

ARTICLE IN PRESS

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

JOURNAL OF
Econometrics1
3 ELSEVIER

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

www.elsevier.com/locate/jeconom5
7
9 Identification and information in monotone binary
models11
13 Thierry Magnac^{a,*}, Eric Maurin^b15 ^a*INRA-LEERNA and IDEI, 21 Allée de Brienne, 31000 Toulouse, France*^b*CREST-INSEE, 15, Boulevard Gabriel Péri, 92245 Malakoff Cedex, France*17
19 **Abstract**

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

© 2006 Published by Elsevier B.V.

JEL classification: C14; C25*Keywords:* Binary models; Semiparametric methods; Very exogeneous regressor

43 *Corresponding author.

E-mail addresses: magnac@cict.fr (T. Magnac), maurin@ensae.fr (E. Maurin).*URL:* <http://www.idei.fr/vitae/magnac.html>.

1 1. Introduction

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

$$5 \quad y = 1(x\beta + v + \varepsilon > 0), \quad (\text{LV})$$

7 where errors, ε , are uncorrelated with a set of instrumental variables z (i.e., $E(z'\varepsilon) = 0$), and conditionally independent of the continuous regressor, v , (i.e., partial independence, $8 F_\varepsilon(\varepsilon | x, z, v) = F_\varepsilon(\varepsilon | x, z)$). 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.

13 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 15 that any binary outcome can be analysed with a latent model satisfying partial independence provided that it is monotone in v (i.e., $\Pr(y = 1 | v, x, z)$ is monotone in v). 17 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 19 assumption. Additional restrictions are needed for identification. Such an additional restriction is the assumption, introduced by [Lewbel \(2000\)](#), that the support of the special 21 regressor is large ($\text{Supp}(v) \supseteq \text{Supp}(-x\beta - \varepsilon)$). Our second result is that the combination of uncorrelated instruments, partial-independence and large-support assumptions provides 23 *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, 25 we show that the large-support assumption can only be used when the conditional probability of success $\Pr(y = 1 | v, x, z)$ increases from 0 to 1 over the support of v , which is 27 admittedly restrictive. Large-support conditions are actually quite common in the literature about semi-parametric limited-dependent variable models (see e.g., [Manski, 29 1975, 1985; Han, 1987; Horowitz, 1992; Cavanagh and Sherman, 1998](#)), but they represent a potential obstacle to empirical applications.

31 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 33 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 , 35 which increases the usefulness of the setting in empirical work.

Making identification restrictions as weak as possible is not the only concern when 37 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 39 contributions. We establish the semi-parametric efficiency bound for the parameter under partial independence and uncorrelated instrument assumptions by using the framework 41 proposed by [Severini and Tripathi \(2001\)](#). It is noteworthy that the special-regressor estimator proposed by [Lewbel \(2000\)](#) achieves this bound (under some regularity 43 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 45 conjectured by [Lewbel \(2000\)](#).

Generally speaking, the set of identifying restrictions analysed in this paper provides 47 interesting means to overcome [Manski's](#) fundamental impossibility result according to

1 which an uncorrelated-error restriction (i.e., $E(x'\varepsilon) = E(v'\varepsilon) = 0$) or even a mean-
 independence restriction (i.e., $E(\varepsilon|x, v) = 0$) is not sufficient for identifying β no matter
 3 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
 5 than standard parametric models or than the semi-parametric methods that are based on
 the properties of statistical independence (i.e., $F_\varepsilon(\varepsilon|x, v) = F_\varepsilon(\varepsilon)$) or of single-index
 7 sufficiency (i.e., $F_\varepsilon(\varepsilon|x, v) = F_\varepsilon(\varepsilon|x\beta + v)$) see e.g. Cosslett (1983), Ruud (1983), Han
 (1987), Powell et al. (1989), Ichimura (1993), Klein and Spady (1993) who provide
 9 estimators of β under statistical independence or index sufficiency.

The quantile-independence assumption does not provide just identification of the
 11 parameter of interest, but permits slightly more general forms of conditional
 heteroskedasticity than the exclusion restrictions used in this paper. Still, the fact remains
 13 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
 15 (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
 17 advances have recently been proposed by Chen (2002) who suggests strengthening the
 median-independence assumption into conditional symmetry and a weak restriction on
 19 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
 21 *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
 23 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
 25 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),
 27 Lewbel (2003).

The paper is organized as follows.

29 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
 31 random variables (y, x, v, z) such that the conditional probability of success $\Pr(y =$
 $1 | v, x, z)$ increases from 0 to 1 when v varies over its support.

33 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
 35 assumption for obtaining just identification of β . The only condition that $\Pr(y =$
 $1 | v, x, z)$ should satisfy is to be monotone in v .

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

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

41

43 2. The set-up and the equivalence result

45 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
 47 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 $S_{x,z} \subset \mathbb{R}^p \times \mathbb{R}^q$. The dimension of the set $S_{x,z}$ is $r \leq p + q$ where $p + q - r$ are the potential overlaps and functional dependencies.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, $\text{rank}(E(z'x)) = p$.

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 $F_v(\cdot | x, z)$ 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, $(F_v(\cdot | x, z) \cdot F_{x,z})$, 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).

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

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

$$3 \quad F_{\varepsilon}(\cdot | v, x, z) = F_{\varepsilon}(\cdot | x, z).$$

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

7 (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 :
 9 $E(z'\varepsilon) = 0$.

Regarding (L.1), it should be noted that Powell (1994) discusses partial independence
 11 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
 13 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,
 15 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
 17 models where ε represents unobserved preference variation, prices satisfy the partial
 independence condition when goods are homogenous and markets are competitive. In
 19 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
 21 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
 23 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
 25 year significantly influences children's outcomes in primary school. Given that this variable
 is plausibly independent from children's unobserved ability, partial independence is
 27 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
 29 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
 31 special regressor in a duration model.

As it turns out, the partial independence assumption provides an identifying restriction
 33 which can be applied in contexts which are economically interesting. In his recent
 contribution, Lewbel (2000) constructs an estimator of β by combining partial
 35 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
 37 next section describes the class of binary phenomena that may actually be analysed
 through this set-up.

39

41 2.1. The equivalence result

43 Consider $(\beta, F_{\varepsilon}(\cdot | x, z))$, 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
 45 generated by $(\beta, F_{\varepsilon}(\cdot | x, z))$ through the binary transformation (LV). The following lemma
 shows that this conditional distribution necessarily increases from 0 to 1 when v varies over
 47 its support.

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

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

5 (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:

$$7 \quad \Pr(y_i = 1 | v_l, x, z) = 0, \quad \Pr(y_i = 1 | v_h, x, z) = 1.$$

9 **Proof.** See Appendix A.1.

11 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.

15 Summing up, if we denote:

$$17 \quad \mathcal{M}_{NP} = \{\Pr(y = 1 | v, x, z) \text{ satisfying monotonicity (NP.1),}$$

$$19 \quad \text{and support (NP.2) conditions}\}$$

and

$$21 \quad \mathcal{M}_L = \{\Pr(y = 1 | v, x, z) \text{ generated through (LV)}$$

$$23 \quad \text{by some } (\beta, F_e(\cdot | x, z)) \text{ satisfying (L.1–L.3)}\}$$

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

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

$$31 \quad E(z'x) \cdot \beta = E(z' \tilde{y}), \tag{1}$$

33 where

$$35 \quad \tilde{y} = \frac{y - \mathbf{1}(v > 0)}{f(v | x, z)} \tag{2}$$

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

39 **Proof.** See Appendix A.2.

