Marginal Effects for Probit and Tobit with Endogeneity

Kirill S. Evdokimov* Ilze Kalnina† Andrei Zeleneev‡

This version: August 2024

Abstract

When evaluating partial effects, it is important to distinguish between structural endogeneity and measurement errors. In contrast to linear models, these two sources of endogeneity affect partial effects differently in nonlinear models. We study this issue focusing on the Instrumental Variable (IV) Probit and Tobit models. We show that even when a valid IV is available, failing to differentiate between the two types of endogeneity can lead to either under- or over-estimation of the partial effects. We develop simple estimators of the bounds on the partial effects and provide easy to implement confidence intervals that correctly account for both types of endogeneity. We illustrate the methods in a Monte Carlo simulation and an empirical application.

Keywords. (Average) Partial Effects, Instrumental Variable, Control Variable, Errors-in-Variables, Counterfactuals.

*Universitat Pompeu Fabra and Barcelona School of Economics: kirill.evdokimov@upf.edu.
†North Carolina State University: ikalnin@ncsu.edu.
‡University College London: a.zeleneev@ucl.ac.uk.

1 Introduction

Probit and Tobit are some of the most popular nonlinear models in applied economics. When a covariate is endogenous, IV-Probit and IV-Tobit models can be used for instrumental variable (IV) estimation of the coefficients (Smith and Blundell, 1986, Rivers and Vuong, 1988).[1]

A covariate can be endogenous for two reasons. First, the covariate can be correlated with the individual's unobserved characteristics (unobserved heterogeneity). Second, mismeasurement of the covariate also results in endogeneity (Errors-in-Variables, EiV). We will refer to these two types of endogeneity as the structural endogeneity and the EiV. In many empirical settings both sources of endogeneity need to be addressed simultaneously.

The goal of this paper is to characterize the partial effects in these classic models allowing for both types of endogeneity, and to emphasize the importance of distinguishing between the two types. We provide the expressions for the partial effects and average partial effects that correctly account for the two kinds of endogeneity. Although the two sources of endogeneity cannot be precisely distinguished using the observed data, we use the constraints of the model to obtain bounds on the amounts of endogeneity that can be attributed to each source. This allows us to characterize sharp bounds on the true partial effects and average partial effects, allowing for both types of endogeneity. We also provide simple estimators of these bounds and corresponding valid confidence intervals that are easy to calculate.

The primary objects of interest of this paper are the partial effects of the covariates, rather than the regression coefficients (the coefficients on the covariates). Denote by X_i^* the potentially endogenous covariate, and by X_i its mismeasured observed version. Estimation of the coefficients on X_i^* and other covariates is a simpler task than estimation of the partial effects of X_i^*. In particular, in the IV-Probit, IV-Tobit, and related models, to identify and estimate these coefficients it is sufficient to simply consider X_i as the endogenous regressor of interest without needing to distinguish between the types of endogeneity.[2]

In nonlinear models, the need to differentiate between the two kinds of endogeneity arises because structural endogeneity and EiV play different roles. In particular, partial effects of covariates are averaged with respect to the distribution of the individual unobserved heterogeneity. On the other hand, one aims to remove the impact of the measurement errors, since they are not properties of individuals but a deficiency in the measurement process.

[1] For example, in Stata, these estimators are *ivprobit* and *ivtobit*.

[2] In particular, Smith and Blundell (1986) and Rivers and Vuong (1988) simply consider X_i as endogenous. Similarly, in a recent paper, Chesher, Kim, and Rosen (2023) provide a sharp identified set for the coefficients on the covariates in a Tobit model with endogeneity under weak assumptions. The approaches of these papers implicitly allow for mismeasured covariates, as long as the focus is only on the regression coefficients.

The textbook treatment of the problem often focuses only on the first type of endogeneity, treating endogeneity as purely structural. When X_i^* is mismeasured, the partial effect of X_i^* in nonlinear models differs from the effect of X_i one would calculate using the standard formulas that assume the endogeneity is purely structural.

To identify the partial effects of X_i^* and other covariates one needs to identify the distribution of the true unobserved heterogeneity not contaminated by the measurement error. It turns out that this distribution is only partially identified. Thus, even though the IV-Probit and IV-Tobit methods consistently estimate the coefficients on all regressors regardless of the sources of endogeneity, the effects of the covariates on the outcomes are only partially identified. The width of the identified set depends on how hard it is to disentangle structural endogeneity and EiV for the data at hand. Importantly, we find that naively ignoring the distinction between the two types of endogeneity can result in both under- and over-estimation of the magnitude of the partial effects by these IV estimators.[3]

IV-Probit and IV-Tobit can be interpreted as control variable estimators. Partial effects in general control variable models were considered by Blundell and Powell (2003), Chesher (2003), Imbens and Newey (2009), and Wooldridge (2005, 2015), among others. These control variable methods focus on structural endogeneity exclusively but do not consider EiV. The problems of estimation with structural endogeneity or measurement errors are studied by two large but (mostly) distinct literatures in econometrics, see, e.g., Matzkin (2013) and Schennach (2020) for reviews. In nonlinear models, accounting for both types of endogeneity is challenging, see, e.g., Schennach (2022). Exceptions include Adusumilli and Otsu (2018); Song, Schennach, and White (2015); Schennach, White, and Chalak (2012); Hahn and Ridder (2017). These papers obtain point identification results when the distribution of the measurement error is either known or can be recovered from repeated measurements using the lemma of Kotlarski (1967). Such datasets, however, are relatively rare.

Control variable methods in nonlinear models typically require the endogenous variable to be continuously distributed (e.g., see Imbens and Newey, 2009, for a discussion). This limitation also applies to our framework: the mismeasured endogenous variable X_i is assumed to be continuously distributed. Other covariates can be discrete. Note that properly accounting for both types of endogeneity of X_i is essential for characterizing ceteris paribus effects of all covariates, including the discrete ones.

The advantage of gaussian nonlinear models is their simplicity and transparency, which makes them a convenient starting point in an empirical analysis. Our approach in particular provides the researchers with a simple way to gauge the importance of properly accounting

[3] Wooldridge (2010), page 586, alludes to the potential importance of the sources of endogeneity for the partial effects in IV-Probit, but does not elaborate.

for the two types of endogeneity, which is essential given the ubiquity of both in economic applications. In addition, for the settings where relaxing gaussianity is important, we develop an extension of our approach that allows both the first stage unobservables and the measurement errors to be non-gaussian.

The rest of the paper is organized as follows. The analysis of partial effects in the Probit and Tobit models is virtually identical, thus we first focus on Tobit in Sections 2-3. Section 4 then considers the Probit model. Section 5 extends the analysis to cover the average partial effects and other counterfactuals. Section 6 provides some Monte Carlo simulation results. Section 7 presents an empirical application. Section 8 relaxes gaussianity assumptions.

2 The Model

The Tobit model is often used for estimation of economic models with a "corner solution," i.e., models where the outcome variable Y_i is forced to be non-negative. The examples of such dependent variables Y_i include the amounts of charitable contributions, hours worked, or monthly consumption of cigarettes.

First, consider the standard Tobit model with exogenous covariates and without EiV:

$$Y_i = m\left(\theta_{01} X_i^* + \theta_{02}' W_i + U_i^*\right), \quad \text{where } m\left(s\right) = \max(s, 0), \tag{2.1}$$

the individual unobserved heterogeneity U_i^* has a normal distribution $N\left(0, \sigma_{U^*}^2\right)$ and is independent from the covariates X_i^* and W_i. We use the asterisk to denote variables that will be affected by the EiV, as we explain in detail below.

We collect the covariates in a vector $H_i^* = (X_i^*, W_i')'$, so (2.1) can be written as

$$Y_i = m\left(\theta_0' H_i^* + U_i^*\right), \qquad H_i^* = \left(X_i^*, W_i'\right)', \quad \theta_0 = \left(\theta_{01}, \theta_{02}'\right)'.$$

We denote the standard normal cumulative distribution and density functions by Φ and ϕ.[4]

In the Tobit model, one is usually interested in the partial effects (marginal effects) of covariates H_i^* on $\mathrm{E}\left(Y_i | H_i^*\right)$ and $\mathrm{P}\left(Y_i > 0 | H_i^*\right)$. For concreteness we consider partial effects of the continuously distributed covariates.

The partial effect of the j^{th} covariate on the mean $\mathrm{E}\left(Y_i | H_i^* = h\right)$ at a given h is

$$PE_j^{\text{Tob}}\left(h\right) = \frac{\partial}{\partial h_j} \int m\left(\theta_0' h + u\right) f_{U^*}\left(u\right) du = \Phi\left(\frac{\theta_0' h}{\sigma_{U^*}}\right) \theta_{0j}. \tag{2.2}$$

[4] Most of the analysis in Sections 2 and 3 equally applies to the Probit model. For simplicity of exposition, we focus on the Tobit model for the moment, and then discuss Probit in Section 4.

The partial effect of the j^{th} covariate on the probability $P\left(Y_i > 0 | H_i^* = h\right)$ is

$$PE_j^{\text{Pr}}\left(h\right) = \frac{\partial}{\partial h_j} \int 1\left\{\theta_0' h + u > 0\right\} f_{U^*}\left(u\right) du = \phi\left(\frac{\theta_0' h}{\sigma_{U^*}}\right) \frac{\theta_{0j}}{\sigma_{U^*}}. \tag{2.3}$$

These formulas for the PE_j are standard, see, e.g., Wooldridge (2010), for detailed calculations. Most often one considers the partial effects at the means of the covariates $h = E\left[H_i^*\right]$ or partial effects averaged with respect to the distribution of H_i^*.

