birth only. French statisticians 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.

The problem is that discrete (or interval-valued) covariates tend to render inference in regressions very difficult. When all covariates are discrete or when only interval data are available, point identification of parameters of popular index models is lost whatever the identifying restrictions (Manski (1988)).[2] When all covariates (denoted $x$) are discrete, Bierens and Hartog (1988) have shown that there exists actually an infinite number of single-index representations for the mean regression of a dependent variable, $y$, i.e. $E(y \mid x) = \varphi_{\theta}(x\theta)$. Specifically, under weak conditions, the set of observationally equivalent parameters $\theta$ is dense in its domain of variation, $\Theta$.

A recent contribution by Manski and Tamer (2002) considers a less general framework

[1] We thank participants at seminars at LSE and CEMFI for helpful comments. The usual disclaimer applies.

[2] Though other parameters of interest such as the non-parametric mean regression might remain identified (see Angrist, 2001, and discussion).

where the non-parametric mean regression $E(y \mid x)$ is assumed monotonic with respect to at least one particular regressor, say $v$. They show that this assumption restricts the magnitude of under-identification when the "special" regressor, $v$, is not perfectly observed, i.e., when interval-data only are available on $v$. Under a quantile-independence assumption, what is identified is a non-empty, convex set of observationally equivalent values that they characterize. In other words, they achieve set-identification. Among other results, they also show that identified "sets" can be estimated by a modified maximum score technique (Manski, 1985).

In this paper, we explore the route of another weak identifying restriction in the semi-parametric binary models that has recently been introduced by Lewbel (2000). Consider the binary response model,

$$y = 1(x\beta + v + \varepsilon > 0)$$

where $y$ is the observed binary dependent variable, $x = (x_1, ..., x_p)$ are covariates, $v$ is an observed continuous explanatory variable (whose coefficient is set equal to 1 by normalisation) and $\varepsilon$ is an unobserved random variable. Lewbel proposed a simple estimator of $\beta$ under the combination of an uncorrelated-error assumption (i.e., $E(x'\varepsilon) = 0$) with a partial independence assumption (i.e., $F_\varepsilon(\varepsilon \mid x, v) = F_\varepsilon(\varepsilon \mid x)$) and a large support assumption ($Supp(-x\beta - \varepsilon) \subset Supp(v)$).[3] By adapting the partial independence assumption, Lewbel also developed an IV version of his estimator when $\varepsilon$, though correlated with $x$, is uncorrelated with a set of instrumental variables $z$. Recently, Honoré and Lewbel (2002) presented a fixed-effect version of this estimator. Generally speaking, these estimators are very appealing: they permit general form of endogeneity and conditional heteroskedasticity; Their implementation only requires estimating a conditional density function and a linear regression which means that no optimization is needed; They are root-n consistent under general conditions. Moreover, we showed in Magnac and Maurin (2003) that the set of latent models satisfying uncorrelated-errors (UE), large-support (LS) and partial-independence (PI) assumptions is isomorphic to the set of monotone-in-v non-parametric models where the

[3] In these semi-parametric models, the identification of $\beta$ requires that the distribution of the regressors has a sufficiently rich support. See Horowitz (1998) for a discussion on identification under quantile-independence when the support of the regressors is bounded. Magnac and Maurin (2003) provide alternative identifying assumptions under which $\beta$ remains identified under partial independence even when $v$ does not satisfy the large support assumption.

probability of success $E(y \mid x, v)$ varies between 0 and 1 (inclusive) over the support of $v$. As it turns out, the partial-independence assumption is congruent to the monotonicity assumption made by Manski & Tamer (2002).

In this paper, we investigate how these properties are translated when the special regressor is not continuous. To begin with, we show that the class of binary outcomes which can be analyzed through latent models satisfying (UE), (PI) and (LS) has exactly the same structure when $v$ is discrete as when it is continuous. Specifically, any binary outcome such that the probability of success increases from 0 and 1 over the support of $v$ can be analyzed through such latent models. The structural parameters of latent models satisfying (UE), (PI) and (LS) are "set"-identified if, and only if, the probability of success (conditional on $x$) is observed increasing from 0 to 1 (inclusive) over the support of $v$. In the discrete case, the identification is not exact anymore, however. The uncorrelated-error, large-support and partial-independence assumptions do not restrict the model parameters to a singleton (as when $v$ is continuous and perfectly observed), but to a non-empty, convex set. We explain how simple linear regression methods provide estimates of the bounds of the intervals in which lie each scalar components $\beta_k$ of parameter $\beta$.

We next ask whether it is possible to relax the large-support assumption (LS). This specific assumption restricts the domain of application of the latent models to the analysis of phenomena such that low-$v$ (high-$v$) persons all give the same response, namely $y = 0$ ($y = 1$), which is arguably restrictive. In Magnac and Maurin (2003), we studied the continuous case and proposed alternative assumptions on the distribution of $\epsilon$ which combination with (UE) and (PI) restores (exact) identification, whatever the support of $v$. In the discrete case, the question is whether it is possible to restore set-identification when the support of $v$ is not large, i.e., when the support of $(-x\beta - \varepsilon)$ is not included in the support of $v$. The answer is positive. As a matter of fact, the only additional assumption that is needed for set-identification (on top of UE and PI) is that the support of $(-x\beta - \varepsilon)$ is included in some finite interval $[v_0, v_{K+1}]$, regardless of whether this set coincides or not with the support of $v$ actually observed in the data. Put differently, what is necessary for set-identification is not to observe $E(y \mid x, v) = 0$ ($E(y \mid x, v) = 1$) at the minimum (maximum) value of $v$ observed in the available data, but to be able to impose these conditions as priors on the data generating process of $y$ at values of $v$ that are not necessarily observed in the data: As examplified below, this result makes it possible to analyze a class of binary phenomena which is substantially more general than when the support of $v$ is assumed to be large.

We next analyse the case where $v$ is continuous, but only observed by intervals. In such a case, the uncorrelated-error, large-support and partial-independence assumptions still restricts the model parameters to a non-empty, convex set (as in the discrete case), but the shape of this set and the methods for estimating it are somehow different from the discrete case.

Lastly, we analyze the case where some information is available on the distribution of $v$ within intervals. Most interestingly, the "size" of the identification set diminishes as the distribution of the special regressor within intervals becomes closer to uniformity. When $v$ is uniformly distributed within intervals, the identification set is a singleton and the parameter of interest $\beta$ is exactly identified. 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 identification set, one should censor the data in such a way that the distribution of the censored variable is the closest as possible to a uniform distribution within the resulting intervals.

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 reports Monte Carlo experiments and the last section concludes. All proofs are in appendices.

Since the case where the $x$ are endogenous is not more complex than the case where they are exogenous, we will consider right from the start the endogenous case where $\varepsilon$, though potentially correlated with the variables $x$, is uncorrelated with a set of instruments $z$.

# 2 The Set-Up

Let the "data" be given by the distribution of the following random variable[4]:

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

[4] We only consider random samples and we do not subscript individual observations by $i$.

where $y$ is a binary outcome, while $v$, $x$ and $z$ are covariates and instrumental variables which role and properties are specified below. We first introduce some regularity conditions on the distribution of $\omega$. 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.[5] The condition of full rank, $\text{rank}(E(z'x)) = p$, holds.

R.iii: (Special Regressor) The support of the conditional distribution of $v$ conditional on $(x, z)$ is $\Omega_v \subset \mathbb{R}$ almost everywhere-$F_{x,z}$ (a.e $F_{x,z}$). This conditional distribution; denoted $F_v(. \mid x, z)$; is defined a.e. $F_{x,z}$. In the remainder we will assume either $\Omega_v = \{v_1, ..., v_K\}$ (discrete case) or $\Omega_v = [v_1, v_K[$ (interval case) where $v_1$ and $v_K$ are finite. In both cases $\Omega_v^0$ will denote $[v_1, v_K[$.

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

Assumption R.i defines a binary model where there are $p$ explanatory variables and $q$ instrumental variables (assumption R.ii). Given assumption R.ii, we could denote the functionally independent description of $(x, z)$ as $u$ and this notation could be used interchangeably with $(x, z)$.[6] In assumption R.iii; the support of the special regression, $v$, is assumed to be independent of variables $(x, z)$. If this support is an interval in $\mathbb{R}$ (including $\mathbb{R}$ itself) and $v$ is perfectly observed, we are back to the case studied by Lewbel (2000) and Magnac & Maurin (2003). In the next section (section 3), this support is assumed to be discrete so that the special regressor is said to be discrete. In section 4, the support is assumed continuous, but $v$ is observed imperfectly, through censoring. 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.

[5] With no loss of generality, the $p$ explanatory variables $x$ can partially overlap with the $q \geq p$ instrumental variables $z$. Variables $(x, z)$ may also be functionally dependent (for instance $x$, $x^2$, $\log(x)$,...). A collection $(x_1, ., x_K)$ 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 $K$ which probability measure is equal to 1).

[6] Denoting $(x, z)$ as $u$ is used by Lewbel (2000) and leads to more exact arguments below at the cost of an additional notation. We prefer to stick to the more parsimonious notation $(x, z)$.

Assuming that the data satisfy $R.i - R.iv$, the basic issue adressed in this paper is whether they can be generated by the following semi-parametric latent variable index structure :

$$y = 1\{x\beta + v + \varepsilon > 0\}; \quad \text{(LV)}$$

where $1\{A\}$ is the indicator function that equals one if A is true and zero otherwise and where the random shock $\varepsilon$ satisfies the properties introduced by Lewbel (2000) and Honoré and Lewbel (2002),

Assumption L(atent)

(L.1) (Partial independence) The conditional distribution of $\varepsilon$ given covariates $x$ and variables $z$ is independent of the covariate $v$:

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

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

(L.2) (Large support) The support of $-x\beta - \varepsilon$ is a subset of $\Omega_v^0$ as defined in R(iii).

(L.3) (Moment condition) The random shock $\varepsilon$ is uncorrelated with variables $z$: $E(z'\varepsilon) = 0$.

The index parameter $\beta \in R^p$ is the unknown parameter of interest. The distribution function of the error term, $\varepsilon$, is also unknown and may be considered as a nuisance parameter. Assumptions $L.1 - L.3$ and some examples are commented in Lewbel (2000) or Magnac and Maurin (2003). Once $v$ is perfectly observed and continuously distributed, the latter paper shows that assumptions $L.1 - L.3$ are sufficient for exact identification of both $\beta$ and $F_\varepsilon(. \mid x, z)$. There is only a minor difference between assumptions $(L.1 - L.3)$ and the set-up introduced by Lewbel (2000), namely we do not restrain the distribution function $F_\varepsilon$ to 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[7].

In the remainder, any $(\beta, F_\varepsilon(. \mid x, z))$ satisfying $(L.1 - L.3)$ is called a latent model. Identification is studied in the set of all such $(\beta, F_\varepsilon(. \mid x, z))$.

[7] Given that $F_\varepsilon(. \mid x, z)$ is potentially discrete and assuming that 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 $-x\beta - v_K$.

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

To begin with, we are going to explore the properties that a conditional probability distribution $\Pr(y = 1 \mid v, x, z)$ necessarily satisfies when it is generated by a latent model $(\beta; F_\epsilon(\cdot \mid x, z))$ that satisfies conditions (L1-L3). The issue is to make explicit the class of binary outcomes which can actually be analyzed through the latent models under consideration.

## 3.1 Characterizing the Conditional Distribution

As defined by (L.1) and (L.2), partial-independence and large-support assumptions restrict the class of binary outcomes that can actually be analyzed. Restrictions are characterized by the following lemma :

**Lemma 1** Under partial independence (L.1) and large support (L.2) conditions:

(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) (Support) The conditional probability $\Pr(y_i = 1 \mid v, x, z)$ varies from 0 to 1 when $v$ varies over its support:

$$\Pr(y_i = 1 \mid v = v_1, x, z) = 0; \quad \Pr(y_i = 1 \mid v = v_K, x, z) = 1.$$

*Proof.* See Appendix A ■

If a binary outcome does not satisfy (NP.1) or (NP.2) then there exists no latent model generating the reduced form $\Pr(y = 1 \mid v, x, z)$. In other words, the monotonicity condition (NP.1) and the support condition (NP.2) are necessary conditions for the identification of the latent models considered in this paper. The next section studies whether the reciprocal holds true, i.e. whether (NP.1) and (NP.2) are sufficient conditions for identification.

## 3.2 Set-identification

We consider a binary reduced-form $\Pr(y = 1 \mid v, x, z)$ satisfying the monotonicity condition (NP1) and the support condition (NP2) and ask whether there exists a latent model $(\beta;$

$F_\varepsilon(. \mid x, z))$ generating this reduced-form through the latent variable transformation $(LV)$. To anticipate, we show that the answer is positive. The admissible latent model is not unique, however. There are many possible latent models which parameters are observationally equivalent.

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, for $k \in \{2, .., K-1\}$:

$$
\begin{aligned}
\delta_k &= (v_{k+1} - v_{k-1})/2 \\
p_k(x, z) &= \Pr(v = v_k \mid x, z).
\end{aligned}
$$

Using these notations, the counterpart adapted to the discrete case of the transformation of the binary response variable introduced by Lewbel (2000) is defined as:[8]

$$
\begin{aligned}
\tilde{y} &= \frac{\delta_k . y}{p_k(x, z)} - \frac{v_K + v_{K-1}}{2} \text{ if } v = v_k, \text{ for } k \in \{2, .., K-1\}, \qquad (1)\\
\tilde{y} &= -\frac{v_K + v_{K-1}}{2} \text{ if } v = v_1 \text{ or } v = v_K,
\end{aligned}
$$

If $v$ was continuous, the set of latent models satisfying $(L.1 - L.3)$ and generating $\Pr(y = 1 \mid v, x, z)$ through transformation $(LV)$ would be reduced to a singleton (Magnac and Maurin, 2003) and the parameter of interest $\beta$ would be uniquely defined by the instrumental regression of the transformation of the dependent variable on covariates. When $v$ is discrete, the identification of $\beta$ is not exact anymore as stated in the following theorem:

**Theorem 2** *Consider a conditional probability distribution, $\Pr(y = 1 \mid v = v_k, x, z)$, denoted $G_k(x, z)$, which satisfies conditions of monotonicity $(NP.1)$ and support $(NP.2)$. The two following statements are equivalent,*

*(i) there exists a vector of parameters $\beta$ and there exists a latent random variable $\varepsilon$ such that the latent model $(\beta, F_\varepsilon(. \mid x, z))$ satisfies conditions $(L.1 - L.3)$ and such that $\{G_k(x, z)\}_{k=1,..,K}$ is its image through the transformation $(LV)$;*

*(ii) there exists a vector of parameters $\beta$ and there exists a measurable function $u(x, z)$ from $S_{x,z}$ to $R$ which takes its values in the interval (a.e.$F_{x,z}$)*

$$
I(x, z) = ]-\Delta(x, z), \Delta(x, z)],
$$

[8] For 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. Obviously, this argument might need some adaptation in practice in finite samples.

where

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

and such that,

$$E(z'(x\beta - \nu)) = E(z'u(x,z)). \qquad (2)$$

Proof. See Appendix A. ∎

Theorem 2 characterizes the set of all observationally equivalent values of parameter $\beta$. We shall denote $B$ this identification set as the set of parameters $\beta$ which satisfies equation (2). As discussed in Appendix A, the proof of Theorem 2 also leads to a characterization of the set of observationally equivalent distribution functions $F_\varepsilon(. \mid x,z)$.

