IDENTIFICATION IN BIVARIATE BINARY-CHOICE MODELS WITH ELLIPTICAL INNOVATIONS

Koen Jochmans*
CORE

April 15, 2011

Abstract: In this note it is shown that the index coefficients and location parameters in the standard triangular binary-choice model are identified under an assumption of symmetry on the joint density of the latent disturbances. Identification of average effects follows. The implied restrictions suggest semiparametric rank estimators that are $\sqrt{n}$-consistent and asymptotically normal under standard conditions.

JEL classification: C14, C31, C35
Keywords: binary choice, rank estimation, simultaneity, symmetry, truncation.

Introduction

A difficult issue in microeconometrics is the non- and semiparametric identification of causal parameters in triangular limited dependent-variable models (see, e.g., Chesher, 2007). When endogenous variables exhibit discrete variation, such parameters are generally set- rather than point identified (see, e.g., Chesher, 2005, 2010). The problem appears most severe in a bivariate binary-choice model—a cornerstone model for empirical practice—where the attention has shifted toward inferring local average treatment-effect (LATE) parameters (Imbens and Angrist, 1994). Vytlacil and Yildiz (2007) showed how both the average structural function and the average treatment effect can be recovered in a specification featuring weak separability and large-support conditions. Yildiz (2004) suggested a multistep estimator for a linear-index version of their model. Shaikh and Vytlacil (2010) showed that omitting the support condition results in set identification. Here, I complement these analyses with the finding that point identification in the benchmark bivariate binary-choice model can also be achieved when the disturbances are known to be drawn from an elliptical distribution. The analogy principle leads to rank-based estimators whose large-sample properties are easy to analyze.

1 Information and identification

Let the observable random variable $W \equiv (Y, D, X, Z)$ have distribution P, supported on $\mathscr{W} \equiv \{0,1\} \times \{0,1\} \times \mathscr{X} \times \mathscr{Z}$. The canonical bivariate binary-choice model (see, e.g., Heckman, 1978) takes the form

$$Y = 1\{X\beta + D\delta \geq \mu_U + U\}, \quad D = 1\{Z\gamma \geq \mu_V + V\}, \quad (U,V) \perp (X,Z), \tag{1}$$

for conformable unknown χ and z vectors β and γ, and scalars δ, μ_U, and μ_V. Assume that the density of (U,V), $f_{U,V}$, is absolutely continuous and symmetric, that is, $f_{U,V}(u,v) = f_{U,V}(-u,-v)$. The centering

*Address: U.C. Louvain, Center for Operations Research and Econometrics, Voie du Roman Pays 34, B-1348 Louvain-la-Neuve, Belgium. Tel. +32 10 474329; E-mail: koen.jochmans@uclouvain.be.

of the density at zero is innocuous due to the inclusion of the location parameters μ_V and μ_U. The first components of both β and γ are normalized to unity. This is without loss of generality provided that their true value is non-zero. Consequently, identification and estimation statements concerning β and γ below relate to their last $\chi - 1$ and $z - 1$ components, respectively.

The specification in (1) entails sign restrictions that provide non-trivial information on the unknown parameters in both equations. A familiar support condition on X and Z implies these to be sufficiently informative to point identify the index coefficients and the location parameters.

First-stage equation. Because of the triangular structure of (1), identifying γ and μ_V poses little complication. On letting F_V be the marginal distribution of V,

$$E[D|Z = z] = F_V(z\gamma - \mu_V) = 1 - F_V(\mu_V - z\gamma).$$

The first equality follows from the independence of V and Z. The second transition stems from the fact that symmetry of $f_{U,V}$ implies symmetry of its marginals. Let $\operatorname{sgn}\{\cdot\}$ be the sign function. Then, for each (z_1, z_2) in $\mathscr{Z} \otimes \mathscr{Z}$,

$$\operatorname{sgn}\{E[D|Z = z_1] - E[D|Z = z_2]\} = \operatorname{sgn}\{(z_1 - z_2)\gamma\} \tag{2}$$

because $F_V(\cdot)$ is monotone (Han, 1987), and

$$\operatorname{sgn}\{E[D|Z = z_1] - E[(1 - D)|Z = z_2]\} = \operatorname{sgn}\{(z_1 + z_2)\gamma - 2\mu_V\} \tag{3}$$

because $F_V(\cdot)$ is symmetric (Chen, 2000). Suppose that the first component of Z has an everywhere-positive Lebesgue density given realizations of the remaining $z - 1$ components and suppose that $\mathscr{Z}$ is not contained in a linear subspace of $\mathscr{R}^z$. Then (2) and (3) point identify γ and μ_V (see, e.g., Manski, 1985, Han, 1987).

