MONASH University

MONASH BUSINESS SCHOOL

ISSN 1440-771X

Department of Econometrics and Business Statistics

http://business.monash.edu/econometrics-and-business-statistics/research/publications

Binary Outcomes, OLS, 2SLS and IV Probit

Chuhui Li, Donald S. Poskitt, Frank Windmeijer and Xueyan Zhao

December 2019

Working Paper 05/19

Binary Outcomes, OLS, 2SLS and IV Probit

Chuhui Li[a], Donald S. Poskitt[a], Frank Windmeijer[b] and Xueyan Zhao[a]

[a]Dept of Econometrics and Business Statistics,
Monash University, Melbourne, Australia
[b]Dept of Economics, University of Bristol,
Bristol, UK

December 9, 2019

1 Introduction

In this paper, we first answer the following basic question. Let a binary dependent variable be generated by the following threshold specification

$$Y = 1\left(\beta_0 + \beta_1 X - U \geq 0\right),$$

where X is a normally distributed explanatory variable, U a continuous unit variance error with pdf f_U, and X and U independently distributed. Then what is the estimand for the OLS estimator in the linear probability model?

The linear probability model (LPM) for Y is given by

$$Y = \delta_0 + \delta_1 X + \varepsilon.$$

We show in Section 2 that the OLS estimator for δ_1 is a consistent estimator of the mean of the average marginal effects, given by

$$\eta_x = \beta_1 E_X\left[f_U\left(\beta_0 + \beta_1 X\right)\right].$$

Hence the OLS estimand of δ_1 for this setup is a meaningful object with a clear interpretation.

Given this result, we then move to an instrumental variables (IV) setup with endogenous X and normal instrumental variable Z. We show in Section 3 that the IV, or two-stage least squares (2SLS) estimator for δ_1 in the linear probability model is then *not* a consistent estimator for the mean of the average marginal effects, but estimates a different causal parameter, which is equal to the mean of the average marginal effects of Z, η_z, scaled by the linear effect γ_1 from the first-stage relationship $X = Z\gamma_1 - V$.

This result can be seen as an extension of the well-known ATE/LATE results for the binary treatment and instrument case, where OLS in the LPM is an estimator for the average treatment effect (ATE) under exogeneity, but the IV/Wald estimator identifies a local average treatment effect (LATE); see Angrist, Imbens, and Rubin (1996) and the discussion in Clarke and Windmeijer (2012). It highlights the fact that one cannot simply translate OLS and IV results developed for structural linear models to situations with a nonlinear data generating process like the threshold model for binary dependent variables above. In particular, tests for weak instruments, where the null of weak instruments is related to the relative bias of the IV estimator to that of the OLS estimator, clearly do not translate to this setting.

A leading application to which our setting and findings apply is Nunn and Qian (2014). Their binary outcome variable is an indicator whether or not there is a conflict in a given country during a given time period. The continuous endogenous explanatory variable of interest is the quantity of wheat aid shipped from the US, and the continuous instrument is the amount of US wheat production in the previous year interacted with the country specific average probability of receiving any US food aid over the period. The estimation method is linear 2SLS and the robust F-statistic is reported as a measure of instrument strength.

We apply and develop our findings further for the normal IV probit model in Section 4. The model structure with normally distributed variables enables us to derive exact results. We establish in Section 4.1 that for the two-step control function estimation procedure of Rivers and Vuong (1988), a double averaging over the marginal distributions of the first stage errors V and explanatory variable X leads to a consistent estimator of η_x. In contrast, a single averaging over the joint distribution of V and X, as for example proposed in Wooldridge (2010), is a consistent estimator of the same estimand as that of the 2SLS estimator, η_z/γ_1.

2 OLS in LPM for Binary Response Model with Exogenous Normal Explanatory Variable

Following the setup and notation as in Arellano (2008), we consider a binary outcome Y, an explanatory variable X and unit variance continuous error U, related via a binary index model

$$Y = 1\left(\beta_0 + \beta_1 X - U \geq 0\right). \tag{1}$$

Potential outcomes in this model are given by

$$Y\left(x\right) = 1\left(\beta_0 + \beta_1 x - U \geq 0\right).$$

Therefore, for an individual with error U, the effect on the potential outcome of a change in X from x to x' is given by

$$Y\left(x'\right) - Y\left(x\right) = 1\left(\beta_0 + \beta_1 x' - U \geq 0\right) - 1\left(\beta_0 + \beta_1 x - U \geq 0\right).$$

The average effect, over the distribution of U, is then given by

$$E_U\left[Y\left(x'\right) - Y\left(x\right)\right] = F_U\left(\beta_0 + \beta_1 x'\right) - F_U\left(\beta_0 + \beta_1 x\right),$$

where F_U is the cdf of U. For continuous X, the average marginal effects are given by

$$\eta\left(x\right) = \frac{\partial E_U\left[Y\left(x\right)\right]}{\partial x} = \frac{\partial F_U\left(\beta_0 + \beta_1 x\right)}{\partial x} = \beta_1 f_U\left(\beta_0 + \beta_1 x\right),$$

where f_U is the pdf of U.

A potential object of interest is the mean of the average marginal effects, taken over the distribution of X, and given by

$$\eta_x = \beta_1 E_X\left[f_U\left(\beta_0 + \beta_1 X\right)\right]. \tag{2}$$

Throughout the paper, we will investigate the properties of various estimators in this simple model design with a normally distributed X, as specified in the following assumption.

Assumption 1 $X \sim N\left(0, \sigma_x^2\right)$.

Consider the linear probability model specification

$$Y = \delta_0 + \delta_1 X + \varepsilon. \tag{3}$$

We will first investigate what quantity the OLS estimator of δ_1 estimates when the data generation process for Y is given by (1), under the exogeneity assumption that U is independent of X, or $U \perp X$.

Assumption 2 *Exogeneity,* $U \perp X$.

Under Assumption 2, it follows that $E[Y|X] = F_U(\beta_0 + \beta_1 X)$. The OLS estimand $\delta_{1,OLS}$ for δ_1 in (3) is then given by

$$\delta_{1,OLS} = \frac{Cov(X,Y)}{Var(X)} = \frac{E_X[XE[Y|X]]}{\sigma_x^2} = \frac{E_X[XF_U(\beta_0 + \beta_1 X)]}{\sigma_x^2}. \tag{4}$$

The equivalence of $\delta_{1,OLS}$ and η_x is stated in the next proposition.

Proposition 1 *Consider the binary outcome model (1) with X and U satisfying Assumptions 1 and 2. Then the OLS estimand $\delta_{1,OLS}$ as defined in (4) is equal to the mean of the average marginal effects η_x as defined in (2), $\delta_{1,OLS} = \eta_x$.*

Proof. See Appendix ■

The implication of Proposition (1) is that the OLS estimator of δ_1 in the linear probability model (3) is a consistent estimator of η_x, a meaningful object.

3 2SLS in LPM for Binary Response Model with Normal Instrument

Next, we allow for endogeneity, relaxing Assumption 2. We have a normally distributed instrument Z available, as specified in the following assumption.

Assumption 3 $Z \sim N(0, \sigma_z^2)$.

