The Stata Journal (2008)
8, Number 2, pp. 190–220

SNP and SML estimation of univariate and bivariate binary-choice models

Giuseppe De Luca
University of Rome "Tor Vergata"
Rome, Italy
delucgius@yahoo.it

Abstract. We discuss the semi-nonparametric approach of Gallant and Nychka (1987, *Econometrica* 55: 363–390), the semiparametric maximum likelihood approach of Klein and Spady (1993, *Econometrica* 61: 387–421), and a set of new Stata commands for semiparametric estimation of three binary-choice models. The first is a univariate model, while the second and the third are bivariate models without and with sample selection, respectively. The proposed estimators are $\sqrt{n}$ consistent and asymptotically normal for the model parameters of interest under weak assumptions on the distribution of the underlying error terms. Our Monte Carlo simulations suggest that the efficiency losses of the semi-nonparametric and the semiparametric maximum likelihood estimators relative to a maximum likelihood correctly specified estimator of a parametric probit are rather small. On the other hand, a comparison of these estimators in non-Gaussian designs suggests that semi-nonparametric and semiparametric maximum likelihood estimators substantially dominate the parametric probit maximum likelihood estimator.

Keywords: st0144, snp, snp2, snp2s, sml, sml2s, binary-choice models, semi-nonparametric approach, SNP estimation, semiparametric maximum likelihood, SML estimation, Monte Carlo simulation

1 Introduction

The parameters of discrete-choice models are typically estimated by maximum likelihood (ML) after imposing assumptions on the distribution of the underlying error terms. If the distributional assumptions are correctly specified, then parametric ML estimators are known to be consistent and asymptotically efficient. However, as discussed at length in the semiparametric literature, departures from the distributional assumptions may lead to inconsistent estimation. This problem has motivated the development of several semiparametric estimation procedures which consistently estimate the model parameters under less restrictive distributional assumptions. Semiparametric estimation of binary-choice models has been considered by Manski (1975); Cosslett (1983); Gallant and Nychka (1987); Powell, Stock, and Stoker (1989); Horowitz (1992); Ichimura (1993); and Klein and Spady (1993), among others.

In this article, I discuss the semi-nonparametric (SNP) approach of Gallant and Nychka (1987), the semiparametric maximum likelihood (SML) approach of Klein and Spady (1993), and a set of new Stata commands for semiparametric estimation of univariate and bivariate binary-choice models. The SNP approach of Gallant and Nychka

(1987), originally proposed for estimation of density functions, was adapted to estimation of univariate and bivariate binary-choice models by Gabler, Laisney, and Lechner (1993) and De Luca and Peracchi (2007), respectively.[1] A generalization of the SML estimator of Klein and Spady (1993) for semiparametric estimation of bivariate binary-choice models with sample selection was provided by Lee (1995). The SNP and SML approaches differ from the parametric approach because they can handle a broader class of error distributions. The SNP and SML approaches differ from each other in how they approximate the unknown distributions. The SNP approach uses a flexible functional form to approximate the unknown distribution while the SML approach uses kernel functions.

The remainder of the article is organized as follows. In section 2, I briefly review parametric specification and ML estimation of three binary-choice models of interest. SNP and SML estimation procedures, the underlying identifiability restrictions, and the asymptotic properties of the corresponding estimators are discussed in sections 3 and 4, respectively. Section 5 describes the syntax and the options of the Stata commands, while section 6 provides some examples. Monte Carlo evidence on the small-sample performances of the SNP and SML estimators relative to the parametric probit estimator is presented in section 7.

2 Parametric ML estimation

A univariate binary-choice model is a model for the conditional probability of a binary indicator. This model is typically represented by the following threshold crossing model,

$$Y^* = \alpha + \boldsymbol{\beta}^\top \boldsymbol{X} + U \tag{1}$$

$$Y = 1(Y^* \geq 0) \tag{2}$$

where Y^* is a latent continuous random variable, $\boldsymbol{X}$ is a k vector of exogenous variables, $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta})$ is a $(k+1)$ vector of unknown parameters, and U is a latent regression error. The latent variable Y^* is related to its observable counterpart Y through the observation rules (2), where $1\{A\}$ is the indicator function of the event A. If the latent regression error U is assumed to follow a standardized Gaussian distribution, then model (1)–(2) is known as a probit model.[2] In this case, the log-likelihood function for a random sample of n observations $(Y_1, \boldsymbol{X}_1), \ldots, (Y_n, \boldsymbol{X}_n)$ is of the form

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{n} Y_i \ln \pi_i(\boldsymbol{\theta}) + (1 - Y_i) \ln \{1 - \pi_i(\boldsymbol{\theta})\} \tag{3}$$

where $\pi_i(\boldsymbol{\theta}) = \Pr(Y_i = 1|\boldsymbol{X}) = \Phi(\mu_i)$ is the conditional probability of observing a positive outcome, $\Phi(\cdot)$ is the standardized Gaussian distribution function, and $\mu_i =$

---

[1] Our Stata command for SNP estimation of univariate binary-choice models can be considered a specific version of the command provided by Stewart (2004) for SNP estimation of ordered-choice models. Nevertheless, the proposed routine is faster, more accurate, and allows more estimation options.

[2] The normalization of the variance is necessary because the vector of parameters $\boldsymbol{\theta}$ can be identified only up to a scale coefficient.

$\alpha + \boldsymbol{\beta}^\top \boldsymbol{X}_i$. An ML estimator of the parameter vector $\boldsymbol{\theta}$ can be obtained by maximizing the log-likelihood function (3) over the parameter space $\boldsymbol{\Theta} = \Re^{k+1}$.

If we are interested in modeling the joint probability of two binary indicators Y_1 and Y_2, a simple generalization of model (1)–(2) is the following bivariate binary-choice model

$$Y_j^* = \alpha_j + \boldsymbol{\beta}_j^\top \boldsymbol{X}_j + U_j \qquad j = 1,2 \tag{4}$$

$$Y_j = 1(Y_j^* \geq 0) \qquad j = 1,2 \tag{5}$$

where the Y_j^* are latent variables for which only the binary indicators Y_j can be observed, the $\boldsymbol{X}_j$ are k_j vectors of (not necessary distinct) exogenous variables, the $\boldsymbol{\theta}_j = (\alpha_j, \boldsymbol{\beta}_j)$ are $(k_j + 1)$ vectors of unknown parameters, and the U_j are latent regression errors. When U_1 and U_2 have a bivariate Gaussian distribution with zero means, unit variances, and correlation coefficient ρ, model (4)–(5) is known as a bivariate probit model. Because Y_1 and Y_2 are fully observable, the vectors of parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ can always be estimated consistently by separate estimation of two univariate probit models, one for Y_1 and one for Y_2. However, when the correlation coefficient ρ is different from zero, it is more efficient to estimate the two equations jointly by maximizing the log-likelihood function

$$\begin{aligned} L(\boldsymbol{\theta}) = \sum_{i=1}^n & Y_{i1} Y_{i2} \ln \pi_{i11}(\boldsymbol{\theta}) + Y_{i1}(1 - Y_{i2}) \ln \pi_{i10}(\boldsymbol{\theta}) + \\ & (1 - Y_{i1}) Y_{i2} \ln \pi_{i01}(\boldsymbol{\theta}) + (1 - Y_{i1})(1 - Y_{i2}) \ln \pi_{i00}(\boldsymbol{\theta}) \end{aligned} \tag{6}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \rho)$, and the probabilities underlying the four possible realizations of the two binary indicators Y_1 and Y_2 are given by[3]

$$\begin{aligned} \pi_{11}(\boldsymbol{\theta}) &= \Pr(Y_1 = 1, Y_2 = 1) = \Phi_2(\mu_1, \mu_2; \rho) \\ \pi_{10}(\boldsymbol{\theta}) &= \Pr(Y_1 = 1, Y_2 = 0) = \Phi(\mu_1) - \Phi_2(\mu_1, \mu_2; \rho) \\ \pi_{01}(\boldsymbol{\theta}) &= \Pr(Y_1 = 0, Y_2 = 1) = \Phi(\mu_2) - \Phi_2(\mu_1, \mu_2; \rho) \\ \pi_{00}(\boldsymbol{\theta}) &= \Pr(Y_1 = 0, Y_2 = 0) = 1 - \Phi(\mu_1) - \Phi(\mu_2) + \Phi_2(\mu_1, \mu_2; \rho) \end{aligned}$$

where $\Phi_2(\cdot, \cdot; \rho)$ is the bivariate Gaussian distribution function with zero means, unit variances, and correlation coefficient ρ, and $\mu_j = \alpha_j + \boldsymbol{\beta}_j^\top \boldsymbol{X}_j$. An ML estimator $\widehat{\boldsymbol{\theta}}$ maximizes the log-likelihood function (6) over the parameter space $\boldsymbol{\Theta} = \Re^{k_1+k_2+2} \times (-1, 1)$.

Consider a bivariate binary-choice model with sample selection where the indicator Y_1 is always observed, while the indicator Y_2 is assumed to be observed only for the subsample of n_1 observations (with $n_1 < n$) for which $Y_1 = 1$. The model can be written as

---

[3] In the following, the suffix i and the explicit conditioning on the vector of covariates $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are suppressed to simplify notation.

$$Y_j^* = \alpha_j + \boldsymbol{\beta}_j^\top \boldsymbol{X}_j + U_j \qquad j = 1, 2 \tag{7}$$

$$Y_1 = 1(Y_1^* \geq 0) \tag{8}$$

$$Y_2 = 1(Y_2^* \geq 0) \qquad \text{if } Y_1 = 1 \tag{9}$$

When the latent regression errors U_1 and U_2 have a bivariate Gaussian distribution with zero means, unit variances, and correlation coefficient ρ, model (7)–(9) is known as a bivariate probit model with sample selection. Unlike the case of full observability, the presence of sample selection has two important implications. First, ignoring the potential correlation between the two latent regression errors may lead to inconsistent estimates of $\boldsymbol{\theta}_2 = (\alpha_2, \boldsymbol{\beta}_2)$ and inefficient estimates of $\boldsymbol{\theta}_1 = (\alpha_1, \boldsymbol{\beta}_1)$. Second, identifiability of the model parameters requires imposing at least one exclusion restriction on the two sets of exogenous covariates $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ (Meng and Schmidt 1985). Construction of the log-likelihood function for joint estimation of the overall vector of model parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \rho)$ is straightforward after noticing that the data identify only three possible events: $(Y_1 = 1, Y_2 = 1)$, $(Y_1 = 1, Y_2 = 0)$, and $(Y_1 = 0)$. Thus the log-likelihood function for a random sample of n observations is

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{n} Y_{i1} Y_{i2} \ln \pi_{i11}(\boldsymbol{\theta}) + Y_{i1}(1 - Y_{i2}) \ln \pi_{i10}(\boldsymbol{\theta}) + (1 - Y_{i1}) \ln \pi_{i0}(\boldsymbol{\theta}_1) \tag{10}$$

where $\pi_0 = \pi_{00} + \pi_{01}$. An ML estimator $\widehat{\boldsymbol{\theta}}$ maximizes the log-likelihood function (10) over the parameter space $\boldsymbol{\Theta} = \Re^{k_1+k_2+2} \times (-1, 1)$.

