Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Journal of Econometrics ■ (■■■■) ■■■–■■■

JOURNAL OF Econometrics

www.elsevier.com/locate/jeconom

Identification and information in monotone binary models

Thierry Magnac[a,*], Eric Maurin[b]

[a] *~~INRA-LEERNA and IDEI,~~ 21 Allée de Brienne, 31000 Toulouse, France*

[b] *~~CREST-INSEE, 15, Boulevard Gabriel Péri, 92245 Malakoff Cedex,~~ France*

Abstract

This paper considers binary response models where errors are uncorrelated with a set of instrumental variables and are independent of a continuous regressor v, conditional on all other variables. It is shown that these exclusion restrictions are not sufficient for identification and that additional identifying assumptions are needed. Such an assumption, introduced by Lewbel [Semiparametric qualitative response model estimation with unknown heteroskedasticity or instrumental variables. Journal of Econometrics 97, 145–177], is that the support of the continuous regressor is large, but we show that it significantly restricts the class of binary phenomena which can be analysed. We propose an alternative additional assumption under which β remains just identified and the estimation unchanged. This alternative assumption does not impose specific restrictions on the data, which broadens the scope of the estimation method in empirical work. The semiparametric efficiency bound of the model is also established and an existing estimator is shown to achieve that bound. The efficient estimator uses a plug-in density estimate. It is shown that plugging in the true density rather than an estimate is inefficient. Extensions to ordered choice models are provided.

JEL classification: C14; C25

Keywords: Binary models; Semiparametric methods; Very exogeneous regressor

*Corresponding author.

E-mail addresses: magnac@cict.fr (T. Magnac), maurin@ensae.fr (E. Maurin).

URL: http://www.idei.fr/vitae/magnac.html.

0304-4076/$ - see front matter © 2006 Published by Elsevier B.V.
doi:10.1016/j.jeconom.2006.06.005

1. Introduction

Let y be a binary outcome, x a vector of covariates and v a continuous covariate. This paper considers the following binary response model:

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

where errors, ε, are uncorrelated with a set of instrumental variables z (i.e., $\mathrm{E}(z'\varepsilon) = 0$), and conditionally independent of the continuous regressor, v, (i.e., partial independence, $F_\varepsilon(\varepsilon \mid x, z, v) = F_\varepsilon(\varepsilon \mid x, z)$). As discussed below, these exclusion restrictions arise naturally in many economic models and the purpose of this paper is to analyse the conditions under which they can be used in empirical applications, as well as the conditions under which they provide a means for identifying the structural parameters of the latent model.

It is shown that the partial independence assumption does not impose very strong restrictions on the data and can be used in a wide range of cases. Specifically, it is shown that any binary outcome can be analysed with a latent model satisfying partial independence provided that it is monotone in v (i.e., $\Pr(y = 1 \mid v, x, z)$ is monotone in v). The problem is that partial-independence is not sufficient for identification of the parameter of interest β, even when it is combined with the uncorrelated instrument assumption. Additional restrictions are needed for identification. Such an additional restriction is the assumption, introduced by Lewbel (2000), that the support of the special regressor is large ($\mathrm{Supp}(v) \supseteq \mathrm{Supp}(-x\beta - \varepsilon)$). Our second result is that the combination of uncorrelated instruments, partial-independence and large-support assumptions provides *exact* identification of β. Yet, it is shown that the large-support assumption significantly restricts the class of binary phenomena which can be analysed through (LV). Specifically, we show that the large-support assumption can only be used when the conditional probability of success $\Pr(y = 1 \mid v, x, z)$ increases from 0 to 1 over the support of v, which is admittedly restrictive. Large-support conditions are actually quite common in the literature about semi-parametric limited-dependent variable models (see e.g., Manski, 1975, 1985; Han, 1987; Horowitz, 1992; Cavanagh and Sherman, 1998), but they represent a potential obstacle to empirical applications.

This is why we propose an alternative to the large-support condition. Parameter β remains just identified and the estimation unchanged when a symmetry condition on the tails of the errors ε holds. This alternative to the large-support assumption can be used when the conditional probability of success does not vary from 0 to 1 over the support of v, which increases the usefulness of the setting in empirical work.

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. This is where the paper presents two further contributions. We establish the semi-parametric efficiency bound for the parameter under partial independence and uncorrelated instrument assumptions by using the framework proposed by Severini and Tripathi (2001). It is noteworthy that the special-regressor estimator proposed by Lewbel (2000) achieves this bound (under some regularity conditions). The efficient estimator uses a plug-in density. It is shown that plugging in the true density, when it is known, rather than an estimate is inefficient. This finding was conjectured by Lewbel (2000).

Generally speaking, the set of identifying restrictions analysed in this paper provides interesting means to overcome Manski's fundamental impossibility result according to which an uncorrelated-error restriction (i.e., $\mathrm{E}(x'\varepsilon) = \mathrm{E}(v'\varepsilon) = 0$) or even a mean-independence restriction (i.e., $\mathrm{E}(\varepsilon \mid x, v) = 0$) is not sufficient for identifying β no matter what conditions on the support of (v, x) are adopted (see Manski, 1988). Also, the set-up used in this paper imposes much weaker distributional assumptions on the error terms than standard parametric models or than the semi-parametric methods that are based on the properties of statistical independence (i.e., $F_\varepsilon(\varepsilon \mid x, v) = F_\varepsilon(\varepsilon)$) or of single-index sufficiency (i.e., $F_\varepsilon(\varepsilon \mid x, v) = F_\varepsilon(\varepsilon \mid x\beta + v)$) see e.g. Cosslett (1983), Ruud (1983), Han (1987), Powell et al. (1989), Ichimura (1993), Klein and Spady (1993) who provide estimators of β under statistical independence or index sufficiency.

The quantile-independence assumption does not provide just identification of the parameter of interest, but permits slightly more general forms of conditional heteroskedasticity than the exclusion restrictions used in this paper. Still, the fact remains that very few empirical studies use the corresponding maximum score estimation method, as developed by Manski (1975, 1985) or its smoothed 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) who suggests strengthening the median-independence assumption into conditional symmetry and a weak restriction on conditional heteroskedasticity. Estimation can be proven to be root-n consistent, though optimization is still needed. Under the identifying restrictions used in this paper, the *special-regressor* estimator developed by Lewbel (2000) can be directly obtained without optimization and is root-n consistent. The implementation of the estimation method is quite simple. It only requires the estimation of a conditional density and a linear regression. Honoré and Lewbel (2002) extends this method to estimating binary choice models using panel data and allowing for individual effects. Recent empirical applications of this estimation method include Anton et al. (2001) , Lewbel et al. (2001) , Maurin (2002), Lewbel (2003).

The paper is organized as follows.

Section 2 provides the equivalence result between the set of latent models satisfying uncorrelated instruments, partial independence and large-support conditions and the set of random variables (y, x, v, z) such that the conditional probability of success $\Pr(y = 1 \mid v, x, z)$ increases from 0 to 1 when v varies over its support.

Section 3 shows that uncorrelated instruments and partial independence alone are not sufficient for identification of β. We propose an alternative to the large-support assumption for obtaining just identification of β. The only condition that $\Pr(y = 1 \mid v, x, z)$ should satisfy is to be monotone in v.

In Section 4, we state the semi-parametric efficiency bound and the efficiency comparison between two estimators using estimated or true density functions.

Section 5 provides extensions of the equivalence result to ordered choice models and Section 6 concludes. All proofs are in the Appendices.

2. The set-up and the equivalence result

Let the "data" be given by the distribution of the following random variable where, for simplicity, we only consider random samples and we do not subscript individual observations by i:

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

Variable 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. Their respective role in the latent model will be clarified below. We first introduce some regularity conditions on the distribution of ω, which will be assumed valid in the rest of the text.

Assumption R(egularity).

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

R.ii. (*Covariates and instruments*): The support of the distribution of (x, z) is a compact set $S_{x,z} \subset R^p \times R^q$. The dimension of the set $S_{x,z}$ is $r \leqslant p + q$ where $p + q - r$ are the potential overlaps and functional dependencies.[1] The probability measure, $dF_{x,z}$, is supposed to be absolutely continuous with respect to a product of Lebesgue and discrete measures so as to allow continuous, discrete or mixed regressors. Finally, $rank(\mathrm{E}(z'x)) = p$.

R.iii. (*Special regressor*): The support of the conditional distribution of v conditional on (x, z) is $]v_L, v_H[$ almost everywhere (a.e.) $F_{x,z}$. Moreover, $v_L < 0 < v_H$ and v_L and v_H can be infinite. The conditional distribution is denoted $F_v(.\,|\,x, z)$ and is defined a.e. $F_{x,z}$. It is absolutely continuous with respect to the Lebesgue measure and its density $f(v\,|\,x, z)$ is continuous and bounded away from zero except possibly on the boundary of the support of v.

R.iv. (*Functional independence of v and* (x, z)): There is no subspace of $]v_L, v_H[\times S_{x,z}$ of dimension strictly less than $r + 1$ which probability measure, $(F_v(.\,|\,x, z).F_{x,z})$, 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). Denoting (x, z) as u may lead to less ambiguous arguments below at the cost of additional notation. We prefer to stick to the more parsimonious notation (x, z). Assumption *R.iii* defines what is meant by the continuity of the special regressor v. The support of v might be made dependent on (x, z) with no loss of generality. 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 in Lewbel (2000) and Honoré and Lewbel (2002). The second one is a non-parametric binary model. Let us start with the latent binary model:

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

where $1(A)$ is the indicator function that equals one if A is true and zero otherwise and $\beta \in \mathbb{R}^p$ is the vector of coefficients of interest. The distribution of the random error ε satisfies the following properties as in Lewbel (2000):

Assumption L(atent).