The first-stage relationship between X and Z is given by

$$X = Z\gamma_1 - V, \tag{5}$$

with $\gamma_1 \neq 0$, $Var(V) = \sigma_v^2$ and $Cov(U,V) = \sigma_{uv}$. Z is independent of both U and V,

Assumption 4 *Exogeneity of Z, $Z \perp (U, V)$.*

Then consider the 2SLS estimator of δ_1 in the linear model (3). Its estimand is given by

$$\delta_{1,2sls} = \frac{Cov\,(Z, Y)}{Cov\,(Z, X)} = \frac{E_Z\,[ZE\,[Y|Z]]}{\gamma_1 \sigma_z^2}. \tag{6}$$

From (1) and (5) it follows that

$$\begin{aligned} Y &= 1\,(\beta_0 + \beta_1 X - U \geq 0) \\ &= 1\,(\beta_0 + \beta_1 \gamma_1 Z - (U + \beta_1 V) \geq 0)\,. \end{aligned}$$

Let W_{β_1} be the standardised random variable

$$W_{\beta_1} = \frac{U + \beta_1 V}{\sqrt{1 + \beta_1^2 \sigma_v^2 + 2\beta_1 \sigma_{uv}}},$$

with cdf $F_{W_{\beta_1}}$ and pdf $f_{W_{\beta_1}}$. As before for x, we now get for the potential outcomes, setting $Z = z$,

$$Y\,(z) = 1\,(\beta_0^* + \beta_1^* \gamma_1 z - W_{\beta_1} \geq 0)\,,$$

where $\beta_j^* = \beta_j / \sqrt{1 + \beta_1^2 \sigma_v^2 + 2\beta_1 \sigma_{uv}}$ for $j = 0, 1$. The average effect on the potential outcome of a change in the value of Z from z to z', over the distribution of W_{β_1}, is then given by

$$E_{W_{\beta_1}}\,[Y\,(z') - Y\,(z)] = F_{W_{\beta_1}}\,(\beta_0^* + \beta_1^* \gamma_1 z') - F_{W_{\beta_1}}\,(\beta_0^* + \beta_1^* \gamma_1 z)\,.$$

and the average marginal effect for Z at z,

$$\eta\,(z) = \frac{\partial E_{W_{\beta_1}}\,[Y\,(z)]}{\partial z} = \frac{\partial F_{W_{\beta_1}}\,(\beta_0^* + \beta_1^* \gamma_1 z)}{\partial z} = \beta_1^* \gamma_1 f_{W_{\beta_1}}\,(\beta_0^* + \beta_1^* \gamma_1 z)\,.$$

The mean of these average marginal effects over the distribution of Z is then given by

$$\eta_z = \beta_1^* \gamma_1 E_Z \left[f_{W_{\beta_1}}\,(\beta_0^* + \beta_1^* \gamma_1 Z) \right]. \tag{7}$$

The relationship between η_z and $\delta_{1,2sls}$ is given in the next proposition.

Proposition 2 *Consider the binary outcome model (1), with instrumental variable Z related to X as in (5) and satisfying Assumptions 3 and 4. Let η_z be as defined in (7) and the 2SLS estimand $\delta_{1,2sls}$ as in (6). Then $\delta_{1,2sls}$ is given by*

$$\delta_{1,2sls} = \frac{\eta_z}{\gamma_1} = \beta_1^* E_Z \left[f_{W_{\beta_1}}\,(\beta_0^* + \beta_1^* \gamma_1 Z) \right].$$

Proof. Follows directly from expression (6) and the proof of Proposition 1, which establishes that $E_Z[ZE[Y|Z]]/\sigma_z^2 = \eta_z$. ■

It is clear that, in general, $\delta_{1,2sls} \neq \eta_x$, unless $\beta_1 = 0$, as then $\eta_x = \eta_z = 0$. We can therefore use the 2SLS estimation results to test the null hypothesis $H_0 : \beta_1 = 0$. From (2) it follows that this is equivalent to testing $H_0 : \eta_x = 0$. It is also clear that the weak instruments testing procedure of Stock and Yogo (2005) based on the relative bias of the 2SLS estimator relative to that of the OLS estimator does not apply here, as the procedures estimate different causal parameters in general, with again the exception at $\beta_1 = 0$. We will further address these issues in Section 5 below.

4 Probit Model

The probit model specifies U as a standard normal variable, and therefore, under exogeneity Assumption 2, $E[Y|X] = \Phi(\beta_0 + \beta_1 X)$, where $\Phi(.)$ is the standard normal cdf. Because of symmetry, we can now write model (1) equivalently as $Y = 1(\beta_0 + \beta_1 X + U \geq 0)$.

The mean of the average marginal effects is then given by

$$\eta_{xp} = \beta_1 E_X[\phi(\beta_0 + \beta_1 X)] \tag{8}$$

where $\phi(.)$ is the standard normal pdf.

Let $\widehat{\beta}_{0p}$ and $\widehat{\beta}_{1p}$ be the probit ML estimators of β_0 and β_1 then η_{xp} can be consistently estimated by

$$\widehat{\eta}_{xp} = \widehat{\beta}_{1p}\left(\frac{1}{n}\sum_{i=1}^{n}\phi\left(\widehat{\beta}_{0p} + \widehat{\beta}_{1p}X_i\right)\right). \tag{9}$$

As derived in Proposition 1, the OLS estimator $\widehat{\delta}_{1,OLS}$ of δ_1 in the linear specification (3) is an alternative consistent estimator of η_{xp}, but note that $\widehat{\delta}_{1,OLS}$ is robust to departures of the normality assumption $U \sim N(0,1)$.

With the density now fully specified in (8), we can further simplify the expression for η_{xp} using the following lemma:

Lemma 1 *Under the normality assumption of X, Assumption 1, it follows that*

$$E_X[\phi(\beta_0 + \beta_1 X)] = \frac{1}{\sqrt{1+\beta_1^2\sigma_x^2}}\phi\left(\frac{\beta_0}{\sqrt{1+\beta_1^2\sigma_x^2}}\right).$$

Proof. See Appendix. ■

It follows then from Lemma 1 that

$$\eta_{xp} = \beta_1 E_X \left[\phi \left(\beta_0 + \beta_1 X\right)\right] = \frac{\beta_1}{\sqrt{1+\beta_1^2\sigma_x^2}} \phi \left(\frac{\beta_0}{\sqrt{1+\beta_1^2\sigma_x^2}} \right). \tag{10}$$

Next, we allow for endogeneity. The relationship between X and instrument Z is given by

$$X = \gamma_0 + \gamma_1 Z + V, \tag{11}$$

with $\gamma_1 \neq 0$. We assume that both Z and V are normally distributed, so that the conditions for IV probit maximum likelihood estimation are fulfilled:

Assumption 5

$$\begin{pmatrix} U \\ V \\ Z \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sigma_{uv} & 0 \\ \sigma_{uv} & \sigma_v^2 & 0 \\ 0 & 0 & \sigma_z^2 \end{pmatrix} \right).$$

It follows from Assumptions 1 and 5 that $\gamma_0 = 0$.

