Endogeneity Bias Modeling Using Observables EL40046 Revision

Antonio F. Galvao[*]
Gabriel Montes-Rojas[†]
Suyong Song[‡]

January 26, 2017

Abstract

This paper proposes an alternative solution to the endogeneity problem by explicitly modeling the joint interaction of the endogenous variables and the unobserved causes of the dependent variable as a function of additional observables. We derive identification of the parameters, develop an estimator, and establish its consistency and asymptotic normality.

Keywords: Endogeneity, instrumental variables, proxy variables.

JEL Classification: C10, C26

Words: 1970

[*]Department of Economics, University of Iowa, W284 Pappajohn Business Building, 21 E. Market Street, Iowa City, IA 52242. E-mail: antonio-galvao@uiowa.edu

[†]CONICET-Universidad de San Andrés, Vito Dumas 284, Victoria, B1644, Provincia de Buenos Aires, Argentina. Email: gmontesrojas@udesa.edu.ar

[‡]Corresponding author. Department of Economics, University of Iowa, W360 Pappajohn Business Building, 21 E. Market Street, Iowa City, IA 52242. E-mail: suyong-song@uiowa.edu

1 Introduction

The problem of endogeneity occupies a substantial amount of research in theoretical and applied econometrics. The most popular solutions are instrumental variables (IV) (see e.g. Hausman, 1983; Angrist and Krueger, 2001, for surveys) and proxy variables approach (see e.g. Olley and Pakes, 1996; Levinsohn and Petrin, 2003). These solutions rely on exogenous information derived from an additional exclusion restriction. In applications, the type of restriction chosen determines the nature of the model to be used, i.e. the instrument or the *proxy* variable. However, in many empirical applications, there is frequently disagreement and concern about the exclusion restrictions imposed, and instruments and proxies selections. The potential IV are often argued to be invalid since they are still correlated with the error term (see, e.g., Bound, Jaeger, and Baker (1995) and Hahn and Hausman (2002)) while the conditions for identification using proxy variables are many times implausible.

Recently there has been an expanding literature on analyzing endogeneity when IV and proxy variables models fail. This literature explores alternative moment conditions and exclusion restrictions. For instance, Altonji, Elder, and Taber (2005a,b, 2008) develop a strategy to extract information from observables about the endogeneity bias. They construct an index of observables, which can be used to identify the endogenous variable parameter, in combination with prior knowledge about the sign of the bias and a condition on the relationship between included (observable) and excluded (non-observable) variables. Chalak and White (2011) define a new class of extended IV, and introduce notions of conditioning and conditional extended IV which allow use of non-traditional instruments, as they may be endogenous. Chalak (2012) achieves identification of parameters by employing restrictions on the magnitude and sign of confounding instead of using traditional IV. Nevo and Rosen (2012) provide bounds for the parameters when the standard exogeneity assumption on IV fails, by assuming the correlation between the instruments and the error term has the same sign as the correlation between the endogenous regressor and the error term and that the instruments are less correlated with the error term than is the endogenous regressor. Montes-Rojas and Galvao (2014) exploit information on the structure of endogeneity and use prior information in a Bayesian framework to infer about the potential heterogeneity in parameter estimators.

This paper proposes an alternative solution to the endogeneity problem by explicitly modeling the joint interaction of the endogenous variables and the unobserved causes of the dependent variable as a function of additional observables. Identification uses the endogeneity structure of the model to build an alternative moment condition which is based on the non-zero conditional expectation implied by the endogeneity. That is, rather than imposing a sign on the endogeneity effect or exploring the bounds derived from its potential magnitude, we work with an alternative moment restriction. The intuition on the main identification condition of the new procedure is that, by using the proposed condition, the econometrician is able to model the endogeneity bias using the additional observable variables. Our framework allows for situations in which there are no valid standard IV or proxy variables available, but there exist additional variables that happen to be related to both the endogenous variable and the unobserved causes of the dependent variable. We develop a simple estimator based on the identification, and establish its consistency and asymptotic normality.

Many potential empirical applications might benefit from the proposed approach, especially those where the potential IV might still be related to the unobservables, or the proposed proxy variable does not satisfy all the requirements. Consider the errors-in-variables setting to motivate its empirical relevance. Many empirical applications rely on lagged mismeasured variables as IV to solve the implied endogeneity (see e.g. Biorn, 2000). This would fail if the measurement error is persistent because the instruments (i.e. lagged mismeasured variables) would still be correlated with the error term. More reliable estimates could be obtained by modeling the joint interaction of the mismeasured variable and the error term as a function of lagged mismeasured variables.

