Moments of IV and JIVE Estimators

by

Russell Davidson

GREQAM
Centre de la Vieille Charité
2 rue de la Charité
13236 Marseille cedex 02, France

Department of Economics
McGill University
Montreal, Quebec, Canada
H3A 2T7

Russell.Davidson@mcgill.ca

and

James G. MacKinnon

Department of Economics
Queen's University
Kingston, Ontario, Canada
K7L 3N6

jgm@econ.queensu.ca

Abstract

We develop a new method, based on the use of polar coordinates, to investigate the existence of moments for instrumental variables and related estimators in the linear regression model. For generalized IV estimators, we obtain familiar results. For JIVE, we obtain the new result that this estimator has no moments at all. Simulation results illustrate the consequences of its lack of moments.

JEL codes: C100, C120, C300

This research was supported, in part, by grants from the Social Sciences and Humanities Research Council of Canada.

June, 2006.

1. Introduction

It is well-known that the LIML estimator of a single equation from a linear simultaneous equations model has no moments, and that the generalized IV (2SLS) estimator has as many moments as there are overidentifying restrictions. For a recent survey, see Mariano (2001); key papers include Fuller (1977) and Kinal (1980). In this paper, we propose a new method, based on the use of polar coordinates, to study the existence, or non-existence, of moments of IV estimators. This approach is considerably easier than existing ones. The principal new result of the paper is that the JIVE estimator proposed by Angrist, Imbens, and Krueger (1999) and Blomquist and Dahlberg (1999) has no moments. This is a result that those who have studied the finite-sample properties of JIVE by simulation, including Hahn, Hausman, and Keuersteiner (2004) and Davidson and MacKinnon (2006), have suspected for some time.

In the next section, we discuss a simple model with just one endogenous variable on the right-hand side and develop some simple expressions for the IV estimate of the coefficient of that variable. We also show how the JIVE estimator can be expressed in a way that is quite similar to a generalized IV estimator. Then, in Section 3, we rederive standard results about the existence of moments for IV and K–class estimators in a novel way. In Section 4, we show that the JIVE estimator has no moments. In Section 5, we show that the results for the simple model extend to a more general model in which there are exogenous variables in the structural equation. In Section 6, we present some simulation results which illustrate some of the consequences of nonexistence of moments for these estimators. Section 7 concludes.

2. A Simple Model

The simplest model that we consider has a single endogenous variable on the right-hand side and no exogenous variables. This model is written as

$$\begin{aligned} y_t &= \beta x_t + \sigma_u u_t, \\ x_t &= \sigma_v(a w_{t1} + v_t), \end{aligned} \tag{1}$$

where we have used some unconventional normalizations that will later be convenient. The structural disturbances u_t and the reduced form disturbances v_t are assumed to be serially independent and bivariate normal:

$$\begin{bmatrix} u_t \\ v_t \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right).$$

The n–vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ have typical elements u_t and v_t, respectively. The n–vector $\boldsymbol{w}_1$, with typical element w_{t1}, is to be interpreted as an instrumental variable. As such, it is taken to be exogenous. The disturbance u_t in the structural equation that defines y_t can be expressed as $u_t = \rho v_t + u_{t1}$, where v_t and u_{t1} are independent, with $u_{t1} \sim \mathrm{N}(0, 1 - \rho^2)$.

The simple instrumental variables, or IV, estimator of the parameter β in model (1) is

$$\hat{\beta}_0 = \frac{\boldsymbol{w}_1{}^\top \boldsymbol{y}}{\boldsymbol{w}_1{}^\top \boldsymbol{x}}, \tag{2}$$

where the n–vectors $\boldsymbol{y}$ and $\boldsymbol{x}$ have typical elements y_t and x_t, respectively. It is well known that this estimator has no moments, since there are no overidentifying restrictions, as indicated by the "0" subscript in $\hat{\beta}_0$.

When there are overidentifying restrictions, $\boldsymbol{W}$ denotes an $n \times l$ matrix of exogenous instruments, with $\boldsymbol{w}_i$ denoting its i^{th} column. The generalized IV, or 2SLS, estimator, that makes use of these instruments is

$$\hat{\beta}_{l-1} = \frac{\boldsymbol{x}^\top \boldsymbol{P}_{\boldsymbol{W}} \boldsymbol{y}}{\boldsymbol{x}^\top \boldsymbol{P}_{\boldsymbol{W}} \boldsymbol{x}}, \tag{3}$$

where $\boldsymbol{P}_{\boldsymbol{W}} \equiv \boldsymbol{W}(\boldsymbol{W}^\top \boldsymbol{W})^{-1}\boldsymbol{W}^\top$ is the orthogonal projection on to the span of the columns of $\boldsymbol{W}$. The estimator is indexed by $l-1$, the degree of overidentification, which is also the number of moments that it possesses. Note that (1) may be taken to be the DGP (data-generating process) even when there are overidentifying restrictions. This involves no loss of generality, because we can think of the instrument vector $\boldsymbol{w}_1$ as a particular linear combination of the columns of the matrix $\boldsymbol{W}$.

It is clear from (3) that the generalized IV, or 2SLS, estimator depends on $\boldsymbol{W}$ only through the linear span of its columns. There is therefore no loss of generality if we assume that $\boldsymbol{W}^\top \boldsymbol{W} = \mathbf{I}$, the $l \times l$ identity matrix. In that case,

$$\boldsymbol{x}^\top \boldsymbol{P}_{\boldsymbol{W}} \boldsymbol{x} = \sum_{i=1}^{l} (\boldsymbol{w}_i^\top \boldsymbol{x})^2 = \sigma_v^2 \Big((a + \boldsymbol{w}_1{}^\top \boldsymbol{v})^2 + \sum_{i=2}^{l} (\boldsymbol{w}_i^\top \boldsymbol{v})^2 \Big), \tag{4}$$

where the last expression follows from (1). If the true value of the parameter β is β^0, then we can also see that

$$\begin{aligned} \boldsymbol{x}^\top \boldsymbol{P}_{\boldsymbol{W}}(\boldsymbol{y} - \boldsymbol{x}\beta^0) &= \sigma_u \sigma_v \Big((a + \boldsymbol{w}_1{}^\top \boldsymbol{v}) \boldsymbol{w}_1{}^\top \boldsymbol{u} + \sum_{i=2}^{l} (\boldsymbol{w}_i^\top \boldsymbol{v})(\boldsymbol{w}_i^\top \boldsymbol{u}) \Big) \\ &= \rho \sigma_u \sigma_v \Big((a + \boldsymbol{w}_1{}^\top \boldsymbol{v}) \boldsymbol{w}_1{}^\top \boldsymbol{v} + \sum_{i=2}^{l} (\boldsymbol{w}_i^\top \boldsymbol{v})^2 \Big) \\ &\quad + \sigma_u \sigma_v \Big((a + \boldsymbol{w}_1{}^\top \boldsymbol{v}) \boldsymbol{w}_1{}^\top \boldsymbol{u}_1 + \sum_{i=2}^{l} (\boldsymbol{w}_i^\top \boldsymbol{v})(\boldsymbol{w}_i^\top \boldsymbol{u}_1) \Big), \end{aligned} \tag{5}$$

where $\boldsymbol{u}_1$ is the n–vector with typical element u_{t1}.