Before moving on to a more detailed discussion of the characteristics of set $B$, it is possible to provide an clarifying sketch of its proof by analyzing the trivial case, $K = 2$. Consider $(\beta, F_\varepsilon)$ satisfying $(L.1 - L.3)$ and its associated reduced form $G_k(x,z)$ for $k = 1,2$. By Lemma 1 and (NP2), trivially, $G_1(x,z) = 0$ and $G_2(x,z) = 1$. Restriction $(L.2)$ implies:

$$v_1 \leq -(x\beta + \varepsilon) < v_2$$

When $K = 2$, $\nu$ is equal to $-(v_2 + v_1)/2$ whatever $v$ and the previous condition can be rewritten:

$$-(v_2 - v_1)/2 \leq -(x\beta + \varepsilon) + \nu < (v_2 - v_1)/2 = \Delta(x,z)$$

Hence, if we define $u(x,z) = -E(\nu - x\beta - \varepsilon \mid x,z)$, it belongs to $]-\Delta(x,z), \Delta(x,z)]$ and satisfies (2), as stated by Theorem 2.

Reciprocally, assume that there exists $u(x,z)$ in $]-(v_2 - v_1)/2, (v_2 - v_1)/2]$ which satisfies condition (2). Consider a random variable $\lambda$ taking values in $]0,1]$ and such that:

$$E(\lambda \mid x,z) = \frac{1}{2} + \frac{u(x,z)}{v_2 - v_1}$$

Then consider the random variable, $\varepsilon = -x\beta - (1 - \lambda)v_1 - \lambda v_2$. By construction, it satisfies $v_1 < -(x\beta + \varepsilon) \leq v_2$. Hence, the model $(\beta, F_\varepsilon)$ satisfies $(L.1 - L.2)$ and generates $G_1(x,z) = 0$ and $G_2(x,z) = 1$ through (LV). The only remaining condition to check is $(L.3)$, namely $\varepsilon$ is uncorrelated with $z$. It is shown using condition (2) and the definition of $\lambda$.

Figure 1 provides an illustration of the results stated in Theorem 2. Given some $(x,z)$, the nodes represent the conditional probability distribution $G(v,x,z)$ as a function of the special regressor, $v$. In this example, $v$ satisfies ($NP.2$), namely the conditional probability is equal to 0 at the lower bound ($v = -1$) and equal to 1 at the upper bound ($v = 1$). The other observed values are at $v = -.5, 0, 0.5$. By construction, if $(\beta, F_\varepsilon)$ generates $G$ through ($LV$), it satisfies $1 - F_\varepsilon(-x\beta - v \mid x, z) = G(v, x, z)$. Hence, the only compatible distribution functions of the shock $\varepsilon$ are such that $1 - F_\varepsilon(-x\beta - v \mid x, z)$ is passing through the nodes at $v = -.5, 0, 0.5$. The only other restrictions are 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 $\varepsilon$.

The following section builds on Theorem 2 to provide a more detailed description of $B$, the set of observationally equivalent parameters.

## 3.3 Bounds on Structural Parameters and Overidentification

This section builds on Theorem 2 to provide a more 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 will briefly indicate how the results could be extended to the case where the number of instruments $z$ is larger that the number of $x$.

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 $\beta$ for any function $u(x, z)$ Because equation (2) is linear in $\beta$, the set $B$ is convex. Also it is non-empty, since it necessarily contains the pseudo-true value $\beta^*$ associated with the moment condition, $E(z'(x\beta^* - \tilde{y})) = 0$ when $u(x, z) = 0$.

The set $B$ can be described as a neighborhood of $\beta^*$ which size depends on the distances $(v_k - v_{k-1})$ between the different elements of the support of $v$. Specifically, $\beta^*$ can be interpreted as the specific value that $\beta$ would take if these distances were negligible. First, Theorem 2 makes possible to obtain very simple upper bounds for the potential bias that affects the result of the IV regression of $\tilde{y}$ on $x$. Denoting the half-length of the largest interval as

$$\Delta_M = \max_{k \in \{1,\dots,K\}} (v_k - v_{k-1})/2;$$

we have:

**Corollary 3** The identification set $B$ is non empty and convex. It contains the pseudo-true value $\beta^*$ defined as:

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

and any $\beta \in B$ satisfies,

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

where $W = E(x'z)(E(z'z))^{-1}E(z'x)$.

*Proof.* See Appendix A. ■

Corollary 3 shows that $B$ lies within an ellipsoïd whose size is bounded by $\Delta_M$. Notice that in the specific case where the different $v_k$ are equidistant (i.e., $\forall k = 3, \dots, K$, $v_k - v_{k-1} = v_2 - v_1$), $\Delta_M = \frac{v_2 - v_1}{2}$ and the half-length between two successive points provides an upper bound for the size of the ellipsoïd.

Returning to the general case, the maximum-length index, $\Delta_M$, can be taken as a measure of distance to continuity of the distribution function of $v$ (or of its support $\Omega_v$). For a latent model $(\beta; F_\varepsilon(. \mid x; z))$, corollary 3 proves that, for a sequence of support $\Omega_v$ indexed by $\Delta_M$; we have:

$$\lim_{\Delta_M \to 0} B = \{\beta^*\};$$

and exact identification is restored.

Identification set $B$ can be projected onto its elementary dimensions to better characterize the specific sets within which lie the different individual parameters. It can be done using the usual rules of projection. Let

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

represents the projected set corresponding to the last coefficient (say). All scalar parameters belonging to this set, are observationally equivalent to the $p$th component of the true parameter.

**Corollary 4** $B_p$ is an interval centered at $\beta^*_p$, the p-th component of $\beta^*_p$. 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 residual of the IV regression of $x_p$ onto the other components of $x$ using instruments $z$.

*Proof.* See Appendix A. ■

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 $E(\tilde{y}_\Delta \mid x,z)$ where $\tilde{y}_\Delta = \frac{\mu_k.y}{p_k(x,z)} + \frac{v_K - v_{K-1}}{2}$, with $\mu_k = \frac{(v_k - v_{k-1} - (v_{k+1} - v_k))}{2}$ for $k = 2,..,K-1$ and $\mu_1 = \mu_K = 0$ (as shown at the end of the proof of corollary 5). Hence, $E(|\tilde{x}_p|\,\Delta(x,z))$ 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 $\tilde{y}_\Delta$, [2] an estimation of the residual $\tilde{x}_p$ and [3] the linear regression of $\tilde{y}_\Delta$ on $|\tilde{x}_p|$.

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 2 may have no solutions at all (i.e., some overidentification restrictions may be not true).

Consider $z_A$, a random vector which dimension is the same as random vector $x$, defined by:

$$z_A = Az$$

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 identification 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:

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

As previously, this set is convex because it is the intersection of convex sets. What changes is that it can be empty which refutes the maintained assumption (L.3). This argument would form the basis for optimizing the choice of $A$ or for constructing test procedures of overidentifying restrictions in such a partial identification framework. The question is open whether the usual results hold. Finally, we can always project this set onto its elementary dimensions. The intersection of the projections is the projection of the intersections.

For the sake of simplicity, we shall proceed in the rest of the paper using the assumption that $p = q$ which is worthwhile investigating first.

## 3.4 Priors on The Range of Variation

Theorem 2 and its corollaries characterize the set of parameters (denoted $B$) that are observationally equivalent to the true parameter under the assumption that the conditional probability $\Pr(y = 1 \mid v; x; z)$ increases from 0 to 1 when $v$ varies over its support. This condition represents a potentially important limitation in empirical applications. A more careful look at Theorem 2 shows that it is possible to relax this assumption and to characterize the identification set in a substantially more general framework.

Because of $(NP.2)$, one key aspect of Theorem 2 is that there is no variation in the dependent variable $y$ at the top and bottom values of $v$ (i.e., $v_1$ and $v_K$). It is either always equal to 0 or always equal to 1. Knowing $\Pr(v = v_1 \mid x; z)$ or $\Pr(v = v_K \mid x; z)$ does not provide any additional information on the parameters of interest. In fact, the previous argument about identification is untouched and $B$ can be identified even in the extreme case where $\Pr(v = v_1 \mid x; z) = \Pr(v = v_K \mid x; z) = 0$; when $v_1$ and $v_K$ are outside the true support of $v$. In other words, it is not necessary to actually observe $\Pr(y = 1 \mid v; x; z)$ varying from zero to one to identify the set $B$, it is only necessary to impose this condition as a prior on the data generating process of $y$ at values of $v$ that are not observed in the available data. Some economic examples are given below.

To be more specific, consider the following reformulation of $(L.2)$;

$(L.2bis)$ There exist two finite real numbers $v_0$ and $v_{K+1}$; with $v_0 < v_1$ and $v_{K+1} > v_K$; such that the support of $-x\beta - \varepsilon$ is included in $[v_0; v_{K+1}[$.

Condition $(L.2bis)$ clearly relaxes condition $(L.2)$: Under $(L.2bis)$, $\Pr(y = 1 \mid v; x; z)$ does not necessarily vary from zero to one when $v$ varies over its support $\Omega_v = \{v_1, \ldots, v_K\}$, so that $\Pr(y = 1 \mid v; x; z)$ does not necessarily satisfy condition $(NP.2)$ anymore. Condition

$(L.2bis)$ imposes $(NP.2)$ as a prior on $Pr(y = 1 \mid v; x; z)$ for values of $v$, $v_0$ and $v_{K+1}$, that are actually not observed in the data.

It is straightforward to check that $B$ can be identified under $(L.2bis)$ following exactly the same route as under $(L.2)$. The only change is to replace $v_1$ by $v_0$ and $v_K$ by $v_{K+1}$.

**Corollary 5** Consider a conditional probability distribution, $Pr(y = 1 \mid v = v_k; x; z)$, denoted $G_k(x; z)$, which satisfies the monotonicity condition $(NP.1)$. The two following statements are equivalent,

(i) there exists a vector of parameters $\beta$ and there exists a latent random variable $\varepsilon$ such that the latent model $(\beta; F_\varepsilon(. \mid x; z))$ satisfies conditions $(L.1 - L.2bis - L.3)$ and such that $\{G_k(x; z)\}_{k=1,...,K}$ is its image through the transformation $(LV)$;

(ii) there exists a vector of parameters $\beta$ and there exists a measurable function $u(x; z)$ from $S_{x,z}$ to $R$ which takes its values in the interval (a.e. $F_{x,z}$)

$$I(x; z) = ]-\Delta(x; z); \Delta(x; z)],$$

where

$$\Delta(x; z) = \frac{1}{2}\sum_{k=1}^{K+1}[(v_k - v_{k-1})(G_k(x; z) - G_{k-1}(x; z))],$$

and such that,

$$E(z'(x\beta - \varepsilon)) = E(z'u(x; z)). \qquad (3)$$

This corollary states that identification remains possible even when the support of the special regressor is not large and when the probability of observing $y = 1$ does not vary from zero to one. The cost is that the identification set depends on priors (i.e, $v_0$ and $v_{K+1}$) which location might be debatable.

An example of potential application is the analysis of the probability of buying an object (a bottle of water, say) as a function of an experimentally-set price $v$. Specifically, each individual is faced with a price which is under experimental control and can take only two values $v_1$ and $v_2$. Though we only observe two prices, we can plausibly assume that for a sufficiently small (large) $v_0$ $(v_3)$ the probability of buying the object is 1 (0) whatever the characteristics of the individuals. Hence, the problem can be redefined with the support of $v$ being $\{v_0; v_1; v_2; v_3\}$ and with the additional assumption that $Pr(y = 1 \mid v; x; z)$ varies from zero to one when $v$ varies over its support.

Other (non-experimental) examples include the analysis of the probability of entry (or exit) into such basic institutions as the labor market or the school system. Consider for instance the school-leaving probability in a typical developed country, with $v$ representing individuals' age at the end of the year. We can plausibly speculate that (NP:2) is satisfied when (say) $v_0 = 15$ years and $v_{K+1} = 30$ years. Using these priors and assuming that the school-leaving latent propensity may be written $(x\beta + v + \varepsilon)$, we can provide valuable inference on $\beta$ even if our sample of observations consists in individuals aged from 20 to 25 years and such that the observed probability of school leaving of the 20 (25) years' old is strictly greater (lower) than 0 (1)[9].

# 4 Interval Data

In this section, we consider the case where $v$ is coninuous, but observed by intervals only. We show that the set of parameters observationally equivalent to the true structural parameter has a similar structure as in the discrete case. It is a convex set and, when there are no overidentifying restrictions ($p = q$), it is not empty. It contains the pseudo-true 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 the special regressor $v$ within-intervals, the identification set can be shrinked. 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 identification set is a singleton and the parameter of interest $\beta$ is exactly identified.

## 4.1 Identification Set: the General Case

Data is now characterized by a random variable $(y; v; v^*; x; z)$ where $v^*$ is the result of censoring $v$ by interval. 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:

[9] Under slightly different structural assumptions, this example can also be used in the section dealing with interval data when age is treated as a censored continuous variable.

(i) (Interval Data) The support of $v^*$ conditional on $(x, z)$ is $\{1, ..., K - 1\}$ 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 $[v_1, v_K[$. The distribution function of $v$ conditional on $x, z, v^*$ is denoted $F_v(. \mid v^*, x, z)$ and is assumed to be absolutely continuous. Its density function denoted $f_v(. \mid v^*, x, z)$ is strictly positive and bounded.

Within this framework, 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,

$$F_\epsilon(. \mid v, v^*, x, z) = F_\epsilon(. \mid x, z) \qquad (L.1^*)$$

The conditional probability distributions $Pr(y = 1 \mid v^*, x, z)$ generated through $(LV)$ by such latent models clearly satisfy condition $(NP.1)$. It can be shown using the same argument as in Lemma 1. In contrast, condition $(NP.2)$ is not anymore a consequence of $(L.2)$. When the special regressor is censored by intervals, the binary outcomes that can be analyzed through our latent models do not necessarily satisfy condition $(NP.2)$. We will drop this restriction from the definition of the class of binary reduced forms under consideration.

As previously, we consider a conditional probability function $Pr(y = 1 \mid v^*, x, z)$ which satisfies $(NP.1)$ and we search for a latent model generating this reduced form through transformation $(LV)$. In analogy with the discrete case, we begin by 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 :

$$\tilde{y} = \frac{\delta(v^*)}{p_{v^*}(x, z)} y - v_K \qquad (4)$$

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

With these notations, the following theorem analyses the degree of underidentification of the structural parameter $\beta$.

**Theorem 6** Consider $Pr(y = 1 \mid v^*; x; z)$ (denoted $G_{v^*}(x;z)$) a conditional distribution function satisfying the monotonicity condition (NP:1). The two following statements are equivalent,

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

a. $(\beta; F_\varepsilon(. \mid x; z))$ satisfies (L:1*; L:2; L:3)

b. $G_{v^*}(x;z)$ is the image of $(\beta; F_\varepsilon(. \mid x; z))$ through the transformation (LV);

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

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

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

and such that,

$$E(z'(x\beta - \tilde{y})) = E(z'u^*(x;z))). \tag{5}$$

*Proof.* See Appendix B ■

The identification set has the same general structure in the interval-data case as in the discrete case. It is a non-empty convex set which contains the pseudo-true value corresponding to the moment condition $E(z'(x\beta - \tilde{y})) = 0$.

