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

n° 2003-07

Identification et Information in Monotone Binary Models

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

[1] INRA, Paris-Jourdan and CREST-INSEE, ENS – 48 boulevard Jourdan, 75014 PARIS. France.
thierry.magnac@ens.fr http://www.inra.fr/Internet/Departements/ESR/UR/lea/equipe/magnac/magnac.htm
[2] CREST-INSEE, 15 boulevard Gabriel Péri, 92245 Malakoff Cedex. France. maurin@ensae.fr

Identification & Information in Monotone Binary Models

Thierry Magnac[*] Eric Maurin[†]

First version: February 2003
Comments welcome

[*]INRA, Paris-Jourdan and CREST-INSEE, ENS - 48, boulevard Jourdan - 75014 Paris, France. Thierry.Magnac@ens.fr,
http://www.inra.fr/Internet/Departements/ESR/UR/lea/equipe/magnac/magnac.htm

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

Résumé: Soient y une variable binaire, v une variable explicative continue et x d'autres variables explicatives. On suppose que la fonction de répartition du vecteur aléatoire $w = (y, v, x)$ satisfait deux conditions: de monotonicité ($E(y \mid v, x)$ est monotone en v) et de support étendu ($E(y \mid v, x)$ varie de 0 à 1 quand v varie sur son support. Dans ce cadre, ce papier étudie les méthodes d'inférence du paramètre β dans le modèle semiparmaétrique binaire $y = 1(x\beta + v + \epsilon > 0)$. On montre que les restrictions de moment que Lewbel (2000) propose permettent la juste identification du paramètre d'interêt, β. En d'autres termes, l'absence de corrélation entre régresseurs et erreurs ($E(x'\epsilon) = 0$) et une hypothèse d'indépendance partielle ($F_\epsilon(\epsilon \mid v, x) = F_\epsilon(\epsilon \mid x)$) sont des conditions nécessaires et suffisantes pour l'identification. On montre aussi que cette méthode d'estimation atteint la borne d'efficacité semiparamétrique. Pourtant, les hypothèses d'absence de corrélation et d'indépendance partielle ne sont pas suffisantes quand le support de v n'est pas assez étendu. On propose des restrictions identifiantes naturelles pour lesquelles β demeure juste identifié. Des simulations par Monte-Carlo montrent que l'estimation est satisfaisante dans des échantillons assez petits. Des extensions aux modèles à choix discrets ordonnées est aussi proposée.

Mots-clés: Modèles binaires, Méthodes Semiparamétriques; Bornes d'Efficacité.
Classification JEL: C14, C25.

Abstract: Let, y, a binary outcome, v a continuous explanatory variable and x some other explanatory variables. Assume that the population distribution of the random variable $w = (y, v, x)$ satisfies Monotone (1) and Large Support (2) assumptions: (1) $E(y \mid v, x)$ is monotone in v and (2) $E(y \mid v, x)$ varies from 0 to 1 when v varies over its support. Within this framework, this paper studies inference on the parameters of the semiparametric binary regression model $y = 1(x\beta + v + \epsilon > 0)$. It shows that the moment restrictions that Lewbel (2000) proposed lead to exact identification of the parameter of interest, β. In other words, an uncorrelated-error restriction ($E(x'\epsilon) = 0$) combined with a partial-independance assumption ($F_\epsilon(\epsilon \mid v, x) = F_\epsilon(\epsilon \mid x)$) are sufficient and necessary for identification. We also show that Lewbel's moment estimator attains the semi-parametric efficiency bound in the set of latent models that he considers. Yet, uncorrelated-error and partial-independence assumptions are not sufficient to identify β when the support of v is not sufficiently rich. We propose intuitive additional identifying assumptions under which β remains just identified. Monte-Carlo experiments show that the estimation performs well in moderately small samples. An extension to ordered choice models is also provided.

Keywords: Binary models, Semiparametric methods, Efficiency bounds.
JEL Classification: C14, C25

1 Introduction[1]

Consider the latent binary response model,

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

where y is a binary variable, $1(A)$ is an indicator function that equals one if A is true and zero otherwise, x is a vector of covariates, v is a continuous covariate whose coefficient is set equal to 1 by convention and ϵ is an unobserved random variable. It was proved by Manski (1988) that a mean-independence restriction (i.e. $E(\epsilon \mid x, v) = 0$) is not sufficient for identifying β whatever conditions on the support of (v, x) are adopted. As a consequence, an uncorrelated-error assumption (i.e. $E(v'\varepsilon) = E(x'\varepsilon) = 0$) is not sufficient either. In contrast, Manski (1988) shows that, provided that the support of v is sufficiently rich, a quantile-independence assumption (i.e., for a given α, $\Pr(\epsilon < 0 \mid x, v) = \alpha$ for any x, v) is sufficient for identifying β. In a recent paper, however, Lewbel (2000) provides a very simple estimator of β under the combination of an uncorrelated-error restriction (i.e. $E(x'\varepsilon) = 0$), a partial independence assumption $((F_\epsilon(\epsilon \mid x, v) = F_\epsilon(\epsilon \mid x))$ and a large support assumption (i.e., for almost any x, the support of v contains the support of $-(x\beta + \epsilon)$). Under these assumptions, Lewbel considers the following transformation of the dependent variable, $\widetilde{y} = \dfrac{y - I(v > 0)}{f(v \mid x)}$, where $f(v \mid x)$ is the density of v conditional on x, and show that β can be consistently estimated by the linear regression of $\widetilde{y}$ on x. Lewbel also provides an instrumental variable version of this estimator when errors are correlated with x but uncorrelated with a set of instrumental variables z. Honoré and Lewbel (2002) extends this method to estimating binary choice panel data models with fixed effects.

All in all, the partial independence hypothesis is an appealing identifying restriction. It overcomes Manski's fundamental impossibility result and allows for the estimation of β under very general forms of endogeneity and conditional heteroskedasticity (see e.g., Maurin, 2002, for a recent empirical application). We are not aware, however, of any literature dealing with the identification of this model. In particular, it remains unclear whether the partial independence hypothesis imposes restrictions on the range of binary phenomena that may actually be analyzed through model (LV) and whether these restrictions are testable or not. In this paper, we first adress this issue and prove that the partial independence hypothesis is actually general enough to provide a semi-parametric estimator in a fairly wide class of

[1] We thank Arthur Lewbel for very helpful discussions and participants at seminars at CREST for helpful comments. The usual disclaimer applies.

binary choice models. Namely, we prove that the set of semi-parametric latent binary models satisfying the assumptions of Lewbel (2000), that is, uncorrelated errors[2], partial independence and large support for v, is one-to-one with the set of non-parametric binary models where: 1. The probability of success is monotone with respect to the continuous regressor, v. 2. The probability of success varies from 0 to 1 over the support of the continuous regressor. Under these two conditions, the unconditional moment restrictions that Lewbel derived are shown to be the only restrictions on the parameter of interest β in the latent model and therefore β is just identified.

Making identification restrictions as weak as possible is not the only concern when estimating binary choice models. The simplicity of the approach and its efficiency properties should also be taken into account. Parametric methods such as Probit or Logit are still the most commonly used methods in empirical work even if economic theory can hardly justify their assumptions. Popular semi-parametric methods are based on the properties of statistical independence (i.e., $F_\epsilon(\epsilon \mid x, v) = F_\epsilon(\epsilon)$) or of single-index sufficiency (i.e., $F_\epsilon(\epsilon \mid x, v) = F_\epsilon(\epsilon \mid x\beta + v)$). They use weaker distributional assumption than standard parametric models, but still impose strong constraints on the distribution of ϵ (see e.g. Cosslett (1983), Ruud (1983), Powell, Stock and Stoker (1989), Ichimura (1993), Klein and Spady (1993) who provide estimators of β under statistical independence or index sufficiency). The quantile-independence assumption permits much more general forms of conditional heteroskedasticity. Still, the fact remains that very few empirical studies use the corresponding maximum score estimation method, as developed by Manski (1975, 1985) or its smoothened version developed by Horowitz (1992). The numerical methods needed for optimizing the score may be one cause of underutilization, the lower than root-n rate of convergence might be another reason. Some advances have recently been proposed by Chen (2002) by stengthening the median-independence assumption into conditional symmetry and a weak restriction on conditional heteroskedasticity. Estimation can then be proved to be root-n consistent though optimisation is still needed. In contrast, Lewbel estimator can be directly obtained without optimization and is root-n consistent. The implementation of the estimation method is actually quite simple. It only requires the estimation of a conditional density (i.e., $f(v \mid x)$) and a linear regression. As far as we know, there is no result about efficiency though.

It is where this paper presents a second contribution. We prove that Lewbel's estimator

[2] For the moment, errors are supposed to be uncorrelated with explanatory variables, x. See the main text for the general case with instrumental variables.

attains the semi-parametric efficiency bound in the set of latent models under consideration. This result is derived using our equivalence result reported above though it is not straightforward since the unknown non-parametric component of $\widetilde{y}$ (i.e., $f(v \mid x)$) is also the density function with respect to which the moment restriction (i.e., $E(x'[\tilde{y} - x\beta]) = 0$) is defined, and we cannot directly use the analysis of Chamberlain (1992). We evaluate the semi-parametric efficiency bound using the formal derivation framework proposed by Severini and Tripathi (2001). Also, in the specific case where $f(v \mid x)$ is known, we prove that the estimate of β has a smaller variance when we use the estimated $\widehat{f}(v \mid x)$ rather than the true $f(v \mid x)$ to transform y. This paradoxical finding was conjectured in Lewbel (2000) and should be reminiscent of the central result of Hirano, Imbens and Ridder (2002).

These results emphasize that the partial independence assumption is definitely worthwhile considering when analyzing binary responses. Deep structural parameters are identified without imposing particularly strong restrictions on the range of phenomena that can be analyzed. As it turns out, the weakness of Lewbel's setting is not so much the partial independence assumption as such, but the accompanying assumption on the support of the continuous regressor. In general, the identification of semiparametric binary choice models is lost when the support of the continuous regressor is not rich enough (Manski, 1988) and the partial independence hypothesis does not overcome this problem. Under Lewbel's assumptions, the support of v should include the support of $-(x\beta + \epsilon)$. As already said, this assumption restricts the domain of application of this method to the analyses of phenomena such that the probability of success actually varies from 0 to 1 when v varies over its support. Analysing probabilities of entry or exit as a function of age (v) and other covariates (x) provides examples where the assumption may hold. Namely, if the support of the age variable is sufficiently large, the youngest persons should all respond 0 and the oldest, 1. In contrast, it is not possible to analyze phenomena such that at least some low-v persons and some high-v persons give the same response, which is admittedly restrictive. The analysis of fertility is one such example since, at all fecund ages, the probability of birth is never equal to 0 or to 1.

To address this issue, our paper proposes some additional intuitive identifying restrictions. One such restriction is related to the conditional symmetry of the tails of the distribution of individual propensities to respond. Even though Manski (1988) showed that conditional symmetry has no identifying power over and above median-independence, uncorrelated errors and conditional symmetry of the tails of the distribution are shown to be sufficient. For instance, under the partial independence hypothesis, the binary response under consideration may be understood as the result of the comparison of an individual propensity to respond $(x\beta+\epsilon)$ and a continuous cost function $(-v)$. The response is observed if and only if the cost is less than the propensity, which leads to the binary latent structure (see Lewbel, Linton and McFadden, 2001, for other interpretations and examples adapted to experimental design). We show that the identification of β through unconditional moment restrictions remains possible provided that we impose a symmetric distribution of the very high and very low propensities, i.e. propensities that are either so high or so low that the responses do not depend on the specific value taken by v. Necessary and sufficient conditions, though less intuitive, are also provided. These conditions provides exact identification of the parameters of interest and they are not testable.

Finally, we design Monte Carlo experiments in moderately small samples between 100 and 1000 in simple and more elaborate cases including endogeneity of covariates and/or heteroskedasticity of regressors. The sensitivity of the estimation method to the distribution of the continuous regressor v is analyzed as well as its sensitivity to the size of the support of v. Monte-Carlo results support our claim that the estimation method is worth considering even when samples are small. Lastly we explore whether our results hold for other linear latent variable models such as those considered in Lewbel (1998) or Lewbel (2000). In general the answer is no, but we find interesting exceptions. In particular, there exists an interesting set of ordered discrete choice models which, under the partial independence hypothesis, is actually one-to-one with a very general class of monotone ordered response models.

Section 2 introduces the framework and presents the equivalence result. Section 3 computes the efficiency bound and reports additional results about the variance of the estimator. In Section 4, we study (non) identification when the condition on the support is not satisfied and we provide additional sets of identifying restrictions. We report Monte Carlo experiments in Section 5. Section 6 provides extensions to ordered choices and Section 7 concludes.

2 The Set-up and the Equivalence Result

Let the "data" be given by the distribution of the following random variable[3]:

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

[3] For simplicity, we only consider random samples and we do not subscript individual observations by i.

where y is the binary variable, v is the continuous regressor, x are the "structural" explanatory variables and z are the instruments. At this point, explanatory and instrumental variables cannot be distinguished since no model has been written so far. Their respective role will be clarified below in the latent model. We first introduce some regularity conditions on the distribution of ω. They will be assumed valid in the rest of the text.

Assumption R(egularity):

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

R.ii. (Covariates & Instruments) The support of the distribution, $F_{x,z}$ of (x, z) is $S_{x,z} \subset \mathbb{R}^p \times \mathbb{R}^q$ where the p explanatory variables x can partially overlap with the $q \geq p$ instrumental variables z with no loss of generality. Variables (x, z) may be functionally dependent[4] (for instance x, x^2, $\log(x)$,...). The dimension of the set $S_{x,z}$ is $r \leq p + q$ where $p + q - r$ are the potential overlaps and functional dependencies. Finally, $rank(E(z'x)) = p$.

R.iii. (Continuous Regressor) The support of the conditional distribution of v conditional on (x, z) is $[v_L, v_H]$ almost everywhere $F_{x,z}$. Moreover, $v_L < 0 < v_H$ and v_L and v_H can be infinite. The conditional distribution is denoted $F_v(. \mid x, z)$ and is defined almost everywhere $F_{x,z}$. Furthermore, for any interval $I \subset [v_L, v_H]$ of positive Lebesgue measure, $Pr(v \in I \mid x, z) > 0$ a.e. $F_{x,z}$.

R.iv. (Functional independence of v and (x, z)) There is no subspace of $[v_L, v_H] \times S_{xz}$ of dimension strictly less than $r + 1$ which probability measure, $(F_v(. \mid x, z).F_{x,z})$, is equal to 1.

The first two assumptions define a binary model where there are p explanatory variables and q instrumental variables (assumption *R.ii*). According to 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).[5] Assumption *R.iii* defines the continuity assumption of continuous regressor v. Note that mass points are allowed because what matters is that the distribution of y given v, x and z can be defined almost everywhere in $[v_L, v_H]$. Assumption *R.iv* avoids the degenerate case where v and (x, z) are functionally dependent.

We now consider two possible formulations of the distribution of y conditional on v and (x, z) and show that they are equivalent. The first formulation is a semi-parametric latent index binary model as Lewbel (2000) and Honoré and Lewbel (2002) set it up. The second

[4] 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).

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

one is a non-parametric binary model. Let us start with the latent binary model:

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

where $1(A)$ is the indicator function that equals one if A is true and zero otherwise and $\beta \in \mathbb{R}^p$ the vector of coefficients of interest. The distribution of the random error ϵ_i satisfies the following properties as in Lewbel (2002):

Assumption L(atent) or L(ewbel)

$(L.1)$ *(Partial independence)* The conditional distribution of ϵ given covariates x and variables z is independent of the continuous regressor v:

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

The support of ε is denoted $\Omega_\varepsilon(x, z)$ and its distribution function $F_\varepsilon(. \mid x, z)$ is supposed to be absolutely continuous. Denote $f_\varepsilon(. \mid x, z)$ its density function.

$(L.2)$ *(Large support)* The support of $-x\beta - \varepsilon$ is a subset of $[v_L, v_H]$.