The paper is organized as follows. Section 2 presents the econometric model and establishes identification. Section 3 develops a consistent estimator, and establishes its asymptotic properties.

2 The Model

Consider the following structural model

$$y_i = \boldsymbol{x}_{1i}\boldsymbol{\beta}_1 + \boldsymbol{x}_{2i}\boldsymbol{\beta}_2 + \epsilon_i, \quad i = 1, ..., n, \tag{1}$$

where $\boldsymbol{\beta}_1$ is a p_1-vector, $\boldsymbol{\beta}_2$ is a p_2-vector, and ϵ_i is a scalar innovation term. Define $\boldsymbol{\beta} = [\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top]^\top$. We assume that $\boldsymbol{x}_{2i}$ is endogenous, and correlated with the innovation term ϵ_i in (1), such that $E[\boldsymbol{x}_{2i}^\top \epsilon_i] \neq 0$. In addition, $\boldsymbol{x}_{1i}$ is exogenous with $E[\boldsymbol{x}_{1i}^\top \epsilon_i] = 0$. The endogeneity in $\boldsymbol{x}_2$ produces endogeneity bias. To solve the endogeneity problem we will model the interaction of the endogenous variable and the error term, $\boldsymbol{x}_{2i}\epsilon_i$ and establish identification of $\boldsymbol{\beta}$ under some mild conditions. For simplicity, throughout we consider the case where $p_2 = 1$, i.e., there is only one endogenous variable, x_{2i}. Extension to the multivariate case is straightforward.

The following equation formalizes modeling endogeneity,

$$E(x_2\epsilon \mid \boldsymbol{z}, \boldsymbol{x}) = \boldsymbol{z}\boldsymbol{\phi}. \tag{2}$$

Equation (2) considers a linear model only for simplicity, but it could be extended to a nonparametric model (e.g., method of sieve). It explicitly models the endogeneity of x_2 using variables $\boldsymbol{z}$. In this case, by modeling endogeneity we mean to model the term $x_2\epsilon$.

When $\boldsymbol{\phi} \neq \mathbf{0}$, we can interpret the exogenous variable $\boldsymbol{z}$ as a noisy measure of the common cause(s) of x_2 and ϵ, which is related to the *joint* interaction of the endogenous variable and the unobservables. Our identification strategy requires observable variables, $\boldsymbol{z}$.

The proposed identification is related to the control function approach. When the correlation between x_2 and ϵ is modeled, equation (2) can be rewritten as $x_2 E(\epsilon \mid \boldsymbol{z}, \boldsymbol{x}) = \boldsymbol{z}\boldsymbol{\phi}$, hence, we have $E(\epsilon \mid \boldsymbol{z}, \boldsymbol{x}) = \frac{\boldsymbol{z}}{x_2}\boldsymbol{\phi}$. Therefore, the conditional expected value of the unobserved error term is a function of the "normalized" variables , i.e., $\frac{\boldsymbol{z}}{x_2}$. The emphasis is however on the nature of $\boldsymbol{z}$, which provide information about the joint interaction of the endogenous variable and the error term.

We are interested in identifying and estimating the parameters $\boldsymbol{\beta}$ in equation (1). In practice, $\boldsymbol{\phi}$ is unknown, and it is important to note that this parameter cannot be directly estimated from equation (2) because ϵ is unobservable. Define $\boldsymbol{\theta} \equiv [\boldsymbol{\beta}_1^\top, \boldsymbol{\alpha}^\top]^\top$ with $\boldsymbol{\alpha} \equiv [\beta_2^\top, \boldsymbol{\phi}^\top]^\top$. To ease the notation, define $\widetilde{y}$ and $\widetilde{x}_2$ after netting out the exogenous regressor $\boldsymbol{x}_1$ and multiplying the resulting objects by x_2. Thus, $\widetilde{y} = x_2(y - \boldsymbol{x}_1 E(\boldsymbol{x}_1^\top \boldsymbol{x}_1)^{-1} E(\boldsymbol{x}_1^\top y))$ and $\widetilde{\boldsymbol{x}} = [\widetilde{x}_2, \boldsymbol{z}]$, with $\widetilde{x}_2 = x_2(x_2 - \boldsymbol{x}_1 E(\boldsymbol{x}_1^\top \boldsymbol{x}_1)^{-1} E(\boldsymbol{x}_1^\top x_2))$. Let $\widetilde{\boldsymbol{z}}$ be a set of variables induced by conditioning variables $[\boldsymbol{z}, \boldsymbol{x}]$. Note that in this case we are obtaining the residual projection on $\boldsymbol{x}_1$. Consider the following assumptions.