## 4.2 Inference Using Additional Information on the Distribution Function of the Special Regressor

We now study how additional information helps to shrink the identification set. There are many instances where there exists additional information on the conditional distribution function of $v$ within intervals. It may correspond to the case where $v$ is observed at the initial stage of a survey or a census, but then dropped from the files that are made available to researchers for confidentiality reasons. Only interval-data information and information (estimates for instance) about the conditional distribution function of $v$ remains available. This framework may also correspond to the case where the conditional distribution function of $v$ is available in one database that does not contains information on $y$ while the information on $y$ is available in an another database[10] which contains only interval information on $v$.

To analyse these situations, we complete the statistical model by assuming that we have full information on the conditional distribution of $v$ :

$(NP:3)$ : The conditional distribution function of $v$ is known and denoted $\Phi(v \mid x; z; v^*)$.

The first question is whether this additional information reduces the identification set. The second question is whether there exists an optimal way of censoring $v$ and chosing the intervals for defining $v^*$: Knowing how identification is related to the conditional distribution $\Phi(v \mid x; z; v^*)$ may provide interesting guidelines to control censorship.

The first unsurprising result is that additional knowledge on $\Phi(v \mid x; z; v^*)$ actually helps to shrink the identification set. The second - more surprising - result is 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 to uniformity. The construction of these indexes is in three steps. To begin with, for any $v \in ]v_k; v_{k+1}[$, note that,

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

As $\Phi$ is absolutely continuous and its density is positive everywhere $(ID(ii))$, we can divide the previous expression by $\Phi(v \mid v^* = k; x; z)$ or $1 - \Phi(v \mid v^* = k; x; z)$, to obtain the two inequalities:

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

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

Given these inequalities, we are in position to define the two following indices:

[10] Angrist and Krueger (1992) or Arellano and Meghir (1992) among others developped two-sample IV techniques for such data design in the linear case.

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

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

where the strict inequalities stem from assumption ID(ii); i.e., the density function associated to $\Phi$ is positive and bounded. Using these notations, we have the following theorem:

**Theorem 7** Consider $\beta$ a vector of parameters, $Pr(y = 1 \mid v^*, x, z)$ (denoted $G_{v^*}(x,z)$) a conditional distribution function satisfying the monotonicity condition (NP.1) and $\Phi(v \mid v^*, x, z)$ a conditional distribution function. The two following statements are equivalent,

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

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

b. $G_{v^*}(x,z)$ is the image of $(\beta, F_\varepsilon(. \mid x, z))$ through the transformation (LV);

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

$$\underline{\Delta}^*_\Phi(x,z) = \sum_{k=1,\ldots,K-1} (v_{k+1} - v_k)\min(\xi_k^L(x,z), 0)(G_k(x,z) - G_{k-1}(x,z))$$

$$\overline{\Delta}^*_\Phi(x,z) = \sum_{k=1,\ldots,K-1} (v_{k+1} - v_k)\max(\xi_k^U(x,z), 0)(G_{k+1}(x,z) - G_k(x,z))$$

and such that,

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

*Proof.* See Appendix B ■

Given that $\min(\xi_k^L(x,z), 0) \in ]-1, 0]$ and $\max(\xi_k^U(x,z), 0) \in [0, 1[$, the identification set characterized by Theorem 7 is clearly smaller than the identification set characterized by Theorem 6 when no information is available on $v$. Also, Theorem 7 makes clear that the size of identification 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 this distance is abolished and $v$ is piece-wise uniform, the identification set clearly boils down to a singleton.

**Corollary 8** The identification 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 ■

One intuition of such a result is the following. When the identification set is a singleton, the moment condition that defines parameter $\beta$ is the same as the moment condition that would define $\beta$ if $v$ was replaced by a piece-wise uniform measurement $v_0$ [11] In general, the replacement of $v$ by such a piece-wise measurement produces an auxiliary model which does not satisfy the partial independence assumption[12]. What Corollary 8shows is that partial independence holds when $v$ itself is piece-wise uniform. This interpretation is developed in the appendix at the end of the proof of Corollary 8.

Corollary 8 corresponds to the "best" case. 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 $\beta$. The answer is positive. Specifically, the closer to uniformity the conditional distribution of $v$ is, the smaller the identification set is.

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

**Definition 9** $\Phi_2(v \mid x, z, v^*)$ is closer to uniformity than $\Phi_1(v \mid x, z, v^*)$, when a.e. $F_{x,z}$ and for any $k \in \{1, \ldots, K-1\}$:

$$\min(\xi^L_{k,1}(x,z), 0) \leq \min(\xi^L_{k,2}(x,z), 0)$$
$$\max(\xi^U_{k,1}(x,z), 0) \geq \max(\xi^U_{k,2}(x,z), 0).$$

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

[11] Such a measurement $v_0$ is drawn conditionally on $v^*$ in a uniform distribution in $[v_k, v_{k+1}[$.

[12] The auxiliary model is: $y = 1\{v_0 + x\beta + \varepsilon_0 > 0\}$,
where by construction: $\varepsilon_0 = \varepsilon + v - v_0$. The "special regressor" $v_0$ is now continuous and the transformation of Lewbel can be constructed. The new residual $\varepsilon_0$ does not necessarily satisfies partial independent however. The conditions under which the standard Lewbel procedure leads to consistent estimates need to be investigated.

Using this definition:

**Corollary 10** Let $\Phi(v \mid v^* = k; x; z)$ any conditional distribution. Let $B$ the associated region of identification for $\beta$. Then:

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

*Proof.* Straightforward using Theorem 7. ■

Assuming that we have some control on the construction on $v^*$ (i.e., on the data on $v$ that are made available to researchers), this result show that, in order to minimize the length of the interval, it has simply to be constructed in a way that minimizes the distance between the uniform distribution and the distribution of $v$ conditional on $v^*$ (and other regressors). Consider for instance the case of date of birth. 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 this variable does not vary significantly within months, meaning it is likely 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 Identification Set

The results about how to project the identification set in the discrete case can be easily extended to the case of interval data. Specifically, $B$ can be projected onto its elementary dimensions using the same usual rules of projection as in Corollary 4. As in the discrete case though, we focus on the leading case of no overidentifying restrictions ($p = q$).

Let:

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

represent the projected set corresponding to the last (say) coefficient. All scalar parameters belonging to this set, are observationally equivalent to the pth component of the true parameter. We denote $\beta^*$ the solution of equation (5) when function $u^*(x; z) = 0$:

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

To begin with, we consider the case where no information is available on the distribution of $v$ and state the corollary to Theorem 6.

**Corollary 11** $B_p$ is an interval which center is $\beta^*_p$, where $\beta^*_p$ represents the p-th component of $\beta^*$. Specifically, we have,

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

where :

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

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

with $\tilde{x}_p$ is the residual of the projection of $x_p$ onto the other components of $x$.

*Proof.* See Appendix B. ■

The corresponding corollary to Theorem 7 has exactly the same structure as Corollary 11, with $\underline{\Delta}^*_\Phi$ and $\overline{\Delta}^*_\Phi$ replacing $\underline{\Delta}^*$ and $\overline{\Delta}^*$.

Generally speaking, the estimation of $B_p$ requires the estimation of $\varsigma_{L,p}$ and $\varsigma_{U,p}$. At the end of the proof of corrolary 11, we show that these scalars can be estimated through simple regresssions. Specifically, let us denote $\tilde{y}_L = \dfrac{\theta_{L,v^*}y}{p_k(x,z)} + v_K - v_{K-1}$, where $\theta_{L,k} = \dfrac{(v_{k+2} - v_{k+1} - (v_{k+1} - v_k))}{2}$ for $k = 2, \ldots, K-1$ and where $v_{K+1} = v_K$ by convention. Similarly, define $\tilde{y}_U = \dfrac{\theta_{U,v^*}y}{p_k(x,z)} + v_K - v_{K-1}$, where $\theta_{U,k} = \dfrac{(v_k - v_{k-1} - (v_{k+1} - v_k))}{2}$ for $k = 2, \ldots, K-1$ and where $v_0 = v_1$.

Using these notations, $\varsigma_{L,p}$ is the regression coefficient of $(1\{\tilde{x}_p > 0\}\tilde{y}_{L,p} + 1\{\tilde{x}_p \leq 0\}\tilde{y}_{U,p})$ on $\tilde{x}_p$ and $\varsigma_{U,p}$ is the regression coefficient of $(1\{\tilde{x}_p \leq 0\}\tilde{y}_{L,p} + 1\{\tilde{x}_p > 0\}\tilde{y}_{U,p}))$ on $\tilde{x}_p$. Most interestingly, when all intervals have the same length, $\tilde{y}_L$ and $\tilde{y}_U$ are equal and constant and the length of the one-dimensional identification region is then proportional to this constant.

# 5 Monte Carlo Experiments

In this section, we present simple Monte Carlo experiments in order to analyze how our (set) estimators perform in medium-sized samples (i.e., 100 to 1000 observations). The simulated model is $y = 1\{1 + v + x_2 + \varepsilon > 0\}$. 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 cases where the special regressor $v$ is discrete or interval-valued.

Specifically, the construction of the special regressor $v$, the covariate $x_2$, the instrument $z$ and the random shock $\varepsilon$ proceeds in two steps. To begin with, consider four random variables such as: $e_1$ 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:

$$\eta = 2e_2 + \alpha e_4; \qquad x_2 = e_1 + e_4$$

$$\varepsilon = \rho(e_1 - .5) + e_3; \qquad z = e_4.$$

where $\alpha$ is a parameter that makes the random shock a non-normal variate and $\rho$ 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.

In the discrete case, we choose $v_1$ and $v_K = -v_1$ at the $2.5$ and $97.5$ percentiles of the distribution of $\eta$. The other points of the support of $v$ are denoted $v_2, .., v_{K-1}$. With these notations, $v$ is defined as (where $v_{K+1} = \infty$):

$$v = v_k \text{ if } \eta \in [v_k; v_{k+1}[ \text{ and } k = 2, .., K$$

$$v = v_1 \text{ if } \eta < v_2$$

To comply with assumption L.2, we then truncate $x_2 + \varepsilon$ by a method of acceptation and rejection in order that $1 + x_2 + \varepsilon + v_1 > 0$ and $1 + x_2 + \varepsilon + v_K < 0$.

In the interval case, $v$ is defined by truncating $\eta$ to the 95% symmetric interval around 0, denoted $[v_1; v_K]$. We do that by a method of acceptation and rejection. To comply with assumption L.2, we then truncate $x_2 + \varepsilon$ by the same method of acceptation and rejection than before so that $1 + x_2 + \varepsilon + v_1 > 0$ and $1 + x_2 + \varepsilon + v_K < 0$. We then construct the censored $K - 1$ intervals in the obvious way:

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

## 5.1 Presentation of results

Tables 1 to 8 report various Monte Carlo experiments in cases where the data are discrete or are interval-valued. In all tables, we make the sample size vary using 100, 200, 500 or 1000 observations. The number of Monte Carlo replications is equal to 1000 in all experiments. Additional replications do not affect any estimates (resp. standard errors) by more than a 1% margin of error (resp. 3%). We report results in two panels. In the top panel, we report estimates of the lower and upper bounds of both coefficients (intercept and variable) by recentering them at zero instead of their true values which are equal to one. In the bottom panel, we compute the average of the estimates of the lower and upper bounds, $E(\hat{\mu}_b + \hat{\mu}_u)/2$; the adjusted length of the interval, $E(\hat{\mu}_u - \hat{\mu}_b)/2\sqrt{3}$, and the average sampling error defined as:

$$(\sigma_u^2 + \sigma_b^2 + \sigma_u \sigma_u)/3$$

where $\sigma_u$ and $\sigma_b$ are estimated standard errors of the estimated lower and upper bounds. These three statistics provide an interesting decomposition of the mean square error uniformly integrated over the interval $[\hat{\mu}_b; \hat{\mu}_u]$:

$$\begin{aligned}
MSEI &= E\left[\int_{\hat{\mu}_b}^{\hat{\mu}_u} (\mu - \mu_0)^2 \frac{d\mu}{\hat{\mu}_u - \hat{\mu}_b}\right] \\
&= \frac{1}{3} E\left[\frac{(\hat{\mu}_u - \mu_0)^3 - (\hat{\mu}_b - \mu_0)^3}{\hat{\mu}_u - \hat{\mu}_b}\right] \\
&= \frac{1}{3} E\left[(\hat{\mu}_u - \mu_0)^2 + (\hat{\mu}_b - \mu_0)^2 + (\hat{\mu}_b - \mu_0)(\hat{\mu}_u - \mu_0)\right]
\end{aligned}$$

Let $\bar{\mu}_i = E(\hat{\mu}_i)$, $i = u; b$, the expected values of the estimates, and $\bar{\mu}_m = (\bar{\mu}_u + \bar{\mu}_b)/2$ the average center of the interval. We then have:

$$\begin{aligned}
MSEI &= (\bar{\mu}_m - \mu_0)^2 + \frac{1}{3} E\left[(\hat{\mu}_u - \bar{\mu}_m)^2 + (\hat{\mu}_b - \bar{\mu}_m)^2 + (\hat{\mu}_b - \bar{\mu}_m)(\hat{\mu}_u - \bar{\mu}_m)\right] \\
&= (\bar{\mu}_m - \mu_0)^2 + \frac{1}{3}\left((\bar{\mu}_u - \bar{\mu}_b)/2\right)^2 \\
&\quad \frac{1}{3} E\left[(\hat{\mu}_u - \bar{\mu}_u)^2 + (\hat{\mu}_b - \bar{\mu}_b)^2 + (\hat{\mu}_b - \bar{\mu}_b)(\hat{\mu}_u - \bar{\mu}_u)\right]
\end{aligned}$$

The first term is the square of a "decentering" term which can be interpreted as a bias term. The second term is the square of the "adjusted" length, which can be interpreted as the "uncertainty" due to partial identification instead of point identification. The third term is an average of standard errors and can then be interpreted as sample variability. These three terms are reported in the bottom panel for both coefficients as well as root mean square error, $MSEI^{1/2}$.

## 5.2 Discrete Data

In experiments reported in Tables 1 to 4, the data are discrete. We make some parameters vary in these tables: The bandwidth in Table 1, the degree of non normality in Table 2, the degree of endogeneity in Table 3 and the number of points in the support of the special regressor in Table 4. In all cases, the true value of the parameter belongs to the interval built up around the estimates of the lower and upper bounds. Horowitz and Manski (2000) and Imbens and Manski (2003) for an alternative, rigorously de…ne con…dence intervals when identi…cation is partial. We here report con…dence intervals for bounds only. In cases where the number of points is …xed (Tables 1 to 3), the stability of the estimated length of the interval across experiments is a noticeable result. It almost never vary by more than a relative factor of 10%.

In Table 1, we experimented with di¤erent bandwidths. As said, interval lengths are stable, though intervals can be severely decentered for the intercept term. Increasing the sample size or the bandwidth recenters the interval around the true value. Increasing the bandwidth decenters interval estimates for the coe¢cient of the variable towards the negative numbers though at a much lesser degree. Finally, the mean square error (MSEI) for the intercept decreases with the bandwidth while it has a U-shape form for the coe¢cient of the variable. We have tried to look for a data-driven choice of the bandwidth by minimizing this quantity but it was unconclusive. A larger bandwidth seems to be always preferred. Some further research is clearly needed on this issue.