Due to the endogeneity problem, the probit estimator $\widehat{\eta}_{xp}$ as defined in (9) and the OLS estimator $\widehat{\delta}_{1,OLS}$ are no longer consistent estimators of η_{xp}. From (11) and Assumption 5 it follows that

$$\begin{pmatrix} X \\ U \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \sigma_{uv} \\ \sigma_{uv} & 1 \end{pmatrix} \right)$$

and so

$$U = \psi X + E,$$

where $\psi = \sigma_{uv}/\sigma_x^2$ and $E \sim N\left(0, 1-\tau^2\right)$ with $\tau = \sigma_{uv}/\sigma_x$. Therefore

$$E\left[Y|X\right] = \Phi \left(\frac{\beta_0 + \left(\beta_1 + \psi\right) X}{\sqrt{1-\tau^2}} \right) = \Phi \left(\widetilde{\beta}_0 + \widetilde{\beta}_1 X \right),$$

where $\widetilde{\beta}_0 = \beta_0/\sqrt{1-\tau^2}$ and $\widetilde{\beta}_1 = \left(\beta_1 + \psi\right)/\sqrt{1-\tau^2}$. It follows then from Lemma 1 that the probit model in this design with endogeneity estimates as the mean of the average marginal effects

$$\begin{aligned} \widetilde{\eta}_{xp} &= \widetilde{\beta}_1 E_X \left[\phi \left(\widetilde{\beta}_0 + \widetilde{\beta}_1 X \right) \right] \\ &= \frac{\widetilde{\beta}_1}{\sqrt{1+\widetilde{\beta}_1^2\sigma_x^2}} \phi \left(\frac{\widetilde{\beta}_0}{\sqrt{1+\widetilde{\beta}_1^2\sigma_x^2}} \right), \end{aligned} \tag{12}$$

which is the estimand of the probit based estimator $\widehat{\eta}_{xp}$. It follows directly from Proposition 1 that then also $\delta_{1,ols} = \widetilde{\eta}_{xp}$. Clearly, $\widetilde{\eta}_{xp} \neq \eta_{xp}$ unless $\sigma_{uv} = 0$, as then $\psi = \tau = 0$.

For this probit model specification, we get the following expression for the estimand $\delta_{1,2sls}$.

Proposition 3 *For model specifications (1), (11) and under Assumption 5, the 2SLS estimand $\delta_{1,2sls}$ for the 2SLS estimator of δ_1 in the linear probability model (3) as defined in (6) is given by*

$$\delta_{1,2sls} = \frac{\beta_1}{\sqrt{1+\beta_1^2\sigma_x^2+2\beta_1\sigma_{uv}}}\phi\left(\frac{\beta_0}{\sqrt{1+\beta_1^2\sigma_x^2+2\beta_1\sigma_{uv}}}\right). \tag{13}$$

Hence $\delta_{1,2sls} \neq \eta_{xp}$ unless $\beta_1 = 0$ and/or $\sigma_{uv} = 0$.

Proof. See Appendix. ■

As an illustration, Figure 1 plots the bias $\delta_{1,ols} - \eta_{xp}$ and the difference $\delta_{1,2sls} - \eta_{xp}$ as a function of σ_{uv}, for values of $\beta_1 = 1$, $\beta_0 = -0.2$, $\sigma_x^2 = 1$, resulting in $\eta_{xp} = 0.279$. For this design, the OLS bias is negative for negative values of σ_{uv} and positive for positive values of σ_{uv}. This is the opposite for the 2SLS difference, which has a positive (negative) difference for negative (positive) values of σ_{uv}. The difference of $\delta_{1,2sls}$ from η_{xp} is especially large at the more negative values of σ_{uv}.

4.1 ML and Two-Step Estimation

For an iid sample $\{Y_i, X_i, Z_i\}_{i=1}^n$, the IV probit maximum likelihood estimator is consistent and asymptotic normal under Assumption 5. The ML estimator for $\theta = (\beta_0, \beta_1, \gamma_0, \gamma_1, \rho, \sigma_v)$ is given by

$$\widehat{\theta}_{ml} = \arg\min_{\theta}\sum_{i=1}^{n}\log\left(\Phi\left(G_i\left(\theta\right)\right)^{Y_i}\left(1-\Phi\left(G_i\left(\theta\right)\right)\right)^{1-Y_i}\frac{1}{\sigma_v}\phi\left(\frac{X_i-\gamma_0-\gamma_1 Z_i}{\sigma_v}\right)\right),$$

where

$$G_i\left(\theta\right) = \frac{\beta_0+\beta_1 X_i+\left(\rho/\sigma_v\right)\left(X_i-\gamma_0-\gamma_1 Z_i\right)}{\sqrt{\left(1-\rho^2\right)}},$$

with $\rho = \sigma_{uv}/\sigma_v$.

The causal average marginal effect is therefore consistently estimated by

$$\widehat{\eta}_{ml}\left(x\right) = \widehat{\beta}_{1,ml}\phi\left(\widehat{\beta}_{0,ml}+\widehat{\beta}_{1,ml}x\right),$$

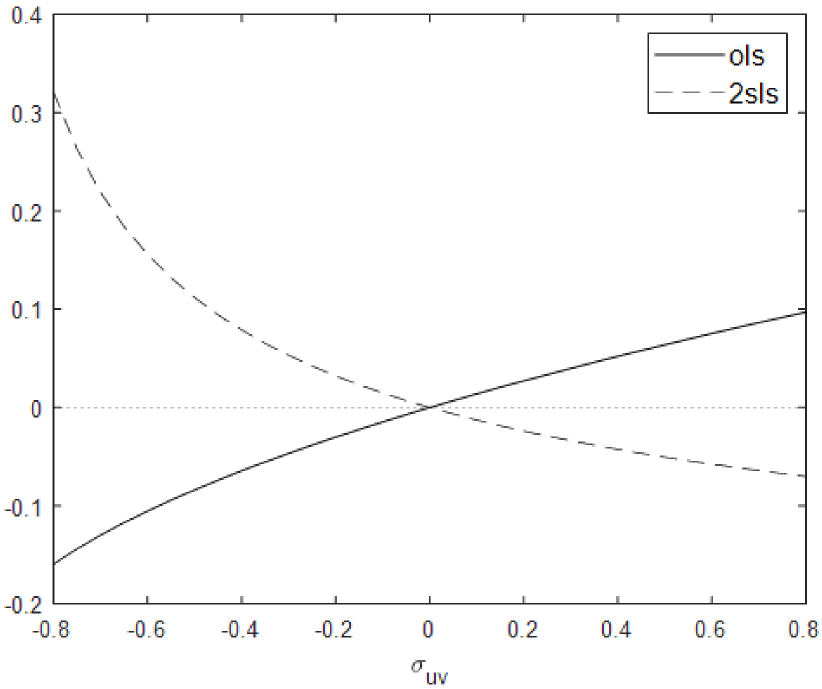

Figure 1: Bias $\delta_{1,ols} - \eta_{xp}$ and difference $\delta_{1,2sls} - \eta_{xp}$ as a function of σ_{uv}. $\beta_0 = -0.2$; $\beta_1 = 1$; $\sigma_x^2 = 1$; $\eta_{xp} = 0.279$.

and the mean of the average marginal effects by

$$\widehat{\eta}_{x,ml} = \widehat{\beta}_{1,ml} \left(\frac{1}{n} \sum_{i=1}^{n} \phi \left(\widehat{\beta}_{0,ml} + \widehat{\beta}_{1,ml} X_i \right) \right). \tag{14}$$

A popular alternative estimation method is the two-step control function approach of Rivers and Vuong (1988). From Assumption 5 it follows that

$$U = \omega V + W,$$

where $\omega = \sigma_{uv}/\sigma_v^2 = \rho/\sigma_v$, and $W \sim N\left(0, 1 - \rho^2\right)$.