[1]With no loss of generality, the p explanatory variables x can partially overlap with the $q \geqslant p$ instrumental variables z. Variables (x, z) may also be functionally dependent (for instance x, x^2, $\log(x), \ldots$). 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).

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

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

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

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

(L.3) (*Uncorrelated instruments*): The random shock ε is uncorrelated with variables z: $\mathrm{E}(z'\varepsilon)=0$.

Regarding (L.1), it should be noted that Powell (1994) discusses partial independence assumptions (calling them exclusion restrictions) in the context of other semi-parametric models, i.e., without combining them with (L.2) or (L.3). Generally speaking, partial independence assumptions are akin to exogeneity assumptions and arise in many economic models. For example, in a labour supply model where ε represents unobserved ability, partial independence is satisfied by any variable that affects or is correlated with labour supply decisions but not with ability (such as government benefits). In consumer demand models where ε represents unobserved preference variation, prices satisfy the partial independence condition when goods are homogenous and markets are competitive. In contingent valuation studies, where ε stands for unobserved taste variation, v can be the bid that is determined by experimental design, and so may be constructed by the researcher to satisfy the necessary exclusion and support restrictions. Lewbel et al. (2001) provide an empirical application for this case. Other empirical applications using the partial independence assumption include Maurin (2002) who estimate an education production function using date of birth (within the year) as special regressor v. Date of birth within the year significantly influences children's outcomes in primary school. Given that this variable is plausibly independent from children's unobserved ability, partial independence is plausible too. Cogneau and Maurin (2002) study demand for education in primary schools in Madagascar, using the same special regressor. Lewbel (2003) studies the probability of obtaining a university degree using the cost of attending a local public college (relative to local unskilled wages) as a special regressor. Anton et al. (2001) use an individual's age as a special regressor in a duration model.

As it turns out, the partial independence assumption provides an identifying restriction which can be applied in contexts which are economically interesting. In his recent contribution, Lewbel (2000) constructs an estimator of β by combining partial independence, with uncorrelated instruments and large-support assumptions. However, the scope of this method and whether it provides (over) identification of β is unclear. The next section describes the class of binary phenomena that may actually be analysed through this set-up.

2.1. The equivalence result

Consider $(\beta, F_\varepsilon(.\,|\,x,z))$, a latent structure satisfying partial independence, support and moment conditions (L.1–L.3) and denote $\Pr(y=1\,|\,v,x,z)$ the conditional distribution generated by $(\beta, F_\varepsilon(.\,|\,x,z))$ through the binary transformation (LV). The following lemma shows that this conditional distribution necessarily increases from 0 to 1 when v varies over its support.

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* $[v_L, v_H]$ *such that*:

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

Proof. See Appendix A.1.

Condition (NP.1) is a direct consequence of the fact that v is an exogenous regressor positively affecting the propensity of success ($y = 1$). As for condition (NP.2), it is a direct consequence of the large-support hypothesis, which implies that the propensity of success of persons with the lowest (largest) v is always negative (positive) regardless of their unobserved and observed characteristics.

Summing up, if we denote:

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

and

$$\mathscr{M}_{L} = \{\Pr(y = 1 \mid v, x, z) \text{ generated through (LV)} \\ \text{by some } (\beta, F_{\varepsilon}(. \mid x, z)) \text{ satisfying } (L.1-L.3)\}$$

we have just proved that $\mathscr{M}_L \subset \mathscr{M}_{NP}$. Let us analyse the condition under which $\mathscr{M}_{NP} \subset \mathscr{M}_L$.

Lemma 2. *Let* $\Pr(y = 1 \mid v, x, z)$ *be a conditional probability satisfying monotonicity* (NP.1) *and support* (NP.2) *conditions. Any latent model* $(\beta, F_{\varepsilon}(. \mid x, z))$ *satisfying* (L.1–L.3) *and generating* $\Pr(y = 1 \mid v, x, z)$ *through transformation* (LV) *necessarily satisfies the following moment conditions*:

$$\mathrm{E}(z'x).\beta = \mathrm{E}(z'\tilde{y}), \tag{1}$$

where

$$\tilde{y} = \frac{y - \mathbf{1}(v>0)}{f(v \mid x, z)} \tag{2}$$

is the transform of y *introduced by* Lewbel (2000).

Proof. See Appendix A.2.

When there exist as many instruments as explanatory variables ($q = p$), condition (1) defines a unique parameter β and $\mathscr{M}_{NP} \subset \mathscr{M}_L$. In contrast, when there are more instruments than explanatory variables ($q > p$), it can happen that condition (1) has no solution, as in the usual linear model. To address this issue, we have to complete the setting by the following regularity condition:

($R.v$) The distribution of $\omega = (y, v, x, z)$ is such that condition (1) has a solution.

Under ($R.v$), this solution is unique, $\mathscr{M}_{NP} \subset \mathscr{M}_L$ and, taken together, Lemmas 1 and 2 prove our first basic result:

Theorem 3. *Under regularity conditions* (*Ri-Rv*), *the set of latent models defined by conditions* (*L*.1–*L*.3) *and transformation* (LV) *is one-to-one with the set of conditional probabilities satisfying* (*NP*.1–*NP*.2).

Put differently, any statistical model in $\mathcal{M}_{NP}$ is generated by a unique structural model in $\mathcal{M}_L$ and reciprocally, any structural model in $\mathcal{M}_L$ generates a unique statistical distribution of the binary outcome satisfying (*NP*.1–*NP*.2). Conditional on (*NP*.1–*NP*.2), the parameters of interest in the structural model are just identified, in that they are defined by a unique function of the joint distribution of the data.[2]

2.2. Discussion

Theorem 3 sheds some light on the deep nature of the partial independence hypothesis (*L*.1). This theorem shows that combining (*L*.1) with a large-support assumption such as (*L*.2) and an uncorrelated-error condition such as (*L*.3) is exactly what is needed to overcome Manski's underidentification result, according to which an uncorrelated-error restriction (i.e., $\mathrm{E}(x'\varepsilon) = \mathrm{E}(v'\varepsilon) = 0$) or even a mean-independence restriction (i.e., $\mathrm{E}(\varepsilon \mid x, v) = 0$) is not sufficient for identifying β no matter what conditions on the support of (v, x) are adopted (see 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 3, while adding more than (*L*.1) to (*L*.2) and (*L*.3) would generate testable overidentifying restrictions.[3]

It should be noted that the partial independence assumption is closely connected with the control function assumption used by Blundell and Powell (2004). Transposing Blundell and Powell's model into our framework involves splitting up regressors $x = (z_1, y_2)$ into a set of exogenous regressors, z_1, and a set of endogenous regressors, y_2. The complete list of instruments comprises v and $z = (z_1, z_2)$ and the model is $y = 1(v + z_1\beta_1 + y_2\beta_2 + \varepsilon > 0)$. There is also an auxiliary first-stage regression where the error term is defined as, $u = y_2 - \mathrm{E}(y_2 \mid v, z)$.

Using these notations, Blundell and Powell's identifying assumption can be written

$$F_\varepsilon(. \mid y_2, z, v) \equiv F_\varepsilon(. \mid u, z, v) = F_\varepsilon(. \mid u),$$

whereas the partial independence assumption used in this paper is

$$F_\varepsilon(. \mid y_2, z, v) = F_\varepsilon(. \mid y_2, z) \equiv F_\varepsilon(. \mid u + \mathrm{E}(y_2 \mid v, z), z).$$

Consequently, Blundell and Powell's model requires a stronger exclusion restriction if $\mathrm{E}(y_2 \mid v, z)$ does not depend on v. However, generally the models are not nested.[4] Other

[2]Apart from over-identifying restrictions provided by supernumerary instruments, Powell (1994) proposes a definition of semi-parametric (versus non-parametric) modelling that exploits the distinction between just-identification and over-identification. According to Powell (1994), a model can be said to be "non-parametric" whenever the parameters are just identified, i.e., defined by a unique function of the joint distribution of the data. In that specific sense, our model is non-parametric.

[3]For instance, strengthening (*L*.3) into a mean-independence restriction $\mathrm{E}(\varepsilon \mid z) = 0$ generates additional restrictions. We conjecture that most results derived in this paper can be extended to this case in the way that they are in the usual linear model.

[4]Although as pointed out by a referee, the partial independence assumption could be rewritten so that ε and v are independent conditionally on y_2, z_1 and $\mathrm{E}(y_2 \mid z, v)$ (instead of z_2). The partial independance assumption is then strictly weaker than the assumption used by Blundell and Powell.

differences are that the control function approach requires the endogenous regressor y_2 to be continuous, although it does not require a large-support assumption or that the regressor v should be continuous.

Regarding the comparison with quantile-independence assumptions, it should be noted that quantile independence assumes that one quantile of ε is independent of all covariates, whereas the partial independence assumption used in this paper is equivalent to assuming that all quantiles of ε are independent of one covariate. In this crude sense, both assumptions seem comparably restrictive.

Another difference is that partial independence yields just identification of β while quantile independence imposes testable overidentifying restrictions.[5] Also, the partial independence hypothesis makes it possible to estimate the distribution of the unobserved residuals while the quantile-independence assumption does not. This property may be of particular interest for evaluating the impact of the covariates on the probability of observing $y = 1$ (Lewbel et al., 2001). The price to pay is that partial independence 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 (L.2) which implies that $v + x\beta$ takes both positive and negative value for any x when v varies over its support. Lastly, it should be noted that the endogeneity of covariates can be also accommodated in a quantile independence setting (Hong and Tamer, 2003) so that the two methods are on par in this respect.