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

Let $\mathcal{M}_L^*$ the set of latent models which elements $(\beta,\ F_\varepsilon(. \mid x, z))$ satisfy partial independence, support and moment conditions $(L.1 - L.3)$. Using transformation (LV), we obtain an image set of conditional distributions $\Pr(y = 1 \mid v, x, z)$ that we denote:

$$\mathcal{M}_L = \mathrm{Im}_{LV}(\mathcal{M}_L^*)$$

These conditional distributions necessarily satisfy the following conditions:

Lemma 1 *Under partial independence* $(L.1)$ *and large support* $(L.2)$ *conditions, we necessarily have:*

$(NP.1)$ *(Monotonicity) The conditional probability $Pr(y_i = 1 \mid v, x, z)$ is increasing and absolutely continuous in v a.e. $F_{x,z}$.*

$(NP.2)$ *(Support) There exist (a.e. $F_{x,z}$) two values $v_l(x, z)$ and $v_h(x, z)$ (possibly infinite) in the support $[v_L, v_H]$ such that:*

$$\Pr(y_i = 1 \mid v_l, x, z) = 0 \quad \Pr(y_i = 1 \mid v_h, x, z) = 1$$

Proof. Write:

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

As $dF_\varepsilon(\epsilon \mid x, z) \geq 0$ and F_ε is absolutely continuous, the first conclusion follows.

Second, for almost any (x, z), as the support of $-x\beta - \varepsilon$ is a subset of $[v_L, v_H]$ that we denote $[v_l(x, z), v_h(x, z)]$, we have for all $\varepsilon \in \Omega_\varepsilon(x, z)$:

$$v_L \leq v_l(x, z) \leq -(x\beta + \varepsilon) \leq v_h(x, z) \leq v_H$$

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

$$v_l(x, z) + x\beta + \varepsilon \leq 0 \quad v_h(x, z) + x\beta + \varepsilon \geq 0$$

The second conclusion follows. ■

Summing up, if we denote the set:

$$\mathcal{M}_{NP} = \{F(y \mid v, x, z) \text{ satisfying monotonicity } (NP.1), \text{ and support } (NP.2) \text{ conditions}\}$$

we have just proved that $\mathcal{M}_L \subset \mathcal{M}_{NP}$. Let us prove the reciprocal, $\mathcal{M}_{NP} \subset \mathcal{M}_L$:

Lemma 2 *Let $Pr(y = 1 \mid v, x, z)$ (denoted $G(v, x, z)$) be a conditional probability satisfying monotonicity $(NP.1)$ and support $(NP.2)$ conditions. Then, there exists a unique element $(\beta, F_\varepsilon(. \mid x, z))$ in $\mathcal{M}_L^*$ such that $Pr(y = 1 \mid v, x, z)$ is its image throught the transformation (LV). In particular, parameter β is uniquely defined by the following equation:*

$$E(z'x).\beta + E(z' \int v \frac{\partial G}{\partial v} dv) = 0$$

Proof. Consider $G(v, x, z)$ satisfying $(NP.1)$ and $(NP.2)$. According to the support condition $(NP.2)$, there exists (a.e. $F_{x,z}$) two values $v_l(x, z)$ and $v_h(x, z)$ in $[v_L, v_H]$ such that $G(v_l(x, z), x, z) = 0$ and $G(v_h(x, z), x, z) = 1$. Assume that there exists $(\beta, F_\varepsilon(. \mid x, z))$ *in* $\mathcal{M}_L^*$ such that $G(v, x, z)$ is its image throught the transformation (LV). Define the support of the random variable ε as:

$$\Omega_\varepsilon^{(\beta)}(x, z) = [-(v_h(x, z) + x\beta), -(v_l(x, z) + x\beta)] \tag{1}$$

which is a subset of $[-(v_H + x\beta), -(v_L + x\beta)]$ By definition of (LV), $(\beta, F_\varepsilon(. \mid x, z))$ satisfies,

$$\begin{aligned} G(v, x, z) &= \int_{v+x\beta+\varepsilon>0, \varepsilon\in\Omega_\varepsilon^{(\beta)}(x,z)} f_\varepsilon(\varepsilon \mid x, z) d\varepsilon = \int_{-(v+x\beta)}^{-(v_l+x\beta)} f_\varepsilon(\varepsilon \mid x, z) d\varepsilon \\ &= 1 - F_\varepsilon(-(v + x\beta) \mid x, z). \end{aligned}$$

which implies for any $\varepsilon \in \Omega_{\varepsilon}^{(\beta)}(x,z)$ that:

$$f_{\varepsilon}(\varepsilon \mid x,z) = \frac{\partial G}{\partial v}(-(x\beta+\varepsilon),x,z). \tag{2}$$

The $\frac{\partial G}{\partial v}$ function is defined almost everywhere (F_v) since (a) by the monotonicity assumption ($NP1$), $G(v,x,z)$ is absolutely continuous in $v \in [v_L, v_H]$ (Billingsley, 1995) and (b) v varies continuously ($R.iii$, $R.iv$).

Furthermore, condition ($L.3$) implies:

$$\begin{aligned} E(z'\varepsilon) &= 0 \\ &= E_{x,z}(z' \int \varepsilon f_{\varepsilon}(\varepsilon \mid x,z) d\varepsilon) \\ &= -E_{x,z}(z' \int (x\beta+v)\frac{\partial G}{\partial v} dv \\ &= -E(z'x)\beta - E_{x,z}(z' \int v \frac{\partial G}{\partial v} dv) \end{aligned} \tag{3}$$

where the notation $E_{x,z}$ means that the expectation is taken with respect to the subscript variables only (if there is some ambiguity) and the integrals are taken on the support of each variable. Because of $R.iii$, $E(z'x)$ is of rank equal to the dimension of β. The previous equation therefore uniquely defines β.

Thus if $(\beta, F_{\varepsilon}(. \mid x,z))$ exists, it is defined by (1), (2) and (3). Reciprocally, consider $(\beta, F_{\varepsilon}(. \mid x,z))$ in $\mathcal{M}_L^*$ which satisfies (1), (2) and (3). Its image through (LV) is G. This completes the proof. ■

Before discussing these results, remark that the equation determining β cannot be easily used as an estimating equation because of the term $\int v \frac{\partial G}{\partial v} dv$. Some change of variables leads however to the much simpler Lewbel's estimating equation and this is proven now.

Proposition 3 *Let the Lewbel transform of y be:*

$$\tilde{y} = \frac{y - \mathbf{1}(v>0)}{f(v \mid x,z)} \tag{4}$$

then under monotonicity and support conditions, $(NP.1 - NP.2)$:

$$E_{x,z}(z' \int v \frac{\partial G}{\partial v} dv) = -E(z'\tilde{y})$$

Proof. See Lewbel (2000, page 115) or appendix A. ■

Returning to the main argument, lemmas 1 and 2 prove therefore that:

Theorem 4 *The class of latent models defined by independence, support and moment conditions* $(L.1 - L.3)$ *and transformation (LV) is one-to-one with the class of monotone and bounded support binary models defined by conditions* $(NP.1 - NP.2)$

Theorem 4 sheds some light on the deep nature of the partial independence hypothesis $(L.1)$. This theorem shows that combining $(L.1)$ with a rich support assumption such as $(L.2)$) and an uncorrelated-errors condition such as $(L.3)$ is exactly what is needed to overcome the underidentification result of Manski (1988). Adding $(L.1)$ to $(L.2)$ and $(L.3)$ provides a framework where β is just identified. Adding $(L.1)$ to $(L.3)$ only would not be sufficient as shown in section 4, while adding more than $(L.1)$ to $(L.2)$ and $(L.3)$ would generate testable overidentifying constraints.

Given this result, as far as identification is concerned, it is not straightforward to evaluate the relative merits of Lewbel's partial independence framework and Manski's quantile-independence setting, i.e. the two semi-parametric approaches that permit the most general form of dependence between error distribution and covariates in binary models.

A first argument is that Lewbel's framework accomodates endogeneous covariates while Manski's does not (at least to our knowledge). In contrast, the partial independence framework requires conditions on the support of the covariates that are stronger than the conditions required under quantile-independence. As shown by Horowitz (1998), a sufficient support condition for estimating β under quantile-independence is that, for a set of x of positive mass, $v + x\beta$ takes both positive and negative values when v varies over its support. It is weaker (and in some cases strictly weaker) than $(L2)$ which implies that $v + x\beta$ takes both positive and negative value for **any** x when v varies over its support.

The quantile-independence assumption requires a weaker support condition than partial independence but it may very well provide several alternative restrictions for identifying β and generate testable overidentifying restrictions. Assume for instance that x is discrete and belongs to $\{0, 1, ...K\}$ while the density of v is positive everywhere on the real line. Under median-independence, β is identified by looking simply to the change in sign of $(P(y = 1 \mid x = k, v) - .5)$ for any k such that $v + k\beta$ takes positive and negative value when v varies over its support. More specifically, if v_k is the value where $(P(y = 1 \mid x = k, v) - .5)$ changes sign, β is equal to $\frac{-v_k}{k}$. When $v + x\beta$ takes positive and negative values when v varies over its support for at least two values of x (say, 1 and 2) median-independence provides at least two ways of identifying β and at least one testable restriction (namely $\frac{v_2}{2} = v_1$). Such restrictions do not exist in Lewbel's setting since identification is exact.

All in all, the partial independance framework requires stronger support conditions on the one hand. On the other hand, it yields exact identification while quantile independance assumption might impose testable restrictions on the distribution of the covariates. This trade-off between assumptions on disturbances and assumptions on explanatory variables that exist when choosing between partial independence and quantile-independence is already mentioned by Manski (1988) as something peculiar to the study of identification in the discrete case.

Assuming partial independence is not the only way to overcome Manski's result. A one-to-one mapping such as the one exhibited in Theorem 4 also exists between monotone binary models and latent linear variable models when some other relationship between random shock ε and variable v is assumed. It simply happens that partial independence between ε and variable v is the simplest type of relationship that one can think of and one that can be easily implemented in experimental settings.

3 Information

Identification is not the only concern when choosing among different estimation methods, information also is. Lewbel's estimator is root-n consistent and asymptotically normal and its variance-covariance matrix is relatively easy to compute (Lewbel, 2000). We do not know much however about how precise it is with respect to other estimators. In the next section, we take some steps further and prove that the specific estimator of β proposed by Lewbel attains the semi-parametric efficiency bound and is semi-parametrically efficient. One possible source of inefficiency comes from the relationship between the unknown non-parametric component of $\widetilde{y}$ (i.e., $f(v \mid x, z)$) and the density function with respect to which the moment restriction (i.e., $E(z' [\tilde{y} - x\beta]) = 0$) is defined. Thus, the general framework investigated by Chamberlain (1992) needs to be amended and the semi-parametric efficiency of our moment estimator has to be checked by hand. Finally, we also show that it is more efficient to use an estimate of the conditional density function than its true value when it is known.

3.1 The Estimating Equation

Denote the function of interest:

$$m(y, v, x, z; \beta) = z' \left[\frac{y - \mathbf{1}(v > 0)}{f(v \mid x, z)}) - x\beta \right] = z' [\tilde{y} - x\beta],$$

since the estimate is based on the unconditional moment conditions:

$$E\left[m(y,v,x,z;\beta_0)\right]=0. \tag{5}$$

From regularity conditions (R.i-iv.),

$$E\left[mm'\right]=\Omega_0$$

is of full rank, L. It is because $E\left[mm'\right]=E(z'z.E\left[\tilde{y}-x\beta\mid z\right]^2)$ and because $E\left[\tilde{y}-x\beta\mid z\right]\neq 0$ on a set of positive measure $F_{x,z}$.

Note also that the derivative of function m with respect to β is constant and equal to:

$$-E(z'x),$$

so that the moment conditions are linear. If $f(v\mid x,z)$ were known, the semi-parametric efficiency bound for estimating solutions of unconditional moment restrictions would apply (Chamberlain, 1987). The GMM efficiency bound would be:

$$(E(x'z)\Omega_0^{-1}E(z'x))^{-1},$$

and the efficient estimate would then be obtained as usual. In our case however, the density $f(v\mid x,z)$ is unknown. Results reported by Chamberlain (1992) cannot directly be applied because the unknown non parametric component is in our case also a density function with respect to which the unconditional moment restriction is taken. The extension is however shown to hold below.

For simplicity, we shall consider an estimation in two steps. First, we begin with the estimation of parameter $\pi_0=E(z'x).\beta_0$. Second we estimate parameter β_0 using minimum distance and this estimate of π_0. In the first step, the unconditional moment restriction that we consider is:

$$E(\tilde{g}(y,v,x,z;\pi_0))=E(z'\tilde{y}-\pi_0)=0. \tag{6}$$

We derive the efficiency bound and the matrix of variance-covariance of Lewbel's estimate for π_0 in the next subsection and appendix. The efficiency bound and variance-covariance matrices for β_0 are derived next as in Newey and McFadden (1994), for instance. Namely, if V_π is the variance-covariance matrix of whatever estimate of π_0 then, under the usual regularity conditions, the variance-covariance matrix of the corresponding estimate of β_0 is given by:

$$(E(x'z).V_\pi^{-1}.E(z'x))^{-1}$$

3.2 The Semiparametric Efficiency Bound

The density function (with respect to products of Lebesgue and counting measures) of the random vector $w = (y, v, x, z)$, as defined by regularity conditions R, is rewritten as:

$$\begin{aligned} f(y, v, x, z) &= f(y \mid v, x, z).f(v \mid x, z).f(x, z) \\ &= \phi_1^2(y \mid v, x, z).\psi^2(v \mid x, z).\phi_2^2(x, z) \end{aligned}$$

to conform with the technique of Severini and Tripathi (2001) to derive efficiency bounds. The "structural" parameter of interest is π and the "reduced form" fonctionals describing the random variable are ϕ_1, ϕ_2, ψ which are assumed to belong to the following sets:

$$\Phi_1 = \{\phi_1 : \{0,1\}\times[v_L, v_H]\times S_{x,z} \to \mathbb{R}, \sum_{y=0,1} \phi_1^2(y \mid v, x, z) = 1, \phi_1^2(y \mid v, x, z) \geq 0\}$$

$$\begin{aligned} \Phi_2 = \{\phi_2 \in L^2(S_{x,z}), \quad & \int_{S_{x,z}} \phi_2^2(x, z) dx dz = 1, \phi_2^2(x, z) > 0, \\ & \phi_2^2(x, z) \text{ is bounded and continuous}\} \end{aligned}$$

$$\begin{aligned} \Psi = \{\psi : [v_L, v_H]\times S_{x,z} \to \mathbb{R}, \quad & \int_{[v_L, v_H]} \psi^2(v \mid x, z) dv = 1, \psi^2(v \mid x, z) > 0 \text{ a.e. } F_{x,z}, \\ \psi^2(v \quad \mid \quad & x, z) \text{ is bounded and continuous a.e. } F_{x,z}\} \end{aligned}$$

Let $(\dot{\phi}_1, \dot{\phi}_2, \dot{\psi})$ denote a vector in the vector space $L^2(\{0,1\}\times[v_L, v_H]\times S_{x,z})$ which is tangent, at the true value $(\phi_1^0, \phi_2^0, \psi^0)$ and corresponding π_0, to the set E of all $(\phi_1, \phi_2, \psi) \in \Phi_1 \times \Phi_2 \times \Psi$ which satisfies the unconditional moment condition given above (for all π). The tangent space is the smallest linear space which is closed in the L^2 norm and contains all such $(\dot{\phi}_1, \dot{\phi}_2, \dot{\psi})$. As shown in Severini and Tripathi (2001), under regularity conditions, these tangent spaces are the product of the following subspaces:

$$\overline{linT(\Phi_1, \phi_1^0)} = \{\dot{\phi}_1 : \{0,1\}\times[v_L, v_H]\times S_{x,z} \to \mathbb{R} : \sum_{y=0,1} \phi_1^0(y \mid v, x, z).\dot{\phi}_1 = 0\}$$

$$\overline{linT(\Phi_2, \phi_2^0)} = \{\dot{\phi}_2 \in L^2(S_{x,z}) : \int_{S_{x,z}} \phi_2^0 \dot{\phi}_2 dx dz = 0\}$$

Let an arbitrary $c \in \mathbb{R}^K$. We compute the semi-parametric efficiency bound for the parameter:

$$\rho(\phi_1^0, \phi_2^0, \psi^0) = c'\pi_0$$

that is the minimum bound for any parametric path indexed by $t \in [0, t_0]$ on the set E of interest such that:

$$\rho(\phi_{1t}, \phi_{2t}, \psi_t) = c'\pi_t$$

We later derive the efficiency bound at π_0 from this "directional" bound.