Assumption 1

$$(i)\ E(\boldsymbol{x}_1^\top \epsilon) = \mathbf{0};$$

$$(ii)\ E(x_2\epsilon \mid \boldsymbol{z}, \boldsymbol{x}) = \boldsymbol{z}\boldsymbol{\phi}.$$

Assumption 2 *$E(\boldsymbol{x}_1^\top \boldsymbol{x}_1)$ and $E(\widetilde{\boldsymbol{z}}^\top \widetilde{\boldsymbol{x}})$ are non-singular.*

Assumptions 1 and 2 allow identification of the parameters of interest. Assumption 1 (i) simply states that $\boldsymbol{x}_1$ are exogenous regressors. Assumption 1 (ii) is the main identification condition. It is new in the literature and deserves further discussion. Condition 1 (*ii*) explicitly models the interaction between the endogenous variable and the unobserved causes of the dependent variable using a parametric model specification. It states that $\boldsymbol{z}$ are able to capture the information on the endogeneity term. The intuition behind this assumption is that once one controls for $\boldsymbol{z}$, $\boldsymbol{x}$ is not related to the interaction term $x_2\epsilon$. In other words, the endogeneity bias implied by the non-zero conditional expectation of the interaction term can be specified as a function of $\boldsymbol{z}$.

It is important to notice the restrictions this assumption imposes relative to IV approach in the literature. For simplicity we consider a model with only one (endogeneous) covariate $y = x\beta + \epsilon$. In our case, the additional equation can be rewritten as $x\epsilon = z\phi + u$ where u is the orthogonal projection of $x\epsilon$ on z. Our required moment conditions are two: $E[zu] = 0$ and $E[x^2u] = 0$. The IV model requires dependence between the endogenous regressor and the instrumental variables, which are restricted to be uncorrelated with the error term. This could be written as an additional equation $x = z\phi + u$ (where now u is the orthogonal projection of x on z) with also two moment conditions $E[zu] = 0$ and $E[z\epsilon] = 0$. Our method is able to allow the additional variable(s) z to still be correlated with the error term, ϵ, and also the endogenous variable to be correlated with u, the residual (unexplained) component in the additional equation. As a result, the difference between our proposed model and traditional IV approach rests on different model specifications; researchers fail to identify parameters if an incorrect method is employed to control for the endogeneity in each case. In our case we model endogeneity, the correlation of x and ϵ, i.e. $x\epsilon$.

We now return to the general structural equation (1) and general identification. For the sake of clarity, we focus on exactly identified model motivated by the conditional moment restriction of equation (2). The following theorem formalizes the identification results of $\boldsymbol{\theta}$, with $\boldsymbol{\theta} \equiv [\boldsymbol{\beta}_1^\top, \boldsymbol{\alpha}^\top]^\top$ and $\boldsymbol{\alpha} \equiv [\beta_2^\top, \boldsymbol{\phi}^\top]^\top$.

Theorem 1 *Suppose Assumption 1 holds. Then, $\boldsymbol{\theta}$ is fully identified with*

$$\boldsymbol{\alpha} = E(\widetilde{\boldsymbol{z}}^\top \widetilde{\boldsymbol{x}})^{-1} E(\widetilde{\boldsymbol{z}}^\top \widetilde{y}), \ \boldsymbol{\beta}_1 = E(\boldsymbol{x}_1^\top \boldsymbol{x}_1)^{-1} E(\boldsymbol{x}_1^\top y) - E(\boldsymbol{x}_1^\top \boldsymbol{x}_1)^{-1} E(\boldsymbol{x}_1^\top x_2)\beta_2,$$

if and only if Assumption 2 holds.

