SEMIPARAMETRIC FOURIER-DEPENDENT SIEVE IV ESTIMATOR (SPIV)

October 9, 2017

Nir Billfeld* Moshe Kim†

Instrumental variables are intended to correct for misspecifications largely stemming from endogeneity problems or omission of relevant important covariates correlated with some of the other included covariates. The validity of the IV estimator relies on the orthogonality with respect to the random disturbance. However, in cases of endogenously truncated data as well as in other instances (e.g, censored data) which is very frequently the nature of data used in empirical research, there exists severe contamination in the disturbance due to the endogenous selection process. The endogenous selection process generates a co-movement between the IV and the disturbance which is related to the variation in the selection equation's covariates. This contamination propagates additional bias introduced into the parameter estimates of the various covariates. Consequently, not only that the conventional IV does not solve the problem it is intended to but rather introduces additional bias into the parameter estimates of the various covariates of the substantive equation. Our empirical implementation shows that even under mild correlation between the random disturbances, the resulting bias in the estimated parameter of the endogenous covariate in the substantive equation can amount to almost 10 times the true parameter value for 500 observations and can amount to 5 times the true parameter value in a sample of 10,000 observations. We offer a semi-parametric Fourier-dependent *Sieve IV* (*SPIV*) estimator correcting for both truncation as well as endogeneity biases. The proposed estimator removes the hurdle which prevents orthogonality under truncation or other misspecifications. Using Monte Carlo simulations attest to the high accuracy of our offered semi-parametric *Sieve IV* estimator as expressed by the $\sqrt{n}$ consistency. These results have been verified by utilizing 2,000,000 different distribution functions, practically generating 100 millions realizations to generate the various data sets.

Keywords: Endogenous truncation, Semi-parametric *Sieve* IV, Fourier series.

*University of Haifa nbillfeld@staff.haifa.ac.il

†University of Haifa kim@econ.haifa.ac.il

1. Introduction

It is quiet common that researchers in almost any field of research do not have information regarding the entire data distribution function, a phenomenon referred to as truncation. As such, the truncated data set is most likely of different characteristics than the non-truncated full distribution, leading potentially to biased estimates which challenge the causal interpretation of the covariates in the model on the dependent variable. The problem is further aggravated when this truncation is actually non-random and rather endogenously propagated by various decision units (observations) since it implies the joint dependence of the random disturbance and the covariates in the truncated environment.

Social scientists however, are typically interested in capturing behavioral relationships expressed by casual links between variables (the dependent variable and each of the covariates), rather than merely by their correlations. In order to make causal inferences,[1] they utilize models which often involve *instrumental variables* (IV) (Angrist et al., 1996).[2] The purpose of employing IV is to guarantee a covariate exhibits a causal relationship with the dependent variable as well as solving problems arising from various misspecification.

Endogenous covariates cannot represent causal relationships since some of the variation in the dependent variable of interest stems from a common variation in both the endogenous covariate as well the random disturbance which is not easily decomposable. Application of a proper instrumental variable generates variation in the endogenous covariate without introducing variation in the random disturbance and hence orthogonal to it. Thus, IV contributes to exogeneity and therefore has been extremely popular in empirical work. Note that IV can provide a strategy that can be used to deal with the problem of omitted-variables bias in a wide range of single-equation regression applications, such as models with mismeasured regressors (Durbin, 1954) and the estimation of treatment effects (Angrist and Imbens, 1995; Heckman and Hotz, 1989; Heckman and Robb, 1985).[3]

The primary purpose of the present paper is to show that a valid IV cannot perform its intended task in a truncated environment, because it may not be orthogonal to the random

[1] For a specific form of causality due to treatment effect see Angrist and Kuersteiner (2004)'s definition.

[2] These are variables that are excluded from some equations and included in others, and therefore are correlated with some outcomes only through their effect on other variables (Angrist et al., 1996).

[3] See Angrist and Imbens (1995) for a very elaborate discussion of IV use in the literature.

disturbance in such an environment.[4] Once we have analytically proven that the IV estimator is no longer valid in a truncated environment, we introduce a correction procedure, referred to here as semi-parametric Fourier-dependent *Sieve IV* estimator ($SPIV$) which corrects for both sources of bias: the endogeneity of covariates as well as the endogenous self-selection biases.

Utilizing a semi-parametric estimation procedure such as Sieve or kernel estimators allows us to obtain a distribution free estimator, without having to specify a joint distribution function for the random disturbances in the model. Such an estimator is more general and robust than its parametric counterpart, since it is less dependent on the data generation process. This is important due to the fact that the parametric distributional assumptions are rarely satisfied in practice. However, kernel estimators involve computational burden due to the necessity to find the optimal bandwidth (Ichimura, 1993) and the fact that choosing the wrong bandwidth might leads to bias estimates (Lewbel and Linton, 2007). Unlike kernel estimators, Sieve estimators are based on a combination of basis functions and are global estimators in the sense that the entire function is characterized by a finite set of parameters, such as trigonometric functions (Fourier) and Hermit polynomials. Sieve estimators, being independent of bandwidth,[5] are attractive alternatives to the kernel estimators. Among the various Sieve estimators is the Fourier series, a functional of the *Orthonormal* polynomials sequence family, which allows for efficient estimation of functions with non-smoothness, discontinuities in derivatives, sharp spikes and discontinuities in the function itself. Thus, it is useful in non-parametric regression for approximating a much broader class of functions (Ogden, 2012) than the kernel approach. The major insight of our approach is the realization that generally in a truncated environment there exists co-movement between the instrumental variable and the substantive equation's random disturbance, while this does not exist in the non-truncated environment. This co-movement is propagated by the endogenous selection equation's covariates. In this context our offered estimator is a generic IV as it removes the contamination propagated by any co-movement of instrumental variable and the random disturbance generated by truncation or censoring. In our particular specific treatment the

[4] Note that we deal with endogenously *truncated* sample selection model to differentiate from *censored* sample selection models (Heckman, 1979; Newey, 2009; Powell, 2001) where there exists information pertaining to the non-participants.

[5] Kernel estimators are bandwidth dependent, for example.

co-movement stems from the endogenous selection process which determines truncation.

Instrumental variables are routinely employed in various semi-parametric models as well as non-parametric models using methods which are generalization of the parametric non-linear instrumental variables.[6] These models can be either partially linear (Florens et al., 2012; Robinson, 1988), including the partially linear single index models as a special case using fitted values obtained from the first stage regression (Zhou et al., 2016),[7] or completely non-linear (e.g, Chen et al. (2014); Newey (2013); Newey and Powell (2003).) using control function approach[8] with additivity (Newey et al., 1999) or without additivity (Imbens and Newey, 2009). A quite common method is the conditional moment method which has been introduced by Ai and Chen (2003); Chen and Pouzo (2009) and can be applied also in GMM estimation (Chen and Liao, 2015).[9,10] Additional implementation of IV is present in endogenous sample selection models allowing for a causal interpretation in the presence of endogenous covariates. The identification strategy in these models relies on either the conditional independence of the IV and the selection variable (which is not truncated, and thus is fully observed) given the dependent variable of interest (d'Haultfoeuille, 2010) or the population (non-truncated) distribution function of at least one of the selection equation's covariates (Lewbel, 2007).

However, neither the availability of an auxiliary data set from the entire population nor the conditional independence assumption (which limits the availability of valid instrumental variables) is required in the present article, as the instrumental variable and the selection variable are modeled as conditionally dependent (through variation in the selection equation's covariates) given the substantive equation's dependent variable. Consequently, our present model facilitates finding (and implementing) a valid IV as it is subjected to fewer restrictions such as joint independence.

The key difference between censored and truncated sample selection models with endoge-

[6] The parametric non-linear instrumental variables approach includes the non-linear two stage least squares ($NLLS$) (Amemiya, 1974), the generalized IV ($GIVE$) (Sargan, 1958) and the generalized method of moments (GMM) (Hansen, 1982). For the linear version see Angrist and Imbens (1995).

[7] See Liang et al. (2010) for a detail discussion about the partially linear index model estimation.

[8] The control function approach includes the residual obtained from the first stage regression as an additional covariate in the substantive equation to control for the omitted variables which generates the endogeneity.

[9] See Blundell and Powell (2003) for a further discussion on the various IV methods in semi-parametric and non-parametric models.

[10] Other approach is appropriate only for an invertible index-model (Ahn et al., 2015; Das et al., 2003).

nous self-selection is that in the former, only the dependent variable of interest is truncated, while in the later the entire data set is truncated, including the endogenous covariate being estimated in the first stage regression. Consequently, the substantive equation is not the only equation affected by the endogenous truncation. In fact, the first stage equation is affected as well due to a contamination in its random disturbance with the selection equation's covariates (as depicted graphically in figure 1). In other words, there is a co-movement between its residual and its covariates which is related to a variation in the selection equation's covariates generating endogeneity in the first stage regression's covariates.

In the present article, we are particularly interested in causal interpretation of covariates which jointly co-move with the random disturbance under endogenously truncated data distribution function. In such a model, there are two sources of bias: (i) an endogenous truncation bias and (ii) bias generated by endogenous covariates in the substantive equation. The former bias is due to the joint dependence of the random disturbances in the substantive and selection equations (Heckman, 1979).[11] Decomposing the random disturbance in the substantive equation we uncover these sources of bias and show analytically that the conventional IV fails to provide causal interpretation of the endogenous covariates under endogenously truncated environment. To alleviate this, we offer a semiparametric Fourier dependent Sieve IV estimator enabling the correction of these two biases.

We run Monte Carlo simulations to measure the magnitude of the potential bias in the parameters' estimates under endogenous truncation obtained by employing a conventional IV to eliminate the endogeneity bias. Our empirical implementation shows that even under mild correlation between the random disturbances, the resulting bias in the estimated parameter of the endogenous covariate in the substantive equation can amount to almost 10 times the true parameter value for 500 observations and can amount to 5 times the true parameter value in a sample of 10,000 observations.

2. Motivation

IV estimators are routinely employed to resolve bias introduced by endogenous covariates due to the joint dependence of the covariates and the random disturbance. In the conventional

[11] Such a dependence generates a misspecification in the participants' substantive equation, leading to potential bias in its parameters' estimates (Heckman, 1979).

IV implementation the covariates endogeneity is intrinsic in the model (population regression), in the sense that the cause for this joint dependence is unrelated to sample selection. However, in many cases this joint dependence is generated partially (or entirely) by sample selection generating an endogenously truncated data set which entails endogeneity bias of its own. Consequently, the parameters estimates will be biased and inconsistent due to not only the covariates' endogeneity problem, but also due to the endogenous self-selection, and the interaction of these. Thus, it is the contention of this paper that under conditions of data truncation, the IV estimator does not remove the problem for which it was designed to solve and introduces additional bias. Our primary argument is that in truncated environment the IV is no longer orthogonal to the regression equation's disturbance due to a co-movement between the IV and the selection equation's covariate(s). It is well known (Ichimura and Lee (1991)) that the substantive equation's distribution function of the disturbance is a conditional distribution given the selection equation's covariates. Simply put, generally there exists a co-movement between the disturbance (of the substantive equation) and the IV, related to the variation in the selection equation's covariate(s). The approach we adapt is to discover the reason for this co-movement and neutralizing it, thereby persevering the necessary orthogonality conditions for the IV. To that end, we develop a semi-parametric IV (SPIV) estimator that corrects for both truncation bias as well as endogeneity bias. The proposed estimator removes the hurdle which prevents orthogonality under truncation. In what follows we formally represent our argument.