Only those $(\dot{\phi}_1, \dot{\phi}_2, \dot{\psi})$ that satisfy the differentiation of the moment conditions shall be used. The moment condition above can also be written as:

$$\sum_{y=0,1} \int_{[v_L,v_H]\times S_{x,z}} z'(y - \mathbf{1}(v > 0))\phi_{1t}^2 \phi_{2t}^2 dv dx dz - \pi_t = 0$$

What is remarkable is that this moment condition does not depend on ψ. Therefore differentiation on any path through the true value π_0 yields:

$$\sum_{y=0,1} \int_{[v_L,v_H]\times S_{x,z}} z'(y - \mathbf{1}(v > 0))2(\dot{\phi}_1 \phi_2^0 + \dot{\phi}_2 \phi_1^0)\phi_1^0 \phi_2^0 dv dx dz - \dot{\pi} = 0 \tag{7}$$

As the objective is to estimate the functional $\rho(\phi_{1t}, \phi_{2t}, \psi_t) = c'\pi_t$ the tangent vectors also have to satisfy (at $(\phi_1^0, \phi_2^0, \psi^0)$ and π_0):

$$\nabla \rho(\phi_{1t}, \phi_{2t}, \psi_t) = c'\dot{\pi}$$

Since this operator is a linear functional on the linear tangent space, Riesz theorem implies that there exists a triplet $(\phi_1^*, \phi_2^*, \psi^*)$ belonging to the tangent space such as:

$$4\sum_{y=0,1} E_{x,z,v}(\dot{\phi}_1 \phi_1^*) + 4E_{x,z}\left(\int_{[v_L,v_H]} \dot{\psi}\psi^* dv\right) + 4\left(\int_{S_{x,z}} \dot{\phi}_2 \phi_2^* dx dz\right) = c'\dot{\pi}$$

As it is valid for all $(\dot{\phi}_1, \dot{\phi}_2, \dot{\psi})$ in the tangent space, we can replace $\dot{\pi}$ by its value given in (7) and identify term by term. It is first obvious that $\psi^* = 0$. Second:

$$\begin{aligned}\phi_1^*(y \mid v, x, z) &= \frac{1}{2}c'z'\left[\frac{y - \mathbf{1}(v > 0)}{(\psi^0(v \mid x, z))^2} - E\left(\frac{y - \mathbf{1}(v > 0)}{(\psi^0(v \mid x, z))^2} \mid v, x, z\right)\right]\phi_1^0 \\ &= \frac{1}{2}c'z'(\tilde{y} - E(\tilde{y} \mid v, x, z))\phi_1^0\end{aligned}$$

where we used that ϕ_1^* and $\dot{\phi}_1$belong to the tangent space and therefore that:

$$\sum_{y=0,1} \phi_1^0(y \mid v, x, z).\phi_1^* = 0 \qquad \sum_{y=0,1} \phi_1^0(y \mid v, x, z).\dot{\phi}_1 = 0$$

in order to get the "centering" second term in the RHS. Furthermore, we have:

$$\begin{aligned}\phi_2^*(x, z) &= \frac{1}{2}c'.z'\left(\sum_{y=0,1} \int_{\mathbb{S}_v} \frac{y - \mathbf{1}(v > 0)}{(\psi^0(v \mid x, z))^2}(\phi_1^0)^2(\psi^0)^2 dv - \pi_0\right)\phi_2^0 \\ &= \frac{1}{2}c'.(z'E(\tilde{y} \mid x, z) - \pi_0)\phi_2^0\end{aligned}$$

where we used that ϕ_1^* and $\dot{\phi}_1$belong to the tangent space and therefore that:

$$\int_{S_{x,z}} \phi_2^0 \phi_2^* dxdz = 0 \qquad \int_{S_{x,z}} \phi_2^0 \dot{\phi}_2 dxdz = 0$$

in order to get the "centering" second term in the RHS.

As shown in Severini and Tripathi (2001), the efficiency bound is thus:

$$\begin{aligned} \left\|\phi_1^{*2}\right\|_F + \left\|\phi_2^{*2}\right\|_F &= c'E\left(z'(\tilde{y} - E\left(\tilde{y} \mid v, x, z\right))^2 z\right) c \\ &\quad + c'.E((z'E\left(\tilde{y} \mid x, z\right) - \pi_0)(z'E\left(\tilde{y} \mid x, z\right) - \pi_0)').c \\ &= c'E\left(z'(\tilde{y} - E\left(\tilde{y} \mid v, x, z\right))^2 z\right) c \\ &\quad + c'.E(z'(E\left(\tilde{y} \mid x, z\right) - x\beta_0)^2 z).c \\ &= c'E(z'(\tilde{y} - E\left(\tilde{y} \mid v, x, z\right) + E\left(\tilde{y} \mid x, z\right) - x\beta_0)^2 z).c \end{aligned}$$

where we used that $\pi_0 = E(z'x).\beta_0$. Thus, the semi parametric efficiency bound at π_0 is:

$$E(z'(\tilde{y} - E\left(\tilde{y} \mid v, x, z\right) + E\left(\tilde{y} \mid x, z\right) - x\beta_0)^2 z).$$

For the paper to be self-contained, we provide again in appendix B a short proof of the variance-covariance of Lewbel's estimate that was derived by Lewbel (2000) and show that it attains the previous bound.

3.3 Plugging-in the True or Estimated Conditional Density?

In this section, we assume that the conditional density $f(v \mid x, z)$ is known. It may correspond to the case where v is under experimental control or the case where one has access to additional external information on the distribution of v (through census information for instance). In such a case, we can consider both transformations, $\tilde{y} = \frac{y - I(v>0)}{f(v|x,z)}$ or $\frac{y - I(v>0)}{\widehat{f}(v|x,z)}$ when constructing the linear regression that leads to the estimation of β, where $f(v \mid x, z)$ is the true distribution and $\widehat{f}(v \mid x, z)$ an estimate of $f(v \mid x, z)$. It was conjectured in Lewbel (2000) that the estimate of β obtained with $\widetilde{y}$ and the true value of the density actually has a larger asymptotic variance than the estimate obtained with $\hat{y}$ and the estimated value of the density. We now offer a proof for this conjecture:

Theorem 5 *The estimate of π_0 defined by the unconditional moment condition (6) (i.e. $E(z'\tilde{y} - \pi_0) = 0$) has a smaller variance when the estimated $\widehat{f}(v \mid x, z)$ is used to transform the dependent variable.*

Proof. When $f(v \mid x, z)$ is unknown and estimated, Lewbel (2000) and Appendix B shows that the variance-covariance matrix of π_0 is the variance-covariance of the random variable:

$$q = z'(\tilde{y} - E(\tilde{y} \mid v, x, z) + E(\tilde{y} \mid x, z) - x\beta_0)$$

When $f(v \mid x, z)$ is known, the variance is the usual GMM variance-covariance matrix of:

$$q_0 = z'(\tilde{y} - x\beta_0)$$

Note that it is the same variable $\tilde{y}$ which is used here since we deal with asymptotics and $\widehat{f}(v \mid x, z)$ is consistent for $f(v \mid x, z)$. Denote:

$$\eta_0 = \tilde{y} - x\beta_0$$

and write:

$$q = z'(\eta_0 - E(\eta_0 \mid v, x, z) + E(\eta_0 \mid x, z))$$

Consider:

$$\eta = \eta_0 - E(\eta_0 \mid v, x, z) + E(\eta_0 \mid x, z)$$

so that we can write:

$$Vq_0 = E(z'.E((\eta_0)^2 \mid v, x, z).z)$$
$$Vq = E(z'.E((\eta)^2 \mid v, x, z).z)$$

Some algebra yields:

$$\begin{aligned}
E((\eta)^2 \quad | \quad x, z, v) &= E\left[(\eta_0 - E(\eta_0 \mid v, x, z) + E(\eta_0 \mid x, z))^2 \mid v, x, z\right] \\
&= E\left[(\eta_0)^2 + (E(\eta_0 \mid v, x, z))^2 + (E(\eta_0 \mid x, z))^2 \mid v, x, z\right] \\
&\quad -2E\left[\eta_0 E(\eta_0 \mid v, x, z) \mid v, x, z\right] + 2E\left[\eta_0 E(\eta_0 \mid x, z) \mid v, x, z\right] \\
&\quad -2E(\eta_0 \mid v, x, z)E(\eta_0 \mid x, z) \\
&= E\left[(\eta_0)^2 \mid v, x, z\right] - (E(\eta_0 \mid v, x, z))^2 + (E(\eta_0 \mid x, z))^2
\end{aligned}$$

Therefore:

$$\Delta = Vq_0 - Vq = E(z'.\left[(E(\eta_0 \mid x, z, v))^2 - (E(\eta_0 \mid x, z))^2\right].z)$$

As we can write:

$$E(\eta_0 \mid x, z, v) = E(\eta_0 \mid x, z) + \eta_1$$

where $E(\eta_1 \mid x, z) = 0$, we have:

$$E(\eta_0 \mid x, z, v)^2 = E(\eta_0 \mid x, z)^2 + (\eta_1)^2 + 2E(\eta_0 \mid x, z)\eta_1$$

and therefore:

$$\begin{aligned}
\Delta &= Vq_0 - Vq = E(z'.\left[(\eta_1)^2 + 2\eta_1 E(\eta_0 \mid x, z)\right].z) \\
&= E(z'.(\eta_1)^2.z) + 2E(z'.\eta_1 E(\eta_0 \mid x, z).z) \\
&= E(z'.(\eta_1)^2.z) + 2E(z'.E(\eta_1 \mid x, z)E(\eta_0 \mid x, z).z) \\
&= E(z'.(\eta_1)^2.z)
\end{aligned}$$

is a semi-definite positive matrix. ■

The result can be understood by using broadly similar arguments to the ones presented in Crépon, Kramarz and Trognon (1998) or the review of similar results reported in Hirano, Imbens and Ridder (2002) when one has to deal with a nuisance parameter – which is here the conditional density f – in a set of moment restrictions. In the case where the distribution is known, there is a new set of equations that are added to the original moments conditions. One can then show that the estimate based on the original set of moment equations where the nuisance parameter is replaced by a consistent estimate is equivalent to the full set of equations and is therefore efficient. In this sense, the estimate that uses the estimated conditional density is "adapted" by construction to the empirical distribution of variables in the sample while the true distribution is not.

All these results make it clear that the partial independence assumption is definitely worthwhile considering when analyzing binary responses. It makes it possible to identify and estimate efficiently deep structural parameters without imposing particularly strong restrictions on the range of phenomena that can be analyzed.

In general however, the identification of β in binary choice models is lost when the support of the regressors is not sufficiently rich. This is true when one uses the index sufficiency or the quantile-independence models (Manski 1988) and it remains true under the partial independence hypothesis. Horowitz (1998) provides some insights into the conditions that can be added to the quantile-independence assumption for β to remain identified when the support of the regressors is bounded. In the next section, we show how to complement the partial independence assumption in order to identify β when the large support assumption (L2) does not hold true.

4 Unrestricted Support and Identification

In this section, we analyze the conditions under which the partial independence assumption remains necessary and sufficient for identifying β. We maintain the monotonicity condition

NP1 but relax the large support condition NP2. The large support assumption implies that for any population of characteristics (x, z), the probability of success actually varies from 0 to 1 when v varies from v_L to v_H. It is certainly not true for all applications of interest and it is the reason why we now turn to study the case where this assumption is relaxed [6]. We have only that:

$$\Pr(y_i = 1 \mid v_L, x, z) = G(v_L, x, z) \geq 0 \quad \Pr(y_i = 1 \mid v_H, x, z) = G(v_H, x, z) \leq 1$$

This framework includes the previous one since inequalities can be both binding. Yet, if $v_H = +\infty$ and $G(v_H, x, z) < 1$ (or if $v_L = -\infty$ and $G(v_L, x, z) > 0$), then there is no latent variable model in $\mathcal{M}_L^*$ which can lead to the conditional probability function $G(v, x, z)$ since the distribution of ε would be defective. It does not seem to be easy to make econometric sense (and amenable to simple testing) of such cases. We shall consider, from now on, only non-defective conditional probability functions that can agree with the latent structure:

$$\lim_{v \to +\infty} G(v, x, z) = 1 \qquad \lim_{v \to -\infty} G(v, x, z) = 0 \tag{NP2'}$$

Note that in the case with unbounded support on the real line, $v_H = +\infty$ and $v_L = -\infty$, $(NP.2')$ implies $(NP.2)$. Cases of interest are therefore $v_L > -\infty$ or/and $v_H < +\infty$, conditions that we shall assume here.

In this section, we first show that the combination of the partial independence and the uncorrelated-error assumptions alone are not sufficient for identifying neither the distribution of the random shock ϵ nor the structural parameters β. Secondly we present two different sets of additional identifying restrictions which lead back to exact identification even when the large support condition does not hold. It is shown that the first set preserves the consistency of Lewbel's estimation procedure.

4.1 Under-Identification of the Distribution of Latent Shocks

Consider a conditional distribution $\Pr(y = 1 \mid v, x, z)$, denoted $G(v, x, z)$, satisfying the monotonicity condition $(NP.1)$ and condition $(NP.2')$. Assume that this conditional probability is the image of the latent model $(\beta, F_\varepsilon(. \mid x, z))$ which satisfies partial independence

[6] The large support assumption is quite natural in many settings though. For instance, it is the case for events that necessarily take place within a specific period of the life-cycle. When y describes such phenomena as primary-school attendance, school-leaving, leaving parental home, the entry into (or the exit from) the labor market (for male workers), age is the foremost candidate to be the special continuous regressor, v, and the large support restriction is satisfied. Young enough children have never attended primary-school and old enough children have all attended primary school for instance.

$(L.1)$ and moment condition $(L.3)$. By definition, for any v in $[v_L, v_H]$, we have:

$$G(v,x,z) = \int_{v+x\beta+\varepsilon,\varepsilon\in\Omega_\varepsilon(x,z)} f_\varepsilon(\varepsilon \mid x,z)dv$$

$$G(v,x,z)-G(v_L,x,z) = \int_{-(v+x\beta)}^{-(v_H+x\beta)} f_\varepsilon(\varepsilon \mid x,z)dv = F_\varepsilon(-(v_L+x\beta) \mid x,z)-F_\varepsilon(-(v+x\beta) \mid x,z).$$

Thus, for any ε in $[-(v_H + x\beta), -(v_L + x\beta)]$, we have necessarily,

$$f_\varepsilon(\varepsilon \mid x,z) = \frac{\partial G}{\partial v}(-(x\beta+\varepsilon),x,z). \tag{8}$$

Yet, in contrast to the bounded support case, the support of ε (conditional on x and z) is not included in $[-(v_H + x\beta), -(v_L + x\beta)]$ if $G(v_H,x,z) < 1$ or $G(v_L,x,z) > 0$. Furthermore, $f_\varepsilon(\varepsilon \mid x,z)$ has no non parametric counterpart for ε in

$$B(x) =]-\infty, -(v_H + x\beta)[\cup]-(v_L + x\beta), +\infty[.$$

The only restrictions on the distribution of ε on $B(x)$ are that:

$$\begin{aligned}\Pr\{\varepsilon &\leq -(v_H + x\beta) \mid x,z\} = 1 - G(v_H,x,z) \\ \Pr\{\varepsilon &> -(v_L + x\beta) \mid x,z\} = G(v_L,x,z)\end{aligned} \tag{9}$$

which are compatible with $(NP.2')$. As a matter of fact, all elements of the set, , $(\beta, F_\varepsilon(. \mid x,z))$, which satisfies conditions (8) and (9), generate G through (LV). This set is not empty and $F_\varepsilon(. \mid x,z)$ cannot be completely identified under the independence $(L.1)$ and moment $(L.3)$ conditions only, even if β is known. The only restrictions on $F(\varepsilon \mid x,z)$ are conditions (8) and (9) and the probability weight (9) is the only information on the distribution of ε within the off-support set $B(x)$. We are now going to show that the parameter of interest β is itself underidentified.