Second-stage equation. Introduce the shorthand notation $Y^* \equiv X\beta - \mu_U$ and $D^* \equiv Z\gamma - \mu_V$ and define $C(\tau) \equiv 1\{-|\tau| < V \leq |\tau|\}$. By index sufficiency,

$$E[Y|Y^* = \iota, D = 1, C(\tau) = 1] = \frac{E[YD|Y^* = \iota, D^* = \tau] - E[YD|Y^* = \iota, D^* = -\tau]}{\operatorname{sgn}\{\tau\} \Pr[C(\tau) = 1]}, \tag{4}$$

$$E[Y|Y^* = \iota, D = 0, C(\tau) = 1] = \frac{E[Y(1 - D)|Y^* = \iota, D^* = \tau] - E[Y(1 - D)|Y^* = \iota, D^* = -\tau]}{\operatorname{sgn}\{-\tau\} \Pr[C(\tau) = 1]}. \tag{5}$$

For each non-zero τ in the support of D^*, the indicator $C(\tau)$ defines the subpopulation of *compliers* (Angrist, Imbens, and Rubin, 1996) associated with shifting the propensity score of D from $-|\tau|$ to $|\tau|$. The relative size of this subpopulation is identified as $\Pr[C(\tau) = 1] = \operatorname{sgn}\{\tau\} \big[E[D|D^* = \tau] - E[D|D^* = -\tau]\big]$. So the conditional expectations for compliers are identified from the right-hand sides of (4) and (5) provided that the value of β can be learned in the population, which is the case (see (8)). Within each complier population (i.e., for each τ), D is exogenous. Moreover, (4) and (5) can be compactly expressed as $E[Y|Y^* = \iota, D = d, C(\tau) = 1] = \Pr[U \leq \iota + d\delta - \mu_U| - |\tau| < V \leq |\tau|]$, so that

$$\text{sgn}\{E[Y|Y^* = \iota_1, D = d_1, C(\tau) = 1] - E[Y|Y^* = \iota_2, D = d_2, C(\tau) = 1]\} = \text{sgn}\{(x_1 - x_2)\beta + (d_1 - d_2)\delta\} \quad (6)$$

for each pair (w_1, w_2) in $\mathscr{W} \otimes \mathscr{W}$. Furthermore, given $C(\tau) = 1$, U is symmetrically distributed around zero, as

$$\Pr[U \leq u|C(\tau) = 1] = \Pr[U \leq u| - |\tau| < V \leq |\tau|] = \Pr[U > -u| - |\tau| < V \leq |\tau|] = \Pr[U > -u|C(\tau) = 1].$$

Consequently,

$$\begin{gathered} \text{sgn}\{E[Y|Y^* = \iota_1, D = d_1, C(\tau) = 1] - E[(1 - Y)|Y^* = \iota_2, D = d_2, C(\tau) = 1]\} \\ = \\ \text{sgn}\{(x_1 + x_2)\beta + (d_1 + d_2)\delta - 2\mu_U\}. \end{gathered} \quad (7)$$

The sign restrictions in (6) and (7) provide information on all parameters of the second-stage equation. Here, restricting attention to compliers can be understood as an artificial-truncation argument, similar in spirit as Powell (1986) and Honoré (1992). Symmetry of the error distribution is restored by correcting for the presence of *always-takers* and *never-takers* (see, again, Angrist, Imbens, and Rubin, 1996). This is apparent from the right-hand side of both (4) and (5).

Identifying the expected value of Y for a complier group given realizations of Y^* in stead of X does not preclude variation in X but requires β to be identified. Index sufficiency implies that

$$\frac{E[YD|Y^* = \iota, D^* = \tau]}{E[D|D^* = \tau]} = \Pr[U \leq \iota + \delta|V \leq \tau], \quad \frac{E[Y(1 - D)|Y^* = \iota, D^* = \tau]}{E[(1 - D)|D^* = \tau]} = \Pr[U \leq \iota|V > \tau],$$

regardless of symmetry. Therefore, given $D = d$ and $D^* = \tau$, variation in X is exogenous and

$$\text{sgn}\{E[Y|Y^* = \iota_1, D = d, D^* = \tau] - E[Y|Y^* = \iota_2, D = d, D^* = \tau]\} = \text{sgn}\{(x_1 - x_2)\beta\} \quad (8)$$

by a standard application of control-function arguments.