When X_i^* is correlated with U_i^* and we observe data (Y_i, X_i^*, W_i, Z_i), the IV-Tobit model can be estimated using instrumental variables Z_i, as proposed by Smith and Blundell (1986), Newey (1987), and Rivers and Vuong (1988). Assume that

$$\begin{aligned} Y_i &= m\left(\theta_{01} X_i^* + \theta_{02}' W_i + U_i^*\right), \quad m\left(s\right) = \max(s, 0), & (2.4) \\ X_i^* &= \pi_{01}' Z_i + \pi_{02}' W_i + V_i^*, \quad \pi_{01}' \neq 0, & (2.5) \end{aligned}$$

where V_i^* is a normal random variable, possibly correlated with U_i^*,

$$\begin{pmatrix} U_i^* \\ V_i^* \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{U^*}^2 & \sigma_{U^*V^*} \\ \sigma_{U^*V^*} & \sigma_{V^*}^2 \end{pmatrix}\right), \tag{2.6}$$

and (U_i^*, V_i^*) is independent from (Z_i, W_i). In this model X_i^* is continuously distributed.

The IV-Tobit model in (2.4)-(2.6) can be estimated using a random sample of (Y_i, X_i^*, W_i, Z_i) in two steps, see, e.g., Wooldridge (2010). First, one estimates V_i^* in equation (2.5) by the residuals $\hat{V}_i^*$ in the regression of X_i^* on (W_i, Z_i). Note that we can write $U_i^* = e_i^* + \theta_{V^*} V_i^*$, where $\theta_{V^*} \equiv \sigma_{V^*U^*}/\sigma_{V^*}^2$, and e_i^* is independent of Z_i, W_i, and V_i^* (and hence of X_i^*). Then, one estimates the standard Tobit model

$$Y_i = m\left(\theta_{01} X_i^* + \theta_{02}' W_i + \theta_{V^*} V_i^* + e_i^*\right),$$

where V_i^* are replaced by their estimates $\hat{V}_i^*$. (Alternatively, the two steps can be combined and all of the parameters can be estimated simultaneously by the Maximum Likelihood Estimator.) The reason this approach works is that equation (2.5) creates a control variable V_i^*, and the inclusion of V_i^* in the above equation makes X_i^* exogenous.

To estimate the partial effects, one would plug the estimates $\widehat{\theta}$ and $\widehat{\sigma}_{U^*}^2$ into equations (2.2)-(2.3) in place of θ_0 and $\sigma_{U^*}^2$.

So far we were assuming that the data has no measurement errors. We now allow X_i^* to be mismeasured, i.e., that instead of X_i^* we observe its noisy measurement X_i:

$$X_i = X_i^* + \varepsilon_i, \qquad \varepsilon_i \sim N\left(0, \sigma_\varepsilon^2\right). \tag{2.7}$$

We assume that $\varepsilon_i \perp (U_i^*, V_i^*, W_i, Z_i)$, i.e., the measurement error is classical. The normality assumption simplifies the analysis but it is not crucial. We relax it in Section 8.

Note that the researcher's object of interest has not changed: the goal is to estimate the partial effects defined in equations (2.2)-(2.3). The structural endogeneity and measurement errors are difficulties that an estimation procedure needs to overcome. In particular, note that we are interested in estimation of the effect of X_i^* and *not* in the effect of the error-laden X_i.[5]

3 Analysis of the Model

First, we use the model in equations (2.4)-(2.7) to obtain the model in terms of the observable X_i. Since $X_i^* = X_i - \varepsilon_i$, we can rewrite (2.4) as

$$\begin{aligned} Y_i &= m\left(\theta_{01} X_i^* + \theta_{02}' W_i + U_i^*\right) = m\left(\theta_{01} X_i + \theta_{02}' W_i - \theta_{01}\varepsilon_i + U_i^*\right) \\ &= m\left(\theta_{01} X_i + \theta_{02}' W_i + U_i\right), \end{aligned}$$

where $U_i \equiv U_i^* - \theta_{01}\varepsilon_i$. Let $V_i \equiv V_i^* + \varepsilon_i$. The model in equations (2.4)-(2.7) can be written as

$$Y_i = m\left(\theta_{01} X_i + \theta_{02}' W_i + U_i\right), \tag{3.8}$$

$$X_i = \pi_{01}' Z_i + \pi_{02}' W_i + V_i, \tag{3.9}$$

$$\begin{pmatrix} U_i \\ V_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_U^2 & \sigma_{UV} \\ \sigma_{UV} & \sigma_V^2 \end{pmatrix}\right). \tag{3.10}$$

The definitions of U_i and V_i imply that

$$\sigma_U^2 = \sigma_{U^*}^2 + \theta_{01}^2 \sigma_\varepsilon^2, \quad \sigma_V^2 = \sigma_{V^*}^2 + \sigma_\varepsilon^2, \quad \sigma_{UV} = \sigma_{U^*V^*} - \theta_{01}\sigma_\varepsilon^2. \tag{3.11}$$

Note that variables X_i, U_i, V_i are the analogs of the true variables X_i^*, U_i^*, V_i^* that arise due to the measurement errors ε_i. In the absence of measurement errors, i.e., when $\varepsilon_i = 0$,

[5] This is similar to the linear regression settings, where one would be interested in the effect of X_i^* on Y_i. The slope coefficient in the OLS regression of Y_i on X_i is not the object of interest because it is subject to the attenuation bias due to the EiV (and also possibly due to the endogeneity of X_i^*).

we have $X_i = X_i^*$, $U_i = U_i^*$, $V_i = V_i^*$.

The model in equations (3.8)-(3.10) can be estimated by MLE or using the control variable two-step approach described earlier. Specifically, both approaches will consistently estimate parameters θ_0 and the covariance matrix of the unobservables in equation (3.10), i.e., σ_U^2, σ_{UV}, and σ_V^2.

Note that because the model is nonlinear, the marginal effects defined in equations (2.2)-(2.3) depend not only on θ_0 but also on $\sigma_{U^*}^2$. Thus, even though the available data (Y_i, X_i, W_i, Z_i) allows immediately estimating θ_0, we cannot obtain the marginal effects because we do not know $\sigma_{U^*}^2$. Naively using an estimate of σ_U^2 in place of $\sigma_{U^*}^2$ would lead to a biased estimate of the partial effects, since $\sigma_U^2 \geq \sigma_{U^*}^2$, as implied by equation (3.11).

The problem with identifying $\sigma_{U^*}^2$ is that the data only allows identification of the 3 parameters σ_U^2, σ_{UV}, and σ_V^2. However, the distribution of the true $(U_i^*, V_i^*, \varepsilon_i)$ is governed by 4 parameters: $\sigma_{U^*}^2$, $\sigma_{U^*V^*}$, $\sigma_{V^*}^2$, and σ_ε^2. Thus, one cannot uniquely determine these 4 parameters from the 3 equations (3.11). In other words, models with different values of σ_ε^2 are observationally equivalent: they correspond to identical distributions of the observables (Y_i, X_i, W_i, Z_i) even though they imply different values of true $\sigma_{U^*}^2$. Thus, one cannot uniquely determine (i.e., point-identify) $\sigma_{U^*}^2$ from the data (Y_i, X_i, W_i, Z_i). Correspondingly, one cannot point-identify the partial effects, which depend on $\sigma_{U^*}^2$.

Equations (3.11) provide restrictions on $\sigma_{U^*}^2$, which we will use to provide bounds on the possible values of true $\sigma_{U^*}^2$, and hence on the values of the partial effects.

Bounds on $\sigma_{U^*}^2$ From equations (3.11) the upper bound on $\sigma_{U^*}^2$ is $\sigma_{U^*}^2 \leq \sigma_U^2$. We now obtain the lower bound on $\sigma_{U^*}^2$. In particular, we look to find the smallest $\sigma_{U^*}^2$ that satisfies equations (3.11), Cauchy-Schwarz inequality $\sigma_{U^*V^*}^2 \leq \sigma_{U^*}^2\sigma_{V^*}^2$, and the non-negativity constraints $\sigma_{U^*}^2 \geq 0$, $\sigma_{V^*}^2 \geq 0$, and $\sigma_\varepsilon^2 \geq 0$. Let $\rho_{UV} = \operatorname{corr}(U_i, V_i)$.

Proposition 3.1 *Suppose $|\rho_{UV}| < 1$ in model (3.8)-(3.10). Then the sharp identified set for $\sigma_{U^*}^2$ is given by*

$$\sigma_{U^*}^2 \in \left[\underline{\sigma}_{U^*}^2, \sigma_U^2\right],$$

where

$$\underline{\sigma}_{U^*}^2 \equiv \max\left\{\frac{\left(\theta_{01}\sigma_{UV} + \sigma_U^2\right)^2}{\sigma_V^2\theta_{01}^2 + 2\sigma_{UV}\theta_{01} + \sigma_U^2},\ \sigma_U^2 - \theta_{01}^2\sigma_V^2\right\}. \tag{3.12}$$

Proposition 3.1 provides the bounds in terms of the quantities that can be estimated using the data (Y_i, X_i, W_i, Z_i). Condition $|\rho_{UV}| < 1$ guarantees that the denominator in the fraction above is positive. The proof of Proposition 3.1 also provides bounds on $\sigma_{U^*V^*}$ and σ_ε^2.

Correct Partial Effects We now use the bounds on $\sigma^2_{U^*}$ from Proposition 3.1 to obtain the bounds on the partial effects, in terms of the parameters that can be recovered from data. For simplicity and concreteness of exposition, we first consider partial effects evaluated at some fixed values of covariates. We consider average partial effects and other counterfactuals of interest in Section 5.

For a given $\sigma^2_{U^*}$, the partial effects for the j^{th} covariate are defined as in equations (2.2)-(2.3),

$$PE_j^{\text{Tob}}\left(h,\sigma^2_{U^*}\right)=\Phi\left(\frac{\theta_0'h}{\sigma_{U^*}}\right)\theta_{0j}\quad\text{and}\quad PE_j^{\text{Pr}}\left(h,\sigma^2_{U^*}\right)=\phi\left(\frac{\theta_0'h}{\sigma_{U^*}}\right)\frac{\theta_{0j}}{\sigma_{U^*}}. \tag{3.13}$$

The lower and upper bounds for partial effects for the j^{th} covariate, $PE_j\left(h\right)$, are computed as

$$\min_{\sigma^2_{U^*}\in\left[\underline{\sigma}^2_{U^*},\sigma^2_U\right]} PE_j\left(h,\sigma^2_{U^*}\right)\quad\text{and}\quad \max_{\sigma^2_{U^*}\in\left[\underline{\sigma}^2_{U^*},\sigma^2_U\right]} PE_j\left(h,\sigma^2_{U^*}\right). \tag{3.14}$$

Function $PE_j^{\text{Tob}}\left(h,\sigma^2_{U^*}\right)$ in (3.13) is a monotone function of $\sigma^2_{U^*}$, so the minimum and maximum in equation (3.14) are achieved on the boundaries of interval $[\underline{\sigma}^2_{U^*},\sigma^2_U]$.