When there exist as many instruments as explanatory variables ($q = p$), condition (1) defines a unique parameter β and $\mathcal{M}_{NP} \subset \mathcal{M}_L$. 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:

45 (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, $\mathcal{M}_{NP} \subset \mathcal{M}_L$ and, taken together, Lemmas 1 and 2 prove our first basic result:

1 **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
 3 probabilities satisfying (NP.1–NP.2).

5 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
 7 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
 9 by a unique function of the joint distribution of the data.²

11 2.2. Discussion

13 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
 15 (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
 17 restriction (i.e., $E(x'\varepsilon) = E(v'\varepsilon) = 0$) or even a mean-independence restriction (i.e.,
 $E(\varepsilon | x, v) = 0$) is not sufficient for identifying β no matter what conditions on the support
 19 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
 21 shown in Section 3, while adding more than (L.1) to (L.2) and (L.3) would generate
 testable overidentifying restrictions.³

23 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
 25 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
 27 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,
 29 $u = y_2 - E(y_2 | v, z)$.

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

$$31 \quad F_\varepsilon(\cdot | y_2, z, v) \equiv F_\varepsilon(\cdot | u, z, v) = F_\varepsilon(\cdot | u),$$

33 whereas the partial independence assumption used in this paper is

$$35 \quad F_\varepsilon(\cdot | y_2, z, v) = F_\varepsilon(\cdot | y_2, z) \equiv F_\varepsilon(\cdot | u + E(y_2 | v, z), z).$$

Consequently, Blundell and Powell's model requires a stronger exclusion restriction if
 37 $E(y_2 | v, z)$ does not depend on v . However, generally the models are not nested.⁴ Other

39 ²Apart 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-
 41 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.

43 ³For instance, strengthening (L.3) into a mean-independence restriction $E(\varepsilon | z) = 0$ 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.

45 ⁴Although 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 $E(y_2 | z, v)$ (instead of z_2). The partial independence assumption is then
 47 strictly weaker than the assumption used by Blundell and Powell.

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

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

9 Another difference is that partial independence yields just identification of β while
 10 quantile independence imposes testable overidentifying restrictions.⁵ Also, the partial
 11 independence hypothesis makes it possible to estimate the distribution of the unobserved
 12 residuals while the quantile-independence assumption does not. This property may be of
 13 particular interest for evaluating the impact of the covariates on the probability of
 14 observing $y = 1$ (Lewbel et al., 2001). The price to pay is that partial independence requires
 15 conditions on the support of the covariates that are stronger than the conditions required
 16 under quantile independence. As shown by Horowitz (1998), a sufficient support condition
 17 for estimating β under quantile-independence is that for a set of x of positive mass, $v + x\beta$
 18 takes both positive and negative values when v varies over its support. It is weaker (and in
 19 some cases strictly weaker) than (L.2) which implies that $v + x\beta$ takes both positive and
 20 negative value for any x when v varies over its support. Lastly, it should be noted that the
 21 endogeneity of covariates can be also accommodated in a quantile independence setting
 22 (Hong and Tamer, 2003) so that the two methods are on par in this respect.

23 24 25 3. Unrestricted support and identification

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

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

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

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

43 ⁶However, the large-support assumption is quite natural in many settings. For instance, it seems a plausible
 44 assumption for events that necessarily take place within a specific period of the life-cycle. When y describes such
 45 phenomena as primary-school attendance, school-leaving, leaving parental home, the entry into (or the exit from)
 46 the labour market (for male workers), age is the most obvious candidate as the special continuous regressor, v , and
 47 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).

1 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 $\Pr(y = 1 | v, x, z)$ is that it should
 3 be zero or one when v is $\pm\infty$. If $v_H = +\infty$ and $\Pr(y = 1 | v_H, x, z) < 1$ (or if $v_L = -\infty$ and
 $\Pr(y = 1 | v_L, x, z) > 0$), no latent variable model in \mathcal{M}_L can lead to the conditional
 5 probability function. It is the reason why we shall exclude this case by setting:

$$7 \quad (NP.2') \lim_{v \rightarrow +\infty} \Pr(y = 1 | v, x, z) = 1 \quad \text{and} \quad \lim_{v \rightarrow -\infty} \Pr(y = 1 | v, x, z) = 0.$$

Observe also that when the support of v coincides with the real line (i.e., $v_H = +\infty$ and
 9 $v_L = -\infty$) (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.

11 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.
 13 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
 15 of additional identifying restrictions leading to exact identification. It is shown that it
 preserves the validity of Lewbel's estimation procedure.

17 3.1. The necessity of additional identifying restrictions

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

$$25 \quad \Pr(y = 1 | v, x, z) = \int_{v+x\beta+\varepsilon > 0, \varepsilon \in \Omega_\varepsilon(x, z)} f_\varepsilon(\varepsilon | x, z) d\varepsilon$$

27 and thus:

$$29 \quad \begin{aligned} & \Pr(y = 1 | v, x, z) - \Pr(y = 1 | v_L, x, z) \\ &= \int_{-(v+x\beta)}^{-(v_L+x\beta)} f_\varepsilon(\varepsilon | x, z) d\varepsilon = F_\varepsilon(-(v_L + x\beta) | x, z) - F_\varepsilon(-(v + x\beta) | x, z). \end{aligned}$$

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

$$33 \quad f_\varepsilon(\varepsilon | x, z) = \frac{\partial}{\partial v} \Pr(y = 1 | v, x, z) |_{v=-(x\beta+\varepsilon)}. \quad (3)$$

In contrast to the large-support case, the support of ε (conditional on x and z) is not
 37 necessarily included in $]-(v_H + x\beta), -(v_L + x\beta)[$ and $f_\varepsilon(\varepsilon | x, z)$ has no non-parametric
 counterpart for ε in

$$39 \quad B(x) =]-\infty, -(v_H + x\beta)[\cup]-(v_L + x\beta), +\infty[.$$

41 The only restrictions on the distribution of ε in $B(x)$ are the following:⁷

$$43 \quad \begin{aligned} \Pr\{\varepsilon \leq -(v_H + x\beta) | x, z\} &= 1 - \Pr(y = 1 | v_H, x, z), \\ \Pr\{\varepsilon > -(v_L + x\beta) | x, z\} &= \Pr(y = 1 | v_L, x, z). \end{aligned} \quad (4)$$

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

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

1 $1 | v, x, z)$ through (LV) necessarily satisfies (3) and (4). Conversely, any latent model
 2 $(\beta, F_\varepsilon(\cdot | x, z))$ satisfying (L.1) and (3) and (4) generate function $\Pr(y = 1 | v, x, z)$ through
 3 (LV). From this, it follows clearly that the partial independence hypothesis is not sufficient
 4 for complete identification of $F_\varepsilon(\cdot | x, z)$, even when β is known. Under (L.1), the only
 5 restrictions on $F_\varepsilon(\cdot | x, z)$ when ε is in $B(x)$ are given by (4), which means that the
 6 distribution of ε conditional on $\varepsilon \in B(x)$ is left completely unidentified.