Proof. In Appendix. ■

In practice, the choice of the simultaneous variables is an important problem. The set of variables included in $\boldsymbol{z}$ is crucial, and the economic theory and empirical findings can be applied to guide the selection of the simultaneous variables and why the identification assumptions are satisfied in each case. An example on how relevant theory and empirical findings help in the selection of $\boldsymbol{z}$ is the returns to education. After the human capital theory of wage determination pioneered by Becker (1964, 1975) and following various empirical results, it is common to model the logarithm of wages as a function of education and other characteristics. However, a major concern regarding return to education has been the presence of ability bias because education and unobserved ability are positively correlated. According to economic theory (e.g. Roy, 1951; Willis and Rosen, 1979) and psychological theory (e.g. Binet, 1905; Cecci, 1991; Ree, Earles, and Teachout, 1994), intelligence quotient (IQ) or other measure of ability can be modeled as a function of both the unobserved ability and education. Thus, under some assumptions, one can use IQ or other measure of ability as simultaneous variables to model the interaction of education and the unobserved ability. In particular, IQ explains not only ability, but also the interaction of ability and education, as individuals with high IQ are likely to have more years of schooling.

3 Estimator

Given the identification result in Theorem 1, an estimator of $\boldsymbol{\theta}$ is

$$\widehat{\boldsymbol{\alpha}} = \left(\frac{1}{n}\sum_{i=1}^{n}\widetilde{\boldsymbol{z}}_i^\top\widehat{\boldsymbol{x}}_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\widetilde{\boldsymbol{z}}_i^\top\widehat{y}_i\right), \tag{3}$$

$$\widehat{\boldsymbol{\beta}}_1 = \left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{1i}^\top\boldsymbol{x}_{1i}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{1i}^\top y_i\right) - \left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{1i}^\top\boldsymbol{x}_{1i}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{1i}^\top x_{2i}\right)\widehat{\beta}_2, \tag{4}$$

where $\widetilde{\boldsymbol{z}}$ is the set of variables generated by conditioning variables, $\widehat{\boldsymbol{x}}$ and $\widehat{y}$ are sample analogs of $\widetilde{\boldsymbol{x}}$ and $\widetilde{y}$, which are obtained by replacing the expectations with sample means, and where $\widehat{\beta}_2$ is the first element of $\widehat{\boldsymbol{\alpha}}$.

The implementation of the proposed estimator can be carried through a sequence of OLS estimations as follows. First, compute the variables $\widehat{x}$ and $\widehat{y}$. To calculate $\widehat{y}$, one first partials out the exogenous regressors by computing the errors from a OLS regression of y on $\boldsymbol{x}_1$, then multiply those by x_2. Computation of $\widehat{x}$ is analogous. Second, estimate $\widehat{\boldsymbol{\alpha}}$ using equation (3) and $\widetilde{z}$, the set of variables generated by the conditioning variables. Finally, given $\widehat{\boldsymbol{\alpha}}$ and consequently $\widehat{\beta}_2$, $\widehat{\boldsymbol{\beta}}_1$ can be estimated from OLS as in equation (4), by using the coefficients of the OLS regression of y on $\boldsymbol{x}_1$ and also the coefficients of the regression of x_2 on $\boldsymbol{x}_1$. These generated variables affect the asymptotic variance-covariance matrix (see e.g. Pagan, 1984), as shown in the derivation of the asymptotic normality below.

The limiting behavior of the estimator, consistency and asymptotic normality, follows.

Theorem 2 *Let assumptions of Theorem 1 hold and the observations* $\{(y_i, \boldsymbol{x}_i, \boldsymbol{z}_i); i = 1, 2, ..., n\}$ *be i.i.d. across i and their fourth moments exist, i.e.,* $E(\|y_i\|^4) < \infty$, $E(\|\boldsymbol{x}_i\|^4) < \infty$, *and* $E(\|\boldsymbol{z}_i\|^4) < \infty$. *Denote* $Q \equiv E(\widetilde{\boldsymbol{z}}_i^\top\widetilde{\boldsymbol{x}}_i)$, $\boldsymbol{C}_1 \equiv E(\boldsymbol{x}_{1i}^\top\boldsymbol{x}_{1i})$, *and* $C_2 \equiv E(\boldsymbol{x}_{1i}^\top x_{2i})$. *Then, as* $n \to \infty$,

$$\widehat{\boldsymbol{\alpha}} \xrightarrow{p} \boldsymbol{\alpha}.$$

In addition, we have that

$$\sqrt{n}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \xrightarrow{d} N(0, Q^{-1}MQ^{-1}),$$

with $M = Var(\widetilde{\boldsymbol{z}}^\top u - Gr(\delta) + Hs(\gamma))$, *where* $G, r(\delta), H,$ *and* $s(\gamma)$ *are defined in the proof. Moreover,*

$$\widehat{\boldsymbol{\beta}}_1 \xrightarrow{p} \boldsymbol{\beta}_1,$$

and

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) \xrightarrow{d} N(0, \boldsymbol{C}_1^{-1} C_2 V_{\beta_2} C_2^\top \boldsymbol{C}_1^{-1}),$$

where V_{β_2} *is the variance of* $\widehat{\beta}_2$.