It follows from (4) and (5) that the expectation of $\hat\beta_{l-1}-\beta^0$, conditional on the vector of reduced form disturbances $\boldsymbol{v}$, is

$$\mathrm{E}(\hat\beta_{l-1}-\beta^0\,|\,\boldsymbol{v}) = \frac{\rho\sigma_u}{\sigma_v}\left(\frac{(a+\boldsymbol{w}_1^\top\boldsymbol{v})\boldsymbol{w}_1^\top\boldsymbol{v}+\sum_{i=2}^{l}(\boldsymbol{w}_i^\top\boldsymbol{v})^2}{(a+\boldsymbol{w}_1^\top\boldsymbol{v})^2+\sum_{i=2}^{l}(\boldsymbol{w}_i^\top\boldsymbol{v})^2}\right). \tag{6}$$

The right-hand side of (6) can be rewritten as $\rho\sigma_u/\sigma_v$ times

$$1-\frac{a(a+\boldsymbol{w}_1^\top\boldsymbol{v})}{(a+\boldsymbol{w}_1^\top\boldsymbol{v})^2+\sum_{i=2}^{l}(\boldsymbol{w}_i^\top\boldsymbol{v})^2}. \tag{7}$$

The estimator $\hat\beta_{l-1}$ has as many moments as the second term above. Observe that the random l–vector with typical component $\boldsymbol{w}_i^\top\boldsymbol{v}$ is distributed as $\mathrm{N}(\boldsymbol{0},\mathbf{I})$, so that the denominator of this second term is distributed as noncentral chi-squared with l degrees of freedom and noncentrality parameter a^2. Notice that this denominator vanishes when $\boldsymbol{w}_1^\top\boldsymbol{v}=-a$ and $\boldsymbol{w}_i^\top\boldsymbol{v}=0$, $i=2,\ldots,l$. This point will be crucial when we discuss the existence of the expectation in (6).

We can replace the matrix $\boldsymbol{P_W}$ in (3) by other matrices $\boldsymbol{A}^\top$ with the property that $\boldsymbol{AW}=\boldsymbol{W}$, and this leads to other estimators of interest. For instance, if we make the choice $\boldsymbol{A}=\mathbf{I}-K\boldsymbol{M_W}$, where $\boldsymbol{M_W}=\mathbf{I}-\boldsymbol{P_W}$ is the orthogonal projection complementary to $\boldsymbol{P_W}$, we obtain a K–class estimator. With $K=1$, of course, we recover the generalized IV estimator (3).

In general, under the assumption that $\boldsymbol{AW}=\boldsymbol{W}$, we see that

$$\boldsymbol{x}^\top\boldsymbol{A}^\top(\boldsymbol{y}-\boldsymbol{x}\beta^0)=\sigma_u\sigma_v(a\rho\boldsymbol{w}_1^\top\boldsymbol{A}^\top\boldsymbol{v}+\rho\boldsymbol{v}^\top\boldsymbol{A}^\top\boldsymbol{v}+a\boldsymbol{w}_1^\top\boldsymbol{A}^\top\boldsymbol{u}_1+\boldsymbol{v}^\top\boldsymbol{A}^\top\boldsymbol{u}_1), \tag{8}$$

of which the expectation conditional on $\boldsymbol{v}$ is

$$\mathrm{E}\big(\boldsymbol{x}^\top\boldsymbol{A}^\top(\boldsymbol{y}-\boldsymbol{x}\beta^0)\,|\,\boldsymbol{v}\big)=\rho\sigma_u\sigma_v(a\boldsymbol{w}_1^\top\boldsymbol{v}+\boldsymbol{v}^\top\boldsymbol{A}\boldsymbol{v}). \tag{9}$$

Similarly,

$$\boldsymbol{x}^\top\boldsymbol{A}^\top\boldsymbol{x}=\sigma_v^2(a^2+a\boldsymbol{v}^\top\boldsymbol{A}^\top\boldsymbol{w}_1+a\boldsymbol{w}_1^\top\boldsymbol{v}+\boldsymbol{v}^\top\boldsymbol{A}\boldsymbol{v}). \tag{10}$$

The factor of $\boldsymbol{A}^\top$ is retained in the second term on the right-hand side because we do not necessarily require that $\boldsymbol{A}$ should be symmetric, so that $\boldsymbol{A}\boldsymbol{w}_1=\boldsymbol{w}_1$ does not necessarily imply that $\boldsymbol{A}^\top\boldsymbol{w}_1=\boldsymbol{w}_1$. It follows from (9) and (10) that

$$\mathrm{E}(\hat\beta_{l-1}-\beta^0\,|\,\boldsymbol{v})=\frac{\rho\sigma_u}{\sigma_v}\left(\frac{a\boldsymbol{w}_1^\top\boldsymbol{v}+\boldsymbol{v}^\top\boldsymbol{A}\boldsymbol{v}}{a^2+a\boldsymbol{v}^\top\boldsymbol{A}^\top\boldsymbol{w}_1+a\boldsymbol{w}_1^\top\boldsymbol{v}+\boldsymbol{v}^\top\boldsymbol{A}\boldsymbol{v}}\right). \tag{11}$$

Both the numerator and the denominator of this conditional expectation vanish if $\boldsymbol{v}=-a\boldsymbol{w}_1$. This condition also implies that the numerator and the denominator of the second term in (7) vanish.

Now we make the change of variables

$$\boldsymbol{v} = \boldsymbol{\xi} - a\boldsymbol{w}_1, \tag{12}$$

whereby the n–vector $\boldsymbol{\xi}$ is distributed as $\mathrm{N}(a\boldsymbol{w}_1, \mathbf{I})$. With this change of variables, the numerator of the factor in parentheses in expression (11) becomes

$$a\boldsymbol{w}_1{}^\top\boldsymbol{\xi} - a^2 + \boldsymbol{\xi}^\top\boldsymbol{A}\boldsymbol{\xi} - a\boldsymbol{\xi}^\top\boldsymbol{w}_1 - a\boldsymbol{w}_1{}^\top\boldsymbol{A}\boldsymbol{\xi} + a^2 = \boldsymbol{\xi}^\top\boldsymbol{A}\boldsymbol{\xi} - a\boldsymbol{w}_1{}^\top\boldsymbol{A}\boldsymbol{\xi},$$

and the denominator becomes

$$a^2 + a\boldsymbol{\xi}^\top\boldsymbol{A}^\top\boldsymbol{w}_1 - a^2 + a\boldsymbol{w}_1{}^\top\boldsymbol{\xi} - a^2 + \boldsymbol{\xi}^\top\boldsymbol{A}\boldsymbol{\xi} - a\boldsymbol{w}_1{}^\top\boldsymbol{A}\boldsymbol{\xi} - a\boldsymbol{\xi}^\top\boldsymbol{w}_1 + a^2 = \boldsymbol{\xi}^\top\boldsymbol{A}\boldsymbol{\xi}.$$

The factor in parentheses in expression (11) therefore simplifies to

$$\frac{\boldsymbol{\xi}^\top\boldsymbol{A}\boldsymbol{\xi} - a\boldsymbol{w}_1{}^\top\boldsymbol{A}\boldsymbol{\xi}}{\boldsymbol{\xi}^\top\boldsymbol{A}\boldsymbol{\xi}} = 1 - a\frac{\boldsymbol{w}_1{}^\top\boldsymbol{A}\boldsymbol{\xi}}{\boldsymbol{\xi}^\top\boldsymbol{A}\boldsymbol{\xi}}, \tag{13}$$

of which expression (7) is a special case.