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

$$12 \quad S = \{\beta \in \mathbb{R}^p \mid \exists F_\varepsilon(\cdot | x, z) \text{ satisfying (L.1) and (L.3)} \\
 13 \quad \text{s.t. } (\beta, F_\varepsilon(\cdot | x, z)) \text{ generates } \Pr(y = 1 | v, x, z)\}$$

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

17 **Proposition 4.** Consider $\Pr(y = 1 | v, x, z)$ satisfying (NP1), (NP2)', but not (NP2). For any
 18 $\lambda_0 > 0$, there exists a latent model $(\beta, F_\varepsilon(\cdot | x, z))$ such that

- 19 (i) $(\beta, F_\varepsilon(\cdot | x, z))$ satisfies (L1), (L3) and generates $\Pr(y = 1 | v, x, z)$ through (LV).
 20 (ii) $(\beta - \beta_0)'(\beta - \beta_0) \geq \lambda_0$,
 21 where β_0 is the value associated with the moment condition $E(z'x) \cdot \beta_0 = E(z'\tilde{y})$.

22 **Proof.** See Appendix B.1.

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

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

34 3.2. Generalizing the special-regressor estimator

35 It should be noted that the set $B(x)$ is the union of two subsets, $B_F(x) = \{\varepsilon : \\
 36 \varepsilon + v_H + x\beta < 0\}$ and $B_S(x) = \{\varepsilon : \varepsilon + v_L + x\beta > 0\}$. An individual in $B_F(x)$ always responds
 37 $y = 0$, even when $-v$ is minimum ($-v_H$). Symmetrically, an individual in $B_S(x)$ always
 38 responds $y = 1$, even when $-v$ is maximum ($-v_L$). The set $B_F(x)$ may be interpreted as the
 39 subset of certain failure and, $B_S(x)$, the subset of certain success. By construction, the data
 40 do not provide any information on the distribution of the propensities of success in $B_F(x)$

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

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

$$9 \quad E(z' y_{v_H}^* \mathbf{1}\{y_{v_H}^* > 0\}) = E(z' y_{v_L}^* \mathbf{1}\{y_{v_L}^* > 0\}), \quad (5)$$

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

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

15 **Proof.** See Appendix B.2.

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

25 If either v_H or v_L is infinite,⁸ condition (5) cannot be satisfied. Let $v_H = +\infty$ (say), then
 27 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
 27 since $E y_{v_L}^* \mathbf{1}\{y_{v_L}^* > 0\} > 0$. Nevertheless as shown in B.2, the bias may affect the intercept
 29 term only.

31 **Proposition 6.** Assume $v_H = +\infty$, $v_L > -\infty$ and consider $\Pr(y = 1 | v, x, z)$ satisfying (NP.1)
 31 and (NP.2'). Let $S \subseteq \mathbb{R}^p$ be the set of parameters β such that there exists a latent model
 33 $(\beta, F_\varepsilon(\cdot | x, z))$ satisfying (L.1), (L.3) and generating $\Pr(y = 1 | v, x, z)$ through (LV) and let
 33 $S_1 = \{(\beta_2, \dots, \beta_p) \in \mathbb{R}^{p-1} \text{ s.t. } \beta_1 \in \mathbb{R}, (\beta_1, \dots, \beta_p) \in B\}$ where β_1 is the intercept coefficient.
 33 S_1 is reduced to a singleton if there exists a constant α independent from z such that

$$35 \quad E(y_{v_L}^* \mathbf{1}\{y_{v_L}^* > 0\} | z) = \alpha, \quad (6)$$

37 where $y_{v_L}^* = (x\beta + v_L + \varepsilon)$ represents the propensity of success for individuals with the lowest
 37 possible v . Under (L.1), (L.3) and (6) the moment condition $(E(z'x) \cdot \beta = E(z'\tilde{y}))$ provides exact
 39 identification of β apart from the intercept coefficient.

41 **Proof.** See Appendix B.2.

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

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

1 4. Information and asymptotic properties

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

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

15 In our set-up, the only identifying restriction is given by the moment (1) and one possible
 16 source of difficulty comes from the relationship between the unknown non-parametric
 17 component of \tilde{y} (i.e., $f(v|x, z)$) and the density function with respect to which the moment
 18 restriction (i.e., $E(z'[\tilde{y} - x\beta]) = 0$) is defined. Given this relationship, the general
 19 framework investigated by Chamberlain (1992) needs to be amended and the semi-
 20 parametric efficiency of the estimators has to be checked by hand.

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

28 In the large-support case, the regularity conditions under which the special-regressor
 29 estimator is root- n consistent, and asymptotically normal, are given in Lewbel (2000),
 30 Appendix B, Conditions B1–B6 and Condition B7 or B7' depending on whether the
 31 support of ε is bounded. It is easy to check that Conditions B1–B6 and Condition B7'
 32 remain applicable⁹ when the large-support assumption does not hold and that—under
 33 these conditions—the proof of root- n consistency and asymptotic normality still holds true
 34 too. In particular, Condition B7' imposes conditions either on the rate at which $\Pr(y =$
 35 $1 | v, x, z)$ tends to 0 or 1 when $|v| \rightarrow \infty$ or on the support of v which are easy to satisfy
 36 when the large-support assumption does not hold true any more.¹⁰

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

41

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

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

1 4.1. The semiparametric efficiency bound

3 The estimate is based on the unconditional moment conditions:

$$5 \quad E[m(y, v, x, z; \beta_0)] = 0, \quad (7)$$

where the function of interest is

$$7 \quad m(y, v, x, z; \beta) = z' \left[\frac{y - \mathbf{1}(v > 0)}{f(v | x, z)} - x\beta \right] = z'[\tilde{y} - x\beta].$$

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

$$11 \quad E[mm'] = \Omega_0$$

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

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

$$17 \quad (E(x'z)\Omega_0^{-1}E(z'x))^{-1},$$

19 and the efficient estimate would then be obtained as usual. In our case, however, the density $f(v | x, z)$ 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.

23 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

$$27 \quad E(\tilde{g}(y, v, x, z; \pi_0)) = E(z'\tilde{y} - \pi_0) = 0. \quad (8)$$

29 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

$$33 \quad (E(x'z).V_\pi^{-1}.E(z'x))^{-1}.$$

35 The bound for V_π is described by the following result.

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

$$39 \quad E(z'(\tilde{y} - E(\tilde{y} | v, x, z) + E(\tilde{y} | x, z) - x\beta_0)^2 z).$$

Proof. See Appendix C.1.

41 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.

45 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

1 variance–covariance of estimator does not depend on the specific assumptions made on
 2 bounds. Whether conditions (L.2) or (NP.2) are satisfied or not, the same properties apply
 3 to Lewbel’s estimate. It is consistent and semi-parametrically efficient under the conditions
 4 of Propositions 5 or 6.