Function $PE_j^{\text{Pr}}\left(h,\sigma^2_{U^*}\right)$ in equation (3.13) is not monotone in $\sigma^2_{U^*}$, but the bounds in equation (3.14) for $PE_j^{\text{Pr}}\left(h\right)$ can also be simplified. The minimum and maximum over $\sigma^2_{U^*}\in[\underline{\sigma}^2_{U^*},\sigma^2_U]$ can be attained only at $\sigma^2_{U^*}=\underline{\sigma}^2_{U^*}$, at $\sigma^2_{U^*}=\sigma^2_U$, and, when $\left(\theta_0'h\right)^2\in[\underline{\sigma}^2_{U^*},\sigma^2_U]$, at $\sigma^2_{U^*}=\left(\theta_0'h\right)^2$. Thus, one only needs to evaluate $PE_j^{\text{Pr}}\left(h,\sigma^2_{U^*}\right)$ at these 2 or 3 points to calculate the minimum and maximum in equation (3.14).

Since $\sigma_U\geq\sigma_{U^*}$, naively using σ_U instead of σ_{U^*} when calculating $PE_j^{\text{Tob}}\left(h\right)$, would lead to attenuation bias when $\theta_0'h>0$, but would bias $PE_j^{\text{Tob}}\left(h\right)$ away from zero when $\theta_0'h<0$, i.e., the EiV would make naive $PE_j^{\text{Tob}}\left(h,\sigma^2_U\right)$ over-estimate the partial effects $PE_j^{\text{Tob}}\left(h\right)$ in the latter case. Likewise, for the probability, naively using $PE_j^{\text{Pr}}\left(h,\sigma^2_U\right)$ can both under- and over-estimate the true partial effect $PE_j^{\text{Pr}}\left(h\right)$.

Estimation Using the standard two-step or MLE approaches described in Section 2, one obtains the estimates of θ_0, σ^2_U, σ^2_V, and σ_{UV} (and of their variance-covariance matrix for inference). Then, from equation (3.12) one obtains the estimate of $\underline{\sigma}^2_{U^*}$.

For a given value of $\sigma^2_{U^*}$, the estimated partial effects would be

$$\widehat{PE}_j^{\text{Tob}}\left(h,\sigma^2_{U^*}\right)=\Phi\left(\frac{\widehat{\theta}'h}{\sigma_{U^*}}\right)\widehat{\theta}_j\quad\text{and}\quad \widehat{PE}_j^{\text{Pr}}\left(h,\sigma^2_{U^*}\right)=\phi\left(\frac{\widehat{\theta}'h}{\sigma_{U^*}}\right)\frac{\widehat{\theta}_j}{\sigma_{U^*}}. \tag{3.15}$$

Then, the estimated bounds on $PE_j(h)$ are

$$\min_{v\in\left[\underline{\widehat{\sigma}}^2_{U^*},\widehat{\sigma}^2_U\right]} \widehat{PE}_j(h,v) \quad \text{and} \quad \max_{v\in\left[\underline{\widehat{\sigma}}^2_{U^*},\widehat{\sigma}^2_U\right]} \widehat{PE}_j(h,v), \tag{3.16}$$

where the minimum and maximum are easily computed using univariate numerical optimization. For the partial effects in equation (3.15), these extrema can also be computed as described under equation (3.14).

For example, one often considers the partial effects at the mean values of covariates taking $h = (\overline{X}, \overline{W}')'$, where $\overline{X}$ and $\overline{W}$ are the sample averages. Note that $\mathrm{E}[X_i] = \mathrm{E}[X_i^*]$.

Inference To provide a simple method for inference about the partial effects, we adopt a Bonferroni approach (e.g., McCloskey, 2017). This approach allows us to avoid computational challenges that often arise in the context of subvector inference in partially identified models. The construction of a $1-\alpha$ confidence interval for a partial effect $PE_j(h)$ proceeds in two steps:

1. Pick $\alpha_1 \in (0,\alpha)$ and construct $CI^{\sigma^2_{U^*}}_{1-\alpha_1}$, a $1-\alpha_1$ confidence interval for $\sigma^2_{U^*}$, based on the bounds provided in Proposition 3.1.
2. Construct a $1-\alpha$ confidence interval for $PE_j(h)$ as the union $CI^{PE_j(h)}_{1-\alpha} = \bigcup_{\sigma^2_{U^*}\in CI^{\sigma^2_{U^*}}_{1-\alpha_1}} CI^{PE_j(h)}_{1-(\alpha-\alpha_1)}(\sigma^2_{U^*})$, where $CI^{PE_j(h)}_{1-(\alpha-\alpha_1)}(\sigma^2_{U^*})$ is a standard $1-(\alpha-\alpha_1)$ confidence interval for $PE_j(h)$ based on $\widehat{PE}_j(h,\sigma^2_{U^*})$ in equation (3.15) for a given $\sigma^2_{U^*}$.

We now provide the implementation details for each step.

Step 1. The confidence interval for $\sigma^2_{U^*}$ is constructed based on the bounds given in Proposition 3.1. As the upper bound, we take $\hat{\sigma}^2_U + z_{1-\alpha_1/2} \times s_{\hat{\sigma}^2_U}$, where $s_{\hat{\sigma}^2_U}$ is the standard error of $\hat{\sigma}^2_U$, and $z_{1-\alpha_1/2}$ is the $1-\alpha_1/2$ quantile of the standard normal distribution. The lower bound is based on $\underline{\hat{\sigma}}^2_{U^*} = \max\{\hat{\xi}_1, \hat{\xi}_2\}$, where $\hat{\xi}_1$ and $\hat{\xi}_2$ are the plug-in estimators of the two terms on the right hand side of equation (3.12). Note that $\hat{\xi}_1$ and $\hat{\xi}_2$ are (generally) jointly asymptotically normal and their asymptotic variance-covariance matrix can be computed using the delta method. Then, as the lower bound of $CI^{\sigma^2_{U^*}}_{1-\alpha_1}$, we take $\max\{\hat{\xi}_1 - c_{1-\alpha_1/2} \times s_{\hat{\xi}_1}, \hat{\xi}_2 - c_{1-\alpha_1/2} \times s_{\hat{\xi}_2}\}$. Here $s_{\hat{\xi}_1}$ and $s_{\hat{\xi}_2}$ are the standard errors of $\hat{\xi}_1$ and $\hat{\xi}_2$, and $c_{1-\alpha_1/2}$ is the $1-\alpha_1/2$ quantile of $\max\{\eta_1,\eta_2\}$, where (η_1,η_2) are jointly normal with unit variances and correlation $\hat{\rho}_{\hat{\xi}_1,\hat{\xi}_2}$, and $\hat{\rho}_{\hat{\xi}_1,\hat{\xi}_2}$ is an estimator of the correlation between $\hat{\xi}_1$ and $\hat{\xi}_2$ (e.g., see Romano and Wolf, 2005). By a standard argument, the confidence interval for $\sigma^2_{U^*}$ given by

$$CI^{\sigma^2_{U^*}}_{1-\alpha_1} = \left[\max\left\{\hat{\xi}_1 - c_{1-\alpha_1/2} \times s_{\hat{\xi}_1}, \hat{\xi}_2 - c_{1-\alpha_1/2} \times s_{\hat{\xi}_2}\right\}, \hat{\sigma}^2_U + z_{1-\alpha_1/2} \times s_{\hat{\sigma}^2_U}\right]$$

has asymptotic coverage at least $1-\alpha_1$ for the true $\sigma^2_{U^*}$. In the numerical illustrations we take $\alpha_1 = \alpha/10$.

Step 2. First, the standard $CI^{PE_j(h)}_{1-(\alpha-\alpha_1)}\left(\sigma^2_{U^*}\right)$ is $\left[l^{PE_j(h)}_{1-(\alpha-\alpha_1)}\left(\sigma^2_{U^*}\right), u^{PE_j(h)}_{1-(\alpha-\alpha_1)}\left(\sigma^2_{U^*}\right)\right]$ constructed by adding and subtracting $z_{1-(\alpha-\alpha_1)/2} \times s_{\widehat{PE}_j(h,\sigma^2_{U^*})}$ from $\widehat{PE}_j(h,\sigma^2_{U^*})$. The standard error $s_{\widehat{PE}_j(h,\sigma^2_{U^*})}$ of $\widehat{PE}_j(h,\sigma^2_{U^*})$ can be computed using the delta method. Then we can construct $CI^{PE_j(h)}_{1-\alpha}$ as

$$CI^{PE_j(h)}_{1-\alpha} = \left[\min_{\sigma^2_{U^*} \in CI^{\sigma^2_{U^*}}_{1-\alpha_1}} l^{PE_j(h)}_{1-(\alpha-\alpha_1)}\left(\sigma^2_{U^*}\right), \max_{\sigma^2_{U^*} \in CI^{\sigma^2_{U^*}}_{1-\alpha_1}} u^{PE_j(h)}_{1-(\alpha-\alpha_1)}\left(\sigma^2_{U^*}\right)\right],$$

where the minimum and maximum are easily calculated using univariate numerical optimization over $\sigma^2_{U^*}$. By the standard Bonferroni argument, the confidence interval $CI^{PE_j(h)}_{1-\alpha}$ has asymptotic coverage of at least $1-\alpha$ for the true partial effect $PE_j(h)$.