Using variation within complier populations allows to disentangle δ and μ_U. By itself, symmetry, at best, provides information on a linear combination of these parameters. This is so because mirroring $f_{U,V}$ around the origin requires changing D. Recall that $E[YD|Y^* = \iota_1, D^* = \tau] = \Pr[U \leq \iota_1 + \delta, V \leq \tau]$ and that

$$E[(1 - Y)(1 - D)|Y^* = \iota_2, D^* = -\tau] = \Pr[U > \iota_2, V > -\tau] = \Pr[U \leq -\iota_2, V \leq \tau].$$

So,

$$\text{sgn}\{E[YD|Y^* = \iota_1, D^* = \tau] - E[(1 - Y)(1 - D)|Y^* = \iota_2, D^* = -\tau]\} = \text{sgn}\{\iota_1 + \iota_2 + \delta - 2\mu_U\} \quad (9)$$

and, by an analogous argument, it is readily established that

$$\text{sgn}\{E[Y(1 - D)|Y^* = \iota_1, D^* = \tau] - E[(1 - Y)D|Y^* = \iota_2, D^* = -\tau]\} = \text{sgn}\{\iota_1 + \iota_2 + \delta - 2\mu_U\}. \quad (10)$$

These moment conditions do not allow to separately learn δ and μ_U.

For the above sign restrictions to be powerful enough to uniquely pin down all of β, δ, and μ_U X needs to be able to shift Y^* sufficiently given $D^* = \tau$ for all τ. Let the first component of X have an everywhere-positive Lebesgue density given realizations of both the $\chi - 1$ remaining components and D^*. Suppose that $\mathscr{X}$ is not contained in a linear subspace of $\mathscr{R}^{\chi}$. Then (6), (7), (8), and (9) point identify all second-stage parameters. The proof to this claim is virtually identical as the argument in the standard case (see, again, Manski, 1985 or Han, 1987).

Average effects. Given identification of the coefficients, one can learn policy parameters that involve averages with respect to the marginal distribution of U. To illustrate, consider the average structural function at $(X = x, D = 1)$, that is, $\int 1\{\iota_1 + \delta \geq u\} \; \mathrm{d}F_U(u)$ for $\iota_1 = x\beta - \mu_U$. By the law of total probability, it can equivalently be expressed as an average over D^* of

$$\int 1\{\iota_1 + \delta \geq u\} \; \mathrm{d}F_U(u|V \leq \tau) \;\; \Pr[D = 1|D^* = \tau] + \int 1\{\iota_2 \geq u\} \; \mathrm{d}F_U(u|V > \tau) \;\; \Pr[D = 0|D^* = \tau],$$

for $\iota_2 = \iota_1 + \delta$. The first integral is nonparametrically identified, and so is the propensity score for D. The second integral can be computed in the population as $E[Y|Y^* = \iota_2, D = 0, D^* = \tau]$, and is thus also identified. Identification of other average-effect parameters follows in the same way.

2 Estimation

Suppose throughout a random sample of size n has been drawn from P. Let $W_i \equiv (Y_i, D_i, X_i, Z_i)$, $i = 1, \ldots, n$, denote the realizations.

First-stage equation. Han (1987) (and later also Cavanagh and Sherman, 1998) used (2) and suggested inferring β by maximizing

$$q_\gamma(g) \equiv \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} D_i(1 - D_j) \; 1\{(Z_i - Z_j)g > 0\} + (1 - D_i)D_j \; 1\{(Z_i - Z_j)g < 0\}$$

with respect to g. Sherman (1993) gave conditions under which doing so leads to a $\sqrt{n}$-consistent and asymptotically-normal estimator. The asymptotic efficiency of this procedure can be improved by the weighting argument in Subbotin (2008).

Chen (2000) utilized (3) and proposed to maximize

$$q_{\gamma,\mu_V}(g, \mu) \equiv \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} D_i D_j \; 1\{(Z_i + Z_j)g > 2\mu\} + (1 - D_i)(1 - D_j) \; 1\{(Z_i + Z_j)g < 2\mu\}$$

either with respect to μ using an asymptotically-linear estimator of γ or jointly over (g, μ). Both these procedures yield $\sqrt{n}$-consistent and asymptotically-normal estimators under the conditions outlined in Chen (2000).

A third option would be to set

$$(\widehat{\gamma}, \widehat{\mu}_V) \equiv \arg\max \, (q_\gamma(g) + q_{\gamma,\mu_V}(g,\mu)).$$

Because (2) is not redundant given (3), this estimator will be more efficient than the simultaneous procedure based on $q_{\gamma,\mu_V}(g,\mu)$ alone.