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

7

8 4.2. *Plugging-in the true or estimated conditional density?*

9

10 In this section, we assume that the conditional density $f(v|x, z)$ is known. It may
 11 correspond to the case where v is under experimental control or the case where one has
 12 access to additional external information on the distribution of v (through census
 13 information for instance). In such a case, we can consider two different transformations,
 14 $\tilde{y} = (y - I(v > 0))/f(v|x, z)$ or $(y - I(v > 0))/\hat{f}(v|x, z)$ when constructing the linear regres-
 15 sion that leads to the estimation of β . Here, $f(v|x, z)$ is the true distribution and $\hat{f}(v|x, z)$ is
 16 an estimate of $f(v|x, z)$. It was conjectured by Lewbel (2000) (and confirmed by Monte-
 17 Carlo experiments) that the estimate of β obtained with \tilde{y} and the true value of the density
 18 actually has a larger asymptotic variance than the estimate obtained with \hat{y} and the
 19 estimated value of the density. We now offer a proof for this conjecture:

20 **Theorem 8.** *The estimate of π_0 defined by the unconditional moment condition (8) (i.e.,*
 21 *$E(z'\tilde{y} - \pi_0) = 0$) has a strictly smaller variance when the estimated $\hat{f}(v|x, z)$ is used to*
 22 *transform the dependent variable than when the true density is used.*

23

24 **Proof.** See Appendix C.3.

25

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

40 5. Extensions

41 Lewbel (1998, 2000, 2003) uses the special-regressor hypothesis to estimate the structural
 42 parameters of other linear latent variable models, $y = L(x\beta + \varepsilon)$, such as the ordered
 43 discrete choice model with constant thresholds or the censored regression model. One
 44 obvious issue is whether the equivalence results given by Theorem 3 can be extended to
 45 these models. In some interesting cases the answer is positive. In other cases, the special

1 regressor setting imposes testable restrictions on the set of statistical phenomena that are
 2 generated by the latent structure.

3 To illustrate the generalization of Theorem 3, we consider the most straightforward
 4 extension of binary responses which are ordered choice models. Assume that the support
 5 of y is now $S_y = \{0, 1, \dots, K\}$ ($K \geq 1$). We consider two definitions of ordered choice
 6 models and discuss each in turn. In the first one, each individual is defined by an ordered
 7 set of propensities (i.e., y_1^*, \dots, y_K^*) and his/her response ($y \in \{0, 1, \dots, K\}$) depends on how
 8 propensities compare with a given cost variable v . In the second model, each individual is
 9 defined by one specific propensity y^* and his/her response depends on how this propensity
 10 compares with an ordered set of thresholds $\alpha_k(v)$. A straightforward extension of Theorem
 11 3 only holds in the first case, whereas structural parameters are overidentified in the second
 12 model.

13

14 5.1. Ordered choices: first model

15

16 Consider the following definition for latent ordered choice models.

17

18 **Definition 9.** Latent ordered discrete choice models are characterized by a set of ordered
 19 latent random variables $\{y_1^*, \dots, y_K^*\}$ where $y_k^* > y_{k+1}^*$. By convention define $y_{K+1}^* = -\infty$. The
 20 observable model is given by

21

22

23

24

25

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

26

27 We consider linear latent models such as

28

29

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

30

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

32

33 This model is a straightforward generalization of (LV). When $K = 1$, the two models
 34 coincide. Such an ordered choice model may typically be used for analysing consumer
 35 behaviour. Suppose that the observed variable y records the number of units of a good that
 36 is bought by consumer i when the offered unit price is $(-v)$. The latent variables, y_k^* , stand
 37 for the willingness to pay for an additional unit of this good when the number of units
 38 bought is $k - 1$. If marginal utility is decreasing, then the marginal-unit willingness to pay
 39 is decreasing, which justifies the ordered choice setting. The fact that an entire array of
 40 unobserved components affect willingness to pay is due to individual differences in the
 41 relation between marginal utility and quantity purchased.

42

43 One of the interesting features of the setting given by (LV1) is that it is equivalent to a
 44 system of K binary latent models given by

45

$$y_k = I(-y_k^* < v). \tag{LV1k}$$

46

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

49

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

50 Reciprocally:

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

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

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

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 $(\varepsilon_1, \dots, \varepsilon_K)$ is therefore:

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

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

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

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

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

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 \mathcal{M}_{LOC}^* of latent ordered models be given by parameters $(\beta_1, \dots, \beta_K) \in \mathbb{R}^K$, distribution functions $(f_1(\varepsilon_1 | x, z), \dots, f_K(\varepsilon_K | x, z)) \in \mathcal{D}^K$, 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:

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

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

Theorem 10. \mathcal{M}_{LOC}^* is one-to-one with \mathcal{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:

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

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

$$\alpha_1(v) = -v \leq \alpha_2(v) \leq \dots \leq \alpha_k(v) \leq \alpha_{k+1}(v) = +\infty, \quad (10)$$

while ε satisfies (L.1)–(L.3).

1 This model is also a direct generalization of (LV). When $K = 1$, the two models coincide.
 2 $x\beta + \varepsilon$ may be interpreted as a propensity to respond as in (LV), but now the response has
 3 several possible levels of intensity. The $\alpha_k(v)$ thresholds may be interpreted as the cost of
 4 responding with intensity k . The only structural assumption about these costs is that they
 5 increase with the intensity of the response.

6 Such a model may describe, for instance, the performance of young children when
 7 starting school, where y^* represents their (latent) schooling ability (plausibly dependent on
 8 family inputs) and the $\alpha_k(v)$ thresholds represent the set of thresholds (plausibly dependent
 9 on v being the birthdate within the year) imposed by the educational system for deciding
 10 who should be held back ($y = 0$), who should be on time ($y = 1$) and who should be ahead
 11 ($y = 2$) at school.¹¹

12 Let $\mathcal{M}_{\text{LOC2}}^*$ be the set of latent ordered discrete choice models where elements $(\beta,$
 13 $F_\varepsilon(\cdot | x, z), \alpha_k(v), k = 2, \dots, K)$ satisfy independence, support and moment conditions
 14 ($L.1-L.3$). Consider also a statistical model $F(y | v, x, z)$ on \mathcal{S}_y such that $\Pr(y_i \geq 1 | v, x, z)$
 15 satisfy conditions ($NP.1-NP.2$) and assume that there exists a latent ordered choice model
 16 $(\beta, F_\varepsilon(\cdot | x, z), \alpha_k(v), k = 2, \dots, K)$ in $\mathcal{M}_{\text{LOC2}}^*$ where the image is $F(y | v, x, z)$.

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

$$21$$

$$22 \quad G_k(v, x, z) = \int_{-\infty}^{-x\beta + \alpha_k(v)} dF(\varepsilon | x, z) = \int_{-\infty}^{-x\beta + \alpha_k(v)} \frac{-\partial G_0}{\partial v} (-(x\beta + \varepsilon), x, z) d\varepsilon$$

$$23$$

$$24 \quad = G_0(-\alpha_k(v), x, z).$$

$$25$$

26 It therefore yields

$$27$$

$$28 \quad \alpha_k(v) = -G_0^{-1}(\cdot, x, z) \circ G_k(v, x, z).$$

$$29$$

30 Thus $F(y | v, x, z)$ is the image of an element of $\mathcal{M}_{\text{LOC2}}^*$ only if $G_0^{-1}(\cdot, x, z) \circ G_k(v, x, z)$ do not
 31 depend on x and z . Put differently, a monotone ordered discrete phenomena can be
 32 analysed as a structural ordered choice model that satisfies the partial independence
 33 hypothesis only if $G_0^{-1} \circ G_k$ does not depend on x and z , which is a testable assumption.
 34 Note finally that the inequalities described by (10) translate into the same inequalities in
 35 the functions G_k that we had in the previous subsection and which are adapted to the
 36 present setting. They do not affect our argument.