3 SNP estimation

The basic idea of SNP estimation is to approximate the unknown densities of the latent regression errors by Hermite polynomial expansions and use the approximations to derive a pseudo-ML estimator for the model parameters. Once we relax the Gaussian distributional assumption, a semiparametric specification of the likelihood function is needed. For the three binary-choice models considered in this article, semiparametric specifications of the log-likelihood functions have the same form as (3), (6), and (10), respectively, with the probability functions replaced by[4]

$$\begin{aligned}
\pi_{11}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) &= 1 - F_1(-\mu_1) - F_2(-\mu_2) + F(-\mu_1, -\mu_2) \\
\pi_{10}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) &= F_2(-\mu_2) - F(-\mu_1, -\mu_2) \\
\pi_{01}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) &= F_2(-\mu_1) - F(-\mu_1, -\mu_2) \\
\pi_{00}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) &= F(-\mu_1, -\mu_2)
\end{aligned}$$

where F_j is the unknown marginal distribution function of the latent regression error U_j, $j = 1, 2$, and F is the unknown joint distribution function of (U_1, U_2).[5]

---

[4] The marginal probability function is defined by $\pi_1(\boldsymbol{\theta}_1) = 1 - F_1(-\mu_1)$.

[5] The probability functions underlying the probit specifications can be easily obtained from these general expressions by exploiting the symmetry of the Gaussian distribution.

Following Gallant and Nychka (1987), we approximate the unknown joint density, f, of the latent regression errors by a Hermite polynomial expansion of the form

$$f^*(u_1, u_2) = \frac{1}{\psi_R}\, \tau_R(u_1, u_2)^2\, \phi(u_1)\, \phi(u_2) \tag{11}$$

where $\phi(\cdot)$ is the standardized Gaussian density, $\tau_R(u_1, u_2) = \sum_{h=0}^{R_1} \sum_{k=0}^{R_2} \tau_{hk} u_1^h u_2^k$ is a polynomial in u_1 and u_2 of order $R = (R_1, R_2)$, and

$$\psi_R = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tau_R(u_1, u_2)^2 \phi(u_1) \phi(u_2)\, du_1\, du_2$$

is a normalization factor that ensures f^* is a proper density. As shown by Gallant and Nychka (1987), the class of densities that can be approximated by this polynomial expansion includes densities with arbitrary skewness and kurtosis but excludes violently oscillatory densities or densities with tails that are too fat or too thin.[6] Our approximation to the joint density function of U_1 and U_2 differs from that originally proposed by Gallant and Nychka (1987) only because the order of the polynomial $\tau_R(u_1, u_2)$ is not restricted to be the same for U_1 and U_2. Although asymptotic properties of the SNP estimator require that both R_1 and R_2 increase with the sample size, there is no reason to impose that $R_1 = R_2$ in finite samples. For instance, different orders of R_1 and R_2 can help account for either departures from Gaussianity along one single component, or different sample sizes on Y_1 and Y_2 arising in the case of sample selection.

Since the polynomial expansion in (11) is invariant to multiplication of the vector of parameters $\tau = (\tau_{00}, \tau_{01}, \ldots, \tau_{R_1 R_2})$ by a scalar, some normalization is needed. Setting $\tau_{00} = 1$, expanding the square of the polynomial in (11) and rearranging terms gives

$$f^*(u_1, u_2) = \frac{1}{\psi_R} \left(\sum_{h=0}^{2R_1} \sum_{k=0}^{2R_2} \tau^*_{hk} u_1^h u_2^k \right) \phi(u_1)\, \phi(u_2)$$

with $\tau^*_{hk} = \sum_{r=a_h}^{b_h} \sum_{s=a_k}^{b_k} \tau_{rs} \tau_{h-r,k-s}$, where $a_h = \max(0, h - R_1)$, $a_k = \max(0, k - R_2)$, $b_h = \min(h, R_1)$, and $b_k = \min(k, R_2)$. Integrating $f^*(u_1, u_2)$ alternatively with respect to u_2 and u_1 gives the following approximations to the marginal densities f_1 and f_2

$$\begin{aligned} f_1^*(u_1) &= \int_{-\infty}^{\infty} f^*(u_1, u_2)\, du_2 \\ &= \frac{1}{\psi_R} \left(\sum_{h=0}^{2R_1} \sum_{k=0}^{2R_2} \tau^*_{hk}\, m_k\, u_1^h \right) \phi(u_1) = \frac{1}{\psi_R} \left(\sum_{h=0}^{2R_1} \gamma_{1h}\, u_1^h \right) \phi(u_1) \end{aligned} \tag{12}$$

$$\begin{aligned} f_2^*(u_2) &= \int_{-\infty}^{\infty} f^*(u_1, u_2)\, du_1 \\ &= \frac{1}{\psi_R} \left(\sum_{h=0}^{2R_1} \sum_{k=0}^{2R_2} \tau^*_{hk}\, m_h\, u_2^k \right) \phi(u_2) = \frac{1}{\psi_R} \left(\sum_{k=0}^{2R_2} \gamma_{2k}\, u_2^k \right) \phi(u_2) \end{aligned} \tag{13}$$

---

[6] Further details on the smoothness conditions defining this class of densities can be found in Gallant and Nychka (1987, 369).

where m_h and m_k are the hth and the kth central moments of the standardized Gaussian distribution, $\gamma_{1h} = \sum_{k=0}^{2R_2} \tau^*_{hk}\, m_k$, $\gamma_{2k} = \sum_{h=0}^{2R_1} \tau^*_{hk}\, m_h$, and $\psi_R = \sum_{h=0}^{2R_1} \gamma_{1h}\, m_h = \sum_{k=0}^{2R_2} \gamma_{2k}\, m_k$. As for the bivariate density function, γ_{10} and γ_{20} are normalized to one by imposing that $\tau_{h0} = \tau_{0k} = 0$ for all $h = 1, \ldots, R_1$ and $k = 1, \ldots, R_2$. Thus, if $\gamma_{1h} = 0$ for all $h \geq 1$, then $\psi_R = 1$ and so the approximation f_1^* coincides with the standard normal density. Similarly, the approximation f_2^* coincides with the standard normal density when $\gamma_{2k} = 0$ for all $k \geq 1$.

Adopting the SNP approximation to the density of the latent regression errors does not guarantee that they have zero mean and unit variance. The zero-mean condition implies that some location restriction needs to be imposed on either the distributions of the error terms, or the systematic part of the model. For the univariate model, Gabler, Laisney, and Lechner (1993) impose restrictions on the SNP parameters to guarantee that the error term has zero mean. For the bivariate model, this approach is quite complex. Therefore, we follow the alternative approach of Melenberg and van Soest (1996) and set the two intercept coefficients α_1 and α_2 to their parametric estimates. The parametric probit and the SNP estimates are not directly comparable because the SNP approximation does not have unit variance. However, as shown in section 6, one can compare the ratio of the estimated coefficients.

After accounting for the above restrictions, the total number of estimated parameters is $(k_1 + R_1)$ in the univariate SNP model and $(k_1 + k_2 + R_1 R_2)$ in the bivariate SNP model. Clearly, such models are not identified if the number of independent probabilities is lower than the number of free parameters to be estimated.[7]

Subject to these identifiability restrictions, integrating the joint density (11) gives the following approximation to the joint distribution function F

$$\begin{aligned} F^*(u_1, u_2) = {} & \Phi(u_1)\Phi(u_2) + \frac{1}{\psi_R}\, A_1^*(u_1, u_2)\phi(u_1)\phi(u_2) \\ & - \frac{1}{\psi_R}\, A_2^*(u_2)\Phi(u_1)\phi(u_2) - \frac{1}{\psi_R}\, A_3^*(u_1)\phi(u_1)\Phi(u_2) \end{aligned}$$

Similarly, integrating the marginal densities (12) and (13) gives the following approximations to the marginal distribution functions F_1 and F_2,

$$\begin{aligned} F_1^*(u_1) &= \Phi(u_1) - \frac{1}{\psi_R}\, A_3^*(u_1)\phi(u_1) \\ F_2^*(u_2) &= \Phi(u_2) - \frac{1}{\psi_R}\, A_2^*(u_2)\phi(u_2) \end{aligned}$$

---

[7] For instance, a univariate SNP model with a single categorical variable X is identified only if X can take at least $(1 + R_1)$ different values. A bivariate SNP model with X is identified only if X can take at least $(2 + R_1 R_2)/3$ different values. A bivariate SNP model with sample selection, in which X_1 and X_2 are two distinct categorical variables with p_1 and p_2 different values, is identified only if $(2 + R_1 R_2) \leq 2\, p_1\, p_2$.

where

$$A_1^*(u_1,u_2)=\sum_{h=0}^{2R_1}\sum_{k=0}^{2R_2}\tau_{hk}^* A_h(u_1)A_k(u_2)$$

$$A_2^*(u_2)=\sum_{h=0}^{2R_1}\sum_{k=0}^{2R_2}\tau_{hk}^* m_h A_k(u_2)$$

$$A_3^*(u_1)=\sum_{h=0}^{2R_1}\sum_{k=0}^{2R_2}\tau_{hk}^* m_k A_h(u_1)$$

with $A_0(u_j)=0$, $A_1(u_j)=1$, and $A_r(u_j)=(r-1)A_{r-2}(u_j)+u_j^{r-1}$, $j=1,2$. These approximations imply that the univariate probit model is always nested in the univariate SNP model, while the bivariate probit model is nested in the corresponding SNP model only if the correlation coefficient ρ is equal to zero. This result is due to two points. First, the leading terms in the SNP approximations to the marginal distribution functions F_1 and F_2 are Gaussian distribution functions and the remaining terms are products of Gaussian densities and polynomials of orders $(2R_1-1)$ and $(2R_2-1)$, respectively. Second, the leading term in the approximation to the joint distribution function F is the product of two Gaussian distribution functions and the remaining terms are complicated functions of u_1 and u_2.

SNP estimators can be obtained by maximizing the pseudo–log-likelihood functions (3), (6), and (10), respectively, in which the unknown distribution functions F, F_1, and F_2 are replaced by their approximations F^*, F_1^*, and F_2^*. As shown by Gallant and Nychka (1987), the resulting pseudo-ML estimators are $\sqrt{n}$ consistent provided that both R_1 and R_2 increase with the sample size. Although Gallant and Nychka (1987) provide consistency results for the SNP estimators, they do not provide distributional theory. However, when R_1 and R_2 are treated as known, inference can be conducted as though the model was estimated parametrically. The underlying assumption is that, for fixed values of R_1 and R_2, the true joint density function f belongs to the class of densities that can be approximated by the Hermite polynomial expansion in (11). Thus the SNP model can be considered as a flexible parametric specification for fixed values of R_1 and R_2, with the choice of R_1 and R_2 as part of the model-selection procedure. In practice, for a given sample size, the values of R_1 and R_2 may be selected either through a sequence of likelihood-ratio tests or by model selection criteria such as the Akaike information criterion or the Bayesian information criterion.

4 SML estimation

The basic idea of the SML estimation procedure is that of maximizing a pseudo–log-likelihood function in which the unknown probability functions are locally approximated by nonparametric kernel estimators.