Proof. In Appendix. ∎

Appendix: Proof of the Results

Proof of Theorem 1

Note that from Assumption 1, $E(x_2\epsilon - \boldsymbol{z}\boldsymbol{\phi} \mid \boldsymbol{z}, \boldsymbol{x}) = E(x_2(y - \boldsymbol{x}_1\boldsymbol{\beta}_1 - x_2\beta_2) - \boldsymbol{z}\boldsymbol{\phi} \mid \boldsymbol{z}, \boldsymbol{x}) = E(x_2 y - x_2\boldsymbol{x}_1\boldsymbol{\beta}_1 - x_2x_2\beta_2 - \boldsymbol{z}\boldsymbol{\phi} \mid \boldsymbol{z}, \boldsymbol{x}) = E(x_2 y - x_2\boldsymbol{x}_1(E(\boldsymbol{x}_1^\top\boldsymbol{x}_1)^{-1}E(\boldsymbol{x}_1^\top y) - E(\boldsymbol{x}_1^\top\boldsymbol{x}_1)^{-1}E(\boldsymbol{x}_1^\top x_2)\beta_2) - x_2x_2\beta_2 - \boldsymbol{z}\boldsymbol{\phi} \mid \boldsymbol{z}, \boldsymbol{x}) = E(x_2(y - \boldsymbol{x}_1 E(\boldsymbol{x}_1^\top\boldsymbol{x}_1)^{-1}E(\boldsymbol{x}_1^\top y)) - x_2(x_2 - \boldsymbol{x}_1 E(\boldsymbol{x}_1^\top\boldsymbol{x}_1)^{-1}E(\boldsymbol{x}_1^\top x_2))\beta_2 - \boldsymbol{z}\boldsymbol{\phi} \mid \boldsymbol{z}, \boldsymbol{x}) = E(\widetilde{y} - \widetilde{\boldsymbol{x}}\boldsymbol{\alpha} \mid \boldsymbol{z}, \boldsymbol{x}) = 0$. We then have $E(\widetilde{\boldsymbol{z}}^\top(\widetilde{y} - \widetilde{\boldsymbol{x}}\boldsymbol{\alpha})) = 0$ or $E(\widetilde{\boldsymbol{z}}^\top\widetilde{y}) = E(\widetilde{\boldsymbol{z}}^\top\widetilde{\boldsymbol{x}})\boldsymbol{\alpha}$. Also note that from $E(\boldsymbol{x}_1^\top\epsilon) = 0$, $E(\boldsymbol{x}_1^\top(y - \boldsymbol{x}_1\boldsymbol{\beta}_1 - x_2\beta_2)) = 0$, $E(\boldsymbol{x}_1^\top y) - E(\boldsymbol{x}_1^\top\boldsymbol{x}_1)\boldsymbol{\beta}_1 - E(\boldsymbol{x}_1^\top x_2)\beta_2 = 0$. This system admits a unique solution $\boldsymbol{\theta}$ if and only if $E(\widetilde{\boldsymbol{z}}^\top\widetilde{\boldsymbol{x}})$ and $E(\boldsymbol{x}_1^\top\boldsymbol{x}_1)$ are non-singular (Assumption 2). *Q.E.D.*