37 Therefore, the ordered discrete choice models with fixed thresholds (i.e., $\alpha_k(v) - \alpha_0(v) =$
 38 γ_k) are *not* one-to-one with the monotone discrete models. The partial independence
 39 hypothesis makes it possible to identify very easily the structural parameters that
 40 characterize these ordered choice models, but this assumption also implies (testable)
 41 restrictions on the set of discrete monotone phenomena which can be analysed with such
 42 models.

43

44

45 ¹¹Maurin (2002) uses the binary approach to estimate the probability to be held back using $v =$ day-of-birth
 46 within the year as a special regressor and interpreting $x\beta + \varepsilon$ as schooling abilities. $-\alpha_2(v)$ can be interpreted as the
 47 ability threshold (defined by the educational system) above which children can be ahead at school.

1 6. Conclusion

3 The first contribution of this paper is to characterize the conditions under which the
 5 identifying assumptions proposed by Lewbel (2000) are justified: $E(y | v, x)$ is monotone in
 7 v and varies from 0 to 1 when v varies over its support. Second, it is shown that the
 9 uncorrelated-error, partial independence and large-support assumptions lead to the exact
 11 identification of the structural parameters of the binary response model. We also prove
 13 that the large-support assumption—which might be unadapted in some instances—can be
 15 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.

23 Acknowledgements

25 We thank the editors, two anonymous referees, Magali Beffy, Pierre Dubois, Guy
 27 Laroque, Oliver Linton, Sandra McNally, Jean-Marc Robin, Bernard Salanié and
 29 particularly Arthur Lewbel for very helpful discussions. We thank participants at the
 ESRC conference in Bristol (2003), at the (EC)² conference in London (2003), ESEM in
 Stockholm (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.

35 Appendix A. Proofs of Section 2

37 A.1. Proof of Lemma 1

39 Write:

$$41 \Pr(y_i = 1 | v, x, z) = \int_{x\beta + v + \varepsilon > 0, \varepsilon \in \Omega_\varepsilon(x, z)} dF_\varepsilon(\varepsilon | x, z)$$

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

45 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
 $\varepsilon \in \Omega_\varepsilon(x, z)$:

$$47 v_L \leq v_l(x, z) < -(x\beta + \varepsilon) < v_h(x, z) \leq v_H$$

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

$$3 \quad v_l(x, z) + x\beta + \varepsilon < 0, \quad v_h(x, z) + x\beta + \varepsilon > 0$$

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

7 A.2. Proof of Lemma 2

9 Consider $G(v, x, z) = \Pr(y = 1 | v, x, z)$ 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 $(\beta, F_\varepsilon(\cdot | x, z))$ in \mathcal{M}_L^* such that $G(v, x, z)$ is its image through the transformation (LV). Define the support of the random variable ε as

$$15 \quad \Omega_\varepsilon(x, z) =] - (v_h(x, z) + x\beta), -(v_l(x, z) + x\beta)[\quad (11)$$

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

$$19 \quad G(v, x, z) = \int_{v+x\beta+\varepsilon > 0, \varepsilon \in \Omega_\varepsilon(x, z)} f_\varepsilon(\varepsilon | x, z) d\varepsilon = \int_{-(v+x\beta)}^{-(v_l+x\beta)} f_\varepsilon(\varepsilon | x, z) d\varepsilon$$

$$21 \quad = 1 - F_\varepsilon(-(v + x\beta) | x, z),$$

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

$$25 \quad f_\varepsilon(\varepsilon | x, z) = \frac{\partial G}{\partial v}(-(x\beta + \varepsilon), x, z). \quad (12)$$

27 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).

31 Furthermore, condition (L.3) implies:

$$33 \quad 0 = E(z' \varepsilon)$$

$$35 \quad = E_{x,z} \left(z' \int \varepsilon f_\varepsilon(\varepsilon | x, z) d\varepsilon \right)$$

$$37 \quad = - E_{x,z} \left(z' \int (x\beta + v) \frac{\partial G}{\partial v} dv \right)$$

$$39 \quad = - E(z' x) \beta - E_{x,z} \left(z' \int v \frac{\partial G}{\partial v} dv \right), \quad (13)$$

41 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 in the usual sense when some linear restrictions are overidentifying (rigorously defined in R.v).

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

1 Finally, we have

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

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

$$19 \quad E_{x,z} \left(z' \int v \frac{\partial G}{\partial v} dv \right) = -E(z' \tilde{y})$$

21 which completes the proof.

25 Appendix B. Proofs of Section 3

27 B.1. Proposition 4

29 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 .

31 Fix β . Chose $F_\varepsilon(\cdot | x, z)$ satisfying partial independence and (3) and (4). By construction, the latent model $(\beta, F_\varepsilon(\cdot | x, z))$ 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):

$$\begin{aligned}
 35 \quad 0 &= E(z' \varepsilon) = E_{x,z} \left(z' \int \varepsilon dF(\varepsilon | x, z) \right) \\
 37 \quad &= E \left(z' \int_{\varepsilon \in B(x)} \varepsilon dF(\varepsilon | x, z) \right) + E \left(z' \int_{-(v_H + x\beta)}^{-(v_L + x\beta)} \varepsilon dF(\varepsilon | x, z) \right) \quad (14)
 \end{aligned}$$

39 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

$$\begin{aligned}
 43 \quad 0 &= E \left(z' \int_{\varepsilon \in B(x)} \varepsilon dF(\varepsilon | x, z) \right) - E \left(z' \int_{v_L}^{v_H} (x\beta + v) \frac{\partial G}{\partial v} dv \right) \\
 45 \quad &= E(z' \varepsilon \mathbf{1}\{\varepsilon \in B(x)\}) - E \left(z' x\beta \int_{v_L}^{v_H} \frac{\partial G}{\partial v} dv \right) - E \left(z' \int_{v_L}^{v_H} v \frac{\partial G}{\partial v} dv \right). \quad (15)
 \end{aligned}$$

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

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

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:

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

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 $b(\cdot) = 0$) and $B(x) = \emptyset$. The last line of condition (16) would give back Lewbel's moment condition (i.e., $E(z' x) \beta = E(z' \tilde{y})$).

