Simultaneous Variables*

Antonio F. Galvao†
Gabriel Montes-Rojas‡
Suyong Song§

February 16, 2015

Abstract

This paper develops practical inference methods for econometric models with endogeneity while relaxing the standard exclusion restrictions. We propose 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. These additional variables are defined as simultaneous variables because they are allowed to still be correlated with the innovation term. We derive identification of the parameters, develop an estimator, and establish its consistency and asymptotic normality. Inference procedures are based on the Wald statistic. We illustrate the methods with an example that examines the investment equation model. Our empirical findings support the idea that the proposed estimator is a useful alternative to the instrumental variables for economic applications in which endogeneity is a relevant concern.

Keywords: Endogeneity, instrumental variables, simultaneous variables, investment equation.

JEL Classification: C10, C26, G31

*The authors would like to express their appreciation to Carolina Caetano, Karim Chalak, Juan Carlos Escanciano, Sergio Firpo, Amit Gandhi, Zhengyuan Gao, Chuan Goh, Fabio Gomes, Susumu Imai, Kyoo il Kim, Roger Koenker, Joris Pinkse, Alex Poirier, Gene Savin, Liang Wang, and seminar participants at the University of Minnesota, University of Illinois at Urbana-Champaign, Universidad de Buenos Aires, University of Iowa, University of Wisconsin-Milwaukee, Central Bank of Brazil, 23rd Midwest Econometrics Group and Universidad de San Andrés for helpful discussions and comments regarding this paper. All the remaining errors are ours.

†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 and Department of Economics, City University London, D308 Social Sciences Bldg, Northampton Square, London EC1V 0HB, UK. Email: `Gabriel.Montes-Rojas.1@city.ac.uk`

§Department of Economics, University of Wisconsin-Milwaukee, Bolton Hall 870, 3210 N. Maryland Ave., Milwaukee, WI 53211. Email: `songs@uwm.edu`

1 Introduction

The problem of endogeneity occupies a substantial amount of research in theoretical and applied econometrics. One of the most popular approaches to solve endogeneity is the instrumental variables (IV). The solution relies on exogenous information derived from an additional exclusion restriction, and requires the additional variables (instruments) to be correlated with the endogenous variable and uncorrelated with the unobserved causes of the response variable. However, in many empirical applications, there are frequently disagreement and concerns about the IV selection. 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; Hahn and Hausman, 2002). For example, consider the following model

$$y = \gamma + x\beta + \epsilon,$$

where $x = v_1 + v_2$, $\epsilon = v_2 + v_3$, and $v_j \sim i.i.d.N(0,1), j = 1,2,3$. In this model, x is correlated with ϵ, and hence, endogenous. Suppose, an additional variable $z = v_1 v_2 + v_2^2$ is available to the researcher. The variable z is correlated with x. However, it is an invalid IV because it is still related to ϵ through v_2. In this paper, we propose a new method to solve the endogeneity of x using a variable as z, which is still correlated with the error term. We provide identification conditions under which z can be utilized to obtain consistent estimates of β and γ.

This paper contributes to the literature by proposing a novel alternative solution to the problem of endogeneity. In doing so, we propose a new identification condition which explicitly models the endogeneity bias. More specifically, the conditional expectation of the joint interaction of the endogenous variables and unobserved causes of the dependent variable is assumed to be a function of additional observable variables. We define such additional variables as *simultaneous variables* because they are simultaneously related to both the endogenous variable and the unobservables. Thus, we develop simple and tractable methods for conducting estimation and inference in econometric models with endogeneity, and allow the additional variables (the simultaneous variables) to still be correlated with the innovation term. Given the structural model and the simultaneous variables, we establish point identification of the parameters of interest. Motivated by the identification result we develop an estimator that is simple to implement in practice, and establish its consistency and asymptotic normality. In addition, we propose practical inference procedures based on the Wald statistic, as well as a test for endogeneity.

Our framework allows for situations in which there are no valid standard IV available, but there exist additional variables that happen to be related to the joint interaction of the endogenous variable and the unobserved causes of the dependent variable. The intuition on the main identification condition of the new procedure is that, by using the proposed restriction, the econometrician is able to approximate the endogeneity bias using the simultaneous variables, such that after controlling for the simultaneous variables, the endogenous variables are not related to the interaction term (between the endogenous variable and the innovation). Thus, the endogeneity bias implied by the non-zero conditional expectation of the interaction term can be effectively specified as a function of the simultaneous variables only.[1]

Monte Carlo studies are conducted to evaluate the finite sample properties of the proposed estimator. The experiments suggest that when simultaneous variables are available, the proposed method produces consistent estimates under the required assumptions. Moreover, the simulations highlight the differences between the simultaneous variables approach and the IV method. The results suggest that when only simultaneous variables exist, our method

[1] Strictly speaking, neither the IV or simultaneous method is more general than the other because the underlying assumptions are non-nested. They differ in the characteristics of the additional variables; in IV case these cause the endogenous variable only, while in our proposed method they are allowed to affect the joint interaction of the endogenous variable and error term.

is able to eliminate the endogeneity bias while the IV returns significantly biased estimates.

The new procedures provide intuitive and practical ways of handling the problem of endogeneity in empirical settings. To motivate and illustrate the applicability of the identification and estimator, we apply the methods to an investment equation model where measurement errors are a relevant concern. In particular, we consider the Fazzari, Hubbard, and Petersen's (1988) investment equation model, where a firm's investment is regressed on observed investment demand (Tobin's q) and cash flows. Concerns about measurement errors in Tobin's q have been emphasized in the empirical investment equation models context (see e.g., Hayashi, 1982; Poterba, 1988) because researchers only observe average q instead of marginal q. Hence, the measurement errors induce endogeneity in the independent variables, and consequently bias in their estimates. We show that our proposed methods are well suited to solve the existing measurement error problem in this literature. It has been common in the literature to employ IV methods using the lags of q as instruments to resolve the endogeneity problem (see e.g., Almeida, Campello, and Galvao, 2010; Lewellen and Lewellen, 2014). However, the approach of using lagged q as IV fails when the measurement errors on the marginal q are systematic. In practice, it is highly likely that current-period measurement error is correlated with the first-order or higher-order lags of the measurement error. This phenomenon of autocorrelation in the measurement errors seems to be more realistic in practice because systematic errors on a way of evaluating average q would be highly likely to be persistent. Contrary to the IV, we show that even under the assumption that the marginal q and measurement error are correlated, the lagged Tobin's q and its square are valid simultaneous variables because they are related to the interaction of the endogenous variable q and the error term. Thus, we can use these simultaneous variables to remove the bias from the measurement error and obtain consistent estimates for the parameters of interest. For comparison, we estimate the model using OLS, IV and simultaneous variables estimators. Our empirical findings support the idea that the simultaneous variable estimator is a useful alternative to instrumental variables models.

The proposed methods relate to recent advances in the literature on various alternative estimation and inference procedures that attempt to solve endogeneity without standard instrumental variables. In a recent paper, Conley, Hansen, and Rossi (2012) present practical methods for performing inference while relaxing the IV exclusion restriction. They use prior information regarding the extent of deviations from the exact exclusion restriction. 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. Gandhi, Kim, and Petrin (2013) propose a generalized control function approach for models where the endogenous variables interact with the error term. Caetano (2015) develops an interesting test procedure for exogeneity of explanatory variables without relying on IV. The test rests on an assumption that the structural function needs to be continuous in the explanatory variable of interest, but it does not require the structural function to be identified under either the null or the alternative hypotheses.

The paper is organized as follows. Section 2 presents the econometric model, the identification results, and comparison with IV method. Section 3 proposes a consistent estimator, establishes its asymptotic properties, and develops inference procedures. Section 4 presents extensions to nonparametric model and a case of general form of endogeneity. Section 5 studies the finite sample performance of the proposed estimator. Section 6 applies the estimator to the investment equation model. Finally, Section 7 concludes the paper. Technical proofs and extension to multiple endogenous variables are included in the Appendix.

2 Modeling endogeneity using observables

In this section, we propose a new methodology to solve the endogeneity bias problem. Identification of the parameters of interest is achieved by modeling the endogeneity bias through a novel additional equation. We also compare the proposed method with the IV framework.

2.1 The model

Consider the following structural model

$$E[y_i|\boldsymbol{x}_{1i}, \boldsymbol{x}_{2i}] = f(\boldsymbol{x}_{1i}, \boldsymbol{x}_{2i}), \quad i = 1, ..., n, \tag{1}$$

where y_i is a scalar dependent variable, $\boldsymbol{x}_{1i}$ is a p_1-vector of exogenous explanatory variables, $\boldsymbol{x}_{2i}$ is a p_2-vector of endogenous regressors, and $f(\cdot, \cdot)$ is an unknown measurable function. Define $\boldsymbol{x}_i = [\boldsymbol{x}_{1i}, \boldsymbol{x}_{2i}]$ as a $(p_1 + p_2)$-vector with the sample covariates. The main focus of this paper is endogeneity and its solution.

For motivation and exposition purposes, we primarily investigate the endogeneity problem in linear regression models. Nevertheless, the results generalize to nonparametric models (see Section 4.1 below for details). Equation (1) can be written as

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

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 (2), such that $E[\boldsymbol{x}_{2i}^\top \epsilon_i] \neq 0$. In addition, $\boldsymbol{x}_{1i}$ is exogenous with $E[\boldsymbol{x}_{1i}^\top \epsilon_i] = 0$. Thus, we also assume that

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i^\top \epsilon_i \xrightarrow{p} E[\boldsymbol{x}^\top \epsilon] \equiv \boldsymbol{B} = [\boldsymbol{0}^\top, \boldsymbol{B}_2^\top]^\top,$$

with $\boldsymbol{B}$ a (p_1+p_2)-vector (finite), and $\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i^\top \boldsymbol{x}_i \xrightarrow{p} E[\boldsymbol{x}^\top \boldsymbol{x}] \equiv \boldsymbol{C}$ (finite and non-singular). The endogeneity in $\boldsymbol{x}_2$ produces an endogeneity bias, the term $\boldsymbol{B}$, in the standard OLS estimator, $\widehat{\boldsymbol{\beta}}^{OLS} = (\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i^\top \boldsymbol{x}_i)^{-1}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i^\top y_i$. Simple algebra and asymptotic calculations show that $\widehat{\boldsymbol{\beta}}_1^{OLS} \xrightarrow{p} \boldsymbol{\beta}_1 - \boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{\delta}_2$, and $\widehat{\boldsymbol{\beta}}_2^{OLS} \xrightarrow{p} \boldsymbol{\beta}_2 + \boldsymbol{\delta}_2$, where $\boldsymbol{\delta}_2 = \boldsymbol{V}_{2\cdot 1}^{-1}\boldsymbol{B}_2$. Note that if the endogeneity term $\boldsymbol{B}$ were known, then the correct estimating equation would be

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i^\top (y_i - \boldsymbol{x}_i\boldsymbol{\beta}) = \boldsymbol{B}, \tag{3}$$

and its solution is $\widetilde{\boldsymbol{\beta}} = (\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i^\top \boldsymbol{x}_i)^{-1}\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i^\top y_i - \boldsymbol{B}) = \widehat{\boldsymbol{\beta}}^{OLS} - (\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i^\top \boldsymbol{x}_i)^{-1}\boldsymbol{B}$. However, since the $\boldsymbol{B}$ is unknown, to solve the endogeneity problem we will instead 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}. The extension to multiple endogenous variables is derived in Appendix B.