Second-stage equation. For each $i = 1, \ldots, n$, construct $\widehat{D}_i^* \equiv Z_i\widehat{\gamma} - \widehat{\mu}_V$. Consider a symmetric univariate kernel function $K\{\cdot\}$ and a bandwidth $\sigma = \sigma(n)$ that satisfies $\lim_{n\to\infty} \sigma = \infty$. Define the weight $\widehat{\omega}_{i,j}^- \equiv \sigma K\{\sigma(\widehat{D}_i^* - \widehat{D}_j^*)\}$. An estimator of β based on (9) in the spirit of Han's (1987) rank estimator is $\widehat{\beta} \equiv \arg\max q_\beta(b)$, for

$$q_\beta(b) \equiv \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j\neq i} \left[Y_i(1-Y_j)\, 1\{(X_i - X_j)b > 0\} + Y_j(1-Y_i)\, 1\{(X_i - X_j)b < 0\}\right] \widehat{\omega}_{i,j}^-.$$

The large-sample behavior of this estimator of β—$\sqrt{n}$-consistency and asymptotic normality under the appropriate conditions—and its asymptotic covariance matrix follow from Theorem 2 and Proposition 2 in Jochmans (2011), respectively.

For each $i = 1, \ldots, n$, use $\widehat{\beta}$ to construct $\widehat{Y}_i^* \equiv X_i\widehat{\beta}$. Observe that, contrary to Y_i^*, $\widehat{Y}_i^*$ does not include and estimate of μ_U. Let $\widehat{\lambda}_{i,j} \equiv \sigma^2 K\{\sigma(\widehat{Y}_i^* - \widehat{Y}_j^*)\}K\{\sigma(\widehat{D}_i^* + \widehat{D}_j^*)\}$. The use of a product kernel is not crucial. To implement the restrictions in (4) and (5), let I be an indicator that can take on the values zero and one. For each i, construct the outcomes

$$\widehat{S}_i(I) \equiv Y_i[ID_i + (1-I)(1-D_i)] - \frac{\frac{1}{n-1}\sum_{j\neq i} Y_j[ID_j + (1-I)(1-D_j)]\, \widehat{\lambda}_{i,j}}{\frac{1}{n-1}\sum_{j\neq i}\widehat{\lambda}_{i,j}},$$

$$\widehat{F}_i(I) \equiv (1-Y_i)[ID_i + (1-I)(1-D_i)] - \frac{\frac{1}{n-1}\sum_{j\neq i}(1-Y_j)[ID_j + (1-I)(1-D_j)]\, \widehat{\lambda}_{i,j}}{\frac{1}{n-1}\sum_{j\neq i}\widehat{\lambda}_{i,j}}.$$

Abstract away from any need for trimming to keep the denominator well defined. Standard smoothness conditions yield $\widehat{S}_i(I) \xrightarrow{p} S_i(I) \equiv Y_i[ID_i + (1-I)(1-D_i)] - E[Y[ID + (1-I)(1-D)]|Y^* = Y_i^*, D^* = -D_i^*]$ and also $\widehat{F}_i(I) \xrightarrow{p} F_i(I)$, in obvious notation. So,

$$E[S(I)|Y^* = Y_i^*, D^* = D_i^*] \propto \text{sgn}\{ID_i^* - (1-I)D_i^*\}\; E[Y|Y^* = Y_i^*, D = I, C(D_i^*) = 1],$$

and similarly for $F(I)$. The factor of proportionality is $\Pr[C(D_i^*) = 1]$ and is irrelevant for our purposes.

The generated outcomes $\widehat{S}_i(I)$ and $\widehat{F}_i(I)$ can be used to construct a criterion function for (δ, μ_U). By analogy to $\widehat{\omega}_{i,j}^-$, let $\widehat{\omega}_{i,j}^+ \equiv \sigma K\{\sigma(\widehat{D}_i^* + \widehat{D}_j^*)\}$. Consider

$$\zeta_{i,j}^{\delta}(d) \equiv \left[\widehat{S}_i(1)\, 1\{\widehat{Y}_i^* - \widehat{Y}_j^* > -d\} + \widehat{S}_j(0)\, 1\{\widehat{Y}_i^* - \widehat{Y}_j^* < -d\}\right] \widehat{\omega}_{i,j}^+ \,\text{sgn}\{\widehat{D}_i^* - \widehat{D}_j^*\},$$