Before investigating the existence or nonexistence of moments of expression (13), we consider the particular case that was rather misleadingly called "jackknife instrumental variables," or JIVE, by Angrist, Imbens, and Krueger (1999) and Blomquist and Dahlberg (1999).[1] This estimator makes use, for its single instrumental variable, of what we may call the vector of omit-one fitted values from the first-stage regression of the endogenous explanatory variable $\boldsymbol{x}$ on the instruments $\boldsymbol{W}$:

$$\boldsymbol{x} = \boldsymbol{W}\boldsymbol{\gamma} + \sigma_v\boldsymbol{v}. \tag{14}$$

The omit-one fitted value for observation t, $\tilde{x}_t$, is defined as $\boldsymbol{W}_t\hat{\boldsymbol{\gamma}}^{(t)}$, where $\boldsymbol{W}_t$ is row t of $\boldsymbol{W}$, and the estimates $\hat{\boldsymbol{\gamma}}^{(t)}$ are obtained by running regression (14) without observation t.

The vector of omit-one estimates $\hat{\boldsymbol{\gamma}}^{(t)}$ is related to the full-sample vector of estimates $\hat{\boldsymbol{\gamma}}$ by the relation

$$\hat{\boldsymbol{\gamma}}^{(t)} = \hat{\boldsymbol{\gamma}} - \frac{1}{1-h_t}(\boldsymbol{W}^\top\boldsymbol{W})^{-1}\boldsymbol{W}_t^\top\hat{u}_t,$$

where the $\hat{u}_t$ are the residuals from regression (14) run on the full sample, and h_t is the t^{th} diagonal element of $\boldsymbol{P}_{\boldsymbol{W}}$; see, for instance, equation (2.63) in Davidson and

[1] Actually, Angrist, Imbens, and Krueger (1999) called the estimator we will study JIVE1, and Blomquist and Dahlberg (1999) called it UJIVE. But it is most commonly just called JIVE.

MacKinnon (2004). Under our assumption that $\boldsymbol{W}^\top \boldsymbol{W} = \mathbf{I}$, we may write $h_t = \|\boldsymbol{W}_t\|^2$. Further, $\boldsymbol{W}_t \hat{\boldsymbol{\gamma}} = (\boldsymbol{P_W x})_t$. It follows then that

$$\tilde{x}_t = \boldsymbol{W}_t \hat{\boldsymbol{\gamma}}^{(t)} = (\boldsymbol{P_W x})_t - \frac{1}{1 - \|\boldsymbol{W}_t\|^2} (\boldsymbol{P_W})_{tt} (\boldsymbol{M_W x})_t = x_t - \frac{(\boldsymbol{M_W x})_t}{1 - \|\boldsymbol{W}_t\|^2}, \tag{15}$$

since the t^{th} diagonal element $(\boldsymbol{P_W})_{tt}$ of $\boldsymbol{P_W}$ is $h_t = \|\boldsymbol{W}_t\|^2$. We assume throughout that $0 < h_t < 1$ with strict inequality, thereby avoiding the potential problem of a zero denominator in (15). This assumption is not at all restrictive, since, if $h_t = 0$, the t^{th} elements of all the instruments vanish, whereas, if $h_t = 1$, the span of the instruments contains the dummy variable for observation t.

For the JIVE estimator, we wish to define the matrix $\boldsymbol{A}$ in such a way that the vector $\tilde{\boldsymbol{x}}$ of omit-one fitted values is equal to $\boldsymbol{Ax}$. By letting the (t, s) element of $\boldsymbol{A}$ be

$$a_{ts} = \frac{1}{1 - \|\boldsymbol{W}_t\|^2} \big((\boldsymbol{P_W})_{ts} - \delta_{ts} \|\boldsymbol{W}_t\|^2 \big), \tag{16}$$

where δ_{ts} is the Kronecker delta, we may check that

$$\sum_{s=1}^{n} a_{ts} x_s = \frac{1}{1 - \|\boldsymbol{W}_t\|^2} \big((\boldsymbol{P_W x})_t - x_t \|\boldsymbol{W}_t\|^2 \big) = x_t - \frac{(\boldsymbol{M_W x})_t}{1 - \|\boldsymbol{W}_t\|^2} = \tilde{x}_t,$$

and that, for $i = 1, \ldots, l$,

$$\sum_{s=1}^{n} a_{ts} w_{si} = \frac{1}{1 - \|\boldsymbol{W}_t\|^2} \big(w_{ti} - w_{ti} \|\boldsymbol{W}_t\|^2 \big) = w_{ti},$$

so that $\boldsymbol{AW} = \boldsymbol{W}$, as required.

With $\tilde{\boldsymbol{x}}$ defined as above, the JIVE estimator is just

$$\hat{\beta}_J = \frac{\tilde{\boldsymbol{x}}^\top \boldsymbol{y}}{\tilde{\boldsymbol{x}}^\top \tilde{\boldsymbol{x}}}, \tag{17}$$

which is similar to the IV estimator (2) and the generalized IV estimator (3).

3. Existence of Moments for IV Estimators

In this section, we study the problem of how many moments exist for the estimators discussed in the previous section under DGPs belonging to the model (1). We begin with the generalized IV estimator $\hat{\beta}_{l-1}$ given in (3), which, although it has properties not shared with many other estimators, allows us to appreciate the relevant issues and the methodology for dealing with them.

We saw that, when there are l instruments, the estimator $\hat{\beta}_{l-1}$ has as many moments as the expression given in equation (7). With the change of variables (12), this expression is just the second term on the right-hand side of equation (13). Since $\boldsymbol{A} = \boldsymbol{P_W}$ here, this second term is, except for the factor of a, equal to

$$\frac{\boldsymbol{w}_1^\top \boldsymbol{\xi}}{\boldsymbol{\xi}^\top \boldsymbol{P_W} \boldsymbol{\xi}} = \frac{z_1}{\sum_{l=1}^{l} z_i^2}, \tag{18}$$

where $z_i \equiv \boldsymbol{w}_i^\top \boldsymbol{\xi}$. The z_i are mutually independent, and they are distributed as standard normal except for z_1, which is distributed as $\mathrm{N}(0, a)$.