Consider first SML estimation of a univariate binary-choice model. Before describing the estimation procedure in detail, we discuss nonparametric identification of the vector of parameters $\boldsymbol{\theta}=(\alpha,\boldsymbol{\beta})$. As for the SNP estimation procedure, the intercept coefficient

α can be absorbed into the unknown distribution function of the error term and is not separately identified. Furthermore, the slope coefficients $\boldsymbol{\beta}$ can only be identified up to a scale parameter. In this case, however, the scale normalization must be based on a continuous variable with a nonzero coefficient and it must be directly imposed on the estimation process.[8] Per Pagan and Ullah (1999), these location-scale normalizations can be obtained by imposing the linear index restriction

$$\pi(\boldsymbol{\theta}) = \Pr(Y = 1 \mid \boldsymbol{X}; \boldsymbol{\theta}) = \Pr\{Y = 1 \mid \upsilon(\boldsymbol{X}; \boldsymbol{\delta})\} = \pi(\boldsymbol{\delta})$$

where $\upsilon(\boldsymbol{X}; \boldsymbol{\delta}) = X_1 + \boldsymbol{\delta}^\top \boldsymbol{X}_2$, X_1 is a continuous variable with a nonzero coefficient, $\boldsymbol{X}_2$ are the other covariates, and $\boldsymbol{\delta} = (\delta_2, \dots, \delta_k)$ is the vector of identifiable parameters with $\delta_j = \beta_j/\beta_1$. The index restriction is also useful to reduce the dimension of the covariate space thereby avoiding the curse of dimensionality problem.

Under the index restriction, one can use Bayes Theorem to write

$$\pi(\boldsymbol{\delta}) = \frac{P\, g\{\upsilon(\boldsymbol{X}; \boldsymbol{\delta}) \mid Y = 1\}}{P\, g\{\upsilon(\boldsymbol{X}; \boldsymbol{\delta}) \mid Y = 1\} + (1 - P)\, g\{\upsilon(\boldsymbol{X}; \boldsymbol{\delta}) \mid Y = 0\}} \tag{14}$$

where $P = \Pr\{Y = 1\}$ is the unconditional probability of observing a positive outcome and $g(\cdot)$ is the conditional density of $\upsilon(\boldsymbol{X}; \boldsymbol{\delta})$ given Y. As in Klein and Spady (1993), a nonparametric estimator of $g_{1\upsilon}\{\upsilon(\boldsymbol{X}; \boldsymbol{\delta})\} = P\, g\{\upsilon(\boldsymbol{X}; \boldsymbol{\delta}) \mid Y = 1\}$ in the numerator of (14) is given by

$$\widehat{g}_{1\upsilon}(\upsilon_i; h_n) = \{(n-1)h_n\}^{-1} \sum_{j \neq i}^{n} y_j \mathcal{K}\left(\frac{\upsilon_i - \upsilon_j}{h_n}\right)$$

where $\upsilon_i = \upsilon(\boldsymbol{X}_i; \boldsymbol{\delta})$, $\mathcal{K}(\cdot)$ is a kernel function, and h_n is a bandwidth parameter which satisfies the restriction $n^{-1/6} < h_n < n^{-1/8}$. A nonparametric estimator of $g_{0\upsilon}\{\upsilon(\boldsymbol{X}; \boldsymbol{\delta})\} = (1 - P)\, g\{\upsilon(\boldsymbol{X}; \boldsymbol{\delta}) \mid Y = 0\}$ in the denominator of (14) can be defined in a similar way by replacing y_j with $(1 - y_j)$. To reduce the bias generated by kernel density estimation, Klein and Spady (1993) suggest using either bias-reducing kernels, or adaptive kernels with a variable and data-dependent bandwidth. For simplicity, we use a Gaussian kernel with a fixed-bandwidth parameter.[9]

An SML estimator $\widehat{\boldsymbol{\delta}}$ maximizes the pseudo–log-likelihood functions (3) where the unknown probability function $\pi(\boldsymbol{\delta})$ is replaced by a nonparametric estimate of the form[10]

$$\widehat{\pi}(\boldsymbol{\delta}) = \frac{\widehat{g}_{1\upsilon}(\upsilon; h_n)}{\widehat{g}_{1\upsilon}(\upsilon; h_n) + \widehat{g}_{0\upsilon}(\upsilon; h_n)} \tag{15}$$

Klein and Spady (1993) show that, under mild regularity conditions, the resulting SML estimator is $\sqrt{n}$ consistent, asymptotically normal, and achieves the semiparametric efficiency bound of Chamberlain (1986) and Cosslett (1987). In establishing the asymptotic

---

[8] See Klein and Spady (1993, assumption C.3b).

[9] Results of preliminary Monte Carlo simulations suggest that using a Gaussian kernel with a fixed bandwidth does not affect the small-sample performance of the SML estimator, by much.

[10] If the densities $g_{1\upsilon}\{\upsilon(\boldsymbol{X}; \boldsymbol{\delta})\}$ and $g_{0\upsilon}\{\upsilon(\boldsymbol{X}; \boldsymbol{\delta})\}$ are estimated by kernel methods with the same bandwidth parameter, then the estimator in (15) corresponds to a Nadaraya–Watson kernel estimator for the expected value of Y conditional on the index $\upsilon(\boldsymbol{X}; \boldsymbol{\delta})$.

properties of this estimator, a trimming function is used to downweight observations for which the corresponding densities are small. Because the Klein and Spady (1993) simulation results suggest that trimming is not important in practical applications, we ignore trimming.

When generalizing the SML estimator to bivariate binary-choice models with sample selection, the relevant issue is nonparametric estimation of the conditional probability $\pi_{1|1}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \Pr(Y_2 = 1 \mid Y_1 = 1, \boldsymbol{X}_1, \boldsymbol{X}_2)$. As for the univariate model, we assume that the model satisfies the double-index restriction

$$\pi_{1|1}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \Pr\{Y_2 = 1 \mid Y_1 = 1, \upsilon(\boldsymbol{X}_1; \boldsymbol{\delta}_1), \upsilon(\boldsymbol{X}_2; \boldsymbol{\delta}_2)\} = \pi_{1|1}(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)$$

where the $\upsilon(\boldsymbol{X}_j; \boldsymbol{\delta}_j) = X_{j1} + \boldsymbol{\delta}_j^\top \boldsymbol{X}_{j2}$, $j = 1, 2$, are linear indexes, the X_{j1} are continuous variables with nonzero coefficients, the $\boldsymbol{X}_{j2}$ are the remaining covariates, and the $\boldsymbol{\delta}_j = (\delta_{j2}, \ldots, \delta_{jk_j})$ are vectors of identifiable parameters with $\delta_{jh} = \beta_{jh}/\beta_{j1}$. As argued by Ichimura and Lee (1991), nonparametric identification of a double-index model requires the existence a distinct continuous variable for each index. Thus, unlike the parametric or the SNP specification of the model, exclusion restrictions should now include some continuous variables.

Subject to these identifiability restrictions, Bayes Theorem implies that

$$\pi_{1|1}(\boldsymbol{\delta}) = \frac{P_{1|1}\, g(\boldsymbol{\upsilon} \mid Y_1 = 1, Y_2 = 1)}{P_{1|1}\, g(\boldsymbol{\upsilon} \mid Y_1 = 1, Y_2 = 1) + (1 - P_{1|1})\, g(\boldsymbol{\upsilon} \mid Y_1 = 1, Y_2 = 0)} \tag{16}$$

where $\boldsymbol{\delta} = (\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)$, $P_{1|1} = \Pr(Y_2 = 1 \mid Y_1 = 1)$, $\boldsymbol{\upsilon} = \{\upsilon(\boldsymbol{X}_1; \boldsymbol{\delta}_1), \upsilon(\boldsymbol{X}_2; \boldsymbol{\delta}_2)\}$, and $g(\cdot)$ is the conditional density of $\boldsymbol{\upsilon}$ given Y_1 and Y_2. A nonparametric estimator of the density $g_{1\boldsymbol{\upsilon}|1}(\boldsymbol{\upsilon}) = P_{1|1}\, g(\boldsymbol{\upsilon} \mid Y_1 = 1, Y_2 = 1)$ in the numerator of (16) is given by

$$\widehat{g}_{1\boldsymbol{\upsilon}|1}(\boldsymbol{\upsilon}_i; h_{n_1}) = \left\{(n_1 - 1)h_{n_1}^2\right\}^{-1} \sum_{j \neq i}^{n_1} y_{1j}\, y_{2j}\, \mathcal{K}_2\left(\frac{\boldsymbol{\upsilon}_i - \boldsymbol{\upsilon}_j}{h_{n_1}}\right)$$

where n_1 is the subsample of observations for which $Y_1 = 1$, and $\mathcal{K}_2(\cdot)$ is the product of two univariate Gaussian kernels with the same bandwidth h_{n_1}. A nonparametric estimator of the density $g_{0\boldsymbol{\upsilon}|1}(\boldsymbol{\upsilon}) = (1 - P_{1|1})\, g(\boldsymbol{\upsilon} \mid Y_1 = 1, Y_2 = 0)$ in the denominator of (16) can be defined in a similar way by replacing y_{2j} with $(1 - y_{2j})$. As before, these nonparametric estimators differ from those adopted by Lee (1995) only because we use Gaussian kernels, instead of bias-reducing kernels. Thus the conditional probability $\pi_{1|1}(\boldsymbol{\delta})$ is estimated by

$$\widehat{\pi}_{1|1}(\boldsymbol{\delta}) = \frac{\widehat{g}_{1\boldsymbol{\upsilon}|1}(\boldsymbol{\upsilon}; h_{n_1})}{\widehat{g}_{1\boldsymbol{\upsilon}|1}(\boldsymbol{\upsilon}; h_{n_1}) + \widehat{g}_{0\boldsymbol{\upsilon}|1}(\boldsymbol{\upsilon}; h_{n_1})}$$

and an SML estimator $\widehat{\boldsymbol{\delta}}$ is obtained by maximizing the log-likelihood function (10), where the unknown probability functions are replaced by

$$\begin{aligned} \widehat{\pi}_0(\boldsymbol{\delta}_1) &= 1 - \widehat{\pi}_1(\boldsymbol{\delta}_1) \\ \widehat{\pi}_{11}(\boldsymbol{\delta}) &= \widehat{\pi}_1(\boldsymbol{\delta}_1)\, \widehat{\pi}_{1|1}(\boldsymbol{\delta}) \\ \widehat{\pi}_{10}(\boldsymbol{\delta}) &= \widehat{\pi}_1(\boldsymbol{\delta}_1)\, \{1 - \widehat{\pi}_{1|1}(\boldsymbol{\delta})\} \end{aligned}$$

Lee (1995) shows that, under mild regularity conditions, the resulting SML estimator is $\sqrt{n}$ consistent and asymptotically normal. Furthermore, its asymptotic variance is very close to the efficiency bound of semiparametric estimators for this type of model.

5 Stata commands

5.1 Syntax of SNP commands

The new Stata commands `snp`, `snp2`, and `snp2s` estimate the parameters of the SNP binary-choice models considered in this article. In particular, `snp` fits a univariate binary-choice model, `snp2` fits a bivariate binary-choice model, while `snp2s` fits a bivariate binary-choice model with sample selection. The general syntax of these commands is as follows:

`snp` *depvar varlist* [*if*] [*in*] [*weight*] [, `noconstant` `offset(`*varname*`)`
`order(`*#*`)` `robust` `from(`*matname*`)` `dplot(`*filename*`)` `level(`*#*`)`
maximize_options]

`snp2` *equation1 equation2* [*if*] [*in*] [*weight*] [, `order1(`*#*`)` `order2(`*#*`)` `robust`
`from(`*matname*`)` `dplot(`*filename*`)` `level(`*#*`)` *maximize_options*]

`snp2s` *depvar varlist* [*if*] [*in*] [*weight*],
`select(`*depvar_s* `=` *varlist_s* [, `offset(`*varname*`)` `noconstant`]`)`
[`order1(`*#*`)` `order2(`*#*`)` `robust` `from(`*matname*`)` `dplot(`*filename*`)` `level(`*#*`)`
maximize_options]

where each equation is specified as

`(` [*eqname*`:`] *depvar* [`=`] *varlist* [, `noconstant` `offset(`*varname*`)`] `)`

`snp`, `snp2`, and `snp2s` are implemented for Stata 9 by using `ml model lf`. These commands share the same features of all Stata estimation commands, including access to the estimation results and the options for the maximization process (see [R] **maximize**). `fweight`s, `pweight`s, and `iweight`s are allowed (see [U] **14.1.6 weight**). Most of the options are similar to those of other Stata estimation commands. A description of the options that are specific to our SNP commands is provided below.