2.2 Identification

Identification of the parameters of interest is achieved by explicitly modeling the interaction of the endogenous variable and the unobserved causes of the dependent variable as a function of additional observable variables. In particular, motivated by equation (3), we consider the case where the variable $x_2\epsilon$ can be modeled using additional variables. The following equation formalizes modeling endogeneity

$$E(x_2\epsilon \mid \boldsymbol{z}, \boldsymbol{x}) = g(\boldsymbol{z}), \tag{4}$$

where $g(\cdot)$ is an unknown smooth function and $\boldsymbol{z}$ a k-vector of additional observable variables. This is a general formulation to model the endogeneity in the linear parametric model. A more general form of endogeneity which takes into account nonlinear dependence between these two variables is considered in Section 4.2 below. For simplicity, assume that $g(\cdot)$ is a known function of $\boldsymbol{z}$ with unknown parameters ϕ such as $g(\boldsymbol{z};\phi)$, but the analysis can be easily extended to the case of unknown functional form of $g(\cdot)$ which is discussed in Section 4.1. For notational simplicity, we suppress the subscript i whenever there is no confusion.

A simple example of (4) that is convenient for exposition and estimation purposes is the following polynomial of $\boldsymbol{z}$,

$$E(x_2\epsilon \mid \boldsymbol{z}, \boldsymbol{x}) = \boldsymbol{Z}\boldsymbol{\phi}, \tag{5}$$

where $\boldsymbol{Z} = [1, \boldsymbol{z}, \boldsymbol{z}^2, \ldots, \boldsymbol{z}^m]$ and $\boldsymbol{\phi} = [\phi_0, \boldsymbol{\phi}_1^\top, \ldots, \boldsymbol{\phi}_m^\top]^\top$ which is a nonzero vector.[2] Equation (5) is explicitly modeling the endogeneity of x_2. In this case, by modeling endogeneity we mean to model the term $x_2\epsilon$. When $\boldsymbol{\phi} \neq \boldsymbol{0}$, 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, $\boldsymbol{z}$. We define such variables as simultaneous variables because they are *simultaneously* related to x_2 and ϵ.

We are interested in identifying and estimating the parameters $\boldsymbol{\beta}$ in equation (2). In practice, $\boldsymbol{\phi}$ is unknown, and it is important to note that this parameter cannot be directly estimated from equation (5) because ϵ is unobservable. Hence, we consider the joint identification and estimation of both $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$. To this end, we impose restrictions on the relationship between the endogenous regressor and the simultaneous variables.

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 $\widetilde{x}_2 = x_2(x_2 - \boldsymbol{x}_1 E(\boldsymbol{x}_1^\top \boldsymbol{x}_1)^{-1} E(\boldsymbol{x}_1^\top x_2))$.

[2] One might want to approximate the unknown function $g(\cdot)$ with one of the sieve bases (e.g., power series, Fourier series, splines, etc.). See, e.g., Chen (2007) for more details on the method of sieve.

Let $\widetilde{\boldsymbol{z}}$ be a set of variables induced by conditioning variables $[\boldsymbol{z}, \boldsymbol{x}]$. 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 us to identify 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 the simultaneous variables are able to capture the information on the endogeneity term. The intuition behind this assumption is that once one controls for the simultaneous variables ($\boldsymbol{z}$), $\boldsymbol{x}$ 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 the simultaneous variables only.

It is important to notice the restriction this assumption imposes relative to the literature. The simultaneous variables model allows for dependence between the error term and simultaneous variables at the expense of restricting the interaction between the endogenous regressor and the error term being a function of the simultaneous variables only. In contrast, the IV model requires dependence between the endogenous regressor and the instrumental variables, which are restricted to be uncorrelated with the error term. Therefore, the new method is able to allow the additional variables to still be correlated with the error term at the cost of requiring the interaction not to be related to the endogenous variable. 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.

Assumption 1 (*ii*) can be interpreted in an alternative form. In particular, it can be restated as

$$x_2\epsilon = \boldsymbol{Z}\boldsymbol{\phi} + u, \text{ with } E(u \mid \boldsymbol{z}, \boldsymbol{x}) = 0. \tag{6}$$

This format provides an auxiliary equation which models endogeneity bias using simultaneous variables, and the condition $E(u \mid \boldsymbol{z}, \boldsymbol{x}) = 0$ states that the innovation u should have conditional mean zero given $\boldsymbol{z}$ and $\boldsymbol{x}$. This requirement relates to standard exogeneity condition in the IV literature, which requires that the innovation term in the first-stage equation needs to be uncorrelated with instrumental variables.[3]

Although the condition $\boldsymbol{\phi} \neq \boldsymbol{0}$ is not a formal requirement, it implicitly arises because of the fact that if $\boldsymbol{\phi} = \boldsymbol{0}$, there is no endogeneity problem and the whole exercise is unnecessary. Assumption 2 is a standard rank condition. Non-singularity of $E(\boldsymbol{x}_1^\top \boldsymbol{x}_1)$ is necessary for the identification of $\boldsymbol{\beta}_1$ and non-singularity of $E(\widetilde{\boldsymbol{z}}^\top \widetilde{\boldsymbol{x}})$ is required for the identification of $\boldsymbol{\alpha}$. The later states that the simultaneous variables are not linearly related to the endogenous regressor.

To motivate our approach and fix the ideas, we consider the following simple example. We discuss the previous assumptions and heuristically show how they deliver identification of the parameters of interest. Let

$$y = \gamma + x\beta + \epsilon,$$

[3] One can also notice yet another additional interpretation of the main identification condition. When the correlation between x_2 and ϵ is modeled, Assumption 1 (ii) 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" simultaneous variables by the endogenous variable , i.e., $\frac{\boldsymbol{Z}}{x_2}$.

where x is endogenous, i.e. $E[x\epsilon] \neq 0$. Note that there are no exogenous regressors for simplicity. Multiplying both sides of the structural model by x we have

$$xy = x\gamma + x^2\beta + x\epsilon.$$

In this simple example, Assumption 1 (ii) can be restated as $E(x\epsilon \mid x, \boldsymbol{z}) = \boldsymbol{Z}\boldsymbol{\phi}$. The previous equation for xy can be written as

$$E(xy - x\gamma - x^2\beta \mid x, \boldsymbol{z}) = \boldsymbol{Z}\boldsymbol{\phi}.$$

As a result, Assumption 1 (ii) provides a moment condition such as

$$E(\widetilde{\boldsymbol{z}}^\top(xy - x\gamma - x^2\beta - \boldsymbol{Z}\boldsymbol{\phi})) = \mathbf{0},$$

where $\widetilde{\boldsymbol{z}}$ is a set of variables generated by the conditioning variables. By selecting $\widetilde{\boldsymbol{z}} \equiv [x, x^2, \boldsymbol{Z}]$, for instance, we have

$$E([x, x^2, \boldsymbol{Z}]^\top(xy - x\gamma - x^2\beta - \boldsymbol{Z}\boldsymbol{\phi})) = \mathbf{0}.$$

By rearranging the above equation, we obtain

$$E(\widetilde{\boldsymbol{z}}^\top\widetilde{y}) = E(\widetilde{\boldsymbol{z}}^\top\widetilde{\boldsymbol{x}})\boldsymbol{\alpha},$$

where $\widetilde{y} \equiv xy$, $\widetilde{\boldsymbol{x}} \equiv [x, x^2, \boldsymbol{Z}]$ and $\boldsymbol{\alpha} \equiv [\gamma, \beta, \boldsymbol{\phi}]^\top$. Finally, by Assumption 2, $E(\widetilde{\boldsymbol{z}}^\top\widetilde{\boldsymbol{x}})$ is invertible, and $\boldsymbol{\alpha}$ can be uniquely identified as

$$\boldsymbol{\alpha} = E(\widetilde{\boldsymbol{z}}^\top\widetilde{\boldsymbol{x}})^{-1}E(\widetilde{\boldsymbol{z}}^\top\widetilde{y}).$$

In this simple illustration, the first condition used to derive identification is that the interaction of x and ϵ can be modeled by observables $\boldsymbol{z}$. The second condition states that $\boldsymbol{z}$ cannot be linearly related to x and x^2 by the invertibility condition on $E(\widetilde{\boldsymbol{z}}^\top\widetilde{\boldsymbol{x}})$.

We now return to the general structural equation (2) and general identification. For the sake of clarity, we first focus on exactly identified model motivated by the conditional moment restriction of equation (5).[4] We consider the over-identified case in Section 4.2 below. 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 A. ■

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 along with empirical findings can be applied to guide the selection of the simultaneous variables and why the identification assumptions are satisfied in each case.

2.3 Comparison with IV approach

This section discusses the differences between the proposed simultaneous variables and the IV approaches. Suppose that x_2 is endogenous in (2). For the standard IV case, in the first-stage we have

$$x_2 = \boldsymbol{x}_1 \boldsymbol{\pi}_1 + z\pi_2 + w,$$

where w is an unobserved component. The instrumental variable, z, is a valid instrument when it satisfies two conditions. First, it is correlated with the endogenous variable, $Cov(x_2, z) \neq 0$, after partialling out $\boldsymbol{x}_1$. Second, the instrument is required to be uncorrelated with the unobserved causes of the dependent variable, $E(\epsilon \mid z) = 0$, after partialling

[4] Based on $E(\rho(x_2\epsilon, \boldsymbol{Z}; \boldsymbol{\theta}) \mid \boldsymbol{z}, \boldsymbol{x}) = 0$, where $\rho(\cdot, \cdot; \boldsymbol{\theta}) = x_2(y - \boldsymbol{x}\boldsymbol{\beta}) - \boldsymbol{Z}\boldsymbol{\phi}$, we obtain the orthogonality condition $E(\widetilde{\boldsymbol{z}}\rho(x_2\epsilon, \boldsymbol{Z}; \boldsymbol{\theta})) = 0$. A consistent estimator of $\boldsymbol{\theta}$ can be achieved by imposing the sample analogue of the population orthogonality condition.

out $\boldsymbol{x}_1$.[5] Assumption 1 (ii) plays a similar role to the first IV condition above such that an instrumental variable should be related to the endogenous variable, i.e. $\pi_2 \neq 0$. The proposed method, however, requires the simultaneous variables to capture information on the interaction term. In addition, the IV requires that $E(w \mid \boldsymbol{x}_1, z) = 0$. This is similar to the simultaneous variables condition presented in equation (6), $E(u \mid \boldsymbol{x}, z) = 0$.

Regarding the second IV restriction, $E(\epsilon \mid z) = 0$, when it is violated, the IV approach fails to solve the endogeneity problem. However, under the assumptions for our methodology z could still be used as a simultaneous variable to control for endogeneity. Assumption 1 (ii) allows z to be correlated with the unobserved causes of the dependent variable, ϵ, by explicitly modeling the endogeneity. The main trade-off is that the simultaneous variables method requires the interaction between the endogenous regressor and the error term to be a function of the simultaneous variables only. Our approach is also different from the conditioning instrumental variables in the model of Chalak and White (2011), where in this case x_2 and ϵ would be uncorrelated once we condition on this type of extended instrument. Furthermore, our method relies on different model and assumptions than the proxy variable approach. In the proxy variables approach, condition $E[\epsilon \mid \boldsymbol{x}, \boldsymbol{z}] = g(\boldsymbol{z})$ controls for endogeneity of x_2 through the equation $y = \boldsymbol{x}_1\boldsymbol{\beta}_1 + x_2\boldsymbol{\beta}_2 + g(\boldsymbol{z}) + w$, where $E[w \mid \boldsymbol{x}, \boldsymbol{z}] = 0$. It is worth noting that our condition $E[x_2\epsilon \mid \boldsymbol{x}, \boldsymbol{z}] = g(\boldsymbol{z})$ is instead based on the interaction between x_2 and ϵ.