4.2 Under-Identification of the Parameter of Interest

If $(\beta, F_\varepsilon(. \mid x,z))$ generates G then $F_\varepsilon(. \mid x,z)$ satisfies conditions (8) and (9). The only remaining restriction on β is given by the moment condition $(L.3)$:

$$\begin{aligned}
0 &= E(z'\varepsilon) = E_{x,z}(z' \int \varepsilon dF(\varepsilon \mid x,z)) &(10)\\
&= E(z' \int_{\varepsilon\in B(x)} \varepsilon dF(\varepsilon \mid x,z)) + E(z' \int_{-(v_H+x\beta)}^{-(v_L+x\beta)} \varepsilon dF(\varepsilon \mid x,z)) &(11)\\
&= E(z' \int_{\varepsilon\in B} \varepsilon dF(\varepsilon \mid x,z)) - E(z' \int_{v_L}^{v_H} (x\beta + v)\frac{\partial G}{\partial v}dv) \\
&= E(z'\varepsilon\mathbf{1}\{\varepsilon \in B(x)\}) - E(z'x\beta \int_{v_L}^{v_H} \frac{\partial G}{\partial v}dv) - E(z' \int_{v_L}^{v_H} v\frac{\partial G}{\partial v}dv) = 0
\end{aligned}$$

The last term can be expressed as in the proof of lemma 3:

$$\begin{aligned}
\int_{v_L}^{v_H} v\frac{\partial G}{\partial v}dv &= \int_{0}^{v_H} v\frac{\partial G}{\partial v}dv + \int_{v_L}^{0} v\frac{\partial G}{\partial v}dv \\
&= [v(G(v,x,z)-1)]_0^{v_H} - \int_0^{v_H}(G(v,x,z)-1)dv \\
&\quad + [vG(v,x,z)]_{v_L}^0 - \int_{v_L}^0 G(v,x,z)dv \\
&= -\left(b(v_H,v_L,x,z) + \int_{v_L}^{v_H}(G(v,x,z)-\mathbf{1}(v>0))dv\right) \\
&= -\left(b(v_H,v_L,x,z) + E(\tilde{y} \mid x,z)\right)
\end{aligned}$$

where:

$$b(v_H,v_L,x,z) = v_H(1-G(v_H,x,z)) + v_L.G(v_L,x,z) \tag{12}$$

is a function of conditional probabilities at the bounds (and can be infinite). Note that it is equal to zero when $G(v_H,x,z)=1$ and $G(v_L,x,z)=0$ (i.e., under $NP.2$).

The moment condition given by equation (10) can then be written as:

$$\begin{aligned}
0 &= E(z'\varepsilon\mathbf{1}\{\varepsilon\in B(x)\}) - E(z'x\{G(v_H,x,z)-G(v_L,x,z)\})\beta \\
&\quad +E(z'b(v_H,v_L,x,z)) + E(z'\tilde{y}) \\
&= E(z'\varepsilon\mathbf{1}\{\varepsilon\in B(x)\}) + E(z'x\{1-G(v_H,x,z)+G(v_L,x,z)\})\beta \\
&\quad +E(z'b(v_H,v_L,x,z)) \\
&\quad -E(z'x)\beta + E(z'\tilde{y})
\end{aligned} \tag{13}$$

If the support condition $(NP.2)$ would hold, we would have $G(v_H,x,z)=1$, $G(v_L,x,z)=0$ (therefore $b(.)=0$) and $B(x)=\emptyset$. The last line of condition (13) would give Lewbel's moment condition back (i.e., $E(z'x)\beta = E(z'\tilde{y})$). Given that $(NP.2)$ does not hold, $E(z'\varepsilon\mathbf{1}\{\varepsilon\in B(x)\})$ is unknown and parameter β cannot be identified. Moreover, $E(z'\varepsilon 1\{\varepsilon\in B(x)\})$ cannot be bounded (put mass points of ε very close to $+$ or $-\infty$) and therefore, β cannot be bounded.

We summarize this result in the following proposition:

Proposition 6 *Under partial independence, moment condition, $(L1)$ and $(L3)$, parameter β is identified if and only if $E(z'\varepsilon\mathbf{1}\{\varepsilon\in B(x)\})$ is known and $E(z'b(v_H,v_L,x,z))$ is finite.*

Proof. As $G(v,x,z)$ is supposed to be known, the term $b(v_H,v_L,x,z)$ is known and finite. The only term which is not known in (13) is $E(z'\varepsilon\mathbf{1}\{\varepsilon\in B(x)\})$. ■

To understand such a condition, an interesting framework is the experimental setting developed by Lewbel, interpreting v as (minus) the unit price and $x\beta + \varepsilon$ as the willingness to pay for an object. Suppose that for every individual, random variables (x, z) are first drawn according to their distribution. In a second stage, random shock ε is drawn according to d.f. $f(\varepsilon \mid x, z)$ and variable v is drawn according to its distribution $F_v(v \mid x, z)$. Note that because of the partial independence condition $(L.1)$, the two last random drawings are independent. Whether the sequence of drawings is sequential or simultaneous does not make any difference. Suppose then that ε is drawn before v. Within this framework, $B(x)$ represents the set of ε such that the binary response is certain. It does not depend on the resolution of uncertainty over v. Namely, there are two cases for elements of $B(x)$. Either $(\varepsilon + v_H + x\beta < 0)$ and therefore for any $v \leq v_H$, $(\varepsilon + v + x\beta < 0)$ and the probability of success is equal to zero. It is the subset of "certain failure", $(B_F(x))$. Or $(\varepsilon + v_L + x\beta > 0)$ and therefore for any $v \geq v_L$, $(\varepsilon + v + x\beta > 0)$ and the probability of success is equal to one. It is the subset of "certain success", $(B_S(x))$. Our basic finding developed in the previous proposition is that we need to specify how individuals are allocated to $B(x)$ (i.e., allocated to certainty) in order to define β. It is because choices of individuals allocated to $B(x)$ are not informative about their willingness to pay, $x\beta + \varepsilon$.

In the remainder of this section, we explore two routes for specifying this allocation process and solving the identification issue. First, we propose an additional assumption on the density of ε in $B(x)$. Second, as what makes things difficult is the absence of definition of the density function of ε in $B(x)$, we propose to replace $(L.3)$ by a moment restriction bearing directly on values $\varepsilon \notin B(x)$.

4.3 Generalizing Lewbel Estimation Method

As shown above, the set $B(x)$ is the union of two "symmetrical" subsets, the subset of certain failure $(B_F(x) = \{\varepsilon : \varepsilon + v_H + x\beta < 0\})$ and the subset of certain success $(B_S(x) = \{\varepsilon : \varepsilon + v_L + x\beta > 0\})$. By construction, we have no means of identifying the distribution of propensities of success and/or failure over these two sub-sets. One of the simplest assumption we can think of is that the distribution of propensities of success within the certain-success subset $B_S(x)$ is identical to the distribution of the propensity of failure within the certain-failure subset $B_F(x)$). Such a symmetry assumption is reminiscent of the symmetrically trimmed least squares of Powell (1984). Most interestingly, this symmetry assumption is sufficient for identification. The following proposition states the necessary and sufficient assumption for exact identification of which symmetry is sufficient.

Proposition 7 *Let $v_H < +\infty$ and $v_L > -\infty$. Denote $y^*_{v_L} = (x\beta + v_L + \varepsilon)$ the propensity of success for individuals with the lowest possible v and denote $y^*_{v_H} = -(x\beta + v_H + \varepsilon)$ the propensity of failure for individuals with the highest possible v. The class of latent models defined by independence $(L.1)$, moment condition $(L.3)$ and transformation (LV) is one-to-one with the class of monotone, continuous and bounded binary models defined by monotonicity $(NP.1)$ and $(NP.2')$ and the parameter β is defined by the usual moment condition (5) if and only if:*

$$E(z'y^*_{v_H}\mathbf{1}\{y^*_{v_H} > 0\}) = E(z'y^*_{v_L}\mathbf{1}\{y^*_{v_L} > 0\}) \tag{14}$$

Proof. See appendix C. ■

In particular, once the distribution of $y^*_{v_L}$ over $B_S(x)$ is the same as the distribution of $y^*_{v_H}$ over $B_{v_L}(x)$, Lewbel estimator is unbiased. Alternatively, it is always possible to choose conditional distributions for $y^*_{v_H}$ and $y^*_{v_L}$ such that equation (14) is satisfied (choose any distribution for a positive $y^*_{v_L}$ and set $y^*_{v_H} = -y^*_{v_L}$).

If either v_H or v_L is infinite[7], this equation cannot be verified. Let $v_H = +\infty$ (say), then the absence of bias means that $E(z'y^*_{v_L}\mathbf{1}\{y^*_{v_L} > 0\})$ should be set to zero which is impossible since $Ey^*_{v_L}\mathbf{1}\{y^*_{v_L} > 0\} > 0$. Nevertheless as shown in Appendix C, the bias may affect the intercept term only.

Proposition 8 *Let $v_H = +\infty$ and $v_L > -\infty$. Denote $y^*_{v_L} = (x\beta + v_L + \varepsilon)$ the propensity of success for individuals with the lowest possible v. and denote $y^*_{v_H} = -(x\beta + v_H + \varepsilon)$ the propensity of failure for individuals with the highest possible v. The class of latent models defined by independence $(L.1)$, moment condition $(L.3)$ and transformation (LV) is one-to-one with the class of monotone, continuous and bounded binary models defined by monotonicity $(NP.1)$ and $(NP.2')$ and the parameter β is defined by the usual moment condition (5) apart from the constant term if:*

$$E(y^*_{v_L}\mathbf{1}\{y^*_{v_L} > 0\} \mid z) = \alpha$$

is a constant which is independent of z.

As for information issues, we should emphasize that Lewbel's estimator remains efficient when the large support hypothesis $(L.2)$ is replaced by a symmetry assumption such as

[7] but not both. If both K and L are infinite, we are back to the case described as restricted support (!), condition $(L.2)$. Theorem 4 applies.

condition (14). As a matter of fact, the derivation of the semi-parametric efficiency bound and of the variance-covariance of estimator do not depend on the specific assumptions made on bounds. Whether conditions $(L.2)$ or $(NP.2)$ are satisfied or not, the same properties apply to Lewbel's estimate. It is consistent and semi-parametrically efficient under the conditions of Propositions 7 or 8.

If one is ready to lose some efficiency then in the asymmetric case described by Proposition 8, one can always use the symmetrical trimming proposed by Powell (1986).

4.4 A Conditional Moment Restriction

When ε belongs to the certainty set $B(x)$, variations in v have no empirical counterpart and there is no way for defining the conditional distribution of ε or the propensity of success. In this specific sense, the certainty set has no possible real-world counterpart. Given that, it may seem more natural to impose restrictions on conditional moments relative to ε and z, conditional on its being within the region of interest (and outside $B(x)$) rather than restrictions on conditional distribution. In particular, once $B(x)$ is non-empty, it may seem preferable to assume:

$$E(z'\varepsilon \mid \varepsilon \notin B(x)) = 0, \tag{L3'}$$

rather than moment condition $(L.3)$. This modification of Lewbel's identifying restriction makes identification possible at the cost of changing the definition of the parameter of interest. Lewbel's basic estimator is not valid anymore, however.

Proposition 9 *The class of latent models defined by independence* $(L.1)$ *and the conditional moment* $(L.3')$ *restrictions and transformation (LV) is one-to-one with the class of monotone binary models defined by* $(NP1)$ *and* $(NP2')$

Proof: See appendix D.

The definition of β however changed. Let:

$$\breve{y} = \frac{\tilde{y} + b(v_H, v_L, x, z)}{G(v_H, x, z) - G(v_L, x, z)}$$

where $\tilde{y}$ is the Lewbel transform of variable y as defined by (4) and where $b(v_H, v_L, x, z)$ is given by equation (12). Then parameter β is defined by the moment restriction:

$$E(z'x).\beta = E(z'\breve{y})$$

5 Monte-Carlo experiments

We present results concerning two Monte Carlo experiments and show that the estimators developed in the case where the large support assumption $(L.2)$ is not satisfied perform well in medium-sized samples. The first experiment corresponds to the simplest model we can think of and we call it the simple experiment. In the second and more sophisticated experiment, we can accomodate endogeneity and heteroskedasticity issues and we call it the complete experiment.

5.1 First Monte Carlo experiment

Consider the simple experiment where the scalar random variable x can take two values 0 and +1 and where $0 < \beta < 1$. Assume also that the coefficient of the constant term $\alpha = 0$ and that the support of v is given by $v_L = -1$ and $v_H = +1$. Let the support of ε be:

$$\Omega_\varepsilon = [-2, 2],$$

and the distribution of ε be piece-wise uniform with $f(\varepsilon \mid x)$ taking constant values over the intervals $[-2,-1[$, $[-1,0[$, $[0,1[$, $[1,2]$. In Appendix E.1 we show the restrictions that has to be imposed on these values in order to impose the uncorrelated-errors condition and the no-bias restriction (i.e., conditions 5 and 14).

Results of the Monte-Carlo experiments are reported in Tables 1 to 4. These experiments were replicated 200 times, the number over which averages and standard errors of estimates remain stable. For each experiment, we report the bias, the empirical standard error, the root mean square error and the mean absolute error. The number of observations varies between 100 and 1000. Other details are reported in Appendix E.1.

In Table 1, we evaluate the influence of the distribution of the continuous regressor v. In the simulations, we consider 4 types of distributions, uniform, Logit based, triangular and assymetric as explained in the Details section of Appendix E.1 but we do not use this information at the estimation stage.

For the uniform, biases are extremely small and the root mean square error is virtually equal to the variance. Biases are more important for the other distributions in small samples of 100 from 8% for the triangular to more than 100% for the assymetric distribution that we use. Yet the part of biases in the RMSE is small and biases become very small when the sample size reaches 200 observations. The reported mean absolute error shows that nothing particular happens in the tails of the distribution except when the sample size is small (100).

The Logit case gives average results and it will consequently be the distribution that we use in the following Tables.

In Table 2, we evaluate the impact of the method used for estimating the conditional density of v. We estimate the model using differing window sizes around the optimized value as suggested by Lewbel (2000) and which construction is explained in Appendix E.3. For every sample size, we make the window size vary between half and twice the optimal window. Optimization does not seem to work well in the current experiments since the RMSE of the estimates is never minimal for this choice of window size. For all sample sizes, the larger the window is, the smaller the RMSE. It might only translate that the optimization is performed over a constant term only while there is an explanatory variable in the experiment reported here. Yet a method of minimization of the RMSE might be advisable in this context.

In Table 3, we vary the average frequency of observations such that the probability of success is certain over all the support of v (the set $B_s(x)$ as described in the previous sections) and the average frequency of observations such that the probability of failure is certain over all the support of v (the set $B_f(x)$ as described in the previous sections). As before, biases are small and do not vary systematically with the average frequencies of "perfect" success or failure. Yet, these observations where the issue can be perfectly predicted in the support of v, do not contribute to the information about the parameter of interest and this can be seen in the increases of RMSEs with the increase in frequencies for almost all sample sizes. All these comments remain valid when we vary the value of the coefficient of interest β (Table 4).