Proof of Theorem 2

Let $\widetilde{\boldsymbol{x}} \equiv [x_2(x_2 - \boldsymbol{x}_1\boldsymbol{\delta}), \boldsymbol{z}]$ and $\widehat{\boldsymbol{x}} \equiv [x_2(x_2 - \boldsymbol{x}_1\widehat{\boldsymbol{\delta}}), \boldsymbol{z}]$ where $\boldsymbol{\delta} \equiv E(\boldsymbol{x}_1^\top\boldsymbol{x}_1)^{-1}E(\boldsymbol{x}_1^\top x_2)$ and $\widehat{\boldsymbol{\delta}} \equiv \left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{1i}^\top\boldsymbol{x}_{1i}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{1i}^\top x_{2i}\right)$. Also let $\widetilde{y} \equiv x_2(y - \boldsymbol{x}_1\boldsymbol{\gamma})$ and $\widehat{y} \equiv x_2(y - \boldsymbol{x}_1\widehat{\boldsymbol{\gamma}})$ where $\boldsymbol{\gamma} \equiv E(\boldsymbol{x}_1^\top\boldsymbol{x}_1)^{-1}E(\boldsymbol{x}_1^\top y)$ and $\widehat{\boldsymbol{\gamma}} \equiv \left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{1i}^\top\boldsymbol{x}_{1i}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{1i}^\top y_i\right)$.

From $\widetilde{y}_i = \widetilde{\boldsymbol{x}}_i \boldsymbol{\alpha} + u_i$, where u_i is i.i.d. innovation, we have

$$\widetilde{y}_i + (\widehat{y}_i - \widehat{y}_i) = (\widetilde{\boldsymbol{x}}_i + \widehat{\boldsymbol{x}}_i - \widehat{\boldsymbol{x}}_i)\boldsymbol{\alpha} + u_i,$$

$$\widehat{y}_i = \widehat{\boldsymbol{x}}_i \alpha + u_i - (\widehat{\boldsymbol{x}}_i - \widetilde{\boldsymbol{x}}_i)\boldsymbol{\alpha} + (\widehat{y}_i - \widetilde{y}_i). \tag{5}$$

Also we have

$$\begin{aligned}\widehat{\boldsymbol{\alpha}} &= \left(\frac{1}{n}\sum_{i=1}^{n} \widetilde{\boldsymbol{z}}_i^\top \widehat{\boldsymbol{x}}_i\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} \widetilde{\boldsymbol{z}}_i^\top \widehat{y}_i\right) \\ &= \boldsymbol{\alpha} + \left(\frac{1}{n}\sum_{i=1}^{n} \widetilde{\boldsymbol{z}}_i^\top \widehat{\boldsymbol{x}}_i\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} \widetilde{\boldsymbol{z}}_i^\top (u_i - (\widehat{\boldsymbol{x}}_i - \widetilde{\boldsymbol{x}}_i)\boldsymbol{\alpha} + (\widehat{y}_i - \widetilde{y}_i))\right).\end{aligned}$$

Then

$$\sqrt{n}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) = \widehat{Q}^{-1} n^{-1/2} \sum_{i=1}^{n} \widetilde{\boldsymbol{z}}_i^\top (u_i - (\widehat{\boldsymbol{x}}_i - \widetilde{\boldsymbol{x}}_i)\boldsymbol{\alpha} + (\widehat{y}_i - \widetilde{y}_i)), \tag{6}$$

where $\widehat{Q} \equiv \frac{1}{n}\sum_{i=1}^{n} \widetilde{\boldsymbol{z}}_i^\top \widehat{\boldsymbol{x}}_i$. By Chebychev's LLN and Slutsky's theorem,

$$\widehat{Q} \equiv \frac{1}{n}\sum_{i=1}^{n} \widetilde{\boldsymbol{z}}_i^\top \widehat{\boldsymbol{x}}_i \xrightarrow{p} E(\widetilde{\boldsymbol{z}}_i^\top \widetilde{\boldsymbol{x}}_i) \equiv Q.$$

As considered in Pagan (1984), equation (5) contains generated regressors and generated dependent variables. So we need to consider errors from these approximations in equation (6).

First, since $E(\widetilde{\boldsymbol{z}}_i^\top u_i) = 0$, we have

$$n^{-1/2} \sum_{i=1}^{n} \widetilde{\boldsymbol{z}}_i^\top u_i = o_p(1).$$

Second, by a mean value expansion,

$$\begin{aligned} n^{-1/2} \sum_{i=1}^{n} \widetilde{\boldsymbol{z}}_i^\top (\widehat{\boldsymbol{x}}_i - \widetilde{\boldsymbol{x}}_i)\alpha &= \left[n^{-1} \sum_{i=1}^{n} \widetilde{\boldsymbol{z}}_i^\top \nabla_{\boldsymbol{\delta}} \widetilde{\boldsymbol{x}}_i \alpha \right] \sqrt{n}(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}) + o_p(1), \\ &= G\sqrt{n}(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}) + o_p(1),\end{aligned}$$

where $G = E[\widetilde{\boldsymbol{z}}_i^\top \nabla_{\boldsymbol{\delta}} \widetilde{\boldsymbol{x}}_i \alpha]$.