Therefore,

$$\begin{aligned} E\left[Y|Z,V\right] &= \Phi\left(\frac{\beta_0 + \beta_1 X + \omega V}{\sqrt{1-\rho^2}}\right) \\ &= \Phi\left(\beta_{0\rho} + \beta_{1\rho} X + \omega_\rho V\right), \end{aligned}$$

where e.g. $\beta_{1\rho} = \beta_1/\sqrt{1-\rho^2}$.

Following Wooldridge (2010, p 588), the average marginal effects are obtained by taking derivatives of the average structural function

$$s(x) = E_V\left[\Phi\left(\beta_{0\rho} + \beta_{1\rho} x + \omega_\rho V\right)\right].$$

Let $\widehat{V}_i = X_i - \widehat{\gamma}_0 - Z_i \widehat{\gamma}_1$ be the OLS residual, then a consistent estimator of $s(x)$ is given by

$$\widehat{s}_{2s}(x) = \frac{1}{n}\sum_{i=1}^{n}\Phi\left(\widehat{\beta}_{0\rho} + \widehat{\beta}_{1\rho} x + \widehat{\omega}_\rho \widehat{V}_i\right),$$

where $\widehat{\beta}_{\rho 0}$, $\widehat{\beta}_{\rho 1}$ and $\widehat{\omega}_\rho$ are the standard probit ML estimators in a probit model for Y_i with X_i and $\widehat{V}_i$ as explanatory variables. It therefore follows that a consistent estimator for the average marginal effect $\eta_p(x)$ is given by

$$\widehat{\eta}_{2s}(x) = \widehat{\beta}_{1\rho}\left(\frac{1}{n}\sum_{i=1}^{n}\phi\left(\widehat{\beta}_{0\rho} + \widehat{\beta}_{1\rho} x + \widehat{\omega}_\rho \widehat{V}_i\right)\right). \tag{15}$$

A consistent estimator of the mean of the average marginal effects η_{xp} is then given by

$$\widehat{\eta}_{x2s} = \widehat{\beta}_{1\rho}\left(\frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n}\phi\left(\widehat{\beta}_{0\rho} + \widehat{\beta}_{1\rho} X_j + \widehat{\omega}_\rho \widehat{V}_i\right)\right). \tag{16}$$

In contrast, Wooldridge (2010, p 589), proposes to estimate an average partial effect as

$$\widehat{\alpha}_{2s} = \widehat{\beta}_{1\rho}\left(\frac{1}{n}\sum_{i=1}^{n}\phi\left(\widehat{\beta}_{0\rho} + \widehat{\beta}_{1\rho} X_i + \widehat{\omega}_\rho \widehat{V}_i\right)\right). \tag{17}$$

This estimator $\widehat{\alpha}_{2s}$ is a a consistent estimator of α, given by

$$\alpha = \beta_{1\rho} E_{X,V}\left[\phi\left(\beta_{0\rho} + \beta_{1\rho} X + \omega_\rho V\right)\right]. \tag{18}$$

The next proposition shows that α is equal to the 2SLS estimand $\delta_{1,2sls}$.

Proposition 4 *For the model specifications (1), (11) and under Assumption 5, let α be as defined in (18) and $\delta_{1,2sls}$ as in (13), then $\alpha = \delta_{1,2sls}$.*

Proof. See Appendix ■

Table 1: Estimation results, $\eta_{xp} = 0.264$, $n = 500$

ρ	-0.75	-0.5	-0.25	0	0.25	0.5	0.75
$\delta_{1,ols}$	0.179	0.211	0.239	0.264	0.287	0.308	0.328
$\widehat{\delta}_{1,ols}$	0.180	0.211	0.239	0.264	0.287	0.308	0.328
	(0.016)	(0.014)	(0.013)	(0.012)	(0.011)	(0.011)	(0.011)
$\widehat{\eta}_{xp}$	0.179	0.211	0.239	0.264	0.287	0.307	0.328
	(0.015)	(0.014)	(0.012)	(0.011)	(0.011)	(0.012)	(0.014)
$\delta_{1,2sls}$	0.449	0.351	0.298	0.264	0.239	0.220	0.205
$\widehat{\delta}_{1,2sls}$	0.451	0.352	0.299	0.265	0.240	0.220	0.205
	(0.045)	(0.040)	(0.038)	(0.036)	(0.035)	(0.034)	(0.033)
$\widehat{\alpha}_{2s}$	0.451	0.352	0.299	0.264	0.239	0.220	0.204
	(0.046)	(0.040)	(0.037)	(0.035)	(0.034)	(0.032)	(0.030)
$\widehat{\eta}_{x2s}$	0.263	0.263	0.262	0.261	0.261	0.260	0.260
	(0.010)	(0.012)	(0.016)	(0.020)	(0.025)	(0.028)	(0.031)

Notes: Means and (standard deviations) of 1000 MC replications

5 Some Monte Carlo Results

To illustrate the findings, we present results from a small Monte Carlo exercise. The data are generated as,

$$\begin{pmatrix} U_i \\ V_i \\ Z_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho & 0 \\ \rho & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right);$$

$$\begin{aligned} X_i &= \gamma_1 Z_i + V_i; \\ Y_i &= 1\left(\beta_0 + \beta_1 X_i - U_i \geq 0\right). \end{aligned}$$

We draw samples of size $n = 500$, and set $\gamma_1 = 0.5$, $\beta_0 = -0.2$ and $\beta_1 = 1$. At these parameter values, the mean of the average marginal effects is given by $\eta = 0.264$. We vary $\rho = -0.75, -0.5, ..., 0.75$, resulting in values for $\delta_{1,ols}$ ranging from 0.179 at $\rho = -0.75$ to 0.328 at $\rho = 0.75$, and values of $\delta_{1,2sls}$ ranging from 0.449 at $\rho = -0.75$, to 0.205 at $\rho = 0.75$. Table 1 presents estimation results for OLS, $\widehat{\delta}_{1,ols}$; $\widehat{\eta}_P$ based on the standard probit estimator; 2SLS, $\widehat{\delta}_{1,2sls}$; and $\widehat{\alpha}$ and $\widehat{\eta}_{2s}$ as defined in (17) and (15) based on the two-step probit estimation results. The results clearly confirm the theoretical results obtained above.

5.1 2SLS Wald Test for $H_0 : \beta_1 = 0$

We next analyse the behaviour of the 2SLS Wald test for testing $H_0 : \beta_1 = 0$. This Wald test is given by

$$W_{2sls} = \frac{\widehat{\delta}_{1,2sls}^2}{V\widehat{a}r_r\left(\widehat{\delta}_{1,2sls}\right)},$$

where $V\widehat{a}r_r\left(\widehat{\delta}_{1,2sls}\right)$ is a robust variance estimator. Figure 2 shows the rejection frequencies of this test at the 5% level as a function of the value of β_1 for values of $\rho = -0.75$; 0; 0.75. The other values of the parameters are as above, including the sample size of $n = 500$. The number of Monte Carlo replications for each value of β_1 is equal to $10,000$. We see that the test has correct size, but that power is affected by the bias of the 2SLS under the alternative. For example, for $\rho = -0.75$, there is a negative bias as shown in Table 1 and power is less than nominal size for small positive values of β_1.