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 $t_0(x, z) \geq \eta > 0$. 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,

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

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

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

1 B.2. Propositions 5 and 6

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

$$5 \quad E(z'\varepsilon\mathbf{1}\{\varepsilon \in B(x)\}) + E(z'x\{1 - G(v_H, x, z) + G(v_L, x, z)\})\beta + E(z'b(v_H, v_L, x, z)) = 0$$

6 \iff

$$7 \quad E(z'\varepsilon\mathbf{1}\{\varepsilon < -(v_H + x\beta)\}) + E(z'\varepsilon\mathbf{1}\{\varepsilon > -(v_L + x\beta)\})$$

$$9 \quad E(z'(x\beta + v_H)\{1 - G(v_H, x, z)\}) + E(z'(x\beta + v_L)\{G(v_L, x, z)\}) = 0$$

10 \iff

$$11 \quad E(z'(x\beta + v_H + \varepsilon)\mathbf{1}\{\varepsilon < -(v_H + x\beta)\}) + E(z'(x\beta + v_L + \varepsilon)\mathbf{1}\{\varepsilon > -(v_L + x\beta)\}) = 0,$$

13 which is equivalent to:

$$15 \quad -E(z'y_{v_H}^*\mathbf{1}\{y_{v_H}^* > 0\}) + E(z'y_{v_L}^*\mathbf{1}\{y_{v_L}^* > 0\}) = 0,$$

17 where $y_{v_L}^* = -(x\beta + v_H + \varepsilon)$ and $y_{v_L}^* = x\beta + v_L + \varepsilon$. It proves Proposition 5.

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

$$21 \quad E(z'y_{v_L}^*\mathbf{1}\{y_{v_L}^* > 0\}).$$

23 If the conditional mean is independent of z :

$$25 \quad E(y_{v_L}^*\mathbf{1}\{y_{v_L}^* > 0\} | z) = \alpha$$

27 then the constant only in β is biased.

29 Appendix C. Proofs of Section 4

31 C.1. Proof of Proposition 7

33 C.1.1. Preliminaries

35 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 .

37 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

$$39 \quad f(y, v, x, z) = f(y | v, x, z)f(v | x, z)f(x, z)$$

$$41 \quad = \phi_1^2(y | v, x, z).\psi^2(v | x, z).\phi_2^2(x, z).$$

43 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:

$$45 \quad \Phi_1 = \{\phi_1 : \{0, 1\} \times v_L, v_H \times S_{x,z} \rightarrow \mathbb{R}, \sum_{y=0,1} \phi_1^2(y | v, x, z) = 1, \phi_1^2(y | v, x, z) \geq 0\},$$

$$\Phi_2 = \left\{ \phi_2 \in L^2(S_{x,z}), \int_{S_{x,z}} \phi_2^2(x, z) dx dz = 1, \quad \phi_2^2(x, z) > 0, \right.$$

$$\left. \phi_2^2(x, z) \text{ is bounded} \right\},$$

$$\Psi = \left\{ \psi \in L^2(]v_L, v_H]), \int_{]v_L, v_H[} \psi^2(v | x, z) dv = 1, \psi^2(v | x, z) > 0, \right.$$

$$\left. \psi^2(v | x, z) \text{ is bounded and continuous} \right\},$$

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 $\text{lin } T(E, (\phi_1^0, \phi_2^0, \psi^0))$ 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 $(\dot{\phi}_1, \dot{\phi}_2, \dot{\psi}) \in L^2(F_{y|v,x,z} \cdot F_{v|x,z} \cdot F_{x,z})$ 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 $t \rightarrow \varphi_t$ 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), $\overline{\text{lin } T(E, (\phi_1^0, \phi_2^0, \psi^0))}$ is the product of the following subspaces:

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

$$\left. \sum_{y=0,1} \phi_1^0(v | v, x, z) \cdot \dot{\phi}_1 = 0 \text{ a.e. }]v_L, v_H[\times S_{x,z} \right\},$$

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

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

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 $\langle \cdot, \cdot \rangle_F$ (and the corresponding norm $\|\cdot\|_F$) with,

$$\begin{aligned} \langle (\dot{\phi}_1, \dot{\phi}_2, \dot{\psi}), (\dot{\phi}'_1, \dot{\phi}'_2, \dot{\psi}') \rangle_F &= 4 \sum_{y=0,1} E_{x,z,v}(\dot{\phi}_1 \dot{\phi}'_1) + 4 E_{x,z} \left(\int_{]v_L, v_H[} \dot{\psi} \dot{\psi}' dv \right) \\ &+ 4 \int_{S_{x,z}} \dot{\phi}_2 \dot{\phi}'_2 dx dz \end{aligned}$$

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

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

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

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

$$5 \quad \rho(\phi_1, \phi_2, \psi) = c' \pi(\phi_1, \phi_2, \psi) = c' \cdot \sum_{y=0,1} \int_{]v_L, v_H[\times S_{x,z}} z'(y - \mathbf{1}(v > 0)) \phi_1^2 \phi_2^2 dv dx dz. \quad (17)$$

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

10 To implement this technique, the issue is to prove that ρ is pathwise differentiable at
 11 $(\phi_1^0, \phi_2^0, \psi^0)$, to prove that the pathwise derivative of $\rho(\phi_1, \phi_2, \psi)$ at $(\phi_1^0, \phi_2^0, \psi^0)$ (denoted
 12 $\nabla \rho_0$) is a continuous linear functional, to find $\phi^* = (\phi_1^*, \phi_2^*, \psi^*)$ the representer of $\nabla \rho_0$ and
 13 to compute $\|\phi^*\|_F$. Severini and Tripathi (2001) show that the lower bound for the
 14 asymptotic variance of root-n consistent regular estimators of $\rho(\phi_1^0, \phi_2^0, \psi^0)$ is actually
 15 $\|(\phi_1^*, \phi_2^*, \psi^*)\|_F$, i.e., the Fisher information norm of $(\phi_1^*, \phi_2^*, \psi^*)$ the representer of $\nabla \rho_0$.

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

18 For some $t_0 > 0$ let $\varphi(t) : t \rightarrow (\phi_{1t}, \phi_{2t}, \psi_t)$ be a curve from $[0, t_0]$ into E such that $\varphi(t)$
 19 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
 20 definition $\dot{\varphi}$ corresponds to the slope of $\varphi(t)$ at $t = 0$ (i.e., $\|(\varphi(t) - \varphi(0))/t - \dot{\varphi}\|_{L^2} \rightarrow 0$
 21 when $t \downarrow 0$).

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

$$25 \quad \nabla \rho_0(\dot{\phi}_1, \dot{\phi}_2, \dot{\psi}) = c' \sum_{y=0,1} \int_{]v_L, v_H[\times S_{x,z}} z'(y - \mathbf{1}(v > 0)) 2(\dot{\phi}_1 \phi_2^0 + \dot{\phi}_2 \phi_1^0) \phi_1^0 \phi_2^0 dv dx dz. \quad (18)$$

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

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

$$33 \quad \nabla \rho_0(\dot{\phi}_1, \dot{\phi}_2, \dot{\psi}) = \langle (\dot{\phi}_1, \dot{\phi}_2, \dot{\psi}), (\phi_1^*, \phi_2^*, \psi^*) \rangle_F.$$

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

$$35 \quad \nabla \rho_0(\dot{\phi}_1, \dot{\phi}_2, \dot{\psi}) = 2c' \sum_{y=0,1} \int_{]v_L, v_H[\times S_{x,z}} z' \left[\frac{(y - \mathbf{1}(v > 0))}{\psi^2} \psi^2 (\phi_2^0)^2 \right] \dot{\phi}_1 \phi_1^0 dv dx dz$$

$$37 \quad + 2c' \int_{S_{x,z}} \dot{\phi}_2 \phi_2^0 \sum_{y=0,1} \int_{]v_L, v_H[} \left[\frac{(y - \mathbf{1}(v > 0))}{\psi^2} \psi^2 (\phi_1^0)^2 \right] dv dx dz.$$