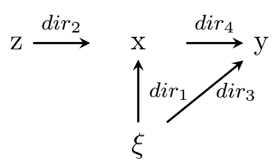
Let $\mathbf{w}$ be a random variable vector, and let x, z and ξ be jointly dependent random variables, such that z and ξ are conditionally independent given $\mathbf{w}$.[12] Denote a random variable y which is constructed as a linear function of x and the random disturbance ξ.

Then, by construction, the co-movement between z and ξ is entirely generated by a variation in $\mathbf{w}$. These five variables $(\mathrm{x}, \mathrm{z}, \mathbf{w}, \mathrm{y}, \xi)$ are interrelated as depicted in figure 1. Only in panel (a) $\mathbf{w}$ doesn't play any role. In panel (c) we assume that x and ξ are both conditionally as well as unconditionally dependent given $\mathbf{w}$ (implying the endogeneity of x is only partially related to $\mathbf{w}$), unlike panel (b) in which x and ξ are conditionally independent given $\mathbf{w}$ although they are unconditionally dependent given this $\mathbf{w}$ (implying the endogeneity of x is

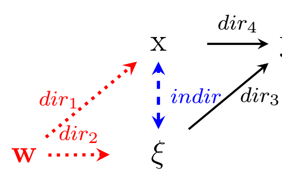
[12] By holding $\mathbf{w}$ constant, z and ξ are not jointly dependent rather, z and ξ are conditionally independent given $\mathbf{w}$.

entirely related to **w**). Figure 1 demonstrates two sources of endogeneity. The first source is due to the presence of endogenous covariate (panel (a)). The second source stems from the endogenous self-selection (panel (b)) represented by **w**, which induces an indirect association between the x and ξ. These two bias sources imply both endogeniety in x and an indirect association between the x and ξ through the covariate **w** (panel (c)). The notation *dir* implies a direct association between two variables which are connected by a thick arrow. The notation *indir* implies an indirect association between two variables (through a mediation covariate **w**) which are connected with a dashed line arrow.

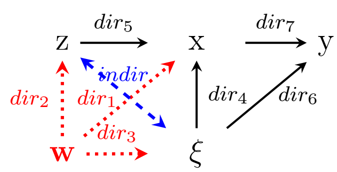
Figure 1: Two sources of endogeneity



(a) An endogenous covariate (x)



(b) Sample selection



(c) Sample selection with endogenous covariate (x)

3. Methodology

As discussed above, IV is based on the following basic requirements: it is correlated with the endogenous covariate, as well as orthogonal to the random disturbance. Additionally, it must satisfy the exclusion restriction according to which in the presence of the endogenous covariate the IV must be excluded from the regression. The IV is allowed to affect the dependent variable only through its affect on the endogenous covariate. However, the orthogonality condition is rarely satisfied in the presence of truncation which is very frequently the nature of data used in empirical research, and therefore the IV will not provide a solution for the endogeneity problem. In what follows, we demonstrate the shortcoming of the conventional IV estimator as well as potential bias generated in an environment of truncated data.

Suppose that there is a population random variable $\omega = (\mathrm{z}; \mathrm{x}_1, \mathbf{x_{-1}}; \mathbf{w})$ and that there is an independent and identically distributed sample $\{z_i, x_{1i}, \boldsymbol{x_{-1_i}}, \boldsymbol{w_i}\}_{i=1}^{N}$ drawn from this population, referred to as the complete data set consisting of N observations.[13] The instrumental variable is z, the endogenous variable is x_1 and the exogenous random variables are $(\mathbf{x_{-1}; w})$, and where $\mathbf{w} \in \mathbb{R}^l$ is a covariates vector.

Let ξ_{1i}, ξ_{2i} and v_i be jointly dependent random disturbances with the respective marginal distribution functions F_{ξ_1}, F_{ξ_2} and F_{v}. Their joint distribution function is $F_{\xi_1,\xi_2,\mathrm{v}}$. The model is semiparametric since neither the marginals nor the joint distribution function are required to be specified by the researcher.

The underlying model is composed of two parts. The first part consists of a selection equation, while the second part consists of the substantive (of interest) equation.

The population (non-truncated) selection equation is defined as:

$$y_{2i}^{*} = \boldsymbol{w}_i^T \boldsymbol{\gamma} + \xi_{2i} \tag{3.1}$$

where $\boldsymbol{\gamma} \in \mathbb{R}^l$ and $\boldsymbol{w_i} \in \mathbb{R}^l$ are the selection equation's coefficients and covariates vector, respectively. The selection equation's random disturbance is denoted by ξ_{2i}.

The substantive's equation and the endogenous variable's equation are defined as system of equations:

$$\begin{bmatrix} y_{1i}^{*} \\ x_{1i}^{*} \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_i^T \\ [\boldsymbol{z}_i^T, \boldsymbol{x}_{-1_i}^T] \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\delta} \end{bmatrix} + \begin{bmatrix} \xi_{1i} \\ \mathrm{v}_i \end{bmatrix} \quad \text{the substantive system of equations} \tag{3.2}$$

where $\boldsymbol{\beta} \in \mathbb{R}^{p_1}$ and $\boldsymbol{\delta} \in \mathbb{R}^{p_2}$ are covariates vectors, x_{1i} is an endogenous variable included in vector $\boldsymbol{x}_i \in \mathbb{R}^{p_1}$, and the exogenous variables are denoted by $\boldsymbol{x}_{-1_i}^T$. The substantive equation's random disturbance is denoted by ξ_{1i} and v_{1i}.

However the variables $y_{1i}^{*}, y_{2i}^{*}, x_{1i}^{*}$ are latent in the truncated environment and their respective observed realizations are denoted by y_{1i}, y_{2i}, x_{1i}, (3.3) and (3.4) defined in to follow.

[13] Capital letters indicate random variables; lower case letters indicate realizations of these random variables.

The variable y_{2i}^* is latent, while is y_{2i} observed and defined as

$$y_{2i} = \begin{cases} 1 & \text{if } y_{2i}^* \geq 0 \\ \text{Unobserved} & \text{if } y_{2i}^* < 0 \end{cases}, \quad \text{the selection equation} \tag{3.3}$$

$$\begin{bmatrix} y_{1i} \\ x_{1i} \end{bmatrix} = \begin{cases} \begin{bmatrix} y_{1i}^* \\ x_{1i}^* \end{bmatrix} & \text{if } y_{2i} = 1 \\ \text{Unobserved} & \text{if } y_{2i} = 0 \end{cases}, \quad \text{the substantive system of equations} \tag{3.4}$$

In the next section we reformulate the substantive equation as a partially linear single index model.

3.1. Semi-parametric selectivity bias correction

Following Robinson (1988), the conditional expectation of the substantive equation in semi-parametric (censored)[14] sample selection models is some generally unknown function $\mathcal{M}_1(.)$ (to be estimated) of the selection equation's covariates variables $\boldsymbol{w_i}$:

$$\mathbb{E}\left[\xi_{1i} | \xi_{2i} > -\boldsymbol{w}_i^T \boldsymbol{\gamma}\right] = \mathcal{M}_1(\boldsymbol{w}_i^T \boldsymbol{\gamma}) \tag{3.5}$$

such that $\boldsymbol{\gamma}$ is the selection equation's coefficients vector. Since y_{1i} is observed only if i is a participant, the substantive equation's dependent variable obtains the following functional form:

$$y_{1i} = \boldsymbol{x}_i^T \boldsymbol{\beta} + \underbrace{\mathcal{M}_1(\boldsymbol{w}_i^T \boldsymbol{\gamma})}_{\text{the bias term}} + \underbrace{\tilde{\epsilon}_{1i}}_{(\xi_{1i} - \mathcal{M}_1(\boldsymbol{w}_i^T \boldsymbol{\gamma}))|\mathrm{y}_{2i}=1} \tag{3.6}$$

The regression equation in (3.5) is referred to as a semi-parametric partially linear regression (SP-NLS), in which the non-linear part is the bias term function. This regression can be estimated semi-parametrically in cases of a truncated sample selection model using a non-

[14] His approach is a generalization of the well-known inverse-mills ratio estimator introduced by Heckman (1979) for the substantive equation's bias term $\mathbb{E}\left[\xi_{1i} | \xi_{2i} > -\boldsymbol{w}_i^T \boldsymbol{\gamma}\right]$ in the case of a censored sample selection model. Note the difference between censored data and truncated data, which is the case we deal with.

linear least squares procedure as suggested by Ichimura (1993).

Both Ichimura (1993)'s and Robinson (1988)'s models involve a kernel function estimation. However, kernel estimates accuracy are sensitive to the bandwidth selected. This entails a potential problem of finding the optimal bandwidth resulting in computational complexity.

One of the drawbacks of the above described methodology is that the kernel estimates' accuracy depend on the bandwidth selection:"Whether there is a way to choose a bandwidth sequence that is optimal for the estimation of the parameters is an open question" (Ichimura and Lee, 1991). Additionally, semi-parametric models which involve a kernel function estimation might lead to biased estimates due to the difficulty of the optimal bandwidth to be found: "The well known bandwidth selection rules used in non-parametric estimation, such as cross validation, are not generally applicable to semi-parametric settings" (Lewbel and Schennach, 2007). Thus, in practice one needs to use a bandwidth that is "slightly" smaller than the optimal bandwidth obtained using the cross-validation procedure. However, this informal method for bandwidth choice may lead to a non-ignorable bias in the estimates (Lewbel and Schennach, 2007).[15] Thus, avoiding the problems involves with kernel estimation our methodology relies on series (Sieve) estimator to approximate the bias term (in (3.6)).

The substantive equation depicted in (3.6) deals with endogenous truncation bias, assuming that in this (substantive) equation the random disturbance and the covariates are not jointly dependent. However, in cases where this random disturbance is jointly dependent with one (or more) of the covariates there are two bias terms. The first one propagated by the endogenous truncation and the second one propagated by the endogenous covariate. Next we present a decomposition Theorem 1 which enables reformulating the substantive equations as a partially linear single index model in the presence of an endogenous covariate.

3.2. Decomposition of the substantive equations

Theorem 1 *Let the underlying model be as depicted in* (3.3) *and* (3.4). *Denote the random disturbances* ε_i *and* ϵ_{1i} *which are constructed as:* $\varepsilon_i = y_{1i}^* - \mathbb{E}[y_{1i}^*|\boldsymbol{x}_i]$ *and* $\epsilon_{1i} = y_{1i} - \mathbb{E}[y_{1i}^*|y_{2i} = 1]$, *respectively. The following requirements must hold: (i)* $y_{1i} = \boldsymbol{x}_i^T\boldsymbol{\beta} + \mathbb{E}[\xi_{1i}\,|\,\boldsymbol{x}_i] + \varepsilon_i \ \forall i \in \{1, ..., N\}$; *(ii)* $y_{1i} = \boldsymbol{x}_i^T\boldsymbol{\beta} + \epsilon_{1i}^*, \quad \mathbb{E}[\epsilon_{1i}|\ y_{2i} = 1] = 0, \quad \epsilon_{1i}^* \equiv \epsilon_{1i} + \mathbb{E}[\xi_{1i}\,|\,\boldsymbol{x}_i] + \mathcal{M}(\boldsymbol{w_i}^T\boldsymbol{\gamma})$,

[15] There is no "scientific protocol" enabling the determination of a proper bandwidth (in semi-parametric estimation). It is known though that incorrect bandwidth choice leads to bias estimates (Lewbel and Schennach, 2007).