Options of SNP commands

`order(#)` specifies the order R to be used in the univariate Hermite polynomial expansion. The default is `order(3)`.

`order1(#)` specifies the order R_1 to be used in the bivariate Hermite polynomial expansion. The default is `order1(3)`.

`order2(#)` specifies the order R_2 to be used in the bivariate Hermite polynomial expansion. The default is `order2(3)`.

`robust` specifies that the Huber/White/sandwich estimator of the covariance matrix is to be used in place of the traditional calculation (see [U] **23.11 Obtaining robust variance estimates**).[11]

`from(`*matname*`)` specifies the name of the matrix to be used as starting values. By default, starting values are the estimates of the corresponding probit specification, namely, the `probit` estimates for `snp`, the `biprobit` estimates for `snp2`, and the `heckprob` estimates for `snp2s`.

`dplot(`*filename*`)` plots the estimated marginal densities of the error terms. A Gaussian density with the same estimated mean and variance is added to each density plot. For the `snp` command, *filename* specifies the name of the density plot to be created. For `snp2` and `snp2s`, three new graphs are created. The first is a plot of the estimated marginal density of U_1 and is stored as *filename*`_1`. The second is a plot of the estimated marginal density of U_2 and is stored as *filename*`_2`. The third is a combination of the two density plots in a single graph and is stored as *filename*.

5.2 Syntax of SML estimators

The new Stata commands `sml` and `sml2s` estimate the parameters of the SML models discussed in this article. `sml` fits a univariate binary-choice model, and `sml2s` fits a bivariate binary-choice model with sample selection. The general syntax of these commands is as follows:

```
sml depvar varlist [if] [in] [weight] [, noconstant offset(varname)
   bwidth(#) from(matname) level(#) maximize_options]
```

```
sml2s depvar varlist [if] [in] [weight],
   select(depvar_s = varlist_s [, offset(varname) noconstant])
   [bwidth1(#) bwidth2(#) from(matname) level(#) maximize_options]
```

`sml` and `sml2s` are implemented for Stata 9 by using `ml model d2` and `ml model d0`, respectively. In this case, `ml model lf` cannot be used because SML estimators violate

---

[11] As pointed out by an anonymous referee, for a finite R, the SNP model can be misspecified and the `robust` option accounts for this misspecification in estimating the covariance matrix of the SNP estimator.

the linear-form restriction.[12] Unlike the SNP commands, `pweight` and `robust` are not allowed with `sml` and `sml2s` commands. Although this may be a drawback of our SML routines, it is important to mention that SML estimators impose weaker distributional assumptions than the SNP estimators and they are also robust to the presence of heteroskedasticity of a general but known form and heteroskedasticity of an unknown form if it depends on the underlying indexes (see Klein and Spady [1993]). A description of the options that are specific to our SML commands is provided below.

Options

`bwidth(#)` specifies the value of the bandwidth parameter h_n. The default is $h_n = n^{-1/6.5}$, where n is the overall sample size.

`bwidth1(#)` specifies the value of the bandwidth parameter h_n used for nonparametric estimation of the selection probability $\widehat{\pi}_1(\delta_1)$. The default is $h_n = n^{-1/6.5}$, where n is the overall sample size.

`bwidth2(#)` specifies the value of the bandwidth parameter h_{n_1} used for nonparametric estimation of the conditional probability $\widehat{\pi}_{1|1}(\delta_1, \delta_2)$. The default is $h_{n_1} = n_1^{-1/6.5}$, where n_1 is the number of selected observations.

`from(`*matname*`)` specifies the name of the matrix to be used as starting values. By default, starting values are the estimates of the corresponding probit specification, namely, the `probit` estimates for `sml` and the `heckprob` estimates for `sml2s`.

5.3 Further remarks

1. SNP and SML estimators typically require large samples. Furthermore, since the log-likelihood functions of these estimators are not globally concave, it is good practice to check for convergence to the global maximum rather than a local one by using the `from` option.

2. Asymptotic properties of the SNP estimators require that the degree R of the Hermite polynomial expansion increases with the sample size. In particular, `snp` generalizes the probit model only if $R \geq 3$ (see Gabler, Laisney, and Lechner [1993]). For `snp2` and `snp2s`, the error terms may have skewness and kurtosis different from those of a Gaussian distribution only if $R_1 \geq 2$ or $R_2 \geq 2$. In practice, the values of R, R_1, and R_2 may be selected either through a sequence of likelihood-ratio tests or by model-selection criteria such as the Akaike information criterion or the Bayesian information criterion (see the `lrtest` command).

3. SML estimation uses Gaussian kernels with a fixed bandwidth. Asymptotic properties of the SML estimators require the bandwidth parameters to satisfy the restrictions $n^{-1/6} < h_n < n^{-1/8}$ and $n_1^{-1/6} < h_{n_1} < n_1^{-1/8}$. In practice, one may

---

[12] An extensive discussion on the alternative Stata ML models can be found in Gould, Pitblado, and Sribney (2006).

either experiment with alternative values of h_n and h_{n_1} in the above range or use a more sophisticated method like generalized cross validation (see Gerfin [1996]).

4. The proposed estimators are more computationally demanding than the corresponding parametric estimators because of both the greater complexity of the likelihood functions and the fact that they are written as ado-files. The number of iterations required by SNP estimators typically increases with the order of the Hermite polynomial expansion. Convergence of SML estimators usually requires a lower number of iterations, but they are more computationally demanding since kernel regression is conducted at each step of the maximization process. For both types of estimators, estimation time further depends on the number of observations and the number of covariates.

6 Examples

This section provides illustrations of the SNP and SML commands using simulated data, which allows us to have a benchmark for the estimation results. The Stata code for our data-generating process is

```
. * Data generating process
. clear all
. set seed 1234
. matrix define sigma=(1,.5\.5,1)
. quietly drawnorm u1 u2, n(2000) corr(sigma) double
. generate double x1=(uniform()*2-1)*sqrt(3)
. generate double x2=(uniform()*2-1)*sqrt(3)
. generate double x3=invchi2(1,uniform())
. generate x4=(uniform()>.5)
. generate y1=(x1-x3+2*x4+u1>0)
. generate y2=(x2+.5*x3-1.5*x4+u2>0)
```

Error terms are generated from a bivariate Gaussian distribution with zero means, unit variances, and a correlation coefficient equal to 0.5. The set of covariates includes four variables: X_1 and X_2 are independently drawn from a standardized uniform distribution on $(-1,1)$, X_3 is drawn from a chi-squared distribution with 1 degree of freedom, and X_4 is drawn from a Bernoulli distribution with a probability of success equal to 0.5. To guarantee identifiability of the model parameters, our data-generating process imposes one exclusion restriction in each equation, namely, X_1 only enters the equation of Y_1, while X_2 only enters the equation of Y_2.

The probit estimates of the first equation are given by

```
. probit y1 x1 x3 x4, nolog
Probit regression                                 Number of obs   =       2000
                                                  LR chi2(3)      =    1492.07
                                                  Prob > chi2     =     0.0000
Log likelihood = -633.69196                       Pseudo R2       =     0.5407

------------------------------------------------------------------------------
          y1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   1.093143   .0539591    20.26   0.000     .9873851    1.198901
          x3 |  -1.153823   .0605857   -19.04   0.000    -1.272569   -1.035077
          x4 |   2.037333   .0988448    20.61   0.000     1.843601    2.231065
       _cons |   .1540149   .0596325     2.58   0.010     .0371375    .2708924
------------------------------------------------------------------------------
Note: 29 failures and 0 successes completely determined.
. nlcom (b3_b1: _b[x3] / _b[x1]) (b4_b1: _b[x4] / _b[x1])
       b3_b1:  _b[x3] / _b[x1]
       b4_b1:  _b[x4] / _b[x1]

------------------------------------------------------------------------------
          y1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       b3_b1 |   -1.05551   .0527265   -20.02   0.000    -1.158852   -.9521678
       b4_b1 |   1.863739   .0887034    21.01   0.000     1.689883    2.037594
------------------------------------------------------------------------------
```

Because of the different scale normalization, estimated coefficients of the probit model are not directly comparable with those of the SNP and SML models. Here we compare the ratio of the estimated coefficients by using the `nlcom` command.

The SNP estimates of the same model, with degree of the univariate Hermite polynomial expansion $R = 4$, are given by

(*Continued on next page*)

```
. snp y1 x1 x3 x4, nolog order(4)
Order of SNP polynomial - R=4
SNP Estimation of Binary-Choice Model              Number of obs   =       2000
                                                   Wald chi2(3)    =      36.86
Log likelihood =  -632.0571                        Prob > chi2     =     0.0000

------------------------------------------------------------------------------
          y1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y1           |
          x1 |   1.600913   .2670701     5.99   0.000     1.077465    2.124361
          x3 |  -1.688176   .2840946    -5.94   0.000    -2.244991    -1.13136
          x4 |   2.990631   .4982394     6.00   0.000       2.0141    3.967163
-------------+----------------------------------------------------------------
       _cons |   .1540149      Fixed
-------------+----------------------------------------------------------------
SNP coefs: 1 |  -.1277219   .0937248    -1.36   0.173    -.3114192    .0559754
           2 |   .1213818   .0883377     1.37   0.169    -.0517569    .2945206
           3 |   .0391799   .0232224     1.69   0.092    -.0063352    .0846951
           4 |   .0170504   .0201136     0.85   0.397    -.0223715    .0564722
------------------------------------------------------------------------------
Likelihood ratio test of Probit model against SNP model:
Chi2(2) statistic =      3.269725      (p-value =  .1949791)
------------------------------------------------------------------------------
Estimated moments of error distribution:
Variance =             2.161259      Standard Deviation =      1.470122
3rd moment =           .7185723      Skewness =                 .226157
4th moment =           12.70987      Kurtosis =                2.720993
------------------------------------------------------------------------------

. nlcom (b3_b1: _b[x3] / _b[x1]) (b4_b1: _b[x4] / _b[x1]), post
       b3_b1:  _b[x3] / _b[x1]
       b4_b1:  _b[x4] / _b[x1]

------------------------------------------------------------------------------
          y1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       b3_b1 |  -1.054508   .0532962   -19.79   0.000    -1.158966   -.9500493
       b4_b1 |   1.868078   .0853722    21.88   0.000     1.700752    2.035405
------------------------------------------------------------------------------

. matrix b0=e(b)
```

Estimated coefficients and standard errors are very close to the corresponding probit estimates. SNP coefficients are not significantly different from zero, and a likelihood-ratio test of the probit model against the SNP model does not reject the Gaussianity assumption. Estimates of skewness and kurtosis are also close to the Gaussian values of 0 and 3, respectively. In general, however, a very large-sample size, of say 10,000 observations, is necessary to obtain accurate estimates of these higher order moments.[13] The `post` option in `nlcom` causes this command to behave like a Stata estimation command. Below we use these normalized estimates of the `snp` command as starting values for the `sml` command:

---

[13] Simulation also indicates that while the skewness and kurtosis converge to those of the true error distribution, the reported variance differs by a scale factor from the variance of the true error distribution.

```
. sml y1 x3 x4, offset(x1) from(b0, copy) nolog
SML Estimator - Klein & Spady (1993)              Number of obs   =       2000
                                                  Wald chi2(2)    =     625.91
Log likelihood = -637.72168                       Prob > chi2     =     0.0000

          y1       Coef.   Std. Err.      z    P>|z|      [95% Conf. Interval]

          x3   -1.065468   .0530123   -20.10   0.000     -1.169371   -.9615662
          x4    1.882163   .0899912    20.91   0.000      1.705784    2.058543
          x1    (offset)
```

Identifiability of the model parameters is obtained by constraining the coefficient of the continuous variable X_1 to one through the `offset` option (the `nonconstant` option is always specified by default). In this case, the bandwidth parameter is set to its default value, namely, $h_n = n^{-1/6.5}$. Overall, estimated coefficients and standard errors are again very close to their `probit` estimates.