3 Estimation and inference

In this section, we develop a simultaneous variables estimator and present the details on its practical implementation. We then study its asymptotic properties by establishing consistency and asymptotic normality. We also develop inference procedures based on the Wald

[5] Strictly speaking, $E(\epsilon z) = 0$ is sufficient for the linear IV model. Similarly, Assumption 1 (ii) for the proposed model can be relaxed to weaker condition which is written as unconditional expectation. But the conditional version is adopted for better motivation and coherence.

statistic. Finally, we propose a simple test for endogeneity.

3.1 Estimation

Given the identification result in Theorem 1, we are able to estimate the parameters of interest. We construct an estimator which is simple to implement in practice. Recall that $\boldsymbol{\theta} \equiv [\boldsymbol{\beta}_1, \boldsymbol{\alpha}]^\top$ with $\boldsymbol{\alpha} \equiv [\beta_2, \boldsymbol{\phi}]^\top$. An estimator of $\boldsymbol{\theta}$ is as following

$$\widehat{\boldsymbol{\alpha}} = \left(\frac{1}{n}\sum_{i=1}^{n} \widetilde{\boldsymbol{z}}_i^\top \widehat{\boldsymbol{x}}_i\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} \widetilde{\boldsymbol{z}}_i^\top \widehat{y}_i\right), \tag{7}$$

$$\widehat{\boldsymbol{\beta}}_1 = \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_{1i}^\top \boldsymbol{x}_{1i}\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_{1i}^\top y_i\right) - \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_{1i}^\top \boldsymbol{x}_{1i}\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_{1i}^\top x_{2i}\right) \widehat{\beta}_2, \tag{8}$$

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 in practice is simple and 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 (7) 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 (8), 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.

3.2 Asymptotic theory

Now we establish the asymptotic properties of the estimator described above. In particular, we show its consistency and asymptotic normality. 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 $C_2 \equiv E(\boldsymbol{x}_{1i}^\top x_{2i})$. The limiting behavior of the estimator is summarized in the following result.

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* $E(\|\boldsymbol{Z}_i\|^4) < \infty$. *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 A. ■

The results given in Theorem 2 show that the limiting distribution of the proposed estimator is standard. This result is the foundation for construction of inference procedures.

3.3 Inference

Inference is very important in practice. Given the results in Theorem 2, general hypotheses on the vector $\boldsymbol{\theta}$ can be accommodated by Wald-type tests. We focus on linear restrictions. The Wald statistic and associated limiting theory provide a natural foundation for testing hypotheses as $R\boldsymbol{\theta} = r$, where R is a full-rank matrix imposing q restrictions on the parameters, and r is a column vector of q elements. The Wald statistic can be constructed as

$$\mathcal{W}_n = n(R\widehat{\boldsymbol{\theta}} - r)^\top [R\widehat{\Omega}R^\top]^{-1}(R\widehat{\boldsymbol{\theta}} - r), \tag{9}$$

where $\widehat{\Omega}$ is a consistent estimator of the variance-covariance matrix of Ω.

Under the null hypothesis $H_0 : R\boldsymbol{\theta} = r$, the statistic $\mathcal{W}_n$ is asymptotically χ^2_q with q-degrees of freedom, where q is the rank of the matrix R. The limiting distribution of the test is summarized in the following result.

Corollary 1 *Under $H_0 : R\theta(\tau) = r$, and the conditions of Theorem 2,*

$$\mathcal{W}_n \overset{a}{\sim} \chi^2_q.$$

Proof. Given a consistent estimator $\widehat{\Omega}$, the proof is a direct application of Theorem 2. ■

In practice, to carry inference and apply a Wald test in (9) one needs a consistent estimator of the asymptotic variance-covariance matrix. As described in the above result, to estimate the asymptotic variance-covariance matrix, we need to estimate both $Var(\widehat{\boldsymbol{\alpha}}) = Q^{-1}MQ^{-1}/n$ and $Var(\widehat{\beta}_1) = \boldsymbol{C}_1^{-1}C_2V_{\beta_2}C_2^\top\boldsymbol{C}_1^{-1}/n$. The later is easily recovered from its sample counterparts, that is, $\widehat{\boldsymbol{C}}_1 = n^{-1}\sum_{i=1}^n \boldsymbol{x}_{1i}^\top\boldsymbol{x}_{1i}$ and $\widehat{C}_2 = n^{-1}\sum_{i=1}^n \boldsymbol{x}_{1i}^\top x_{2i}$. Now, notice that $V_{\widehat{\beta}_2}$ is the first element of the variance-covariance matrix $\widehat{Var}(\widehat{\boldsymbol{\alpha}})$. Finally, for the estimation of the variance-covariance matrix of $\widehat{\boldsymbol{\alpha}}$ one can consider its sample counterpart such as $\widehat{Var}(\widehat{\boldsymbol{\alpha}}) = \widehat{Q}^{-1}\widehat{M}\widehat{Q}^{-1}/n$ with

$$\widehat{Q} = \frac{1}{n}\sum_{i=1}^n \widetilde{\boldsymbol{z}}_i^\top \widehat{\boldsymbol{x}}_i,$$

$$\widehat{M} = \frac{1}{n}\sum_{i=1}^n \left(\widetilde{\boldsymbol{z}}_i^\top \widehat{u}_i - \widehat{G}\widehat{r}_i(\delta) + \widehat{H}\widehat{s}_i(\gamma)\right)\left(\widetilde{\boldsymbol{z}}_i^\top \widehat{u}_i - \widehat{G}\widehat{r}_i(\delta) + \widehat{H}\widehat{s}_i(\gamma)\right)^\top$$

where

$$\widehat{u}_i = \widehat{y}_i - \widehat{\boldsymbol{x}}_i \widehat{\boldsymbol{\alpha}},$$

$$\widehat{G} = \frac{1}{n}\sum_{i=1}^{n}(\widetilde{\boldsymbol{z}}_i^\top \nabla_{\boldsymbol{\delta}} \widetilde{\boldsymbol{x}}_i \widehat{\alpha}) = \frac{1}{n}\sum_{i=1}^{n}(\widetilde{\boldsymbol{z}}_i^\top [-x_{2i}\boldsymbol{x}_{1i}, 0, 0]\widehat{\alpha}),$$

$$\widehat{H} = \frac{1}{n}\sum_{i=1}^{n}(\widetilde{\boldsymbol{z}}_i^\top \nabla_{\boldsymbol{\gamma}} \widetilde{y}_i) = \frac{1}{n}\sum_{i=1}^{n}(\widetilde{\boldsymbol{z}}_i^\top (-x_{2i}\boldsymbol{x}_{1i})),$$

$$\widehat{r}_i(\widehat{\boldsymbol{\delta}}) = \left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{1i}^\top \boldsymbol{x}_{1i}\right)^{-1}\left(\boldsymbol{x}_{1i}^\top (x_{2i} - \boldsymbol{x}_{1i}\widehat{\boldsymbol{\delta}})\right),$$

$$\widehat{s}_i(\widehat{\boldsymbol{\gamma}}) = \left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{1i}^\top \boldsymbol{x}_{1i}\right)^{-1}\left(\boldsymbol{x}_{1i}^\top (y_i - \boldsymbol{x}_{1i}\widehat{\boldsymbol{\gamma}})\right),$$

$$\widehat{\boldsymbol{\delta}} = \left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{1i}^\top \boldsymbol{x}_{1i}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{1i}^\top x_{2i}\right),$$

and

$$\widehat{\boldsymbol{\gamma}} = \left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{1i}^\top \boldsymbol{x}_{1i}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{1i}^\top y_i\right).$$

The practical estimation of these quantities is simple. As for the point estimates, it only relies on simple OLS regressions.

3.4 Endogeneity test

When $\boldsymbol{x}_2$ and ϵ in equation (2) are correlated, the usual OLS estimator is inconsistent. Thus, it is necessary to resort to estimators discussed in (7) and (8) above. In this subsection we consider a test for exogeneity of $\boldsymbol{x}_2$ based on classical principles, as in Rivers and Vuong (1988).

To test exogeneity we consider the null hypothesis of $H_0 : \boldsymbol{\phi} = \mathbf{0}$. This is a joint test for the coefficients $\boldsymbol{\phi}$ in the simultaneous variables model. Notice that under the null hypothesis, from Assumption 1 (ii) and the law of the iterated expectations, $E(x_2\epsilon) = 0$, and therefore, there is no endogeneity problem. Thus, the modified Wald statistic we consider is given by

$$TE = n\widehat{\boldsymbol{\phi}}^\top \widehat{V}(\widehat{\boldsymbol{\phi}})\widehat{\boldsymbol{\phi}}, \tag{10}$$

where $\widehat{V}(\widehat{\boldsymbol{\phi}})$ is a consistent estimator of the variance-covariance matrix corresponding to $\widehat{\boldsymbol{\phi}}$, and $\boldsymbol{\phi}$ is estimated by the simultaneous variables estimator.

Following Rivers and Vuong (1988), under the null hypothesis the test statistic in (10) has an asymptotic central chi-square distribution with m degrees of freedom, where m is the number of simultaneous variables included in the estimation equations (7) and (8).

4 Extensions

4.1 Nonparametric model

We have focused on a simple linear parametric model for the sake of clear exposition and motivation. Nevertheless, the model we consider can be extended to more general additively separable nonparametric models. The structural model (1) is now rewritten as

$$y_i = f(\boldsymbol{x}_{1i}, x_{2i}) + \epsilon_i, \quad i = 1, ..., n, \tag{11}$$

where x_2 is endogenous and ϵ is an innovation term. Then, from the equations (4) and (11) we obtain

$$E(x_2(y - f(\boldsymbol{x}_1, x_2)) \mid \boldsymbol{z}, \boldsymbol{x}) = g(\boldsymbol{z}),$$

or

$$\begin{aligned} E(x_2 y \mid \boldsymbol{z}, \boldsymbol{x}) &= \widetilde{f}(\boldsymbol{x}_1, x_2) + g(\boldsymbol{z}) \qquad (12) \\ &= h(\boldsymbol{w}), \end{aligned}$$

where $\widetilde{f}(\boldsymbol{x}_1, x_2) \equiv x_2 f(\boldsymbol{x}_1, x_2)$ and $\boldsymbol{w} \equiv [\boldsymbol{x}_1, x_2, \boldsymbol{z}]$. This implies that the function $\widetilde{f}(\boldsymbol{x}_1 x_2)$, can be obtained from an additive regression of $x_2 y$ on $(\boldsymbol{x}_1, x_2)$ and $\boldsymbol{z}$. Thus, identification of the function of interest, $f(\boldsymbol{x}_1, x_2)$, rests on the identification of $\widetilde{f}(\boldsymbol{x}_1, x_2)$ through $f(\boldsymbol{x}_1, x_2) = \widetilde{f}(\boldsymbol{x}_1, x_2)/x_2$.