$\forall i \in \{i|y_{2i} = 1\}$.

Proof: By construction $\varepsilon_i = y^*_{1i} - \mathbb{E}[y^*_{1i}|\boldsymbol{x}_i]$, it follows that

$$y^*_{1i} = \mathbb{E}[\boldsymbol{x}_i^T\boldsymbol{\beta} \mid \boldsymbol{x}_i] + \mathbb{E}[\xi_{1i} \mid \boldsymbol{x}_i] + \varepsilon_i \tag{3.7}$$

Using (3.7) we get:

$$\mathbb{E}[y^*_{1i}|y_{2i} = 1] = \mathbb{E}[\varepsilon_i|y_{2i} = 1] + \mathbb{E}\left\{\mathbb{E}[\boldsymbol{x}_i^T\boldsymbol{\beta} \mid \boldsymbol{x}_i]|y_{2i} = 1\right\} + \mathbb{E}\left\{\mathbb{E}[\xi_{1i} \mid \boldsymbol{x}_i]|y_{2i} = 1\right\} \tag{3.8}$$

which is simplified to:

$$\mathbb{E}[y^*_{1i}|y_{2i} = 1] = \mathbb{E}[\varepsilon_i|y_{2i} = 1] + \mathbb{E}[\boldsymbol{x}_i^T\boldsymbol{\beta} \mid \boldsymbol{x}_i] + \mathbb{E}[\xi_{1i} \mid \boldsymbol{x}_i] \tag{3.9}$$

In order to obtain the substantive equation in the truncated environment, we construct $\epsilon_{1i} = y_{1i} - \mathbb{E}[y^*_{1i}|y_{2i} = 1]$ where $\mathbb{E}[y^*_{1i}|y_{2i} = 1]$ is obtained from (3.9).[16] Following Ichimura and Lee (1991), the conditional expectation of ε_i given participation is expressed by some unknown function $\mathcal{M}_1(\cdot)$ as $\mathbb{E}[\varepsilon_i|y_{2i} = 1] = \mathcal{M}_1(\boldsymbol{w}_i^T\boldsymbol{\gamma})$. Thus, we obtain:

$$y_{1i} = \underbrace{\boldsymbol{x}_i^T\boldsymbol{\beta}}_{\substack{\text{substantive}\\ \text{covariates}}} + \overbrace{\underbrace{\mathcal{M}_1(\boldsymbol{w}_i^T\boldsymbol{\gamma})}_{\substack{\text{selection bias}\\ \text{term}}} + \underbrace{\mathbb{E}[\xi_{1i} \mid \boldsymbol{x}_i]}_{\substack{\text{endogeneity bias}\\ \text{term}}} + \underbrace{\epsilon_{1i}}_{\text{white noise}}}^{\epsilon^*_{1i}} \tag{3.10}$$ ■

For sake of brevity we present equation (3.10) which is a decomposition of the substantive equation into its components such as substantive equation's covariates; selection bias term; endogeniety bias term and a stochastic white noise term. It is easy to see that the conventional IV can not be sufficient in eliminating the endogeniety bias $\mathbb{E}[\xi_{1i} \mid \boldsymbol{x}_i]$ in (3.10), since under truncation the endogeneity bias term is actually $\mathbb{E}[\xi_{1i} \mid \boldsymbol{x}_i, y_{2i} = 1]$.

Similarly, we construct $\epsilon_{2i} = x_{1i} - \mathbb{E}[x^*_{1i}|y_{2i} = 1]$ where $\mathbb{E}[x^*_{1i}|y_{2i} = 1]$ satisfies:

$$\mathbb{E}[x^*_{1i}|y_{2i} = 1] = \mathbb{E}[\mathrm{v}_i|y_{2i} = 1] + \mathbb{E}[[\boldsymbol{z}_i^T, \boldsymbol{x}_{-1_i}^T]\boldsymbol{\delta} \mid y_{2i} = 1] \tag{3.11}$$

[16] By construction of y_{1i}, the equality $\mathbb{E}[y_{1i}|y_{2i} = 1] = \mathbb{E}[y^*_{1i}|y_{2i} = 1]$ must be satisfied. It implies that $\mathbb{E}[\epsilon_{1i}|y_{2i} = 1] = \mathbb{E}[y_{1i}|y_{2i} = 1] - \mathbb{E}\left\{\mathbb{E}[y^*_{1i}|y_{2i} = 1]|y_{2i} = 1\right\} = \mathbb{E}[y_{1i}|y_{2i} = 1] - \mathbb{E}[y^*_{1i}|y_{2i} = 1] = 0$.

to get:

$$\epsilon_{2i} = x_{1i} - \mathbb{E}[\mathrm{v}_i | y_{2i} = 1] - \mathbb{E}[[\boldsymbol{z}_i^T, \boldsymbol{x}_{-1_i}^T]\boldsymbol{\delta} \mid y_{2i} = 1] \tag{3.12}$$

We express $\mathbb{E}[\mathrm{v}_i|y_{2i} = 1]$ in (3.12) as $\mathbb{E}[\mathrm{v}_i|y_{2i} = 1] = \mathcal{M}_2(\boldsymbol{w}_i^T\boldsymbol{\gamma})$ where $\mathcal{M}_2(\cdot)$ (see Theorem 4 to follow) is some unknown function and obtain:

$$x_{1i} = \underbrace{[\boldsymbol{z}_i^T, \boldsymbol{x}_{-1_i}^T]\boldsymbol{\delta}}_{\substack{\text{substantive}\\\text{covariates}}} + \overbrace{\underbrace{\mathcal{M}_2(\boldsymbol{w}_i^T\boldsymbol{\gamma})}_{\substack{\text{selection bias}\\\text{term}}} + \underbrace{\epsilon_{2i}}_{\text{white noise}}}^{\epsilon_{2i}^*} \tag{3.13}$$ ■

It is easy to see the joint dependence of ϵ_{2i}^* and ϵ_{1i}^* through the selection bias terms in (3.10) and (3.13).

Next we formulate the relationship between the covariates and dependent variables in the equations to be estimated, in the presence of an endogenous covariate in the substantive equation under truncation.

3.3. Truncated sample selection model with an endogenous covariate

In cases where the substantive equation's dependent variable is a function of an endogenous covariate x_{1i}, both x_{1i} as well as y_{1i} (as in (3.4)) are truncated and therefore, we face a truncated sample selection model with an endogenous covariate.

Thus, the semi-parametric partially linear index model in a truncated environment consists of the following system of equations:

$$\begin{bmatrix} y_{1i} \\ x_{1i} \end{bmatrix} = \begin{cases} \boldsymbol{x}_i^T\boldsymbol{\beta} + \mathcal{M}_1(\boldsymbol{w}_i^T\boldsymbol{\gamma}) + \overbrace{\mathbb{E}[\xi_{1i} \mid \boldsymbol{x}_i] + \underbrace{\epsilon_{1i}}_{\text{white noise}}}^{\epsilon_{1i}^{**}} \\ [\boldsymbol{z}_i^T, \boldsymbol{x}_{-1_i}^T]\boldsymbol{\delta} + \mathcal{M}_2(\boldsymbol{w}_i^T\boldsymbol{\gamma}) + \underbrace{\epsilon_{2i}}_{\text{white noise}} \end{cases} \tag{3.14}$$

where ϵ_{1i} and ϵ_{2i} are two jointly dependent random disturbances,[17] and by construction are

[17] There is dependency of these two random disturbances due to the dependence between v_i and ξ_{1i} (as in (3.4)) in the complete (non-truncated) data.

independent of the random variables vector $\mathbf{w}$.[18] The intrinsic endogeneity in the model is captured by the joint dependence of ϵ_{1i}^{**} and the covariates.[19] The presence of the function $\mathcal{M}_2(.)$ implies that we allow for a dependence between $\mathrm{v_i}$ (the endogenous part of x_i) and the selection equation's random disturbance ξ_{2i} (in (3.3)). In such cases, x_i is endogenous to both random disturbances ϵ_{1i} and ϵ_{2i}.

We assume that the instrumental variable is jointly distributed with all covariates in the data. Our primary interest is to show how the conditional expectation of the random variables product $\mathrm{z} \cdot \xi_1$ given participation (which is a function of a covariates vector $\mathbf{w}$) is affected by the co-movement of each one of these random variables with respect to $\mathbf{w}$. For doing so, we initially simplify the expectation of these random variables product, utilizing the Tower property (Williams, 1991) of conditional expectation as follows:

Let z and $\mathbf{w}$ be continuous random variables, and let s be a discrete variable, indicating participation. These three random variables $\{\mathrm{z}, \mathbf{w}, \mathrm{s}\}$ must satisfy the following property (see Appendix for a formal proof):

$$\mathbb{E}_{\mathbf{w}}\left[\mathbb{E}[\mathrm{z}|\mathbf{w}, \mathrm{s}]|\mathrm{s} = s\right] = \mathbb{E}[\mathrm{z}|\mathrm{s} = s] \tag{3.15}$$

It would be trivial to require that conditioning the instrumental variable z both on random variable $\mathbf{w}$ and a stochastic function of $\mathbf{w}$ denoted by $\mathcal{F}(\mathbf{w}, \varepsilon)$ (given the stochastic component ε is an i.i.d white noise which is independent of z), would be the same as conditioning only on $\mathbf{w}$. Formally:

$$\mathbb{E}[\mathrm{z}|\mathbf{w} = \boldsymbol{w}, \mathcal{F}(\mathbf{w}, \varepsilon)] = \mathcal{G}(\mathbf{w}) \tag{3.16}$$

where $\mathcal{G}(\cdot)$ is some function of w.

The indicator (selection variable) $\mathrm{s} = I(\xi_{2i} > -\mathbf{w}^T\boldsymbol{\gamma})$ is a stochastic function of $\mathbf{w}$, thus, $\mathbb{E}[\mathrm{z}|\mathrm{s}, \mathbf{w} = \boldsymbol{w}] = \mathbb{E}[\mathrm{z}|\mathbf{w} = \boldsymbol{w}]$ using the requirement in (3.16).

In Theorem 2 to follow we present our primary argument: in truncated sample selection models, the orthogonality condition of the instrumental variable with respect to the ran-

[18] Not to be confused with its realization $\boldsymbol{w_i}$.

[19] The intrinsic model's endogeneity is related to the joint dependence of the random disturbance and the covariates *in the population*, unlike a conditional joint dependence of the random disturbance and the covariates given participation in the sample.

dom disturbance might be violated. This violation stems from a dependency between the instrumental variables and the selection equation's covariates.

Theorem 2 *Let ξ_1 and ξ_2 be two jointly distributed random disturbances, and let* z *be a valid instrumental variable satisfying* $\mathbb{E}[\mathrm{z} \cdot \xi_1] = 0$. *Denote a random variables vector* $\mathbf{w} \in \mathbb{R}^l$, *a parameters vector* $\boldsymbol{\gamma} \in \mathbb{R}^l$ *and a truncated environment using the indicator variable* $\mathrm{s} = I(\xi_2 > -\mathbf{w'}\boldsymbol{\gamma})$.