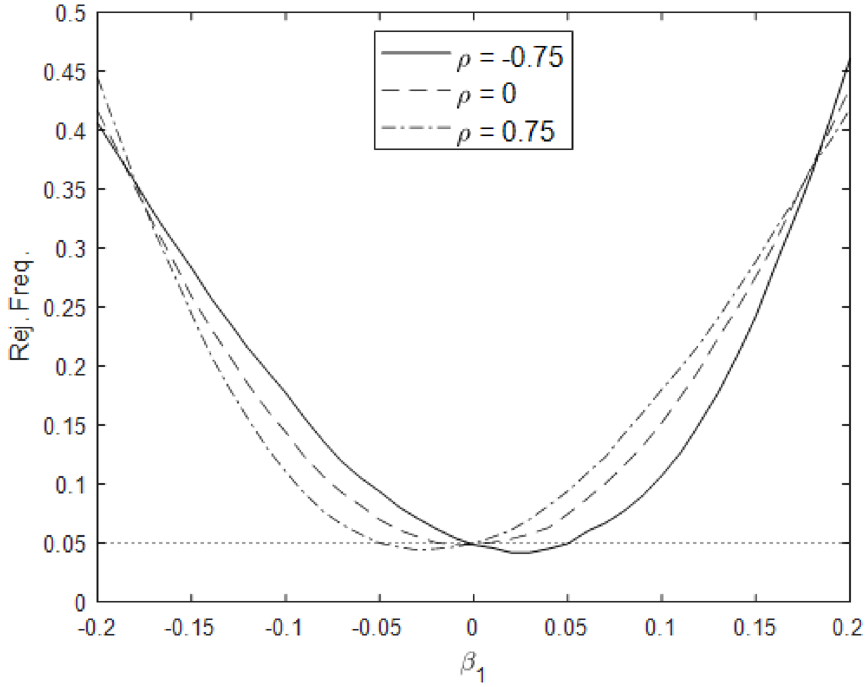

Figure 2: Rejection frequency of Wald test for $H_0 : \beta_1 = 0$ at 5% level.

5.2 2SLS and Weak Instruments

Stock and Yogo (2005) derived weak instruments critical values for the standard F-test for testing $H_0 : \gamma_1 = 0$. The null hypothesis for weak instruments is then formulated in terms of the 2SLS based Wald test size distortion. For example, for a maximal size of 10% at the 5% nominal level, the critical value for a single endogenous variable, just identified model is equal to 16.38. This test procedure is generally quite conservative, as the Wald test size distortion is maximal at $\rho_{\varepsilon v} = 1$, with ε the error in the linear probability model (3). The critical values apply only when both ε and V are conditionally homoskedastic. As ε is conditionally heteroskedastic, the Stock and Yogo (2005) critical values may not apply, see e.g. Bun and de Haan (2010), Montiel Olea and Pflueger (2013) and Andrews (2018). However, the design considered here is conditionally homoskedastic when $\beta_1 = 0$, and we can hence use the Stock and Yogo (2005) weak instrument critical values for the Wald test size distortion under the null $H_0 : \beta_1 = 0$. We can also use the relative bias based critical values when $\beta_1 = 0$ and the model is overidentified, as at that point it is meaningful to consider the bias of the 2SLS estimator, relative to that of the OLS estimator.

Figure 3 shows the rejection frequencies of the 2SLS based Wald test for the true null $H_0 : \beta_1 = 0$ for $\rho = 0.999$ and for different values of the expectation of the concentration parameter $E\left(\mu_n^2\right) = n\gamma_1^2\sigma_z^2/\sigma_v^2$, by varying the value of $\gamma_1 = 0, 0.01, ..., 0.3$. Other values of the parameters are as above, including the sample size of $n = 500$. The rejection frequencies are plotted against the 0.95 quantile of the F-statistic, for $20,000$ replications. We see that the Wald rejection frequencies are larger than 10% for small values of the concentration parameter. However, at a critical value of around 10, the rejection frequency drops below 10%, and so the SY critical value of 16.38 appears conservative.

Figure 4 shows relative bias results for a design with $k_z = 3$ independent standard normally distributed instrumental variables. The first-stage is given by

$$X = cZ'\iota_3 + V \tag{19}$$

where $\iota_3 = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}'$, and the concentration parameter is varied by varying the value c. Again, $\beta_1 = 0$, and $n = 500$, whilst we set $\rho = 0.5$. The Stock-Yogo critical value for the F-statistic for a 10% relative bias is given by 9.08. The results in Figure 4 confirm

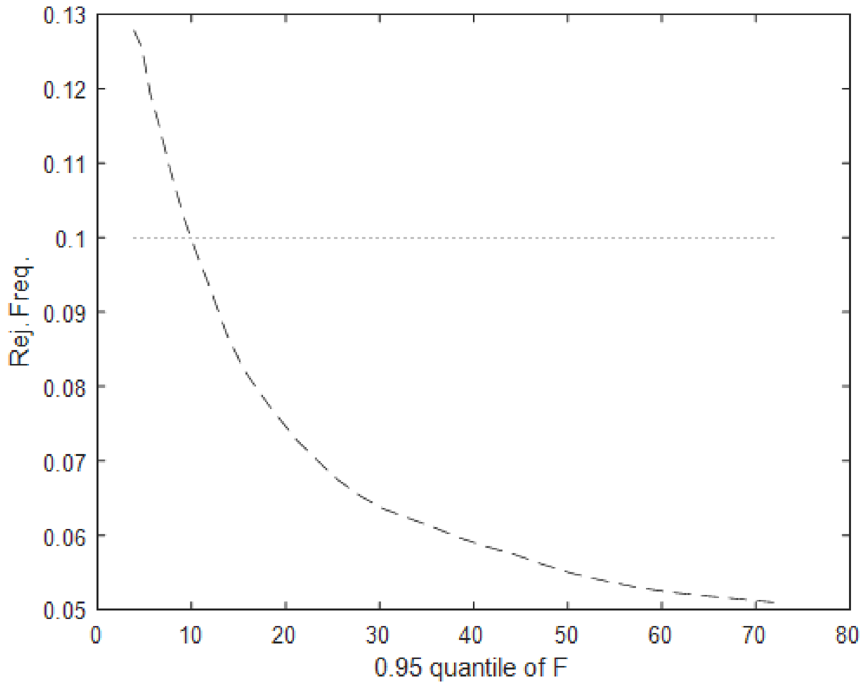

Figure 3: Rejection frequencies of Wald test for $H_0 : \beta_1 = 0$, $\rho = 0.999$.

the Stock-Yogo results at $\beta_1 = 0$, where we find a relative bias of 0.10 at a 5% critical value of the F-statistic of 9.10.

5.3 ML Estimator and Weak Instruments

Whilst the results obtained above for the 2SLS Wald test and relative bias are limited to the value $\beta_1 = 0$, they are useful in practice as the hypothesis $H_0 : \beta_1 = 0$ is often the main hypothesis of interest. For other values of β_1 we need to consider different estimators. Keeping the focus on η_x, we consider the ML IV estimator $\widehat{\eta}_{x,ml}$ as defined in (14) that is a consistent and normal estimator of η_x if the assumptions of the probit IV model hold.