3. Unrestricted support and identification

Generally speaking, the main potential obstacle to empirical application of the latent model under consideration is not so much the partial independence assumption as such, but the accompanying large-support assumption. As shown by the equivalence result, this assumption restricts the domain of application of the latent model to binary phenomena such that the probability of success varies from 0 to 1 when v varies over its support.[6]

The identification of structural parameters in binary choice models can be lost when the support of the regressors is not sufficiently rich. This is true when using the index sufficiency but all regressors are discrete or when using the quantile-independence models (Horowitz, 1998) and it remains true under the partial independence hypothesis. Thus, assuming the existence of large-support, continuous regressors is not uncommon in the literature on semiparametric limited-dependent variable models (see e.g., Han, 1987; Cavanagh and Sherman, 1998; Manski, 1975, 1975; Horowitz, 1992).

In this section, we maintain (L.1) (and (L.3)), but we relax the large-support assumption. In such a case, the conditional distribution $\Pr(y = 1 \mid v, x, z)$ obtained through (LV) still

[5] Namely, the hypersurface in the space of covariates describing the conditional quantile of the dependent variable is linear.

[6] However, the large-support assumption is quite natural in many settings. For instance, it seems a plausible assumption 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 labour market (for male workers), age is the most obvious candidate as the special continuous regressor, v, and the large-support restriction is satisfied. For instance, sufficiently young children have never attended primary-school and sufficiently old children have all attended primary school (in developed countries at least).

satisfies (NP.1), but does not satisfy (NP.2) anymore. More specifically, in the absence of any restrictions on the support of ε, the only restriction on $\Pr(y = 1 \mid v, x, z)$ is that it should be zero or one when v is $\pm\infty$. If $v_H = +\infty$ and $\Pr(y = 1 \mid v_H, x, z) < 1$ (or if $v_L = -\infty$ and $\Pr(y = 1 \mid v_L, x, z) > 0$), no latent variable model in $\mathscr{M}_L$ can lead to the conditional probability function. It is the reason why we shall exclude this case by setting:

$$(NP.2')\lim_{v\to+\infty} \Pr(y = 1 \mid v, x, z) = 1 \quad \text{and} \quad \lim_{v\to-\infty} \Pr(y = 1 \mid v, x, z) = 0.$$

Observe also that when the support of v coincides with the real line (i.e., $v_H = +\infty$ 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 in this section.

In the remainder of the section, we consider statistical models satisfying (NP.1)–($NP.2'$) and we seek the conditions under which the parameters of the latent model β are identified. We first show that the combination of assumptions of partial independence (L.1) and uncorrelated-error (L.3) alone is not sufficient for identifying β. Secondly, we present a set of additional identifying restrictions leading to exact identification. It is shown that it preserves the validity of Lewbel's estimation procedure.

3.1. The necessity of additional identifying restrictions

Consider a conditional distribution $\Pr(y = 1 \mid v, x, z)$ satisfying the monotonicity condition (NP.1) and condition ($NP.2'$). Assume that this conditional probability is the image of a latent model $(\beta, F_\varepsilon(. \mid x, z))$ which satisfies partial independence (L.1). By definition, for any v in $]v_L, v_H[$, we have

$$\Pr(y = 1 \mid v, x, z) = \int_{v+x\beta+\varepsilon>0,\varepsilon\in\Omega_\varepsilon(x,z)} f_\varepsilon(\varepsilon \mid x, z)\,\mathrm{d}\varepsilon$$

and thus:

$$\Pr(y = 1 \mid v, x, z) - \Pr(y = 1 \mid v_L, x, z)$$
$$= \int_{-(v+x\beta)}^{-(v_L+x\beta)} f_\varepsilon(\varepsilon \mid x, z)\,\mathrm{d}\varepsilon = F_\varepsilon(-(v_L + x\beta) \mid x, z) - F_\varepsilon(-(v + x\beta) \mid x, z).$$

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

$$f_\varepsilon(\varepsilon \mid x, z) = \frac{\partial}{\partial v}\Pr(y = 1 \mid v, x, z)|_{v=-(x\beta+\varepsilon)}. \tag{3}$$

In contrast to the large-support case, the support of ε (conditional on x and z) is not necessarily included in $]-(v_H + x\beta), -(v_L + x\beta)[$ and $f_\varepsilon(\varepsilon \mid x, z)$ 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 in $B(x)$ are the following:[7]

$$\Pr\{\varepsilon \leqslant -(v_H + x\beta) \mid x, z\} = 1 - \Pr(y = 1 \mid v_H, x, z),$$
$$\Pr\{\varepsilon > -(v_L + x\beta) \mid x, z\} = \Pr(y = 1 \mid v_L, x, z). \tag{4}$$

Hence, any latent model $(\beta, F_\varepsilon(. \mid x, z))$ satisfying (L.1) and generating function $\Pr(y =$

[7]As the distribution of ε is absolutely continuous, the use of large or strict inequalities is equivalent.

$1 \mid v, x, z)$ through (LV) necessarily satisfies (3) and (4). Conversely, any latent model $(\beta, F_\varepsilon(. \mid x, z))$ satisfying (*L*.1) and (3) and (4) generate function $\Pr(y = 1 \mid v, x, z)$ through (LV). From this, it follows clearly that the partial independence hypothesis is not sufficient for complete identification of $F_\varepsilon(. \mid x, z)$, even when β is known. Under (*L*.1), the only restrictions on $F_\varepsilon(. \mid x, z)$ when ε is in $B(x)$ are given by (4), which means that the distribution of ε conditional on $\varepsilon \in B(x)$ is left completely unidentified.

In settings like index sufficiency models, the identification of the distribution of error terms is not a necessary condition for identifying parameter β. However, it is a necessary condition in the present setting since—when used as an identifying restriction—(*L*.3) is a moment condition which uses the distribution of random shocks over its whole support. Specifically, if S represents the set of observationally equivalent values of the parameter,

$$\begin{aligned} S = \{\beta \in \mathbb{R}^p \mid \exists F_\varepsilon(. \mid x, z) &\text{ satisfying } (L.1) \text{ and } (L.3) \\ &\text{s.t. } (\beta, F_\varepsilon(. \mid x, z)) \text{ generates } \Pr(y = 1 \mid v, x, z)\} \end{aligned}$$

the next proposition states that the size of S is unbounded. It contains an infinite number of elements which value may be chosen arbitrarily differently from the value that β would take if the large-support assumption were true.

Proposition 4. *Consider* $\Pr(y = 1 \mid v, x, z)$ *satisfying* (*NP*1), (*NP*2)$'$, *but not* (*NP*2). *For any* $\lambda_0 > 0$, *there exists a latent model* $(\beta, F_\varepsilon(. \mid x, z))$ *such that*

(i) $(\beta, F_\varepsilon(. \mid x, z))$ *satisfies* (*L*1), (*L*3) *and generates* $\Pr(y = 1 \mid v, x, z)$ *through* (LV).
(ii) $(\beta - \beta_0)'(\beta - \beta_0) \geqslant \lambda_0$,
where β_0 *is the value associated with the moment condition* $\mathrm{E}(z'x).\beta_0 = \mathrm{E}(z'\tilde{y})$.

Proof. See Appendix B.1.

To probe the meaning of Proposition 4, let us interpret $x\beta + \varepsilon$ as the willingness to pay for an object, v as (minus) the unit price of this object and y as the decision to buy it. When the support of v is not large, the most extreme values of the willingness to pay $x\beta + \varepsilon$ are such that we cannot observe prices $(-v)$ which separate individuals whose willingness to pay is larger than the price (they buy the object) from those whose willingness to pay is smaller (they do not buy). This is the reason why the tails of the distribution of $x\beta + \varepsilon$ are not identified and Proposition 4 shows that without additional assumptions on these tails, β is not identified.

In the remainder of this section, we explore an alternative route for restoring identification by the way of an additional assumption on the tails of the distribution of ε (i.e., ε in $B(x)$).

3.2. Generalizing the special-regressor estimator

It should be noted that the set $B(x)$ is the union of two subsets, $B_\mathrm{F}(x) = \{\varepsilon : \varepsilon + v_H + x\beta < 0\}$ and $B_\mathrm{S}(x) = \{\varepsilon : \varepsilon + v_L + x\beta > 0\}$. An individual in $B_\mathrm{F}(x)$ always responds $y = 0$, even when $-v$ is minimum $(-v_H)$. Symmetrically, an individual in $B_\mathrm{S}(x)$ always responds $y = 1$, even when $-v$ is maximum $(-v_L)$. The set $B_\mathrm{F}(x)$ may be interpreted as the subset of certain failure and, $B_\mathrm{S}(x)$, the subset of certain success. By construction, the data do not provide any information on the distribution of the propensities of success in $B_\mathrm{F}(x)$

and $B_S(x)$. The next proposition shows that identification is restored provided that some balance may be assumed between these two distributions.