Assume the following conditions are satisfied: (i) $\mathbb{E}[\xi_1|\mathrm{s} = s, \mathbf{w} = \boldsymbol{w}] = \mathcal{M}(\boldsymbol{w}^T\boldsymbol{\gamma})$; *(ii)* $\mathbb{E}[\mathrm{z}|\mathbf{w} = \boldsymbol{w}, \mathrm{s} = s] = \mathbb{E}[\mathrm{z}|\mathbf{w} = \boldsymbol{w}] = \mathcal{G}(\boldsymbol{w})$; *(iii)* z *and* ξ_1 *are conditionally independent given* $\mathbf{w}$ *and* s; *(v)* $\mathcal{G}$ *and* $\mathcal{M}$ *are linearly dependent in the truncated environment (given* s*).*[20] *Under conditions (i)-(v),* z *is not orthogonal to the random disturbance* ξ_1 *given* s.

Proof: Using the tower property depicted in (3.15), the following must hold:

$$\mathbb{E}[\mathrm{z}\xi_1|\mathrm{s} = s] = \mathbb{E}_{\mathbf{w}}\left[\mathbb{E}[\mathrm{z}\xi_1|\mathbf{w}, \mathrm{s}]|\mathrm{s} = s\right] = \underbrace{\mathbb{E}_{\mathbf{w}}\left[\mathbb{E}_{\mathrm{z}}[\mathrm{z}|\mathbf{w}, \mathrm{s}]\mathbb{E}_{\xi_1}[\xi_1|\mathbf{w}, \mathrm{s}]|\mathrm{s} = s\right]}_{\text{by conditional independence of z and } \xi_1 \text{ given } \mathbf{w} \text{ and s}}$$

$$= \mathbb{E}_{\mathbf{w}}[\mathcal{G}\mathcal{M}|\mathrm{s} = s] = \int_{\boldsymbol{w}} \mathcal{G}(\boldsymbol{w})\mathcal{M}(\boldsymbol{w}^T\boldsymbol{\gamma}) f_{\mathbf{w}|\mathrm{s}=s}(\boldsymbol{w}|\mathrm{s} = s)d\boldsymbol{w}$$

Similarly,

$$\mathbb{E}[\mathrm{z}|\mathrm{s} = s] = \mathbb{E}_{\mathbf{w}}\left[\mathbb{E}[\mathrm{z}|\mathbf{w}, \mathrm{s}]|\mathrm{s} = s\right] = \mathbb{E}_{\mathbf{w}}[\mathcal{G}|\mathrm{s} = s] = \int_{\boldsymbol{w}} \mathcal{G}(\boldsymbol{w}) f_{\mathbf{w}|\mathrm{s}=s}(\boldsymbol{w}|\mathrm{s} = s)d\boldsymbol{w}$$

and

$$\mathbb{E}[\xi_1|\mathrm{s} = s] = \mathbb{E}_{\mathbf{w}}\left[\mathbb{E}[\xi_1|\mathbf{w}, \mathrm{s}]|\mathrm{s} = s\right] = \mathbb{E}_{\mathbf{w}}[\mathcal{M}|\mathrm{s} = s] = \int_{\boldsymbol{w}} \mathcal{M}(\boldsymbol{w}^T\boldsymbol{\gamma}) f_{\mathbf{w}|\mathrm{s}=s}(\boldsymbol{w}|\mathrm{s} = s)d\boldsymbol{w}$$

Thus, since $\mathcal{G}$ and $\mathcal{M}$ are conditionally linearly dependent random variables in the truncated environment (given s), implies that:

$$\int \mathcal{G}(\boldsymbol{w})\mathcal{M}(\boldsymbol{w}^T\boldsymbol{\gamma}) f_{\mathbf{w}|\mathrm{s}=s}(\boldsymbol{w})d\boldsymbol{w} \neq \int \mathcal{G}(\boldsymbol{w}) f_{\mathbf{w}|\mathrm{s}=s}(\boldsymbol{w})d\boldsymbol{w} \int \mathcal{M}(\boldsymbol{w}^T\boldsymbol{\gamma}) f_{\mathbf{w}|\mathrm{s}=s}(\boldsymbol{w})d\boldsymbol{w}$$

[20] $\int \mathcal{G}(\boldsymbol{w})\mathcal{M}(\boldsymbol{w}^T\boldsymbol{\gamma}) f_{\mathbf{w}|\mathrm{s}=s}(\boldsymbol{w})d\boldsymbol{w} \neq \int \mathcal{G}(\boldsymbol{w}) f_{\mathbf{w}|\mathrm{s}=s}(\boldsymbol{w})d\boldsymbol{w} \int \mathcal{M}(\boldsymbol{w}^T\boldsymbol{\gamma}) f_{\mathbf{w}|\mathrm{s}=s}(\boldsymbol{w})d\boldsymbol{w}$.

and consequently:

$$\mathbb{E}[\mathrm{z}\xi_1|\mathrm{s}=s] \neq \mathbb{E}[\mathrm{z}|\mathrm{s}=s]\mathbb{E}[\xi_1|\mathrm{s}=s] \Rightarrow \mathbb{COV}[\mathrm{z},\xi_1|\mathrm{s}=s] \neq 0 \quad \blacksquare$$

Therefore, z is not orthogonal to ξ_1 given s (in the truncated environment).

Theorem 3 *Removing the contamination factor (the bias term) from the residual in the truncated environment, leads to orthogonality of the instrumental variable to the substantive's equation disturbance, such that:* $\mathbb{E}[\mathrm{z}\left[\xi_1 - \mathcal{M}(\mathbf{w}^{\mathrm{T}}\boldsymbol{\gamma})\right]|\mathrm{s}=s] = 0$.

Proof: Express $\mathbb{E}[\mathrm{z}\left[\xi_1 - \mathcal{M}(\mathbf{w}^{\mathrm{T}}\boldsymbol{\gamma})\right]|\mathrm{s}=s]$ as a difference of two conditional expectations:

$$\mathbb{E}[\mathrm{z}\left[\xi_1 - \mathcal{M}(\mathbf{w}^{\mathrm{T}}\boldsymbol{\gamma})\right]|\mathrm{s}=s] = \mathbb{E}[\mathrm{z}\xi_1|\mathrm{s}=s] - \mathbb{E}[\mathrm{z}\mathcal{M}(\mathbf{w}^{\mathrm{T}}\boldsymbol{\gamma})|\mathrm{s}=s]$$

Using the Tower property depicted in (3.15), to get:

$$\mathbb{E}[\mathrm{z}\mathcal{M}(\mathbf{w}^{\mathrm{T}}\boldsymbol{\gamma})|\mathrm{s}=s] = \mathbb{E}_{\mathbf{w}}\left[\mathbb{E}[\mathrm{z}\mathcal{M}(\mathbf{w}^{\mathrm{T}}\boldsymbol{\gamma})|\mathbf{w},\mathrm{s}]|\mathrm{s}=s\right]$$

$$= \underbrace{\mathbb{E}_{\mathbf{w}}\left[\mathbb{E}_{\mathrm{z}}[\mathrm{z}|\mathbf{w},\mathrm{s}]\mathbb{E}[\mathcal{M}(\mathbf{w}^{\mathrm{T}}\boldsymbol{\gamma})|\mathbf{w},\mathrm{s}]|\mathrm{s}=s\right]}_{\text{by conditional independence of z and } \mathcal{M}(\mathbf{w}^T\gamma) \text{ given } \mathbf{w} \text{ and s}} = \mathbb{E}[\mathcal{G}(\mathbf{w})\mathcal{M}(\mathbf{w}^{\mathrm{T}}\boldsymbol{\gamma})|\mathrm{s}=s]$$

But since $\mathbb{E}[\mathrm{z}\xi_1|\mathrm{s}=s] = \mathbb{E}[\mathcal{G}(\mathbf{w})\mathcal{M}(\mathbf{w}^{\mathrm{T}}\boldsymbol{\gamma})|\mathrm{s}=s]$ (proof of Theorem 2), implies that $\mathbb{E}[\mathrm{z}\left[\xi_1 - \mathcal{M}(\mathbf{w}^{\mathrm{T}}\boldsymbol{\gamma})\right]|\mathrm{s}=s] = 0$.

Moreover,

$$\mathbb{COV}\left[\mathrm{z},\xi_1 - \mathcal{M}(\mathbf{w}^T\boldsymbol{\gamma})|\mathrm{s}=s\right]$$

$$= \underbrace{\mathbb{E}[\mathrm{z}\left[\xi_1 - \mathcal{M}(\mathbf{w}^{\mathrm{T}}\boldsymbol{\gamma})\right]|\mathrm{s}=s]}_{0} - \mathbb{E}[\mathrm{z}|\mathrm{s}=s]\underbrace{\mathbb{E}[\xi_1 - \mathcal{M}(\mathbf{w}^{\mathrm{T}}\boldsymbol{\gamma})|\mathrm{s}=s]}_{0} = 0. \quad \blacksquare$$

Therefore, a valid instrumental variable z is orthogonal to the truncated distribution (non-contaminated) disturbance ϵ_{1i} in (3.14), even though z and $\mathbf{w}$ are dependent.

The complete data $\{\boldsymbol{x_i},\boldsymbol{z_i},\boldsymbol{w_i},y_{1i},y_{2i}\}$ is a random sample drawn from $\{\mathbf{x_i},\mathbf{z_i},\mathbf{w_i},\mathrm{y}_{1\mathrm{i}},\mathrm{y}_{2\mathrm{i}}\}$, while the observed data is a subsample consisting of observations which satisfy the selection equation.

The joint dependence of $(\xi_1, \xi_2, \mathrm{v})$ implies the violation of zero mean expectation (under truncation) in the x_{1i} regression equation (3.4) such that $\mathbb{E}[\mathrm{v}|\xi_2 > -\boldsymbol{w'\gamma}] = \mathcal{M}_2(\boldsymbol{w}^T\boldsymbol{\gamma})$. This violation is a precondition for the endogeneity of $\{\boldsymbol{x_{-1i}}, \boldsymbol{z_i}\}$ with respect to $\mathrm{v_i}$ given participation in the regression of x_{1i}.[21] The following Theorem argues that such violation is also obtained in cases where v and ξ_2 are conditionally independent given ξ_1 implying that the co-movement of v and ξ_2 is entirely related to a variation in ξ_1.