Identification of $\widetilde{f}(\boldsymbol{x}_1, x_2)$ can be analyzed by a similar argument of Theorem 2.1 in Newey, Powell, and Vella (1999). To be specific, uniqueness of a conditional expectation is equivalent to the statement that any additive function $\widetilde{f}^*(\boldsymbol{x}_1, x_2) + g^*(\boldsymbol{z})$ satisfying equation (12) must have $\text{Prob}(\widetilde{f}^*(\boldsymbol{x}_1, x_2) + g^*(\boldsymbol{z}) = \widetilde{f}(\boldsymbol{x}_1, x_2) + g(\boldsymbol{z})) = 1$. Thus, identification implies equality of the additive components up to a constant. Intuitively speaking, failure of identification implies a functional relationship between the random vector $(\boldsymbol{x}_1, x_2)$ and $\boldsymbol{z}$. Therefore, a sufficient condition for identification is the absence of an exact relationship between these random variables. We summarize the identification result in the following Lemma.

Lemma 1 $\widetilde{f}(\boldsymbol{x}_1, x_2)$ *is identified, up to an additive constant, if and only if Prob*$(\delta(\boldsymbol{x}_1, x_2) + \lambda(\boldsymbol{z}) = 0) = 1$ *implies there is a constant* c_f *with Prob*$(\delta(\boldsymbol{x}_1, x_2) = c_f) = 1$.

Proof. The result follows from a similar argument to Theorem 2.1 in Newey, Powell, and Vella (1999). ■

4.2 General forms of endogeneity

We have characterized the endogeneity based on the linear dependence (i.e., expectation of the product of x_2 and ϵ). More generally, modeling endogeneity can be defined in terms of any nonlinear dependence between x_2 and ϵ. To be precise, instead of equation (4), we assume

$$E(\Lambda(x_2, \varepsilon; \boldsymbol{\beta}) \mid \boldsymbol{z}, \boldsymbol{x}) = g(\boldsymbol{z}; \boldsymbol{\phi}), \tag{13}$$

where $\Lambda(x_2, \varepsilon; \boldsymbol{\beta})$ is a known measurable function of (x_2, ε) up to an unknown vector of parameters $\boldsymbol{\beta}$ and where $g(\boldsymbol{z}; \boldsymbol{\phi})$ is a known measurable function of $\boldsymbol{z}$ up to an unknown vector of parameters $\boldsymbol{\phi}$. We focus on a nonlinear parametric model as follows:

$$y_i = f(\boldsymbol{x}_{1i}, x_{2i}; \boldsymbol{\beta}) + \epsilon_i, \quad i = 1, ..., n,$$

where $f(\boldsymbol{x}_1, x_2; \boldsymbol{\beta})$ is a known measurable function of $(\boldsymbol{x}_1, x_2)$ up to an unknown vector of parameters $\boldsymbol{\beta}$.

Let $\Psi(y, \boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\theta}) \equiv \Lambda(x_2, y - f(\boldsymbol{x}_1, x_2; \boldsymbol{\beta}); \boldsymbol{\beta}) - g(\boldsymbol{z}; \boldsymbol{\phi})$ with $\boldsymbol{\theta} \equiv [\boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top]^\top$. Then equation (13) can be rewritten as the following moment condition,

$$E(\Psi(y, \boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\theta}) \mid \boldsymbol{z}, \boldsymbol{x}) = 0.$$

Therefore, the orthogonality conditions can be rewritten as

$$E(\widetilde{\boldsymbol{z}} \Psi(y, \boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\theta})) = 0, \tag{14}$$

where $\widetilde{\boldsymbol{z}}$ is a set of variables generated by the conditioning variables $(\boldsymbol{z}, \boldsymbol{x})$. It is important to notice that the model in equation (14) allows for over-identified models where the dimension of $\widetilde{\boldsymbol{z}}$ is larger than the dimension of regressors. Assuming global identification of parameters and full rank condition for matrices associated with Assumption 2, a minimum-distance estimator of the parameter $\boldsymbol{\theta}$ based on equation (14) is obtained by solving the following minimization problem

$$\widehat{\boldsymbol{\theta}} \equiv \text{argmin}_{\boldsymbol{\theta}} n h_n(\boldsymbol{v}_i; \boldsymbol{\theta})^\top \widehat{W} h_n(\boldsymbol{v}_i; \boldsymbol{\theta}),$$

where $\widehat{W}$ is a consistent estimator of a positive definite matrix W, $\boldsymbol{v} \equiv [y, \boldsymbol{x}, \boldsymbol{z}]$, and

$$h_n(\boldsymbol{v}_i; \boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \widetilde{\boldsymbol{z}}_i \Psi(y_i, \boldsymbol{x}_i, \boldsymbol{z}_i).$$

Asymptotic properties of the above estimator can be established by using standard econometric results (see, e.g., Newey, 1990).

5 Monte Carlo simulations

We use simulation experiments to assess the finite sample performance of the proposed simultaneous variables (SV) estimator. For comparison, we also report results for the standard ordinary least squares (OLS) and the instrumental variables (IV) estimators. The resulting estimates are compared in terms of bias, median absolute deviation (MAD) and root mean square error (RMSE). The simulation experiments are based on 5,000 replications and sample sizes of $n \in \{100, 200, 500, 1000\}$.

We consider a simple data generating process (DGP) as

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i, \quad i = 1, 2, ..., n,$$

with $\beta_0 = \beta_1 = \beta_2 = 1$, with covariates generated as $x_1 = v_1 + v_2$ and $x_2 = v_2 + v_3$, where $v_j \sim i.i.d.N(0, 1)$, $j = 1, 2, 3$.

First, we generate a model to evaluate the relative performance of the SV estimator when both x_1 and x_2 variables are exogenous. Thus, there is no need to correct for endogeneity. In this case we generate $\epsilon = v_4$, $v_4 \sim i.i.d.N(0, 1)$, such that there is no correlation between ϵ and x_1 or x_2. In this scenario, we expect the standard OLS estimator, which does not make use of any additional variables for estimation, of both β_1 and β_2 to be approximately unbiased. Both the SV and IV use additional variables. So, we construct different sets of variables to study these estimators. In all the simulation experiments, the structural equation is kept the same and the SV and IV models differ only in the DGP of z. In this first set of simulations, two different cases are considered. We construct a simultaneous variable $z_{sv} = v_2 v_4 + v_5$ (Model SV), and an instrumental variable $z_{iv} = v_2 + v_5$ (Model IV), where $v_5 \sim i.i.d.N(0, 1)$. Then, we use the two variables, $\{z_{sv}, z_{iv}\}$, alternatively for the different estimators, that is, in the role of SV or IV for both estimators. It is important to notice that z_{sv} contains v_2 and v_4 and hence is correlated with $x_2\epsilon$. The rationale for such an experiment is to evaluate what would be the effect of using a variable for the corresponding estimator and (wrongly) for another estimator, when there is no endogeneity.

Table 1 reports the simulation results for β_2 for this case. Note that all estimators are unbiased and consistent when the correct variable is used. As expected OLS is consistent, but also is the SV estimator when z_{sv} is used as simultaneous variable and IV when z_{iv} is used as instrument. Also note that the SV estimator is also consistent when any variable is used, i.e. $\{z_{sv}, z_{iv}\}$. However, the IV estimator only works when a proper instrument is used, and produces biased estimates with significant variance when an incorrect instrument is used instead. That is, if we use z_{sv} as instrument, then by construction, it is not correlated with x_2 (that is, $Cov(z_{sv}, x_2) = Cov(v_2v_4 + v_5, v_2 + v_3) = 0$), and then the IV estimator variance is large with a consequent effect on MAD and RMSE. On the other hand, SV works with both $\{z_{sv}, z_{iv}\}$ because it does not affect the β estimators when $\phi = 0$.

[TABLE 1 ABOUT HERE]

Second, consider a DGP model with endogeneity. In this case, we generate the innovation term as $\epsilon = v_3 + v_4$, where $v_4 \sim i.i.d.N(0, 1)$. This model induces endogeneity because $corr[x_1, x_2] = corr[x_2, \epsilon] = 0.5$ while $corr[x_1, \epsilon] = 0$. Given the correlation between the x_2 regressor and the error term in the DGP, the standard OLS estimators of both β_1 and β_2 should be biased, and in particular, $E[\widehat{\beta}_2^{OLS}] = \frac{5}{3} > 1$ so that the OLS bias is $\frac{2}{3}$. As in the previous case, in all the simulation experiments, the structural equation is kept the same and the models differ only in the DGP of z. Three different cases are considered. Let $v_5 \sim i.i.d.N(0, 1)$. In Model SV1, we generate a DGP following a simultaneous variable framework. From the additional equation $x_2\epsilon = v_3^2 + v_2v_3 + v_2v_4 + v_3v_4$, we set $z = v_3^2 + v_2v_3$. Note that $corr[z, x_2\epsilon - z] = 0$ and $corr[x_2^2, x_2\epsilon - z] = 0$. Thus, Model SV1 satisfies the simultaneous variables requirements, and the proposed estimator is expected to be consistent.

In Model SV2, we consider $z = 1 + x_2\epsilon + v_5$. In this case, z is a noisy measure of $x_2\epsilon$. Finally, Model IV is designed to satisfy the IV method requirements. We construct z as an instrumental variable such as $z = v_2 + v_5$. In this case, the IV estimator is expected to perform the best. These simulation results (for β_2) are presented in Table 2.

[TABLE 2 ABOUT HERE]

The results for models SV1 and SV2 show that, as expected, the proposed SV performs very well in terms of bias, MAD and RMSE. The proposed estimator is approximately unbiased, even for small sample sizes, and the bias, MAD and RMSE decrease as the sample size increases. In contrast, OLS and IV estimators are largely biased, and their biases do not disappear for large n. Finally, Model IV is favorable to the IV estimator. In this case, the IV estimator is approximately unbiased while OLS and SV are not.

Overall, the results show evidence that the SV performs very well in finite sample simulations when a proper simultaneous variable is used. The SV estimator is able to reduce the endogeneity bias when a simultaneous variable exists that is related to the joint interaction of the endogenous variable and the innovations. The last two experiments also show that there are fundamental differences between our estimator and other methods for solving endogeneity such as IV. In fact, the IV estimator works only for Model IV.

6 Empirical application: Investment equation

6.1 Investment model

To provide a formal economic framework for the discussion, we present a standard dynamic investment model as in Abel and Eberly (1994) and Erickson and Whited (2000). The model considers risk-neutral managers choosing investment each period to maximize the expected present value of future profits. Capital is a quasi-fixed factor. Thus, we can write the value function for firm i at time t as

$$V_{it} = E\left[\sum_{j=0}^{\infty}\left(\prod_{s=1}^{j} b_{i,t+s}\right)\Pi(K_{i,t+j},\xi_{i,t+j}) - \psi(I_{i,t+j},K_{i,t+j},\nu_{t,t+j},h_{i,t+j}) - I_{i,t+j}|\Omega_{it}\right], \quad (15)$$

where E is the expectation operator, Ω_{it} is the information set of the manager of firm i at time t, b_{it} is the firm's discount factor at time t; K_{it} is the beginning of period capital stock, I_{it} is investment, $\Pi(K_{it},\xi_{it})$ is the profit function, and $\Pi_K > 0$, $\psi(I_{i,t+j},K_{i,t+j},\nu_{t,t+j},h_{i,t+j}) - I_{i,t+j}$ is the investment adjustment cost function, which is increasing in I_{it} and decreasing in K_{it} and convex in both arguments. The term h_{it} is a vector of variables, such as labor productivity that might affect adjustment costs, and ξ_{it} and ν_{it} are exogenous stocks to the prefect and adjustment cost functions, both are observed by the manager but not by the econometrician at time t. All variables are expressed in real terms, and the relative price of capital is normalized to unity. We note that any variable factors of production have already been maximized out of the problem.