In Table 2, we experimented with di¤erent degrees of non-normality, by making parameter ® vary. If this parameter increases, interval length is very weakly a¤ected. There is some recentering of intervals either towards negative numbers for the intercept or towards positive values for the coe¢cient of the variable. Note that average standard errors and mean square errors also tend to increase with parameter ®.

In Table 3, we experimented with di¤erent degrees of correlation between covariates and errors and therefore the amount of endogeneity. It is the only case where interval length slightly di¤ers across experiments. It increases with the amount of endogeneity. There is also some large decentering of the intervals for small sample sizes (100) but decentering either completely disappears when the sample size is equal to 1000 or is not much a¤ected by varying the degree of endogeneity. As well, standard errors are slightly affected only when the sample size is less than 200.

In Table 4, we experimented with varying the number of points of the discrete support. Theory predicts that interval length should decrease with the number of points of support. In our experiments, it is always true and this decrease is not much affected by sample sizes. We obtain that result by estimating the conditional probability function of $v$ using nearest neighbors (w.r.t. $v$) and using kernels for the other covariates. A preliminary less careful estimation of this probability function led to humps and bumps in the estimates. There can be some strong decentering problems though and there is evidence of a trade-off between the length of the interval and the average standard errors. The latter tend to increase when the number of points in the support increases. No doubt that it is partly due to the way we built up the probability estimates.

## 5.3 Interval Data

In experiments reported in Tables 5 to 8, the data are interval-valued. Similarly to the discrete case, we make the same parameters vary in these tables: The bandwidth in Table 5, the degree of non normality in Table 6, the degree of endogeneity in Table 7 and the number of points in the support of the special regressor in Table 8.

Although the experiments cannot be strictly compared, results are in most cases very similar to the discrete case. The true values of the parameters belong to the confidence interval built up around the estimates of the lower and upper bounds. In cases where the number of points is fixed (Tables 5 to 7), the stability of the length of the interval is again a noticeable result. It almost never vary by more than a relative factor of 10%. The average length seems however to be larger in the interval case than in the discrete case.

In Table 1, results remain very close to those obtained in the discrete case. The interval for the intercept is severely decentered in small samples while the interval for the variable coefficient is decentered in large samples with a slightly larger magnitude than in the discrete case. Similarly, the mean square error is decreasing with the bandwidth or, less frequently has a U-shape form. Again, finding a data-driven bandwidth through minimization of this mean square error is not an easy task. Table 6 has a different flavour. Decentering can be quite severe above all for the coefficient of the variable when the degree of non-normality is large. It is also true at a lesser degree for the intercept. In Table 7 also, results are less systematic than in the discrete case. Interval length either decrease or increase when the degree of endogeneity increases while decentering can be quite severe, much more than in the discrete case. Nevertheless, results are very similar to the discrete case when the number of intervals is varied (Table 8). Interval lengths regularly shrink towards 0 while mean square error increases, yielding evidence on the trade-off between those characteristics.

# 6 Conclusion

In this paper, we explored partial identification of coefficients of binary variable models in the case where the special 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 identification 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, even if we do not provide proofs of consistency and asymptotic properties of the estimates of the bounds of the intervals, those would follow very similar lines to the ones Lewbel (2000) presents. The asymptotic variance-covariance matrix of the bounds can also be derived along similar lines. Finally, one can show that under some conditions (see the companion paper, Magnac and Maurin, 2003), these estimates are efficient in a semi-parametric sense.

REFERENCES


Angrist, J.D., 2001, "Estimation of Limited Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice", Journal of Business and Economic Statistics, 19:2-16 with discussions.

Angrist, J.D., and A. Krueger, 1992, "The E¤ect of Age at School Entry on Educational Attainment : An Application of Instrumental Variables with Moments from Two Samples", Journal of the American Statistical Association.

Arellano, M. and C., Meghir, 1992, "Female Labor Supply and On-the-Job Search: An Empirical Model Estimated Using Complementary Datasets", Review of Economic Studies, 59:537-560.

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..

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 Experimenst with Missing Covariate and Outcome Data", Journal of the American Statistical Association, 95:77-84.

Imbens, G., and C.F., Manski, 2004, "Con…dence Intervals for Partially Identi…ed Parameters", forthcoming Econometrica.

Lewbel, A., 2000, "Semiparametric Qualitative Response Model Estimation with Unknown Heteroskedasticity or Instrumental Variables", Journal of Econometrics, 97:145-77.

Magnac, T., and E., Maurin, 2003, "Identi…cation & Information in Monotone Binary Models", CREST Working Paper, 2003-07.

Manski, C.F., 1985, "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator", Journal of Econometrics, 27:313-33.

Manski, C.F., 1988, "Identi…cation of Binary Response Models", Journal of the American Statistical Association, 83:729-738.

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, 85:301-32.

# A Proofs in Section 3

## A.1 Proof of Lemma 1

Write:

$$\Pr(y_i = 1 \mid v; x; z) = \int_{x\beta + v + \epsilon > 0;\, \varepsilon \in \Omega_\varepsilon(x;z)} dF(\epsilon \mid x; z)$$

As $dF(\epsilon \mid x; z) \geq 0$, monotonicity in $v$ follows.

Secondly, by assumption $L{:}2$, the support of $-x\beta - \varepsilon$ is a subset of $\Omega_v^0 = [v_1; v_K[$:

$$v_1 \leq -(x\beta + \varepsilon) < v_K$$

and therefore for all $\varepsilon \in \Omega_\varepsilon(x; z)$:

$$v_1 + x\beta + \varepsilon \leq 0 \quad v_K + x\beta + \varepsilon > 0$$

The second conclusion follows.

## A.2 Proof of Theorem 2

Let $\{G_k(x; z)\}_{k=1,...,K}$ satisfy (NP:1) and (NP:2). It is an ordered set of functions such that $G_1 = 0$ and $G_K = 1$. Fix $\beta$. We first prove that (i) implies (ii).

(Necessity) Assume that there exists a latent random variable $\varepsilon$ such that $(\beta; F_\varepsilon(. \mid x; z))$ satisfies $(L{:}1 - L{:}3)$ and such that $\{G_k(x; z)\}_{k=1,...,K}$ is its image through transformation $(LV)$: By $(L{:}2)$, the conditional support of $\varepsilon$ given $(x; z)$, is included in $]-(v_K + x\beta); -(v_1 + x\beta)]$ and we can write,

$$\forall k;\, G_k(x; z) = \int_{-(v_k + x\beta)}^{-(v_1 + x\beta)} f_\varepsilon(\varepsilon \mid x; z) d\varepsilon = 1 - F_\varepsilon(-(v_k + x\beta) \mid x; z): \qquad (A.1)$$

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

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 $k = 2; ..; K - 1$. Setting $\delta_1 = \delta_K = 0$; the transformed variable $\tilde{y}$ is $(\frac{\delta_k y}{p_k(x; z)} - s_K)$ where $y = 1\{v > -(x\beta + \epsilon)\}$. Integrate $\tilde{y}$ with respect to $v$ and $\varepsilon$:

$$E(\tilde{y} \mid x; z) = \int_{\Omega(\epsilon \mid x; z)} \left[\sum_{k=1}^{K} \delta_k 1\{v_k > -(x\beta + \epsilon)\}\right] f(\epsilon \mid x; z) d\epsilon - s_K$$

$$= \int_{\Omega(\epsilon \mid x; z)} \left[\sum_{k=2}^{K-1} (s_{k+1} - s_k) 1\{v_k > -(x\beta + \epsilon)\}\right] f(\epsilon \mid x; z) d\epsilon - s_K$$

As the support of $(x\beta + \varepsilon)$ is bounded, we can also define for any value of $w = -x\beta - \varepsilon$; an integer function $j(w)$ in $\{1; ..; K - 1\}$, such that $v_{j(w)} \leq w < v_{j(w)+1}$: By construction,

$v_k > w$ , $k > j(w)$ and $\sum_{k=2}^{K-1}(s_{k+1} - s_k)1\{v_k > w\} = (s_K - s_{j(-(x\beta+\varepsilon))+1})$: Hence, we have :

$$E(\varphi \mid x;z) = \int_{\beta(\varepsilon|x;z)} (s_K - s_{j(-(x\beta+\varepsilon))+1}) f(\varepsilon \mid x;z)d\varepsilon - s_K = -E[s_{j(-x\beta-\varepsilon)+1} \mid x;z]$$
$$= x\beta + E(\varepsilon \mid x;z) - E[s_{j(-x\beta-\varepsilon)+1} + x\beta + \varepsilon \mid x;z]$$
$$= x\beta + E(\varepsilon \mid x;z) - u(x;z) \quad \text{(A.2)}$$

where:

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

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

$$-\frac{v_{j(w)+1} - v_{j(w)}}{2} < \frac{v_{j(w)+1} + v_{j(w)}}{2} - w \le \frac{v_{j(w)+1} - v_{j(w)}}{2}$$

which yields:

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

Hence, we can write,

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

using equation (A.1). By analogy,

$$-\Delta(x;z) < u(x;z) \le \Delta(x;z):$$

Since $G_K(x;z) = 1$ and $G_1(x;z) = 0$; we have $\Delta(x;z) \ge \min_k(\frac{v_k - v_{k-1}}{2})$; meaning $\Delta(x;z) > 0$ and $I(x;z)$ 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 $F_\varepsilon(.\mid x;z)$ satisfying $(L.1 - L.3)$ such that the image of $(\beta; F_\varepsilon(.\mid x;z))$ through $(LV)$ is $\{G_k(x;z)\}_{k=1,..,K}$.

First, let $\eta$ a random variable which support is $]0;1]$; which conditional density given $(v;x;z)$ is independent of $v$ (a.e. $F_{x,z}$) and which is such that:

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

Second, let $\kappa$ a discrete random variable which support is $\{2; ::; K\}$ and which conditional distribution given $(v; x; z)$ is independent of $v$ and is given by:

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

For any $k \in \{2; ::; K\}$, consider $K - 1$ random variables, say $\epsilon(\lambda; k)$ which are constructed from $\lambda$ by:

$$\epsilon(\lambda; k) = -x\beta - \lambda v_{k-1} - (1 - \lambda)v_k$$

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

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

which support is $]-x\beta - v_K; -x\beta - v_1]$; which is absolutely continuous (because $\lambda$ is), and which is independent of $v$ (because both $\lambda$ and $\kappa$ are). It therefore satisfies (L:1) and (L:2). Furthermore, because of (A.4), the image of $(\beta; F_\varepsilon(: \mid x; z))$ through (LV) is $\{G_k(x; z)\}_{k=1;::;K}$ because they satisfy equation (A.1). The last condition to prove is (L:3). Consider, for almost any $(x; z)$,

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

$$\sum_{k=2} \left( \int_{-x\beta - v_k}^{-x\beta - v_{k-1}} \left(\frac{v_k + v_{k-1}}{2} + x\beta + \varepsilon\right) f(\varepsilon \mid x; z; \kappa = k)d\varepsilon \right)$$

$$= \sum_{k=2} E\left(\frac{v_k + v_{k-1}}{2} - \lambda v_{k-1} - (1 - \lambda)v_k \mid x; z\right)(G(v_k; x; z) - G(v_{k-1}; x; z))$$

$$= \sum_{k=2} E(\lambda - 1/2 \mid x; z).(v_k - v_{k-1}).(G_k(x; z) - G_{k-1}(x; z))$$

$$= (u(x; z)/(2\Delta(x; z)))(2\Delta(x; z)) = u(x; z):$$

Therefore:

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

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

Remark: It is worth emphasizing that this proof also provides a characterization of the domain of observationally equivalent distribution functions $F_\varepsilon$, i.e. the set of random variables $\varepsilon$ such that there exists $\beta$ with $(\beta; F_\varepsilon)$ satisfying conditions (L:1 − L:3) and generating $\{G_k(x; z)\}_{k=1;::;K}$.

To begin with, any such $\varepsilon$ can clearly be decomposed into a mixture of two independent variables as in (A.5):

$$\kappa = \sum_{k=2;:::;K} k.1\{\varepsilon \in ]-x\beta - v_k; -x\beta - v_{k-1}]\}$$

$$\lambda = (v_\kappa + \frac{(\varepsilon + x\beta)}{(v_\kappa - v_{\kappa-1})}) \in ]0; 1]$$

By construction, $\kappa$ necessarily satisfies equation (A.4) and $\beta$ correspond to the IV regression coefficient of $\tilde{y} - \Delta(x;z)(2\lambda - 1)$ on $x$.

Reciprocally, any mixture $\varepsilon(\lambda;\kappa) = -x\beta - \lambda v_{\kappa-1} - (1-\lambda)v_{\kappa}$ where $\lambda \in ]0;1]$ and $\kappa \in \{2;\dots;K-1\}$ are two independent variables satisfying equation (A.4). Parameter $\beta$ defined as the IV regression coefficient of $\tilde{y} - \Delta(x;z)(2\lambda - 1)$ on $x$ is such that $(\beta; F_\varepsilon)$ satisfies conditions $(L.1 - L.3)$ and generates $\{G_k(x;z)\}_{k=1;\dots;K}$.

Concluding for any $\varepsilon$; the two following statements are equivalent,

(i) there exists a vector of parameter $\beta$ such that the latent model $(\beta; F_\varepsilon(.\mid x;z))$ verifies conditions $(L.1 - L.3)$ 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 $(\lambda;\kappa)$, conditional on $(x;z)$, such that the support of $\lambda$ is $]0;1]$, the support of $\kappa$ is $\{2;\dots;K\}$; equation (A.4) holds and such that:

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

where:

$$\begin{aligned}\beta_\lambda &= \left[E(x'z)(E(z'z))^{-1}E(z'x)\right]^{-1} \\ &\quad .E(x'z)(E(z'z))^{-1}E(z'(\tilde{y} - \Delta(x;z)(2\lambda - 1)))\end{aligned}$$

## A.3 Proof of Corollary 3

First, $B$ contains $\beta^*$ because $u(x;z) = 0$ takes its values in the admissible set, $I(x;z)$. Second, $B$ is convex because $I(x;z)$ is convex and equation (2) is linear. Furthermore, assume that $(\beta; F_\varepsilon(.\mid x;z))$ satisfies conditions $(L.1 - L.3)$ and generates $G(v;x;z)$ through the transformation $(LV)$. Using Theorem 2, there exists $u(x;z) \in I(x;z)$ such that,

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

and thus using the definition of $W$:

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

Using the generalized Cauchy-Schwarz inequality, we have,

$$E(u'(x;z)z)E(z'z)^{-1}E(z'u(x;z)) \leq E(u^2(x;z)).$$

and by Theorem 2, $E(u^2(x;z)) \leq E(\Delta^2(x;z))$. By definition, $E(\Delta^2(x;z)) \leq \Delta_M^2$ which completes the proof.