We now once more change variables, so as to use the polar coordinates that correspond to the cartesian coordinates z_i, $i = 1, \ldots, l$. The first polar coordinate is denoted r, and it is the positive square root of $\sum_{i=1}^{l} z_i^2$. The other polar coordinates, denoted $\theta_1, \ldots, \theta_{l-1}$, are all angles. They are defined as follows:

$$\begin{aligned} z_1 &= r\cos\theta_1, \\ z_i &= r\cos\theta_i \prod_{j=1}^{i-1} \sin\theta_j, \quad i = 2, \ldots, l-1, \text{ and} \\ z_l &= r \prod_{j=1}^{l-1} \sin\theta_j, \end{aligned}$$

where $0 \le \theta_i < \pi$ for $i = 1, \ldots, l-2$, and $0 \le \theta_{l-1} < 2\pi$. The use of polar coordinates in more than two dimensions is relatively uncommon. Anderson (2003, p. 285) uses somewhat different conventions to obtain an equivalent result. In terms of these polar coordinates, expression (18) is just $\cos\theta_1/r$. It is also easy to show that, for $1 \le j < l$,

$$\sum_{i=j+1}^{l} z_i^2 = r^2 \prod_{i=1}^{j} \sin^2\theta_i. \tag{19}$$

The joint density of the z_i is

$$\frac{1}{(2\pi)^{l/2}} \exp\Bigl(-\tfrac{1}{2}\Bigl((z_1 - a)^2 + \sum_{i=2}^{l} z_i^2\Bigr)\Bigr) = \frac{e^{-a^2/2}}{(2\pi)^{l/2}} \exp\Bigl(-\tfrac{1}{2}(r^2 - 2ar\cos\theta_1)\Bigr).$$

The Jacobian of the transformation to polar coordinates can be shown to be

$$\frac{\partial(z_1,\ldots,z_l)}{\partial(r,\theta_1,\ldots,\theta_{l-1})} = r^{l-1}\sin^{l-2}\theta_1 \sin^{l-3}\theta_2 \ldots \sin\theta_{l-2}\,;$$

see Anderson (2003, p. 286). Consequently, the m^{th}moment of (18), if it exists, is given by the integral

$$\begin{aligned}\frac{e^{-a^2/2}}{(2\pi)^{(l-2)/2}} \int_0^\pi \sin^{l-2}\theta_1\, d\theta_1 \ldots \int_0^\pi \sin\theta_{l-2}\, d\theta_{l-2} \\ \times \cos^m\theta_1 \int_0^\infty r^{l-m-1}\, dr \exp\Bigl(-\tfrac{1}{2}(r^2 - 2ar\cos\theta_1)\Bigr).\end{aligned} \tag{20}$$

Here, the integral with respect to θ_{l-1} has been performed explicitly: Since neither the density nor the Jacobian depends on this angle, the integral with respect to it is just 2π.

Expression (20) can certainly be simplified, but that is not necessary for our conclusion regarding the existence of moments. The integral over r converges if and only if the exponent $l-m-1$ is greater than -1. If not, then it diverges at $r=0$. The angle integrals are all finite, and the joint density is everywhere positive, and so the only possible source of divergence is the singularity with respect to r at $r=0$. Thus moments exist only for $m<l$. This merely confirms existing results, which were first demonstrated in Kinal (1980) by a more complicated argument.

The methodology employing polar coordinates can be used to investigate the properties of other estimators. As an example, we cite another class of estimators examined by Kinal, namely, K–class estimators with exogenous $K<1$. For these estimators, the matrix $\boldsymbol{A}$ is $\mathbf{I} - K\boldsymbol{M_W}$.

We begin by constructing the $n\times n$ orthogonal matrix $\boldsymbol{U}$ with its first l columns identical to those of $\boldsymbol{W}$, the remaining $n-l$ columns constituting an orthonormal basis for the orthogonal complement of the span of the columns of $\boldsymbol{W}$. If we denote those remaining columns by $\boldsymbol{u}_j$, $j=l+1,\ldots,n$, we can define $z_j \equiv \boldsymbol{u}_j^\top\boldsymbol{\xi}$. The z_j are all standard normal and independent of the z_i, for $i=1,\ldots,z_l$. Then we have

$$\boldsymbol{\xi}^\top\boldsymbol{A}\boldsymbol{\xi} = \boldsymbol{\xi}^\top\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{A}\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{\xi}. \tag{21}$$

Observe that

$$\begin{aligned}\boldsymbol{A}\boldsymbol{U} &= (\mathbf{I} - K\boldsymbol{M_W})[\boldsymbol{w}_1\ldots\boldsymbol{w}_l \;\; \boldsymbol{u}_{l+1}\ldots\boldsymbol{u}_n] \\ &= [\boldsymbol{w}_1\ldots\boldsymbol{w}_l \;\; (1-K)\boldsymbol{u}_{l+1}\ldots(1-K)\boldsymbol{u}_n],\end{aligned}$$

and so

$$\boldsymbol{U}^\top\boldsymbol{A}\boldsymbol{U} = \begin{bmatrix}\mathbf{I}_l & \mathbf{O} \\ \mathbf{O} & (1-K)\mathbf{I}_{n-l}\end{bmatrix}. \tag{22}$$

Let $\boldsymbol{z} \equiv \boldsymbol{U}^\top\boldsymbol{\xi}$. This accords with our previous definitions, in that the z_i, $i = 1, \ldots, l$, are the first l components of $\boldsymbol{z}$. Then

$$\boldsymbol{\xi}^\top\boldsymbol{A}\boldsymbol{\xi} = \boldsymbol{z}^\top\boldsymbol{U}^\top\boldsymbol{A}\boldsymbol{U}\boldsymbol{z} = {\boldsymbol{z}_1}^\top\boldsymbol{z}_1 + (1-K){\boldsymbol{z}_2}^\top\boldsymbol{z}_2, \tag{23}$$

where $\boldsymbol{z}_1$ is the l–vector made up of the first l components of $\boldsymbol{z}$, and $\boldsymbol{z}_2$ contains the other components.

The denominator of the second term on the right-hand side of (13) is thus given by (23). The numerator is, as before, just z_1. If we now transform to the polar coordinates for the full n–vector $\boldsymbol{z}$, rather than just for the l components of $\boldsymbol{z}_1$, we have, in terms of these new polar coordinates, that $z_1 = r\cos\theta_1$, as before. Then, from the result (19) applied to our n–dimensional coordinates, we see that

$${\boldsymbol{z}_2}^\top\boldsymbol{z}_2 = \sum_{i=l+1}^{n} z_i^2 = r^2\prod_{j=1}^{l}\sin^2\theta_j,$$

so that

$${\boldsymbol{z}_1}^\top\boldsymbol{z}_1 = \boldsymbol{z}^\top\boldsymbol{z} - {\boldsymbol{z}_2}^\top\boldsymbol{z}_2 = r^2 - {\boldsymbol{z}_2}^\top\boldsymbol{z}_2 = r^2\Big(1 - \prod_{j=1}^{l}\sin^2\theta_j\Big).$$

The right-hand side of (23) can therefore be written as

$$r^2\Big(1 - \prod_{j=1}^{l}\sin^2\theta_j + (1-K)\prod_{j=1}^{l}\sin^2\theta_j\Big) = r^2\Big(1 - K\prod_{j=1}^{l}\sin^2\theta_j\Big).$$

Note that, because $K < 1$, the right-hand side above cannot vanish for any values of the θ_j. The m^{th}moment of the K–class estimator is therefore a multiple integral with finite angle integrals and an integral over r of the function

$$r^{n-m-1}\exp\Big(-\tfrac{1}{2}(r^2 - 2ar\cos\theta_1)\Big).$$

This integral diverges unless $n - m - 1 > -1$, that is, unless $m < n$.