The constructed confidence interval is asymptotically valid as long as (i) the first step confidence interval $CI^{\sigma^2_{U^*}}_{1-\alpha_1}$ covers the true $\sigma^2_{U^*}$ with probability at least $1-\alpha_1$ asymptotically, and (ii) the delta method applies to $\widehat{PE}_j(h,\sigma^2_{U^*})$ for the true $\sigma^2_{U^*}$. Both conditions are satisfied provided that the true $\sigma^2_{U^*}$ is bounded away from zero. Note that in this case $CI^{\sigma^2_{U^*}}_{1-\alpha_1}$ is valid even if $\theta_{01}\sigma_{UV} + \sigma^2_U$ is equal to (or local to) zero, which implies that $CI^{PE_j(h)}_{1-\alpha}$ is also valid.

4 Probit

IV-Probit model is the same as IV-Tobit except $m(s) = 1\{s > 0\}$ in equation (2.4). Since Probit is a binary outcome model, in equation (3.8) one needs to impose a scale normalization, for example, $\|\theta_0\| = 1$ or $\sigma^2_U = 1$. For both of those normalizations, θ_0 and the other reduced form parameters are point identified, and the sharp bounds for $\sigma^2_{U^*}$ are given by Proposition 3.1 as before.

For Probit, we are interested in the partial effects of covariates on the probability of $Y_i = 1$, which are given by $PE^{\text{Pr}}_j(h)$ in equation (2.3). Similarly to the IV-Tobit model, the IV-Probit model can be estimated by MLE or by the two-step approach identical to the one described in Section 2, except the second step uses the standard Probit estimator in place of the Tobit estimator. Then the bounds on $PE^{\text{Pr}}_j(h)$ are estimated as in equation (3.16). Confidence intervals for $PE^{\text{Pr}}_j(h)$ can be computed exactly as described above.

5 Average Partial Effects and Other Counterfactuals

In addition to the partial effects at a given h, researchers are often interested in the Average Partial Effects

$$APE_j^{\mathrm{Tob}} \equiv \mathrm{E}\left[PE_j^{\mathrm{Tob}}\left(H_i^*\right)\right] \text{ and } APE_j^{\mathrm{Pr}} \equiv \mathrm{E}\left[PE_j^{\mathrm{Pr}}\left(H_i^*\right)\right], \tag{5.17}$$

which are the partial effects $PE_j(h)$ averaged with respect to the distribution of $H_i^* = (X_i^*, W_i')'$. Define

$$APE_j^{\mathrm{Tob}}\left(\sigma_{U^*}^2\right) \equiv \mathrm{E}\left[PE_j^{\mathrm{Tob}}\left(H_i^*, \sigma_{U^*}^2\right)\right] \text{ and } APE_j^{\mathrm{Pr}}\left(\sigma_{U^*}^2\right) \equiv \mathrm{E}\left[PE_j^{\mathrm{Pr}}\left(H_i^*, \sigma_{U^*}^2\right)\right].$$

Note that the distribution of X_i^* is not directly observable due to the EiV. Averaging $PE_j(h)$ with respect to the distribution of the observed $H_i = (X_i, W_i')'$ would result in biased estimators of the APEs. To account for this, in the Appendix we show that these APEs can be calculated as

$$APE_j^{\mathrm{Tob}}\left(\sigma_{U^*}^2\right) = \mathrm{E}\left[\Phi\left(\frac{\theta_{01}\pi_{01}' Z_i + (\theta_{01}\pi_{02} + \theta_{02})' W_i}{\sqrt{2\sigma_{U^*}^2 - \sigma_U^2 + \theta_{01}^2\sigma_V^2}}\right)\right]\theta_{0j}, \tag{5.18}$$

$$APE_j^{\mathrm{Pr}}\left(\sigma_{U^*}^2\right) = \mathrm{E}\left[\phi\left(\frac{\theta_{01}\pi_{01}' Z_i + (\theta_{01}\pi_{02} + \theta_{02})' W_i}{\sqrt{2\sigma_{U^*}^2 - \sigma_U^2 + \theta_{01}^2\sigma_V^2}}\right)\right]\frac{\theta_{0j}}{\sqrt{2\sigma_{U^*}^2 - \sigma_U^2 + \theta_{01}^2\sigma_V^2}}. \tag{5.19}$$

Hence, for any given value of $\sigma_{U^*}^2$, these APEs can be estimated by

$$\widehat{APE}_j^{\mathrm{Tob}}\left(\sigma_{U^*}^2\right) = \frac{1}{n}\sum_{i=1}^{n}\Phi\left(\frac{\hat\theta_1\hat\pi_1' Z_i + \left(\hat\theta_1\hat\pi_2 + \hat\theta_2\right)' W_i}{\sqrt{2\sigma_{U^*}^2 - \hat\sigma_U^2 + \hat\theta_1^2\hat\sigma_V^2}}\right)\hat\theta_j,$$

$$\widehat{APE}_j^{\mathrm{Pr}}\left(\sigma_{U^*}^2\right) = \frac{1}{n}\sum_{i=1}^{n}\phi\left(\frac{\hat\theta_1\hat\pi_1' Z_i + \left(\hat\theta_1\hat\pi_2 + \hat\theta_2\right)' W_i}{\sqrt{2\sigma_{U^*}^2 - \hat\sigma_U^2 + \hat\theta_1^2\hat\sigma_V^2}}\right)\frac{\hat\theta_j}{\sqrt{2\sigma_{U^*}^2 - \hat\sigma_U^2 + \hat\theta_1^2\hat\sigma_V^2}}.$$

Finally, the estimated bounds on the APEs are obtained by finding the minimum and maximum over $\sigma_{U^*}^2 \in \left[\underline{\hat\sigma}_{U^*}^2, \hat\sigma_U^2\right]$. These can be easily computed numerically, since $\widehat{APE}_j(v)$ are smooth functions of a scalar argument v. Our two-step approach to inference also applies to the APEs with a minimal modification. The only difference is that in Step 2 the construction of the standard error $s_{\widehat{APE}_j}\left(\sigma_{U^*}^2\right)$ as usual needs to account for the sampling variability in both the parameter estimators and the data entering the expressions for the APEs directly.

It is also straightforward to apply our analysis to other counterfactuals, including partial effects and APEs of discrete covariates, as well as to the ordered Probit and two-sided Tobit models. Proposition 3.1 and the bounds on $\sigma^2_{U^*}$ remain the same, and hence the estimation and inference procedures remain unchanged, except for different formulas in equations (3.13)-(3.16) corresponding to the counterfactuals of interest.

6 Numerical Illustration

We simulate a Tobit model with endogenous and mismeasured X_i^*, as in equations (2.4)-(2.7), with $W_i = 1$, $Z_i \sim N(0,1)$, $(\theta_{01}, \theta_{02}, \sigma_{V^*}, \sigma_{U^*}, \sigma_\varepsilon, \pi_{01}, \pi_{02}) = (2,1,1,1,1,1,0)$, and $n = 1000$. Figure 1 plots the results for the Partial Effects (PEs) of X_i^* at the population mean values of the covariates.

We consider a range of designs corresponding to the true values of $\rho_{U^*V^*} \in [-0.95, 0.95]$ on the horizontal axis. For each $\rho_{U^*V^*}$, the figure shows the true PE ("true"), the true (population) bounds for the PE obtained using Proposition 3.1 ("true bounds"), as well as the medians over the Monte Carlo replications of the estimated lower and upper bounds on the PE ("LB" and "UB") and the corresponding 95% confidence intervals ("CI") based on the two-step IV-Tobit estimator. The true bounds for the PE are calculated using the point identified parameters θ_0, σ^2_U, σ_{UV}, and σ^2_V, see equations (3.8)-(3.10). For comparison, we also include the results for the PE calculated using the standard naive IV-Tobit estimator ("naive") and the corresponding confidence intervals ("CI naive"). The "naive" estimators of the partial effects are $\widehat{PE}_1^{\text{Tob}}(h, \widehat{\sigma}^2_U)$ and $\widehat{PE}_1^{\text{Pr}}(h, \widehat{\sigma}^2_U)$, i.e., they replace $\sigma^2_{U^*}$ with $\widehat{\sigma}^2_U$ in equation (3.15).

Figure 1 (a) shows the bounds on $PE^{\text{Tob}}_{h_1}(h)$, while Figure 1 (b) considers $PE^{\text{Pr}}_1(h)$. As expected, the true PE is between the lower and upper bounds for all values of $\rho_{U^*V^*}$. By construction, the "naive" IV-Tobit estimator of $PE^{\text{Tob}}_1(h)$ coincides with one of the bounds. In both panels, the "naive" estimates are below the true values for every $\rho_{U^*V^*}$, and the "naive" IV-Tobit confidence intervals do not include the true PE. In this design, the identified set and the confidence intervals for the true partial effects are much wider than those of the "naive" estimator. The relative width of the identified set and the confidence intervals depends on the specific parameter values. In contrast to these simulations, in the example of the next section, the identified set is very narrow and the confidence intervals for the partial effects have width similar to those of the naive estimator.