Theorem 4 *Let ξ_1 and ξ_2 be two jointly distributed random disturbances of the substantive and selection equations, respectively. Let* v *be a random variable which depends on ξ_1 such that* v *and ξ_2 are conditionally independent given ξ_1. Denote a random variables vector $\mathbf{w} \in \mathbb{R}^l$ independent of $(\xi_1, \xi_2, \mathrm{v})$ with a realization $\boldsymbol{w}$, a parameters vector $\boldsymbol{\gamma} \in \mathbb{R}^l$ and a truncated environment using the indicator variable $\mathrm{s} = I(\xi_2 > -\boldsymbol{w'\gamma})$.*

Assume the following conditions are satisfied: (i) the conditional expectation of the random disturbance given participation is $\mathbb{E}[\xi_1|\xi_2 > -\boldsymbol{w'\gamma}] = \mathcal{M}_1(\boldsymbol{w}^T\boldsymbol{\gamma})$, Robinson (1988); (ii) $\mathbb{E}[\mathrm{v}|\xi_1, \xi_2 > -\boldsymbol{w'\gamma}] = \mathbb{E}[\mathrm{v}|\xi_1] = \mathcal{H}(\xi_1)$, (endogniety); (iii) $\mathcal{H}(.)$, a monotonic mapping $\mathbb{R} \mapsto \mathbb{R}$. Under conditions (i)-(iii), $\mathbb{E}[\mathrm{v}|\xi_2 > -\boldsymbol{w'\gamma}] \neq \mathbb{E}[\mathrm{v}]$ regardless of the conditional independence of v *and ξ_2 given ξ_1.*

Proof: Applying Tower property to $\mathbb{E}[\mathrm{v}|\mathrm{s} = 1]$:

$$\mathbb{E}[\mathrm{v}|\xi_2 > -\boldsymbol{w'\gamma}] = \mathbb{E}[\mathrm{v}|\mathrm{s} = 1] = \mathbb{E}_{\xi_1}\{\mathbb{E}[\mathrm{v}|\xi_1, \mathrm{s}]|\mathrm{s} = 1\} = \mathbb{E}[\mathcal{H}(\xi_1)|\mathrm{s} = 1] = \mathcal{M}_2(\boldsymbol{w}^T\boldsymbol{\gamma}) \neq \mathbb{E}[\mathrm{v}].$$

It can be shown that ξ_1 mediates between v and s (participation), in that it generates a co-movement between the random variables v and s. The last equality relies on the fact that the random variable $\mathcal{H}(\xi_1)$ is a monotonic mapping of ξ_1, implying dependence on s due to the dependency between ξ_1 and s. ■

Next we show that the conventional *IV* estimator is inconsistent in the presence of a

[21] As been discussed in Heckman (1979), the fact that the conditional disturbance (given participation) in the the substantive equation of x_{1i} is a function of the selection equation's covariates, leads to a potential correlation between the disturbance and the substantive equation's covariates. This correlation implies the endogeneity of the substantive equation's covariates $\{\boldsymbol{x_{-1i}}, \boldsymbol{z_i}\}$ with respect to its random disturbance $\mathrm{v_i}$ given participation.

truncated environment in which both the instrumental variable's as well as the random disturbance's expectation are function of the selection equation's covariates vector $\mathbf{w}$. The proof in (3.4) to follow relies on a linear dependence assumption between these two functions of $\mathbf{w}$. The rational for the linear dependence is due to the fact that the random disturbance's (ξ_1) conditional expectation generally satisfies a monotonicity with respect to the index variable $\mathbf{w}'\boldsymbol{\gamma}$. Therefore, it is enough to assume that on average, z is affected monotonically by the index variable $\mathbf{w}'\boldsymbol{\gamma}$ to generate a linear dependence between z and the expectation of ξ_1.[22]

3.4. The conventional IV estimator's Asymptotic bias

The IV estimator's asymptotic bias is:

$$\widehat{\boldsymbol{\beta}}_{\mathbf{iv}} = (\mathbf{z}^T\mathbf{x})^{-1}\mathbf{z}^T\mathbf{y}_1 = (\mathbf{z}^T\mathbf{x})^{-1}\mathbf{z}^T(\mathbf{x}\boldsymbol{\beta} + \mathcal{M}_1(\mathbf{w}^T\boldsymbol{\gamma}) + \epsilon_{1i}^{**}) \tag{3.17}$$

$$\widehat{\boldsymbol{\beta}}_{\mathbf{iv}} = (\mathbf{z}^T\mathbf{x})^{-1}\mathbf{z}^T(\mathbf{x}\boldsymbol{\beta}) + (\mathbf{z}^T\mathbf{x})^{-1}\mathbf{z}^T\mathcal{M}_1(\mathbf{w}^T\boldsymbol{\gamma}) + (\mathbf{z}^T\mathbf{x})^{-1}\mathbf{z}^T\epsilon_{1i}^{**} \tag{3.18}$$

$$\widehat{\boldsymbol{\beta}}_{\mathbf{iv}} = \boldsymbol{\beta} + (\mathbf{z}^T\mathbf{x})^{-1}\mathbf{z}^T\mathcal{M}_1(\mathbf{w}^T\boldsymbol{\gamma}) + (\mathbf{z}^T\mathbf{x})^{-1}\mathbf{z}^T\epsilon_{1i}^{**} \tag{3.19}$$

$$\underset{N\to\infty}{\text{plim}}\left[\widehat{\boldsymbol{\beta}}_{\mathbf{iv}}\right] = \boldsymbol{\beta} + \underbrace{\underset{n\to\infty}{\text{plim}}\left[(N^{-1}\mathbf{z}^T\mathbf{x})^{-1}\right]\underset{N\to\infty}{\text{plim}}\left[N^{-1}\mathbf{z}^T\mathcal{M}_1(\mathbf{w}^T\boldsymbol{\gamma})\right]}_{\text{Asymptotic bias}} \tag{3.20}$$

Given any correlation between $\mathbf{z}$ and $\mathcal{M}_1(\mathbf{w}^T\boldsymbol{\gamma})$, $\underset{n\to\infty}{\text{plim}}\left[\mathbf{z}^T\mathcal{M}_1(\mathbf{w}^T\boldsymbol{\gamma})\right] \not\to 0$. Thus, the $\widehat{\boldsymbol{\beta}}_{\mathbf{iv}}$ estimator is an inconsistent estimator for $\boldsymbol{\beta}$.

Next we discuss the joint dependence of the covariates and the substantive equations' random disturbances in the truncated data.

3.5. The instrument variable estimator: truncated sample

Denote the truncated data by a sequence of observations $\{y_{1i}, \boldsymbol{x}_i, \boldsymbol{w}_i, \boldsymbol{z}_i\}_{i=1}^n$, such that each observation is an independent realization of the conditional joint distribution function

[22] Both functions are dependent through $\mathbf{w}$ by construction, generally leading to some degree of linear dependence.

of the random variables $\{\mathrm{y}_1, \mathbf{x}, \mathbf{w}, \mathbf{z}\}$ given they are selected into the sample ($\mathrm{y}_2 = 1$). The endogenous random variable is denoted by x_1 and is included in vector $\mathbf{x}$. At least one of the exogenous and endogenous parts of the random variable $\mathrm{x_i}$ depends on the conditional random variable $\xi_{1i}|\xi_{2i} \geq -\mathbf{w}_i^T\boldsymbol{\gamma}$ due to a co-movement between these two random variable with respect to $\mathbf{w}_i^T\boldsymbol{\gamma}$. Moreover, only the endogenous part of $\mathrm{x_i}$ depends on the ξ_{1i} due to a co-movement between these two random variable with respect to $\mathrm{v_i}$. This implies there are two sources of endogeneity to be taken into consideration: the first source is related to the endogenous covariate, while the second source is due to the truncation environment of the data.

Next we present our proposed estimator using two different trigonometric series: Cosine and Fourier.

3.6. Transformation of both Cosine and Fourier series for unknown functions estimation

The functions $\mathcal{M}_1(.)$ and $\mathcal{M}_2(.)$ in (3.14) are approximated using their conditional moment expansion by employing either Cosine or Fourier sequence. The Cosine sequence requires that the support of the index variable in (3.14) will be on the $[0, 1]$ domain, while the Fourier sequence requires that the support will be on the $[-1, 1]$ domain.[23] This requirement does not entail loss of generality, because it is satisfied by utilizing a different monotone transformation function on the index variable (Horowitz, 2014) in each one of the Cosine and Fourier series. The series generated by the transformation is referred to as a transformed Cosine (or Fourier) series.

We express the conditional moments expressions by the functions $\mathcal{M}_1(\cdot)$ and $\mathcal{M}_2(\cdot)$:[24]

$$\mathbb{E}[\xi_1|\xi_2 > -\boldsymbol{w}_i^T\boldsymbol{\gamma}] \equiv \mathcal{M}_1(\boldsymbol{w}_i^T\boldsymbol{\gamma}), \qquad \mathbb{E}[\mathrm{v}|\xi_2 > -\boldsymbol{w}_i^T\boldsymbol{\gamma}] \equiv \mathcal{M}_2(\boldsymbol{w}_i^T\boldsymbol{\gamma}) \tag{3.21}$$

[23]Fourier series decomposes a **periodic** signal into a sum of an infinite number of harmonics (sine and cosine functions) of different frequencies and and amplitudes, while Fourier transform decomposes a **non-periodic** signal into an infinite number of harmonics having different frequencies and amplitudes.

[24]Unlike, the two stage estimation procedure for partially linear single index models in Zhou et al. (2016) in which the instrumental variable equation is estimated by a linear regression, we deal with an endogenous truncated data set consisting of a system of truncated substantive equations. Due to the truncation, each of the equations in this system is modeled as a partially linear single index model. We also depart from the procedure suggested in Zhou et al. (2016) by employing a Sieve estimator instead of a kernel.

Utilizing the Fourier cosine sequence we substitute the functions $\mathcal{M}_1(\cdot)$ and $\mathcal{M}_2(\cdot)$ with their respective conditional moment expansions $\widehat{\mathcal{M}}_1^c(\cdot)$ and $\widehat{\mathcal{M}}_2^c(\cdot)$, defined as:

$$\widehat{\mathcal{M}}_j^c(\boldsymbol{w_i};\boldsymbol{\theta_{jc}}) = \alpha_{jc} + \sum_{k=1}^{K} \delta_{jk}^c \cos(\rho(\boldsymbol{w}_i^T\boldsymbol{\gamma})\pi k), \quad j = 1,2 \tag{3.22}$$

where $\rho(.)$ is some known (arbitrarily chosen) monotonic twice differentiable mapping $\mathbb{R} \mapsto (0,1)$, $\boldsymbol{\theta_{jc}} \equiv \left\{\alpha_{jc}, \boldsymbol{\gamma}, \boldsymbol{\delta_j^c}\right\}$ and $\boldsymbol{\delta_j^c} \equiv [\delta_{j1}^c, ..., \delta_{jK}^c]$, with K being the number of elements in the expansion.

Similarly, utilizing the Fourier sequence we substitute the functions $\mathcal{M}_1(\cdot)$ and $\mathcal{M}_2(\cdot)$ with their respective conditional moment expansions $\widehat{\mathcal{M}}_1^f(\cdot)$ and $\widehat{\mathcal{M}}_2^f(\cdot)$, defined as:

$$\widehat{\mathcal{M}}_j^f(\boldsymbol{w_i};\boldsymbol{\theta_{jf}}) = \alpha_{jf} + \sum_{k=1}^{K} \delta_{1j_k}^f \cos(\lambda(\boldsymbol{w}_i^T\boldsymbol{\gamma})\pi k_1) + \sum_{k=1}^{K} \delta_{2j_k}^j \sin(\lambda(\boldsymbol{w}_i^T\boldsymbol{\gamma})\pi k_2), \quad j = 1,2 \tag{3.23}$$

where $\lambda(.)$ is some known (arbitrarily chosen) monotonic twice differentiable mapping $\mathbb{R} \mapsto (-1,1)$, $\boldsymbol{\theta_{jf}} \equiv \left\{\alpha_{jf}, \boldsymbol{\gamma}, \boldsymbol{\delta_{1j}^f}, \boldsymbol{\delta_{2j}^f}\right\}$ and $\boldsymbol{\delta_{mj}^f} \equiv [\delta_{mj_1}^f, ..., \delta_{jm_K}^f]$, $m = 1,2$ representing Sine or Cosine respectively.