## A.4 Proof of Corollary 4

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 $\tilde{x}_p$:

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

Applying the principle of Frish-Waugh, we have

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

As $\tilde{x}_p$ is a scalar, the maximum (minimum) of $E(\tilde{x}_p u(x))$ when $u(x;z)$ varies in $]-\Delta(x);\Delta(x)]$ is obtained by setting $u(x) = \Delta(x)1\{\tilde{x}_p > 0\} - \Delta(x)1\{\tilde{x}_p \le 0\}$ ($u(x) = -\Delta(x)1\{\tilde{x}_p > 0\} + \Delta(x)1\{\tilde{x}_p \le 0\}$). 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 $\beta_p$ 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

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

This function takes its values in $]-\Delta(x);\Delta(x)[$ and therefore satisfies point (ii) of Theorem 2. Thus, there exists $\beta \in B$ such that its last component is $\beta_p$.

The adaptation to the general IV case uses the generalized transformation:

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

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 $E(\tilde{y}_\Delta \mid x;z)$ where

$$\tilde{y}_\Delta = \frac{\mu_k . y}{p_k(x;z)} + \frac{v_K - v_{K-1}}{2};$$

with

$$\mu_k = \frac{(v_k - v_{k-1} - (v_{k+1} - v_k))}{2} \text{ for } k = 2,..,K-1$$

and $\mu_1 = \mu_K = 0$. Specifically,

$$
\begin{aligned}
\Delta(x,z) &= \frac{1}{2}\sum_{k=2}^{K}[(v_k - v_{k-1})(G_k(x,z) - G_{k-1}(x,z))] \\
&= \frac{1}{2}[(v_2 - v_1)G_2(x,z) + (v_3 - v_2)(G_3(x,z) - G_2(x,z)) + \ldots \\
&\quad \ldots + (v_{K-1} - v_{K-2})(G_{K-1}(x,z) - G_{K-2}(x,z)) + (v_K - v_{K-1})(1 - G_{K-1}(x,z))] \\
&= \frac{1}{2}\sum_{k=2}^{K-1}(v_k - v_{k-1} - (v_{k+1} - v_k))G_k(x,z) + \frac{v_K - v_{K-1}}{2} \\
&= \frac{1}{2}\sum_{k=2}^{K-1}(v_k - v_{k-1} - (v_{k+1} - v_k))E(y \mid v = v_k, x, z) + \frac{v_K - v_{K-1}}{2} \\
&= E\left(\frac{\mu_k . y}{p_k(x,z)} \mid x,z\right) + \frac{v_K - v_{K-1}}{2} = E(\tilde{y}_\Delta \mid x,z)
\end{aligned}
$$

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

# B Proofs in Section 4

## B.1 Proof of Theorem 6

Consider a vector of parameters $\beta$ and a conditional probability distribution $\Pr(y = 1 \mid v^*, x, z)$ (denoted $G_{v^*}(x,z)$) satisfying monotonicity conditions (NP.1).

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

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

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

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

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

which yields the convolution equation:

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

Note that this condition implies:

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

The second inequality is strict because $F_v$ is absolutely continuous and $F_w$ is continuous on the right (CADLAG).

To prove (5), write $E(\tilde{y} \mid x, z)$ as

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

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

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

Evaluate first the inner integral with respect to $v$. As the support of $w$ is included in $[v_1, v_K[$, we can define for any value of $w$ in its support, an integer function $j(w)$ in $\{1, \ldots, K-1\}$, 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,

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

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

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

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

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

Summing over values of $v^*$,

$$\begin{aligned} &\sum_{v^* \in \{v_1,\ldots,v_{K-1}\}} \delta(v^*) \int_{\Omega(v|v^*,x,z)} 1(v > w) f_v(v \mid v^*, x, z) dv \\ &= -F_v(w \mid v_{j(w)}, x, z)(v_{j(w)+1} - v_{j(w)}) + v_K - v_{j(w)}. \end{aligned}$$

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

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

where

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

Integrating (B.4) with respect to $x, z$ and using condition (L.3) yields condition (5).

To finish the proof, upper and lower bounds for $u^*(x,z)$ are now provided. Let write,

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

where:

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

By integration by parts, the first term is:

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

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

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

Using (B.2) and the fact that $F_w(w \mid x,z)$ is continuous on the right, it implies

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

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

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

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

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

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

$$\theta(x;z) = \max\left(\frac{u^*(x;z)}{\overline{\Delta}^*(x;z)}; 1 - \frac{u^*(x;z)}{\underline{\Delta}^*(x;z)}\right).$$

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 $\kappa$ a discrete random variable which support is $\{1;..;K-1\}$; which is independent of $v^*$ (a.e. $F_{x;z}$) and which conditional on $(x;z)$ distribution is:

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

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

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

For instance, $\lambda$ 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:

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

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

To construct $v$, we first introduce a random variable $\eta$ which support is $[0;1[$, which is absolutely continuous, which is defined conditionally on $(k;x;z)$; which is independent of $\lambda$ and such that:

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

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

For instance, when $\lambda$ is chosen discrete with a mass point on $\lambda_0(x;z)$, we simply have to chose $\eta$ such that

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

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 $(\beta; F_w(. \mid x; z))$ through $(LV)$ is $G_{v^*}(x; z)$ because it satisfies equation (B.1):

$$\int_{v_k}^{v_{k+1}} F_w(v \mid x; z).f_v(v \mid k; x; z)dv = H_k(x; z)+$$

$$+(H_{k+1}(x; z) - H_k(x; z)) \int_{v_k}^{v_{k+1}} \Pr(w = (1 - \lambda)v_k + \lambda v_{k+1} \leq v \mid x; z).f_v(v \mid k; x; z)dv =$$

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

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

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

Therefore,

$$\sum_{k=1}^{K-1} (v_{k+1} - v_k)\varphi_k(x; z) = \sum_{k=1}^{K-1} (v_{k+1} - v_k)(H_{k+1}(x; z) - G_k(x; z))$$

$$- \sum_{k=1}^{K-1} (v_{k+1} - v_k)(H_{k+1}(x; z) - H_k(x; z))E(\lambda \mid x; z) = u^*(x; z).$$

using equation (B.9). Plugging (5) in (B.4) yields $E(z'\varepsilon) = 0$ that is $(L.3)$.

## B.2 Proof of Theorem 7

We use large parts of the proof of Theorem 6:

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

$$\varphi_k(x; z) = \int_{v_k}^{v_{k+1}} \left(\Psi(v \mid k; x; z) - \frac{v - v_k}{v_{k+1} - v_k}\right)dF_w(v \mid x; z).$$

We then have first:

$$\begin{aligned}
\varphi_k(x; z) &= \int_{v_k}^{v_{k+1}} \left(1 - \frac{\frac{v - v_k}{v_{k+1} - v_k}}{\Psi(v \mid k; x; z)}\right)\Psi(v \mid k; x; z)dF_w(v \mid x; z) \\
&\leq \sup_{v \in ]v_k; v_{k+1}[} \left(1 - \frac{\frac{v - v_k}{v_{k+1} - v_k}}{\Psi(v \mid k; x; z)}\right) \int_{v_k}^{v_{k+1}} \Psi(v \mid k; x; z)dF_w(v \mid x; z) \\
&= \xi_k^U(x; z) \int_{v_k}^{v_{k+1}} \Psi(v \mid k; x; z)dF_w(v \mid x; z)
\end{aligned}$$

But, using equation (B.1), we have

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

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

$$\varphi_k(x;z) \leq \max(\xi_k^U(x;z);0).(G_{k+1}(x;z) - G_k(x;z)).$$

The derivation of the lower bound follows the same logic:

$$
\begin{aligned}
\varphi_k(x;z) &\geq \inf_{v \in ]v_k;v_{k+1}[} \left(-1 - \frac{\frac{v - v_{k+1}}{v_{k+1} - v_k}}{1 - \Phi(v \mid k;x;z)}\right) \int_{v_k}^{v_{k+1}} (1 - \Phi(v \mid k;x;z))\,dF_w(v \mid x;z) \\
&\geq \min(\xi_k^L(x;z);0))\left[\int_{v_k}^{v_{k+1}} d[(1 - \Phi(v \mid k;x;z))F_w(v \mid x;z)] + G_k(x;z)\right]
\end{aligned}
$$

Hence, using $F_w(v_k \mid x;z) \geq G_k(x;z)$, we have

$$\varphi_k(x;z) \geq \min(\xi_k^L(x;z);0)(G_k(x;z) - G_{k-1}(x;z)).$$

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). Denote $u^*(x;z)$ in $[\underline{\Delta}_\Phi^*(x;z);\overline{\Delta}_\Phi^*(x;z)]$ such that

$$E(z'(x\beta - y)) = E(z'u^*(x;z))$$

We shall prove that there exists a distribution function of the random term $\varepsilon$ which agree with parameter $\beta$ defined by such a moment condition when the distribution function of the special regressor $v$ is $\Phi(v \mid k;x;z)$. As in the proof of Theorem 6, 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 $\{1;..;K-1\}$ such as:

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

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 $\xi^L_k(x;z) < 0$ and $\xi^U_k(x;z) > 0$, the closer $u^*(x;z)$ is from the lower bound $\underline{\Delta}^*_\theta(x;z)$, the closer is $H_k$ to $G_{k-1}$, and the closer $u^*(x;z)$ is from the upper bound $\overline{\Delta}^*_\theta(x;z)$, the closer is $H_k$ to $G_k$.

Decompose now $u^*(x;z)$ into $\psi^*_k(x;z)$ such that:

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

and such that the bounbds on $u^*$ can be translated into:

$$\xi^L_k(x;z)(G_k(x;z) - H_k(x;z)) \leq \psi^*_k(x;z) \leq \xi^U_k(x;z)(H_{k+1}(x;z) - G_k(x;z)) \qquad \text{(B.12)}$$

There are many decompositions of this type. Choose one.

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

$$\Pr(\kappa = k \mid x;z) = H_{k+1}(x;z) - H_k(x;z): \qquad \text{(B.13)}$$

Consider also $K-1$ random variable $\lambda_k$ which support is $]0;1[$; which are independent of $v^*$ (a.e. $F_{x;z}$) and which conditional (on $(x;z)$) expectation is:

$$E(\lambda_k \mid x;z) = \frac{H_{k+1}(x;z) - G_k(x;z) - \psi^*_k(x;z)}{H_{k+1}(x;z) - H_k(x;z)} \qquad \text{(B.14)}$$

and such that:

$$\int_0^1 (\theta_v(\lambda v_k + (1-\lambda)v_{k+1} \mid k;x;z) - \frac{v - v_k}{v_{k+1} - v_k})dF_{\lambda_k}(\lambda \mid x;z) = \frac{\psi^*_k(x;z)}{H_{k+1}(x;z) - H_k(x;z)}$$

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

Finish the proof as in Theorem 6.

## B.3 Proof of Corollary 8

(Necessity) Let the conditional distribution of $v$, $\theta_0$, be piece-wise uniform by intervals, $v^* = k$. Then, for any $k = 1;..;K-1$, $\xi^U_k(x;z) = \xi^L_k(x;z) = 0$. Using Theorem 7 yields that $\underline{\Delta}^*_\theta(x;z) = \overline{\Delta}^*_\theta(x;z) = 0$ and therefore $u^*(x;z) = 0$. Identification of $\beta$ is exact and its value is given by the moment condition (5).

(Sufficiency) By contraposition; Assume that there exists $k \in \{1;..;K-1\}$; 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 $\Psi(v \mid k; x; z)$ is different from a uniform distribution function on $A$ for any $(x; z)$ in $S$. Because $\Psi$ is absolutely continuous (ID(ii)), and for the sake of simplicity assume that:

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

Because $\lambda_k^U(x; z) > 0$; we can always construct a function $u_1^*(x; z)$ which is stricly positive on $S$ satisfying the conditions of Theorem 7. Thus $E(z' u_1^*(x; z)) \neq 0$ and the moment condition (5) can be used to construct parameter $\beta_1$: It implies that the identification set $B$ contains at least two different parameters $\beta$; 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).

Interpretation:

Consider a observable variable $v_0$ drawn conditionally on $v^*$ in a uniform distribution in $[v_k; v_{k+1}[$. Write an auxiliary model as:

$$y = 1\{v_0 + x\beta + \varepsilon_0 > 0\}$$

where by construction:

$$\varepsilon_0 = \varepsilon + v - v_0:$$

Note first that $v$ and $v_0$ are independent conditional on $(v^*; x; z)$. Second, that the auxiliary model now is a binary model with a continuous special regressor. Third, that the discrete-type transformation $\tilde{y}$ of the data is equal up to a constant term to the continuous-type transformation of the data i.e.:

$$\tilde{y} = \frac{v_{k+1} - v_k}{p_{v^*}(x; z)} y - v_K = \frac{y}{f_{v_0}(v_0; v^*; x; z)} - v_K = \bar{y} + cst$$

since by construction:

$$f_{v_0}(v_0; v^*; x; z) = \frac{p_{v^*}(x; z)}{v_{k+1} - v_k}$$

The method of Lewbel (2000) can be applied to the auxiliary model and data $(y; v_0; v^*; x; z)$ to get consistent estimates of parameter $\beta$ if several conditions hold. We shall only check the first of these conditions which is partial independence. What should hold is:

$$F(\varepsilon_0 \mid v_0; v^*; x; z) = F(\varepsilon_0 \mid v^*; x; z)$$

For convenience, omit the conditioning on $(v^*; x; z)$: Thus:

$$\begin{aligned} F(\varepsilon_0 \mid v_0) &= \Pr(\varepsilon + v - v_0 \leq \varepsilon_0 \mid v_0) \\ &= \int f_\varepsilon(\varepsilon \mid v_0) f_v(\varepsilon_0 - (\varepsilon - v_0) \mid v_0) d\varepsilon \end{aligned}$$

As $v_0$ is a random draw $f_\varepsilon(\varepsilon \mid v_0) = f_\varepsilon(\varepsilon)$ and $f_v(v \mid v_0) = f_v(v)$, we have:

$$F(\varepsilon_0 \mid v_0) = \int f_\varepsilon(\varepsilon) f_v(\varepsilon_0 - (\varepsilon - v_0)) d\varepsilon$$

The only dependence on $v_0$ occurs through the density function of $v$ and it is in the case of a uniform distribution only that partial independence holds:

$$F(\varepsilon_0 \mid v_0) = F(\varepsilon_0).$$

The other conditions should be checked and this is the large support one which "creates" the bias in the intercept term.