Third, a similar argument gives us

$$n^{-1/2}\sum_{i=1}^{n}\widetilde{\boldsymbol{z}}_i^{\top}(\widehat{y}_i-\widetilde{y}_i)=\left[n^{-1}\sum_{i=1}^{n}\widetilde{\boldsymbol{z}}_i^{\top}\nabla_{\boldsymbol{\gamma}}\widetilde{y}_i\right]\sqrt{n}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})+o_p(1),$$
$$=H\sqrt{n}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})+o_p(1),$$

where $\nabla_{\boldsymbol{\gamma}}\widetilde{y}_i=-x_2\boldsymbol{x}_1$ and $H=E[\widetilde{\boldsymbol{z}}_i^{\top}\nabla_{\boldsymbol{\gamma}}\widetilde{y}_i]$.

Note that from the definition $\boldsymbol{\delta}$, we can write the following Bahadur representation

$$\sqrt{n}(\widehat{\boldsymbol{\delta}}-\boldsymbol{\delta})=\sqrt{n}\sum_{i=1}^{n}r_i(\delta)+o_p(1),$$

where $r_i(\delta)=\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{1i}^{\top}\boldsymbol{x}_{1i}\right)^{-1}\left(\boldsymbol{x}_{1i}^{\top}(x_{2i}-\boldsymbol{x}_{1i}\boldsymbol{\delta})\right)$, and $E[r_i(\delta)]=0$ by the Law of Iterated Expectations (LIE). In the same way, given the definition of $\boldsymbol{\gamma}$, we can write the following representation

$$\sqrt{n}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})=\sqrt{n}\sum_{i=1}^{n}s_i(\gamma)+o_p(1),$$

where $s_i(\gamma)=\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{1i}^{\top}\boldsymbol{x}_{1i}\right)^{-1}\left(\boldsymbol{x}_{1i}^{\top}(y_i-\boldsymbol{x}_{1i}\boldsymbol{\gamma})\right)$, and $E[s_i(\gamma)]=0$ by LIE.

By combining all terms together, we have

$$\sqrt{n}(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})=Q^{-1}\left\{n^{-1/2}\sum_{i=1}^{n}[\widetilde{\boldsymbol{z}}_i^{\top}u_i-Gr_i(\delta)+Hs_i(\gamma)]\right\}+o_p(1).$$

For the consistency of $\widehat{\boldsymbol{\alpha}}$, we have

$$\widehat{\boldsymbol{\alpha}}\xrightarrow{p}\boldsymbol{\alpha}+Q^{-1}\cdot 0=\boldsymbol{\alpha}.$$

For the asymptotic normality, we have that by the Lindeberg-Lévy Central Limit Theorem,

$$\sqrt{n}(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})\xrightarrow{d}Q^{-1}N(0,M)\equiv N(0,Q^{-1}MQ^{-1}),$$

where $M=Var(\widetilde{\boldsymbol{z}_i}^{\top}u_i-Gr_i(\delta)+Hs_i(\gamma))$.

Similarly,

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_1 &= \left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{1i}^\top \boldsymbol{x}_{1i}\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{1i}(\boldsymbol{x}_{1i}\boldsymbol{\beta}_1 + x_{2i}\beta_2 + \epsilon)\right) - \left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{1i}^\top \boldsymbol{x}_{1i}\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{1i} x_{2i}\right) \widehat{\beta}_2 \\
&= \boldsymbol{\beta}_1 + \left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{1i}^\top \boldsymbol{x}_{1i}\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{1i}^\top x_{2i}\right) \beta_2 - \left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{1i}^\top \boldsymbol{x}_{1i}\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{1i}^\top x_{2i}\right) \widehat{\beta}_2 \\
&= \boldsymbol{\beta}_1 - \left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{1i}^\top \boldsymbol{x}_{1i}\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{1i}^\top x_{2i}\right) (\widehat{\beta}_2 - \beta_2).
\end{aligned}$$