Proposition 5. *Assume* $v_H < +\infty$, $v_L > -\infty$ *and consider* $\Pr(y = 1 \mid v, x, z)$ *satisfying* $(NP.1)$ *and* $(NP.2')$. *Let* $S \subseteq \mathbb{R}^p$ *be the set of parameters* β *such that there exists a latent model* $(\beta, F_\varepsilon(.\mid x, z))$ *satisfying* $(L.1)$, $(L.3)$ *and generating* G *through* (LV). S *is reduced to a singleton and* $\mathrm{E}(z'x).\beta = \mathrm{E}(z'\tilde{y})$ *if and only if*

$$\mathrm{E}(z' y^*_{v_H} \mathbf{1}\{y^*_{v_H} > 0\}) = \mathrm{E}(z' y^*_{v_L} \mathbf{1}\{y^*_{v_L} > 0\}), \tag{5}$$

where $y^*_{v_L} = (x\beta + v_L + \varepsilon)$ *is the propensity of success for individuals with the smallest* v *and where* $y^*_{v_H} = -(x\beta + v_H + \varepsilon)$ *the propensity of failure for individuals with the largest* v.

Under $(L.1)$, $(L.3)$ *and* (5) *the moment condition* $(\mathrm{E}(z'x).\beta = \mathrm{E}(z'\tilde{y}))$ *provides exact identification of* β.

Proof. See Appendix B.2.

One of the simplest assumptions we can think of which implies this condition, is that propensities of success $y^*_{v_L}$ within the certain-success subset $B_S(x)$ and propensities of failure $y^*_{v_H}$ within the certain-failure subset $B_F(x)$ are identically distributed. If this condition is valid, the special-regressor estimator is unbiased. Alternatively, it is always possible to choose conditional distributions for $y^*_{v_H}$ and $y^*_{v_L}$ when they are positive, such that Eq. (5) is satisfied. It is however impossible to tell from the data whether symmetry of the tails or an alternative restriction verifying (5) is valid. All restrictions on the distribution of ε satisfying (5) are observationally equivalent and all lead to the exact identification of β.

If either v_H or v_L is infinite,[8] condition (5) cannot be satisfied. Let $v_H = +\infty$ (say), then the absence of bias means that $\mathrm{E}(z' y^*_{v_L} \mathbf{1}\{y^*_{v_L} > 0\})$ should be set to zero which is impossible since $\mathrm{E} y^*_{v_L} \mathbf{1}\{y^*_{v_L} > 0\} > 0$. Nevertheless as shown in B.2, the bias may affect the intercept term only.

Proposition 6. *Assume* $v_H = +\infty$, $v_L > -\infty$ *and consider* $\Pr(y = 1 \mid v, x, z)$ *satisfying* $(NP.1)$ *and* $(NP.2')$. *Let* $S \subseteq \mathbb{R}^p$ *be the set of parameters* β *such that there exists a latent model* $(\beta, F_\varepsilon(.\mid x, z))$ *satisfying* $(L.1)$, $(L.3)$ *and generating* $\Pr(y = 1 \mid v, x, z)$ *through* (LV) *and let* $S_1 = \{(\beta_2, \ldots, \beta_p) \in \mathbb{R}^{p-1}$ *s.t* $\beta_1 \in \mathbb{R}, (\beta_1, \ldots, \beta_p) \in B\}$ *where* β_1 *is the intercept coefficient.* S_1 *is reduced to a singleton if there exists a constant* α *independent from* z *such that*

$$\mathrm{E}(y^*_{v_L} \mathbf{1}\{y^*_{v_L} > 0\} \mid z) = \alpha, \tag{6}$$

where $y^*_{v_L} = (x\beta + v_L + \varepsilon)$ *represents the propensity of success for individuals with the lowest possible* v. *Under* $(L.1)$, $(L.3)$ *and* (5) *the moment condition* $(\mathrm{E}(z'x).\beta = \mathrm{E}(z'\tilde{y}))$ *provides exact identification of* β *apart from the intercept coefficient.*

Proof. See Appendix B.2.

The long version of this paper ~~(Magnac and Maurin, 2003)~~ reports Monte-Carlo experiments which show that the estimator developed in this sub-section (i.e., when $(L.2)$ is not satisfied) performs quite well in medium-sized samples.

[8] But not both. If both v_H and v_L are infinite, we are back to the case described as restricted support (!), condition $(L.2)$. Theorem 3 applies.

4. Information and asymptotic properties

Identification is not the only concern when choosing among different estimation methods, information is as well. In this section, we establish the semi-parametric efficiency bound of regular estimators of parameter β under partial-independence and uncorrelated instruments. The bound is valid regardless of whether the support of v is large or not, provided the regularity conditions (R) hold true.

Before moving on to the proof of these results, it should be noted that they correspond to a different setting and are different from the seminal results in Cosslett (1987). Specifically, he assumes that the error terms have a zero median and are independent of the regressors. Under these assumptions, he derives the semi-parametric efficiency bound of the parameters, except the intercept, which cannot be estimated at a root-n rate. As we use a moment condition instead of a median condition on the error term, a root-n consistency result for all estimators can be obtained.

In our set-up, the only identifying restriction is given by the moment (1) and one possible source of difficulty 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., $\mathrm{E}(z'[\tilde{y} - x\beta]) = 0$) is defined. Given this relationship, the general framework investigated by Chamberlain (1992) needs to be amended and the semi-parametric efficiency of the estimators has to be checked by hand.

The special-regressor estimator proposed by Lewbel (2000) is constructed by using the empirical counterparts of the moments in Eq. (1). Under regularity conditions provided by Lewbel (2000), this estimator is root-n consistent and asymptotically normal. Our derivation of the efficiency bound shows that it is not possible to construct an estimator which is more efficient than the special-regressor estimator under assumptions $(L.1)$, $(L.3)$ and a large or restricted support assumption. The specific moment estimator of β proposed by Lewbel is semi-parametrically efficient.

In the large-support case, the regularity conditions under which the special-regressor estimator is root-n consistent, and asymptotically normal, are given in Lewbel (2000), Appendix B, Conditions B1–B6 and Condition B7 or B7′ depending on whether the support of ε is bounded. It is easy to check that Conditions B1–B6 and Condition B7′ remain applicable[9] when the large-support assumption does not hold and that—under these conditions—the proof of root-n consistency and asymptotic normality still holds true too. In particular, Condition B7' imposes conditions either on the rate at which $\Pr(y = 1 \mid v, x, z)$ tends to 0 or 1 when $|v| \to \infty$ or on the support of v which are easy to satisfy when the large-support assumption does not hold true any more.[10]

In the remainder of this section, we establish the efficiency bound. Then, we show that it is more efficient to use an estimate of the conditional density function when constructing $\tilde{y}$ rather than the true value of the density when the latter is known.

[9] In contrast, condition (B7) is not applicable when $(NP2)$ does not hold true. As a matter of fact, it assumes that the support of v is large.

[10] Condition B7' ensures that asymptotic trimming leads to an asymptotically equivalent estimator (see also Lewbel, 1998, Appendix B). As the proof of asymptotic properties is only a little more than the original we do not repeat it here.

4.1. The semiparametric efficiency bound

The estimate is based on the unconditional moment conditions:

$$E[m(y, v, x, z; \beta_0)] = 0, \tag{7}$$

where the function of interest is

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

From regularity conditions $(R.i - v)$,

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

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

Observe 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 be directly applied because the unknown non-parametric component is also a density function with respect to which the unconditional moment restriction is taken.

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 the first-step estimate of π_0. In the first step, the unconditional moment restriction that we consider is

$$E(\tilde{g}(y, v, x, z; \pi_0)) = E(z'\tilde{y} - \pi_0) = 0. \tag{8}$$

The efficiency bound and variance–covariance matrices for β_0 are then derived 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

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

The bound for V_π is described by the following result.

Proposition 7. *The semi-parametric efficiency bound for estimating π_0 is*

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

Proof. See Appendix C.1.

For the paper to be self-contained, Appendix C.2 provides the variance–covariance of Lewbel's estimator (as derived by Lewbel, 2000) and proves that it actually attains the previous bound. Under $(L.1–L.3)$, it is not possible to be more efficient than Lewbel's estimator.

It should be emphasized that the special-regressor estimator remains efficient when the large-support hypothesis $(L.2)$ is replaced by a symmetry assumption such as (5). As a matter of fact, the derivation of the semi-parametric efficiency bound and of the variance–covariance of estimator does 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 5 or 6.

If one is ready to lose some efficiency then—in the asymmetric case described by Proposition 6—one can always use the symmetrical trimming proposed by Powell (1986).

4.2. Plugging-in the true or estimated conditional density?

In this section, we assume that the conditional density $f(v\,|\,x,z)$ 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 two different transformations, $\tilde{y} = (y - I(v>0))/f(v\,|\,x,z)$ or $(y - I(v>0))/\hat{f}(v\,|\,x,z)$ when constructing the linear regression that leads to the estimation of β. Here, $f(v\,|\,x,z)$ is the true distribution and $\widehat{f}(v\,|\,x,z)$ is an estimate of $f(v\,|\,x,z)$. It was conjectured by Lewbel (2000) (and confirmed by Monte-Carlo experiments) 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 8. *The estimate of* π_0 *defined by the unconditional moment condition* (8) (*i.e.*, $\mathrm{E}(z'\tilde{y} - \pi_0) = 0$) *has a strictly smaller variance when the estimated* $\widehat{f}(v\,|\,x,z)$ *is used to transform the dependent variable than when the true density is used.*

Proof. See Appendix C.3.