To gain some intuition about the shape of the bounds in Figure 1, notice that when $\theta_{01} > 0$, measurement error in X_i^* introduces a negative correlation between X_i and U_i. Thus, observing $\rho_{UV} > 0$ is only consistent with $\rho_{U^*V^*} > 0$, but not with $\rho_{U^*V^*} \leq 0$, i.e.,

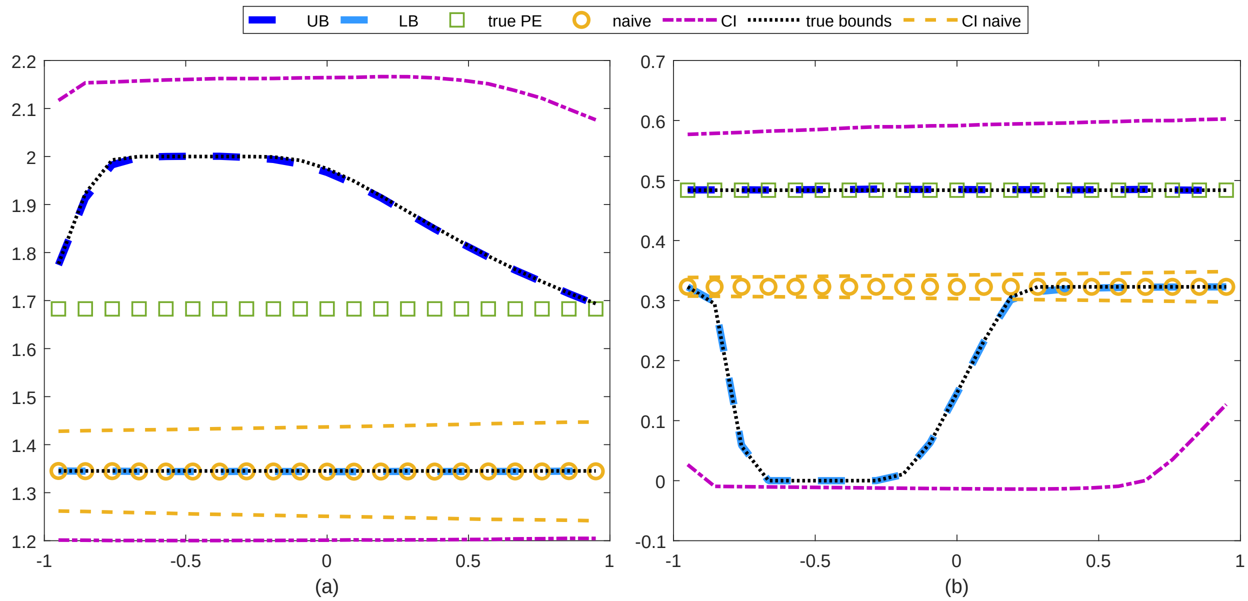

Figure 1: Simulation results for partial effects on: (a) expectations, $PE_{h_1}^{\text{Tob}}(h)$, and (b) probability, $PE_1^{\text{Pr}}(h)$. Values of $\rho_{U^*V^*}$ are on the horizontal axis.

implies positive correlation due to the structural endogeneity. On the other hand, observing $\rho_{UV} < 0$ can be explained both by the effect of EiV combined with $\rho_{U^*V^*} \gtrless 0$, and by $\rho_{U^*V^*} < 0$ without any EiV. Thus, when $\rho_{UV} < 0$ it is harder to disentangle structural endogeneity and EiV. As a consequence, the true and the estimated correct bounds on the PEs in Figure 1 are wider for the negative values of $\rho_{U^*V^*}$.

7 Empirical Application

We illustrate the proposed methods in the classic application estimating the Tobit and Probit models for women's labor force participation. Using NLSY97 we construct an up-to-date dataset similar to the well known dataset of Mroz (1987), who estimates Tobit and related models to explain married women's hours of work. The data contains 1185 women continuously married in 2018, 76% of which report working non-zero hours. For Tobit, the dependent variable is the number of hours worked ($hours$), and for Probit, the dependent variable is working at some point during the year ($1\{hours > 0\}$). In both models, the covariates are age, education, experience, experience squared, nonwife income in thousands ($nwifeinc$), number of children less than six years of age, number of children between 6 and 18 years of age, race-ethnicity indicator variables, and an intercept. Nonwife income could be correlated with the unobserved characteristics (structural endogeneity), and income variables are also known to be frequently mismeasured. Spouse's years of schooling, *speduc*, are used as an instrument for *nwifeinc*. This specification is used in, e.g., Wooldridge (2010).

	Tobit	IV-Tobit	CI for IV-Tobit	[LB, UB] for APE	CI for APE
nwifeinc	-1.40	-4.55	$[-7.78, -1.32]$	$[-4.63, -4.55]$	$[-8.22, -1.12]$
educ	27.6	40.5	$[18.2, 62.9]$	$[40.5, 41.3]$	$[16.6, 66.2]$
exper	171	175	$[124, 226]$	$[175, 178]$	$[119, 236]$
exper2	-0.908	-1.47	$[-3.53, 0.597]$	$[-1.49, -1.47]$	$[-3.71, 0.674]$
age	-105	-90.9	$[-131, -51.1]$	$[-92.5, -90.9]$	$[-135, -49.3]$

Table 1: Tobit. Average Partial Effects on Expectation. $n = 1185$.

	Tobit	IV-Tobit	CI for IV-Tobit	[LB, UB] for APE	CI for APE
nwifeinc	-0.028	-0.094	$[-0.166, -0.023]$	$[-0.094, -0.092]$	$[-0.167, -0.022]$
educ	0.555	0.840	$[0.346, 1.33]$	$[0.822, 0.840]$	$[0.322, 1.35]$
exper	3.43	3.63	$[2.68, 4.59]$	$[3.55, 3.63]$	$[2.63, 4.68]$
exper2	-0.018	-0.030	$[-0.072, 0.012]$	$[-0.030, -0.030]$	$[-0.074, 0.013]$
age	-2.12	-1.88	$[-2.72, -1.05]$	$[-1.88, -1.84]$	$[-2.77, -0.933]$

Table 2: Tobit. Average Partial Effects on Probability. All numbers are multiplied by 100. $n = 1185$.

	Probit	IV-Probit	CI for IV-Probit	[LB, UB] for APE	CI for APE
nwifeinc	-0.033	-0.102	$[-0.211, 0.007]$	$[-0.105, -0.102]$	$[-0.234, 0.014]$
educ	0.321	0.620	$[-0.150, 1.39]$	$[0.620, 0.641]$	$[-0.196, 1.53]$
exper	-0.672	-0.488	$[-1.81, 0.831]$	$[-0.504, -0.488]$	$[-1.99, 0.941]$
exper2	0.208	0.191	$[0.121, 0.260]$	$[0.191, 0.197]$	$[0.119, 0.285]$
age	-2.73	-2.37	$[-3.68, -1.05]$	$[-2.44, -2.37]$	$[-3.94, -0.997]$

Table 3: Probit. Average Partial Effects on Probability. All numbers are multiplied by 100. $n = 1185$.

Tables 1-2 contain the results on average partial effects for Tobit. Table 1 contains the results on average partial effects on expectation, APE^{Tob}, while Table 2 contains average partial effects on probability, APE^{Pr}. In both tables, the first column ("Tobit") provides the average partial effects for different covariates in the standard Tobit MLE where all covariates are assumed to be exogenous. The remaining columns are based on the two-step IV-Tobit estimator, where *speduc* is used to instrument for the endogenous *nwifeinc*. The second column ("IV-Tobit") contains the naive estimators of the average partial effects, followed by the 95% confidence intervals (column "CI for IV-Tobit"). Column "[LB, UB] for APE" provides the proposed estimated bounds for the average partial effects that account for both types of endogeneity. The last column contains the corresponding confidence intervals for the average partial effects.

In both Tables 1 and 2, we observe that the confidence intervals for the correct average partial effects at the mean are only slightly wider than the naive ones of IV-Tobit. In particular, using the correct inference approach does not change any of the conclusions about the effects of the variables being statistically significant.

Table 3 contains the corresponding results for Probit. Again, the confidence intervals for the correct partial effects are not much wider than the naive ones, and as before, using the correct inference approach does not change any of the conclusions about the effects of the variables being statistically significant.

Overall, we find that the proposed valid confidence intervals for the APEs are only slightly wider than those obtained from the naive (and generally invalid) IV-Tobit/IV-Probit estimators. The point estimates of the valid bounds in columns "[LB, UB] for APE" are narrow. We interpret these results as (strong) evidence in favor of practical usefulness our method. Practitioners often worry that methods providing partial identification of parameters of interest achieve robustness at the cost of producing bounds that are too wide to be useful in practice. Our empirical application shows that this concern need not be an issue for our method.

8 Relaxing Distributional Assumptions

We have considered the classic IV-Tobit and IV-Probit settings, which require V_i^* and ε_i to be Gaussian in order for V_i and U_i to be Gaussian, as in Smith and Blundell (1986), Rivers and Vuong (1988), and Wooldridge (2010). However, at least in the model without EiV, the control variable approach does not require V_i^* to be Gaussian. Likewise, we would like to relax the assumption of Gaussianity on ε_i. In this section we propose an approach that weakens the distributional assumptions, while still providing a simple method for computing the identified set for the partial effects of interest. We focus on the IV-Tobit settings.

We model the joint distribution of U_i^* and V_i^* as a mixture of J bivariate normals. Specifically, the joint p.d.f. of U^* and V^* takes the form

$$f_{U^*V^*}(u,v) = \sum_{j=1}^{J} p_{V^*,j}\phi(u,v;\mu_{U^*V^*,j},\Sigma_{U^*V^*,j}),$$

where $p_{V^*,j} > 0$ are the mixing weights, and

$$\mu_{U^*V^*,j} = \begin{pmatrix} 0 \\ \mu_{V^*,j} \end{pmatrix}, \quad \Sigma_{U^*V^*,j} = \begin{pmatrix} \sigma^2_{U^*} & \sigma_{U^*V^*,j} \\ \sigma_{U^*V^*,j} & \sigma^2_{V^*,j} \end{pmatrix},$$

and $\phi(\cdot,\cdot;\mu,\Sigma)$ stands for the p.d.f. of a bivariate normal with mean μ and variance-covariance matrix Σ. Notice that in this parameterization the marginal distribution of U_i^* is $N(0,\sigma^2_{U^*})$ as in the standard Tobit model, whereas the marginal distribution of V_i^* is flexibly modelled as a mixture of J normals. Since $\sigma_{U^*V^*,j}$ can vary over j, this model also allows rich patterns of dependency between U_i^* and V_i^*. Similarly, we model the distribution of ε_i by a mixture of L normals:

$$f_\varepsilon(\varepsilon) = \sum_{\ell=1}^{L} p_{\varepsilon,\ell}\phi(\varepsilon;\mu_{\varepsilon,\ell},\sigma^2_{\varepsilon,\ell}),$$

where $p_{\varepsilon,\ell} > 0$ are the mixing weights, and $\phi(\cdot;\mu,\sigma^2)$ stands for the p.d.f. of a $N(\mu,\sigma^2)$ distribution, and we denote the corresponding c.d.f. by $\Phi(\cdot;\mu,\sigma^2)$.