The firm maximizes the value function in equation (15) subject to the following capital stock accounting constraint

$$K_{i,t+1} = (1-d_i)K_{it} + I_{it}, \quad (16)$$

where d_i is the constant rate of capital depreciation for firm i. For simplicity, we assume the depreciation is constant over time. Let χ_{it} be the sequence of Lagrange multipliers on equation (16). To solve the optimization problem, the first-order condition for maximizing the value of the firm in equation (15) subject to (16) is

$$1 + \psi_I(I_{it},K_{it},\nu_{it},h_{it}) = \chi_{it}, \quad (17)$$

where

$$\chi_{it} = E\left[\sum_{j=0}^{\infty}\left(\prod_{s=1}^{j} b_{i,t+s}\right)(1-d_i)^{j-1}[\Pi_K(K_{i,t+j},\xi_{i,t+j}) - \psi_K(I_{i,t+j},K_{i,t+j},\nu_{t,t+j},h_{i,t+j})]|\Omega_{it}\right]. \quad (18)$$

These equations can be interpreted economically. Equation (17) states the marginal cost of investment equals its expected marginal benefit. The left side captures the adjustment and purchasing costs of capital goods, and the right side represents the expected shadow value of capital, which, as shown in (18), is the expected stream of future marginal benefits from using the capital. These benefits include both the marginal additions to profit and reductions in installation costs. Since we normalize the price of capital goods to unit, χ_{it} represents the quantity "marginal q" (see e.g., Tobin, 1969).

There is a large literature testing the q theory using linear regression models with the rate of investment on χ_{it}. These procedures require a proxy for the unobservable χ_{it} and a functional form for the installation cost function having a partial derivative with respect to I_{it} that is linear in I_{it}/K_{it} and ν_{it}. In this paper, we follow the literature and consider the problem of obtaining χ_{it}. Moreover, we propose to approximate the cost function using a polynomial. This is a class of functions that meets the functional form requirements and is also linearly homogeneous in I_{it} and K_{it}. Thus, we use the following function

$$\psi_K(I_{i,t+j}, K_{i,t+j}, \nu_{t,t+j}, h_{i,t+j}) = (a_1 + a_2\nu_{it})I_{it} + a_3\frac{I_{it}^2}{K_{it}} + K_{it}f(\nu_{it}, h_{it}). \tag{19}$$

Here f is an integrable function, and a_1, a_2, a_3 are constants. It is standard to restrict $a_3 > 0$ to ensure concavity of the value function in the optimization problem. The adjustment cost functions chosen either explicitly or implicitly by researchers who test q theory with linear regressions are variants of (19). Differentiating (19) with respect to I_{it} and substituting the result into (17) yields the following regression equation

$$y_{it} = \alpha_0 + \beta\chi_{it} + u_{it}, \tag{20}$$

where $y_{it} \equiv I_{it}/K_{it}$, $\alpha_0 \equiv -(1 + a_1)/2a_3$, $\beta \equiv 1/2a_3$, and $u_{it} \equiv -a_2\nu_{it}/2a_3$.

Therefore, from the maximization of firm's profit, and an assumption of linearly homogenous cost function, we are able to derive the linear investment model. It is standard in the literature to augment model (20) with cash flow as a regressor. Fazzari, Hubbard, and Petersen (1988) consider an investment equation model, where a firm's investment is regressed on an observed measure of investment demand (Tobin's q) and cash flow. Following Fazzari, Hubbard, and Petersen (1988), investment–cash-flow sensitivities became a standard metric that examines the impact of financing imperfections on corporate investment (Stein, 2003). These empirical sensitivities are also used for drawing inferences about efficiency in internal capital markets (Lamont, 1999; Shin and Stulz, 1998), the effect of agency on corporate spending (Hadlock, 1998; Bertrand and Mullainathan, 2005), the role of business groups in capital allocation (Hoshi, Kashyap, and Scharfstein, 1991), and the effect of managerial characteristics on corporate policies (Bertrand and Schoar, 2003; Malmendier and Tate, 2005). Therefore, the econometric model is represented by the following equation

$$y_{it} = \gamma + \alpha c_{it} + \beta q^*_{it} + \eta_{it}, \tag{21}$$

where $y_{it} \equiv I_{it}/K_{it}$, where I denotes investment and K capital stock, q^* is marginal Tobin's q, $c_{it} \equiv CF_{it}/K_{it}$, where CF is cash flow, and η is the unobserved causes of the investment, which can be interpreted as containing, among other factors, a firm specific economic shock to investment. This shock may affect firms differently because of different perceptions on the aggregate conjecture. As a result, η is a pure idiosyncratic shock to the investment.

6.2 Investment model with measurement errors

Concerns about measurement errors (ME) have been emphasized in the context of the investment equation model. Theory suggests that the correct measure for a firm's investment demand is captured by *marginal* q, but this quantity is unobservable and researchers use instead its measurable counterpart, *average* q. Because average q measures marginal q imperfectly, a measurement problem naturally arises (Hayashi, 1982; Poterba, 1988).

Now we consider the structure of the ME on q^* more carefully. We use a classical ME framework to formalize the discussion. Let e_{it} denote the Tobin's q measurement error. Assume that e_{it} has zero mean, finite variance σ_e^2 and is independent of q_{it}^*. Thus, the average q is given by

$$q_{it} = q_{it}^* + e_{it}. \tag{22}$$

From plugging equation (22) into the investment model in equation (21), we obtain

$$\begin{aligned} y_{it} &= \gamma + \alpha c_{it} + \beta(q_{it} - e_{it}) + \eta_{it} \\ &= \gamma + \alpha c_{it} + \beta q_{it} + \epsilon_{it}, \end{aligned} \tag{23}$$

where $\epsilon_{it} \equiv \eta_{it} - \beta e_{it}$. Therefore, when q^* is unobserved and a mismeasured q is used as a regressor, ME induce endogeneity, i.e., q is correlated with ϵ due to ME.

Therefore, in the standard investment model in (21), if q^* is measured with error, OLS estimates of β will be biased. In addition, given that average q and cash flow are likely to be correlated, the coefficient α is likely to be biased as well. These biases are expected to be reduced by the use of estimators that solve the ME problem. It has been common in the literature to employ instrumental variables estimators to resolve the problem (see, e.g., Almeida, Campello, and Galvao, 2010; Lewellen and Lewellen, 2014).

In order to better evaluate the performance of the alternative estimators in practice, we describe some hypotheses about the effects of measurement error correction on the estimated coefficients β and α from equation (21). Theory does not pin down the exact values that these coefficients should take. Nevertheless, one could argue that the two following conditions should be reasonable.

First, an estimator that solves ME in q^* in a standard investment equation should return a higher estimate for β and a lower estimate for α when compared with standard OLS estimates. ME cause an attenuation bias on the estimate for the coefficient β. In addition, since q and cash flow are likely to be positively correlated, ME should cause an upward bias on the empirical estimate of α under the standard OLS estimation. Thus, denoting the OLS and the measurement error corrected (MEC) estimates, respectively, by (β^{OLS}, α^{OLS}) and (β^{MEC}, α^{MEC}), one should expect:

$$\text{Condition 1:} \quad \beta^{OLS} < \beta^{MEC} \text{ and } \alpha^{OLS} > \alpha^{MEC}.$$

Second, one would expect the coefficient for q to be positive and the cash flow coefficient to be non-negative after controlling for measurement error. The q-theory of investment predicts a positive correlation between investment and q (e.g. Hayashi, 1982). Under this theory, the cash flow coefficient should be zero ("neoclassical view"), after controlling for ME. However, in practice the cash flow coefficient could be positive because of either the presence of financing frictions (e.g. Fazzari, Hubbard, and Petersen, 1988), or the fact that cash flow captures variation in investment opportunities even after we apply a correction for mismeasurement in q.[6] Therefore, one should observe:

$$\text{Condition 2:} \quad \beta^{MEC} > 0 \text{ and } \alpha^{MEC} \geq 0.$$

Notice that these conditions are fairly intuitive. If a particular ME corrected estimator does not deliver these basic results, one should have reasons to question the usefulness of that estimator in applied work.

6.3 Instrumental variables models

One potential solution to the endogeneity problem is the IV method. Conventional IV approaches to solve the measurement error in investment equation models employ lags of q_{it} as instruments for q_{it}.

[6] However, financial constraints are not sufficient to generate a strictly positive cash flow coefficient because the effect of financial constraints is capitalized in stock prices and may thus be captured by variations in q (Chirinko, 1993; Gomes, 2001).

In order to make the instruments valid, such that the IV approach removes the bias, assumptions on the dynamics of measurement errors need to be imposed. Almeida, Campello, and Galvao (2010) discuss such condition in details. In the simplest case, if the measurement error e_{it} in equation (22) is i.i.d. across firms and time, and q_{it} is serially correlated, then, for example, q_{it-2}, q_{it-3}, or $(q_{it-2} - q_{it-3})$ are valid as instruments for q_{it}, because, for instance, in equation (23) q_{it-2} is correlated with q_{it} but uncorrelated with ϵ_{it}. The resulting instrumental variables estimator is consistent.

There are two simple ways to relax the standard assumption of i.i.d. measurement errors. First, one can consider an assumption of time-invariant autocorrelation. Then twice-lagged levels of the observable variables can be used as instruments. Second, under the first-order moving average structure (i.e., MA(1)) for the measurement error, the consistency of the IV estimator requires the use of instruments that are lagged two periods or longer. Additionally, the identification requires the latent regressor to have some degree of autocorrelation (since lagged values are used as instruments). See Biorn (2000) for more details.

However, the approach of using lagged q_{it} as IV's fails when the measurement errors on the marginal q are systematic. In practice, it is highly likely that current-period measurement error is correlated with the first-order or higher-order lags of the measurement error. This phenomenon of autocorrelation in the measurement errors seems to be more realistic in practice because systematic errors on a way of evaluating average q would be highly likely to be persistent. Then, in general e_{it} could depend on $\{e_{it-1}, e_{it-2}, e_{it-3}, ...\}$. Under the presence of autocorrelation in the measurement errors (e.g., AR(1)), the IV approach will not work since the estimates are biased due to the dependence of the instrument (q_{it-1}) and the current-period measurement error (e_{it}) through the first-order lag of measurement error (e_{it-1}).

6.4 A preview of our solution

In this paper we solve the ME endogeneity problem by explicitly modeling the conditional expectation of the interaction $q_{it}\epsilon_{it}$ in (23) given a set of simultaneous variables $\boldsymbol{z}_{it}$. In other words, we model $E(q_{it}\epsilon_{it}|\boldsymbol{z}_{it}, c_{it}, q_{it})$ as a function of observables, $\boldsymbol{z}_{it}$, and estimate the parameters of interest consistently. In this section, we show that, contrary to IV, even under the assumption that the marginal q and the ME are correlated, lagged Tobin's q and its square are valid simultaneous variables, $\boldsymbol{z}_{it}$. This result holds because we can show that these simultaneous variables are simultaneously related to the endogenous variable q and the error term ϵ. Therefore, we use the proposed methods with the following specification for Assumption 1 (ii) and a set of observable variables $\boldsymbol{z}_{it}$ such that

$$E(q_{it}\epsilon_{it}|\boldsymbol{z}_{it}, c_{it}, q_{it}) = \boldsymbol{Z}_{it}\boldsymbol{\phi}, \tag{24}$$

where $\boldsymbol{Z}_{it} = [q_{it-1}, q_{it-1}^2]$.