Replacing the nuisance parameter—the conditional density—by an estimate is more efficient than replacing it by its true value. Hirano et al. (2003) report similar results in the context of treatment models where the nuisance parameter is the inverse of a propensity score, in a set of moment restrictions. They show that using an estimate of the score leads to more efficient estimation of treatment parameters than using the true score. They interpret the estimator with the estimated score as an empirical likelihood estimator where the information about the nuisance function has been efficiently incorporated. Theorem 5 can also be understood by using broadly similar arguments to those presented by Crépon et al. (1998). Consider two sets of moment conditions. The first set depends on the parameters of interest and the nuisance parameters while the second set of moment conditions depend on the nuisance parameters only. The efficient GMM estimates can be derived from the first set of conditions when the nuisance parameters are replaced by their estimated values using the second set of conditions. In contrast, GMM estimates are not generically efficient when the parameters are replaced by their true values.

5. Extensions

Lewbel (1998, 2000, 2003) uses the special-regressor hypothesis to estimate the structural parameters of other linear latent variable models, $y = L(x\beta + \varepsilon)$, 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 3 can be extended to these models. In some interesting cases the answer is positive. In other cases, the special regressor setting imposes testable restrictions on the set of statistical phenomena that are generated by the latent structure.

To illustrate the generalization of Theorem 3, we consider the most straightforward extension of binary responses which are ordered choice models. Assume that the support of y is now $S_y = \{0, 1, \ldots, K\}$ $(K \geqslant 1)$. We consider two definitions of ordered choice models and discuss each in turn. In the first one, each individual is defined by an ordered set of propensities (i.e., $y_1^*, ., y_K^*$) and his/her response ($y \in \{0, 1, \ldots, K\}$) depends 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 $\alpha_k(v)$. A straightforward extension of Theorem 3 only holds in the first case, whereas structural parameters are overidentified in the second model.

5.1. *Ordered choices: first model*

Consider the following definition for latent ordered choice models.

Definition 9. 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

$$y = \sum_{k=1}^{K} kI(v + y_k^* > 0, v + y_{k+1}^* \leqslant 0)$$

$$= \sum_{k=1}^{K} kI(-y_k^* < v \leqslant -y_{k+1}^*). \qquad \text{(LV1)}$$

We consider linear latent models such as

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

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

This model is a straightforward generalization of (LV). When $K = 1$, the two models coincide. Such an ordered choice model may typically be used for analysing consumer behaviour. Suppose that 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^*, stand for the willingness to pay for an additional unit of this good when the number of units bought is $k - 1$. If marginal utility is decreasing, then the marginal-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 relation between marginal utility and quantity purchased.

One of the interesting features 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 an indicator of purchase (any quantity), y_2 is an indicator of 2 or more units purchased and so on y_k is an indicator showing that k or more units were purchased:

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

Reciprocally:

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

Let $\mathscr{M}^*_{\text{OC}}$ be the set of latent ordered discrete choice models where elements $\{(\beta_k, F_{\varepsilon_k}(.\,|\,x,z)), k = 1, \ldots, K\}$ satisfy partial independence, support and moment conditions (L.1–L.3) and the additional inequality restrictions across alternatives:

$$y_k^* = x\beta_k + \varepsilon_k > x\beta_{k+1} + \varepsilon_{k+1} = y_{k+1}^*. \tag{9}$$

These inequalities translate into restrictions 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:

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

The consequences in terms of non-parametric predictions are now straightforward. They consist of (NP.1) and (NP.2) for any choice k. Inequalities (9) in the latent model translate into

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

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

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

which is a sensible assumption in most cases. For instance, the probability of buying more than k units is decreasing with k. These 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 the latter is underidentified. Only the marginal distributions are identified.

We can now summarize these results. Let the set $\mathscr{M}^*_{\text{LOC}}$ of latent ordered models be given by parameters $(\beta_1, ., \beta_K) \in \mathbb{R}^K$, distribution functions $(f_1(\varepsilon_1 \,|\, x, z), ., f_K(\varepsilon_K \,|\, x, z)) \in \mathscr{D}^K$, 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 $\mathscr{M}_{NPOC}$ given by:

$$\mathscr{M}_{NPOC} = \mathscr{M}_{NP}(y_1) \times \cdots \times \mathscr{M}_{NP}(y_K)$$

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

Theorem 10. *$\mathscr{M}^*_{\text{LOC}}$ is one-to-one with $\mathscr{M}_{NPOC}$.*

5.2. Ordered choices: second model

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

$$y = \sum_{k=1}^{K} kI(\alpha_k(v) < x\beta + \varepsilon \leqslant \alpha_{k+1}(v)), \tag{LV2}$$

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

$$\alpha_1(v) = -v \leqslant \alpha_2(v) \leqslant \cdots \leqslant \alpha_k(v) \leqslant \alpha_{K+1}(v) = +\infty, \tag{10}$$

while ε satisfies (L.1)–(L.3).

This model is also a direct generalization of (LV). When $K=1$, the two models coincide. $x\beta+\varepsilon$ may be interpreted as a propensity to respond as in (LV), but now the response has 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 intensity of the response.

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 $\alpha_k(v)$ thresholds represent the set of thresholds (plausibly dependent on v being the birthdate within the year) imposed by the educational system for deciding who should be held back ($y=0$), who should be on time ($y=1$) and who should be ahead ($y=2$) at school.[11]

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

Let us denote $G_0(v,x,z)=\mathrm{P}(y=0\,|\,v,x,z)$. By definition, $-G_0$ belongs to $\mathcal{M}^*_{NP}$. Thus, using Theorem 3, we can exactly identify the parameter of interest β and the distribution of the error term ε. In particular, we necessarily have $f_\varepsilon(.\,|\,x,z)=(\partial G_0/\partial v)(-(x\beta+\varepsilon),x,z)$. For any $k\geqslant 1$, define now $G_k(v,x,z)=\mathrm{P}(y\leqslant k\,|\,v,x,z)$. We have

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

It therefore yields

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

Thus $F(y\,|\,v,x,z)$ is the image of an element of $\mathcal{M}^*_{\text{LOC2}}$ 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 analysed as a structural ordered choice model that satisfies the partial independence hypothesis only if $G_0^{-1}\circ G_k$ does not depend on x and z, which is a testable assumption. Note finally that the inequalities described by (10) translate into the same inequalities in the 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 thresholds (i.e., $\alpha_k(v)-\alpha_0(v)=\gamma_k$) 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 which can be analysed with such models.

[11] Maurin (2002) uses the binary approach to estimate the probability to be held back using v = day-of-birth within the year as a special regressor and interpreting $x\beta+\varepsilon$ 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.

6. Conclusion

The first contribution of this paper is to characterize the conditions under which the identifying assumptions proposed by Lewbel (2000) are justified: $\mathrm{E}(y \mid v, x)$ is monotone in v and varies from 0 to 1 when v varies over its support. Second, it is shown that the uncorrelated-error, partial independence and large-support assumptions lead to the exact identification of the structural parameters of the binary response model. We also prove that the large-support assumption—which might be unadapted in some instances—can be replaced by an alternative credible restriction which is the conditional symmetry of the tails of the error distribution. Furthermore, we show that Lewbel's moment estimator attains the semi-parametric efficiency bound in the corresponding class of latent models. We propose an extension to ordered choice models. All in all, Lewbel's moment estimator is shown to be consistent in a fairly wide class of binary choice models. This class includes all monotone binary data where the probability of success varies in an interval which is strictly included in [0, 1].

It would be interesting to extend our results to other settings, such as the analyses of truncated regressions (Khan and Lewbel, 2003), treatment effects (Lewbel, 2003) or panel data (Honoré and Lewbel, 2002). We are currently exploring another route by relaxing the assumption that partial independence holds with respect to a regressor which is continuous (Magnec and Maurin, 2004). We consider that v is discrete or has been made discrete and show that bounds of a convex set containing β are identified.

Acknowledgements

We thank the editors, two anonymous referees, Magali Beffy, Pierre Dubois, Guy Laroque, Oliver Linton, Sandra McNally, Jean-Marc Robin, Bernard Salanié and particularly Arthur Lewbel for very helpful discussions. We thank participants at the ESRC conference in Bristol (2003), at the $(\mathrm{EC})^2$ conference in London (2003), ESEM in Stockhom (2003) , and at seminars of CREST, University of Toulouse, University of Concordia, Laval University, University Carlos III, Princeton, Boston College and GREQAM for their helpful comments. Most of the paper was written when Thierry Magnac was at INRA-Paris Jourdan and CREST-INSEE. The usual disclaimers apply.

Appendix A. Proofs of Section 2

A.1. Proof of Lemma 1

Write:

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

As $\mathrm{d}F_\varepsilon(\varepsilon \mid x, z) \geqslant 0$ and F_ε is absolutely continuous, the first conclusion follows.

Second, for almost any (x, z), the support of $-x\beta - \varepsilon$ is a subset of $]v_L, v_H[$ that we denote $]v_l(x, z), v_h(x, z)[$. Suppose first that both bounds are finite. We have for all $\varepsilon \in \Omega_\varepsilon(x, z)$:

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

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

$$v_l(x,z) + x\beta + \varepsilon < 0, \quad v_h(x,z) + x\beta + \varepsilon > 0$$

The second conclusion follows. If bounds are infinite then the expressions in the lemma should be replaced by suitable limits.