41 Hence, we have,

$$43 \quad \nabla \rho_0(\dot{\phi}_1, \dot{\phi}_2, \dot{\psi}) = 2c' \sum_{y=0,1} \left(\int_{]v_L, v_H[\times S_{x,z}} z' [\tilde{y} \psi^2 (\phi_2^0)^2] dv dx dz \right) \dot{\phi}_1 \phi_1^0$$

$$45 \quad + 2c' \int_{S_{x,z}} \left(\sum_{y=0,1} \int_{]v_L, v_H[} z' [\tilde{y} \psi^2 (\phi_1^0)^2] dv \right) \dot{\phi}_2 \phi_2^0 dx dz.$$

47

1 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
 2 fact that $\int_{S_{x,z}} \phi_2^0 \dot{\phi}_2 dx dz = 0$ 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,
 3 we can see that any $(\phi_1^*, \phi_2^*, \psi^*)$ such that

$$5 \quad \phi_1^*(y | v, x, z) = \frac{1}{2} c' z \tilde{y} \phi_1^0 + A_1(x, v, z) \phi_1^0,$$

$$7 \quad \phi_2^*(x, z) = \frac{1}{2} c' \cdot (z' E(\tilde{y} | x, z) \phi_2^0 + A_2 \phi_2^0).$$

$$9 \quad \psi^*(v | x, z) = 0,$$

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

$$11 \quad \langle (\dot{\phi}_1, \dot{\phi}_2, \dot{\psi}), (\phi_1^*, \phi_2^*, \psi^*) \rangle_F = \nabla \rho_0(\dot{\phi}_1, \dot{\phi}_2, \dot{\psi}).$$

13 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., $\int_{S_{x,z}} \phi_2^0 \phi_2^* dx dz = 0$ and $\sum_{y=0,1} \phi_1^0(y | v, x, z) \cdot \phi_1^* = 0$. These two imply,

$$15 \quad A_1(x, v, z) = -\frac{1}{2} c' E(z' \tilde{y} | v, x, z) \quad \text{and} \quad A_2 = -\frac{1}{2} c' \cdot E(z' \tilde{y}) = -\frac{1}{2} c' \cdot \pi_0.$$

17 Thus, we necessarily have

$$19 \quad \phi_1^*(y | v, x, z) = \frac{1}{2} c' z' (\tilde{y} - E(\tilde{y} | v, x, z)) \phi_1^0,$$

$$21 \quad \phi_2^*(x, z) = \frac{1}{2} c' \cdot (z' E(\tilde{y} | x, z) - \pi_0) \phi_2^0.$$

$$23 \quad \psi^*(v | x, z) = 0.$$

We have just found a vector $(\phi_1^*, \phi_2^*, \psi^*)$ in the tangent space which satisfies
 25 $\nabla \rho_0(\dot{\phi}_1, \dot{\phi}_2, \dot{\psi}) = \langle (\dot{\phi}_1, \dot{\phi}_2, \dot{\psi}), (\phi_1^*, \phi_2^*, \psi^*) \rangle_F$ for any $(\dot{\phi}_1, \dot{\phi}_2, \dot{\psi})$ in the tangent space. Using
 again the Riesz–Fréchet theorem, this result proves that the linear operator $\nabla \rho_0$ is
 27 continuous and that $(\phi_1^*, \phi_2^*, \psi^*)$ is its representer.

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

$$29 \quad \begin{aligned} \|\phi_1^{*2}\|_F + \|\phi_2^{*2}\|_F &= c' E(z' (\tilde{y} - E(\tilde{y} | v, x, z))^2 z) c \\ &\quad + c' E((z' E(\tilde{y} | x, z) - \pi_0)(z' E(\tilde{y} | x, z) - \pi_0)') c \\ &= c' E(z' (\tilde{y} - E(\tilde{y} | v, x, z))^2 z) c \\ &\quad + c' E(z' (E(\tilde{y} | x, z) - x\beta_0)^2 z) c \\ &= c' E(z' (\tilde{y} - E(\tilde{y} | v, x, z) + E(\tilde{y} | x, z) - x\beta_0)^2 z) c, \end{aligned}$$

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

$$37 \quad E(z' (\tilde{y} - E(\tilde{y} | v, x, z) + E(\tilde{y} | x, z) - x\beta_0)^2 z).$$

39

41 C.2. 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
 43 any differentiable path indexed by t and where $t = 0$ gives π_0 . For simplicity, denote u as
 the functionally independent representation of (x, z) :

$$47 \quad \pi_t = \int z' \frac{y - \mathbf{1}\{v > 0\}}{f_i(v | u)} f_i(\varepsilon, v, u) d\varepsilon dv du.$$

1 Therefore

$$3 \quad \pi_t = \int z'(y - \mathbf{1}\{v > 0\})f_t(\varepsilon | v, u)f_t(u) d\varepsilon dv du.$$

5 Under regularity conditions given by [Newey \(1994\)](#), formal differentiation with respect to t yields:

$$7 \quad \left. \frac{\partial \pi_t}{\partial t} \right|_{t=0} = \int z'(y - \mathbf{1}\{v > 0\}) \frac{\partial}{\partial t} (f_t(\varepsilon | v, x, z)f_t(x, z)) d\varepsilon dv du$$

$$9 \quad = \int z'(y - \mathbf{1}\{v > 0\}) \left(\frac{\partial}{\partial t} \ln f_t(\varepsilon | v, u) + \frac{\partial}{\partial t} \ln f_t(u) \right) f_0(\varepsilon | v, u)f_0(u) d\varepsilon dv du,$$

$$11 \quad \left. \frac{\partial \pi_t}{\partial t} \right|_{t=0} = \mathbb{E} \left[z' \frac{y - \mathbf{1}\{v > 0\}}{f_0(v | u)} \cdot \left(\frac{\partial}{\partial t} \ln f_t(\varepsilon | v, u) + \frac{\partial}{\partial t} \ln f_t(u) \right) \right]$$

$$13 \quad = \mathbb{E} \left[z' \tilde{y} \cdot \left(\frac{\partial}{\partial t} \ln f_t(\varepsilon, v, u) - \frac{\partial}{\partial t} \ln f_t(v, u) + \frac{\partial}{\partial t} \ln f_t(u) \right) \right]$$

$$15 \quad = \mathbb{E}[z' \tilde{y} \cdot S(\varepsilon, v, u)] - \mathbb{E} \left[z' \tilde{y} \cdot \frac{\partial}{\partial t} \ln f_t(v, u) \right] + \mathbb{E} \left[z' \tilde{y} \cdot \frac{\partial}{\partial t} \ln f_t(u) \right]$$

$$17 \quad = \mathbb{E}[z' \tilde{y} \cdot S(\varepsilon, v, u)] - \mathbb{E}[z' \mathbb{E}(\tilde{y} | v, u) \cdot S(v, u)] + \mathbb{E}[z' \mathbb{E}(\tilde{y} | u) \cdot S(u)],$$