4. Existence of Moments for JIVE

We now move on to the JIVE estimator, for which the matrix $\boldsymbol{A}$ is given in (16) by its typical element. As above, we construct an orthogonal matrix $\boldsymbol{U}$, of which the first l columns are those of $\boldsymbol{W}$. Since $\boldsymbol{AW} = \boldsymbol{W}$, the first l columns of $\boldsymbol{AU}$ are also just the columns of $\boldsymbol{W}$. In order to construct the remaining $n-l$ columns of $\boldsymbol{U}$, consider the vectors $\boldsymbol{\eta}_i$, $i = 1, \ldots, n$, defined so as to have typical element $u_{ti}/(1-h_t)$, where $h_t = \|\boldsymbol{W}_t\|^2$ as before. Choose the vectors $\boldsymbol{u}_j$, $j = l+1, \ldots, l+k$, as an orthonormal basis of the space spanned by the vectors $\boldsymbol{M_W \eta}_i$, $i = 1, \ldots, l$. Clearly, $k \leq l$. The remaining orthonormal basis vectors, $\boldsymbol{u}_j$, $j = l+k+1, \ldots, n$, are then orthogonal to both the $\boldsymbol{w}_i$, $i = 1, \ldots, l$, and the $\boldsymbol{u}_j$, $j = l+1, \ldots, l+k$, and so also to the $\boldsymbol{\eta}_i$, $i = 1, \ldots, l$.

With this construction, the t^{th} element of $\boldsymbol{Au}_j$, for $j = l+1, \ldots, n$, is

$$-\frac{u_{tj}h_t}{1-h_t} = u_{tj} - \frac{u_{tj}}{1-h_t},$$

so that $\boldsymbol{Au}_j = \boldsymbol{u}_j - \boldsymbol{\eta}_j$. Define the $n \times (n-l)$ matrix $\boldsymbol{Z}$ by the relation $\boldsymbol{U} = [\boldsymbol{W} \quad \boldsymbol{Z}]$. Then, putting our various results together, we see that

$$\boldsymbol{U}^\top \boldsymbol{AU} = \begin{bmatrix} \boldsymbol{W}^\top \\ \boldsymbol{Z}^\top \end{bmatrix} [\boldsymbol{W} \quad (\boldsymbol{Z} - \boldsymbol{H})], \tag{24}$$

where $\boldsymbol{H} \equiv [\boldsymbol{H}_2 \quad \boldsymbol{H}_3]$, with $\boldsymbol{H}_2 = [\boldsymbol{\eta}_{l+1} \ldots \boldsymbol{\eta}_{l+k}]$ and $\boldsymbol{H}_3 = [\boldsymbol{\eta}_{l+k+1} \ldots \boldsymbol{\eta}_n]$. Consider the product $\boldsymbol{W}^\top \boldsymbol{H}_3$. Element (i, j) of this matrix is

$$\sum_{t=1}^{n} \frac{w_{ti}u_{tj}}{1-h_t} = \boldsymbol{\eta}_i^\top \boldsymbol{u}_j = 0,$$

since we saw above that the $\boldsymbol{u}_j$ are orthogonal to the $\boldsymbol{\eta}_i$.

These results show that equation (24) can be written in partitioned form as

$$\boldsymbol{U}^\top \boldsymbol{AU} = \mathbf{I} - \begin{bmatrix} \mathbf{O} & \boldsymbol{D}_{12} & \mathbf{O} \\ \mathbf{O} & \boldsymbol{D}_{22} & \boldsymbol{D}_{23} \\ \mathbf{O} & \boldsymbol{D}_{32} & \boldsymbol{D}_{33} \end{bmatrix}, \tag{25}$$

where element (i, j) of the $n \times n$ matrix $\boldsymbol{D}$ is

$$d_{ij} = \sum_{t=1}^{n} \frac{u_{ti}u_{tj}}{1-h_t}. \tag{26}$$

The index 1 of the partition refers to elements in the range $1, \ldots, l$, the index 2 to those in the range $l+1, \ldots, l+k$, and the index 3 to the other elements.

As before, we let $\boldsymbol{z} = \boldsymbol{U}^\top\boldsymbol{\xi}$, and we partition $\boldsymbol{z}^\top$ as $[\boldsymbol{z}_1^\top \;\; \boldsymbol{z}_2^\top \;\; \boldsymbol{z}_3^\top]$ conformably with the partition of $\boldsymbol{D}$. Then, from (21),

$$\boldsymbol{\xi}^\top\boldsymbol{A}\boldsymbol{\xi} = \boldsymbol{z}_1^\top(\boldsymbol{z}_1 - \boldsymbol{D}_{12}\boldsymbol{z}_2) + [\boldsymbol{z}_2^\top \;\; \boldsymbol{z}_3^\top]\begin{bmatrix} \mathbf{I} - \boldsymbol{D}_{22} & -\boldsymbol{D}_{23} \\ -\boldsymbol{D}_{32} & \mathbf{I} - \boldsymbol{D}_{33} \end{bmatrix}\begin{bmatrix} \boldsymbol{z}_2 \\ \boldsymbol{z}_3 \end{bmatrix}. \tag{27}$$

Next, we see that

$$\boldsymbol{w}_1^\top\boldsymbol{A}\boldsymbol{\xi} = \boldsymbol{u}_1^\top\boldsymbol{A}\boldsymbol{U}\boldsymbol{z} = z_1 - \boldsymbol{d}^\top\boldsymbol{z}_2. \tag{28}$$

Here the $1 \times k$ row vector $\boldsymbol{d}^\top$ denotes the top row of the matrix $\boldsymbol{D}_{12}$, and so (28) holds because $\boldsymbol{u}_1^\top\boldsymbol{A}\boldsymbol{U}$ is just the top row of (25).

Consider next the expectation of $\boldsymbol{w}_1^\top\boldsymbol{A}\boldsymbol{\xi}/\boldsymbol{\xi}^\top\boldsymbol{A}\boldsymbol{\xi}$ conditional on $\boldsymbol{z}_2$ and $\boldsymbol{z}_3$, recalling that the components of $\boldsymbol{z}$ are mutually independent. This conditional expectation, should it exist, can therefore be computed using the marginal density of the vector $\boldsymbol{z}_1$. We may apply a linear transformation to this vector which leaves the first component, z_1, unchanged and rotates the remaining $l-1$ components in such a way that, for the value of $\boldsymbol{z}_2$ on which we are conditioning, all components of the vector $\boldsymbol{D}_{12}\boldsymbol{z}_2$ vanish except the first two, which we denote by δ_1 and δ_2, respectively. Thus $\boldsymbol{d}^\top\boldsymbol{z}_2$, which is the first component of $\boldsymbol{D}_{12}\boldsymbol{z}_2$, is equal to δ_1. Since the components of $\boldsymbol{z}_1$ except for the first are multivariate standard normal, and since the first component is unaffected by the rotation of the other components, the joint density is also unaffected by the transformation.