The design is as above, with $k_z = 3$ and $n = 1000$. We again vary the information content by varying the value of c in (19). We consider the values $\beta_1 = 0.3$ and $\beta_1 = 0.6$, with values of η_x approximately equal to 0.11 and 0.20 respectively. We further first set the value of ρ equal to 0.5. The number of Monte Carlo replications is equal to $20,000$, for each value of c.

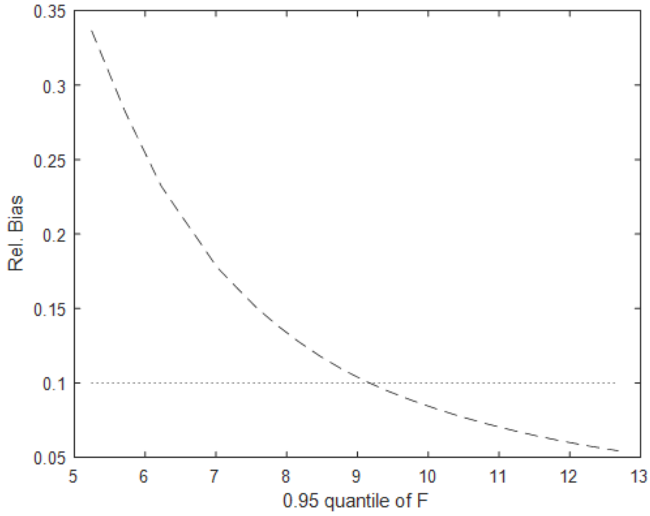

Figure 4: Relative bias, $\beta_1 = 0$, $k_z = 3$.

Figure 5 shows the bias of of the OLS estimator $\widehat{\delta}_{1,ols}$ and the ML IV estimator $\widehat{\eta}_{x,ml}$ as a function of the mean value of the first-stage F-statistic. The bias of the OLS estimator is substantially smaller when $\beta_1 = 0.6$, $\eta_x \approx 0.20$ than when $\beta_1 = 0.3$, $\eta_x \approx 0.11$, whereas the ML IV bias is slightly larger when β_1 is 0.6. It follows from this that the bias of $\widehat{\eta}_{x,ml}$, relative to that of $\widehat{\delta}_{1,ols}$, or equivalently $\widehat{\eta}_{xp}$, is much larger for the larger value of η_x at the same values of the F-statistic. This is confirmed in Figure 6. Hence the Stock and Yogo (2005) critical values for the 2SLS bias relative to the OLS bias in the linear homoskedastic setting do not apply here.

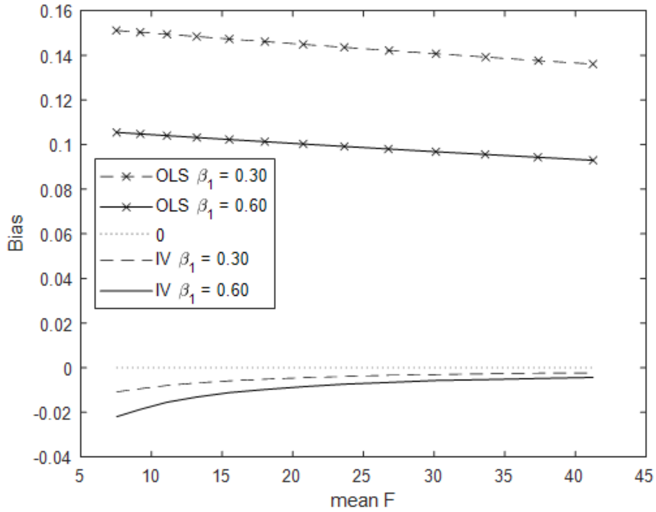

Figure 5: Bias of OLS and ML IV estimators of η_x, $\rho = 0.5$.

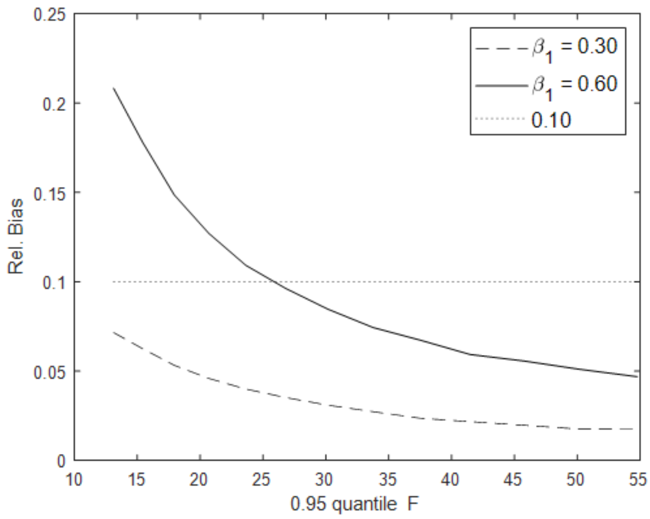

Figure 6: Relative bias of estimators of η_x, $\rho = 0.5$.

Figure 7 shows the size behaviour of the Wald test, testing the true value of η_x. The behaviour of the test is not affected by the different values of η_x and the critical value of the F-test for a size of 10% at the 5% level is approximately 28, larger than the Stock and Yogo (2005) critical value of 22.30 for the linear case.

Finally, Figure 8 shows the Wald test rejection frequencies for the $\beta_1 = 0.3$, $\eta_x = 0.11$ case for different values of the correlation, $\rho = 0.3, 0.5, 0.7, 0.9$. Interestingly, the critical values of the F-test decline with increasing ρ, the critical value at $\rho = 0.9$ approximately equal to 21.

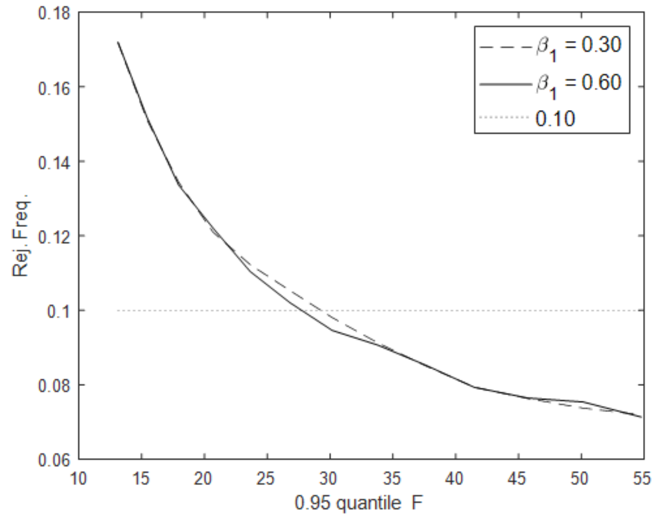

Figure 7: Rejection frequency of Wald test for $H_0 : \eta_x = \eta_{x0}$, $\rho = 0.5$.