A.2. Proof of Lemma 2

Consider $G(v,x,z) = \Pr(y = 1 \mid v,x,z)$ 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* $\mathscr{M}_L^*$ such that $G(v,x,z)$ is its image through the transformation (LV). Define the support of the random variable ε as

$$\Omega_\varepsilon(x,z) =]-(v_h(x,z)+x\beta), -(v_l(x,z)+x\beta)[\tag{11}$$

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

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

which implies for any $\varepsilon \in \Omega_\varepsilon(x,z)$ that

$$f_\varepsilon(\varepsilon \mid x,z) = \frac{\partial G}{\partial v}(-(x\beta+\varepsilon),x,z). \tag{12}$$

The $\partial G/\partial v$ 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:

$$\begin{aligned} 0 &= \mathrm{E}(z'\varepsilon) \\ &= \mathrm{E}_{x,z}\left(z' \int \varepsilon f_\varepsilon(\varepsilon\mid x,z)\,\mathrm{d}\varepsilon\right) \\ &= -\mathrm{E}_{x,z}\left(z' \int (x\beta+v)\frac{\partial G}{\partial v}\,\mathrm{d}v\right) \\ &= -\mathrm{E}(z'x)\beta - \mathrm{E}_{x,z}\left(z' \int v\frac{\partial G}{\partial v}\,\mathrm{d}v\right), \end{aligned} \tag{13}$$

where the notation $\mathrm{E}_{x,z}$ 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 $R.iii$, $\mathrm{E}(z'x)$ is of rank equal to the dimension of β. The previous equation therefore uniquely defines β in the usual sense when some linear restrictions are overidentifying (rigorously defined in $R.v$).

Thus if $(\beta, F_\varepsilon(.\mid x,z))$ exists, it is defined by (11)–(13). Reciprocally, consider $(\beta, F_\varepsilon(.\mid x,z))$ in $\mathscr{M}_L^*$ which satisfies (11)–(13). Its image through (LV) is $G(v,x,z)$.

Finally, we have

$$\int v \frac{\partial G}{\partial v}\,\mathrm{d}v = \int_0^{v_H} v \frac{\partial G}{\partial v}\,\mathrm{d}v + \int_{v_L}^{0} v \frac{\partial G}{\partial v}\,\mathrm{d}v$$
$$= [v(G(v,x,z)-1)]_0^{v_H} - \int_0^{v_H} (G(v,x,z)-1)\,\mathrm{d}v$$
$$+ [vG(v,x,z)]_{v_L}^{0} - \int_{v_L}^{0} G(v,x,z)\,\mathrm{d}v$$
$$= -\int_{v_L}^{v_H} (G(v,x,z) - \mathbf{1}(v>0))\,\mathrm{d}v$$
$$= -\int_{v_L}^{v_H} (\mathrm{E}(y\,|\,v,x,z) - \mathbf{1}(v>0))\,\mathrm{d}v$$
$$= -\int_{v_L}^{v_H} \mathrm{E}(\tilde{y}\,|\,v,x,z).\mathrm{d}F_v(v\,|\,x,z) = -\mathrm{E}(\tilde{y}\,|\,x,z)$$

and therefore (also Lewbel, 2000, p. 115):

$$\mathrm{E}_{x,z}\left(z' \int v \frac{\partial G}{\partial v}\,\mathrm{d}v\right) = -\mathrm{E}(z'\tilde{y})$$

which completes the proof.

Appendix B. Proofs of Section 3

B.1. Proposition 4

Consider $G(v,x,z) = \Pr(y=1\,|\,v,x,z)$ satisfying $NP1$, $NP2'$, but not $NP2$. The purpose is to show that G may be generated by a latent model with an arbitrarily large β .

Fix β. Chose $F_\varepsilon(.\,|\,x,z)$ satisfying partial independence and (3) and (4). By construction, the latent model $(\beta, F_\varepsilon(.\,|\,x,z))$ satisfies (L.1) and generates $G(v,x,z)$ through (LV). The only remaining restriction on β is given by the moment condition (L.3):

$$0 = \mathrm{E}(z'\varepsilon) = \mathrm{E}_{x,z}\left(z' \int \varepsilon\,\mathrm{d}F(\varepsilon\,|\,x,z)\right)$$
$$= \mathrm{E}\left(z' \int_{\varepsilon\in B(x)} \varepsilon\,\mathrm{d}F(\varepsilon\,|\,x,z)\right) + \mathrm{E}\left(z' \int_{-(v_H+x\beta)}^{-(v_L+x\beta)} \varepsilon\,\mathrm{d}F(\varepsilon\,|\,x,z)\right) \tag{14}$$

Thus, using the fact that $f_\varepsilon(\varepsilon\,|\,x,z) = (\partial G/\partial v)(-\varepsilon - x\beta)$ for $\varepsilon \in]-(v_H+x\beta), -(v_L+x\beta)[$ we have

$$0 = \mathrm{E}\left(z' \int_{\varepsilon\in B(x)} \varepsilon\,\mathrm{d}F(\varepsilon\,|\,x,z)\right) - \mathrm{E}\left(z' \int_{v_L}^{v_H} (x\beta + v)\frac{\partial G}{\partial v}\,\mathrm{d}v\right)$$
$$= \mathrm{E}(z'\varepsilon\mathbf{1}\{\varepsilon \in B(x)\}) - \mathrm{E}\left(z'x\beta \int_{v_L}^{v_H} \frac{\partial G}{\partial v}\,\mathrm{d}v\right) - \mathrm{E}\left(z' \int_{v_L}^{v_H} v\frac{\partial G}{\partial v}\,\mathrm{d}v\right). \tag{15}$$

The last term can be expressed as in the proof of Lemma A.2:

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

where

$$\begin{aligned}
b(v_H,v_L,x,z) &= -([v(G(v,x,z)-1)]_0^{v_H} + +[vG(v,x,z)]_{v_L}^{0}) \\
&= v_H(1-G(v_H,x,z)) + v_L G(v_L,x,z)
\end{aligned}$$

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 Eq. (14) can be written as:

$$\begin{aligned}
0 &= \mathrm{E}(z'\varepsilon\mathbf{1}\{\varepsilon\in B(x)\}) - \mathrm{E}(z'x\{G(v_H,x,z)-G(v_L,x,z)\})\beta \\
&\quad + \mathrm{E}(z'b(v_H,v_L,x,z)) + \mathrm{E}(z'\tilde{y}) \\
&= \mathrm{E}(z'\varepsilon\mathbf{1}\{\varepsilon\in B(x)\}) + \mathrm{E}(z'x\{1-G(v_H,x,z)+G(v_L,x,z)\})\beta \\
&\quad + \mathrm{E}(z'b(v_H,v_L,x,z)) \\
&\quad - \mathrm{E}(z'x)\beta + \mathrm{E}(z'\tilde{y}).
\end{aligned} \tag{16}$$

It should be noted that if the support condition (*NP*.2) were true, 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 (16) would give back Lewbel's moment condition (i.e., $\mathrm{E}(z'x)\beta=\mathrm{E}(z'\tilde{y})$).

Given that (*NP*.2) does not hold, either v_L or v_H are finite. Suppose that $v_H<\infty$ so that $-(v_H+x\beta)-t_0(x,z)$ belongs to $B(x)$ for any measurable function $t_0(x,z)\geqslant\eta>0$. Choose the conditional distribution of ε in $B(x)$ such that there is a mass $1-G(v_H,x,z)$ in a small neighbourhood of $-(v_H+x\beta)-t_0(x,z)$ and a mass $1-G(v_L,x,z)$ in a small neighbourhood of $-(v_L+x\beta)$ included in $B(x)$ (possibly a mass 0 at $-\infty$ because of (*NP*.2′)). As the neighbourhoods can be chosen arbitrarily small, we can consider that all the mass is concentrated at two points in $B(x)$, $-(v_H+x\beta)-t_0(x,z)$ and $-(v_L+x\beta)$. Using this specific distribution of ε, (16) may be rewritten, after some manipulation,

$$\mathrm{E}(z'x)(\beta-\beta_0) = -\mathrm{E}(z't_0(x,z)),$$

where β_0 is the value of the parameter associated with the moment condition $\mathrm{E}(z'x).\beta_0=\mathrm{E}(z'\tilde{y})$.

As $\mathrm{E}(z'x)$ is full rank (*R.ii*) then, for all λ_0, there exists $t_0(x,z)$ such that $(\beta-\beta_0)'(\beta-\beta_0)\geqslant\lambda_0$ which concludes the proof.

B.2. Propositions 5 and 6

Eq. (16) proves that the special-regressor estimator is biased except if:

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

$\Longleftrightarrow$

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

$\Longleftrightarrow$

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

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 $y^*_{v_L}=x\beta+v_L+\varepsilon$. It proves Proposition 5.

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:

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

then the constant only in β is biased.

Appendix C. Proofs of Section 4

C.1. Proof of Proposition 7

C.1.1. Preliminaries

We begin by introducing some notations and by presenting the main result of Severini and Tripathi (2001). In the following, we will apply this result to derive the efficiency bound for estimating π_0.

Firstly, 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

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

The "structural" parameter of interest is $\pi(\phi_1,\phi_2,\psi)=E(z'\tilde{y})$. The "reduced form" functionals describing the random variable are ϕ_1, ϕ_2, ψ which are assumed to belong to the following sets:

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

$$\Phi_2 = \left\{ \phi_2 \in L^2(S_{x,z}), \int_{S_{x,z}} \phi_2^2(x,z)\,\mathrm{d}x\,\mathrm{d}z = 1, \quad \phi_2^2(x,z) > 0, \right.$$
$$\left. \phi_2^2(x,z) \text{ is bounded} \right\},$$

$$\Psi = \left\{ \psi \in L^2(]v_L, v_H[), \int_{]v_L,v_H[} \psi^2(v \mid x,z)\,\mathrm{d}v = 1, \psi^2(v \mid x,z) > 0, \right.$$
$$\left. \psi^2(v \mid x,z) \text{ is bounded and continuous} \right\},$$

where all assumptions are derived from the regularity conditions (*R.ii* and *iii*).