Given the non-linear function $\tilde{\mathcal{M}}_j^G(\boldsymbol{w_i};\boldsymbol{\theta_{jG}})$ with $G \in \{c, f\}$ for $j = 1,2$ an index model can be estimated (Racine et al., 2014):

$$(\widehat{\boldsymbol{\eta_j}}, \widehat{\boldsymbol{\theta_{jG}}}) = \arg \min_{(\boldsymbol{\eta_j},\boldsymbol{\theta_{jG}})\in\Theta\times\Delta_K} \frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{y}_{ji} - \boldsymbol{\chi}_{ji}^T\boldsymbol{\eta}_j - \tilde{\mathcal{M}}_j^G(\boldsymbol{w_i};\boldsymbol{\theta_{jG}})\right)^2 \tag{3.24}$$

where $\boldsymbol{y}_{ji}$ is the j'th equation's dependent variable; K is the number of elements in the expansion; $\boldsymbol{\chi_{ji}}$ and $\boldsymbol{\eta_j}$ represent the covariates set and the parameter set, respectively in the linear part of the j'th equation.

Given that the expectation of the objective function in (3.24) is finite for all values of the parameters $(\boldsymbol{\eta_j}, \boldsymbol{\theta_{jG}})$[25]

$$\mathbb{E}\left[\sup_{(\boldsymbol{\eta_j},\boldsymbol{\theta_{jG}})\in\Theta\times\Delta_K} \frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{y}_{ji} - \boldsymbol{\chi}_{ji}^T\boldsymbol{\eta}_j - \tilde{\mathcal{M}}_j^G(\boldsymbol{w_i};\boldsymbol{\theta_{jG}})\right)^2\right] < \infty \tag{3.25}$$

[25]This assumption can be relaxed using a positive weight function $\mathcal{K}(x)$ on $(0,\infty)$ in the nonlinear minimization (see, Racine et al. (2014)).

The objective function to be estimated is some unknown non-linear function of the index variable (Racine et al., 2014). However, in our implementation we will use a partially linear index function, due to the linearity of the substantive equation with respect to its covariates and the non-linearity of the bias term function.

Next we discuss the two steps estimation procedure to be employed for the correction of both endogeneity and truncation bias propagated by truncation.

3.7. The estimation procedure

In this section we introduce a two step estimation procedure to eliminate the two sources of bias discussed. To eliminate the endogeneity bias term we adapt a similar approach to the two step procedure in Zhou et al. (2016) for a partially linear single index model estimation in which the first stage is a regression of the endogenous covariate on all the exogenous covariates and the instrumental variable. In the second stage the endogenous covariate is substituted with the fitted values obtained from the first stage. However, the estimation approach in Zhou et al. (2016) cannot be implemented in truncated environment since it treats the first stage regression as a linear population regression (as if the entire covariates distribution function is observed). We alleviate this by modeling both the first as well as the second stage equations as endogenously truncated equations. In order to eliminate the endogenous truncation bias we control for this source of bias by including the truncation bias term as an additional covariate in the substantive equations as depicted in (3.14). Thus, the partial linearity is applied to both the first as well as the second stage equations.

In the first stage we regress the endogenous covariate on the instrumental and exogenous variables, by minimizing the partially linear index model:

$$(\widehat{\boldsymbol{\delta}}, \widehat{\boldsymbol{\theta_{1f}}}) = \underset{(\boldsymbol{\delta}, \boldsymbol{\theta_{1f}}) \in \Theta \times \Delta_K}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} \Big(x_{1i} - \left[\boldsymbol{x}_{-1_i}^T, \boldsymbol{z}_i^T \right] \boldsymbol{\delta} - \widehat{\mathcal{M}}_1(\boldsymbol{w_i}; \boldsymbol{\theta_{1f}}) \Big)^2 \tag{3.26}$$

In the second stage, the endogenous variable is replaced by its predicted value obtained from the first stage in (3.26), and we minimize the following function:

$$(3.27)\quad (\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta_{2f}}}) = \underset{(\beta,\boldsymbol{\theta_{2f}})\in\Theta\times\Delta_K}{\arg\min} \frac{1}{n}\sum_{i=1}^{n}\Big(y_{1i} - \left[\widehat{x_{1i}}, \boldsymbol{x}_{-1_i}^T\right]\boldsymbol{\beta} - \widehat{\mathcal{M}}_2(\boldsymbol{w_i}; \boldsymbol{\theta_{2f}})\Big)^2$$

As can be seen in (3.27) the two sources of endogeneity bias are dealt with: (i) the bias propagated by the endogenous covariate is alleviated by utilizing the covariate set $\left[\widehat{x_{1i}}, \boldsymbol{x}_{-1_i}^T\right]$ consisting entirely of exogenous covariates; (ii) the bias propagated by the endogenous truncation is alleviated by controlling for the selection bias term.

Next we present Monte Carlo simulation to examine our semi-parametric IV estimator's performance in a truncated environment.

4. Simulation

In this section we generate multiple random data sets to be used for the examination of our model's performance using different sample sizes.

First, we discuss the procedure for the data generation process (DGP).

4.1. Data generation process

Denote the sample size by $N \in \{500, 2000, 3000, 5000, 8000, 10000\}$. In order to not restrict the data generation process to the family of symmetric unimodal distribution functions, a mixture of distribution functions is utilized to generate each of the selection model's disturbances which are jointly dependent (as will be discussed in section 4.1.1 to follow). In order to verify that our proposed model performs well under different data generating processes (DGP), we construct a data set consisting of 2,000,000 distribution functions,[26] practically generating 100 millions realizations which are not i.i.d. By construction, each observation is randomly drawn from a unique mixture of distribution functions.

[26]The estimates obtained given the various data distribution functions will be supplied upon request.

4.1.1. The disturbances' joint distribution function

Each triple of disturbances $\{\xi_{1i}, \xi_{2i}, \mathrm{v}_i\}$ is randomly and independently drawn from $F_{\xi_1,\xi_2,\mathrm{v}}$, which is the substantive and participation equations' disturbances joint distribution function. The aforementioned joint density function consists of two components: a Copula function[27] characterizing the disturbances' dependence structure and three marginal distribution functions F_{ξ_1}, F_{ξ_2} and F_{v}. In order to verify our model's performance in the presence of random disturbances' distribution functions which are not restricted to the family of symmetric and unimodal distribution functions, each one of the sample selection model's disturbances ξ_1 and ξ_2 is marginally distributed according to a mixture of three different distribution functions: (i) a normal distribution function with expectation and standard deviation parameters (μ, σ_a) denoted by $\mathcal{N}(\mu, \sigma_a^2)$; (ii) a normal distribution function with expectation and standard deviation parameters $(-\mu, \sigma_b)$ denoted by $\mathcal{N}(-\mu, \sigma_b^2)$; (iii) a gamma distribution function with scale and shape parameters $(\mu\varphi, \varphi)$ denoted by $\Gamma_{\text{Gamma}}(\mu\varphi, \varphi)$[28]. This mixture distribution function is defined as:

$$(4.1) \qquad \begin{cases} \mathrm{v} \sim \mathcal{N}(0, \sigma_{\mathrm{v}}^2) \\ \xi_{\mathrm{j}} \sim 0.4\mathcal{N}(\mu, \sigma_a^2) + 0.5\mathcal{N}(-\mu, \sigma_b^2) + 0.1\Gamma_{\text{Gamma}}(\mu\varphi, \varphi), \quad j = 1, 2. \end{cases}$$

where $\mathbb{E}[\xi_{\mathrm{j}}] = 0$ and $\mathbb{E}[\mathrm{v}] = 0$.

The parameters set $(\mu, \sigma_a, \sigma_b, \varphi, \sigma_{\mathrm{v}}) = (4, 2.5, 1.5, 2, 1)$ is arbitrarily chosen. Due to its simplicity, the Clayton Copula (as will be discussed in section 4.1.2 to follow) with a degree of dependence parameter is set to equal 1, assuring the disturbances are highly correlated is used for controlling the dependence structure.

Next we employ a function characterizing the dependence properties of the Copula (McNeil and Nešlehová, 2009), referred to as *a generator function* to construct of the joint dependence of the random disturbances in (4.1).

[27] Any continuous joint distribution function can be characterized by a set of marginal distribution functions and a joint distribution function determining the dependence structure which is referred to as a Copula function. (Sklar's Theorem (Sklar, 1959)).

[28] The scale and shape parameters implies the expectation and standard deviation parameters are $(\mu, \sqrt{\mu/\varphi})$, respectively.

4.1.2. Archimedean Copula function

An Archimedean Copula is a Copula characterized with a non-increasing, continuous generator function ψ: $[0, \infty] \rightarrow [0, 1]$ which satisfies $\psi(0) = 1$, $\psi(\infty) = 0$ and is strictly decreasing on $[0, \inf\{t : \psi(t) = 0\}]$. In particular, we are interested in the d dimensional Archemdean Copula family (3 in the present case[29]) which has the simple algebraic form (McNeil and Nešlehová, 2009):[30]

$$\mathcal{C}(u_1, .., u_d) = \psi(\psi^{-1}(u_1), ..., \psi^{-1}(u_d)), \quad (u_1, ..., u_d) \in [0, 1]^d \tag{4.2}$$

where ψ is a specific function known as the generator of $\mathcal{C}$. To generate the disturbances, the Clayton Copula's generator $\psi(t) = (1 + t)^{-1/\theta}$ is chosen.

The covariates random variables vector $[\mathrm{z}, \mathrm{x}_2, \mathrm{w}_1, \mathrm{w}_2]$ is jointly normally distributed with an expectation vector $\boldsymbol{\mu} = [0, 0, 0, 0]^T$ and a covariance matrix $\mathbf{\Sigma_{4\times4}}$:

$$\begin{bmatrix} \mathrm{z} \\ \mathrm{x}_2 \\ \mathrm{w}_1 \\ \mathrm{w}_2 \end{bmatrix} \sim \mathcal{N}_4(\boldsymbol{\mu}, \mathbf{\Sigma_{4\times4}}), \qquad \mathbf{\Sigma_{4\times4}} = \begin{bmatrix} \sigma^2_{\mathrm{z}} & \sigma_{\mathrm{z},\mathrm{x}_2} & \sigma_{\mathrm{z},\mathrm{w}_1} & \sigma_{\mathrm{z},\mathrm{w}_2} \\ \sigma_{\mathrm{z},\mathrm{x}_2} & \sigma^2_{\mathrm{x}_2} & \sigma_{\mathrm{x}_2,\mathrm{w}_1} & \sigma_{\mathrm{x}_2,\mathrm{w}_2} \\ \sigma_{\mathrm{z},\mathrm{w}_1} & \sigma_{\mathrm{x}_2,\mathrm{w}_1} & \sigma^2_{\mathrm{w}_1} & \sigma_{\mathrm{w}_1,\mathrm{w}_2} \\ \sigma_{\mathrm{z},\mathrm{w}_2} & \sigma_{\mathrm{x}_2,\mathrm{w}_2} & \sigma_{\mathrm{w}_1,\mathrm{w}_2} & \sigma^2_{\mathrm{w}_2} \end{bmatrix} \tag{4.3}$$

The arbitrarily chosen covariance matrix is:

$$\mathbf{\Sigma_{4\times4}} = \begin{bmatrix} 1 & 0.4 & 0.8 & -0.6 \\ 0.4 & 1.264 & 0.36 & -0.48 \\ 0.8 & 0.36 & 2 & -0.4 \\ -0.6 & -0.48 & -0.4 & 2 \end{bmatrix} \tag{4.4}$$

We generate the data y_{1i}, y_{2i}, x_{1i} according to the following data generation process (DGP)

[29] $d = 3$ representing the three dimensional vector of random disturbances $(\mathrm{v}_{\mathrm{i}}, \xi_{1i}, \xi_{2i})$.