5.2 Second Monte Carlo experiment

Consider β a vector of parameters. The sample that we construct consists in a variable v, in a scalar variable x and scalar instrument z where:

$$x = z\gamma_z + \varepsilon_x$$

Consider two random disturbances. The first one is constructed from:

$$\varepsilon_0 = \varepsilon_x.\alpha.\gamma_0 + \exp(\gamma_s + x\gamma_x).\alpha_0$$

where α is distributed uniformly over $[0, 1]$ and α_0 is distributed uniformly over $[-0.5, 0.5]$. Parameters $(\gamma_0, \gamma_s, \gamma_x, \gamma_z)$ index the Monte Carlo experiments. A second random drawing ε_1 is shown below to be less important because it only affects the tails of the distribution. We will consider that ε_1 is the absolute value of a zero mean and unit variance normal variate. It is therefore a positive random variable

Consider the following device to construct the final disturbance ε in the Monte Carlo simulations. Fix $\delta \in \mathbb{R}^L$:

$$\begin{array}{rl} \varepsilon = \varepsilon_0 + z\delta & \text{if } -(x\beta + v_H) < \varepsilon_0 + z\delta \leq -(x\beta + v_L) \\ \varepsilon = -(x\beta + v_H) - \varepsilon_1 & \varepsilon_0 + z\delta \leq -(x\beta + v_H) \\ \varepsilon = -(x\beta + v_L) + \varepsilon_1 & \varepsilon_0 + z\delta > -(x\beta + v_L) \end{array}$$

In terms of the model, the last two regimes respectively describe the certain failure set, $B_F(x) = \{\varepsilon \leq -(x\beta + v_H)\}$, and the certain success set, $B_S(x) = \{\varepsilon > -(x\beta + v_L)\}$. As these two last regimes will lead to a simulation $y_s = 0$ or $y_s = 1$ it is why the distribution of ε_1 is less important provided that the unbiasedness condition (14) is satisfied. Namely the simulations are given by:

$$y_s = \mathbf{1}\{v + x\beta + \varepsilon > 0\}$$

The random term ε should verify the conditions of the model that are the moment restriction (5) and the unbiasedness condition (14). In Appendix E.2 we explain how to find a value of δ such that these conditions hold.

Results of the Monte-Carlo experiments are reported in Tables 5 to 8. These experiments were replicated 200 times as in the simple case. Even if an intercept is present, we only report the OLS and IV estimates of the coefficient β of the explanatory variable x. Other details are explained in Appendix E.2.

In Table 5, we report results when the explanatory variable is exogenous and there is no heteroskedasticity. The OLS estimate behaves very well and differences across different distributions for the continuous regressor v are much smaller than in the simple experiment. The largest bias when the sample size is 100 is equal to 20% for the triangular distribution. The reported mean absolute errors show that the tails of the distribution are thin and that the RMSE is hardly explained by outliers. As expected, almost all of the RMSE is due to the variance of the estimates. Also as expected, the bias of the 2SLS estimate is much larger, can reach 40% for 100 observations and remains significant for samples of 1000 (around 30% for the triangular). The part explained by the bias in the RMSE can be almost equal to the part of the variance for 100 observations.

In Table 6 we make the degree of endogeneity vary through coefficient γ_0. We adopt the Logit distribution for v as in the simple experiment. As expected, the bias of OLS increases when the degree of endogeneity increases, whatever the number of observations. In contrast, the 2SLS estimate performs better when the degree of endogeneity increases and the bias is generally lower by a factor equal to 1/3. Overall, these tables confirm that this estimation method performs well even when potential specification errors are present. In Tables 7 we introduce heteroskedasticity. As expected, the bias of the OLS estimate is quite insensitive to the presence of heteroskedasticity though it tends to slightly increase in absolute value in large samples. It is the reverse property that holds for the 2SLS estimate though, as they are opposite in signs, both biases tend to be lower in relative value. More surprisingly, the RMSE is very insensitive to heteroskedasticity which may mean that the degree of it is not "sufficient". Finally, we combine both endogeneity and heteroskedasticity specification problems in Table 8. Results are mainly resulting from the sum of these two problems and interactions are weak.

6 Extensions

Lewbel (1998) and Lewbel (2000) use the continuous regressor hypothesis to estimate the structural parameters of other linear latent variable models, $y = L(x\beta + \epsilon)$, such as the ordered discrete choice model with constant thresholds or the censored regression model. One obvious issue is whether the equivalence results given by Theorem 4 can be extended to these models. In some interesting cases the answer is positive. In other cases, the continuous regressor hypothesis imposes testable restrictions on the set of monotone statistical phenomena that may be generated by the latent structure.

To illustrate the generalization of Theorem 4, we consider the most straighforward extension of binary responses which are ordered choice models. Assume that the support of y is now $S_y = \{0, 1, ..., K\}$ $(K \geq 1)$. The discussion will be split into two according to two definitions of ordered choice models. In the first one, each individual is defined by an ordered set of propensities (i.e., $y_1^*, ., y_K^*$) and his/her response ($y \in \{0, 1, ..., K\}$) depend on how propensities compare with a given cost variable v. In the second model, each individual is defined by one specific propensity y^* and his/her response depends on how this propensity compares with an ordered set of thresholds $\alpha_k(v)$.We are going to show that a straightforward extension of theorem 4 only holds in the first case. In the second model, structural parameters are overidentified.

6.1 Ordered Choices: First Model

Let us consider the following definition for latent ordered choice models.

Definition 10 *Latent ordered discrete choice models are characterized by a set of ordered latent random variables* $\{y_1^*, ., y_K^*\}$ *where* $y_k^* > y_{k+1}^*$. *By convention define* $y_{K+1}^* = -\infty$. *The observable model is given by:*

$$
\begin{aligned}
y &= \sum_{k=1}^{K} I(v + y_k^* > 0, v + y_{k+1}^* \leq 0), \qquad \text{(LV1)} \\
&= \sum_{k=1}^{K} I(-y_k^* < v \leq -y_{k+1}^*).
\end{aligned}
$$

To conform with the binary model, we only consider linear latent models such as:

$$\forall k = 1, ., K \qquad y_k^* = x\beta_k + \varepsilon_k,$$

and where every random shock $\varepsilon_1, ., \varepsilon_K$ satisfy $(L.1 - L.3)$.

This model is also a direct generalization of (LV). When $K = 1$, the two models coincide. Such an ordered choice model may typically be used for analyzing the schooling choices that are made at the end of compulsory school where (a) y is the number of year of post-compulsory education, (b) v is the finanacial cost of each year of post-compulsory schooling while (c) y_k^* is the marginal net return to the $k - th$ year of post-compulsory education (a plausibly decreasing function of k and which plausibly depends on preferences and social background).

Another possible application is when the observed variable y records the number of units of a good that is bought by consumer i when the offered unit price is $(-v)$, the latent variables, y_k^*, stands for the average willingness to pay for a unit of this good when the number of units bought is k. If marginal utility is decreasing, then the unit willingness to pay is decreasing which justifies the ordered choice setting. The fact that an entire array of unobserved components affect willingness of pay is due to individual differences in the relations between marginal utility and quantity purchased.

One of the nice consequence of the setting given by $(LV1)$ is that it is equivalent to a system of K binary latent models given by:

$$y_k = I(-y_k^* < v), \qquad \text{(LV1k)}$$

For instance, y_1, is the indicator of purchase (any quantity), y_2 is the indicator of 2 or more units purchased and so on y_k is the indicator that more than k units were purchased:

$$y_k = I(y \geq k).$$

Reciprocally:

$$y = \sum_{k=1}^{K} y_k.$$

Within this framework, let $\mathcal{M}_{OC}^*$ be the set of latent ordered discrete choice models which elements $(\beta_k,\ F_{\varepsilon_k}(. \mid x, z), k = 1, ..., K)$ satisfy partial independence, support and moment conditions $(L.1 - L.3)$ and the additional inequality restrictions across alternatives:

$$y_k^* = x\beta_k + \varepsilon_k > x\beta_{k+1} + \varepsilon_{k+1} = y_{k+1}^* \tag{15}$$

Those inequalities do not translate into restrictions on the marginal distributions of ε_k but only on the joint distribution of $(\varepsilon_k, \varepsilon_{k+1})$. Let $\Omega_k(x, z)$ be the support of ε_k as defined in the first section. The support of $(\varepsilon_1, ., \varepsilon_K)$ is therefore:

$$\Omega(\beta, x, z) = \{(\varepsilon_1, ., \varepsilon_K) \in \Omega_1 \times .. \times \Omega_K \mid \forall k; x\beta_k + \varepsilon_k > x\beta_{k+1} + \varepsilon_{k+1}\}$$

The consequences in terms of non-parametric predictions are now straightforward. They consist in $(NP.1)$ and $(NP.2)$ for any choice k. Inequalities (15) in the latent model translate into:

$$y_k = \mathbf{1}\{-(x\beta_k + \varepsilon_k) < v\} \geq \mathbf{1}\{-(x\beta_{k+1} + \varepsilon_{k+1}) < v\} = y_{k+1}$$

with some strict inequalities for a positive mass of v. Thus:

$$E(y_k \mid v, x, z) = G_k(v, x, z) > G_{k+1}(v, x, z) = E(y_{k+1} \mid v, x, z)$$

which is a sensible assumption in most cases. For instance, the probability of buying more than k units is decreasing with k. Those inequalities do not translate into restrictions on the marginal distributions of ε_k but only on the joint distribution of $(\varepsilon_k, \varepsilon_{k+1})$ and their joint distribution is underidentified. Only the marginals are.

We can now summarize these results. Let the set $\mathcal{M}_{LOC}^*$ of latent ordered models be given by parameters $(\beta_1, ., \beta_K) \in \mathbb{R}^K$, distribution functions $(f_1(\varepsilon_1 \mid x, z), ., f_K(\varepsilon_K \mid x, z)) \in \mathcal{D}^K$, a family of set $\Omega(\beta, x, z) \subset \mathbb{R}^K$, and the transformation (LV1) such that they verify $(L.1-L.3)$. Let the set $\mathcal{M}_{NPOC}$ given by:

$$\mathcal{M}_{NPOC} = \mathcal{M}_{NP}(y_1) \times .. \times \mathcal{M}_{NP}(y_K)$$

that satisfy $(NP.1)$ and $(NP.2)$ and where $\forall k; G_k(v, x, z) > G_{k+1}(v, x, z)$. Then:

Theorem 11 *$\mathcal{M}_{LOC}^*$ is one-to-one with $\mathcal{M}_{NPOC}$.*

6.2 Ordered Choices: Second Model

Let us now consider the following semi-parametric latent model which is defined with respect to the unobserved heterogeneity component:

$$y = \sum_{k=1}^{K} I(\alpha_k(v) < x\beta + \epsilon \leq \alpha_{k+1}(v)), \tag{LV2}$$

where the thresholds $\alpha_k(v)$, $k = 1, ..., K+1$, satisfy,

$$\alpha_1(v) = -v \leq \alpha_2(v) \leq ... \prec \alpha_k(v) \leq \alpha_{K+1}(v) = +\infty, \tag{16}$$

while ϵ satisfies $(L.1)$, $(L.2)$ and $(L.3)$.

This model also is a direct generalization of (LV). When $K = 1$, the two models coincide. The $x\beta + \epsilon$ may be interpreted as a propensity to respond as in (LV), but the response has now several possible levels of intensity. The $\alpha_k(v)$ thresholds may be interpreted as the cost of responding with intensity k. The only structural assumption about these costs is that they increase with the response's intensity.

Such a model may describe for instance the performance of young children when starting school where y^*represents their (latent) schooling ability (plausibly dependent on family inputs) and the $\alpha_k(v)$ the set of thresholds (plausibly dependent on v being the birthdate within the year) imposed by the educational system for deciding who has to be held back $(y = 0)$, who has to be on time $(y = 1)$ and who has to be ahead $(y = 2)$ at school[8].

Let $\mathcal{M}^*_{LOC2}$ be the set of latent ordered discrete choice models which elements $(\beta,\ F_\varepsilon(. \mid x, z), \alpha_k(v), k = 2, ..., K)$ satisfy independence, support and moment conditions $(L.1 - L.3)$. Consider also a statistical model $F(y \mid v, x, z)$ on S_y such that $Pr(y_i \geq 1 \mid v, x, z)$ satisfy conditions $(NP.1 - NP.2)$ and assume that there exists a latent ordered choice model $(\beta,$ $F_\varepsilon(. \mid x, z), \alpha_k(v), k = 2, ..., K)$ in $\mathcal{M}^*_{LOC2}$ which image is $F(y \mid v, x, z)$.

Let us denote $G_0(v, x, z) = P(y = 0 \mid v, x, z)$. By definition, $-G_0$ belongs in $\mathcal{M}^*_{NP}$. Thus, using Theorem 4, we can exactly identify the parameter of interest β and the distribution of errors ε. In particular, we necessarily have $f_\varepsilon(. \mid x, z) = \frac{\partial G_0}{\partial v}(-(x\beta + \varepsilon), x, z)$.For any $k \geq 1$, define now $G_k(v, x, z) = P(y \leq k \mid v, x, z)$. We have,

[8] Maurin (2002) uses the binary approach to estimate the probability to be held back using v = date of birth within the year as a special regressor and interpreting $x\beta + \epsilon$ as schooling abilities. $-\alpha_2(v)$ can be interpreted as the ability threshold (defined by the educational system) above which children can be ahead at school.

$$
\begin{aligned}
G_k(v,x,z) &= \int_{-\infty}^{-x\beta+\alpha_k(v)} dF(\varepsilon \mid x,z) = \int_{-\infty}^{-x\beta+\alpha_k(v)} \frac{-\partial G_0}{\partial v}(-(x\beta+\varepsilon),x,z)d\varepsilon \\
&= G_0(-\alpha_k(v),x,z).
\end{aligned}
$$

It therefore yields:

$$
\alpha_k(v) = -G_0^{-1}(.,x,z) \circ G_k(v,x,z).
$$

Thus $F(y \mid v,x,z)$ is the image of an element of $\mathcal{M}^*_{LOC2}$ only if $G_0^{-1}(.,x,z) \circ G_k(v,x,z)$ do not depend on x and z. Put differently, a monotone ordered discrete phenomena can be analyzed as a structural ordered choice model that satisfies the partial independence hypothesis only under the testable assumption according which $G_0^{-1} \circ G_k$ do not depend on x and z. Note finally that inequalities described by (16) translate into the same inequalities on functions G_k that we had in the previous subsection and which are adapted to the present setting. They do not affect our argument.

Therefore, the ordered discrete choice models with fixed thesholds (i.e., $\alpha_k(v) - \alpha_0(v) = \gamma_k$) considered in Lewbel (2000) or in Lewbel (1998) are *not* one-to-one with the monotone discrete models. The partial independence hypothesis makes it possible to identify very easily the structural parameters that characterize these ordered choice models, but this assumption also implies (testable) restrictions on the set of discrete monotone phenomena that can be analyzed with such models.

7 Conclusion

In this paper, we show that the uncorrelated-error, partial independence and large-support assumptions proposed by Lewbel (2000) are necessary and sufficient for identifying the structural parameters of a very general class of monotone binary response phenomena. Furthermore, we show that Lewbel's moment estimator attains the semi-parametric efficiency bound in the corresponding class of latent models. We also show that the large support assumption -which might be unadapted in some instances- can be amended using additional credible restrictions such as conditional symmetry of the tails of a distribution. We also report Monte Carlo experiments which work well and propose an extension to ordered choice models.

It would be interesting to extend our results to other settings, such as the analyses of selection models (Kahn and Lewbel, 2002), treatment effects (Lewbel, 2002) or panel data (Honoré and Lewbel, 2002). We are currently exploring another route by relaxing the assumption that partial independence holds with respect to the regressor which is continuous. We consider that v is discrete or has been discretized and show that bounds of an interval containing β are identified in this case.

REFERENCES

Billingsley, P., 1995, *Probability and Measure*, Wiley: New York.

Chamberlain, G., 1987, "Asymptotic Efficiency with Conditional Moment Restrictions", *Journal of Econometrics,* 34:305-34.