By Chebychev's LLN,

$$\begin{aligned}
\widehat{\boldsymbol{C}}_1 \equiv \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{1i}^\top \boldsymbol{x}_{1i} \xrightarrow{p} E(\boldsymbol{x}_{1i}^\top \boldsymbol{x}_{1i}) \equiv \boldsymbol{C}_1, \\
\widehat{C}_2 \equiv \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{1i}^\top x_{2i} \xrightarrow{p} E(\boldsymbol{x}_{1i}^\top x_{2i}) \equiv C_2,
\end{aligned}$$

we have

$$\widehat{\boldsymbol{\beta}}_1 \xrightarrow{p} \boldsymbol{\beta}_1 - \boldsymbol{C}_1^{-1} C_2 Q_{\beta_2}^{-1} \cdot 0 = \boldsymbol{\beta}_1,$$

where Q_{β_2} is the element in the Q matrix that corresponds to the estimation of β_2.

Note that

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) = -\left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{1i}^\top \boldsymbol{x}_{1i}\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{1i}^\top x_{2i}\right) \sqrt{n}(\widehat{\beta}_2 - \beta_2).$$

Thus, we have

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) \xrightarrow{d} \boldsymbol{C}_1^{-1} C_2 N(0, V_{\beta_2}) \equiv N(0, \boldsymbol{C}_1^{-1} C_2 V_{\beta_2} C_2^\top \boldsymbol{C}_1^{-1}).$$

Q.E.D.

References

Altonji, J. G., T. E. Elder, and C. R. Taber (2005a): "An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schooling," *Journal of Human Resources*, 40, 791–821.

——— (2005b): "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 113, 151–184.

——— (2008): "Using Selection on Observed Variables to Assess Bias from Unobservables When Evaluating Swan-Ganz Catheterization," *American Economic Review Papers and Proceedings*, 98, 345–350.

Angrist, J., and A. Krueger (2001): "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments," *Journal of Economic Perspectives*, 15, 69–85.

Becker, G. S. (1964): *Human Capital: A Theoretical Analysis with Special Reference to Education.* Columbia University Press for NBER, New York, New York.

——— (1975): *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education.* Columbia University Press for NBER, New York, New York.

Binet, A. (1905): "New Methods for the Diagnosis of the Intellectual Level of Subnormals," *L'Anne Psychologique*, 12, 191–244.

Biorn, E. (2000): "Panel Data with Measurement Errors: Instrumental Variables and GMM Procedures Combining Levels and Difference," *Econometric Reviews*, 19, 391–424.

Bound, J., D. A. Jaeger, and R. M. Baker (1995): "Problems With Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous

Explanatory Variables is Weak," *Journal of the American Statistical Association*, 90, 443–450.

Cecci, S. J. (1991): "How Much Does Schooling Influence General Intelligence and its Cognitive Components? A Reassessment of the Evidence," *Developmental Psychology*, 27, 703–722.

Chalak, K. (2012): "Identification of Average Random Coefficients under Magnitude and Sign Restrictions on Confounding," mimeo, University of Virginia.

Chalak, K., and H. White (2011): "An Extended Class of Instrumental Variables for the Estimation of Causal Effects," *Canadian Journal of Economics*, 44, 1–51.

Hahn, J., and J. Hausman (2002): "A New Specification Test for the Validity of Instrumental Variables," *Econometrica*, 70, 163–189.

Hausman, J. A. (1983): "Specification and Estimation of Simultaneous Equation Models," in *Handbook of Econometrics*, ed. by Z. Griliches, and M. D. Intriligator, vol. II, pp. 391–448. North-Holland Press.

Levinsohn, J., and A. Petrin (2003): "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 70, 317–342.

Montes-Rojas, G., and A. Galvao (2014): "Bayesian endogeneity bias modeling," *Economics Letters*, 122(1), 36–39.

Nevo, A., and A. M. Rosen (2012): "Identification with Imperfect Instruments," *Review of Economics and Statistics*, 94, 659–671.

Olley, S., and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263–1298.

Pagan, A. (1984): "Econometric Issues in the Analysis of Regressions with Generated Regressors," *International Economic Review*, 25, 221–247.

Ree, M. J., J. A. Earles, and M. S. Teachout (1994): "Predicting Job Performance: Not Much More than g," *Journal of Applied Psychology*, 79, 518–524.

Roy, A. D. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3, 135–146.

Willis, R. J., and S. Rosen (1979): "Education and Self-Selection," *Journal of Political Economy*, 87, S7–S36.