In the next example, we provide parametric estimates of the bivariate binary-choice model for Y_1 and Y_2

```
. biprobit (y1=x1 x3 x4) (y2=x2 x3 x4), nolog
Seemingly unrelated bivariate probit              Number of obs   =       2000
                                                  Wald chi2(6)    =    1242.93
Log likelihood = -1384.2384                       Prob > chi2     =     0.0000

                   Coef.   Std. Err.      z    P>|z|      [95% Conf. Interval]

y1
          x1    1.101044   .0526561    20.91   0.000        .99784    1.204248
          x3   -1.151508   .0589227   -19.54   0.000     -1.266994   -1.036021
          x4    2.056867   .0982408    20.94   0.000      1.864318    2.249415
       _cons    .1432707   .0589297     2.43   0.015      .0277707    .2587708

y2
          x2    1.045511   .0457917    22.83   0.000      .9557608    1.135261
          x3    .4806406   .0356907    13.47   0.000      .4106882     .550593
          x4   -1.551646   .0813526   -19.07   0.000     -1.711094   -1.392198
       _cons    .0184006   .0544944     0.34   0.736     -.0884065    .1252077

     /athrho    .6076755   .0732266     8.30   0.000      .4641541     .751197

         rho    .5424888   .0516764                      .4334638    .6358625

Likelihood-ratio test of rho=0:     chi2(1) =  79.0783    Prob > chi2 = 0.0000
```

(*Continued on next page*)

```
. nlcom (b3_b1: [y1]_b[x3] / [y1]_b[x1]) (b4_b1: [y1]_b[x4] / [y1]_b[x1])
>       (b3_b2: [y2]_b[x3] / [y2]_b[x2]) (b4_b2: [y2]_b[x4] / [y2]_b[x2])
```

(*output omitted*)

```
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       b3_b1 |  -1.045833   .0506567   -20.65   0.000    -1.145118   -.9465474
       b4_b1 |   1.868106   .0861391    21.69   0.000     1.699276    2.036935
       b3_b2 |   .4597184   .0332726    13.82   0.000     .3945054    .5249315
       b4_b2 |  -1.484103   .0749728   -19.80   0.000    -1.631047   -1.337159
------------------------------------------------------------------------------
```

The SNP estimates with $R_1 = R_2 = 3$ are given by

```
. snp2 (y1=x1 x3 x4, noconstant) (y2=x2 x3 x4, noconstant), dplot(gr) nolog
Order of SNP polynomial - (R1,R2)=(3,3)
SNP Estimation of Bivariate Model                 Number of obs   =       2000
                                                  Wald chi2(3)    =     155.43
Log likelihood = -1382.3065                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y1           |
          x1 |   1.554566   .1293488    12.02   0.000     1.301047    1.808085
          x3 |  -1.617294   .1431821   -11.30   0.000    -1.897926   -1.336662
          x4 |   2.971796   .2476522    12.00   0.000     2.486407    3.457186
-------------+----------------------------------------------------------------
y2           |
          x2 |   1.743243   .1550252    11.24   0.000     1.439399    2.047087
          x3 |   .7931422   .0845287     9.38   0.000      .627469    .9588155
          x4 |  -2.603726     .22436   -11.61   0.000    -3.043464   -2.163989
-------------+----------------------------------------------------------------
Intercepts:  |
      _cons1 |          0      Fixed
      _cons2 |          0      Fixed
-------------+----------------------------------------------------------------
SNP coefs:   |
       g_1_1 |   -.467446    .337281    -1.39   0.166    -1.128505    .1936125
       g_1_2 |  -.0437985   .0702888    -0.62   0.533    -.181562     .0939651
       g_1_3 |   .2417127   .0936064     2.58   0.010     .0582476    .4251778
       g_2_1 |   .0275117   .0667043     0.41   0.680    -.1032263    .1582497
       g_2_2 |   .1097933   .0351317     3.13   0.002     .0409364    .1786502
       g_2_3 |  -.0127886   .0201542    -0.63   0.526    -.0522901     .026713
       g_3_1 |   .1368238    .082591     1.66   0.098    -.0250516    .2986991
       g_3_2 |   .0312873   .0186252     1.68   0.093    -.0052175     .067792
       g_3_3 |  -.0309619   .0215084    -1.44   0.150    -.0731175    .0111938
------------------------------------------------------------------------------
Estimated moments of errors distribution
    Main equation                           Selection equation
    Standard Deviation = 1.652599           Standard Deviation = 1.426339
    Variance =           2.731084           Variance =           2.034443
    Skewness =          -.0735322           Skewness =           .1274959
    Kurtosis =            2.56411           Kurtosis =           2.554007
------------------------------------------------------------------------------
Estimated correlation coefficient
    rho =                .4974266
------------------------------------------------------------------------------
```

```
. nlcom (b3_b1 :[y1]_b[x3] / [y1]_b[x1])  (b4_b1 :[y1]_b[x4] / [y1]_b[x1])
>       (b3_b2 :[y2]_b[x3] / [y2]_b[x2])  (b4_b2 :[y2]_b[x4] / [y2]_b[x2])
  (output omitted)

                    Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]

       b3_b1    -1.040351   .0503783   -20.65   0.000     -1.13909   -.9416109
       b4_b1     1.911656   .0838935    22.79   0.000     1.747228    2.076084
       b3_b2     .4549809   .0305409    14.90   0.000     .3951219      .51484
       b4_b2    -1.493611   .0723319   -20.65   0.000    -1.635379   -1.351843
```

By specifying the `noconstant` options, the intercept coefficients are normalized to zero and starting values are set to the estimates of the bivariate probit model with no intercept. Once differences in the scale of the error terms are taken into account, the estimated coefficients of `biprobit` and `snp2` seem to be very close. As explained in section 3, the bivariate probit model is nested in the bivariate SNP model only if the correlation coefficient ρ is equal to zero. Accordingly, a likelihood-ratio test for the Gaussianity of the error terms cannot be used. Furthermore, it is important to notice that `snp2` and `snp2s` do not provide standard errors and confidence intervals for the estimated correlation coefficient. If this is a parameter of interest, inference can be carried out via the bootstrap, although this alternative can be computationally demanding. The estimated correlation coefficient is indeed provided as an estimation output in `e(rho)`. Figure 1 shows the plots of the two estimated marginal densities obtained by specifying the `dplot` option.

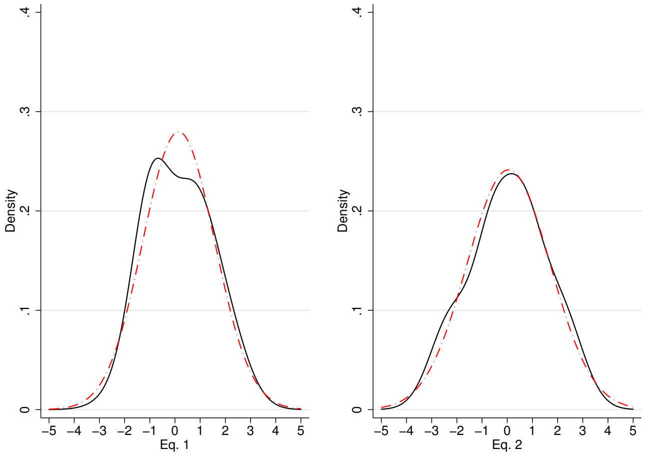

Figure 1. Semiparametric estimates of the error marginal densities

In the next example, we introduce selectivity in the equation for Y_2 and present parametric ML estimates of the resulting bivariate binary-choice model with sample selection.

```
. quietly replace y2=. if y1<1
. heckprob y2 x2 x3 x4, select(y1=x1 x3 x4) nolog
Probit model with sample selection              Number of obs      =      2000
                                                Censored obs       =       919
                                                Uncensored obs     =      1081
                                                Wald chi2(3)       =    260.84
Log likelihood = -1029.019                      Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y2           |
          x2 |   1.093629   .0681056    16.06   0.000     .9601448    1.227114
          x3 |   .4040404   .0848008     4.76   0.000     .2378339    .5702469
          x4 |  -1.609261   .1514939   -10.62   0.000    -1.906183   -1.312338
       _cons |    .046692   .1134511     0.41   0.681     -.175668     .269052
-------------+----------------------------------------------------------------
y1           |
          x1 |   1.089935   .0536583    20.31   0.000     .9847668    1.195103
          x3 |  -1.149565   .0601002   -19.13   0.000    -1.267359   -1.031771
          x4 |   2.044188   .0983742    20.78   0.000     1.851378    2.236998
       _cons |   .1475302   .0592512     2.49   0.013     .0313999    .2636604
-------------+----------------------------------------------------------------
     /athrho |   .6923304   .1696845     4.08   0.000     .3597548    1.024906
-------------+----------------------------------------------------------------
         rho |    .599477   .1087045                       .344998    .7718572
------------------------------------------------------------------------------
LR test of indep. eqns. (rho = 0):   chi2(1) =    19.62   Prob > chi2 = 0.0000
------------------------------------------------------------------------------
. nlcom (b3_b1 :[y1]_b[x3] / [y1]_b[x1])  (b4_b1 :[y1]_b[x4] / [y1]_b[x1])
>       (b3_b2 :[y2]_b[x3] / [y2]_b[x2])  (b4_b2 :[y2]_b[x4] / [y2]_b[x2])
  (output omitted)
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       b3_b1 |  -1.054709   .0524603   -20.10   0.000     -1.15753   -.9518892
       b4_b1 |   1.875514   .0885771    21.17   0.000     1.701906    2.049122
       b3_b2 |   .3694491   .0734796     5.03   0.000     .2254319    .5134664
       b4_b2 |  -1.471486   .1124007   -13.09   0.000    -1.691788   -1.251185
------------------------------------------------------------------------------
```