Chamberlain, G., 1992, "Efficiency Bounds for Semiparametric Regression", *Econometrica, 60:567-96.*

Chen, S., 2002, "Semiparametric Estimation of a Heteroscedastic Binary Choice Model", working paper, Hong Kong University.

Coslett, S.R., 1983, "Distribution-Free Maximum Likelihood Estimator of the Binary Choice Model", *Econometrica*, 51:765-782.

Crépon, B., F., Kramarz and A.,Trognon, 1998, "Parameters of Interest, Nuisance Parameters and Orthogonality Conditions: an Application to Autoregressive Error Component Models, *Journal of Econometrics*, 82:135-156.

Hirano, K., G.W., Imbens and G., Ridder, 2002, "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score", forthcoming *Econometrica.*

Honoré, B., and A., Lewbel, 2002, "Semiparametric Binary Choice Panel Data Models without Strict Exogeneity", *Econometrica*, 70:2053-2063.

Horowitz, J., 1992, "A Smoothed Maximum Score Estimator for the Binary Response Model", *Econometrica*, 60:505-31

Horowitz, J., 1998, *Semiparametric methods in Econometrics*, Springer: Berlin.

Ichimura, H., 1993, "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Model", *Journal of Econometrics*, 58:71-120.

Klein, R.W. and R.H. Spady, 1993, "An Efficient Semiparametric Estimator for Binary Response Models", *Econometrica*, 61:387-421.

Lewbel, A., 1998, "Semiparametric Latent Variable Model Estimation with Endogenous or Mismeasured Regressors", *Econometrica*, 66:105-21.

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

Lewbel, A., O., Linton and D., McFadden, 2001, "Estimating Features of a Distribution from Binomial Data", Working Paper.

Manski, C.F., 1975, "Maximum Score Estimation of the Stochastic Utility Model of Choice", *Journal of Econometrics*, 3:205-28.

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, "Identification of Binary Response Models", *Journal of the American Statistical Association,* 83:729-738.

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.

Newey, W.K., 1994, "The Asymptotic Variance of Semiparametric Estimators", *Econometrica, 62:1349-82.*

Newey, W.K., and D., MacFadden, 1994, "Large Sample Estimation", *Handbook of Econometrics*, 4:

Powell, J., 1986, "Symmetrically Trimmed Least Squares Estimation for Tobit Models", *Econometrica, 54:1435-60.*

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

Powell, J.L., J.H., Stock and T.M, Stoker, 1989, "Semi-parametric Estimation of Index Coefficients", *Econometrica*, 57:1403-30.

Ruud, P.A., 1983, "Sufficient Conditions for the Consistency of Maximum Likelihood Estimation Despite Misspecification of Distribution in Multinomial Discrete Choice Models", *Econometrica,* 51:225-228.

Severini, T.A. and G., Tripathi, 2001, "A Simplified Approach to Computing Efficiency Bounds in Semiparametric Models", *Journal of Econometrics,* 102:23-66..

Vytlacil, E., 2002, "Independence, Monotonicity, and Latent Index Models: An Equivalence Result", *Econometrica*, 70:331-41.

Appendices

A Proof of Lemma 3

$$
\begin{aligned}
\int v\frac{\partial G}{\partial v}dv &= \int_0^{v_H} v\frac{\partial G}{\partial v}dv + \int_{v_L}^{0} v\frac{\partial G}{\partial v}dv \\
&= [v(G(v,x,z)-1)]_0^{v_H} - \int_0^{v_H}(G(v,x,z)-1)dv \\
&\quad + [vG(v,x,z)]_{v_L}^{0} - \int_{v_L}^{0} G(v,x,z)dv \\
&= -\int_{v_L}^{v_H}(G(v,x,z)-\mathbf{1}(v>0))dv \\
&= -\int_{v_L}^{v_H}(E(y\mid v,x,z)-\mathbf{1}(v>0))dv \\
&= -\int_{v_L}^{v_H} E(\tilde{y}\mid v,x,z).dF_v(v\mid x,z) = -E(\tilde{y}\mid x,z)
\end{aligned}
$$

and the proof follows.■

B The variance-covariance of Lewbel estimate

As in Newey (1994), consider the estimation of the parameter of interest $\pi_t = E(z'\tilde{y})$ on any differentiable path indexed by t and where $t=0$ gives π_0. For simplicity, denote u is the functionally independent representation of (x,z):

$$
\pi_t = \int z'\frac{y-\mathbf{1}\{v>0\}}{f_t(v\mid u)}f_t(\varepsilon,v,u)d\varepsilon dv du
$$

Therefore:

$$
\pi_t = \int z'(y-\mathbf{1}\{v>0\})f_t(\varepsilon\mid v,u)f_t(u)d\varepsilon dv du
$$

Under regularity conditions given by Newey (1994), formal differentiation with respect to t yields:

$$
\begin{aligned}
\left.\frac{\partial \pi_t}{\partial t}\right|_{t=0} &= \int z'(y-\mathbf{1}\{v>0\})\frac{\partial}{\partial t}(f_t(\varepsilon\mid v,x,z)f_t(x,z))d\varepsilon dv du \\
&= \int z'(y-\mathbf{1}\{v>0\})\left(\frac{\partial}{\partial t}\ln f_t(\varepsilon\mid v,u) + \frac{\partial}{\partial t}\ln f_t(u)\right) f_0(\varepsilon\mid v,u)f_0(u)d\varepsilon dv du
\end{aligned}
$$

$$
\begin{aligned}
\left.\frac{\partial \pi_t}{\partial t}\right|_{t=0} &= E\left[z'\frac{y-\mathbf{1}\{v>0\}}{f_0(v\mid u)}.\left(\frac{\partial}{\partial t}\ln f_t(\varepsilon\mid v,u)+\frac{\partial}{\partial t}\ln f_t(u)\right)\right] \\
&= E\left[z'\tilde{y}.\left(\frac{\partial}{\partial t}\ln f_t(\varepsilon,v,u)-\frac{\partial}{\partial t}\ln f_t(v,u)+\frac{\partial}{\partial t}\ln f_t(u)\right)\right] \\
&= E\left[z'\tilde{y}.S(\varepsilon,v,u)\right]-E\left[z'\tilde{y}.\frac{\partial}{\partial t}\ln f_t(v,u)\right]+E\left[z'\tilde{y}.\frac{\partial}{\partial t}\ln f_t(u)\right] \\
&= E\left[z'\tilde{y}.S(\varepsilon,v,u)\right]-E\left[z'E(\tilde{y}\mid v,u).S(v,u)\right]+E\left[z'E(\tilde{y}\mid u).S(u)\right]
\end{aligned}
$$

where $S(\varepsilon,v,u)=\frac{\partial}{\partial t}\ln f_t(\varepsilon,v,u)$ is the score of the model evaluated at the true value (respectively $S(v,u)=\frac{\partial}{\partial t}\ln f_t(v,u)$ and $S(u)=\frac{\partial}{\partial t}\ln f_t(u)$. As for any function $\phi(v,u):$

$$E(\phi(v,u)S(v,u))=E(\phi(v,u)S(\varepsilon,v,u))$$

we therefore have:

$$\left.\frac{\partial \pi_t}{\partial t}\right|_{t=0}=E\left[z'(\tilde{y}-E(\tilde{y}\mid v,u)+E(\tilde{y}\mid u)).S(\varepsilon,v,u)\right]$$

and the variance covariance of $\hat{\pi}$ is the variance of q:

$$q=z'(\tilde{y}-E(\tilde{y}\mid v,u)+E(\tilde{y}\mid u)-x\beta_0)$$

since $Eq=0$ and where we used that $\pi_0=E(z'x)\beta_0$ ■

C Proof of Proposition 7

Equation (13) proves that Lewbel's estimator is biased except if:

$$
\begin{aligned}
E(z'\varepsilon\mathbf{1}\{\varepsilon\in B(x)\})+E(z'x\{1-G(v_H,x,z)+G(v_L,x,z)\})\beta & \\
+E(z'b(v_H,v_L,x,z)) &= 0
\end{aligned}
$$

$\Longleftrightarrow$

$$
\begin{aligned}
E(z'\varepsilon\mathbf{1}\{\varepsilon<-(v_H+x\beta)\})+E(z'\varepsilon\mathbf{1}\{\varepsilon>-(v_L+x\beta)\}) & \\
E(z'(x\beta+v_H)\{1-G(v_H,x,z)\}+E(z'(x\beta+v_L)\{G(v_L,x,z)\} &= 0
\end{aligned}
$$

$\Longleftrightarrow$

$$E(z'(x\beta+v_H+\varepsilon)\mathbf{1}\{\varepsilon<-(v_H+x\beta)\})+E(z'(x\beta+v_L+\varepsilon)\mathbf{1}\{\varepsilon>-(v_L+x\beta)\})=0$$

which is equivalent to:

$$-E(z'y^*_{v_H}\mathbf{1}\{y^*_{v_H}>0\})+E(z'y^*_{v_L}\mathbf{1}\{y^*_{v_L}>0\})=0$$

where $y^*_{v_L}=-(x\beta+v_H+\varepsilon)$ and $y^*_{v_L}=x\beta+v_L+\varepsilon$.

If $v_H=+\infty$ and using the support condition $(NP.2')$, the bias is characterized by the quantity:

$$E(z'y^*_{v_L}\mathbf{1}\{y^*_{v_L}>0\})$$

If the conditional mean is independent of $z:$

$$E(y^*_{v_L}\mathbf{1}\{y^*_{v_L}>0\}\mid z)=\alpha$$

then the constant only in β is biased. ■

D Proof of Proposition 9

We replace the unconditional moment restriction given by L3, by the following conditional moment restriction:

$$\begin{aligned} E(z'\varepsilon \quad &| \quad \varepsilon \notin B(x)) = 0 \\ &\iff E(z'\varepsilon \mid v_H + x\beta + \varepsilon > 0, v_L + x\beta + \varepsilon < 0) = 0 \end{aligned}$$

The conditioning event means that we are only considering unobserved heterogeneity terms such that there exists observable values (v, x) such that $v + x\beta + \varepsilon = 0$. Similar developments to equation (10) lead to:

$$-E(z'x)\beta + E(z'\frac{\tilde{y} + b(v_H, v_L, x, z)}{G(v_H, x, z) - G(v_L, x, z)}) = 0$$

If we define:

$$\breve{y} = \frac{\tilde{y} + b(v_H, v_L, x, z)}{G(v_H, x, z) - G(v_L, x, z)}$$

estimates of β can be obtained by regressing $\breve{y}$ on x using instruments z.

There is an alternative form to $\breve{y}$. It is obtained by noting that:

$$\int_{v_L}^{v_H} (G(v, x, z) - \mathbf{1}(v > 0))dv = b + \int_{v_L}^{v_H} (G(v, x, z) - \mathbf{1}(v > b))dv$$

Therefore if:

$$\hat{y} = \frac{y - \mathbf{1}\{v > -b(v_H, v_L, x, z)\}}{f(v \mid x)(G(v_H, x, z) - G(v_L, x, z))}$$

then:

$$Ez'\hat{y} = Ez'\breve{y}$$

■

E Details of the Monte Carlo experiments

E.1 First Monte-Carlo experiment

Let the support of ε be:

$$\Omega_\varepsilon = [-2, 2],$$

and the distribution of ε be piece-wise uniform and given by its conditional distributions. If $x = 1$:

$$\begin{aligned} f_1(\varepsilon) &= \mu_+ \quad \text{if } 0 \leq \varepsilon < 1 \\ f_1(\varepsilon) &= \mu_- \quad \text{if } -1 \leq \varepsilon < 0 \\ f_1(\varepsilon) &= \lambda_+ \quad \text{if } \varepsilon \geq 1 \\ f_1(\varepsilon) &= \lambda_- \quad \text{if } \varepsilon < -1 \end{aligned}$$

and if $x = 0$,

$$f_0(\varepsilon) = \frac{\mu_+ + \mu_-}{2} \quad \text{if } -1 \leq \varepsilon < 1$$
$$f_0(\varepsilon) = \frac{\lambda_+ + \lambda_-}{2} \quad \text{if } \varepsilon \geq 1 \text{ or } \varepsilon < -1$$

We show how to impose three restrictions among which restrictions (5) and (14).

Restrictions First, $f_i(.)$ are density functions if:

$$\mu_+ + \mu_- + \lambda_+ + \lambda_- = 1 \tag{17}$$

Second, the moment conditions (5), $E(\varepsilon) = 0$ and $E(x\varepsilon) = 0$, hold in our case if $E(\varepsilon \mid x = 1) = 0$ and $E(\varepsilon \mid x = 0) = 0$. The first condition implies that:

$$3\lambda_- + \mu_- = 3\lambda_+ + \mu_+ \tag{18}$$

and the second condition is always satisfied by symmetry of $f_0(\varepsilon)$.

Furthermore, the support of $x\beta + \varepsilon$ is $[-2, 2]$ if $x = 0$ and $[-2 + \beta, 2 + \beta]$ if $x = 1$.The first interval is always larger than the support of v, and as $\beta < 1$, the second interval is as well so that:

$$\Pr(y = 1 \mid v) \in (0, 1)$$

The support condition($L2$) is therefore not verified.

Finally, the unbiasedness conditions are:

$$-E(x(x\beta + 1 + \varepsilon)\mathbf{1}\{-(x\beta + 1 + \varepsilon) > 0\}) = E(x(x\beta - 1 + \varepsilon)\mathbf{1}\{x\beta - 1 + \varepsilon > 0\}$$
$$-E((x\beta + 1 + \varepsilon)\mathbf{1}\{-(x\beta + 1 + \varepsilon) > 0\}) = E((x\beta - 1 + \varepsilon)\mathbf{1}\{x\beta - 1 + \varepsilon > 0\}$$

Because x takes two values only, it is again equivalent to:

$$-E((x\beta+1+\varepsilon)\mathbf{1}\{-(x\beta+1+\varepsilon) > 0\} \mid x = 1) = E((x\beta-1+\varepsilon)\mathbf{1}\{x\beta-1+\varepsilon > 0\} \mid x = 1)$$
$$-E((x\beta+1+\varepsilon)\mathbf{1}\{-(x\beta+1+\varepsilon) > 0\} \mid x = 0) = E((x\beta-1+\varepsilon)\mathbf{1}\{x\beta-1+\varepsilon > 0\} \mid x = 0)$$

thus:

$$-E((\beta + 1 + \varepsilon)\mathbf{1}\{\varepsilon < -(\beta + 1)\} \mid x = 1) = E((\beta - 1 + \varepsilon)\mathbf{1}\{\varepsilon > 1 - \beta\} \mid x = 1)$$
$$-E((1 + \varepsilon)\mathbf{1}\{-(1 + \varepsilon) > 0\} \mid x = 0) = E((-1 + \varepsilon)\mathbf{1}\{-1 + \varepsilon > 0\} \mid x = 0)$$

The second condition is always satisfied since RHS and LHS are equal to $\dfrac{3(\lambda_+ + \lambda_-)}{4}$. The first condition is more informative. Since $\beta \in]0, 1[$ its left handside is equal to:

$$-\int_{-2}^{-(\beta+1)} (\beta + 1 + \varepsilon)\lambda_- d\varepsilon = -\int_{-1+\beta}^{0} u\lambda_- du = \lambda_- . \frac{(\beta - 1)^2}{2}$$

Moreover, its right hand side is equal to:

$$\begin{aligned}
\int_{1-\beta}^{1} (\beta - 1 + \varepsilon)\mu_+ d\varepsilon + \int_{1}^{2} (\beta - 1 + \varepsilon)\lambda_+ d\varepsilon &= \int_{0}^{\beta} u\mu_+ du + \beta\lambda_+ + \int_{0}^{1} u\lambda_+ du \\
&= \mu_+ \frac{\beta^2}{2} + \beta\lambda_+ + \lambda_+/2
\end{aligned}$$