Since ε_i is independent from (U_i^*,V_i^*), the joint distribution of $(U_i,V_i) = (U_i^* - \theta_{01}\varepsilon_i, V_i^* + \varepsilon_i)$ is a mixture of $J \times L$ bivariate normals, and its p.d.f. is given by

$$f_{UV}(u,v) = \sum_{j=1}^{J}\sum_{\ell=1}^{L} p_{V^*,j}p_{\varepsilon,\ell}\phi(u,v;\mu_{UV,j\ell},\Sigma_{UV,j\ell}), \text{ where}$$

$$\mu_{UV,j\ell} = \begin{pmatrix} \mu_{U,\ell} \\ \mu_{V,j\ell} \end{pmatrix} = \begin{pmatrix} -\theta_{01}\mu_{\varepsilon,\ell} \\ \mu_{V^*,j} + \mu_{\varepsilon,\ell} \end{pmatrix}, \quad \Sigma_{UV,j\ell} = \begin{pmatrix} \sigma^2_{U,j\ell} & \sigma_{UV,j\ell} \\ \sigma_{UV,j\ell} & \sigma^2_{V,j\ell} \end{pmatrix},$$

$$\sigma^2_{U,j\ell} = \sigma^2_{U^*} + \theta^2_{01}\sigma^2_{\varepsilon,\ell}, \quad \sigma^2_{V,j\ell} = \sigma^2_{V^*,j} + \sigma^2_{\varepsilon,\ell}, \quad \sigma_{UV,j\ell} = \sigma_{U^*V^*,j} - \theta_{01}\sigma^2_{\varepsilon,\ell}. \tag{8.20}$$

We impose standard constraints $\sum_{j=1}^{J} p_{V^*,j} = 1$ and $\sum_{\ell=1}^{L} p_{\varepsilon,\ell} = 1$. In addition, we need location normalizations on the distributions of V_i^* and ε_i. We follow the standard approach and assume that $E[V_i^*] = 0$ and $E[\varepsilon_i] = 0$, which are imposed by the restrictions $\sum_{j=1}^{J} p_{V^*,j}\mu_{V^*,j} = 0$ and $\sum_{\ell=1}^{L} p_{\varepsilon,\ell}\mu_{\varepsilon,\ell} = 0$. Alternatively, it is possible to normalize the medians instead of the means. For example, the restriction $\sum_{\ell=1}^{L} p_{\varepsilon,\ell}\Phi(0;\mu_{\varepsilon,\ell},\sigma^2_{\varepsilon,\ell}) = 1/2$ imposes normalization $\text{median}(\varepsilon_i) = 0$, which in particular allows ε_i to be a non-classical measurement error.[6]

This model naturally generalizes the classic Gaussian model considered in the previous sections. Taking $J = 1$ restricts V_i^* to be Gaussian. Taking $L = 1$ corresponds to ε_i being

[6] It is also possible to impose a normalization on the modes of the distributions of ε_i and/or V_i^*, although this is less convenient computationally.

Gaussian.

Thus, we consider the IV-Mixed-Tobit model that consists of equations (3.8)-(3.9) and replaces equation (3.10) with the assumption that

$$f_{UV}(u,v)=\sum_{k=1}^{K}p_k\phi(u,v;\mu_{UV,k},\Sigma_{UV,k}),\qquad \mu_{UV,k}=\begin{pmatrix}\mu_{U,k}\\ \mu_{V,k}\end{pmatrix},\quad \Sigma_{UV,k}=\begin{pmatrix}\sigma^2_{U,k} & \sigma_{UV,k}\\ \sigma_{UV,k} & \sigma^2_{V,k}\end{pmatrix}, \tag{8.21}$$

where $\sum_{k=1}^{K}p_k\mu_{UV,k}=(0,0)'$ and $\sum_{k=1}^{K}p_k=1$.

First, we discuss identification of the parameters of the above IV-Mixed-Tobit model. As in Section 3, π_0 is immediately identified by the first stage regression of X_i on Z_i and W_i. Identification of the remaining parameters is less straightforward and is established by the following proposition. This identification result appears to be new.

Proposition 8.2 *Suppose $\pi_{01}\neq 0$ and $E[(Z_i',W_i')'(Z_i',W_i')]$ has full rank. Then parameters θ_0, π_0, and $\{p_k,\mu_{UV,k},\Sigma_{UV,k}\}_{k=1}^{K}$ are identified (up to relabelling of the mixture components indexed by k).*

Proposition 8.2 establishes identification of parameters $\theta_0,\pi_0,\{p_k,\mu_{UV,k},\Sigma_{UV,k}\}_{k=1}^{K}$. These parameters can be estimated by the MLE. The log-likelihood can be optimized using the EM algorithm commonly employed for estimation of mixture models. Alternatively, for moderate K, it is also feasible to optimize the log-likelihood directly since both the log-likelihood and its Jacobian are available in closed form. From a practical perspective, we do not recommend models with large K, because Gaussian mixture models are known to be highly flexible even with relatively small K.

Our goal is to provide bounds on the partial effects. To do this, as in Section 3, the key is to construct the identified set for $\sigma^2_{U^*}$. Notice that, for every k, $\Sigma_{UV,k}$ is equal to $\Sigma_{UV,j\ell}$ for some j and ℓ satisfying the restrictions given by equation (8.20). Since these restrictions have the same structure as the ones in equation (3.11), we can apply the result of Proposition 3.1 with a given $\Sigma_{UV,k}$ to construct an identified set for $\sigma^2_{U^*}$ given by

$$\mathcal{I}_k=\left[\underline{\sigma}^2_{U^*,k},\sigma^2_{U,k}\right],$$

where $\underline{\sigma}^2_{U^*,k}$ is computed as in equation (3.12) with $\sigma^2_{U,k}$, $\sigma^2_{V,k}$ and $\sigma_{UV,k}$ in place of σ^2_U, σ^2_V, and σ_{UV}, respectively. Hence, we can construct an identified set for $\sigma^2_{U^*}$ by intersecting $\mathcal{I}_k$ for $k\in\{1,\dots,K\}$, i.e., the bounds for $\sigma^2_{U^*}$ can be constructed as

$$\sigma^2_{U^*}\in\mathcal{I}\equiv\bigcap_k\mathcal{I}_k=\left[\max_k\underline{\sigma}^2_{U^*,k},\min_k\sigma^2_{U,k}\right]. \tag{8.22}$$

Proposition 8.3 *The identified set for $\sigma^2_{U^*}$ given by equation (8.22) is sharp.*

Once the identified set for $\sigma^2_{U^*}$ is constructed, the rest of the analysis follows as in the Gaussian model in Sections 2-3. In particular, sharp bounds for the partial effects are constructed as in equation (3.14) with $\mathcal{I}$ as the identified set for $\sigma^2_{U^*}$. That is, the lower and upper bounds for partial effects for the j^{th} covariate, $PE_j(h)$, are given by

$$\min_{\sigma^2_{U^*}\in\mathcal{I}} PE_j\left(h,\sigma^2_{U^*}\right) \quad \text{and} \quad \max_{\sigma^2_{U^*}\in\mathcal{I}} PE_j\left(h,\sigma^2_{U^*}\right).$$

Note that set $\mathcal{I}$ is a closed interval, as in the original IV-Tobit model. Thus, the discussion concerning the implementation of the identified sets for the partial effects that follows equation (3.14) also applies to the IV-Mixed-Tobit model of this section.

9 Conclusion

Both structural endogeneity and mismeasurement of covariates are pervasive in economic data. Estimating partial effects and other counterfactuals using such data is an important practical problem. This paper addresses the problem in the classic Probit and Tobit models. The relative simplicity of these nonlinear models allows for transparent analysis and practical solutions. The paper provides simple estimators and confidence intervals that are easy to implement. The paper also shows how the proposed solutions can be extended to settings with non-Gaussian unobservables.

Acknowledgements

Evdokimov gratefully acknowledges the support from the Spanish MICIU/AEI via grants RYC2020-030623-I, funded by MICIU/AEI/10.13039/501100011033 and by "ESF Investing in your future", PID2019-107352GB-I00, PID2022-140825NB-I00, and Severo Ochoa Programme CEX2019-000915-S. Zeleneev gratefully acknowledges the generous funding from the UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee (Grant Ref: EP/X02931X/1).

References

Adusumilli, K. and T. Otsu (2018): "Nonparametric Instrumental Regression with Errors in Variables," *Econometric Theory*, 34, 1256–1280.

Blundell, R. and J. L. Powell (2003): "Endogeneity in Nonparametric and Semiparametric Regression Models," in *Advances in Economics and Econometrics*, ed. by M. Dewatripont, L. P. Hansen, and S. J. Turnovsky, 312–357.

Chesher, A. (2003): "Identification in Nonseparable Models," *Econometrica*, 71, 1405–1441.

Chesher, A., D. Kim, and A. M. Rosen (2023): "IV methods for Tobit models," *Journal of Econometrics*, 235, 1700–1724.

Hahn, J. and G. Ridder (2017): "Instrumental variable estimation of nonlinear models with nonclassical measurement error using control variables," *Journal of Econometrics*, 200, 238–250.

Imbens, G. and W. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity," *Econometrica*, 77, 1481–1512.

Kotlarski, I. (1967): "On Characterizing The Gamma And The Normal Distribution," *Pacific Journal Of Mathematics*, 20, 69–76.

Matzkin, R. L. (2013): "Nonparametric Identification in Structural Economic Models," *Annual Review of Economics*, 5, 457–486.

McCloskey, A. (2017): "Bonferroni-based size-correction for nonstandard testing problems," *Journal of Econometrics*, 200, 17–35.

Mroz, T. A. (1987): "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions," *Econometrica*, 55, 765.

Newey, W. K. (1987): "Efficient estimation of limited dependent variable models with endogenous explanatory variables," *Journal of Econometrics*, 36, 231–250.

Rivers, D. and Q. H. Vuong (1988): "Limited Information Estimators And Exogeneity Tests For Simultaneous Probit Models," *Journal of Econometrics*, 39, 347–366.