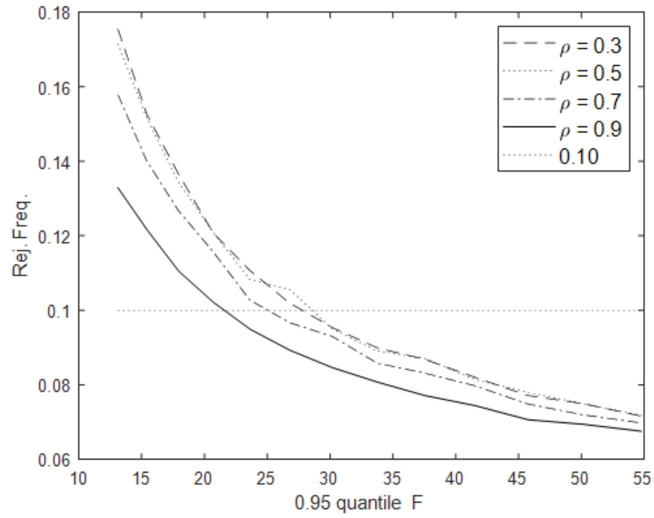

Figure 8: Rejection frequencies Wald test, $\beta_1 = 0.3$, $\eta_x = 0.11$, for various values of ρ.

6 Conclusions

Binary outcome models are frequently used in empirical studies, and linear probability models are often specified and estimated by the two stage least squares instrumental variables estimator when the continuous treatment variable is endogenous. This linearisation of binary models is particularly popular when the issue of weak instruments is a potential concern, where the Stock and Yogo (2005) test, suitable for linear models, is then applied to detect a weak instrument problem. This paper presents both theoretical and Monte Carlo results to show the implications and consequences of this popular linearisation of binary models when the data generating process (DGP) is a threshold crossing latent equation model such as Probit. The key results from the paper can be summarised as follows.

1. When a normally distributed treatment variable X is exogenous, and the true DGP is an additive threshold crossing model, the OLS estimator for the coefficient of X in the LPM is a consistent estimator for η_x, the average marginal effect (AME) of X averaged over the distribution of X.

2. When X is endogenous and the data are generated from a Probit specification, the estimand $\delta_{1,2sls}$ for the 2SLS estimator of the LPM with instrument Z is not the same as η_{xp}, the more desired mean of the AME of X over the distribution X. The 2SLS in fact estimates a different quantity akin to the LATE concept for linear models.

3. The 2SLS estimand is shown to have an interesting connection with the Rivers and Vuong (1988) two-step control function estimator. If the Rivers and Vuong (1988) two-step control function estimator is used, the single averaging over all observed pairs of x_i and estimated residual $\widehat{v}_i$ (as suggested in Wooldridge 2010, p. 588) gives a consistent estimate of the 2SLS estimand δ_{2sls}, whilst counterfactual double averaging over $\widehat{v}_i$ and x_j estimates the correct mean AME of X, i.e. η_{xp}.

Our paper also presents some interesting results for the special case of $\beta_1 = 0$ and so the true AME of X is equal to zero, which has important implications for empirical researchers. We show that if $\beta_1 = 0$, $\delta_{1,2sls} = \eta_{xp}$, so 2SLS is a consistent estimator for the mean AME of X, and the Wald test for $H_0 : \beta_1 = 0$ has the correct size. Under the special case of $\beta_1 = 0$, our Monte Carlo results also show that the Stock and Yogo test for the null of weak instruments is conservative.

These results have interesting implications for empirical practitioners, where the hypothesis the researchers often wish to test is $H_0 : \beta_1 = 0$, or there is a zero treatment effect of X. If they wish to use LPM-2SLS to estimate a threshold crossing model, and instruments are not weak, then the size of the Wald test is controlled. So if $H_0 : \beta_1 = 0$ is rejected, one can be relatively confident that the treatment variable X has non-zero effect on outcome Y, but one cannot be certain about the estimated magnitude of the estimated mean AME of X, as when $\beta_1 \neq 0$, the 2SLS estimator and inference no longer apply to η_x. Therefore, the message for researchers is clear. When an endogenous continuous treatment variable is present for binary outcome variable models, the 2SLS estimator in the LPM does not estimate the mean AME of the treatment in general if the true model is a non-linear threshold crossing model such as the IV-Probit. The Stock and Yogo weak instruments test also does not apply. See Frazier, Renault, Zhang, and Zhao (2019) for a proposed weak IV test for IV-Probit type of models, and Magnusson (2010) for weak IV robust inference.

References

Andrews, I. (2018): "Valid Two-Step Identification-Robust Confidence Sets for GMM," *The Review of Economics and Statistics*, 100, 337–348.

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.

Arellano, M. (2008): "Binary Models with Endogenous Explanatory Variables," Tech. rep., CEMFI.

Bun, M. and M. de Haan (2010): "Weak Instruments and the First Stage F-statistic in IV Models with a Nonscalar Error Covariance Structure," Discussion Paper 2010/02, University of Amsterdam.

Clarke, P. S. and F. Windmeijer (2012): "Instrumental Variable Estimators for Binary Outcomes," *Journal of the American Statistical Association*, 107, 1638–1652.

Frazier, D., E. Renault, L. Zhang, and X. Zhao (2019): "Weak instrument test in discrete choice models," paper presented at the International Association for Applied Econometrics conference, 25-28 June, Nicosia, Cyprus.

Magnusson, L. (2010): "Inference in limited dependent variable models robust to weak identification," *The Econometrics Journal*, 13, S56–S79.

Montiel Olea, J. L. and C. Pflueger (2013): "A Robust Test for Weak Instruments," *Journal of Business & Economic Statistics*, 31, 358–369.

Nunn, N. and N. Qian (2014): "US Food Aid and Civil Conflict," *American Economic Review*, 104, 1630–1666.

Rivers, D. and Q. H. Vuong (1988): "Limited information estimators and exogeneity tests for simultaneous probit models," *Journal of Econometrics*, 39, 347–366.

Stock, J. H. and M. Yogo (2005): "Testing for Weak Instruments in Linear IV Regression," in *Identification and Inference for Econometric Models*, ed. by D. W. K. Andrews and J. H. Stock, Cambridge University Press, 80–108.

Wooldridge, J. M. (2010): *Econometric Analysis of Cross Section and Panel Data*, The MIT Press, second edition ed.

Appendix

Proof. Proposition 1

$$
\begin{aligned}
\delta_{1,OLS} &= \int_{-\infty}^{\infty} xF_U\left(\beta_0+\beta_1 x\right) f_X\left(x\right) dx/\sigma_x^2 \\
&= \frac{1}{\sqrt{2\pi\sigma_x^2}}\int_{-\infty}^{\infty} F_U\left(\beta_0+\beta_1 x\right)\frac{x}{\sigma_x^2}\exp\left(-\frac{x^2}{2\sigma_x^2}\right)dx \\
&= -\frac{1}{\sqrt{2\pi\sigma_x^2}}\int_{-\infty}^{\infty} F_U\left(\beta_0+\beta_1 x\right)\frac{d}{dx}\left(\exp\left(-\frac{x^2}{2\sigma_x^2}\right)\right)dx \\
&= -f_X\left(x\right)F_U\left(\beta_0+\beta_1 x\right)\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} f_X\left(x\right)\frac{d}{dx}\left(F_U\left(\beta_0+\beta_1 x\right)\right)dx \\
&= \beta_1\int_{-\infty}^{\infty} f_U\left(\beta_0+\beta_1 x\right) f_X\left(x\right)dx \\
&= \beta_1 E_X\left[f_U\left(\beta_0+\beta_1 X\right)\right] \\
&= \eta_x.
\end{aligned}
$$