## B.4 Proof of Corollary 11

Same as Corollary 3 except that the maximisation of $E(\tilde{x}_p u^*(x;z))$ is obtained when:

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

and the minimization of such an expression is obtained when:

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

Furthermore, we have:

$$\begin{aligned}
\overline{\Delta}^*(x;z) &= \sum_{k=1}^{K-1} [(v_{k+1} - v_k)(G_{k+1}(x;z) - G_k(x;z))] \\
&= [(v_2 - v_1)(G_2(x;z) - G_1(x;z) + (v_3 - v_2)(G_3(x;z) - G_2(x;z)) + \dots \\
&\quad \dots + (v_{K-1} - v_{K-2})(G_{K-1}(x;z) - G_{K-2}(x;z)) + (v_K - v_{K-1})(1 - G_{K-1}(x;z))] \\
&= -(v_2 - v_1)G_1(x;z) + \sum_{k=2}^{K-1} (v_k - v_{k-1} - (v_{k+1} - v_k))G_k(x;z) + v_K - v_{K-1} \\
&= \sum_{k=1}^{K-1} (v_k - v_{k-1} - (v_{k+1} - v_k))E(y \mid v = v_k; x; z) + v_K - v_{K-1} \\
&= E(\frac{\theta_{U,k} y}{p_k(x;z)} \mid x;z) + v_K - v_{K-1} = E(\tilde{y}_U \mid x;z)
\end{aligned}$$

where by convention $v_0 = v_1$. Similarly:

$$\begin{aligned}
\underline{\Delta}^*(x;z) &= \sum_{k=1}^{K-1} [(v_{k+1} - v_k)(G_{k-1}(x;z) - G_k(x;z))] \\
&= [-(v_2 - v_1)G_1(x;z) + (v_3 - v_2)(G_1(x;z) - G_2(x;z)) + \dots \\
\dots + (v_{K-1} - v_{K-2}) &\quad .(G_{K-3}(x;z) - G_{K-2}(x;z)) + (v_K - v_{K-1})(G_{K-2}(x;z) - G_{K-1}(x;z))] \\
&= \sum_{k=1}^{K-2} (v_{k+2} - v_{k+1} - (v_{k+1} - v_k))G_k(x;z) - (v_K - v_{K-1})G_{K-1}(x;z) \\
&= \sum_{k=1}^{K-1} (v_{k+2} - v_{k+1} - (v_{k+1} - v_k))E(y \mid v = v_k; x; z) \\
&= E(\frac{\theta_{L,k} y}{p_k(x;z)} \mid x;z) = E(\tilde{y}_L \mid x;z)
\end{aligned}$$

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

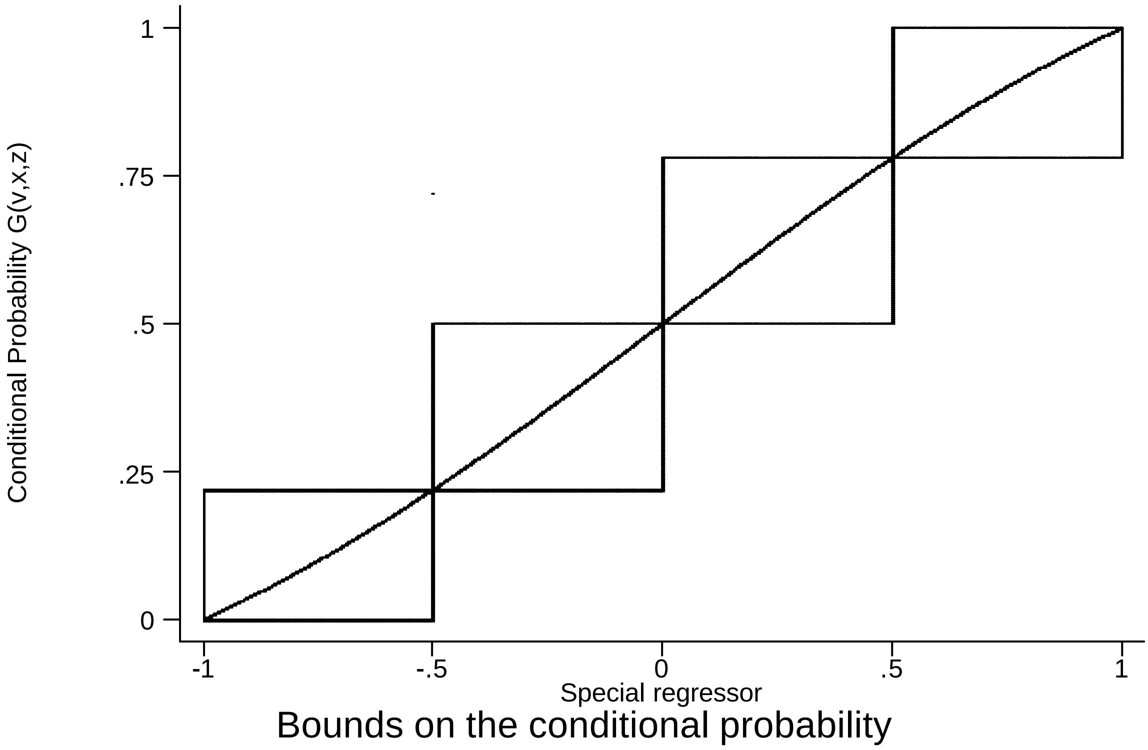


Figure 1: A graphical argument for set-identi…cation

Table 1: Simple experiment: Sensitivity to Bandwidth

Lower and upper estimated bounds with standard errors

| | | Intercept | | | | Variable | | | |
|---|---|---|---|---|---|---|---|---|---|
| Nobs | Bwidth | LB | SE | UB | SE | LB | SE | UB | SE |
| 100 | 1.0 | 0.40 | 0.53 | 1.26 | 0.54 | -0.42 | 0.58 | 0.37 | 0.59 |
| 100 | 1.5 | 0.21 | 0.42 | 1.07 | 0.43 | -0.39 | 0.49 | 0.38 | 0.50 |
| 100 | 3.0 | 0.02 | 0.33 | 0.88 | 0.33 | -0.39 | 0.40 | 0.34 | 0.41 |
| 100 | 5.0 | -0.02 | 0.35 | 0.84 | 0.35 | -0.40 | 0.41 | 0.32 | 0.41 |
| 200 | 1.0 | 0.11 | 0.25 | 0.97 | 0.25 | -0.28 | 0.33 | 0.46 | 0.34 |
| 200 | 1.5 | -0.06 | 0.22 | 0.79 | 0.22 | -0.32 | 0.27 | 0.41 | 0.27 |
| 200 | 3.0 | -0.23 | 0.18 | 0.63 | 0.18 | -0.36 | 0.24 | 0.35 | 0.24 |
| 200 | 5.0 | -0.26 | 0.19 | 0.60 | 0.19 | -0.38 | 0.26 | 0.32 | 0.26 |
| 500 | 1.0 | -0.22 | 0.12 | 0.63 | 0.12 | -0.31 | 0.16 | 0.40 | 0.16 |
| 500 | 1.5 | -0.31 | 0.12 | 0.54 | 0.12 | -0.35 | 0.14 | 0.36 | 0.14 |
| 500 | 3.0 | -0.38 | 0.11 | 0.47 | 0.11 | -0.39 | 0.14 | 0.31 | 0.14 |
| 500 | 5.0 | -0.40 | 0.11 | 0.45 | 0.11 | -0.40 | 0.15 | 0.30 | 0.15 |
| 1000 | 1.0 | -0.34 | 0.08 | 0.51 | 0.08 | -0.35 | 0.10 | 0.36 | 0.10 |
| 1000 | 1.5 | -0.39 | 0.08 | 0.46 | 0.08 | -0.37 | 0.09 | 0.33 | 0.09 |
| 1000 | 3.0 | -0.43 | 0.07 | 0.43 | 0.07 | -0.41 | 0.10 | 0.29 | 0.10 |
| 1000 | 5.0 | -0.44 | 0.07 | 0.42 | 0.07 | -0.41 | 0.11 | 0.29 | 0.11 |

Error Decomposition: Decentering, Adjusted Length and Sampling Error

| | | Intercept | | | | Variable | | | |
|---|---|---|---|---|---|---|---|---|---|
| Nobs | Bwidth | Dec | AL | ASE | RMSEI | Dec | AL | ASE | RMSEI |
| 100 | 1.0 | 0.83 | 0.25 | 0.53 | 1.02 | -0.02 | 0.23 | 0.59 | 0.63 |
| 100 | 1.5 | 0.64 | 0.25 | 0.43 | 0.81 | -0.01 | 0.22 | 0.50 | 0.54 |
| 100 | 3.0 | 0.45 | 0.25 | 0.33 | 0.61 | -0.03 | 0.21 | 0.41 | 0.46 |
| 100 | 5.0 | 0.41 | 0.25 | 0.35 | 0.59 | -0.04 | 0.21 | 0.41 | 0.46 |
| 200 | 1.0 | 0.54 | 0.25 | 0.25 | 0.65 | 0.09 | 0.22 | 0.34 | 0.41 |
| 200 | 1.5 | 0.36 | 0.25 | 0.22 | 0.49 | 0.04 | 0.21 | 0.27 | 0.34 |
| 200 | 3.0 | 0.20 | 0.25 | 0.18 | 0.37 | -0.01 | 0.20 | 0.24 | 0.31 |
| 200 | 5.0 | 0.17 | 0.25 | 0.19 | 0.35 | -0.03 | 0.20 | 0.26 | 0.33 |
| 500 | 1.0 | 0.20 | 0.25 | 0.12 | 0.34 | 0.04 | 0.21 | 0.16 | 0.26 |
| 500 | 1.5 | 0.12 | 0.25 | 0.12 | 0.30 | 0.01 | 0.20 | 0.14 | 0.25 |
| 500 | 3.0 | 0.04 | 0.25 | 0.11 | 0.27 | -0.04 | 0.20 | 0.14 | 0.25 |
| 500 | 5.0 | 0.03 | 0.25 | 0.11 | 0.27 | -0.05 | 0.20 | 0.15 | 0.26 |
| 1000 | 1.0 | 0.08 | 0.25 | 0.08 | 0.27 | 0.00 | 0.20 | 0.09 | 0.23 |
| 1000 | 1.5 | 0.04 | 0.25 | 0.08 | 0.26 | -0.02 | 0.20 | 0.09 | 0.22 |
| 1000 | 3.0 | -0.00 | 0.25 | 0.07 | 0.26 | -0.06 | 0.20 | 0.10 | 0.23 |
| 1000 | 5.0 | -0.01 | 0.25 | 0.07 | 0.26 | -0.06 | 0.20 | 0.11 | 0.24 |

Notes: The number of discrete values is equal to 10. The simple experiment refers to the case where ® = ½ = 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 constant bandwidth that is used. Dec stands for decentering of the mid-point of the interval that is, (UB+LB)/2. AL is the adjusted length of the interval, $(UB-LB)/2\sqrt{3}$. ASE is the sampling variability of bounds as de…ned in the text. The identity $Dec^2 + AL^2 + ASE^2 = RMSEI^2$; is shown in the text. RMSEI is the root mean square.error integrated over the identi…cation set.

Table 2: Sensitivity to Normality

Lower and upper estimated bounds with standard errors

| | | Intercept | | | | Variable | | | |
|---|---|---|---|---|---|---|---|---|---|
| Nobs | Alpha | LB | SE | UB | SE | LB | SE | UB | SE |
| 100 | 0.00 | 0.02 | 0.33 | 0.88 | 0.33 | -0.39 | 0.40 | 0.34 | 0.41 |
| 100 | 0.33 | -0.06 | 0.35 | 0.81 | 0.35 | -0.29 | 0.41 | 0.44 | 0.42 |
| 100 | 0.67 | -0.17 | 0.35 | 0.72 | 0.35 | -0.24 | 0.44 | 0.52 | 0.44 |
| 100 | 1.00 | -0.34 | 0.36 | 0.60 | 0.36 | -0.19 | 0.44 | 0.60 | 0.45 |
| 200 | 0.00 | -0.23 | 0.18 | 0.63 | 0.18 | -0.36 | 0.24 | 0.35 | 0.24 |
| 200 | 0.33 | -0.31 | 0.20 | 0.56 | 0.20 | -0.31 | 0.24 | 0.41 | 0.25 |
| 200 | 0.67 | -0.42 | 0.20 | 0.47 | 0.20 | -0.27 | 0.24 | 0.46 | 0.25 |
| 200 | 1.00 | -0.59 | 0.20 | 0.36 | 0.20 | -0.25 | 0.25 | 0.52 | 0.25 |
| 500 | 0.00 | -0.38 | 0.11 | 0.47 | 0.11 | -0.39 | 0.14 | 0.31 | 0.14 |
| 500 | 0.33 | -0.45 | 0.11 | 0.41 | 0.11 | -0.33 | 0.14 | 0.38 | 0.14 |
| 500 | 0.67 | -0.57 | 0.11 | 0.33 | 0.11 | -0.29 | 0.14 | 0.44 | 0.15 |
| 500 | 1.00 | -0.73 | 0.11 | 0.21 | 0.11 | -0.28 | 0.15 | 0.49 | 0.15 |
| 1000 | 0.00 | -0.43 | 0.07 | 0.43 | 0.07 | -0.41 | 0.10 | 0.29 | 0.10 |
| 1000 | 0.33 | -0.50 | 0.07 | 0.37 | 0.07 | -0.35 | 0.09 | 0.36 | 0.10 |
| 1000 | 0.67 | -0.61 | 0.08 | 0.28 | 0.08 | -0.32 | 0.10 | 0.41 | 0.10 |
| 1000 | 1.00 | -0.77 | 0.08 | 0.17 | 0.08 | -0.29 | 0.10 | 0.47 | 0.11 |

Error Decomposition: Decentering, Adjusted Length and Sampling Error

| | | Intercept | | | | Variable | | | |
|---|---|---|---|---|---|---|---|---|---|
| Nobs | Alpha | Dec | AL | ASE | RMSEI | Dec | AL | ASE | RMSEI |
| 100 | 0.00 | 0.45 | 0.25 | 0.33 | 0.61 | -0.03 | 0.21 | 0.41 | 0.46 |
| 100 | 0.33 | 0.37 | 0.25 | 0.35 | 0.57 | 0.07 | 0.21 | 0.41 | 0.47 |
| 100 | 0.67 | 0.27 | 0.26 | 0.35 | 0.51 | 0.14 | 0.22 | 0.44 | 0.51 |
| 100 | 1.00 | 0.13 | 0.27 | 0.36 | 0.47 | 0.20 | 0.23 | 0.44 | 0.54 |
| 200 | 0.00 | 0.20 | 0.25 | 0.18 | 0.37 | -0.01 | 0.20 | 0.24 | 0.31 |
| 200 | 0.33 | 0.12 | 0.25 | 0.20 | 0.34 | 0.05 | 0.21 | 0.25 | 0.32 |
| 200 | 0.67 | 0.02 | 0.26 | 0.20 | 0.33 | 0.10 | 0.21 | 0.25 | 0.34 |
| 200 | 1.00 | -0.12 | 0.27 | 0.20 | 0.35 | 0.14 | 0.22 | 0.25 | 0.36 |
| 500 | 0.00 | 0.04 | 0.25 | 0.11 | 0.27 | -0.04 | 0.20 | 0.14 | 0.25 |
| 500 | 0.33 | -0.02 | 0.25 | 0.11 | 0.27 | 0.02 | 0.20 | 0.14 | 0.25 |
| 500 | 0.67 | -0.12 | 0.26 | 0.11 | 0.31 | 0.07 | 0.21 | 0.14 | 0.26 |
| 500 | 1.00 | -0.26 | 0.27 | 0.11 | 0.39 | 0.11 | 0.22 | 0.15 | 0.29 |
| 1000 | 0.00 | -0.00 | 0.25 | 0.07 | 0.26 | -0.06 | 0.20 | 0.10 | 0.23 |
| 1000 | 0.33 | -0.07 | 0.25 | 0.07 | 0.27 | 0.00 | 0.20 | 0.10 | 0.23 |
| 1000 | 0.67 | -0.17 | 0.26 | 0.08 | 0.32 | 0.05 | 0.21 | 0.10 | 0.24 |
| 1000 | 1.00 | -0.30 | 0.27 | 0.08 | 0.41 | 0.09 | 0.22 | 0.11 | 0.26 |

Notes: See Table 1 for main comments. Speci…cs are: The bandwidth is equal to 3.0. The Alpha column refers to the increasing amount of non-normaliity.