[30] Knowing the distribution corresponding to a generator ψ, Marshall and Olkin (1988) presented a sampling algorithm for exchangeable Archimedean copulas which does not require the knowledge of the copula density. This algorithm is therefore applicable to large dimensions (Hofert, 2008).

(Escanciano, 2017):

$$\text{(4.5)} \qquad \text{DGP1}: \begin{cases} y_{1i}^{*} = \alpha_1 + \beta_1 x_{1i} + \beta_2 x_{2i} + \xi_{1\mathrm{i}} \\ y_{2i}^{*} = \alpha_2 + \gamma_1 w_{1i} + \gamma_2 w_{2i} + \xi_{2\mathrm{i}} \\ x_{1i}^{*} = \delta_1 z_i + \delta_2 x_{2i} + \mathrm{v_i} \end{cases}$$

where each i element in the sequence $\{x_{2i}, z_i, w_{1i}, w_{2i}\}_{i=1}^{N}$ is an independent realization of the random variables $(\mathrm{x}_2, \mathrm{z}, \mathrm{w}_1, \mathrm{w}_2)$. We choose the parameter setting $[\alpha_1, \alpha_2, \beta_1, \beta_2, \delta_1, \delta_2, \gamma_1, \gamma_2] = [2, 0.5, 1, 1.25, 0.5, 1, 2, -1]$.

The truncated data set is characterized by the following equations:

$$\text{(4.6)} \qquad \begin{bmatrix} y_{1i} \\ x_{1i} \end{bmatrix} = \begin{cases} \begin{bmatrix} \alpha_1 + \beta_1 x_{1i} + \beta_2 x_{2i} + \xi_{1\mathrm{i}} \\ \delta_1 z_i + \delta_2 x_{2i} + \mathrm{v_i} \end{bmatrix} & \text{if } y_{2i}^{*} \geq 0 \\ \text{Unobserved} & \text{if } y_{2i}^{*} < 0 \end{cases},$$

where x_{1i} be an endogenous variable included in vector $\boldsymbol{x}_i \in \mathbb{R}^p$, in which all the elements (except for x_i) are exogenous variables and $\boldsymbol{\beta} \in \mathbb{R}^p$ is a covariates vector. The substantive equation's random disturbance is denoted by ξ_{1i}.

4.2. Simulations result

We have randomly generated for each sample size $N \in \{500, 2000, 3000, 5000, 8000, 10000\}$, $5,000$ data sets using the data generation process elaborated on in 4.1. For a given number of observations N, different models are estimated: (i) an *OLS* estimator utilizing sample consisting of random realizations from the complete distribution function without correcting for the endogeneity of x_{1i} covariate; (ii) a conventional *IV* estimator correcting for the endogeneity of x_{1i} covariate using the aforementioned entire distribution function; (iii) a conventional *IV* estimator correcting for the endogeneity of x_{1i} covariate applied to a truncated portion of the data distribution function consisting of participants only (without correcting for the self-selection bias); (v) truncated sample model's estimates using the developed *Sieve* (SPIV), correcting for both truncation as well as endogeneity biases.

Table I presents summary statistics of models (i)'s and (ii)'s estimates, while table II presents summary statistics of models (iii)'s and (v)'s estimates. In table III different convergence measures of these estimates are presented.

TABLE I

Monte Carlo Simulation - Non-truncated (complete) data set

True Parameter[a]	Estimate	Model Setup					
		Sample size					
		500	2000	3000	5000	8000	10000
Full sample OLS estimates							
$\beta_1 = 1$	Mean	2.6783	2.6791	2.6799	2.6818	2.6803	2.6814
	Median	2.6820	2.6798	2.6796	2.6817	2.6811	2.6807
	Std	0.1573	0.0774	0.0629	0.0486	0.0383	0.0348
$\beta_2 = 1.25$	Mean	-0.6960	-0.6952	-0.6957	-0.6977	-0.6964	-0.6982
	Median	-0.7005	-0.6955	-0.6975	-0.6971	-0.6971	-0.6977
	Std	0.2434	0.1206	0.0983	0.0761	0.0602	0.0543
Full sample conventional IV's estimates							
$\beta_1 = 1$	Mean	0.9722	0.9944	0.9967	1.0003	0.9996	1.0018
	Median	0.9862	1.0011	0.9988	1.0016	1.0020	1.0037
	Std	0.4392	0.2139	0.1754	0.1341	0.1064	0.0966
$\beta_2 = 1.25$	Mean	1.2806	1.2558	1.2537	1.2499	1.2503	1.2472
	Median	1.2611	1.2467	1.2517	1.2491	1.2470	1.2458
	Std	0.5457	0.2613	0.2173	0.1650	0.1320	0.1193

Note: [a] The parameters that are used in the data generation process.
We estimate by ordinary least squares (OLS) method the parameters for the full sample and truncated sample sample without correction for the selectivity bias, and compute the standard deviation in every random sample consisting of N observations. Then, we calculate for these estimates the mean, median and standard deviation (Std.) over all data sets. The standard deviations are obtained using the estimates from the Monte-Carlo simulations.

Entires in table I indicate that regardless of the sample size the means of the OLS estimates is biased, such that $\beta_1 = 2.68$ and $\beta_2 = -0.69$, while the mean of the full sample IV's estimates are $\beta_1 = 0.996$ and $\beta_2 = 1.25$. The standard deviation obtained for β_1 (the endogenous covariate's coefficient) using the IV estimator is almost 3 times larger than in the OLS estimator and decreases from 0.4392 to 0.0966 when the sample size increases from 500 to $10,000$ observations.

TABLE II

Monte Carlo Simulation - Truncated data set

True Parameter[a]	Estimate	Model Setup					
		Sample size					
		500	2000	3000	5000	8000	10000
Truncated sample conventional IV's estimates							
$\beta_1 = 1$	Mean	0.1085	0.1695	0.1772	0.1800	0.1827	0.1873
	Median	0.1619	0.1815	0.1865	0.1826	0.1893	0.1908
	Std	0.7627	0.3538	0.2949	0.2254	0.1793	0.1646
$\beta_2 = 1.25$	Mean	2.0885	2.0240	2.0172	2.0137	2.0105	2.0058
	Median	2.0199	2.0081	2.0031	2.0101	2.0066	2.0025
	Std	0.8886	0.4138	0.3454	0.2647	0.2100	0.1924
Truncated sample model's estimates (First stage *Sieve* SPIV-NLS)							
$\delta_1 = 0.5$	Mean	0.4725	0.4815	0.4857	0.4917	0.4968	0.4986
	Median	0.4716	0.4809	0.4863	0.4927	0.4979	0.4992
	Std	0.0734	0.0401	0.0350	0.0274	0.0209	0.0177
$\delta_2 = 1$	Mean	0.9957	0.9975	0.9979	0.9989	0.9993	0.9999
	Median	0.9957	0.9971	0.9978	0.9988	0.9993	0.9998
	Std	0.0536	0.0264	0.0214	0.0166	0.0131	0.0117
Truncated sample model's estimates (Second stage *Sieve* SPIV-NLS)							
$\beta_1 = 1$	Mean	1.0636	1.0088	1.0089	0.9997	1.0028	1.0030
	Median	1.0831	1.0181	1.0215	1.0074	1.0052	1.0057
	Std	0.7498	0.3782	0.3100	0.2290	0.1779	0.1605
$\beta_2 = 1.25$	Mean	1.1675	1.2352	1.2387	1.2486	1.2462	1.2458
	Median	1.1442	1.2257	1.2286	1.2408	1.2445	1.2451
	Std	0.8297	0.4153	0.3423	0.2566	0.2002	0.1804

Note: [a] The parameters that are used in the data generation process.
We estimate by ordinary least squares (OLS) method the parameters for the full sample and truncated sample sample without correction for the selectivity bias, and compute the standard deviation in every random sample consisting of N observations. Then, we calculate for these estimates the mean, median and standard deviation (Std.) over all data sets. The standard deviations are obtained using the estimates from the Monte-Carlo simulations.

Entires[31] in table II indicate that regardless of the sample size the means of the truncated sample IV's estimates are biased (ranges from five to ten folds difference) compared to the estimate that would have been emerged. Note that the parameters estimates hardly improve their accuracy as sample size increases. This is due to the presence of two bias sources. The mean estimate of β_1 (the endogenous covariate's parameter) obtained from implementing our proposed methodology, basically mimics the results obtained using a random sample from the entire data distribution function for sample sizes above $2,000$ observations. The standard deviations of this estimate for sample sizes of 500 and $10,000$ observations are 0.7498 and

[31]For sake of brevity we have omitted the estimates of the nuisance parameters which can be furnished upon request.

0.1605, respectively. While, for sample size of $5,000$ observation (or above), the mean estimate of β_2 (the exogenous covariate's parameter) approximates the estimate obtained by employing the conventional IV using a random sample from the entire data distribution function. This estimate is biased even for $10,000$ observation.

We conduct sensitivity test to measure the influence of an increase in number of observations on the accuracy of the truncated sample's estimates.

The first accuracy measure we use is the standardized root mean square error, $RMSE_j$, to measure the bias in the truncated regression estimate relative to its true parameter value that would have been obtained in an un-truncated distribution, defined as:

$$RMSE_j\left(\Omega\right)=\left(\frac{1}{\Omega}\sum_{i=1}^{\Omega}\left(\frac{\hat{\beta}^s_{i,j}-\beta^s_j}{\beta^s_j}\right)^2\right)^{1/2}, \tag{4.7}$$

where $\hat{\beta}^s_{i,j}$ and β^s_j stand for the substantive (s) equation's j'th coefficient estimated in the i'th sample and the coefficient in the theoretical model that would have been obtained in the entire population, respectively. Ω is the number of data sets generated for the Monte-Carlo simulations, which is 5000 data sets (each one consists of N observations).

Another measure is based on a similar formula to the one described in (4.7), and is intended to find the relative accuracy of the truncated sample' estimates in comparison to full sample estimates, defined as:

$$R_j\left(\Omega\right)=\left(\frac{1}{\Omega}\sum_{i=1}^{\Omega}\left(\frac{\hat{\beta}^{ts}_{i,j}-\hat{\beta}^s_{i,j}}{\hat{\beta}^s_{i,j}}\right)^2\right)^{1/2}, \tag{4.8}$$

where $\hat{\beta}^{ts}_{i,j}$ and $\hat{\beta}^s_{i,j}$ stand for the substantive (s) equation's j'th coefficient estimated using the truncated (t) sample and the full sample, respectively. This measures evaluates the relative model's performance in the truncated sample with respect to the conventional IV using the full sample.

The last estimates' accuracy measure is the δ coefficient used for the calculation of the estimators' standard deviations convergence rate n^δ with respect to the sample size. It depicts the speed of standard deviation's shrinkage which is due to increasing sample size. This coefficient is calculated based on the following ratio:

$$\delta=\frac{\ln\left(\sigma_1/\sigma_2\right)}{\ln\left(n_2/n_1\right)}, \tag{4.9}$$

where σ_1 and σ_2 are the estimate's standard deviations calculated for data sets with n_1 and n_2 number observations, respectively (calculated for a given estimate).