$$\zeta_{i,j}^{\mu_U}(\mu) \equiv \left[\widehat{F}_i(0)\, 1\{\widehat{Y}_i^* + \widehat{Y}_j^* > 2\mu\} + \widehat{S}_j(0)\, 1\{\widehat{Y}_i^* + \widehat{Y}_j^* < 2\mu\}\right] \widehat{\omega}_{i,j}^- \,\text{sgn}\{\widehat{D}_i^* + \widehat{D}_j^*\},$$

whose average over the empirical product measure serve as the data counterparts to (6) when either μ_U or δ is differenced out. While these random variables suffice to infer the remaining unknowns, they can be complemented with the information in (9) and (10), which is non-redundant. Define

$$\zeta_{i,j}^{\delta,\mu_U}(d,\mu) \equiv \left[\widehat{S}_i(1)\ 1\{\widehat{Y}_i^* + \widehat{Y}_j^* > 2(\mu-d)\} + \widehat{F}_j(1)\ 1\{\widehat{Y}_i^* + \widehat{Y}_j^* < 2(\mu-d)\}\right] \widehat{\omega}_{i,j}^{-}\ \mathrm{sgn}\{\widehat{D}_i^* + \widehat{D}_j^*\},$$
$$\zeta_{i,j}^{\mu_U,\delta}(\mu,d) \equiv \left[\ Y_i \quad 1\{\widehat{Y}_i^* + \widehat{Y}_j^* > 2\mu - d\ \ \} + (1-Y_j)\ 1\{\widehat{Y}_i^* + \widehat{Y}_j^* < 2\mu - d\}\right] \widehat{\omega}_{i,j}^{+} \quad 1\{D_i \neq D_j\}.$$

An objective function that can form the basis for inference on (δ, μ_U) then follows as

$$q_{\delta,\mu_U}(d,\mu) \equiv \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j\neq i} \zeta_{i,j}^{\delta}(\mu) + \zeta_{i,j}^{\mu_U}(\mu) + \zeta_{i,j}^{\delta,\mu_U}(d,\mu) + \zeta_{i,j}^{\mu_U,\delta}(\mu,d),$$

and the $\sqrt{n}$-consistency and asymptotic normality of its maximizer follow from arguments similar to those for $\widehat{\beta}$. Like with the first-stage equation, estimation of β and (δ, μ_U) can also be done by maximizing jointly over all second-stage coefficients using $\widehat{\beta}$ from above to perform the matching. This procedure can be iterated in the matching parameter. $q_{\delta,\mu_U}(d,\mu)$ could also be complemented by $q_\beta(b)$ were such a strategy to be followed.

References

Angrist, J. D., G. W. Imbens, and D. B. Rubin. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91:444–455.

Cavanagh, C. and R. P. Sherman. (1998). Rank estimators for monotonic index models. *Journal of Econometrics*, 84:351–381.

Chen, S. (2000). Rank estimation of a location parameter in the binary choice model. *Journal of Econometrics*, 98:317–334.

Chesher, A. (2005). Nonparametric identification under discrete variation. *Econometrica*, 73:1525–1550.

Chesher, A. (2007). Identification of nonadditive structural functions. In R. W. Blundell, W. K. Newey, and T. Persson (eds.), *Advances In Economics and Econometrics*, volume III. Econometric Society, Cambridge University Press.

Chesher, A. (2010). Instrumental variable models for discrete outcomes. *Econometrica*, 78:575–601.

Heckman, J. J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica*, 46:931–959.

Han, A. K. (1987). Non-parametric analysis of a generalized regression model: The maximum rank correlation estimator. *Journal of Econometrics*, 35:303–316.

Honoré, B. E. (1992). Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects. *Econometrica*, 60:553–565.

Imbens, G. W. and J. D. Angrist. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62:467–475.

Jochmans, K. (2011). Pairwise-comparison estimation with nonparametric controls. Unpublished manuscript.

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

Powell, J. L. (1986). Symmetrically trimmed least squares estimation for tobit models. *Econometrica*, 54:1435–1460.

Shaikh, A. M. and E. J. Vytlacil. (2010). Partial identification in triangular systems of equations with binary dependent variables. *Econometrica*, forthcoming.

Sherman, R. P. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica*, 61:123–137.

Subbotin, V. (2008). Weighted rank estimators. Unpublished manuscript.

Vytlacil, E. and N. Yildiz. (2007). Dummy endogenous variables in weakly separable models. *Econometrica*, 75:757–779.

Yildiz, N. (2004) Estimation of binary choice models with linear index and dummy endogenous variables. Unpublished manuscript.