Table 3: Sensitivity to Endogeneity

Lower and upper estimated bounds with standard errors

| | | Intercept | | | | Variable | | | |
|---|---|---|---|---|---|---|---|---|---|
| Nobs | Rho | LB | SE | UB | SE | LB | SE | UB | SE |
| 100 | 0.00 | 0.02 | 0.33 | 0.88 | 0.33 | -0.39 | 0.40 | 0.34 | 0.41 |
| 100 | 0.33 | -0.31 | 0.55 | 0.54 | 0.55 | -0.63 | 0.57 | 0.34 | 0.59 |
| 100 | 0.67 | -0.32 | 0.55 | 0.53 | 0.55 | -0.65 | 0.57 | 0.33 | 0.59 |
| 100 | 1.00 | -0.34 | 0.54 | 0.52 | 0.54 | -0.67 | 0.56 | 0.31 | 0.57 |
| 200 | 0.00 | -0.23 | 0.18 | 0.63 | 0.18 | -0.36 | 0.24 | 0.35 | 0.24 |
| 200 | 0.33 | -0.12 | 0.24 | 0.73 | 0.24 | -0.49 | 0.30 | 0.33 | 0.30 |
| 200 | 0.67 | -0.13 | 0.24 | 0.72 | 0.24 | -0.50 | 0.30 | 0.33 | 0.30 |
| 200 | 1.00 | -0.14 | 0.24 | 0.71 | 0.24 | -0.51 | 0.30 | 0.32 | 0.30 |
| 500 | 0.00 | -0.38 | 0.11 | 0.47 | 0.11 | -0.39 | 0.14 | 0.31 | 0.14 |
| 500 | 0.33 | -0.35 | 0.11 | 0.50 | 0.11 | -0.42 | 0.14 | 0.33 | 0.14 |
| 500 | 0.67 | -0.36 | 0.11 | 0.50 | 0.11 | -0.43 | 0.14 | 0.33 | 0.14 |
| 500 | 1.00 | -0.37 | 0.11 | 0.49 | 0.11 | -0.44 | 0.14 | 0.33 | 0.14 |
| 1000 | 0.00 | -0.43 | 0.07 | 0.43 | 0.07 | -0.41 | 0.10 | 0.29 | 0.10 |
| 1000 | 0.33 | -0.43 | 0.07 | 0.43 | 0.07 | -0.43 | 0.09 | 0.31 | 0.09 |
| 1000 | 0.67 | -0.43 | 0.07 | 0.42 | 0.07 | -0.44 | 0.09 | 0.31 | 0.09 |
| 1000 | 1.00 | -0.44 | 0.08 | 0.42 | 0.08 | -0.44 | 0.10 | 0.30 | 0.10 |

Error Decomposition: Decentering, Adjusted Length and Sampling Error

| | | Intercept | | | | Variable | | | |
|---|---|---|---|---|---|---|---|---|---|
| Nobs | Rho | Dec | AL | ASE | RMSEI | Dec | AL | ASE | RMSEI |
| 100 | 0.00 | 0.45 | 0.25 | 0.33 | 0.61 | -0.03 | 0.21 | 0.41 | 0.46 |
| 100 | 0.33 | 0.12 | 0.25 | 0.55 | 0.61 | -0.15 | 0.28 | 0.58 | 0.66 |
| 100 | 0.67 | 0.10 | 0.25 | 0.55 | 0.61 | -0.16 | 0.28 | 0.58 | 0.66 |
| 100 | 1.00 | 0.09 | 0.25 | 0.54 | 0.60 | -0.18 | 0.28 | 0.57 | 0.66 |
| 200 | 0.00 | 0.20 | 0.25 | 0.18 | 0.37 | -0.01 | 0.20 | 0.24 | 0.31 |
| 200 | 0.33 | 0.30 | 0.25 | 0.24 | 0.46 | -0.08 | 0.24 | 0.30 | 0.39 |
| 200 | 0.67 | 0.30 | 0.25 | 0.24 | 0.45 | -0.09 | 0.24 | 0.30 | 0.39 |
| 200 | 1.00 | 0.29 | 0.25 | 0.24 | 0.45 | -0.10 | 0.24 | 0.30 | 0.40 |
| 500 | 0.00 | 0.04 | 0.25 | 0.11 | 0.27 | -0.04 | 0.20 | 0.14 | 0.25 |
| 500 | 0.33 | 0.07 | 0.25 | 0.11 | 0.28 | -0.05 | 0.22 | 0.14 | 0.26 |
| 500 | 0.67 | 0.07 | 0.25 | 0.11 | 0.28 | -0.05 | 0.22 | 0.14 | 0.26 |
| 500 | 1.00 | 0.06 | 0.25 | 0.11 | 0.28 | -0.06 | 0.22 | 0.14 | 0.27 |
| 1000 | 0.00 | -0.00 | 0.25 | 0.07 | 0.26 | -0.06 | 0.20 | 0.10 | 0.23 |
| 1000 | 0.33 | 0.00 | 0.25 | 0.07 | 0.26 | -0.06 | 0.21 | 0.09 | 0.24 |
| 1000 | 0.67 | -0.00 | 0.25 | 0.07 | 0.26 | -0.07 | 0.22 | 0.09 | 0.24 |
| 1000 | 1.00 | -0.01 | 0.25 | 0.08 | 0.26 | -0.07 | 0.22 | 0.10 | 0.25 |

Notes: See Table 1 for main comments. Speci…cs are: The bandwidth is equal to 3.0. The Rho column refers to the increasing amount of endogeneity.

Table 4: Sensitivity to the Number of Discrete Points

Lower and upper estimated bounds with standard errors

| | | Intercept | | | | Variable | | | |
|---|---|---|---|---|---|---|---|---|---|
| Nobs | Points | LB | SE | UB | SE | LB | SE | UB | SE |
| 100 | 5 | -0.87 | 0.31 | 1.05 | 0.31 | -0.83 | 0.36 | 0.77 | 0.36 |
| 100 | 10 | 0.02 | 0.33 | 0.88 | 0.33 | -0.39 | 0.40 | 0.34 | 0.41 |
| 100 | 20 | -0.05 | 0.56 | 0.36 | 0.55 | -0.25 | 0.48 | 0.13 | 0.49 |
| 100 | 40 | -1.23 | 0.59 | -0.97 | 0.55 | -0.40 | 0.51 | -0.13 | 0.53 |
| 200 | 5 | -0.94 | 0.19 | 0.98 | 0.19 | -0.83 | 0.23 | 0.76 | 0.23 |
| 200 | 10 | -0.23 | 0.18 | 0.63 | 0.18 | -0.36 | 0.24 | 0.35 | 0.24 |
| 200 | 20 | 0.16 | 0.32 | 0.57 | 0.32 | -0.19 | 0.32 | 0.15 | 0.33 |
| 200 | 40 | -0.20 | 0.40 | 0.00 | 0.40 | -0.18 | 0.35 | 0.01 | 0.35 |
| 500 | 5 | -0.98 | 0.11 | 0.94 | 0.11 | -0.85 | 0.13 | 0.72 | 0.13 |
| 500 | 10 | -0.38 | 0.11 | 0.47 | 0.11 | -0.39 | 0.14 | 0.31 | 0.14 |
| 500 | 20 | 0.00 | 0.12 | 0.41 | 0.12 | -0.18 | 0.17 | 0.15 | 0.17 |
| 500 | 40 | 0.22 | 0.20 | 0.41 | 0.20 | -0.09 | 0.22 | 0.08 | 0.22 |
| 1000 | 5 | -1.00 | 0.08 | 0.92 | 0.08 | -0.87 | 0.09 | 0.71 | 0.09 |
| 1000 | 10 | -0.43 | 0.07 | 0.43 | 0.07 | -0.41 | 0.10 | 0.29 | 0.10 |
| 1000 | 20 | -0.13 | 0.07 | 0.28 | 0.07 | -0.20 | 0.10 | 0.13 | 0.11 |
| 1000 | 40 | 0.13 | 0.09 | 0.33 | 0.09 | -0.10 | 0.13 | 0.07 | 0.13 |

Error Decomposition: Decentering, Adjusted Length and Sampling Error

| | | Intercept | | | | Variable | | | |
|---|---|---|---|---|---|---|---|---|---|
| Nobs | Points | Dec | AL | ASE | RMSEI | Dec | AL | ASE | RMSEI |
| 100 | 5 | 0.09 | 0.55 | 0.31 | 0.64 | -0.03 | 0.46 | 0.36 | 0.59 |
| 100 | 10 | 0.45 | 0.25 | 0.33 | 0.61 | -0.03 | 0.21 | 0.41 | 0.46 |
| 100 | 20 | 0.16 | 0.12 | 0.55 | 0.59 | -0.06 | 0.11 | 0.48 | 0.50 |
| 100 | 40 | -1.10 | 0.08 | 0.55 | 1.23 | -0.27 | 0.08 | 0.50 | 0.58 |
| 200 | 5 | 0.02 | 0.55 | 0.19 | 0.59 | -0.03 | 0.46 | 0.23 | 0.51 |
| 200 | 10 | 0.20 | 0.25 | 0.18 | 0.37 | -0.01 | 0.20 | 0.24 | 0.31 |
| 200 | 20 | 0.36 | 0.12 | 0.32 | 0.50 | -0.02 | 0.10 | 0.32 | 0.34 |
| 200 | 40 | -0.10 | 0.06 | 0.40 | 0.42 | -0.09 | 0.05 | 0.35 | 0.36 |
| 500 | 5 | -0.02 | 0.55 | 0.11 | 0.57 | -0.06 | 0.46 | 0.13 | 0.48 |
| 500 | 10 | 0.04 | 0.25 | 0.11 | 0.27 | -0.04 | 0.20 | 0.14 | 0.25 |
| 500 | 20 | 0.20 | 0.12 | 0.12 | 0.26 | -0.01 | 0.10 | 0.17 | 0.19 |
| 500 | 40 | 0.32 | 0.06 | 0.20 | 0.38 | -0.01 | 0.05 | 0.22 | 0.22 |
| 1000 | 5 | -0.04 | 0.55 | 0.08 | 0.56 | -0.08 | 0.45 | 0.09 | 0.47 |
| 1000 | 10 | -0.00 | 0.25 | 0.07 | 0.26 | -0.06 | 0.20 | 0.10 | 0.23 |
| 1000 | 20 | 0.07 | 0.12 | 0.07 | 0.16 | -0.04 | 0.10 | 0.10 | 0.15 |
| 1000 | 40 | 0.23 | 0.06 | 0.09 | 0.25 | -0.02 | 0.05 | 0.13 | 0.14 |

Notes: See Table 1 for main comments. Specifics are: The bandwidth is equal to 3.0. The Discrete column refers to the number of points in the support of v.

Table 5: Simple experiment, Interval Data: Sensitivity to Bandwidth

Lower and upper estimated bounds with standard errors

| | | Intercept | | | | Variable | | | |
|---|---|---|---|---|---|---|---|---|---|
| Nobs | Bwidth | LB | SE | UB | SE | LB | SE | UB | SE |
| 100 | 1.000 | -0.133 | 0.384 | 1.240 | 0.449 | -0.652 | 0.516 | 0.571 | 0 .507 |
| | 1.500 | -0.290 | 0.314 | 1.110 | 0.369 | -0.677 | 0.433 | 0.541 | 0 .430 |
| | 3.000 | -0.447 | 0.248 | 0.941 | 0.316 | -0.663 | 0.356 | 0.505 | 0 .340 |
| | 5.000 | -0.480 | 0.245 | 0.845 | 0.361 | -0.646 | 0.355 | 0.466 | 0 .343 |
| 200 | 1.000 | -0.383 | 0.194 | 1.099 | 0.233 | -0.643 | 0.283 | 0.657 | 0 .280 |
| | 1.500 | -0.512 | 0.169 | 0.948 | 0.187 | -0.674 | 0.243 | 0.596 | 0 .234 |
| | 3.000 | -0.634 | 0.147 | 0.793 | 0.150 | -0.708 | 0.212 | 0.506 | 0 .200 |
| | 5.000 | -0.663 | 0.145 | 0.756 | 0.140 | -0.726 | 0.220 | 0.475 | 0 .208 |
| 500 | 1.000 | -0.616 | 0.095 | 0.809 | 0.099 | -0.672 | 0.131 | 0.566 | 0 .126 |
| | 1.500 | -0.678 | 0.088 | 0.728 | 0.087 | -0.692 | 0.118 | 0.516 | 0 .108 |
| | 3.000 | -0.731 | 0.084 | 0.656 | 0.081 | -0.719 | 0.118 | 0.460 | 0 .107 |
| | 5.000 | -0.742 | 0.084 | 0.641 | 0.080 | -0.725 | 0.124 | 0.448 | 0 .117 |
| 1000 | 1.000 | -0.702 | 0.061 | 0.691 | 0.059 | -0.690 | 0.078 | 0.503 | 0.076 |
| | 1.500 | -0.735 | 0.059 | 0.647 | 0.056 | -0.706 | 0.072 | 0.468 | 0.070 |
| | 3.000 | -0.762 | 0.058 | 0.610 | 0.055 | -0.725 | 0.079 | 0.433 | 0.075 |
| | 5.000 | -0.767 | 0.058 | 0.604 | 0.054 | -0.729 | 0.084 | 0.426 | 0.082 |

Error Decomposition: Decentering, Adjusted Length and Sampling Error

| | | Intercept | | | | Variable | | | |
|---|---|---|---|---|---|---|---|---|---|
| Nobs | Bwidth | 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 | 0.4 61 |
| | 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 ® = ½ = 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 constant bandwidth that is used. Dec stands for decentering of the mid-point of the interval that is, (UB+LB)/2. AL is the adjusted length of the interval, $(UB\text{-}LB)/2\sqrt{3}$. ASE is the sampling variability of bounds as de…ned in the text. The identity $Dec^2 + AL^2 + ASE^2 = RMSEI^2$ is shown in the text. is shown in the text. RMSEI is the root mean square.error integrated over the identi…cation set.