Romano, J. P. and M. Wolf (2005): "Stepwise Multiple Testing as Formalized Data Snooping," *Econometrica*, 73, 1237–1282.

Schennach, S. (2022): “Measurement Systems,” *Journal of Economic Literature*, 60, 1223–1263.

Schennach, S., H. White, and K. Chalak (2012): “Local indirect least squares and average marginal effects in nonseparable structural systems,” *Journal of Econometrics*, 166, 282–302.

Schennach, S. M. (2020): “Mismeasured and unobserved variables,” in *Handbook of Econometrics*, Elsevier, 487–565.

Smith, R. J. and R. W. Blundell (1986): “An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply,” *Econometrica*, 54, 679.

Song, S., S. M. Schennach, and H. White (2015): “Estimating nonseparable models with mismeasured endogenous variables,” *Quantitative Economics*, 6, 749–794.

Wooldridge, J. M. (2005): “Unobserved Heterogeneity and Estimation of Average Partial Effects,” in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, ed. by D. Andrews and J. Stock, Cambridge University Press, 27–55.

——— (2010): *Econometric Analysis of Cross Section and Panel Data, Second Edition*, The MIT Press.

——— (2015): “Control Function Methods in Applied Econometrics,” *Journal of Human Resources*, 50, 420–445.

A Appendix

A.1 Proof of Proposition 3.1

Note that $\sigma^2_{U^*V^*} \leq \sigma^2_{U^*}\sigma^2_{V^*}$ combined with equation (3.11) implies

$$\begin{aligned} 0 &\leq \sigma^2_{U^*}\sigma^2_{V^*} - \sigma^2_{U^*V^*} = \left(\sigma^2_U - \theta^2_{01}\sigma^2_\varepsilon\right)\left(\sigma^2_V - \sigma^2_\varepsilon\right) - \left(\sigma_{UV} + \theta_{01}\sigma^2_\varepsilon\right)^2 \\ &= \sigma^2_U\sigma^2_V - \sigma^2_{UV} - \sigma^2_\varepsilon\left(\sigma^2_V\theta^2_{01} + 2\sigma_{UV}\theta_{01} + \sigma^2_U\right), \text{ and hence} \\ 0 &\leq \sigma^2_\varepsilon \leq \frac{\sigma^2_U\sigma^2_V - \sigma^2_{UV}}{\sigma^2_V\theta^2_{01} + 2\sigma_{UV}\theta_{01} + \sigma^2_U}. \end{aligned} \tag{A.1}$$

Since $|\rho_{UV}| < 1$, the denominator in this fraction is positive. Since $\sigma^2_\varepsilon \leq \sigma^2_V$, let

$$\overline{\sigma}^2_\varepsilon \equiv \min\left\{\frac{\sigma^2_U\sigma^2_V - \sigma^2_{UV}}{\sigma^2_V\theta^2_{01} + 2\sigma_{UV}\theta_{01} + \sigma^2_U}, \sigma^2_V\right\}.$$

Then

$$\underline{\sigma}_{U^*}^2 \leq \sigma_{U^*}^2 \leq \sigma_U^2, \qquad \underline{\sigma}_{U^*}^2 \equiv \max\left\{\sigma_U^2 - \theta_{01}^2\overline{\sigma}_\varepsilon^2, 0\right\}. \tag{A.2}$$

Note that

$$\sigma_U^2 - \theta_{01}^2 \frac{\sigma_U^2\sigma_V^2 - \sigma_{UV}^2}{\sigma_V^2\theta_{01}^2 + 2\sigma_{UV}\theta_{01} + \sigma_U^2} = \frac{\sigma_U^4 + 2\sigma_U^2\sigma_{UV}\theta_{01} + \theta_{01}^2\sigma_{UV}^2}{\sigma_V^2\theta_{01}^2 + 2\sigma_{UV}\theta_{01} + \sigma_U^2} = \frac{\left(\theta_{01}\sigma_{UV} + \sigma_U^2\right)^2}{\sigma_V^2\theta_{01}^2 + 2\sigma_{UV}\theta_{01} + \sigma_U^2}.$$

Thus, $\underline{\sigma}_{U^*}^2$ in equation (A.2) can be equivalently written as

$$\underline{\sigma}_{U^*}^2 \equiv \max\left\{\frac{\left(\theta_{01}\sigma_{UV} + \sigma_U^2\right)^2}{\sigma_V^2\theta_{01}^2 + 2\sigma_{UV}\theta_{01} + \sigma_U^2}, \sigma_U^2 - \theta_{01}^2\sigma_V^2\right\},$$

where the fraction is always non-negative, ensuring that $\underline{\sigma}_{U^*}^2 \geq 0$. Similarly, equation (3.11) implies that $\sigma_{U^*V^*}$ is bounded between σ_{UV} and $\sigma_{UV} + \theta_{01}\overline{\sigma}_\varepsilon^2$.

Notice that, by construction, $[\underline{\sigma}_{U^*}^2, \sigma_U^2]$ is a valid identified set, i.e., the requirement $\sigma_{U^*}^2 \in [\underline{\sigma}_{U^*}^2, \sigma_U^2]$ is necessary: any $\sigma_{U^*}^2 \notin [\underline{\sigma}_{U^*}^2, \sigma_U^2]$ would violate at least one of the necessary primitive requirements considered above.

Next we show that the constructed identified set for $\sigma_{U^*}^2$ is sharp, i.e., for any $\sigma_{U^*}^2 \in [\underline{\sigma}_{U^*}^2, \sigma_U^2]$ there exist compatible $\sigma_{V^*}^2 \geq 0$, $\sigma_\varepsilon^2 \geq 0$ and $\sigma_{U^*V^*}$ satisfying $\sigma_{U^*V^*}^2 \leq \sigma_{U^*}^2\sigma_{V^*}^2$ consistent with the distribution of the observables, i.e., such that equation (3.11) holds. First, notice that if $\theta_{01} = 0$, $\underline{\sigma}_{U^*}^2 = \sigma_U^2$, so the identified set for $\underline{\sigma}_{U^*}^2$is a singleton and, hence, it is sharp. Suppose that $\theta_{01} \neq 0$. Consider any $\sigma_{U^*}^2 \in [\underline{\sigma}_{U^*}^2, \sigma_U^2]$. Solving for σ_ε^2, $\sigma_{V^*}^2$ and $\sigma_{U^*V^*}$ from equation (3.11) gives

$$\sigma_\varepsilon^2 = (\sigma_U^2 - \sigma_{U^*}^2)/\theta_{01}^2, \quad \sigma_{V^*}^2 = \sigma_V^2 - (\sigma_U^2 - \sigma_{U^*}^2)/\theta_{01}^2, \quad \sigma_{U^*V^*} = \sigma_{UV} + (\sigma_U^2 - \sigma_{U^*}^2)/\theta_{01}. \tag{A.3}$$

Together, these define a (valid) data-generating process (2.4)-(2.7) that is observationally equivalent to the data-generating process in equations (3.8)-(3.10). Thus, the identified set $[\underline{\sigma}_{U^*}^2, \sigma_U^2]$ is sharp. □

A.2 Derivation of equations (5.18) and (5.19)

$$
\begin{aligned}
APE_j^{\text{Tob}}\left(\sigma_{U^*}^2\right) &= \mathrm{E}\left[\Phi\left(\frac{\theta_0' H_i^*}{\sigma_{U^*}}\right)\theta_{0j}\right] = \mathrm{E}\left[\Phi\left(\frac{\theta_{01}X_i^* + \theta_{02}' W_i}{\sigma_{U^*}}\right)\right]\theta_{0j} \\
&= \mathrm{E}\left[\Phi\left(\frac{\theta_{01}\pi_{01}' Z_i + \left(\theta_{01}\pi_{02} + \theta_{02}\right)' W_i}{\sigma_{U^*}} + \frac{\theta_{01}V_i^*}{\sigma_{U^*}}\right)\right]\theta_{0j} \\
&= \mathrm{E}\left[\Phi\left(\frac{\theta_{01}\pi_{01}' Z_i + \left(\theta_{01}\pi_{02} + \theta_{02}\right)' W_i}{\sigma_{U^*}\sqrt{1 + \theta_{01}^2\sigma_{V^*}^2/\sigma_{U^*}^2}}\right)\right]\theta_{0j} \\
&= \mathrm{E}\left[\Phi\left(\frac{\theta_{01}\pi_{01}' Z_i + \left(\theta_{01}\pi_{02} + \theta_{02}\right)' W_i}{\sqrt{2\sigma_{U^*}^2 - \sigma_U^2 + \theta_{01}^2\sigma_V^2}}\right)\right]\theta_{0j},
\end{aligned}
$$

where the penultimate equality comes from taking expectation with respect to V_i^*, and the last equality follows from equation (3.11).

Derivation of equation (5.19) is similar:

$$
\begin{aligned}
APE_j^{\text{Pr}}\left(\sigma_{U^*}^2\right) &= \mathrm{E}\left[\phi\left(\frac{\theta_0' H_i^*}{\sigma_{U^*}}\right)\frac{\theta_{0j}}{\sigma_{U^*}}\right] = \mathrm{E}\left[\phi\left(\frac{\theta_{01}\pi_{01}' Z_i + \left(\theta_{01}\pi_{02} + \theta_{02}\right)' W_i + \theta_{01}V_i^*}{\sigma_{U^*}}\right)\right]\frac{\theta_{0j}}{\sigma_{U^*}} \\
&= \frac{1}{\sqrt{1 + \theta_{01}^2\sigma_{V^*}^2/\sigma_{U^*}^2}}\mathrm{E}\left[\phi\left(\frac{\theta_{01}\pi_{01}' Z_i + \left(\theta_{01}\pi_{02} + \theta_{02}\right)' W_i}{\sigma_{U^*}\sqrt{1 + \theta_{01}^2\sigma_{V^*}^2/\sigma_{U^*}^2}}\right)\right]\frac{\theta_{0j}}{\sigma_{U^*}} \\
&= \mathrm{E}\left[\phi\left(\frac{\theta_{01}\pi_{01}' Z_i + \left(\theta_{01}\pi_{02} + \theta_{02}\right)' W_i}{\sqrt{2\sigma_{U^*}^2 - \sigma_U^2 + \theta_{01}^2\sigma_V^2}}\right)\right]\frac{\theta_{0j}}{\sqrt{2\sigma_{U^*}^2 - \sigma_U^2 + \theta_{01}^2\sigma_V^2}}. \qquad \square
\end{aligned}
$$

A.3 Proof of Proposition 8.2

First, parameters π_0 are immediately identified from the first stage regression.