Now we discuss the choice of these the simultaneous variables $\boldsymbol{Z}$ and show they are related to the interaction term $q\epsilon$. The ME bias in the OLS estimator arises from the fact that $E[q_{it}\epsilon_{it}] \neq 0$. Note that

$$q_{it}\epsilon_{it} = (q_{it}^* + e_{it})(\eta_{it} - \beta e_{it}) = (q_{it}^* + e_{it})\eta_{it} - \beta(q_{it}^* e_{it} + e_{it}^2). \tag{25}$$

From the maximization of firm's profit, we can derive the linear investment model (21) by assuming linearly homogenous cost function. As a result, η is a pure idiosyncratic shock to the investment and the model implies that there is no correlation between (q^*, e) and η. Furthermore, η can be assumed to be i.i.d. and with zero mean in equation (21). Thus, our main concern is the second term in equation (25), $(q_{it}^* e_{it} + e_{it}^2)$.

We need to show that the simultaneous variables, lagged Tobin's q and its square, are valid variables, that is, they are related to the interaction term $q_{it}\epsilon_{it}$. To establish the result, let both q^* and e be autocorrelated such that lags of q can be used to eliminate the endogeneity. Because both the endogenous variable q and the error term ϵ are functions of q^* and e, a polynomial function of the lags of q may be able to model $(q^*_{it}e_{it} + e^2_{it})$. Thus, the main idea is to model this explicitly.

Consider the example where both q^* and e follow AR(1) processes:

$$q^*_{it} = \rho^q_0 + \rho^q_1 q^*_{it-1} + w^q_{it}$$
$$e_{it} = \rho^e_1 e_{it-1} + w^e_{it},$$

where $|\rho^q_1| < 1$, $|\rho^e_1| < 1$ and w^q and w^e are i.i.d and zero mean processes.[7] Then,

$$\begin{aligned} q^*_{it}e_{it} + e^2_{it} &= (\rho^q_0 + \rho^q_1 q^*_{it-1} + w^q_{it})(\rho^e_1 e_{it-1} + w^e_{it}) + (\rho^e_1 e_{it-1} + w^e_{it})^2 \\ &= \rho^q_0\rho^e_1 e_{it-1} + \rho^q_1\rho^e_1 q^*_{it-1}e_{it-1} + (\rho^e_1)^2 e^2_{it-1} + \varpi_{it}, \end{aligned}$$

where $\varpi_{it} \equiv \rho^q_0 w^e_{it} + \rho^q_1 q^*_{it-1} w^e_{it} + \rho^e_1 w^q_{it} e_{it-1} + w^q_{it} w^e_{it} + (w^e_{it})^2 + 2\rho^e_1 e_{it-1} w^e_{it}$ and $E[\varpi_{it} \mid e_{it-1}, q_{it-1}] = 0$.

Now consider a set of variables that contains a polynomial of the lags of q_{it}. For instance, consider a quadratic polynomial of q_{it-1}. Thus, we have

$$q_{it-1} = q^*_{it-1} + e_{it-1},$$
$$q^2_{it-1} = (q^*_{it-1} + e_{it-1})^2 = q^{*2}_{it-1} + 2q^*_{it-1}e_{it-1} + e^2_{it-1}.$$

From equation (25) and the above derivations, one can recognize that $\boldsymbol{Z}_{it} = [q_{it-1}, q^2_{it-1}]$ contains valuable information on $(q^*_{it}e_{it} + e^2_{it})$ and consequently $q_{it}\epsilon_{it}$. This supports the idea that the information content provided by the simultaneous variables is sufficient enough to explain the dependence between the endogenous q_{it} and the error term ϵ_{it}, such that equation (24) should be satisfied. In general, one could consider a higher order polynomial of the lags of q_{it} for firm i obtaining the set of simultaneous variables as $\boldsymbol{Z}_{it} = [1, q_{it-1}, q^2_{it-1}, ..., q^m_{it-1}, q_{it-2}, q^2_{it-2}, ..., q^m_{it-2}, ..., q_{it-j}, ..., q^m_{it-j}]$, where m and j are integers.

[7]There is no constant term in the process for e_{it} without loss of generality.

6.5 Data

The data are taken from COMPUSTAT and cover 1970 to 2005, and the data collection process follows that of Almeida and Campello (2007). The sample consists of manufacturing firms with fixed capital of more than $ 5 million (with 1976 as the base year for the cpi), and the sample firms have growth of less than 100% in both assets and sales. Summary statistics for investment, q, and cash flow are presented in Table 3. These statistics are similar to those reported by Almeida and Campello (2007), among other papers. To save space, we omit the discussion of these descriptive statistics.

[TABLE 3 ABOUT HERE]

6.6 Estimates and tests

We estimate the above investment equation using OLS, instrumental variables (IV), and simultaneous variables models. The results for OLS and IV are summarized in Table 4 and for the simultaneous variables in Table 5. For the IV we use different combinations of lags of Tobin's q as instruments. In the same fashion, we use lags of the regressors as simultaneous variables.

Regarding the OLS estimates, we obtain the standard result in the literature that both q and CF attract positive coefficients which are statistically significant (see Table 4). In the OLS specification, we obtain a q coefficient of 0.0626, and a cash flow coefficient of 0.1307, which are likely to be biased. These results are consistent with those in the literature. In particular, Erickson and Whited (2000) and Almeida, Campello, and Galvao (2010) report very similar results for OLS estimates.

Now we consider the IV estimates. We first test the hypothesis of no first-order autocorrelation in the residuals of the simple OLS regression using a simple Durbin-Watson test (DW). We find that the coefficient for the AR(1) model (of the residuals) is 0.474 with standard error of 0.007, and the DW statistic is 1.051, which rejects the null hypothesis of no first-order autocorrelation. This result evidences that lags of q are likely invalid instruments for the IV case. In spite of the testing results, the point estimates are in Table 4. As expected, following the IV procedure, we obtain estimates for q and cash flow that do not satisfy the conditions discussed above, although they are statistically significant. In particular, IV estimators do not reduce the cash flow coefficient relative to that obtained by the standard OLS, while the q coefficients even decrease. For example, in the third column of Table 4 (IV2), which uses q_{t-1} and q_{t-1}^2 as instruments, the q coefficient is 0.0496 and CF coefficient is 0.1333. We also examine the robustness of the estimators to variations in the set of instruments used in the estimation, including sets that feature longer lags of the variables and the squares of them in the model. Columns (IV3)-(IV8) of Table 4 present these results and show that the q and CF coefficients from IV estimators do not improve substantially with the selection of instruments used. When the first-order of q is not included as instrument, the q coefficients are even worse. Therefore, given the concern about measurement error, these results show that the IV estimates are inconsistent with Conditions 1 and 2 above, and there is strong evidence that the instrumental variables approach is not able to solve the measurement error problem in this investment equation example.

The results for the simultaneous variables case are presented in Table 5. All coefficients for q and CF are statistically different from zero. By comparison, the simultaneous variables approach delivers results that are consistent with Conditions 1 and 2. The q coefficient is almost twice as large as the standard OLS estimate, depending on the set of variables used. At the same time, the cash flow coefficient goes down by about 10% from the standard OLS value. In particular, the results in column S2, using q_{t-1} and q_{t-1}^2 as simultaneous variables, show that the q coefficient increases from 0.0626 for OLS to 0.1107 for the simultaneous variables approach, while the cash flow coefficient drops from 0.1307 for OLS to 0.1208 for the simultaneous variables approach. These results suggest that the proposed estimator does a fairly reasonable job at addressing the measurement error problem. It is worth noting that choosing proper simultaneous variables is important to obtain consistent estimates. When q_{t-1} and q_{t-1}^2 are chosen (column S2), the proposed estimator obtains the most reasonable results. When other further lags are added (column S8), the results do not change substantially. When q_{t-1} and q_{t-1}^2 are omitted, however, the q coefficients reduce. This issue is analogous to the instrumental variables approach of which consistency heavily relies on the selection of valid instruments. As a result, it would be interesting and important to study the optimal selection of simultaneous variables. We leave it as a future research. Nevertheless, the estimates from our proposed approach are always consistent with Conditions 1 and 2, under different sets of simultaneous variables.[8]

[TABLE 4 ABOUT HERE]

[TABLE 5 ABOUT HERE]

In all, our empirical tests support the idea that simultaneous variables estimators are likely to outperform the IV estimators in economics and finance applications in which autocorrelated measurement error is a relevant concern.

7 Conclusion

This paper proposes a new methodology for solving endogeneity bias in econometric models. The main identification condition explicitly models endogeneity bias in a direct way as a function of additional variables. These, however, play a different role from the instrumental

[8]For the robustness check of the performances of the proposed estimator, higher-order moments such as the cubes of the lags of q have been included, but they do not change the results.

variables, and hence the new proposal can be viewed as a complement to the IV. Under the specified conditions, we establish identification of the parameters of interest, propose an estimator and derive its consistency and asymptotic normality of the proposed estimator. We also develop inference procedures based on the Wald statistics. Furthermore, we illustrate the usefulness of the proposed methods by revisiting the investment equation model. The scope of potential applications using the new techniques is large because it relies on mild conditions for identification of parameters.

Many issues remain to be investigated in future research. There are many variants of the model that would extend the structure for the simultaneous variables. A natural extension would be to panel data. The analysis of the performance of the methods and more general testing procedures is also important direction. Applications to program evaluation studies would appear to be a natural laboratory for further development of simultaneous variables methodology.

Appendix

A. Proof of the Theorems

Proof of Theorem 1

Note that from Assumption 1, $E(x_2\epsilon - \boldsymbol{Z}\boldsymbol{\phi} \mid \boldsymbol{z}, \boldsymbol{x}) = E(x_2(y - \boldsymbol{x}_1\boldsymbol{\beta}_1 - x_2\beta_2) - \boldsymbol{Z}\boldsymbol{\phi} \mid \boldsymbol{z}, \boldsymbol{x}) = E(x_2 y - x_2\boldsymbol{x}_1\boldsymbol{\beta}_1 - x_2 x_2\beta_2 - \boldsymbol{Z}\boldsymbol{\phi} \mid \boldsymbol{z}, \boldsymbol{x}) = E(x_2 y - x_2\boldsymbol{x}_1(E(\boldsymbol{x}_1^\top\boldsymbol{x}_1)^{-1}E(\boldsymbol{x}_1^\top y) - E(\boldsymbol{x}_1^\top\boldsymbol{x}_1)^{-1}E(\boldsymbol{x}_1^\top x_2)\beta_2) - x_2 x_2\beta_2 - \boldsymbol{Z}\boldsymbol{\phi} \mid \boldsymbol{z}, \boldsymbol{x}) = E(x_2(y - \boldsymbol{x}_1 E(\boldsymbol{x}_1^\top\boldsymbol{x}_1)^{-1}E(\boldsymbol{x}_1^\top y)) - x_2(x_2 - \boldsymbol{x}_1 E(\boldsymbol{x}_1^\top\boldsymbol{x}_1)^{-1}E(\boldsymbol{x}_1^\top x_2))\beta_2 - \boldsymbol{Z}\boldsymbol{\phi} \mid \boldsymbol{z}, \boldsymbol{x}) = E(\widetilde{y} - \widetilde{\boldsymbol{x}}\boldsymbol{\alpha} \mid \boldsymbol{z}, \boldsymbol{x}) = 0$. 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

$$\begin{aligned} \widetilde{y}_i + (\widehat{y}_i - \widehat{y}_i) &= (\widetilde{\boldsymbol{x}}_i + \widehat{\boldsymbol{x}}_i - \widehat{\boldsymbol{x}}_i)\boldsymbol{\alpha} + u_i, \\ \widehat{y}_i &= \widehat{\boldsymbol{x}}_i\alpha + u_i - (\widehat{\boldsymbol{x}}_i - \widetilde{\boldsymbol{x}}_i)\boldsymbol{\alpha} + (\widehat{y}_i - \widetilde{y}_i). \end{aligned} \tag{26}$$

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

$$\sqrt{n}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) = \widehat{Q}^{-1} n^{-1/2} \sum_{i=1}^{n} \widetilde{\boldsymbol{z}}_i^{\top} (u_i - (\widehat{\boldsymbol{x}}_i - \widetilde{\boldsymbol{x}}_i)\boldsymbol{\alpha} + (\widehat{y}_i - \widetilde{y}_i)), \tag{27}$$

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 (26) contains generated regressors and generated dependent variables. So we need to consider errors from these approximations in equation (27).