In the remainder, E will denote $\Phi_1 \times \Phi_2 \times \Psi$ and $lin\,T(\mathrm{E}, (\phi_1^0, \phi_2^0, \psi^0))$ the space tangent to E at the true value $(\phi_1^0, \phi_2^0, \psi^0)$. This tangent space is the smallest linear space which is closed in the L^2-norm and which contains all $(\dot{\phi}_1, \dot{\phi}_2, \dot{\psi}) \in L^2(F_{y \mid v,x,z}.F_{v \mid x,z}.F_{x,z})$ that are tangent to E at $(\phi_1^0, \phi_2^0, \psi^0)$. A vector $\dot{\varphi}$ is said to be tangent to E at $\varphi_0 = (\phi_1^0, \phi_2^0, \psi^0)$ if there exists a $t_0 > 0$ and a curve $t \longrightarrow \varphi_t$ from $[0, t_0]$ into E which reaches $(\phi_1^0, \phi_2^0, \psi^0)$ at $t = 0$ and such that $\dot{\varphi}$ is the slope of φ_t at $t = 0$ (i.e., $\lim_{t \downarrow 0} \| \frac{\varphi_t - \varphi_0}{t} - \dot{\varphi} \|_{L^2} = 0$).

As shown in Severini and Tripathi (2001), $\overline{lin\,T(\mathrm{E}, (\phi_1^0, \phi_2^0, \psi^0))}$ is the product of the following subspaces:

$$\overline{lin\,T(\Phi_1, \phi_1^0)} = \left\{ \dot{\phi}_1 \in L^2(\{0,1\} \times]v_L, v_H[\times S_{x,z}), \right.$$
$$\left. \sum_{y=0,1} \phi_1^0(y \mid v,x,z).\dot{\phi}_1 = 0 \text{ a.e. }]v_L, v_H[\times S_{x,z} \right\},$$

$$\overline{lin\,T(\Phi_2, \phi_2^0)} = \left\{ \dot{\phi}_2 \in L^2(S_{x,z}), \int_{S_{x,z}} \phi_2^0 \dot{\phi}_2\,\mathrm{d}x\,\mathrm{d}z = 0 \right\},$$

$$\overline{lin\,T(\Psi, \psi^0)} = \left\{ \dot{\psi} \in L^2(]v_L, v_H[\times S_{x,z}), \int_{]v_L,v_H[} \psi\psi^0\,\mathrm{d}v = 0 \text{ a.e. } S_{x,z} \right\}.$$

Following Severini and Tripathi (2001), for any $(\dot{\phi}_1, \dot{\phi}_2, \dot{\psi})$ and $(\dot{\phi}_1', \dot{\phi}_2', \dot{\psi}')$ elements of the tangent space, the Fisher information inner product on the tangent space will be denoted $\langle .,. \rangle_F$ (and the corresponding norm $\|.\|_F$) with,

$$\langle (\dot{\phi}_1, \dot{\phi}_2, \dot{\psi}), (\dot{\phi}_1', \dot{\phi}_2', \dot{\psi}') \rangle_F = 4 \sum_{y=0,1} E_{x,z,v}(\dot{\phi}_1 \dot{\phi}_1') + 4E_{x,z}\left(\int_{]v_L,v_H[} \dot{\psi}\dot{\psi}'\,\mathrm{d}v \right)$$
$$+ 4 \int_{S_{x,z}} \dot{\phi}_2 \dot{\phi}_2'\,\mathrm{d}x\,\mathrm{d}z$$

$$\|(\dot{\phi}_1, \dot{\phi}_2, \dot{\psi})\|_F^2 = \langle (\dot{\phi}_1, \dot{\phi}_2, \dot{\psi}), (\dot{\phi}_1, \dot{\phi}_2, \dot{\psi}) \rangle_F.$$

Since the tangent space is a closed subspace of $L^2(F_{y \mid v,x,z}.F_{v \mid x,z}.F_{x,z})$, the tangent space with this inner product is a Hilbert space. Hence, the Riesz–Frechet theorem implies that for any continuous linear functional L on $(lin\,T(\mathrm{E}, (\phi_1^0, \phi_2^0, \psi^0)), \langle .,. \rangle_F)$ there exists a unique

l^* in $\overline{lin\, T(\mathrm{E},(\phi_1^0,\phi_2^0,\psi^0))}$ such that for any $\dot{l}$ in $\overline{lin\, T(\mathrm{E},(\phi_1^0,\phi_2^0,\psi^0))}$, we have $L(\dot{l}) = \langle \dot{l}, l^* \rangle_F$. The l^* vector is called the representer of L .

For any arbitrary $c \in \mathbb{R}^q$,we will consider $\rho : E \to \mathbb{R}$,

$$\rho(\phi_1,\phi_2,\psi) = c'\pi(\phi_1,\phi_2,\psi) = c'.\sum_{y=0,1}\int_{]v_L,v_H[\times S_{x,z}} z'(y-\mathbf{1}(v>0))\phi_1^2\phi_2^2\,\mathrm{d}v\,\mathrm{d}x\,\mathrm{d}z. \quad (17)$$

To simplify the problem, the usual strategy is to first compute the efficiency bound for estimators of the scalar $\rho(\phi_1^0,\phi_2^0,\psi^0)$. As c is arbitrary, it is straightforward to deduce the efficiency bound for estimators of π_0.

To implement this technique, the issue is to prove that ρ is pathwise differentiable at $(\phi_1^0,\phi_2^0,\psi^0)$, to prove that the pathwise derivative of $\rho(\phi_1,\phi_2,\psi)$ at $(\phi_1^0,\phi_2^0,\psi^0)$ (denoted $\nabla\rho_0$) is a continuous linear functional, to find $\phi^* = (\phi_1^*,\phi_2^*,\psi^*)$ the representer of $\nabla\rho_0$ and to compute $\|\phi^*\|_F$. Severini and Tripathi (2001) show that the lower bound for the asymptotic variance of root-n consistent regular estimators of $\rho(\phi_1^0,\phi_2^0,\psi^0)$ is actually $\|(\phi_1^*,\phi_2^*,\psi^*)\|_F$, i.e., the Fisher information norm of $(\phi_1^*,\phi_2^*,\psi^*)$ the representer of $\nabla\rho_0$.

C.1.2. The representer of $\nabla\rho_0$ and the efficiency bound

For some $t_0>0$ let $\varphi(t) : t \longrightarrow (\phi_{1t},\phi_{2t},\psi_t)$ be a curve from $[0,t_0]$ into E such that $\varphi(t)$ reaches $(\phi_1^0,\phi_2^0,\psi^0)$ when $t=0$, and has a tangent vector $\dot{\varphi} = (\dot{\phi}_1,\dot{\phi}_2,\dot{\psi})$ at $(t=0)$. By definition $\dot{\varphi}$ corresponds to the slope of $\varphi(t)$ at $t=0$ (i.e., $\|(\varphi(t)-\varphi(0))/t - \dot{\varphi}\|_{L^2} \longrightarrow 0$ when $t \downarrow 0$).

Consider $\nabla\rho_0$: $\overline{lin\, T(\mathrm{E},(\phi_1^0,\phi_2^0,\psi^0))} \to \mathbb{R}$ with

$$\nabla\rho_0(\dot{\phi}_1,\dot{\phi}_2,\dot{\psi}) = c'\sum_{y=0,1}\int_{]v_L,v_H[\times S_{x,z}} z'(y-\mathbf{1}(v>0))2(\dot{\phi}_1\phi_2^0+\dot{\phi}_2\phi_1^0)\phi_1^0\phi_2^0\,\mathrm{d}v\,\mathrm{d}x\,\mathrm{d}z. \quad (18)$$

By construction $\nabla\rho_0$ is clearly such that $|(\rho(\varphi(t))-\rho(\varphi(0)))/t - \nabla\rho_0(\dot{\varphi})| \longrightarrow 0$ when $t \downarrow 0$ for any $\varphi(t)$. Hence, ρ is pathwise differentiable at $(\phi_1^0,\phi_2^0,\psi^0)$ and its derivative is the linear functional $\nabla\rho_0$.