Next we make use of the l–dimensional polar coordinates that correspond to the transformed $\boldsymbol{z}_1$. In terms of these, expression (28) becomes $r\cos\theta_1 - \delta_1$, and (27) becomes

$$r^2 - \delta_1 r\cos\theta_1 - \delta_2 r\sin\theta_1\cos\theta_2 - [\boldsymbol{z}_2^\top \;\; \boldsymbol{z}_3^\top]\begin{bmatrix} \boldsymbol{D}_{22} - \mathbf{I} & \boldsymbol{D}_{23} \\ \boldsymbol{D}_{32} & \boldsymbol{D}_{33} - \mathbf{I} \end{bmatrix}\begin{bmatrix} \boldsymbol{z}_2 \\ \boldsymbol{z}_3 \end{bmatrix}.$$

Thus the conditional expectation we wish to evaluate can be written, if it exists, as a multiple integral with finite angle integrals and an integral over the radial coordinate r with integrand

$$r^{l-1}\frac{r\cos\theta_1 - \delta_1}{r^2 - \delta_1 r\cos\theta_1 - \delta_2 r\sin\theta_1\cos\theta_2 - b^2}\exp\Big(-\tfrac{1}{2}(r^2 - 2ar\cos\theta_1)\Big). \tag{29}$$

The quantity b^2 is defined as

$$b^2 = [\boldsymbol{z}_2^\top \;\; \boldsymbol{z}_3^\top]\begin{bmatrix} \boldsymbol{D}_{22} - \mathbf{I} & \boldsymbol{D}_{23} \\ \boldsymbol{D}_{32} & \boldsymbol{D}_{33} - \mathbf{I} \end{bmatrix}\begin{bmatrix} \boldsymbol{z}_2 \\ \boldsymbol{z}_3 \end{bmatrix}, \tag{30}$$

and it is indeed positive, since the matrix in its definition is positive definite, as can be seen by noting from (26) that a typical element of the matrix $\boldsymbol{D} - \mathbf{I}$ is

$$\sum_{t=1}^{n} u_{ti}u_{tj}\frac{h_t}{1 - h_t},$$

where we use the fact that $\sum_{t=1}^{n} u_{ti}u_{tj} = \delta_{ij}$ by the orthonormality of the $\boldsymbol{u}_j$. Thus $\boldsymbol{D} - \mathbf{I} = \boldsymbol{U}^\top \boldsymbol{Q}\boldsymbol{U}$, where $\boldsymbol{Q}$ is the diagonal matrix with typical element $h_t/(1 - h_t)$. Since we assume that $0 < h_t < 1$, $\boldsymbol{Q}$ is positive definite. This implies that $\boldsymbol{D} - \mathbf{I}$ and the lower-right block of $\boldsymbol{D} - \mathbf{I}$ that appears in (30) are also positive definite.

Unlike what we found for the conventional IV estimator and the K–class estimator with $K < 1$, the denominator of (29) does not vanish at $r = 0$. However, it does have a simple pole for a positive value of r. Let $\delta_1 \cos\theta_1 + \delta_2 \sin\theta_1 \cos\theta_2 = d$. Then the denominator can be written as $r^2 - rd - b^2$, which has zeros at

$$r = \frac{d \pm \sqrt{d^2 + 4b^2}}{2}.$$

The discriminant is obviously positive, so that the roots are real, one being positive and the other negative, whatever the sign of d. The positive zero causes the integral over r to diverge, from which we conclude that the JIVE estimator has no moments.

It is illuminating to rederive this result for the special case in which the design of the instruments is perfectly balanced, in the sense that $h_t = l/n$ for all t. This is the case, for instance, if the instruments are all seasonal dummies and the sample contains an integer number of years. A major simplification arises from the fact that $k = 0$ in this circumstance. This follows because, for $i = 1, \ldots, l$, $\boldsymbol{\eta}_i$ is proportional to $\boldsymbol{w}_i$, so that $\boldsymbol{M_W}\boldsymbol{\eta}_i = \mathbf{0}$. Thus block 2 of the threefold partition introduced above disappears.

The second simplification is that the matrix $\boldsymbol{A}$ becomes $\boldsymbol{P_W} - \big(l/(n - l)\big)\boldsymbol{M_W}$, as can be seen from (16) by setting $\|\boldsymbol{W}_t\|^2 = l/n$. The denominator $\boldsymbol{\xi}^\top\boldsymbol{A}\boldsymbol{\xi}$ can then be written as

$$\sum_{i=1}^{l} (\boldsymbol{w}_i^\top\boldsymbol{\xi})^2 - \frac{l}{n-l}\|\boldsymbol{M_W}\boldsymbol{\xi}\|^2.$$

The first term is just $\|\boldsymbol{z}_1\|^2 = r^2$, and minus the second term, which replaces the b^2 of (30), is clearly positive. Thus $\boldsymbol{\xi}^\top\boldsymbol{A}\boldsymbol{\xi} = r^2 - b^2 = (r + b)(r - b)$, and the singularity at $r = b$ is what causes the divergence.

5. Exogenous Explanatory Variables in the Structural Equation

In most econometric models, the structural equation contains exogenous explanatory variables in addition to the endogenous one, and these extra explanatory variables are included in the set of instrumental variables. In this section, we briefly indicate how to extend our previous results to this more general case.

We extend the model (1) as follows:

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{x}\beta + \boldsymbol{W}_2\boldsymbol{\gamma} + \sigma_u\boldsymbol{u}, \\ \boldsymbol{x} &= \sigma_v(\boldsymbol{W}_1\boldsymbol{a}_1 + \boldsymbol{W}_2\boldsymbol{a}_2 + \boldsymbol{v}). \end{aligned} \tag{31}$$

Here $\boldsymbol{W}_2$ has l' columns, and the full set of instruments is contained in the matrix $\boldsymbol{W} = [\boldsymbol{W}_1 \;\; \boldsymbol{W}_2]$, where without loss of generality $\boldsymbol{W}^\top\boldsymbol{W} = \mathbf{I}$, as before. Again without loss of generality, we assume that $\boldsymbol{W}_1\boldsymbol{a}_1 = a\boldsymbol{w}_1$, where $\boldsymbol{w}_1$ is the first column of $\boldsymbol{W}_1$.

We will show in a moment that all of the estimators we have considered so far can still, when applied to the model (31), be expressed as solutions to estimating equations of the form

$$\boldsymbol{x}^\top\boldsymbol{A}^\top(\boldsymbol{y} - \boldsymbol{x}\beta) = 0, \tag{32}$$

where the matrix $\boldsymbol{A}$ now must satisfy the requirements that $\boldsymbol{A}\boldsymbol{W}_1 = \boldsymbol{W}_1$ and that $\boldsymbol{A}\boldsymbol{W}_2 = \boldsymbol{A}^\top\boldsymbol{W}_2 = \mathbf{O}$. Assuming that these requirements are met, we can see easily enough that equations (8), (9), and (10) are still satisfied. Consequently, as before, the estimator has as many moments as the expression $\boldsymbol{w}_1^\top\boldsymbol{A}\boldsymbol{\xi}/\boldsymbol{\xi}^\top\boldsymbol{A}\boldsymbol{\xi}$ that appears in (13).