TABLE III

Monte Carlo Simulation - Convergence measures

Parameter	Model's estimates					
	Number of observations					
	500	2000	3000	5000	8000	10000
RMSE measure						
Full sample conventional IV's estimates						
β_1	0.4400	0.2139	0.1754	0.1341	0.1063	0.0966
β_2	0.4372	0.2090	0.1739	0.1320	0.1056	0.0955
Truncated sample conventional IV's estimates						
β_1	1.1732	0.9027	0.8740	0.8504	0.8367	0.8292
β_2	0.9774	0.7021	0.6731	0.6466	0.6312	0.6239
Truncated sample model's estimates (First stage SPIV-NLS)						
δ_1	0.1567	0.0883	0.0756	0.0573	0.0423	0.0355
δ_2	0.0537	0.0265	0.0215	0.0166	0.0131	0.0117
Truncated sample model's estimates (Second stage SPIV-NLS)						
β_1	0.7525	0.3782	0.3101	0.2290	0.1779	0.1605
β_2	0.6670	0.3324	0.2740	0.2053	0.1602	0.1443
$\boldsymbol{R_j(n)}$ **measure - relative to full sample IV**						
Truncated sample conventional IV's estimates						
β_1	5.0168	0.9620	0.9059	0.8665	0.8459	0.8371
β_2	1.8410	0.7048	0.6722	0.6471	0.6305	0.6265
Truncated sample model's estimates (Second stage SPIV-NLS)						
β_1	4.5390	0.3719	0.2911	0.2042	0.1571	0.1359
β_2	1.2101	0.2986	0.2375	0.1733	0.1345	0.1186
$\boldsymbol{\delta}$ **consistency measure** ($\boldsymbol{n^{\delta}} \equiv$ **the convergence rate**)						
Truncated sample conventional IV's estimates						
β_1	-	0.7627	0.3538	0.2949	0.2254	0.1793
Truncated sample model's estimates (Second stage SPIV-NLS)						
β_1	-	0.4937	0.4904	0.5929	0.5372	0.4613

Note: We estimate by ordinary least squares (OLS) method the parameters for the full sample and truncated sample sample without correction for the selectivity bias, and compute the standard deviation in every random sample consisting of N observations. Then, we calculate for these estimates the mean, median and standard deviation (Std.) over all data sets. The standard deviations are obtained using the estimates from the Monte-Carlo simulations.

The entires in table III indicate that the root mean squares error (RMSE) measure of the estimates obtained by employing the conventional IV estimator using a random sample from the entire data distribution function, is getting smaller as the sample size increases as can be expected. However, applying the same procedure to the truncated data set leads to RMSE measures which are in the range of 4 to 8 fold larger, given a sample sizes of $2,000$ to $10,0000$ observations, respectively. This is indeed a huge bias generated by the conventional IV which is not immune to truncation bias. Additionally, the RMSE measures show negligible improvements as a function of number of observations for the conventional IV, whereas there is a huge improvement of the RMSE as a function of number of observations for the *Sieve* $SPIV$ estimator provided by our model. Another measure of performance is the r measure.

It is evident that the estimate of *Sieve* $SPIV$ is a $\sqrt{n}$ consistent as depicted by the δ consistency measure, while the truncated data conventional IV is poorly functioning in terms of consistency as is shown by entires in table III. In other words the consistency is not asymptoticly improved.

5. Conclusion

This paper extends the literature on instrumental variables for endogenously truncated data. We introduce a two stage estimation procedure, utilizing a Fourier-dependent Sieve estimator ($SPIV$) to capture the bias term. We provide analytical proof showing that the conventional IV estimator does not perform the task it was intended to and introduces an additional unintended bias into the parameters' estimates of the substantive equation. The instrumental variable is endogenous by itself in the context of endogenously truncated data. This endogeneity is related to a co-movement between the instrumental variable and the substantive equation's random disturbance, generated by mediating covariates. The offered truncation-proof IV is shown to be a proper IV estimator under endogenous truncation. Monte-Carlo application attests to the $SPIV$ estimator's high accuracy and its $\sqrt{n}$ consistency. These results have been verified by utilizing 2,000,000 different distribution functions (not restricted to the unimodal symmetric family), practically generating 100 millions realizations to construct the covariates' data sets which are not i.i.d. The various distribution functions attest to a very high accuracy of the model as depicted by the parameter estimates which quiet accurately mimic the true parameters.

Appendix

Proof of Tower property:

$$\mathbb{E}_{\mathbf{w}}\left[\mathbb{E}[\mathrm{z}|\mathbf{w},\mathrm{s}]|\mathrm{s}=s\right] = \int_{\boldsymbol{w}}\int_{z} z f_{\mathrm{z}|\mathbf{w},\mathrm{s}}(z|\mathbf{w}=\boldsymbol{w},\mathrm{s}=s)dz f_{\mathbf{w}|\mathrm{s}}(\boldsymbol{w}|\mathrm{s}=s)d\boldsymbol{w}$$

$$= \int_{\boldsymbol{w}}\int_{z} z \frac{f_{\mathrm{z},\mathbf{w}|\mathrm{s}=s}(z,\boldsymbol{w}|\mathrm{s}=s)}{f_{\mathbf{w}|\mathrm{s}=s}(\boldsymbol{w}|\mathrm{s}=s)} dz f_{\mathbf{w}|\mathrm{s}=s}(\boldsymbol{w}|\mathrm{s}=s)d\boldsymbol{w} = \int_{\boldsymbol{w}}\int_{z} z f_{\mathrm{z},\mathbf{w}|\mathrm{s}}(z,\boldsymbol{w}|\mathrm{s}=s)dz d\boldsymbol{w}$$

$$= \int_{z} z f_{\mathrm{z}|\mathrm{s}}(z|\mathrm{s}=s)dz = \mathbb{E}[\mathrm{z}|\mathrm{s}=s] \qquad \blacksquare$$

References

Ahn, H., Ichimura, H., Powell, J. L., and Ruud, P. A. (2015). Simple estimators for invertible index models. *Manuscript. University of California, Berkeley.*

Ai, C. and Chen, X. (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica*, 71(6):1795–1843.

Amemiya, T. (1974). The nonlinear two-stage least-squares estimator. *Journal of econometrics*, 2(2):105–110.

Angrist, J. D. and Imbens, G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American statistical Association*, 90(430):431–442.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434):444–455.

Angrist, J. D. and Kuersteiner, G. M. (2004). Semiparametric causality tests using the policy propensity score. Technical report, National Bureau of Economic Research.

Blundell, R. and Powell, J. L. (2003). Endogeneity in nonparametric and semiparametric regression models. *Econometric society monographs*, 36:312–357.

Chen, X., Chernozhukov, V., Lee, S., and Newey, W. K. (2014). Local identification of nonparametric and semiparametric models. *Econometrica*, 82(2):785–809.

Chen, X. and Liao, Z. (2015). Sieve semiparametric two-step gmm under weak dependence. *Journal of Econometrics*, 189(1):163–186.

Chen, X. and Pouzo, D. (2009). Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals. *Journal of Econometrics*, 152(1):46–60.

Das, M., Newey, W. K., and Vella, F. (2003). Nonparametric estimation of sample selection models. *The Review of Economic Studies*, 70(1):33–58.

d'Haultfoeuille, X. (2010). A new instrumental method for dealing with endogenous selection. *Journal of Econometrics*, 154(1):1–15.

Durbin, J. (1954). Errors in variables. *Revue de l'institut International de Statistique*, pages 23–32.

Escanciano, J. C. (2017). A simple and robust estimator for linear regression models with strictly exogenous instruments. *The Econometrics Journal*.

Florens, J.-P., Johannes, J., and Van Bellegem, S. (2012). Instrumental regression in partially linear models. *The Econometrics Journal*, 15(2):304–324.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, pages 1029–1054.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, 47(1):153–161.

Heckman, J. J. and Hotz, V. J. (1989). Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training. *Journal of the American statistical Association*, 84(408):862–874.

Heckman, J. J. and Robb, R. (1985). Alternative methods for evaluating the impact of interventions: An overview. *Journal of econometrics*, 30(1):239–267.

Hofert, M. (2008). Sampling archimedean copulas. *Computational Statistics & Data Analysis*, 52(12):5163–5174.

Horowitz, J. L. (2014). Adaptive nonparametric instrumental variables estimation: Empirical choice of the regularization parameter. *Journal of Econometrics*, 180(2):158–173.

Ichimura, H. (1993). Semiparametric least squares (sls) and weighted sls estimation of single-index models. *Journal of Econometrics*, 58(1):71–120.

Ichimura, H. and Lee, L.-F. (1991). Semiparametric least squares estimation of multiple index models: single equation estimation. In *Nonparametric and semiparametric methods in econometrics and statistics: Proceedings of the Fifth International Symposium in Economic Theory and Econometrics. Cambridge*, pages 3–49.

Imbens, G. W. and Newey, W. K. (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, 77(5):1481–1512.

Lewbel, A. (2007). Endogenous selection or treatment model estimation. *Journal of Econometrics*, 141(2):777–806.

Lewbel, A. and Linton, O. (2007). Nonparametric matching and efficient estimators of homothetically separable functions. *Econometrica*, 75(4):1209–1227.

Lewbel, A. and Schennach, S. M. (2007). A simple ordered data estimator for inverse density weighted expectations. *Journal of Econometrics*, 136(1):189–211.

Liang, H., Liu, X., Li, R., and Tsai, C. L. (2010). Estimation and testing for partially linear single-index models. *The Annals of Statistics*, 38(6):3811.

Marshall, A. W. and Olkin, I. (1988). Families of multivariate distributions. *Journal of the American statistical association*, 83(403):834–841.

McNeil, A. J. and Nešlehová, J. (2009). Multivariate archimedean copulas, d-monotone functions and -norm symmetric distributions. *The Annals of Statistics*, pages 3059–3097.

Newey, W. K. (2009). Two-step series estimation of sample selection models. *The Econometrics Journal*, 12(s1):S217–S229.

Newey, W. K. (2013). Nonparametric instrumental variables estimation. *The American Economic Review*, 103(3):550–556.

Newey, W. K. and Powell, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5):1565–1578.

Newey, W. K., Powell, J. L., and Vella, F. (1999). Nonparametric estimation of triangular simultaneous equations models. *Econometrica*, 67(3):565–603.

Ogden, T. (2012). *Essential wavelets for statistical applications and data analysis*. Springer Science & Business Media.

Powell, J. L. (2001). Semiparametric estimation of censored selection models. In *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya*, volume 165, page 96. Cambridge University Press.

Racine, J., Su, L., and Ullah, A. (2014). *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*. Oxford University Press.

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society*, pages 931–954.

Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica: Journal of the Econometric Society*, pages 393–415.

Sklar, M. (1959). *Fonctions de répartition à n dimensions et leurs marges*. Université Paris 8.

Williams, D. (1991). *Probability with martingales*. Cambridge university press.

Zhou, Y., Yang, Y., Han, J., and Zhao, P. (2016). Estimation for partially linear single-index instrumental variables models. *Communications in Statistics-Simulation and Computation*, 45(10):3629–3642.