The `snp2s` estimates of the same model with $(R_1, R_2) = (4, 3)$ are given by

```
. snp2s y2 x2 x3 x4, select(y1=x1 x3 x4) order1(4) order2(3) nolog
Order of SNP polynomial - (R1,R2)=(4,3)
SNP Estimation of Sequential Bivariate Model      Number of obs   =       2000
                                                  Wald chi2(3)    =     110.18
Log likelihood = -1024.4739                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y2           |
          x2 |   1.837312   .1807611    10.16   0.000     1.483027    2.191597
          x3 |    .685451   .1467516     4.67   0.000     .3978231    .9730789
          x4 |  -2.681849   .2982197    -8.99   0.000    -3.266349   -2.097349
-------------+----------------------------------------------------------------
y1           |
          x1 |   1.582025   .1916435     8.26   0.000      1.20641    1.957639
          x3 |  -1.695678   .1991834    -8.51   0.000     -2.08607   -1.305286
          x4 |   3.011402   .3429733     8.78   0.000     2.339186    3.683617
-------------+----------------------------------------------------------------
Intercepts:  |
      _cons1 |   .1475302      Fixed
      _cons2 |    .046692      Fixed
-------------+----------------------------------------------------------------
SNP coefs:   |
       g_1_1 |  -.4411874   .4443872    -0.99   0.321     -1.31217    .4297954
       g_1_2 |  -.0775538   .1175507    -0.66   0.509     -.307949    .1528413
       g_1_3 |   .2447965    .099868     2.45   0.014     .0490589    .4405341
       g_2_1 |   .2157904   .2817236     0.77   0.444    -.3363777    .7679585
       g_2_2 |   .1268945   .0864948     1.47   0.142    -.0426323    .2964212
       g_2_3 |  -.0950307   .0726311    -1.31   0.191    -.2373851    .0473237
       g_3_1 |    .113475   .0886566     1.28   0.201    -.0602888    .2872388
       g_3_2 |   .0453493   .0369828     1.23   0.220    -.0271357    .1178343
       g_3_3 |  -.0294287   .0219723    -1.34   0.180    -.0724937    .0136362
       g_4_1 |  -.0449806   .0489556    -0.92   0.358    -.1409318    .0509705
       g_4_2 |   -.005844   .0185705    -0.31   0.753    -.0422415    .0305535
       g_4_3 |   .0187026   .0113096     1.65   0.098    -.0034638     .040869
------------------------------------------------------------------------------
Estimated moments of errors distribution
    Main equation                           Selection equation
    Standard Deviation = 1.723961           Standard Deviation = 1.473068
    Variance =           2.972042           Variance =            2.16993
    Skewness =          -.0676901           Skewness =           .1437971
    Kurtosis =           2.503351           Kurtosis =           2.862437
------------------------------------------------------------------------------
Estimated correlation coefficient
    rho =               .4984005
------------------------------------------------------------------------------

. nlcom (b3_b1 :[y1]_b[x3] / [y1]_b[x1])  (b4_b1 :[y1]_b[x4] / [y1]_b[x1])
>       (b3_b2 :[y2]_b[x3] / [y2]_b[x2])  (b4_b2 :[y2]_b[x4] / [y2]_b[x2])
```

(*output omitted*)

```
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       b3_b1 |   -1.07184   .0522701   -20.51   0.000    -1.174288   -.9693929
       b4_b1 |   1.903511   .0850225    22.39   0.000      1.73687    2.070152
       b3_b2 |   .3730727   .0669923     5.57   0.000     .2417701    .5043752
       b4_b2 |  -1.459659    .104647   -13.95   0.000    -1.664763   -1.254555
------------------------------------------------------------------------------
```

As a final example, we provide estimates obtained from the `sml2s` command by setting $h_n = n^{-1/6.5}$ and $h_{n_1} = n_1^{-1/6.02}$.

```
. quietly summarize y2
. local bw2=1/(r(N)^(1/6.02))
. sml2s y2 x3 x4, select(y1=x3 x4, offset(x1)) offset(x2) bwidth2(`bw2´) nolog
Two-stage SML estimator - Lee (1995)            Number of obs   =       2000
                                                Wald chi2(2)    =     154.17
Log likelihood =  -1044.401                     Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y2           |
          x3 |   .3691167   .0789709     4.67   0.000     .2143366    .5238969
          x4 |  -1.463612   .1178916   -12.41   0.000    -1.694675   -1.232549
          x2 |   (offset)
-------------+----------------------------------------------------------------
y1           |
          x3 |  -1.063704   .0488601   -21.77   0.000    -1.159468   -.9679396
          x4 |   1.889661   .0845938    22.34   0.000      1.72386    2.055462
          x1 |   (offset)
------------------------------------------------------------------------------
```

The first two lines provide a simple way to specify alternative values for the bandwidth parameters. Here we are implicitly assuming that the number of nonmissing observations on Y_1 is equal to the size of the estimation sample. If this is not the case, because of missing data on the covariates, the `summarize` command on the first line should be appropriately restricted to the relevant estimation sample.

7 Monte Carlo simulations

To investigate the finite sample properties of the SNP and SML estimators, we conducted a set of Monte Carlo simulations. The aim of this experiment is to investigate both the efficiency losses of the these estimators relative to the parametric probit ML estimator in the Gaussian case, and the effectiveness of the proposed estimators under different non-Gaussian distributional assumptions.

Overall, our Monte Carlo experiment consists of four simulation designs and three sample sizes (500, 1000, and 2000). In each design, simulated data were generated from the following bivariate latent regression model

$$
\begin{aligned}
Y_1^* &= \beta_{11} X_{11} + \beta_{12} X_{12} + U_1 \\
Y_2^* &= \beta_{21} X_{21} + \beta_{22} X_{22} + U_2
\end{aligned}
$$

where the true parameters are $\beta_{11} = \beta_{21} = \beta_{22} = 1$ and $\beta_{12} = -1$. The regressors X_{11} and X_{21} were independently drawn from a uniform distribution with support $(-1, 1)$, while X_{12} and X_{22} were independently drawn from a chi-squared distribution with 1 and 3 degrees of freedom, respectively. All the regressors were standardized to have zero means and unit variances. Simulation designs differ because of the distributional assumptions made on the latent regression errors U_1 and U_2 (see table 1).

Table 1. Theoretical moments by simulation design

	Design 1	Design 2	Design 3	Design 4
Skewness of U_1	0	0.66	0	0.68
Skewness of U_2	0	−0.80	0	−1.19
Kurtosis of U_1	3	3	2.60	4.01
Kurtosis of U_2	3	3	2.00	5.13
Correlation coefficient	−0.5	−0.5	−0.5	−0.5

In Design 1, the error terms were generated from a bivariate Gaussian distribution with zero means, unit variances, and correlation coefficient $\rho = -0.5$. In Designs 2–4, the error terms were generated from a mixture of two bivariate Gaussian distributions with equal covariance matrices,

$$f(U_1, U_2; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \pi f_1(U_1, U_2; \boldsymbol{m}_1, \boldsymbol{\Omega}) + (1-\pi) f_2(U_1, U_2; \boldsymbol{m}_2, \boldsymbol{\Omega})$$

where π is the mixing probability, and the $f_j(\cdot, \cdot; \boldsymbol{m}_j, \boldsymbol{\Omega})$, $j = 1, 2$, are bivariate Gaussian densities with mean $\boldsymbol{m}_j = (m_{j1}, m_{j2})$ and covariance matrix

$$\boldsymbol{\Omega} = \begin{bmatrix} \omega_{11}^2 & \omega_{12} \\ & \omega_{22}^2 \end{bmatrix}$$

The theoretical moments of this bivariate mixture are

$$\begin{aligned} E(U_j) &= \pi\, m_{1j} + (1-\pi) m_{2j} \\ E(U_j^2) &= \omega_{jj}^2 + \pi\, m_{1j}^2 + (1-\pi)\, m_{j2}^2 \\ E(U_j^3) &= \pi\,(3\,\omega_{jj}^2\, m_{1j} + m_{1j}^3) + (1-\pi)\,(3\,\omega_{jj}^2\, m_{2j} + m_{2j}^3) \\ E(U_j^4) &= 3\,\omega_{jj}^4 + \pi\,(6\,\omega_{jj}^2\, m_{1j}^2 + m_{1j}^4) + (1-\pi)\,(6\,\omega_{jj}^2\, m_{2j}^2 + m_{2j}^4) \\ E(U_1 U_2) &= \omega_{12} + \pi\, m_{11}\, m_{12} + (1-\pi)\, m_{21}\, m_{22} \end{aligned}$$

By varying the mixing probability π and the parameters of the two Gaussian components $f_1(U_1, U_2; \boldsymbol{m}_1, \boldsymbol{\Omega})$ and $f_2(U_1, U_2; \boldsymbol{m}_2, \boldsymbol{\Omega})$, one can then define a family of bivariate mixtures with given skewness, kurtosis, and correlation coefficient.[14] Table 1 gives the skewness and kurtosis used in each design. The latent regression errors were generated from an asymmetric and mesokurtic distribution in Design 2, a symmetric and platykurtic distribution in Design 3, and an asymmetric and leptokurtic distribution in Design 4.[15] Error terms were then standardized to have zero means, unit variances,

---

[14] Although bivariate mixture distributions allow us to control the level of skewness, kurtosis, and correlation coefficient in each design, it is difficult to assess whether or not these error structures are nested into the SNP model for a finite value of R. For this reason, our simulation design may be biased against the SNP estimator.

[15] To investigate the small-sample behavior of the three estimators under different levels of skewness and kurtosis, error terms were always generated with stronger departures from Gaussianity in the distribution of U_2.

and correlation coefficient $\rho = -0.5$ in each design. Stata code for the data-generating process of the non-Gaussian designs is

```
. * Data generating process - non-Gaussian designs
. clear all
. set seed 1234
. matrix define var=(`v1´,`cov´\`cov´,`v2´)
. matrix define mu1=(`mu11´,`mu21´)
. matrix define mu2=(`mu12´,`mu22´)
. quietly drawnorm u11 u21, n(`sample´) m(mu1) cov(var) double
. quietly drawnorm u12 u22, n(`sample´) m(mu2) cov(var) double
. quietly generate d1=(uniform()<`pi´)
. quietly generate double u1=u11  if d1==1
. quietly generate double u2=u21  if d1==1
. quietly replace u1=u12  if d1==0
. quietly replace u2=u22  if d1==0
. local m1=(`pi´*`mu11´) + (1-`pi´)*(`mu12´)
. local m2=(`pi´*`mu21´) + (1-`pi´)*(`mu22´)
. local sd1=sqrt(`v1´+`pi´*(1-`pi´)*(`mu12´-`mu11´)^2)
. local sd2=sqrt(`v2´+`pi´*(1-`pi´)*(`mu22´-`mu21´)^2)
. quietly replace u1=(u1-`m1´)/`sd1´
. quietly replace u2=(u2-`m2´)/`sd2´
. quietly generate double x11=(uniform()*2-1)*sqrt(3)
. quietly generate double x21=(uniform()*2-1)*sqrt(3)
. quietly generate double x12=(invchi2(1,uniform())-1)/sqrt(2)
. quietly generate double x22=(invchi2(1,uniform())-3)/sqrt(6)
. quietly generate double y1s=`b11´*x11+`b12´*x12+u1
. quietly generate double y2s=`b21´*x21+`b22´*x22+u2
```

where the mixing probability `pi` and the set of parameters (`mu11, mu12, mu21, mu22, v1, v2, cov`) are chosen to obtain the selected levels of skewness, kurtosis, and correlation coefficient (see Preston [1953]).

Throughout the study, comparability of the probit, SNP, and SML estimators is obtained by imposing the scale normalization $\beta_{11} = \beta_{21} = 1$. For the parametric probit and the SNP estimators the normalization is imposed on the estimation results by taking the ratio of the estimated coefficients β_{12}/β_{11} and β_{23}/β_{21}, while for the SML estimator the normalization is directly imposed on the estimation process by constraining the coefficients of X_{11} and X_{21} to one. We always used the default starting values for the SNP and SML estimators. Furthermore, SNP and SML estimation were performed with prespecified values of R and h_n, respectively. To save computational time, no check was undertaken to investigate convergence to the global maximum rather than a local one, and we used rule-of-thumb values for R and h_n.