Let $Q_i \equiv \pi_{01}' Z_i + \pi_{02}' W_i$, so $X_i = Q_i + V_i$. Since π_0 is identified, Q_i is effectively observed. First, notice that the p.d.f. of the joint distribution of Y_i^* and X_i given Q_i and W_i is given by

$$
\begin{aligned}
f_{Y^*X|QW}\left(y, x|q, w\right) &= f_{Y^*V|QW}\left(y, x - q|q, w\right) \\
&= f_{UV}\left(y - \left(\theta_{01}\left(q + v\right) + \theta_{02}' w\right), v|q, w\right)|_{v = x - q} \\
&= f_{UV}\left(y - \left(\theta_{01}x + \theta_{02}' w\right), x - q\right) \\
&= \sum_{k=1}^{K} p_k \phi(y - \left(\theta_{01}x + \theta_{02}' w\right), x - q; \mu_{UV,k}, \Sigma_{UV,k}).
\end{aligned}
$$

Then,

$$f_{YX|QW}(y,x|q,w) = \begin{cases} f_{Y^*X|QW}(y,x|q,w) \text{ if } y>0; \\ \int_{-\infty}^{0} f_{Y^*X|QW}(t,x|q,w)\,dt \text{ if } y=0. \end{cases}$$

Consider the part of function $f_{YX|QW}(y,x|q,w)$ for $y>0$. Since all the involved variables are observed, this function is identified nonparametrically for all $y>0$ and x,q,w (in the support of the respective random variables). Since function $f_{Y^*X|QW}(y,x|q,w)$ is entire (for any fixed x,q,w), identification of $f_{YX|QW}(y,x|q,w)$ for $y>0$ implies that $f_{Y^*X|QW}(y,x|q,w)$ is identified for all y and x,q,w.

Thus, identification of the model using the data on (Y_i,X_i,Z_i,W_i) is equivalent to identification of the model using the data on (Y_i^*,X_i,Z_i,W_i). Then, θ_0 is identified by the linear IV regression argument. Hence, $U_i = Y_i^* - \theta_{01}X_i - \theta_{02}'W_i$ and $V_i = X_i - Q_i$ are effectively observed, and their joint density $f_{UV}(u,v)$ (without any truncation) is identified, which in turn implies that $\{p_k,\mu_{UV,k},\Sigma_{UV,k}\}_{k=1}^{K}$ are identified. □

A.4 Proof of Proposition 8.3

First, notice that $\mathcal{I}$ is a valid identified set, i.e., $\sigma^2_{U^*}\in\mathcal{I}$ is a necessary requirement. Indeed, in the proof of Proposition 3.1, we demonstrated that $\sigma^2_{U^*}\in\mathcal{I}_k$ is a necessary requirement (notice that the same argument applies because $\Sigma_{UV,k}$ is equal to $\Sigma_{UV,j\ell}$ for some j and ℓ satisfying the requirements in equation (8.20)). Since this has to hold for all $k\in\{1,\ldots,K\}$, $\sigma^2_{U^*}\in\bigcap_k\mathcal{I}_k=\mathcal{I}$ is also a necessary requirement.

Next, we demonstrate that $\mathcal{I}$ is sharp. As in the proof of Proposition 3.1, if $\theta_{01}=0$, the identified set is a singleton, and the statement is trivial. Below, we consider $\theta_{01}\neq 0$.

Let $\{p_{V^*,j0},\mu_{U^*V^*,j0},\Sigma_{U^*V^*,j0}\}_{j=1}^{J}$ and $\left\{p_{\varepsilon,\ell 0},\mu_{\varepsilon,\ell 0},\sigma^2_{\varepsilon,\ell 0}\right\}_{\ell=1}^{L}$ denote the true values of the (structural) parameters that map into the joint distribution of (U_i,V_i) according to equation (8.20).

To demonstrate that $\mathcal{I}$ is sharp, we will show that for any $\sigma^2_{U^*}\in\mathcal{I}$, there exist a compatible set of parameters $\{p_{V^*,j},\mu_{U^*V^*,j},\Sigma_{U^*V^*,j}\}_{j=1}^{J}$ and $\left\{p_{\varepsilon,\ell},\mu_{\varepsilon,\ell},\sigma^2_{\varepsilon,\ell}\right\}_{\ell=1}^{L}$ mapping into the same distribution of (U_i,V_i) as the true parameters and hence producing the same distributions of observables.

In, particular we will start with showing that for any $\sigma^2_{U^*}\in\mathcal{I}$ there exist compatible $\sigma^2_{V^*,j}\geq 0$, $\sigma^2_{\varepsilon,\ell}\geq 0$, and $\sigma_{U^*V^*,j}$ satisfying $\sigma^2_{U^*V^*,j}\leq\sigma^2_{U^*}\sigma^2_{V^*,j}$ mapping into the same collection of $\Sigma_{UV,j\ell}$ for $j\in\{1,\ldots,J\}$ and $\ell\in\{1,\ldots,L\}$.

Take any $\sigma^2_{U^*}\in\mathcal{I}$. First, notice that, for any (j,ℓ), there exists $k(j,\ell)$ such that $\Sigma_{k(j,\ell)}=\Sigma_{UV,j\ell}$ satisfying the requirements in equation (8.20). Since $\sigma^2_{U^*}\in\mathcal{I}_{k(j,\ell)}$, there exist unique

$\sigma^2_{V^*,j\ell}$, $\sigma^2_{\varepsilon,j\ell}$, and $\sigma_{U^*V^*,j\ell}$ consistent with $\Sigma_{UV,j\ell} = \Sigma_{k(j,\ell)}$ (as demonstrated in the proof of Proposition 3.1). This triplet is computed using equation (A.3) for any (j, ℓ).

Next, we need to show that these triplets are also internally consistent along the j and ℓ dimensions, i.e., that $\sigma^2_{V^*,j\ell} = \sigma^2_{V^*,j\ell'} = \sigma^2_{V^*,j}$ and $\sigma_{U^*V^*,j\ell} = \sigma_{U^*V^*,j\ell'} = \sigma_{U^*V^*,j}$ for any $\ell, \ell' \in \{1, \ldots, L\}$, and, similarly, $\sigma^2_{\varepsilon,j\ell} = \sigma^2_{\varepsilon,j'\ell} = \sigma^2_{\varepsilon,\ell}$ for any $j, j' \in \{1, \ldots, J\}$. Notice that according to equation (8.20), we also have

$$\begin{pmatrix} \sigma^2_{U,j\ell} & \sigma_{UV,j\ell} \\ \sigma_{UV,j\ell} & \sigma^2_{V,j\ell} \end{pmatrix} = \begin{pmatrix} \sigma^2_{U^*0} + \theta^2_{01}\sigma^2_{\varepsilon,\ell 0} & \sigma_{U^*V^*,j0} - \theta_{01}\sigma^2_{\varepsilon,\ell 0} \\ \sigma_{U^*V^*,j0} - \theta_{01}\sigma^2_{\varepsilon,\ell 0} & \sigma^2_{V^*,j0} + \sigma^2_{\varepsilon,\ell 0} \end{pmatrix}.$$

Using the relationship above and the expressions for $\sigma^2_{V^*,j\ell}$ and $\sigma_{U^*V^*,j\ell}$ obtained in the proof of Proposition 3.1, for any ℓ we have

$$\sigma^2_{V^*,j\ell} = \sigma^2_{V,j\ell} - (\sigma^2_{U,j\ell} - \sigma^2_{U^*})/\theta^2_{01} = \sigma^2_{V^*,j0} - (\sigma^2_{U^*0} - \sigma^2_{U^*})/\theta^2_{01} = \sigma^2_{V^*,j},$$
$$\sigma_{U^*V^*,j\ell} = \sigma_{UV,j\ell} + (\sigma^2_{U,j\ell} - \sigma^2_{U^*})/\theta_{01} = \sigma_{U^*V^*,j0} + (\sigma^2_{U^*0} - \sigma^2_{U^*})/\theta_{01} = \sigma_{U^*V^*,j},$$

where the last equalities in both lines provide consistent definitions of $\sigma^2_{V^*,j}$ and $\sigma_{U^*V^*,j}$, which do not depend on ℓ. Similarly, for any j, we have

$$\sigma^2_{\varepsilon,j\ell} = (\sigma^2_{U,j\ell} - \sigma^2_{U^*})/\theta^2_{01} = \sigma^2_{\varepsilon,\ell 0} + (\sigma^2_{U^*0} - \sigma^2_{U^*})/\theta^2_{01} = \sigma^2_{\varepsilon,\ell},$$

where the last equality defines $\sigma^2_{\varepsilon,\ell}$, which does not depend on j.

Finally, notice that, by construction, $\{p_{V^*,j0}, \mu_{U^*V^*,j0}, \Sigma_{U^*V^*,j}\}_{j=1}^J$ and $\left\{p_{\varepsilon,\ell 0}, \mu_{\varepsilon,\ell 0}, \sigma^2_{\varepsilon,\ell}\right\}_{\ell=1}^L$ map into the same joint distribution of (U_i, V_i) as the true parameters $\{p_{V^*,j0}, \mu_{U^*V^*,j0}, \Sigma_{U^*V^*,j0}\}_{j=1}^J$ and $\left\{p_{\varepsilon,\ell 0}, \mu_{\varepsilon,\ell 0}, \sigma^2_{\varepsilon,\ell 0}\right\}_{\ell=1}^L$. Since we picked an arbitrary $\sigma^2_{U^*} \in \mathcal{I}$, this completes the proof that $\mathcal{I}$ is sharp. □