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

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

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,

$$\widehat{\boldsymbol{C}}_1 \equiv \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{1i}^\top \boldsymbol{x}_{1i} \xrightarrow{p} E(\boldsymbol{x}_{1i}^\top \boldsymbol{x}_{1i}) \equiv \boldsymbol{C}_1,$$

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

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.

B. Extension to Multiple Endogenous Variables

The proposed methodology can be easily extended to the case of multiple endogenous variables. Suppose that there are multiple endogenous causes, $\boldsymbol{x}_2$. We then have the following model for each individual

$$y = \boldsymbol{x}_1\boldsymbol{\beta}_1 + \boldsymbol{x}_2\boldsymbol{\beta}_2 + \epsilon, \tag{28}$$

where $\boldsymbol{x}_2 \equiv [x_{21}, x_{22}, ..., x_{2p_2}]$ is a p_2-dimensional vector of endogenous regressors and $\boldsymbol{\beta}_2 \equiv [\beta_{21}, \beta_{22}, \cdots, \beta_{2p_2}]^{\top}$ is a p_2-dimensional vector of corresponding coefficients. Suppose that the following equations hold.

$$\begin{bmatrix} x_{21}\epsilon \\ x_{22}\epsilon \\ \vdots \\ x_{2p_2}\epsilon \end{bmatrix} = \begin{bmatrix} \boldsymbol{Z}_1\boldsymbol{\phi}_1 + u_1 \\ \boldsymbol{Z}_2\boldsymbol{\phi}_2 + u_2 \\ \vdots \\ \boldsymbol{Z}_{p_2}\boldsymbol{\phi}_{p_2} + u_{p_2} \end{bmatrix}, \tag{29}$$

where $(\boldsymbol{Z}_1, \boldsymbol{Z}_2, \cdots, \boldsymbol{Z}_{p_2})$ are simultaneous variables for each equation and where $E(u_j \mid \boldsymbol{z}_j, \boldsymbol{x}_1, x_{2j}) = 0$ for $j = 1, 2, ..., p_2$. Let each $\boldsymbol{Z}_j$ be of dimension k_j and $\boldsymbol{\phi}_j$ of dimension k_j for $j = 1, 2, ..., p_2$, and $K = k_1 + k_2 + ... + k_{p_2}$. This give us a total of $p_1 + p_2 + K$ parameters to be estimated.

These parameters will be estimated by the same number of moment conditions. First, using exogeneity of $\boldsymbol{x}_1$, p_1 moment conditions

$$E(\boldsymbol{x}_1^\top \epsilon) = 0.$$

Then plugging equation (28) into (29) gives

$$\begin{bmatrix} x_{21}(y - \boldsymbol{x}_1\boldsymbol{\beta}_1 - \boldsymbol{x}_2\boldsymbol{\beta}_2) \\ x_{22}(y - \boldsymbol{x}_1\boldsymbol{\beta}_1 - \boldsymbol{x}_2\boldsymbol{\beta}_2) \\ \vdots \\ x_{2p_2}(y - \boldsymbol{x}_1\boldsymbol{\beta}_1 - \boldsymbol{x}_2\boldsymbol{\beta}_2) \end{bmatrix} = \begin{bmatrix} \boldsymbol{Z}_1\boldsymbol{\phi}_1 + u_1 \\ \boldsymbol{Z}_2\boldsymbol{\phi}_2 + u_2 \\ \vdots \\ \boldsymbol{Z}_{p_2}\boldsymbol{\phi}_{p_2} + u_{p_2} \end{bmatrix}.$$

Using $E(\boldsymbol{x}_1^\top \epsilon) = 0$, we have $\boldsymbol{\beta}_1 = E(\boldsymbol{x}_1^\top \boldsymbol{x}_1)^{-1}E(\boldsymbol{x}_1^\top y) - E(\boldsymbol{x}_1^\top \boldsymbol{x}_1)^{-1}E(\boldsymbol{x}_1^\top \boldsymbol{x}_2)\boldsymbol{\beta}_2$. Thus we obtain

$$\begin{bmatrix} x_{21}(y - \boldsymbol{x}_1(E(\boldsymbol{x}_1^\top \boldsymbol{x}_1)^{-1}E(\boldsymbol{x}_1^\top y) - E(\boldsymbol{x}_1^\top \boldsymbol{x}_1)^{-1}E(\boldsymbol{x}_1^\top \boldsymbol{x}_2)\boldsymbol{\beta}_2) - \boldsymbol{x}_2\boldsymbol{\beta}_2) \\ x_{22}(y - \boldsymbol{x}_1(E(\boldsymbol{x}_1^\top \boldsymbol{x}_1)^{-1}E(\boldsymbol{x}_1^\top y) - E(\boldsymbol{x}_1^\top \boldsymbol{x}_1)^{-1}E(\boldsymbol{x}_1^\top \boldsymbol{x}_2)\boldsymbol{\beta}_2) - \boldsymbol{x}_2\boldsymbol{\beta}_2) \\ \vdots \\ x_{2p_2}(y - \boldsymbol{x}_1(E(\boldsymbol{x}_1^\top \boldsymbol{x}_1)^{-1}E(\boldsymbol{x}_1^\top y) - E(\boldsymbol{x}_1^\top \boldsymbol{x}_1)^{-1}E(\boldsymbol{x}_1^\top \boldsymbol{x}_2)\boldsymbol{\beta}_2) - \boldsymbol{x}_2\boldsymbol{\beta}_2) \end{bmatrix} = \begin{bmatrix} \boldsymbol{Z}_1\boldsymbol{\phi}_1 + u_1 \\ \boldsymbol{Z}_2\boldsymbol{\phi}_2 + u_2 \\ \vdots \\ \boldsymbol{Z}_{p_2}\boldsymbol{\phi}_{p_2} + u_{p_2} \end{bmatrix},$$

or

$$\begin{bmatrix} x_{21}(y - \boldsymbol{x}_1 E(\boldsymbol{x}_1^\top \boldsymbol{x}_1)^{-1}E(\boldsymbol{x}_1^\top y)) \\ x_{22}(y - \boldsymbol{x}_1 E(\boldsymbol{x}_1^\top \boldsymbol{x}_1)^{-1}E(\boldsymbol{x}_1^\top y)) \\ \vdots \\ x_{2p_2}(y - \boldsymbol{x}_1 E(\boldsymbol{x}_1^\top \boldsymbol{x}_1)^{-1}E(\boldsymbol{x}_1^\top y)) \end{bmatrix} = \begin{bmatrix} x_{21}(\boldsymbol{x}_2 - \boldsymbol{x}_1 E(\boldsymbol{x}_1^\top \boldsymbol{x}_1)^{-1}E(\boldsymbol{x}_1^\top \boldsymbol{x}_2))\boldsymbol{\beta}_2 + \boldsymbol{Z}_1\boldsymbol{\phi}_1 + u_1 \\ x_{22}(\boldsymbol{x}_2 - \boldsymbol{x}_1 E(\boldsymbol{x}_1^\top \boldsymbol{x}_1)^{-1}E(\boldsymbol{x}_1^\top \boldsymbol{x}_2))\boldsymbol{\beta}_2 + \boldsymbol{Z}_2\boldsymbol{\phi}_2 + u_2 \\ \vdots \\ x_{2p_2}(\boldsymbol{x}_2 - \boldsymbol{x}_1 E(\boldsymbol{x}_1^\top \boldsymbol{x}_1)^{-1}E(\boldsymbol{x}_1^\top \boldsymbol{x}_2))\boldsymbol{\beta}_2 + \boldsymbol{Z}_{p_2}\boldsymbol{\phi}_{p_2} + u_{p_2} \end{bmatrix}.$$

By rearranging the equations we have

$$\begin{bmatrix} \widetilde{y}_1 \\ \widetilde{y}_2 \\ \vdots \\ \widetilde{y}_{p_2} \end{bmatrix} = \begin{bmatrix} \widetilde{\boldsymbol{x}}_1\boldsymbol{\alpha}_1 + u_1 \\ \widetilde{\boldsymbol{x}}_2\boldsymbol{\alpha}_2 + u_2 \\ \vdots \\ \widetilde{\boldsymbol{x}}_{p_2}\boldsymbol{\alpha}_{p_2} + u_{p_2} \end{bmatrix},$$

where $\widetilde{y}_j \equiv x_{2j}(y - \boldsymbol{x}_1 E(\boldsymbol{x}_1^\top \boldsymbol{x}_1)^{-1}E(\boldsymbol{x}_1^\top y))$ and $\widetilde{\boldsymbol{x}}_j \equiv [x_{2j}(\boldsymbol{x}_2 - \boldsymbol{x}_1 E(\boldsymbol{x}_1^\top \boldsymbol{x}_1)^{-1}E(\boldsymbol{x}_1^\top \boldsymbol{x}_2)), \boldsymbol{Z}_j]$, and $\boldsymbol{\alpha}_j \equiv [\boldsymbol{\beta}_2^\top, \boldsymbol{\phi}_j^\top]^\top$ for $j = 1, 2, ..., p_2$.

As a result, we have the following system of equations

$$\widetilde{\boldsymbol{Y}} = \widetilde{\boldsymbol{X}}\boldsymbol{\alpha} + \boldsymbol{u},$$

where

$$\widetilde{\boldsymbol{Y}} = \begin{bmatrix} \widetilde{y}_1 \\ \widetilde{y}_2 \\ \vdots \\ \widetilde{y}_{p_2} \end{bmatrix}, \widetilde{\boldsymbol{X}} = \begin{bmatrix} \widetilde{\boldsymbol{x}}_1 & 0 & \cdots & 0 \\ 0 & \widetilde{\boldsymbol{x}}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \widetilde{\boldsymbol{x}}_{p_2} \end{bmatrix}, \boldsymbol{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{p_2} \end{bmatrix}$$

with $\boldsymbol{\alpha} \equiv [\boldsymbol{\alpha}_1^\top, \boldsymbol{\alpha}_1^\top, ..., \boldsymbol{\alpha}_{p_2}^\top]^\top$.