We now search for the Riesz–representer of $\nabla\rho_0$, i.e., the vector $(\phi_1^*,\phi_2^*,\psi^*)$ in the tangent space such that, for any $(\dot{\phi}_1,\dot{\phi}_2,\dot{\psi})$ in the tangent space.

$$\nabla\rho_0(\dot{\phi}_1,\dot{\phi}_2,\dot{\psi}) = \langle(\dot{\phi}_1,\dot{\phi}_2,\dot{\psi}),(\phi_1^*,\phi_2^*,\psi^*)\rangle_F.$$

First, notice that $\nabla\rho_0(\dot{\phi}_1,\dot{\phi}_2,\dot{\psi})$ can be rewritten

$$\nabla\rho_0(\dot{\phi}_1,\dot{\phi}_2,\dot{\psi}) = 2c'\sum_{y=0,1}\int_{]v_L,v_H[\times S_{x,z}} z'\left[\frac{(y-\mathbf{1}(v>0))}{\psi^2}\psi^2(\phi_2^0)^2\right]\dot{\phi}_1\phi_1^0\,\mathrm{d}v\,\mathrm{d}x\,\mathrm{d}z$$
$$+\,2c'\int_{S_{x,z}}\dot{\phi}_2\phi_2^0\sum_{y=0,1}\int_{]v_L,v_H[}\left[\frac{(y-\mathbf{1}(v>0))}{\psi^2}\psi^2(\phi_1^0)^2\right]\mathrm{d}v\,\mathrm{d}x\,\mathrm{d}z.$$

Hence, we have,

$$\nabla\rho_0(\dot{\phi}_1,\dot{\phi}_2,\dot{\psi}) = 2c'\sum_{y=0,1}\left(\int_{]v_L,v_H[\times S_{x,z}} z'[\tilde{y}\psi^2(\phi_2^0)^2]\,\mathrm{d}v\,\mathrm{d}x\,\mathrm{d}z\right)\dot{\phi}_1\phi_1^0$$
$$+\,2c'\int_{S_{x,z}}\left(\sum_{y=0,1}\int_{]v_L,v_H[} z'[\tilde{y}\psi^2(\phi_1^0)^2]\,\mathrm{d}v\right)\dot{\phi}_2\phi_2^0\,\mathrm{d}x\,\mathrm{d}z.$$

Comparing this expression with the expression of $\langle(\dot\phi_1,\dot\phi_2,\dot\psi),(\phi_1^*,\phi_2^*,\psi^*)\rangle_F$ and using the fact that $\int_{S_{x,z}}\phi_2^0\dot\phi_2\,\mathrm{d}x\,\mathrm{d}z=0$ and $\sum_{y=0,1}\phi_1^0\dot\phi_1=0$ for any $\dot\phi_1$ and $\dot\phi_2$ in the tangent space, we can see that any $(\phi_1^*,\phi_2^*,\psi^*)$ such that

$$\phi_1^*(y\,|\,v,x,z)=\tfrac{1}{2}c'z'\tilde{y}\phi_1^0+A_1(x,v,z)\phi_1^0,$$

$$\phi_2^*(x,z)=\tfrac{1}{2}c'.(z'E(\tilde{y}\,|\,x,z)\phi_2^0+A_2\phi_2^0.$$

$$\psi^*(v\,|\,x,z)=0,$$

for some function $A_1(x,v,z)$ and some intercept A_2, is such that

$$\langle(\dot\phi_1,\dot\phi_2,\dot\psi),(\phi_1^*,\phi_2^*,\psi^*)\rangle_F=\nabla\rho_0(\dot\phi_1,\dot\phi_2,\dot\psi).$$

To determine $A_1(x,v,z)$ and A_2, we impose that $(\phi_1^*,\phi_2^*,\psi^*)$ belongs to the tangent space, i.e., $\int_{S_{x,z}}\phi_2^0\phi_2^*\,\mathrm{d}x\,\mathrm{d}z=0$ and $\sum_{y=0,1}\phi_1^0(y\,|\,v,x,z).\phi_1^*=0$. These two imply,

$$A_1(x,v,z)=-\tfrac{1}{2}c'\mathrm{E}(z'\tilde{y}\,|\,v,x,z)\quad\text{and}\quad A_2=-\tfrac{1}{2}c'.E(z'\tilde{y})=-\tfrac{1}{2}c'.\pi_0.$$

Thus, we necessarily have

$$\phi_1^*(y\,|\,v,x,z)=\tfrac{1}{2}c'z'(\tilde{y}-\mathrm{E}(\tilde{y}\,|\,v,x,z))\phi_1^0,$$

$$\phi_2^*(x,z)=\tfrac{1}{2}c'.(z'\mathrm{E}(\tilde{y}\,|\,x,z)-\pi_0)\phi_2^0.$$

$$\psi^*(v\,|\,x,z)=0.$$

We have just found a vector $(\phi_1^*,\phi_2^*,\psi^*)$ in the tangent space which satisfies $\nabla\rho_0(\dot\phi_1,\dot\phi_2,\dot\psi)=\langle(\dot\phi_1,\dot\phi_2,\dot\psi),(\phi_1^*,\phi_2^*,\psi^*)\rangle_F$ for any $(\dot\phi_1,\dot\phi_2,\dot\psi)$ in the tangent space. Using again the Riesz–Frechet theorem, this result proves that the linear operator $\nabla\rho_0$ is continuous and that $(\phi_1^*,\phi_2^*,\psi^*)$ is its representer.

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

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

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

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

C.2. The variance–covariance of Lewbel estimate

As in Newey (1994), consider the estimation of the parameter of interest $\pi_t=\mathrm{E}(z'\tilde{y})$ 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):

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

Therefore

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

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

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

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

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

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

we therefore have

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

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

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

since $\mathrm{E}q = 0$ and where we used that $\pi_0 = \mathrm{E}(z'x)\beta_0$

C.3. Theorem 8

When $f(v \mid x, z)$ is unknown and estimated, Lewbel (2000) and Appendix C.2 shows that the variance–covariance matrix of the estimator of π_0 is the variance–covariance of the random variable:

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

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

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

Consider

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

so that we can write

$$Vq_0 = \mathrm{E}(z'.\mathrm{E}((\eta_0)^2 \mid v, x, z).z),$$

$$Vq = \mathrm{E}(z'.\mathrm{E}((\eta)^2 \mid v, x, z).z).$$

Some algebra yields

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

Therefore

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

As we can write

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

where $\mathrm{E}(\eta_1 \mid x, z) = 0$, we have

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

and therefore:

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

is a semi-definite positive matrix.

Finally observe that

$$\mathrm{E}((\eta_1)^2 \mid z) = V\left[\frac{G(v, x, z) - \mathbf{1}\{v > 0\}}{f(v \mid x, z)} \mid z\right]$$

is strictly positive if v varies over its support and G is continuous. If $\mathrm{E}(z'z)$ has full rank, Δ is definite positive.

References

Anton, A.A., Fernandez-Sainz, A., Rodriguez-Poo, J., 2001. Semiparametric estimation of a duration model. Oxford Bulletin of Economics and Statistics 63, 517–533.

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

Blundell, R., Powell, J.L., 2004. Endogeneity in semiparametric binary response models. Review of Economic Studies 71, 655–679.

Cavanagh, C., Sherman, R.P., 1998. Rank estimators for monotone index models. Journal of Econometrics 84, 351–381.

Chamberlain, G., 1987. Asymptotic efficiency with conditional moment restrictions. Journal of Econometrics 34, 305–334.

Chamberlain, G., 1992. Efficiency bounds for semiparametric regression. Econometrica 60, 567–596.

Chen, S., 2002. Semiparametric estimation of a heteroscedastic binary choice model. Working Paper, Hong Kong University.

Cogneau, D., Maurin, E., 2002. Parental income and school attendance in a low-income country, Unpublished manuscript.

Cosslett, S.R., 1983. Distribution-free maximum likelihood estimator of the binary choice model. Econometrica 51, 765–782.

Cosslett, S.R., 1987. Efficiency bounds for distribution-free estimators of the binary choice and the censored regression models. Econometrica 55, 559–585.

Crépon, B., Kramarz, F., Trognon, A., 1998. Parameters of interest, nuisance parameters and orthogonality conditions: an application to autoregressive error component models. Journal of Econometrics 82, 135–156.

Han, A.K., 1987. Non-parametric analysis of a generalized regression model. Journal of Econometrics 35, 303–316.

Hirano, K., Imbens, G.W., Ridder, G., 2003. Efficient estimation of average treatment effects using the estimated propensity score. Econometrica 71, 1161–1189.

Hong, H., Tamer, E., 2003. Endogenous binary choice model with median restrictions. Economics Letters 80, 219–225.

Honoré, B., Lewbel, A., 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, 505–531.

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.

Khan, S., Lewbel, A., 2003. Weighted and two stage least squares estimation of semiparametric truncated regression models. Working Paper, Boston College.

Klein, R.W., Spady, R.H., 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, 105–121.

Lewbel, A., 2000. Semiparametric qualitative response model estimation with unknown heteroskedasticity or instrumental variables. Journal of Econometrics 97, 145–177.

Lewbel, A., 2003. Endogenous selection or treatment model estimation. Working Paper, Boston College.

Lewbel, A., Linton, O., McFadden, D., 2001. Estimating features of a distribution from binomial data. Working Paper, CeMMAP.

Magnac, T., Maurin, E., 2003. Identification & information in monotone binary models. Working Paper, CREST, no. 2003–07, ⟨http://www.crest.fr/doctravail/document/2003-07.pdf⟩.

Magnac, T., Maurin, E., 2004. Partial identification in monotone binary models: discrete and interval valued regressors. Working Paper, CREST, no. 2004-11.

Manski, C.F., 1975. Maximum score estimation of the stochastic utility model of choice. Journal of Econometrics 3, 205–228.

Manski, C.F., 1985. Semiparametric analysis of discrete response: asymptotic properties of the maximum score estimator. Journal of Econometrics 27, 313–333.

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, 301–332.

Newey, W.K., 1994. The asymptotic variance of semiparametric estimators. Econometrica 62, 1349–1382.

Newey, W.K., McFadden, D., 1994. Large sample estimation. In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, vol. 4. North-Holland, Amsterdam, pp. 2111–2245.

Powell, J., 1986. Symmetrically trimmed least squares estimation for tobit models. Econometrica 54, 1435–1460.

Powell, J., 1994. Estimation of semiparametric models. In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, vol. 4. North-Holland, Amsterdam, pp. 2444–2521.

Powell, J.L., Stock, J.H., Stoker, T.M., 1989. Semi-parametric estimation of index coefficients. Econometrica 57, 1403–1430.

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., Tripathi, G., 2001. A simplified approach to computing efficiency bounds in semiparametric models. Journal of Econometrics 102, 23–66.