21 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
 23 $\phi(v, u)$

$$25 \quad \mathbb{E}(\phi(v, u)S(v, u)) = \mathbb{E}(\phi(v, u)S(\varepsilon, v, u))$$

we therefore have

$$27 \quad \left. \frac{\partial \pi_t}{\partial t} \right|_{t=0} = \mathbb{E}[z'(\tilde{y} - \mathbb{E}(\tilde{y} | v, u) + \mathbb{E}(\tilde{y} | u)) \cdot S(\varepsilon, v, u)]$$

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

$$31 \quad q = z'(\tilde{y} - \mathbb{E}(\tilde{y} | v, u) + \mathbb{E}(\tilde{y} | u) - x\beta_0)$$

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

35 C.3. Theorem 8

37 When $f(v | x, z)$ 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
 39 random variable:

$$41 \quad q = z'(\tilde{y} - \mathbb{E}(\tilde{y} | v, x, z) + \mathbb{E}(\tilde{y} | x, z) - x\beta_0).$$

When $f(v | x, z)$ is known, the variance is the usual GMM variance–covariance matrix of

$$43 \quad q_0 = z'(\tilde{y} - x\beta_0).$$

Note that it is the same variable \tilde{y} which is used here since we deal with asymptotics and
 45 $\hat{f}(v | x, z)$ is consistent for $f(v | x, z)$. Denote

$$47 \quad \eta_0 = \tilde{y} - x\beta_0$$

1 and write

$$3 \quad q = z'(\eta_0 - E(\eta_0 | v, x, z) + E(\eta_0 | x, z)).$$

4 Consider

$$5 \quad \eta = \eta_0 - E(\eta_0 | v, x, z) + E(\eta_0 | x, z)$$

6 so that we can write

$$7 \quad Vq_0 = E(z'.E((\eta_0)^2 | v, x, z).z),$$

$$9 \quad Vq = E(z'.E((\eta)^2 | v, x, z).z).$$

10 Some algebra yields

$$\begin{aligned} 13 \quad E((\eta)^2 | x, z, v) &= E[(\eta_0 - E(\eta_0 | v, x, z) + E(\eta_0 | x, z))^2 | v, x, z] \\ &= E[(\eta_0)^2 + (E(\eta_0 | v, x, z))^2 + (E(\eta_0 | x, z))^2 | v, x, z] \\ 15 \quad &\quad - 2E[\eta_0 E(\eta_0 | v, x, z) | v, x, z] + 2E[\eta_0 E(\eta_0 | x, z) | v, x, z] \\ 17 \quad &\quad - 2E(\eta_0 | v, x, z)E(\eta_0 | x, z) \\ 19 \quad &= E[(\eta_0)^2 | v, x, z] - (E(\eta_0 | v, x, z))^2 + (E(\eta_0 | x, z))^2. \end{aligned}$$

Therefore

$$21 \quad \Delta = Vq_0 - Vq = E(z'.[(E(\eta_0 | x, z, v))^2 - (E(\eta_0 | x, z))^2].z).$$

22 As we can write

$$23 \quad E(\eta_0 | x, z, v) = E(\eta_0 | x, z) + \eta_1,$$

24 where $E(\eta_1 | x, z) = 0$, we have

$$27 \quad E(\eta_0 | x, z, v)^2 = E(\eta_0 | x, z)^2 + (\eta_1)^2 + 2E(\eta_0 | x, z)\eta_1$$

28 and therefore:

$$\begin{aligned} 29 \quad \Delta &= Vq_0 - Vq = E(z'.[(\eta_1)^2 + 2\eta_1 E(\eta_0 | x, z)].z) \\ 31 \quad &= E(z'.(\eta_1)^2.z) + 2E(z'.\eta_1 E(\eta_0 | x, z).z) \\ 33 \quad &= E(z'.(\eta_1)^2.z) + 2E(z'.E(\eta_1 | x, z)E(\eta_0 | x, z).z) \\ &= E(z'.(\eta_1)^2.z) \end{aligned}$$

34 is a semi-definite positive matrix.

35 Finally observe that

$$37 \quad E((\eta_1)^2 | z) = V \left[\frac{G(v, x, z) - \mathbf{1}\{v > 0\}}{f(v | x, z)} \mid z \right]$$

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

42 References

- 43 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.
- 44 Billingsley, P., 1995. Probability and Measure. Wiley, New York.

- 1 Blundell, R., Powell, J.L., 2004. Endogeneity in semiparametric binary response models. *Review of Economic Studies* 71, 655–679.
- 3 Cavanagh, C., Sherman, R.P., 1998. Rank estimators for monotone index models. *Journal of Econometrics* 84, 351–381.
- 5 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.
- 7 Chen, S., 2002. Semiparametric estimation of a heteroscedastic binary choice model. Working Paper, Hong Kong University.
- 9 Cogneau, D., Maurin, E., 2002. Parental income and school attendance in a low-income country, Unpublished manuscript.
- 11 Cosslett, S.R., 1983. Distribution-free maximum likelihood estimator of the binary choice model. *Econometrica* 51, 765–782.
- 13 Cosslett, S.R., 1987. Efficiency bounds for distribution-free estimators of the binary choice and the censored regression models. *Econometrica* 55, 559–585.
- 15 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.
- 17 Han, A.K., 1987. Non-parametric analysis of a generalized regression model. *Journal of Econometrics* 35, 303–316.
- 19 Hirano, K., Imbens, G.W., Ridder, G., 2003. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* 71, 1161–1189.
- 21 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.
- 23 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.
- 25 Ichimura, H., 1993. Semiparametric least squares (SLS) and weighted SLS estimation of single-index model. *Journal of Econometrics* 58, 71–120.
- 27 Khan, S., Lewbel, A., 2003. Weighted and two stage least squares estimation of semiparametric truncated regression models. Working Paper, Boston College.
- 29 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.
- 31 Lewbel, A., 2000. Semiparametric qualitative response model estimation with unknown heteroskedasticity or instrumental variables. *Journal of Econometrics* 97, 145–177.
- 33 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.
- 35 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>).
- 37 Magnac, T., Maurin, E., 2004. Partial identification in monotone binary models: discrete and interval valued regressors. Working Paper, CREST, no. 2004-11.
- 39 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.
- 41 Manski, C.F., 1988. Identification of binary response models. *Journal of the American Statistical Association* 83, 729–738.
- 43 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.
- 45 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.
- 47

- 1 Powell, J., 1986. Symmetrically trimmed least squares estimation for tobit models. *Econometrica* 54, 1435–1460.
- 2 Powell, J., 1994. Estimation of semiparametric models. In: Engle, R., McFadden, D. (Eds.), *Handbook of*
3 *Econometrics*, vol. 4. North-Holland, Amsterdam, pp. 2444–2521.
- 4 Powell, J.L., Stock, J.H., Stoker, T.M., 1989. Semi-parametric estimation of index coefficients. *Econometrica* 57,
5 1403–1430.
- 6 Ruud, P.A., 1983. Sufficient conditions for the consistency of maximum likelihood estimation despite
7 misspecification of distribution in multinomial discrete choice models. *Econometrica* 51, 225–228.
- 8 Severini, T.A., Tripathi, G., 2001. A simplified approach to computing efficiency bounds in semiparametric
9 models. *Journal of Econometrics* 102, 23–66.

UNCORRECTED PROOF