Summary Summarizing we have 3 equations:

$$\begin{aligned}
\mu_+ + \mu_- + \lambda_+ + \lambda_- &= 1 \\
3\lambda_- + \mu_- &= 3\lambda_+ + \mu_+ \\
\lambda_-.\frac{(\beta-1)^2}{2} &= \mu_+\frac{\beta^2}{2} + \beta\lambda_+ + \lambda_+/2
\end{aligned}$$

and inequations:

$$\mu_+, \mu_-, \lambda_+, \lambda_- \geq 0, 0 < \beta \leq 1$$

Therefore:

$$\lambda_-.\frac{(\beta-1)^2}{2} = (\frac{1-4\lambda_+ + 2\lambda_-}{2})\frac{\beta^2}{2} + \beta\lambda_+ + \lambda_+/2$$

Thus:

$$2\lambda_-.(\beta^2 + 1 - 2\beta) = (1 - 4\lambda_+ + 2\lambda_-)\beta^2 + 4\beta\lambda_+ + 2\lambda_+$$

$$\lambda_-.(2-4\beta) = \lambda_+.(2 + 4\beta - 4\beta^2) + \beta^2$$

$$\lambda_- = \lambda_+.\frac{1+2\beta-2\beta^2}{1-2\beta} + \frac{\beta^2}{2-4\beta} \tag{19}$$

If we restrict the domain of variation of β to $[0, 1/2[$, both coefficients are positive and therefore the domain of variation of λ_+ is the real positive line. Considering the other equations:

$$\begin{aligned}
\mu_+ + \mu_- + \frac{2-2\beta^2}{1-2\beta}\lambda_+ &= \frac{2-4\beta-\beta^2}{2-4\beta} \\
\lambda_+.\frac{3+6\beta-6\beta^2}{1-2\beta} + \frac{3\beta^2}{2-4\beta} + \mu_- &= 3\lambda_+ + \mu_+
\end{aligned}$$

thus:

$$\begin{aligned}
2\mu_- + \lambda_+.\frac{2+12\beta-8\beta^2}{1-2\beta} &= \frac{1-2\beta-2\beta^2}{1-2\beta} \\
\lambda_+.\frac{-2+12\beta-4\beta^2}{1-2\beta} + \frac{1-2\beta+\beta^2}{1-2\beta} &= 2\mu_+
\end{aligned}$$

and therefore:

$$\mu_- = \frac{1}{1-2\beta}(1/2 - \beta - \beta^2 - \lambda_+.(1 + 6\beta - 4\beta^2)) \tag{20}$$

$$\mu_+ = \frac{1}{1-2\beta}(\frac{1-2\beta+\beta^2}{2} + \lambda_+(-1 + 6\beta - 2\beta^2)) \tag{21}$$

that we can also write:

$$\begin{aligned}
\mu_- &= a_- + \lambda_+.b_- \\
\mu_+ &= a_+ + \lambda_+.b_+
\end{aligned}$$

On the domain $\beta \in [0, 1/2[$, $b_- < 0$ and $a_- > 0$ only if $\beta < \frac{-1+\sqrt{3}}{2}$. We shall restrict the domain of variation of β to $[0, \frac{-1+\sqrt{3}}{2}[$. On this domain, a_+ is always positive though b_+ is negative at $\beta = 0$ and positive at $\beta < \frac{-1+\sqrt{3}}{2}$. The domain of variation of λ_+ is therefore $[0, u_+]$ where:

$$u_+ = \min\left(\frac{1/2 - \beta - \beta^2}{1 + 6\beta - 4\beta^2}, \frac{1/2 - \beta + \beta^2/2}{1 - 6\beta + 2\beta^2}\right) \text{ if } 1 - 6\beta + 2\beta^2 > 0$$

$$u_+ = \frac{1/2 - \beta - \beta^2}{1 + 6\beta - 4\beta^2} \text{ if } 1 - 6\beta + 2\beta^2 < 0$$

It can be proven that the first bound applies and therefore:

$$u_+ = \frac{1/2 - \beta - \beta^2}{1 + 6\beta - 4\beta^2}$$

Example In the case where $\beta = 1/4$, equations (19), (20) and (21) yield:

$$\begin{aligned} \lambda_- &= \frac{11}{4}\lambda_+ + 1/16 \\ \mu_- &= 3/8 - \frac{9}{2}\lambda_+ \\ \mu_+ &= \frac{3}{4}\lambda_+ + \frac{9}{16} \end{aligned}$$

All parameters are positive if $\lambda_+ < u_+ = 1/12$.

Details Different distributions of v are used in the experiments (Table 1):

1. **Uniform:** $v \rightsquigarrow \mathcal{U}[-1, 1]$

2. **Logit:** The support of the distribution is $[-1, 1]$ and the distribution $F(.)$ is given by:

   $$3v = \log\left[(F(v) + c)/(1 - F(v) + c)\right]$$

   where $c = 1/(\exp(3) - 1)$

3. **Triangular:** The support of the distribution is $[-1, 1]$ and the distribution $F(.)$ is given by:

   $$v = [2.\mathbf{1}\{F(v) < 1/2\} - 1]\,[2.(F(v) - 1/2)]^2$$

4. **Assymetric:** The support of the distribution is $[-1, 1]$ and the distribution $F(.)$ is given by:

   $$3v = \log\left[(F(v)^2 + c)/(1 - F(v)^2 + c)\right]$$

   where $c = 1/(\exp(3) - 1)$

Kernel: All estimations of the conditional distribution of $f(v \mid x, z)$ are performed using a bi-weight kernel function:

$$K(u) = 15/16.(1 - u^2)^2\mathbf{1}\{\,|u| \leq 1\}$$

E.2 Second experiment

The random term ε should satisfy two constraints that are the moment restriction (5) and the unbiasedness condition (14) that is:

$$\begin{array}{rcl} E(z'\varepsilon) & = & 0 \\ -E(z'(x\beta + v_H + \varepsilon)\mathbf{1}\{-(x\beta + v_H + \varepsilon) & > & 0\}) = E(z'(x\beta + v_L + \varepsilon)\mathbf{1}\{x\beta + v_L + \varepsilon > 0\}) \end{array}$$

Consider the second one. By definition, it yields:

$$E(z'\varepsilon_1\mathbf{1}\{\varepsilon_0 + z\delta \in B_F(x)\}) = E(z'\varepsilon_1\mathbf{1}\{\varepsilon_0 + z\delta \in B_S(x)\})$$

As ε_0 and ε_1 are independent drawings conditional on z, it is always satisfied for any value of δ. Consider now the first moment condition:

$$\begin{array}{rcl} E(z'\varepsilon) & = & E(z'(\varepsilon_0 + z\delta)\mathbf{1}\{\varepsilon_0 + z\delta \in B_M(x)\} \\ & + & E(z'(-(x\beta + v_H) - \varepsilon_1)\mathbf{1}\{\varepsilon_0 + z\delta \in B_F(x)\}) \\ & + & E(z'(-(x\beta + v_L) + \varepsilon_1)\mathbf{1}\{\varepsilon_0 + z\delta \in B_S(x)\}) \end{array}$$

where we denoted:

$$B_M(x) = \mathbb{R}/(B_F(x) \cup B_S(x))$$

By the same argument as above, we can eliminate ε_1:

$$\begin{array}{rcl} E(z'\varepsilon) & = & E(z'(\varepsilon_0 + z\delta)\mathbf{1}\{\varepsilon_0 + z\delta \in B_M(x)\}) \\ & - & E(z'(x\beta + v_H)\mathbf{1}\{\varepsilon_0 + z\delta \in B_F(x)\}) \\ & - & E(z'(x\beta + v_L)\mathbf{1}\{\varepsilon_0 + z\delta \in B_S(x)\}) \end{array}$$

and we must find δ such that this moment is equal to zero. We can then rewrite:

$$\begin{array}{rcl} 0 & = & E(z'(\varepsilon_0 + z\delta) - E(z'(\varepsilon_0 + z\delta + x\beta + v_H)\mathbf{1}\{\varepsilon_0 + z\delta \in B_F(x)\}) \\ - & & E(z'(\varepsilon_0 + z\delta + x\beta + v_L)\mathbf{1}\{\varepsilon_0 + z\delta \in B_S(x)\}) \end{array}$$

Consider the following algorithm. Let $\delta_0 = 0$. Find δ_n such that:

$$\begin{array}{rcl} 0 & = & E(z'(\varepsilon_0 + z\delta_n) - E(z'(\varepsilon_0 + z\delta_{n-1} + x\beta + v_H)\mathbf{1}\{\varepsilon_0 + z\delta_{n-1} \in B_F(x)\}) \\ - & & E(z'(\varepsilon_0 + z\delta_{n-1} + x\beta + v_L)\mathbf{1}\{\varepsilon_0 + z\delta_{n-1} \in B_S(x)\}) \end{array}$$

that is:

$$\begin{array}{rcl} \delta_n & = & -\left[E(z'z)\right]^{-1}\{E(z'\varepsilon_0) - E(z'(\varepsilon_0 + z\delta_{n-1} + x\beta + v_H)\mathbf{1}\{\varepsilon_0 + z\delta_{n-1} \in B_F(x)\}) \\ - & & E(z'(\varepsilon_0 + z\delta_{n-1} + x\beta + v_L)\mathbf{1}\{\varepsilon_0 + z\delta_{n-1} \in B_S(x)\}\} \end{array}$$

If the algorithm converges, it converges to a value which satisfies the different conditions above. To find parameter δ as a function of the Monte Carlo parameters $(\gamma_0, \gamma_s, \gamma_x, \gamma_z)$ and the distribution L, we use a very large sample consisting of many replications of the original sample and drawings of ε_0 and ε_1.

Some details are also reported in the previous Appendix E.1.

E.3 Optimizing the Window Size

For choosing the window, we adapt the procedure reported by Lewbel (2000). It is an attempt to exploit the identity:

$$\delta = E((\mathbf{1}\{\delta + v + \varepsilon > 0\} - \mathbf{1}\{v > 0\})/f(v \mid x, z))$$

where ε is any zero mean homoskedastic random shock satisfying conditions $(L.1 - L.3)$.

1. We divide the interval between the 5th percentile and the 95th percentile of the distribution of v into I intervals and consider any corresponding value of δ_i at the mid interval.

2. By simulation we compute the average, $\bar{\delta}_i$,of the argument given above within the expectation using a uniformly distributed ε satisfying conditions $(L.1 - L.3)$ for any δ_i. We compute the following distance:

$$S(h) = \sum_{i=1}^{I} \frac{(\bar{\delta}_i - \delta_i)^2}{\hat{\sigma}_i^2}$$

   where $\hat{\sigma}_i$ is the empirical standard error of $\bar{\delta}_i$ and where h is any window size.

3. We repeat the previous step for window size h around the Silverman window:

$$h_0 = (4/(k+2))^p n^{-p}$$

   where k is the number of covariates, n is the number of observations and $p = 1/(k+4)$. The number of window size h is equal to 30 and the range is between $h_0/4$ and $4h_0$.

Table 1: Baseline Monte Carlo experiment (Simple version)

			α :True value = 0.00				β : True value = 0.25			
Distribution	**Nobs**	**Window**	**Bias**	**StdErr**	**RMSE**	**MAE**	**Bias**	**StdErr**	**RMSE**	**MAE**
Uniform	100	0.3781	0.0003	0.1121	0.1119	0.0904	0.0245	0.1989	0.1999	0.1573
	200	0.3766	0.0017	0.0723	0.0722	0.0570	0.0162	0.1327	0.1334	0.1041
	500	0.3531	-0.0013	0.0425	0.0424	0.0341	0.0100	0.0798	0.0802	0.0623
	1000	0.2385	0.0001	0.0301	0.0300	0.0238	0.0055	0.0537	0.0539	0.0441
Logit	100	0.2690	0.0382	0.4204	0.4211	0.1323	-0.0651	0.7398	0.7408	0.3000
	200	0.2817	0.0359	0.2859	0.2875	0.0980	-0.0195	0.3176	0.3174	0.1595
	500	0.2345	0.0022	0.0522	0.0522	0.0420	0.0039	0.0921	0.0919	0.0742
	1000	0.1697	0.0026	0.0355	0.0355	0.0282	0.0024	0.0592	0.0591	0.0484
Triangular	100	0.3235	0.0082	0.2864	0.2858	0.1452	0.0163	0.3781	0.3775	0.2447
	200	0.3291	-0.0022	0.0951	0.0949	0.0705	0.0123	0.1621	0.1621	0.1289
	500	0.2740	-0.0009	0.0511	0.0510	0.0387	0.0102	0.0884	0.0887	0.0715
	1000	0.2041	0.0013	0.0347	0.0347	0.0266	0.0002	0.0594	0.0593	0.0478
Assymetric	100	0.1054	-0.1777	2.2279	2.2294	0.3583	0.2757	2.4128	2.4225	0.6840
	200	0.0918	-0.1911	1.5399	1.5479	0.2724	0.0088	1.8123	1.8078	0.5196
	500	0.2345	0.0072	0.0587	0.0590	0.0469	-0.0056	0.1113	0.1112	0.0844
	1000	0.1697	0.0020	0.0422	0.0422	0.0335	0.0043	0.0751	0.0750	0.0569

Notes: 200 replications. All details are reported in Appendix E.1. **Distribution** corresponds to different choices for the distribution of v. **Nobs** is the number of observations. The **Window** size is optimized according to the device explained in Appendix E.3. **Bias** is the average of differences between the estimates and the true value. **Stderr** is the empirical standard error of the estimates, **RMSE** is the root mean square error and **MAE** is the mean absolute error. The average probabilities of perfect success and perfect failure (see text and Appendix E.1) are equal to 0.15 and 0.12.

Table 2: Monte Carlo experiment (Simple version): Window sensitivity

		α :True value = 0.00				β : True value = 0.25			
Nobs	**Window**	**Bias**	**StdErr**	**RMSE**	**M AE**	**Bias**	**StdErr**	**RMSE**	**MAE**
100	0.1345	0.0251	0.8366	0.8349	0.2004	-0.0197	1.2544	1.2 514	0.5309
	0.2018	0.0086	0.1626	0.1624	0.1108	0.0442	0.4109	0.41 22	0.2668
	0.2690	0.0382	0.4204	0.4211	0.1323	-0.0651	0.7398	0.7 408	0.3000
	0.3363	0.0605	0.4381	0.4411	0.1527	-0.0456	0.4812	0.4 821	0.2544
	0.4035	0.0199	0.1750	0.1757	0.1103	-0.0117	0.2596	0.2 592	0.1962
200	0.1408	0.0161	0.1653	0.1656	0.0842	0.0022	0.3002	0.29 94	0.1743
	0.2112	0.0124	0.3995	0.3987	0.1055	0.0046	0.4218	0.42 08	0.1737
	0.2817	0.0359	0.2859	0.2875	0.0980	-0.0195	0.3176	0.3 174	0.1595
	0.3521	0.0134	0.1023	0.1030	0.0720	-0.0024	0.1656	0.1 652	0.1273
	0.4225	0.0088	0.0853	0.0856	0.0658	0.0007	0.1494	0.14 90	0.1181
500	0.1172	0.0027	0.0570	0.0569	0.0448	0.0045	0.1074	0.10 72	0.0854
	0.1759	0.0026	0.0542	0.0542	0.0430	0.0144	0.1765	0.17 66	0.0869
	0.2345	0.0022	0.0522	0.0522	0.0420	0.0039	0.0921	0.09 19	0.0742
	0.2931	0.0020	0.0499	0.0498	0.0407	0.0028	0.0895	0.08 93	0.0721
	0.3517	0.0018	0.0485	0.0484	0.0395	0.0027	0.0886	0.08 84	0.0713
1000	0.0849	0.0034	0.0384	0.0385	0.0295	0.0070	0.0765	0.0 766	0.0548
	0.1273	0.0032	0.0363	0.0363	0.0284	0.0030	0.0623	0.0 622	0.0504
	0.1697	0.0026	0.0355	0.0355	0.0282	0.0024	0.0592	0.0 591	0.0484
	0.2122	0.0022	0.0354	0.0354	0.0281	0.0022	0.0585	0.0 584	0.0476
	0.2546	0.0019	0.0352	0.0351	0.0279	0.0023	0.0581	0.0 580	0.0471

Notes: 200 replications. All details are reported in Appendix E.1. The distribution of v is Logit. **Nobs** is the number of observations. The **Window** size is varying between one half and 2 times the optimized value according to the device explained in Appendix E.3. **Bias** is the average of differences between the estimates and the true value. **Stderr** is the empirical standard error of the estimates, **RMSE** is the root mean square error and **MAE** is the mean absolute error.