Consider first the generalized IV estimator. The estimating equations for both β and $\boldsymbol{\gamma}$ are given by

$$\begin{aligned} \boldsymbol{x}^\top\boldsymbol{P_W}(\boldsymbol{y} - \boldsymbol{x}\beta - \boldsymbol{W}_2\gamma) &= 0 \quad \text{and} \\ \boldsymbol{W}_2^\top(\boldsymbol{y} - \boldsymbol{x}\beta - \boldsymbol{W}_2\gamma) &= \mathbf{0}. \end{aligned} \tag{33}$$

There is no factor of $\boldsymbol{P_W}$ in the second equation here, because $\boldsymbol{P_W}\boldsymbol{W}_2 = \boldsymbol{W}_2$. If we premultiply this second equation by $\boldsymbol{x}^\top\boldsymbol{W}_2(\boldsymbol{W}_2^\top\boldsymbol{W}_2)^{-1}$, it becomes

$$\boldsymbol{x}^\top\boldsymbol{P}_2(\boldsymbol{y} - \boldsymbol{x}\beta - \boldsymbol{W}_2\gamma) = 0,$$

where $\boldsymbol{P}_2$ projects orthogonally on to the span of the columns of $\boldsymbol{W}_2$. Subtracting this last equation from the first one in (33), we obtain the generalized IV estimating equation for β alone:

$$\boldsymbol{x}^\top\boldsymbol{P}_1(\boldsymbol{y} - \boldsymbol{x}\beta) = 0,$$

where $\boldsymbol{P}_1$ projects orthogonally on to the span of the columns of $\boldsymbol{W}_1$. Observe that, because of the orthonormality of the columns of $\boldsymbol{W}$, $\boldsymbol{P_W} - \boldsymbol{P}_2 = \boldsymbol{P}_1$, and $\boldsymbol{P}_1\boldsymbol{W}_2 = \mathbf{O}$. For the IV estimator, then, $\boldsymbol{A} = \boldsymbol{P}_1$, which does indeed satisfy the required conditions that $\boldsymbol{A}\boldsymbol{W}_1 = \boldsymbol{W}_1$ and $\boldsymbol{A}\boldsymbol{W}_2 = \boldsymbol{A}^\top\boldsymbol{W}_2 = \mathbf{O}$.

Thus, for the IV estimator, the expression $\boldsymbol{w}_1^\top\boldsymbol{A}\boldsymbol{\xi}/\boldsymbol{\xi}^\top\boldsymbol{A}\boldsymbol{\xi}$ becomes

$$\frac{\boldsymbol{w}_1^\top\boldsymbol{\xi}}{\boldsymbol{\xi}^\top\boldsymbol{P}_1\boldsymbol{\xi}} = \frac{z_1}{\sum_{i=1}^{l} z_i^2},$$

just as in (18), where now l is the dimension of $\boldsymbol{W}_1$. The remainder of the analysis of the IV estimator follows exactly as before.

The K–class estimator with exogenous $K < 1$ can readily be seen to be defined by the estimating equation

$$\boldsymbol{x}^\top(\boldsymbol{M}_2 - K\boldsymbol{M_W})(\boldsymbol{y} - \boldsymbol{x}\beta) = 0,$$

where $\boldsymbol{M}_2 = \mathbf{I} - \boldsymbol{P}_2$. Thus $\boldsymbol{A} = \boldsymbol{M}_2 - K\boldsymbol{M}_{\boldsymbol{W}}$, and it is easily verified that $\boldsymbol{A}\boldsymbol{W}_1 = \boldsymbol{W}_1$ and $\boldsymbol{A}\boldsymbol{W}_2 = \boldsymbol{A}^\top\boldsymbol{W}_2 = \mathbf{0}$. The matrix $\boldsymbol{U}^\top\boldsymbol{A}\boldsymbol{U}$, given before by (22), is now

$$\boldsymbol{U}^\top\boldsymbol{A}\boldsymbol{U} = \begin{bmatrix} \mathbf{I}_l & \mathbf{O}_{l,l'} & \mathbf{O}_{l,n'} \\ \mathbf{O}_{l',l} & \mathbf{O}_{l',l'} & \mathbf{O}_{l',n'} \\ \mathbf{O}_{n',l} & \mathbf{O}_{n',l'} & (1-K)\mathbf{I}_{n'} \end{bmatrix},$$

where $n' = n - l - l'$; recall that $\boldsymbol{W}_2$ has l' columns. It follows that $\boldsymbol{\xi}^\top\boldsymbol{A}\boldsymbol{\xi}$ is still given by (23), where $\boldsymbol{z}_2$ now contains the n' components defined by the dimensions that are orthogonal to all the instruments. Our earlier analysis remains good except that moments of order m exist only if $m < n - l'$.

For the JIVE estimator, the vector $\tilde{\boldsymbol{x}}$ of omit-one fitted values is defined exactly as before, using the full matrix $\boldsymbol{W}$ of instruments. In this case, the estimating equation (32) for β becomes

$$\tilde{\boldsymbol{x}}^\top\boldsymbol{M}_2(\boldsymbol{y} - \boldsymbol{x}\beta) = 0.$$

Previously, as can be seen from (17), it was just $\tilde{\boldsymbol{x}}^\top(\boldsymbol{y} - \boldsymbol{x}\beta) = 0$. Thus we see that the new $\boldsymbol{A}$ matrix is $\boldsymbol{M}_2$ times the old one. The old matrix satisfies $\boldsymbol{A}\boldsymbol{W} = \boldsymbol{W}$ by construction, and so the new one satisfies the required conditions that

$$\boldsymbol{M}_2\boldsymbol{A}\boldsymbol{W} = \boldsymbol{M}_2[\boldsymbol{W}_1 \quad \boldsymbol{W}_2] = [\boldsymbol{W}_1 \quad \mathbf{O}] \quad \text{and} \quad \boldsymbol{A}^\top\boldsymbol{M}_2\boldsymbol{W}_2 = \mathbf{O}.$$

For the $n' = n - l'$ dimensions orthogonal to $\boldsymbol{W}_2$, the columns of the orthogonal matrix $\boldsymbol{U}$ can be constructed just as before. In addition, the l' columns of $\boldsymbol{W}_2$ complete the set of columns of $\boldsymbol{U}$. Then, since (with the new $\boldsymbol{A}$) $\boldsymbol{A}\boldsymbol{W}_2 = \mathbf{O}$ and $\boldsymbol{W}_2^\top\boldsymbol{A} = \mathbf{O}$, the l' rows and columns of equation (25) corresponding to $\boldsymbol{W}_2$ are replaced by zeros. Thus equations (27) and (28) still hold with appropriate redefinitions of the vectors $\boldsymbol{z}_1$, $\boldsymbol{z}_2$, and $\boldsymbol{z}_3$ so as to exclude the eliminated l' dimensions. This means that $\boldsymbol{z}_1$ has l components, $\boldsymbol{z}_2$ has k, and $\boldsymbol{z}_3$ has $n' - l - k$. The analysis based on (27) and (28), leading to the conclusion that the JIVE estimator has no moments, then proceeds unaltered.

6. Consequences of Nonexistence of Moments

The fact that an estimator has no moments does not mean that it is necessarily a bad estimator, although it does suggest that extreme estimates are likely to be encountered relatively often. However, when the sample size is large enough and, for cases like the simultaneous equations case we are considering here, the instruments are strong enough, this may not be a problem in practice.

In some ways, the lack of moments is more of a problem for investigators performing Monte Carlo experiments than it is for practitioners actually using the estimator. Suppose we perform N replications of a Monte Carlo experiment and obtain N realizations $\hat{\beta}_j$ of the estimator $\hat{\beta}$. It is natural to estimate the population mean of $\hat{\beta}$ by using the sample mean of the $\hat{\beta}_j$, which converges as $N \to \infty$ to the population mean if the latter exists. However, when the estimator has no first moment, what one is trying to estimate does not exist, and the sequence of sample means does not converge.

To illustrate this, we performed several simulation experiments based on the DGP (1). Both standard errors (σ_u and σ_v) were equal to 1.0, the sample size was 50, ρ was 0.8, β was 1, and a took on two different values, 0.5 and 1.0. There were five instruments, and hence four overidentifying restrictions. For each value of a, 29 different experiments were performed for various values of N, starting with $N = 1000$ and then multiplying N by a factor of (approximately) $\sqrt{2}$ as many times as necessary until it reached 16,384,000.