Tables 2–4 focus on the univariate binary-choice model for Y_2 and present summary statistics for the simulation results from 1000 replications with sample sizes 500, 1000, and 2000, respectively.[16] The normalization restrictions imply that there is only one free parameter in the model whose true value is 1. SNP estimation was performed under three alternative choices of R (with $R = 3, 4, 5$) as degree of the univariate Hermite polynomial expansion, while SML estimation was performed under three alternative values of the bandwidth parameter $h_n = n^{-1/\delta}$ (with $\delta = 6.02, 6.25, 6.5$). According to our simulation results, efficiency losses of the SNP and the SML estimators in the Gaussian design (Design 1) are rather small. In particular, the relative efficiency of the SNP estimator relative to the probit estimator ranges between 74% and 89%, while the relative efficiency of the SML estimator relative to the probit estimator ranges between 78% and 83%.[17]

A comparison of the three estimators in the non-Gaussian designs further suggests that SNP and SML estimators substantially dominate the probit estimator, specially in Designs 2 and 4 where error terms are generated from asymmetric distributions. First, the bias of the probit estimator is about 10% in Design 2 and about 6.5% in Design 4, while the bias of SNP and SML estimators never exceed 1.5%. Second, the ratios between the mean squared estimates (MSE) of the probit estimator and the MSEs of the two semiparametric estimators range between 1.7 and 5.3 in Design 2, and between 1.2 and 3.3 in Design 4. As expected, efficiency gains of the SNP and SML estimators relative to the probit estimator always increase as the sample size becomes larger. Third, the actual rejection rate of the Wald test for the probit estimate being equal to the true value of the parameter is quite far from the nominal value of 5%, while the actual rejection rates of the Wald tests for the SNP and SML estimates converge to their nominal values as the sample size becomes larger.

---

16. For each simulation design and selected sample size, we provide average and standard deviation of the estimates, mean square error of each comparable estimator, and rejection rate of the Wald test for each estimated coefficient being equal to its true value.

17. Results on the SNP estimator are consistent with the simulation results of Klein and Spady (1993) who find a relative efficiency of 78% on different simulation designs.

Table 2. Simulation results for the univariate binary-choice model ($n = 500$)

Des. 1	Est	SD	MSE	RR	Des. 2	Est	SD	MSE	RR
probit	1.0039	.1452	.0211	.0566	probit	.9017	.1598	.0352	.1776
snp$_3$	1.0041	.1581	.0250	.0627	snp$_3$	1.0039	.1446	.0209	.0737
snp$_4$	1.0104	.1624	.0265	.0688	snp$_4$	.9930	.1443	.0209	.0898
snp$_5$	1.0074	.1684	.0284	.1021	snp$_5$	.9931	.1438	.0207	.0979
sml$_1$	1.0157	.1633	.0269	.0890	sml$_1$	.9855	.1405	.0199	.0838
sml$_2$	1.0204	.1634	.0271	.0829	sml$_2$	.9876	.1387	.0194	.0757
sml$_3$	1.0248	.1619	.0268	.0728	sml$_3$	.9900	.1377	.0191	.0676
Des. 3	Est	SD	MSE	RR	Des. 4	Est	SD	MSE	RR
probit	1.0091	.1495	.0224	.0443	probit	.9354	.1461	.0255	.1227
snp$_3$	1.0206	.1690	.0290	.0622	snp$_3$	.9991	.1397	.0195	.0563
snp$_4$	1.0125	.1718	.0297	.0802	snp$_4$	1.0027	.1389	.0193	.0724
snp$_5$	1.0162	.1773	.0317	.0970	snp$_5$	1.0003	.1404	.0197	.0785
sml$_1$	1.0432	.1675	.0299	.0665	sml$_1$	.9749	.1457	.0219	.1066
sml$_2$	1.0487	.1684	.0307	.0675	sml$_2$	.9784	.1439	.0212	.0986
sml$_3$	1.0544	.1698	.0318	.0643	sml$_3$	.9821	.1423	.0206	.0915

Table 3. Simulation results for the univariate binary-choice model ($n = 1000$)

Des. 1	Est	SD	MSE	RR	Des. 2	Est	SD	MSE	RR
probit	1.0048	.1059	.0112	.0592	probit	.9064	.1303	.0257	.2452
snp$_3$	1.0010	.1120	.0125	.0602	snp$_3$	1.0104	.1054	.0112	.0701
snp$_4$	1.0065	.1149	.0133	.0582	snp$_4$	1.0030	.1038	.0108	.0801
snp$_5$	1.0040	.1189	.0142	.0772	snp$_5$	1.0011	.1032	.0107	.0761
sml$_1$	1.0119	.1171	.0138	.0762	sml$_1$	.9977	.1044	.0109	.0801
sml$_2$	1.0158	.1168	.0139	.0702	sml$_2$	.9994	.1037	.0107	.0771
sml$_3$	1.0197	.1164	.0139	.0612	sml$_3$	1.0011	.1030	.0106	.0771
Des. 3	Est	SD	MSE	RR	Des. 4	Est	SD	MSE	RR
probit	1.0022	.1029	.0106	.0484	probit	.9353	.1109	.0165	.1595
snp$_3$	1.0092	.1124	.0127	.0494	snp$_3$	1.0000	.0982	.0096	.0562
snp$_4$	1.0038	.1149	.0132	.0575	snp$_4$	1.0074	.0975	.0096	.0632
snp$_5$	1.0031	.1175	.0138	.0676	snp$_5$	1.0052	.0972	.0095	.0702
sml$_1$	1.0294	.1130	.0136	.0494	sml$_1$	.9876	.1024	.0106	.1013
sml$_2$	1.0343	.1138	.0141	.0434	sml$_2$	.9904	.1012	.0103	.0943
sml$_3$	1.0396	.1151	.0148	.0464	sml$_3$	.9931	.1002	.0101	.0832

Table 4. Simulation results for the univariate binary-choice model ($n = 2000$)

Des. 1	Est	SD	MSE	RR	Des. 2	Est	SD	MSE	RR
probit	1.0005	.0717	.0051	.0502	probit	.8969	.1199	.0250	.4320
snp$_3$	.9957	.0753	.0057	.0592	snp$_3$	1.0030	.0687	.0047	.0530
snp$_4$	1.0009	.0762	.0058	.0532	snp$_4$	.9981	.0676	.0046	.0540
snp$_5$	.9991	.0776	.0060	.0633	snp$_5$	.9974	.0677	.0046	.0560
sml$_1$	1.0035	.0783	.0061	.0602	sml$_1$	.9932	.0694	.0049	.0600
sml$_2$	1.0067	.0779	.0061	.0582	sml$_2$	.9944	.0687	.0048	.0530
sml$_3$	1.0102	.0778	.0062	.0582	sml$_3$	.9956	.0681	.0047	.0520
Des. 3	Est	SD	MSE	RR	Des. 4	Est	SD	MSE	RR
probit	1.0012	.0724	.0052	.0511	probit	.9304	.0940	.0137	.2593
snp$_3$	1.0047	.0783	.0061	.0521	snp$_3$	.9936	.0666	.0045	.0460
snp$_4$	1.0029	.0792	.0063	.0541	snp$_4$	1.0017	.0645	.0042	.0440
snp$_5$	1.0012	.0806	.0065	.0581	snp$_5$	.9995	.0644	.0042	.0450
sml$_1$	1.0226	.0806	.0070	.0591	sml$_1$	.9889	.0694	.0049	.0761
sml$_2$	1.0266	.0813	.0073	.0611	sml$_2$	.9908	.0684	.0048	.0711
sml$_3$	1.0310	.0825	.0078	.0571	sml$_3$	.9928	.0676	.0046	.0641

Tables 5–7 provide simulation results of the bivariate binary-choice model for Y_1 and Y_2. The normalization restrictions now imply that there are two free parameters in the model, one in equation 1 whose true value is -1 and one in equation 2 whose true value is 1. In this set of simulations, we compare performances of the bivariate probit estimator with those of the SNP estimator with $R_1 = R_2 = 4$. As for the univariate model, we find that efficiency losses of the SNP estimator in the Gaussian cases are very small. In this case, however, a larger sample size is usually needed to obtain substantial reductions in the MSE. Most of the efficiency gains typically occur for the coefficients of the second equation where there are stronger departures from Gaussianity (see table 1). Although rejection rates of the Wald tests for the SNP estimates are better than those for the bivariate probit estimates, they are still far from their nominal values even with a sample size $n = 2000$. This poor coverage of the SNP estimator is likely to be due to the incorrect choice of R_1 and R_2. In other words, the bivariate distribution of the latent regression errors may not be nested in the SNP model for the selected values of R_1 and R_2. For this kind of model misspecification, the coverage of the SNP estimator could be improved by using the Huber/White/sandwich estimator of the covariance matrix. Here our Monte Carlo simulations are based on the traditional calculation of the covariance matrix to make the results of the SNP estimator comparable with those of the SML estimators.[18]

---

[18] As explained in section 5.2, the SML commands do not support the `robust` option for the Huber/White/sandwich estimator of the covariance matrix.

Table 5. Simulation results for the bivariate binary-choice model ($n = 500$)

Des. 1	Est	SD	MSE	RR	Des. 2	Est	SD	MSE	RR
bipr$_1$	-1.0095	.1207	.0147	.0543	bipr$_1$	$-.9684$	.1193	.0152	.0775
snp2$_1$	-1.0086	.1315	.0174	.0834	snp2$_1$	$-.9756$	.1279	.0170	.1016
bipr$_2$	1.0025	.1413	.0200	.0492	bipr$_2$	.9056	.1550	.0329	.1640
snp2$_2$	.9933	.1507	.0228	.1055	snp2$_2$	.9500	.1669	.0304	.2072
Des. 3	Est	SD	MSE	RR	Des. 4	Est	SD	MSE	RR
bipr$_1$	-1.0163	.1209	.0149	.0437	bipr$_1$	$-.9681$	.1241	.0164	.1000
snp2$_1$	$-.9988$	.1307	.0171	.0863	snp2$_1$	$-.9875$	.1338	.0180	.1010
bipr$_2$	1.0096	.1441	.0208	.0416	bipr$_2$	.9376	.1416	.0239	.1230
snp2$_2$	1.0053	.1575	.0248	.1102	snp2$_2$	.9583	.1362	.0203	.1360

Table 6. Simulation results for the bivariate binary-choice model ($n = 1000$)

Des. 1	Est	SD	MSE	RR	Des. 2	Est	SD	MSE	RR
bipr$_1$	-1.0067	.0841	.0071	.0530	bipr$_1$	$-.9599$	.0902	.0098	.1020
snp2$_1$	-1.0081	.0898	.0081	.0670	snp2$_1$	$-.9673$	.0919	.0095	.1070
bipr$_2$	1.0056	.1014	.0103	.0590	bipr$_2$	.9105	.1261	.0239	.2360
snp2$_2$	.9962	.1082	.0117	.0910	snp2$_2$	.9537	.1201	.0166	.1900
Des. 3	Est	SD	MSE	RR	Des. 4	Est	SD	MSE	RR
bipr$_1$	-1.0135	.0843	.0073	.0453	bipr$_1$	$-.9615$	.0906	.0097	.1160
snp2$_1$	$-.9975$	.0897	.0081	.0866	snp2$_1$	$-.9908$	.0894	.0081	.0880
bipr$_2$	1.0021	.0985	.0097	.0433	bipr$_2$	.9382	.1076	.0154	.1540
snp2$_2$	.9975	.1029	.0106	.0816	snp2$_2$	.9610	.0985	.0112	.1230