Table 3: Monte Carlo simple experiment: Sensitivity to the support of v

	Probabilities		α :True value = 0.00				β : True value = 0.25			
Nobs	**Success**	**Failure**	**Bias**	**StdErr**	**RMSE**	**MAE**	**Bias**	**StdErr**	**RMSE**	**MAE**
100	0.08	0.04	0.0079	0.0706	0.0708	0.0471	-0.0030	0.5205	0.5192	0.1849
200			0.0124	0.0758	0.0766	0.0339	0.0106	0.1575	0.1574	0.0972
500			0.0053	0.0274	0.0278	0.0211	0.0032	0.0712	0.0711	0.0554
1000			0.0012	0.0188	0.0188	0.0144	0.0034	0.0462	0.0462	0.0364
100	0.13	0.09	0.0230	0.4092	0.4088	0.1085	-0.0164	0.6678	0.6663	0.2551
200			0.0293	0.2752	0.2761	0.0833	-0.0061	0.3048	0.3041	0.1420
500			-0.0002	0.0450	0.0449	0.0362	0.0102	0.0867	0.0871	0.0690
1000			-0.0001	0.0300	0.0300	0.0237	0.0047	0.0565	0.0565	0.0450
100	0.17	0.15	0.0240	0.4563	0.4558	0.1558	-0.0751	0.8328	0.8341	0.3558
200			0.0353	0.2866	0.2880	0.1073	-0.0190	0.3223	0.3220	0.1716
500			0.0045	0.0588	0.0588	0.0474	-0.0008	0.0983	0.0981	0.0770
1000			0.0026	0.0396	0.0396	0.0316	0.0000	0.0628	0.0627	0.0515
100	0.21	0.20	0.0131	0.4693	0.4683	0.1816	-0.0694	0.8298	0.8306	0.3737
200			0.0209	0.2083	0.2088	0.1127	-0.0186	0.2750	0.2750	0.1850
500			0.0022	0.0659	0.0658	0.0527	-0.0021	0.1021	0.1019	0.0798
1000			0.0014	0.0479	0.0478	0.0373	-0.0021	0.0704	0.0702	0.0569

Notes: 200 replications. All details are reported in Appendix E.1. The distribution of v is Logit. **Nobs** is the number of observations. The **Window** size is optimized according to the device explained in Appendix E.3. **Bias** is the average of differences between the estimates and the true value. **Stderr** is the empirical standard error of the estimates, **RMSE** is the root mean square error and **MAE** is the mean absolute error. **Success** (resp. **Failure**) probabilities are the frequency of observations such success (i.e. $y_i = 1$, resp. failure) is certain whatever value v takes.

Table 4: Sensitivity to the support of v (2nd version)

	Probabilities		α :True value = 0.00				β : True value = 0.10			
Nobs	**Success**	**Failure**	**Bias**	**StdErr**	**RMSE**	**MAE**	**Bias**	**StdErr**	**RMSE**	**MAE**
100	0.03	0.01	-0.0007	0.0315	0.0314	0.0078	0.0109	0.1724	0.1724	0.1251
200			0.0001	0.0144	0.0143	0.0051	0.0119	0.1308	0.1310	0.0790
500			0.0003	0.0080	0.0079	0.0040	0.0014	0.0585	0.0584	0.0473
1000			-0.0005	0.0057	0.0057	0.0037	0.0012	0.0398	0.0397	0.0330
100	0.12	0.11	0.0382	0.4204	0.4211	0.1323	-0.0803	0.7299	0.7325	0.2802
200			0.0359	0.2859	0.2875	0.0980	-0.0305	0.3130	0.3137	0.1529
500			0.0022	0.0522	0.0522	0.0420	-0.0060	0.0904	0.0904	0.0723
1000			0.0026	0.0355	0.0355	0.0282	-0.0031	0.0583	0.0582	0.0468
100	0.21	0.22	0.0137	0.4763	0.4753	0.1957	-0.0919	0.8325	0.8355	0.3794
200			0.0235	0.2122	0.2129	0.1167	-0.0301	0.2819	0.2828	0.1855
500			0.0024	0.0671	0.0669	0.0538	-0.0051	0.1063	0.1062	0.0858
1000			0.0001	0.0497	0.0495	0.0385	-0.0044	0.0736	0.0735	0.0596

Notes: 200 replications. All details are reported in Appendix E.1. The distribution of v is Logit. **Nobs** is the number of observations. The **Window** size is optimized according to the device explained in Appendix E.3. **Bias** is the average of differences between the estimates and the true value. **Stderr** is the empirical standard error of the estimates, **RMSE** is the root mean square error and **MAE** is the mean absolute error. **Success** (resp. **Failure**) probabilities are the frequency of observations such success (i.e. $y_i = 1$, resp. failure) is certain whatever value v takes.

Table 5: Baseline Monte Carlo experiment: Complete version

			$\beta = 0.25$: OLS estimate				$\beta = 0.25$: 2SLS estimate			
Distribution	**Nobs**	**Window**	**Bias**	**StdErr**	**RMSE**	**MAE**	**Bias**	**StdErr**	**RMSE**	**MAE**
Uniform	100	0.6642	0.0246	0.1980	0.1985	0.1488	0.0986	0.2789	0.2945	0.2246
	200	0.7629	0.0196	0.0970	0.0984	0.0775	0.0698	0.1448	0.1601	0.1232
	500	0.5997	0.0005	0.0628	0.0625	0.0517	0.0569	0.0806	0.0984	0.0785
	1000	0.5343	0.0003	0.0382	0.0380	0.0307	0.0460	0.0534	0.0703	0.0573
Logit	100	0.6402	-0.0251	0.1775	0.1784	0.1419	0.0570	0.2638	0.2686	0.2060
	200	0.6559	-0.0209	0.1104	0.1118	0.0841	0.0328	0.1520	0.1547	0.1271
	500	0.5447	-0.0176	0.0764	0.0780	0.0601	0.0382	0.0959	0.1028	0.0800
	1000	0.4525	-0.0090	0.0574	0.0579	0.0401	0.0450	0.0697	0.0827	0.0633
Triang	100	0.5922	0.0553	0.2457	0.2507	0.2006	0.1053	0.2926	0.3096	0.2352
	200	0.5062	0.0482	0.1785	0.1840	0.1365	0.0910	0.2586	0.2730	0.2064
	500	0.4529	0.0384	0.0883	0.0959	0.0743	0.0977	0.1217	0.1556	0.1186
	1000	0.4035	0.0298	0.0920	0.0963	0.0545	0.0726	0.0992	0.1226	0.0851
Assymetric	100	0.5442	0.0031	0.2742	0.2729	0.2092	0.1211	0.3710	0.3885	0.2966
	200	0.6345	0.0056	0.1307	0.1301	0.0968	0.0622	0.1848	0.1941	0.1552
	500	0.5263	-0.0031	0.0853	0.0850	0.0670	0.0478	0.1247	0.1329	0.1050
	1000	0.4689	0.0024	0.0717	0.0714	0.0480	0.0669	0.1175	0.1347	0.0976

Notes: 200 replications. All details are reported in Appendix E.2. **Distribution** corresponds to different choices for the distribution of v. **Nobs** is the number of observations. The **Window** size is optimized according to the device explained in Appendix E.3. **Bias** is the average of differences between the estimates and the true value. **Stderr** is the empirical standard error of the estimates, **RMSE** is the root mean square error and **MAE** is the mean absolute error.

Table 6: Complete Monte Carlo experiment: Sensitivity to endogeneity

		$\beta = 0.25$: OLS estimate				$\beta = 0.25$: 2SLS estimate			
Nobs	**Endogeneity**	**Bias**	**StdErr**	**RMSE**	**MAE**	**Bias**	**StdErr**	**RMSE**	**MAE**
100	0.0000	-0.0251	0.1775	0.1784	0.1419	0.0570	0.2638	0.2 686	0.2060
	0.3333	0.0419	0.1638	0.1683	0.1260	0.0576	0.2590	0.26 41	0.2034
	0.6667	0.1115	0.1663	0.1995	0.1548	0.0573	0.2509	0.25 61	0.1981
	1.0000	0.1609	0.1719	0.2349	0.1908	0.0448	0.2531	0.25 58	0.1992
200	0.0000	-0.0209	0.1104	0.1118	0.0841	0.0328	0.1520	0.1 547	0.1271
	0.3333	0.0406	0.1042	0.1113	0.0856	0.0267	0.1411	0.14 29	0.1154
	0.6667	0.1065	0.1061	0.1499	0.1217	0.0301	0.1355	0.13 81	0.1079
	1.0000	0.1562	0.1091	0.1903	0.1618	0.0249	0.1392	0.14 08	0.1129
500	0.0000	-0.0176	0.0764	0.0780	0.0601	0.0382	0.0959	0.1 028	0.0800
	0.3333	0.0434	0.0705	0.0825	0.0612	0.0280	0.0953	0.09 89	0.0770
	0.6667	0.1057	0.0746	0.1292	0.1072	0.0256	0.1014	0.10 41	0.0817
	1.0000	0.1596	0.0724	0.1751	0.1596	0.0234	0.1043	0.10 64	0.0849
1000	0.0000	-0.0090	0.0574	0.0579	0.0401	0.0450	0.0697	0. 0827	0.0633
	0.3333	0.0520	0.0545	0.0751	0.0587	0.0364	0.0706	0.0 791	0.0613
	0.6667	0.1093	0.0561	0.1228	0.1093	0.0293	0.0710	0.0 765	0.0590
	1.0000	0.1633	0.0527	0.1715	0.1633	0.0222	0.0715	0.0 745	0.0577

Notes: 200 replications. All details are reported in Appendix E.2. The distribution of v is assumed to be Logit. **Nobs** is the number of observations. The **Window** size is optimized according to the device explained in Appendix E.3. **Bias** is the average of differences between the estimates and the true value. **Stderr** is the empirical standard error of the estimates, **RMSE** is the root mean square error and **MAE** is the mean absolute error. The **endogeneity** parameter is γ_0 as explained in the text. $\gamma_0 = 0$ stands for the exogeneity case.

Table 7: Complete Monte Carlo experiment: Sensitivity to heteroskedasticity

		$\beta = 0.25$: OLS estimate				$\beta = 0.25$: 2SLS estimate			
Nobs	**Heteroskedasticity**	**Bias**	**StdErr**	**RMSE**	**MAE**	**Bias**	**StdErr**	**RMSE**	**MAE**
100	0.0000	-0.0251	0.1775	0.1784	0.1419	0.0570	0.2638	0.2686	0.2060
	0.3333	-0.0186	0.1872	0.1872	0.1466	0.0767	0.2732	0.2825	0.2146
	0.6667	-0.0269	0.1946	0.1955	0.1541	0.0763	0.2721	0.2813	0.2144
	1.0000	-0.0223	0.2057	0.2058	0.1603	0.0582	0.2583	0.2635	0.2074
200	0.0000	-0.0209	0.1104	0.1118	0.0841	0.0328	0.1520	0.1547	0.1271
	0.3333	-0.0212	0.1145	0.1159	0.0889	0.0342	0.1545	0.1575	0.1245
	0.6667	-0.0286	0.1207	0.1235	0.0949	0.0318	0.1558	0.1583	0.1226
	1.0000	-0.0324	0.1238	0.1274	0.0981	0.0282	0.1594	0.1611	0.1230
500	0.0000	-0.0176	0.0764	0.0780	0.0601	0.0382	0.0959	0.1028	0.0800
	0.3333	-0.0187	0.0837	0.0854	0.0637	0.0339	0.1055	0.1103	0.0853
	0.6667	-0.0288	0.0812	0.0858	0.0645	0.0272	0.1077	0.1106	0.0845
	1.0000	-0.0301	0.0816	0.0866	0.0655	0.0230	0.1094	0.1112	0.0872
1000	0.0000	-0.0090	0.0574	0.0579	0.0401	0.0450	0.0697	0.0827	0.0633
	0.3333	-0.0101	0.0639	0.0644	0.0447	0.0432	0.0771	0.0880	0.0700
	0.6667	-0.0183	0.0626	0.0649	0.0456	0.0327	0.0740	0.0806	0.0644
	1.0000	-0.0251	0.0582	0.0631	0.0454	0.0251	0.0710	0.0749	0.0610

Notes: 200 replications. All details are reported in Appendix E.2. The distribution of v is assumed to be Logit. **Nobs** is the number of observations. The **Window** size is optimized according to the device explained in Appendix E.3. **Bias** is the average of differences between the estimates and the true value. **Stderr** is the empirical standard error of the estimates, **RMSE** is the root mean square error and **MAE** is the mean absolute error. The **heteroskedasticity** parameter is γ_x as explained in the text. $\gamma_x = 0$ describes the homoskedastic case.

Table 8: Sensitivity to Endogeneity and Heteroskedasticity

		$\beta = 0.25$: OLS estimate				$\beta = 0.25$: 2SLS estimate			
Endogeneity	**Heteroskedasticity**	**Bias**	**StdErr**	**RMSE**	**M AE**	**Bias**	**StdErr**	**RMSE**	**MAE**
0.00	0.00	-0.0176	0.0764	0.0780	0.0601	0.0382	0.0959	0.1028	0.0800
	0.33	-0.0187	0.0837	0.0854	0.0637	0.0339	0.1055	0.1103	0.0853
	0.67	-0.0288	0.0812	0.0858	0.0645	0.0272	0.1077	0.1106	0.0845
	1.00	-0.0301	0.0816	0.0866	0.0655	0.0230	0.1094	0.1112	0.0872
0.33	0.00	0.0434	0.0705	0.0825	0.0612	0.0280	0.0953	0.0989	0.0770
	0.33	0.0370	0.0840	0.0914	0.0675	0.0267	0.1071	0.1099	0.0853
	0.67	0.0291	0.0834	0.0879	0.0646	0.0240	0.1056	0.1078	0.0811
	1.00	0.0214	0.0830	0.0853	0.0657	0.0188	0.1084	0.1095	0.0859
0.67	0.00	0.1057	0.0746	0.1292	0.1072	0.0256	0.1014	0.1041	0.0817
	0.33	0.0982	0.0797	0.1262	0.1036	0.0238	0.0992	0.1016	0.0808
	0.67	0.0854	0.0818	0.1180	0.0960	0.0172	0.1112	0.1120	0.0843
	1.00	0.0697	0.0842	0.1090	0.0884	0.0088	0.1166	0.1163	0.0865
1.00	0.00	0.1596	0.0724	0.1751	0.1596	0.0234	0.1043	0.1064	0.0849
	0.33	0.1560	0.0787	0.1745	0.1565	0.0227	0.1061	0.1080	0.0875
	0.67	0.1356	0.0823	0.1584	0.1401	0.0056	0.1100	0.1096	0.0880
	1.00	0.1210	0.0880	0.1494	0.1289	0.0015	0.1192	0.1186	0.0919

Notes: 200 replications. All details are reported in Appendix E.2. The distribution of v is assumed to be Logit. **T**he number of observations is 500. The **Window** size is optimized according to the device explained in Appendix E.3. **Bias** is the average of differences between the estimates and the true value. **Stderr** is the empirical standard error of the estimates, **RMSE** is the root mean square error and **MAE** is the mean absolute error. The **endogeneity** parameter is γ_0 as explained in the text. $\gamma_0 = 0$ stands for the exogeneity case. The **heteroskedasticity** parameter is γ_x as explained in the text. $\gamma_x = 0$ describes the homoskedastic case.