■

Proof. Lemma 1

A standard integration result is

$$\int_{-\infty}^{\infty}\exp\left(-\left(bx+ax^2\right)\right)=\sqrt{\frac{\pi}{a}}\exp\left(\frac{b^2}{4a}\right),$$

with $a>0$. The result then follows as

$$E_X\left[\phi\left(\beta_0+\beta_1 X\right)\right] = \frac{1}{2\pi\sigma_x}\int_{-\infty}^{\infty}\exp\left(-\frac{\left(\beta_0+\beta_1 x\right)^2}{2}\right)\exp\left(-\frac{x^2}{2\sigma_x^2}\right)$$

$$
\begin{aligned}
&= \frac{1}{2\pi\sigma_x}\exp\left(-\frac{\beta_0^2}{2}\right)\int_{-\infty}^{\infty}\exp\left(-\left(\beta_0\beta_1 x + \left(\frac{1+\beta_1^2\sigma_x^2}{2\sigma_x^2}\right)x^2\right)\right) \\
&= \frac{1}{2\pi\sigma_x}\sqrt{\frac{\pi}{\frac{1+\beta_1^2\sigma_x^2}{2\sigma_x^2}}}\exp\left(-\frac{\beta_0^2}{2}\right)\exp\left(\frac{(\beta_0\beta_1)^2}{4\left(\frac{1+\beta_1^2\sigma_x^2}{2\sigma_x^2}\right)}\right) \\
&= \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{1+\beta_1^2\sigma_x^2}}\exp\left(-\frac{\beta_0^2}{2\left(1+\beta_1^2\sigma_x^2\right)}\right) = \frac{1}{\sqrt{1+\beta_1^2\sigma_x^2}}\phi\left(\frac{\beta_0}{\sqrt{1+\beta_1^2\sigma_x^2}}\right).
\end{aligned}
$$

■

Proof. Proposition 3

As the distribution of U is symmetric, we write the model as

$$
\begin{aligned}
Y &= 1\left(\beta_0 + \beta_1 X + U \geq 0\right) \\
&= 1\left(\beta_0 + \beta_1\gamma_1 Z + \left(U + \beta_1 V\right) \geq 0\right).
\end{aligned}
$$

It follows that

$$
E\left[Y|Z\right] = \Phi\left(\frac{\beta_0 + \beta_1\gamma_1 Z}{\sqrt{1+\beta_1^2\sigma_v^2 + 2\beta_1\sigma_{uv}}}\right) = \Phi\left(\beta_0^* + \beta_1^*\gamma_1 Z\right),
$$

with

$$
\beta_0^* = \frac{\beta_0}{\sqrt{1+\beta_1^2\sigma_v^2 + 2\beta_1\sigma_{uv}}};\ \beta_1^* = \frac{\beta_1}{\sqrt{1+\beta_1^2\sigma_v^2 + 2\beta_1\sigma_{uv}}} \tag{20}
$$

Therefore, from the proof of Proposition 1 and Lemma 1 it follows that

$$
\begin{aligned}
\delta_{1,2sls} &= \frac{E_Z\left[Z\Phi\left(\beta_0^* + \beta_1^*\gamma_1 Z\right)\right]}{\gamma_1\sigma_z^2} \\
&= \beta_1^* E_Z\left[\phi\left(\beta_0^* + \beta_1^*\gamma_1 Z\right)\right] \\
&= \frac{\beta_1^*}{\sqrt{1+\beta_1^*\gamma_1^2\sigma_z^2}}\phi\left(\frac{\beta_0^*}{\sqrt{1+\beta_1^{*2}\gamma_1^2\sigma_z^2}}\right).
\end{aligned}
$$

Then the result follows, as

$$
\begin{aligned}
\frac{\beta_1^*}{\sqrt{1+\beta_1^{*2}\gamma_1^2\sigma_z^2}} &= \frac{\beta_1}{\sqrt{1+\beta_1^2\sigma_v^2 + 2\beta_1\sigma_{uv} + \gamma_1^2\beta_1^2\sigma_z^2}} = \frac{\beta_1}{\sqrt{1+\beta_1^2\left(\gamma_1^2\sigma_z^2 + \sigma_v^2\right) + 2\beta_1\sigma_{uv}}} \\
&= \frac{\beta_1}{\sqrt{1+\beta_1^2\sigma_x^2 + 2\beta_1\sigma_{uv}}}
\end{aligned}
$$

and

$$
\frac{\beta_0^*}{\sqrt{1+\beta_1^{*2}\gamma_1^2\sigma_z^2}} = \frac{\beta_0}{\sqrt{1+\beta_1^2\sigma_x^2 + 2\beta_1\sigma_{uv}}}.
$$

■

Proof. Proposition 4

We can write

$$\begin{aligned} E_{X,V}\left[\phi\left(\beta_{0\rho}+\beta_{1\rho}X+\omega_\rho V\right)\right] &= E_{Z,V}\left[\phi\left(\beta_{0\rho}+\beta_{1\rho}\gamma_1 Z+\left(\beta_{1\rho}+\omega_\rho\right)V\right)\right] \\ &= E_Z\left[E_{V|Z}\left[\phi\left(\beta_{0\rho}+\beta_{1\rho}\gamma_1 Z+\left(\beta_{1\rho}+\omega_\rho\right)V\right)|Z\right]\right] \end{aligned}$$

From Lemma 1 it follows that

$$E_{V|Z}\left[\phi\left(\beta_{0\rho}+\beta_{1\rho}\gamma_1 Z+\left(\beta_{1\rho}+\omega_\rho\right)V\right)|Z\right]=\frac{1}{\sqrt{1+\left(\beta_{1\rho}+\omega_\rho\right)^2\sigma_v^2}}\phi\left(\frac{\beta_{0\rho}+\beta_{1\rho}\gamma_1 Z}{\sqrt{1+\left(\beta_{1\rho}+\omega_\rho\right)^2\sigma_v^2}}\right).$$

Further,

$$\begin{aligned} \left(\beta_{\rho 1}+\omega_\rho\right)^2\sigma_v^2 &= \left(\frac{\beta_1+\omega}{\sqrt{1-\rho^2}}\right)^2\sigma_v^2=\frac{\beta_1^2+2\beta_1\omega+\omega^2}{1-\rho^2}\sigma_v^2 \\ &= \frac{\beta_1^2\sigma_v^2+2\beta_1\sigma_{uv}+\rho^2}{1-\rho^2}, \end{aligned}$$

and so

$$1+\left(\beta_{1\rho}+\omega_\rho\right)^2\sigma_v^2=\frac{1+\beta_1^2\sigma_v^2+2\beta_1\sigma_{uv}}{1-\rho^2}.$$

Therefore,

$$\begin{aligned} &\beta_{1\rho}E_{V|Z}\left[\phi\left(\beta_{0\rho}+\beta_{1\rho}\gamma_1 Z+\left(\beta_{1\rho}+\omega_\rho\right)V\right)|Z\right] \\ =\ &\frac{\beta_1}{\sqrt{1+\beta_1^2\sigma_v^2+2\beta_1\sigma_{uv}}}\phi\left(\frac{\beta_0+\beta_1\gamma_1 Z}{\sqrt{1+\beta_1^2\sigma_v^2+2\beta_1\sigma_{uv}}}\right) \end{aligned}$$

and hence

$$\alpha=\frac{\beta_1^*}{\gamma_1}E_Z\left[\phi\left(\beta_0^*+\beta_1^* Z\right)\right]=\delta_{1,2sls},$$

with β_0^* and β_1^* as defined in (20). ■