Table 6: Sensitivity to Normality, Interval Data

Lower and upper estimated bounds with standard errors

| | | Intercept | | | | Variable | | | |
|---|---|---|---|---|---|---|---|---|---|
| Nobs | Alpha | LB | SE | UB | SE | LB | SE | UB | SE |
| 100 | 0.000 | -0.641 | 0.244 | 0.916 | 0.203 | -0.820 | 0.502 | 0.605 | 0 .314 |
| 100 | 0.333 | -0.674 | 0.232 | 0.602 | 0.306 | -0.697 | 0.285 | 0.351 | 0 .238 |
| 100 | 0.667 | -0.659 | 0.165 | 0.563 | 0.352 | -0.403 | 0.351 | 0.569 | 0 .337 |
| 100 | 1.000 | -0.673 | 0.238 | 0.758 | 0.306 | -0.136 | 0.483 | 1.077 | 0 .332 |
| 200 | 0.000 | -0.666 | 0.120 | 0.805 | 0.084 | -0.695 | 0.162 | 0.582 | 0 .152 |
| 200 | 0.333 | -0.720 | 0.087 | 0.799 | 0.104 | -0.639 | 0.199 | 0.664 | 0 .131 |
| 200 | 0.667 | -0.688 | 0.184 | 0.800 | 0.212 | -0.502 | 0.194 | 0.730 | 0 .153 |
| 200 | 1.000 | -0.797 | 0.183 | 0.776 | 0.264 | -0.160 | 0.165 | 1.106 | 0 .161 |
| 500 | 0.000 | -0.714 | 0.032 | 0.660 | 0.098 | -0.728 | 0.133 | 0.460 | 0 .139 |
| 500 | 0.333 | -0.819 | 0.084 | 0.618 | 0.083 | -0.650 | 0.134 | 0.614 | 0 .087 |
| 500 | 0.667 | -0.864 | 0.094 | 0.634 | 0.106 | -0.552 | 0.066 | 0.747 | 0 .096 |
| 500 | 1.000 | -0.891 | 0.097 | 0.663 | 0.093 | -0.486 | 0.165 | 0.889 | 0 .162 |
| 1000 | 0.000 | -0.790 | 0.039 | 0.598 | 0.032 | -0.742 | 0.087 | 0.460 | 0.070 |
| 1000 | 0.333 | -0.816 | 0.033 | 0.602 | 0.047 | -0.635 | 0.088 | 0.590 | 0.056 |
| 1000 | 0.667 | -0.844 | 0.064 | 0.606 | 0.049 | -0.484 | 0.086 | 0.771 | 0.071 |
| 1000 | 1.000 | -0.937 | 0.042 | 0.597 | 0.039 | -0.430 | 0.102 | 0.941 | 0.097 |

Error Decomposition: Decentering, Adjusted Length and Sampling Error

| | | Intercept | | | | Variable | | | |
|---|---|---|---|---|---|---|---|---|---|
| Nobs | Alpha | Dec | AL | ASE | RMSEI | Dec | AL | ASE | RMSEI |
| 100 | 0 | 0.137 | 0.449 | 0.220 | 0.519 | -0.107 | 0.411 | 0.401 | 0. 584 |
| 100 | 0.33 | -0.036 | 0.368 | 0.243 | 0.443 | -0.173 | 0.302 | 0.246 | 0 .427 |
| 100 | 0.66 | -0.048 | 0.353 | 0.219 | 0.418 | 0.083 | 0.281 | 0.328 | 0. 440 |
| 100 | 1 | 0.043 | 0.413 | 0.238 | 0.479 | 0.470 | 0.350 | 0.404 | 0.7 12 |
| 200 | 0 | 0.069 | 0.424 | 0.100 | 0.442 | -0.057 | 0.369 | 0.130 | 0. 395 |
| 200 | 0.33 | 0.039 | 0.438 | 0.094 | 0.450 | 0.013 | 0.376 | 0.164 | 0.4 11 |
| 200 | 0.66 | 0.056 | 0.429 | 0.196 | 0.475 | 0.114 | 0.356 | 0.160 | 0.4 07 |
| 200 | 1 | -0.011 | 0.454 | 0.223 | 0.506 | 0.473 | 0.365 | 0.161 | 0. 619 |
| 500 | 0 | -0.027 | 0.397 | 0.067 | 0.403 | -0.134 | 0.343 | 0.130 | 0 .391 |
| 500 | 0.33 | -0.100 | 0.415 | 0.083 | 0.435 | -0.018 | 0.365 | 0.110 | 0 .381 |
| 500 | 0.66 | -0.115 | 0.433 | 0.100 | 0.459 | 0.098 | 0.375 | 0.081 | 0. 396 |
| 500 | 1 | -0.114 | 0.448 | 0.094 | 0.472 | 0.202 | 0.397 | 0.163 | 0. 474 |
| 1000 | 0 | -0.096 | 0.401 | 0.036 | 0.414 | -0.141 | 0.347 | 0.077 | 0.383 |
| 1000 | 0.33 | -0.107 | 0.409 | 0.040 | 0.425 | -0.022 | 0.354 | 0.071 | 0.361 |
| 1000 | 0.66 | -0.119 | 0.419 | 0.056 | 0.439 | 0.144 | 0.362 | 0.078 | 0 .398 |
| 1000 | 1 | -0.170 | 0.443 | 0.039 | 0.476 | 0.255 | 0.396 | 0.098 | 0 .481 |

Notes: See Table 5 for main comments. Speci…cs are: The bandwidth is equal to 3.0. The Alpha column refers to the increasing amount of non-normality.

Table 7: Sensitivity to Endogeneity, Interval Data

Lower and upper estimated bounds with standard errors

| | | Intercept | | | | Variable | | | |
|---|---|---|---|---|---|---|---|---|---|
| Nobs | Rho | LB | SE | UB | SE | L B | SE | UB | SE |
| 100 | 0.000 | -0.447 | 0.248 | 0.941 | 0.316 | -0.663 | 0.356 | 0.505 | 0 .340 |
| 100 | 0.333 | -0.589 | 0.441 | 0.245 | 0.484 | -0.600 | 0.516 | 0.362 | 0 .522 |
| 100 | 0.667 | -0.602 | 0.434 | 0.235 | 0.479 | -0.626 | 0.515 | 0.343 | 0 .522 |
| 100 | 1.000 | -0.625 | 0.429 | 0.214 | 0.475 | -0.659 | 0.513 | 0.315 | 0 .515 |
| 200 | 0.000 | -0.634 | 0.147 | 0.793 | 0.150 | -0.708 | 0.212 | 0.506 | 0 .200 |
| 200 | 0.333 | -0.583 | 0.168 | 0.649 | 0.254 | -0.684 | 0.274 | 0.461 | 0 .265 |
| 200 | 0.667 | -0.592 | 0.168 | 0.640 | 0.250 | -0.697 | 0.279 | 0.453 | 0 .268 |
| 200 | 1.000 | -0.604 | 0.168 | 0.630 | 0.248 | -0.714 | 0.281 | 0.442 | 0 .268 |
| 500 | 0.000 | -0.731 | 0.084 | 0.656 | 0.081 | -0.719 | 0.118 | 0.460 | 0 .107 |
| 500 | 0.333 | -0.726 | 0.085 | 0.657 | 0.085 | -0.757 | 0.125 | 0.501 | 0 .120 |
| 500 | 0.667 | -0.734 | 0.086 | 0.648 | 0.085 | -0.768 | 0.125 | 0.494 | 0 .118 |
| 500 | 1.000 | -0.744 | 0.086 | 0.635 | 0.086 | -0.782 | 0.127 | 0.482 | 0 .118 |
| 1000 | 0.000 | -0.762 | 0.058 | 0.610 | 0.055 | -0.725 | 0.079 | 0.433 | 0.075 |
| 1000 | 0.333 | -0.762 | 0.058 | 0.610 | 0.055 | -0.764 | 0.081 | 0.467 | 0.077 |
| 1000 | 0.667 | -0.769 | 0.059 | 0.601 | 0.056 | -0.774 | 0.082 | 0.461 | 0.078 |
| 1000 | 1.000 | -0.781 | 0.059 | 0.589 | 0.056 | -0.790 | 0.083 | 0.451 | 0.078 |

Error Decomposition: Decentering, Adjusted Length and Sampling Error

| | | Intercept | | | | Variable | | | |
|---|---|---|---|---|---|---|---|---|---|
| Nobs | Rho | Dec | AL | ASE | RMSEI | Dec | AL | ASE | RMSEI |
| 100 | 0.000 | 0.247 | 0.401 | 0.266 | 0.541 | -0.079 | 0.337 | 0.327 | 0. 476 |
| 100 | 0.333 | -0.172 | 0.241 | 0.448 | 0.537 | -0.119 | 0.278 | 0.501 | 0 .585 |
| 100 | 0.667 | -0.183 | 0.242 | 0.442 | 0.536 | -0.141 | 0.280 | 0.500 | 0 .590 |
| 100 | 1.000 | -0.205 | 0.242 | 0.437 | 0.540 | -0.172 | 0.281 | 0.495 | 0 .594 |
| 200 | 0.000 | 0.079 | 0.412 | 0.144 | 0.444 | -0.101 | 0.350 | 0.191 | 0. 412 |
| 200 | 0.333 | 0.033 | 0.356 | 0.193 | 0.406 | -0.112 | 0.331 | 0.249 | 0. 429 |
| 200 | 0.667 | 0.024 | 0.355 | 0.190 | 0.404 | -0.122 | 0.332 | 0.253 | 0. 435 |
| 200 | 1.000 | 0.013 | 0.356 | 0.190 | 0.404 | -0.136 | 0.334 | 0.254 | 0. 441 |
| 500 | 0.000 | -0.038 | 0.400 | 0.080 | 0.410 | -0.129 | 0.340 | 0.104 | 0 .378 |
| 500 | 0.333 | -0.034 | 0.399 | 0.083 | 0.409 | -0.128 | 0.363 | 0.113 | 0 .401 |
| 500 | 0.667 | -0.043 | 0.399 | 0.083 | 0.410 | -0.137 | 0.364 | 0.113 | 0 .405 |
| 500 | 1.000 | -0.055 | 0.398 | 0.083 | 0.410 | -0.150 | 0.365 | 0.113 | 0 .410 |
| 1000 | 0.000 | -0.076 | 0.396 | 0.055 | 0.407 | -0.146 | 0.334 | 0.071 | 0.372 |
| 1000 | 0.333 | -0.076 | 0.396 | 0.055 | 0.407 | -0.148 | 0.356 | 0.073 | 0.392 |
| 1000 | 0.667 | -0.084 | 0.395 | 0.056 | 0.408 | -0.157 | 0.356 | 0.074 | 0.396 |
| 1000 | 1.000 | -0.096 | 0.396 | 0.056 | 0.411 | -0.170 | 0.358 | 0.075 | 0.403 |

Notes: See Table 5 for main comments. Speci…cs are: The bandwidth is equal to 3.0. The Rho column refers to the increasing amount of endogeneity.

Table 8: Sensitivity to the Number of Intervals, Interval Data

Lower and upper estimated bounds with standard errors

| | | Intercept | | | | Variable | | | |
|---|---|---|---|---|---|---|---|---|---|
| Nobs | Intervals | LB | SE | UB | SE | LB | SE | UB | SE |
| 100 | 5 | -1.104 | 0.195 | 0.998 | 0.177 | -1.150 | 0.274 | 0.710 | 0 .242 |
| 100 | 20 | -0.166 | 0.508 | 0.299 | 0.527 | -0.278 | 0.456 | 0.133 | 0.449 |
| 100 | 40 | -1.090 | 0.542 | -0.843 | 0.518 | -0.360 | 0.486 | -0.114 | 0.502 |
| 100 | 80 | -2.153 | 0.551 | -2.020 | 0.525 | -0.533 | 0.533 | -0.379 | 0.565 |
| 200 | 5 | -1.173 | 0.129 | 0.901 | 0.114 | -1.140 | 0.178 | 0.676 | 0 .155 |
| 200 | 20 | 0.025 | 0.252 | 0.640 | 0.305 | -0.290 | 0.307 | 0.209 | 0 .311 |
| 200 | 40 | -0.148 | 0.378 | 0.074 | 0.371 | -0.147 | 0.332 | 0.054 | 0.329 |
| 200 | 80 | -1.198 | 0.387 | -1.078 | 0.384 | -0.300 | 0.331 | -0.177 | 0.335 |
| 500 | 5 | -1.206 | 0.078 | 0.847 | 0.065 | -1.138 | 0.109 | 0.643 | 0 .089 |
| 500 | 20 | -0.203 | 0.100 | 0.530 | 0.110 | -0.351 | 0.150 | 0.258 | 0.149 |
| 500 | 40 | 0.157 | 0.164 | 0.484 | 0.191 | -0.142 | 0.210 | 0.117 | 0 .206 |
| 500 | 80 | 0.003 | 0.246 | 0.124 | 0.248 | -0.066 | 0.213 | 0.036 | 0 .215 |
| 1000 | 5 | -1.215 | 0.054 | 0.831 | 0.045 | -1.139 | 0.076 | 0.629 | 0.064 |
| 1000 | 20 | -0.315 | 0.062 | 0.401 | 0.064 | -0.370 | 0.095 | 0.225 | 0.091 |
| 1000 | 40 | 0.013 | 0.077 | 0.398 | 0.084 | -0.187 | 0.124 | 0.125 | 0.122 |
| 1000 | 80 | 0.225 | 0.133 | 0.385 | 0.142 | -0.074 | 0.151 | 0.052 | 0.151 |

Error Decomposition: Decentering, Adjusted Length and Sampling Error

| | | Intercept | | | | Variable | | | |
|---|---|---|---|---|---|---|---|---|---|
| Nobs | Discrete | Dec | AL | ASE | RMSEI | Dec | AL | ASE | RMSEI |
| 100 | 5.000 | -0.053 | 0.607 | 0.174 | 0.633 | -0.220 | 0.537 | 0.225 | 0 .622 |
| 100 | 20.000 | 0.066 | 0.134 | 0.506 | 0.528 | -0.073 | 0.119 | 0.443 | 0 .464 |
| 100 | 40.000 | -0.967 | 0.071 | 0.511 | 1.096 | -0.237 | 0.071 | 0.475 | 0.536 |
| 100 | 80.000 | -2.087 | 0.038 | 0.499 | 2.146 | -0.456 | 0.044 | 0.504 | 0.681 |
| 200 | 5.000 | -0.136 | 0.599 | 0.115 | 0.625 | -0.232 | 0.524 | 0.146 | 0 .592 |
| 200 | 20.000 | 0.333 | 0.177 | 0.269 | 0.463 | -0.041 | 0.144 | 0.301 | 0 .336 |
| 200 | 40.000 | -0.037 | 0.064 | 0.368 | 0.375 | -0.047 | 0.058 | 0.325 | 0.333 |
| 200 | 80.000 | -1.138 | 0.035 | 0.373 | 1.198 | -0.239 | 0.035 | 0.320 | 0.401 |
| 500 | 5.000 | -0.179 | 0.593 | 0.068 | 0.623 | -0.247 | 0.514 | 0.087 | 0 .577 |
| 500 | 20.000 | 0.164 | 0.212 | 0.103 | 0.286 | -0.047 | 0.176 | 0.143 | 0 .231 |
| 500 | 40.000 | 0.320 | 0.095 | 0.174 | 0.376 | -0.012 | 0.075 | 0.205 | 0 .219 |
| 500 | 80.000 | 0.063 | 0.035 | 0.245 | 0.255 | -0.015 | 0.029 | 0.212 | 0 .215 |
| 1000 | 5.000 | -0.192 | 0.591 | 0.047 | 0.623 | -0.255 | 0.511 | 0.061 | 0.574 |
| 1000 | 20.000 | 0.043 | 0.207 | 0.062 | 0.220 | -0.073 | 0.172 | 0.090 | 0.207 |
| 1000 | 40.000 | 0.205 | 0.111 | 0.079 | 0.247 | -0.031 | 0.090 | 0.121 | 0.154 |
| 1000 | 80.000 | 0.305 | 0.046 | 0.136 | 0.337 | -0.011 | 0.036 | 0.150 | 0.155 |

Notes: See Table 5 for main comments. Speci…cs are: The bandwidth is equal to 3.0. The Intervals column refers to the number of intervals.of v.