Finally, consider the following $p_2 + K$ moment conditions

$$\begin{bmatrix} E(\widetilde{\boldsymbol{x}}_1^\top u_1) = \mathbf{0} \\ E(\widetilde{\boldsymbol{x}}_2^\top u_2) = \mathbf{0} \\ \vdots \\ E(\widetilde{\boldsymbol{x}}_{p_2}^\top u_{p_2}) = \mathbf{0} \end{bmatrix},$$

or equivalently

$$E(\widetilde{\boldsymbol{X}}^\top \boldsymbol{u}) = \mathbf{0}.$$

As seen in the case for a single endogenous regressor, conditions similar to Assumptions 1 and 2 for each endogeneous cause are required to identify the parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\alpha}$.

References

Abel, A. B., and J. C. Eberly (1994): "A Unified Model of Investment Under Uncertainty," *American Economic Review*, 84, 1369–1384.

Almeida, H., and M. Campello (2007): "Financial Constraints, Asset Tangibility and Corporate Investment," *Review of Financial Studies*, 20, 1429–1460.

Almeida, H., M. Campello, and A. Galvao (2010): "Measurement Errors in Investment Equations," *Review of Financial Studies*, 23, 3279–3328.

Bertrand, M., and S. Mullainathan (2005): "Bidding for Oil and Gas Leases in the Gulf of Mexico: A Test of the Free Cash Flow Model?," mimeo, The University of Chicago and MIT.

Bertrand, M., and A. Schoar (2003): "Managing with Style: The Effect of Managers on Firm Policies," *Quarterly Journal of Economics*, 118, 1169–1208.

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.

Caetano, C. (2015): "A Test of Exogeneity without Instrumental Variables," mimeo, University of Rochester.

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.

Chen, X. (2007): "Large Sample Sieve Estimation of Semi-nonparametric Models," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. E. Leamer, vol. 6B. North-Holland.

Chirinko, R. (1993): "Business Fixed Investment Spending: Modeling Strategies, Empirical Results, and Policy Implications," *Journal of Economic Literature*, 31, 1875–1911.

Conley, T. G., C. B. Hansen, and P. E. Rossi (2012): "Plausibly Exogenous," *Review of Economics and Statistics*, 94, 260–272.

Erickson, T., and T. Whited (2000): "Measurement Error and the Relationship between Investment and Q," *Journal of Political Economy*, 108, 1027–1057.

Fazzari, S., R. G. Hubbard, and B. Petersen (1988): "Financing Constraints and Corporate Investment," *Brooking Papers on Economic Activity*, 1, 141–195.

Gandhi, A., K. Kim, and A. Petrin (2013): "Identification and Estimation in Discrete Choice Demand Models when Endogenous Variables Interact with the Error," University of Minnesota, mimeo.

Gomes, J. (2001): "Financing Investment," *American Economic Review*, 91, 1263–1285.

Hadlock, C. (1998): "Ownership, Liquidity, and Investment," *RAND Journal of Economics*, 29, 487–508.

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

Hayashi, F. (1982): "Tobin's Marginal q and Average q: A Neoclassical Interpretation," *Econometrica*, 50, 213–224.

Hoshi, T., A. Kashyap, and D. Scharfstein (1991): "Corporate Structure, Liquidity, and Investment: Evidence from Japanese Industrial Groups," *Quarterly Journal of Economics*, 106, 33–60.

Lamont, O. (1999): "Cash Flow and Investment: Evidence from Internal Capital Markets," *The Journal of Finance*, 52, 83–110.

Lewellen, J., and K. Lewellen (2014): "Investment and Cashflow: New Evidence," *Journal of Financial and Quantitative Analysis*, Forthcoming.

Malmendier, U., and G. Tate (2005): "CEO Overconfidence and Corporate Investment," *The Journal of Finance*, 60, 2661–2700.

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

Newey, W. K. (1990): "Efficient Instrumental Variables Estimation of Nonlinear Models," *Econometrica*, 58, 809–837.

Newey, W. K., J. L. Powell, and F. Vella (1999): "Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica*, 67, 565–603.

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

Poterba, J. (1988): "Financing Constraints and Corporate Investment: Comment," *Brookings Papers on Economic Activity*, 1, 200–204.

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

Shin, H., and R. Stulz (1998): "Are Internal Capital Markets Efficient?," *Quarterly Journal of Economics*, 113, 531–552.

Stein, J. (2003): "Agency Information and Corporate Investment," in *Handbook of the Economics of Finance*, ed. by M. Harris, and R. Stulz. Elsevier/North-Holland, Amsterdam.

Tobin, J. (1969): "A General Equilibrium Approach To Monetary Theory," *Journal of Money, Credit & Banking*, 1, 15–29.

Table 1: DGP with no endogeneity. Monte Carlo results: Bias, Median Absolute Deviation, and Root Mean Squared Error

	Estimator					
Model	OLS	IV	SV	OLS	IV	SV
	$n = 100$			$n = 200$		
	Bias					
z_{sv}	-0.001	3.641	0.002	0.001	-1.104	0.002
z_{iv}	-0.001	0.002	-0.003	0.001	0.000	0.000
	MAD					
z_{sv}	0.056	1.056	0.070	0.041	1.091	0.050
z_{iv}	0.055	0.166	0.085	0.040	0.115	0.060
	RMSE					
z_{sv}	0.083	253.644	0.106	0.059	119.844	0.075
z_{iv}	0.082	0.416	0.126	0.059	0.186	0.090
	$n = 500$			$n = 1000$		
	Bias					
z_{sv}	0.000	9.609	0.000	0.000	-2.191	0.001
z_{iv}	0.000	0.000	0.000	0.000	0.002	0.000
	MAD					
z_{sv}	0.025	1.041	0.032	0.018	1.048	0.022
z_{iv}	0.024	0.075	0.040	0.018	0.052	0.027
	RMSE					
z_{sv}	0.037	797.869	0.048	0.026	101.951	0.034
z_{iv}	0.036	0.115	0.058	0.026	0.080	0.042

Notes: Monte Carlo experiments are based on 5,000 repetitions. Estimates of β_2. z_{sv} and z_{iv} are, respectively, the used simultaneous variables and instrumental variables employed in the estimations.

Table 2: DGP with endogeneity. Monte Carlo results: Bias, Median Absolute Deviation, and Root Mean Squared Error

	Estimator					
Model	OLS	IV	SV	OLS	IV	SV
	$n = 100$			$n = 200$		
	Bias					
z_{sv1}	0.667	0.851	-0.138	0.668	0.644	-0.097
z_{sv2}	0.666	-21.8	-0.019	0.667	0.755	-0.017
z_{iv}	0.668	-0.078	0.808	0.667	-0.037	0.805
	MAD					
z_{sv1}	0.668	0.943	0.205	0.669	0.947	0.143
z_{sv2}	0.664	1.202	0.045	0.666	1.199	0.027
z_{iv}	0.665	0.236	0.806	0.668	0.170	0.804
	RMSE					
z_{sv1}	0.674	18.1	0.307	0.671	63.462	0.225
z_{sv2}	0.673	1557.8	0.085	0.670	25.6	0.057
z_{iv}	0.675	2.737	0.825	0.670	0.281	0.814
	$n = 500$			$n = 1000$		
	Bias					
z_{sv1}	0.666	-0.161	-0.056	0.667	0.570	-0.031
z_{sv2}	0.668	0.510	-0.009	0.666	1.140	-0.005
z_{iv}	0.667	-0.015	0.801	0.667	-0.003	0.802
	MAD					
z_{sv1}	0.666	0.895	0.089	0.667	0.921	0.060
z_{sv2}	0.668	1.179	0.014	0.667	1.185	0.009
z_{iv}	0.668	0.168	0.841	0.667	0.073	0.802
	RMSE					
z_{sv1}	0.668	58.2	0.141	0.667	20.719	0.094
z_{sv2}	0.669	51.1	0.030	0.667	24.214	0.017
z_{iv}	0.669	0.168	0.805	0.667	0.110	0.804

Notes: Monte Carlo experiments are based on 5000 repetitions. Estimates of β_2. z_{sv1}, z_{sv2}, and z_{iv} are, respectively, the used simultaneous variables 1 and 2, and instrumental variables employed in the estimations.

Table 3: Investment example. Descriptive statistics

Variable	Mean	Std. Dev.	Min	Max	Obs
Investment	0.2026	0.1348	0.0000	2.5349	24676
q	0.8755	0.4900	0.1307	22.7939	24676
Cash flow	0.3193	0.6212	-80.3571	14.8021	24676

Table 4: Investment example. OLS and Instrumental Variables approaches (using q_{t-1}, q^2_{t-1} q_{t-2}, q^2_{t-2}, q_{t-3}, and q^2_{t-3} as IV)

	OLS	IV1	IV2	IV3	IV4	IV5	IV6	IV7	IV8
q	0.0626***	0.0478***	0.0496***	0.0325***	0.0373***	0.0241***	0.0288***	0.0470***	0.0488***
	(0.003)	(0.004)	(0.004)	(0.005)	(0.005)	(0.006)	(0.006)	(0.004)	(0.004)
CF	0.1307***	0.1337***	0.1333***	0.1368***	0.1358***	0.1385***	0.1375***	0.1338***	0.1335*
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Const.	0.1053***	0.1165***	0.1151***	0.1282***	0.1245***	0.1345***	0.1310***	0.1171***	0.1157
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.005)	(0.005)	(0.003)	(0.003)
q_{t-1}		0.8348***	1.0091***					0.8133***	1.0149***
		(0.004)	(0.012)					(0.007)	(0.016)
q^2_{t-1}			-0.0699***						-0.0818***
			(0.004)						(0.005)
q_{t-2}				0.5393***	0.8266***			-0.0140*	-0.0461***
				(0.004)	(0.009)			(0.008)	(0.013)
q^2_{t-2}					-0.0806***				0.0070**
					(0.002)				(0.003)
q_{t-3}						0.2829***	0.4717***	0.0304***	0.0526***
						(0.003)	(0.006)	(0.004)	(0.006)
q^2_{t-3}							-0.0308***		-0.0023***
							(0.001)		(0.001)

Notes: Robust standard errors in parentheses. Coefficients and standard errors for the IV are from the first stage regression. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Investment example. Simultaneous variable approach (using q_{t-1}, q_{t-1}^2, q_{t-2}, q_{t-2}^2, q_{t-3}, and q_{t-3}^2 as simultaneous variables)

	S1	S2	S3	S4	S5	S6	S7	S8
q	0.0938***	0.1107***	0.0896***	0.0898***	0.0804***	0.0802***	0.0969***	0.1129***
	(0.026)	(0.028)	(0.022)	(0.022)	(0.019)	(0.019)	(0.026)	(0.028)
CF	0.1243***	0.1208***	0.1252***	0.1251***	0.1270***	0.1271***	0.1237***	0.1204***
	(0.005)	(0.006)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.006)
Const.	0.0816***	0.0687***	0.0848***	0.0846***	0.0918***	0.0920***	0.0792***	0.0671***
	(0.020)	(0.021)	(0.016)	(0.016)	(0.014)	(0.015)	(0.020)	(0.021)
q_{t-1}	-0.0684**	0.0766***					-0.0350	0.0915***
	(0.029)	(0.022)					(0.028)	(0.030)
q_{t-1}^2		-0.0664***						-0.0606***
		(0.017)						(0.017)
q_{t-2}			-0.0546***	-0.0248			-0.0260*	-0.0142
			(0.017)	(0.019)			(0.015)	(0.026)
q_{t-2}^2				-0.0085				-0.0042
				(0.005)				(0.008)
q_{t-3}					-0.0282***	-0.0253**	-0.0093	-0.0071
					(0.009)	(0.011)	(0.009)	0.014)
q_{t-3}^2						-0.0005		0.0007
						(0.001)		(0.003)

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.