Table 7. Simulation results for the bivariate binary-choice model ($n = 2000$)

Des. 1	Est	SD	MSE	RR	Des. 2	Est	SD	MSE	RR
bipr$_1$	-1.0011	.0566	.0032	.0521	bipr$_1$	$-.9574$	.0701	.0067	.1510
snp2$_1$	-1.0025	.0600	.0036	.0571	snp2$_1$	$-.9661$	.0702	.0061	.1250
bipr$_2$	.9991	.0701	.0049	.0470	bipr$_2$	.9015	.1152	.0230	.4050
snp2$_2$	.9926	.0706	.0050	.0691	snp2$_2$	.9513	.0903	.0105	.1970
Des. 3	Est	SD	MSE	RR	Des. 4	Est	SD	MSE	RR
bipr$_1$	-1.0125	.0578	.0035	.0481	bipr$_1$	$-.9540$	.0737	.0075	.1660
snp2$_1$	$-.9984$	.0609	.0037	.0601	snp2$_1$	$-.9897$	.0621	.0040	.0770
bipr$_2$	1.0015	.0706	.0050	.0531	bipr$_2$	.9337	.0908	.0126	.2460
snp2$_2$	.9990	.0691	.0048	.0621	snp2$_2$	.9630	.0739	.0068	.1320

Finally, tables 8–10 provide simulation results of the bivariate binary-choice model with sample selection for Y_1 and Y_2. In this case, selectivity was introduced by setting Y_2 to missing whenever $Y_1 = 0$. As for the bivariate model without sample selection, the normalization restrictions imply that there are two free parameters in the model, one in the selection equation whose true value is -1 and one in the main equation whose true value is 1. In this case, we compare performances of the bivariate probit estimator with sample selection, the SNP estimator with $R_1 = 4$ and $R_2 = 3$, and the SML estimator with $h_n = n^{-1/6.5}$ and $h_{n_1} = n_1^{-1/6.5}$. Our simulation results suggest again that efficiency losses of SNP and SML estimators with respect to a correctly specified probit estimator are rather small in both equations (namely, 87% and 80% in the first equation, and 86% and 70% in the second equation). In the non-Gaussian cases, the probit estimator is instead markedly biased and less efficient than the SNP and SML estimators specially in the presence of asymmetric distributions and relatively large-sample sizes. As before, the actual rejection rates of the Wald tests for the SNP and SML estimates are better than those for the parametric probit estimates, but they are still far from their nominal values of 5%. These coverage problems are likely to be due to the incorrect choice of the degree of the Hermite polynomial expansion and the bandwidth parameters, respectively. Given the computational burden of our Monte Carlo simulations, investigating the optimal choice of these parameters is behind the scope of this article. We leave this topic for future research.

Table 8. Simulation results for the bivariate binary-choice model with sample selection ($n = 500$)

Des. 1	Est	SD	MSE	RR	Des. 2	Est	SD	MSE	RR
heckpr$_1$	−1.0114	.1228	.0152	.0557	heckpr$_1$	−.9682	.1184	.0150	.0774
snp2s$_1$	−1.0096	.1357	.0185	.0902	snp2s$_1$	−.9795	.1291	.0171	.0905
sml2s$_1$	−1.0106	.1399	.0197	.0952	sml2s$_1$	−1.0066	.1338	.0179	.0905
heckpr$_2$	1.0048	.2047	.0419	.0588	heckpr$_2$	.8944	.2178	.0586	.1347
snp2s$_2$	.9895	.2281	.0521	.1631	snp2s$_2$	.9490	.1980	.0418	.1236
sml2s$_2$	1.0112	.2506	.0629	.1581	sml2s$_2$	.9667	.2435	.0604	.1759
Des. 3	Est	SD	MSE	RR	Des. 4	Est	SD	MSE	RR
heckpr$_1$	−1.0184	.1254	.0161	.0483	heckpr$_1$	−.9688	.1233	.0162	.0993
snp2s$_1$	−1.0044	.1361	.0185	.0924	snp2s$_1$	−.9951	.1338	.0179	.1003
sml2s$_1$	−1.0185	.1371	.0192	.0903	sml2s$_1$	−.9990	.1462	.0214	.1274
heckpr$_2$	1.0622	.2282	.0559	.0378	heckpr$_2$	.9133	.2127	.0527	.1474
snp2s$_2$	1.0019	.2221	.0493	.1082	snp2s$_2$	.9617	.2132	.0469	.1685
sml2s$_2$	1.0713	.2897	.0890	.1366	sml2s$_2$	.9477	.2477	.0641	.2247

Table 9. Simulation results for the bivariate binary-choice model with sample selection ($n = 1000$)

Des. 1	Est	SD	MSE	RR	Des. 2	Est	SD	MSE	RR
heckpr$_1$	−1.0070	.0857	.0074	.0512	heckpr$_1$	−.9593	.0897	.0097	.1030
snp2s$_1$	−1.0082	.0906	.0083	.0592	snp2s$_1$	−.9766	.0895	.0086	.0890
sml2s$_1$	−1.0081	.0944	.0090	.0813	sml2s$_1$	−1.0038	.0924	.0086	.0750
heckpr$_2$	1.0073	.1382	.0192	.0572	heckpr$_2$	.8959	.1694	.0395	.1610
snp2s$_2$	.9830	.1534	.0238	.1155	snp2s$_2$	.9554	.1507	.0247	.1080
sml2s$_2$	1.0063	.1611	.0260	.1185	sml2s$_2$	.9766	.1711	.0298	.1510
Des. 3	Est	SD	MSE	RR	Des. 4	Est	SD	MSE	RR
heckpr$_1$	−1.0136	.0853	.0075	.0446	heckpr$_1$	−.9613	.0910	.0098	.1234
snp2s$_1$	−1.0009	.0920	.0085	.0791	snp2s$_1$	−.9947	.0888	.0079	.0782
sml2s$_1$	−1.0129	.0940	.0090	.0751	sml2s$_1$	−.9973	.1013	.0103	.1224
heckpr$_2$	1.0583	.1665	.0311	.0538	heckpr$_2$	.9056	.1563	.0333	.1635
snp2s$_2$	.9974	.1538	.0237	.0974	snp2s$_2$	.9662	.1413	.0211	.1254
sml2s$_2$	1.0555	.1826	.0364	.1156	sml2s$_2$	.9471	.1668	.0306	.1956

Table 10. Simulation results for the bivariate binary-choice model with sample selection ($n = 2000$)

Des. 1	Est	SD	MSE	RR	Des. 2	Est	SD	MSE	RR
heckpr$_1$	−1.0020	.0576	.0033	.0502	heckpr$_1$	−.9580	.0694	.0066	.1420
snp2s$_1$	−1.0045	.0618	.0038	.0592	snp2s$_1$	−.9781	.0663	.0049	.0920
sml2s$_1$	−1.0016	.0641	.0041	.0662	sml2s$_1$	−1.0026	.0658	.0043	.0840
heckpr$_2$	1.0032	.0961	.0093	.0431	heckpr$_2$	.8880	.1471	.0342	.2500
snp2s$_2$	.9836	.1028	.0108	.0953	snp2s$_2$	.9561	.1082	.0136	.1230
sml2s$_2$	.9985	.1158	.0134	.1003	sml2s$_2$	.9784	.1189	.0146	.1250
Des. 3	Est	SD	MSE	RR	Des. 4	Est	SD	MSE	RR
heckpr$_1$	−1.0117	.0585	.0036	.0518	heckpr$_1$	−.9549	.0729	.0074	.1600
snp2s$_1$	−1.0021	.0615	.0038	.0640	snp2s$_1$	−.9939	.0619	.0039	.0650
sml2s$_1$	−1.0080	.0629	.0040	.0701	sml2s$_1$	−.9950	.0693	.0048	.1130
heckpr$_2$	1.0615	.1213	.0185	.0691	heckpr$_2$	.9033	.1316	.0267	.2470
snp2s$_2$	1.0015	.1037	.0108	.0813	snp2s$_2$	.9721	.0983	.0104	.0990
sml2s$_2$	1.0487	.1213	.0171	.0904	sml2s$_2$	.9572	.1224	.0168	.1900

8 Acknowledgments

I thank Franco Peracchi, David Drukker, Vince Wiggins, and two anonymous referees for their valuable comments and suggestions. I also thank Claudio Rossetti for help in programming the univariate SML estimator.

9 References

Chamberlain, G. 1986. Asymptotic efficiency in semiparametric models with censoring. *Journal of Econometrics* 32: 189–218.

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

———. 1987. Efficiency bounds for distribution-free estimators of the binary choice and censored regression models. *Econometrica* 55: 559–586.

De Luca, G., and F. Peracchi. 2007. A sample selection model for unit and item nonresponse in cross-sectional surveys. *CEIS Tor Vergata—Research Paper Series* 33: 1–44.

Gabler, S., F. Laisney, and M. Lechner. 1993. Seminonparametric estimation of binary-choice models with an application to labor-force participation. *Journal of Business and Economic Statistics* 11: 61–80.

Gallant, A. R., and D. W. Nychka. 1987. Semi-nonparametric maximum likelihood estimation. *Econometrica* 55: 363–390.

Gerfin, M. 1996. Parametric and semi-parametric estimation of binary response model of labour market participation. *Journal of Applied Econometrics* 11: 321–339.

Gould, W., J. Pitblado, and W. Sribney. 2006. *Maximum Likelihood Estimation with Stata*. 3rd ed. College Station, TX: Stata Press.

Horowitz, J. L. 1992. A smoothed maximum score estimator for the binary response model. *Econometrica* 60: 505–531.

Ichimura, H. 1993. Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics* 58: 71–120.

Ichimura, H., and L. F. Lee. 1991. Semiparametric least square of multiple index models: single equation estimation. In *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, ed. W. A. Barnett, J. Powell, and G. E. Tauchen, 350–351. Cambridge: Cambridge University Press.

Klein, R. W., and R. H. Spady. 1993. An efficient semiparametric estimator for binary response models. *Econometrica* 61: 387–421.

Lee, L. F. 1995. Semiparametric maximum likelihood estimation of polychotomous and sequential choice models. *Journal of Econometrics* 65: 381–428.

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

Melenberg, B., and A. van Soest. 1996. Measuring the costs of children: Parametric and semiparametric estimators. *Statistica Neerlandica* 50: 171–192.

Meng, C. L., and P. Schmidt. 1985. On the cost of partial observability in the bivariate probit model. *International Economic Review* 26: 71–85.

Pagan, A., and A. Ullah. 1999. *Nonparametric Econometrics*. Cambridge: Cambridge University Press.

Powell, J. L., J. H. Stock, and T. M. Stoker. 1989. Semiparametric estimation of index coefficients. *Econometrica* 57: 1403–1430.

Preston, E. J. 1953. A graphical method for the analysis of statistical distributions into two normal components. *Biometrika* 40: 460–464.

Stewart, M. B. 2004. Semi-nonparametric estimation of extended ordered probit models. *Stata Journal* 4: 27–39.

About the author

Giuseppe De Luca received his PhD in Econometrics and Empirical Economics from the University of Rome Tor Vergata in the summer of 2007. Currently, he works at Istituto per lo Sviluppo della Formazione Professionale dei Lavoratori (ISFOL) as research assistant in econometrics. His research interests are microeconometric theory and applications, nonparametric estimation methods, problems of nonresponse in sample surveys, and models of labor supply.