In the first experiment, $a = 0.5$. For this value of a and a sample size of only 50, the instruments are quite weak. As can be seen in Figure 1, the averages of the IV estimates converge quickly to a value of approximately 1.1802, which involves a rather serious upward bias. In contrast, the averages of the JIVE estimates are highly variables. These averages tend to be less than 1 most of the time, but they show no real pattern. The figure shows two different sets of results, based on different random numbers, for the JIVE estimates. Only one set is shown for IV, because, at the scale on which the figure is drawn, the two sets would be almost indistinguishable.

In the second experiment, $a = 1.0$, and the instruments are therefore a good deal stronger. As seen in Figure 2, the averages of the IV estimates now converge quickly to a value of approximately 1.0489, which involves much less upward bias than before. The averages of the JIVE estimates do not seem to converge, but they vary much less than they did in the first set of experiments, and it is clear that they tend to underestimate β.

In additional experiments that are not reported here, we also tried $a = 0.25$ and $a = 2$. In the former case, the results were quite similar to those in Figure 1, except that the upward bias of the IV estimator was much greater. In the latter case, where the instruments were quite strong, the averages of both the IV and JIVE estimates appeared to converge, to roughly 1.012 for the former and 0.991 for the latter. Thus, based on the simulation results for this case, there was no sign that the JIVE estimator lacks moments. It would presumably require very large values of N to illustrate the lack of moments when the instruments are strong.

7. Conclusions

In this paper, we have proposed a novel way to investigate the existence of moments for instrumental variables and related estimators in the linear regression model. Our method is based on the use of polar coordinates. For generalized IV and K–class estimators, we obtain standard results in a new and simpler way. However, the main result of the paper concerns the estimator called JIVE. We show that this estimator has no moments. Simulation results suggest that, when the instruments are sufficiently weak, JIVE's lack of moments is very evident. However, when the instruments are strong, it may not be apparent.

References

Angrist J. D., G. W. Imbens, and A. B. Krueger (1999). "Jackknife instrumental variables estimation," *Journal of Applied Econometrics*, **14**, 57–67.

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, Third Edition. Hoboken, NJ, John Wiley & Sons.

Blomquist, S., and M. Dahlberg (1999). "Small sample properties of LIML and jackknife IV estimators: Experiments with weak instruments," *Journal of Applied Econometrics*, **14**, 69–88.

Davidson, R., and J. G. MacKinnon (2004). *Econometric Theory and Methods*, New York, Oxford University Press.

Davidson, R., and J. G. MacKinnon (2006). "The case against JIVE," *Journal of Applied Econometrics*, **21**, forthcoming.

Fuller, W. A. (1977). "Some properties of a modification of the limited information estimator," *Econometrica*, **45**, 939–53.

Hahn, J., J. A. Hausman, and G. Kuersteiner (2004). "Estimation with weak instruments: Accuracy of higher order bias and MSE approximations," *Econometrics Journal*, **7**, 272–306.

Kinal, T. W. (1980). "The existence of moments of k-class estimators," *Econometrica*, **48**, 241–49.

Mariano, R. S. (2001). "Simultaneous equation model estimators: Statistical properties and practical implications," Ch. 6 in *A Companion to Econometric Theory*, ed. B. Baltagi, Oxford, Blackwell Publishers, 122–43.

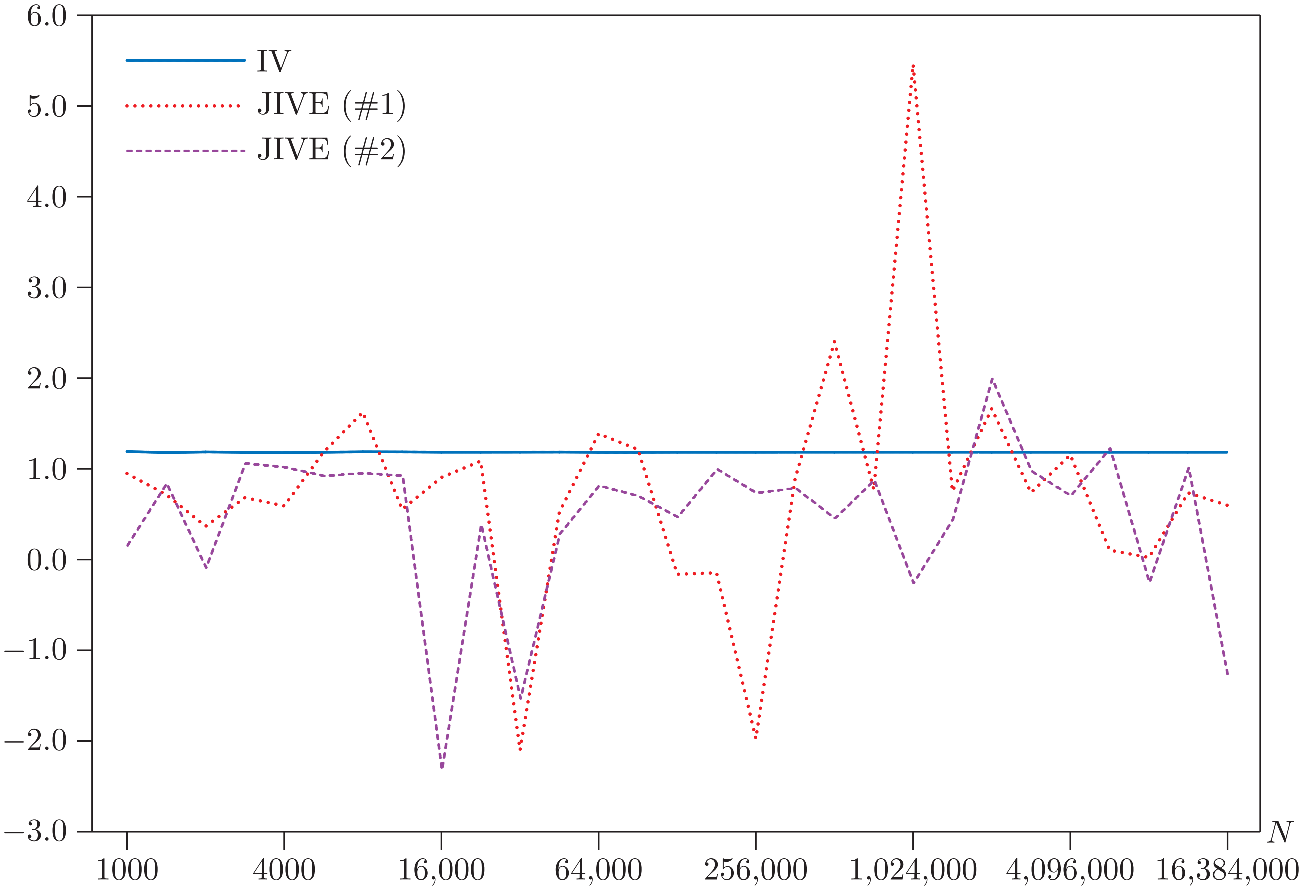

Figure 1. Means of two estimates of β: $n = 50$, $a = 0.5$

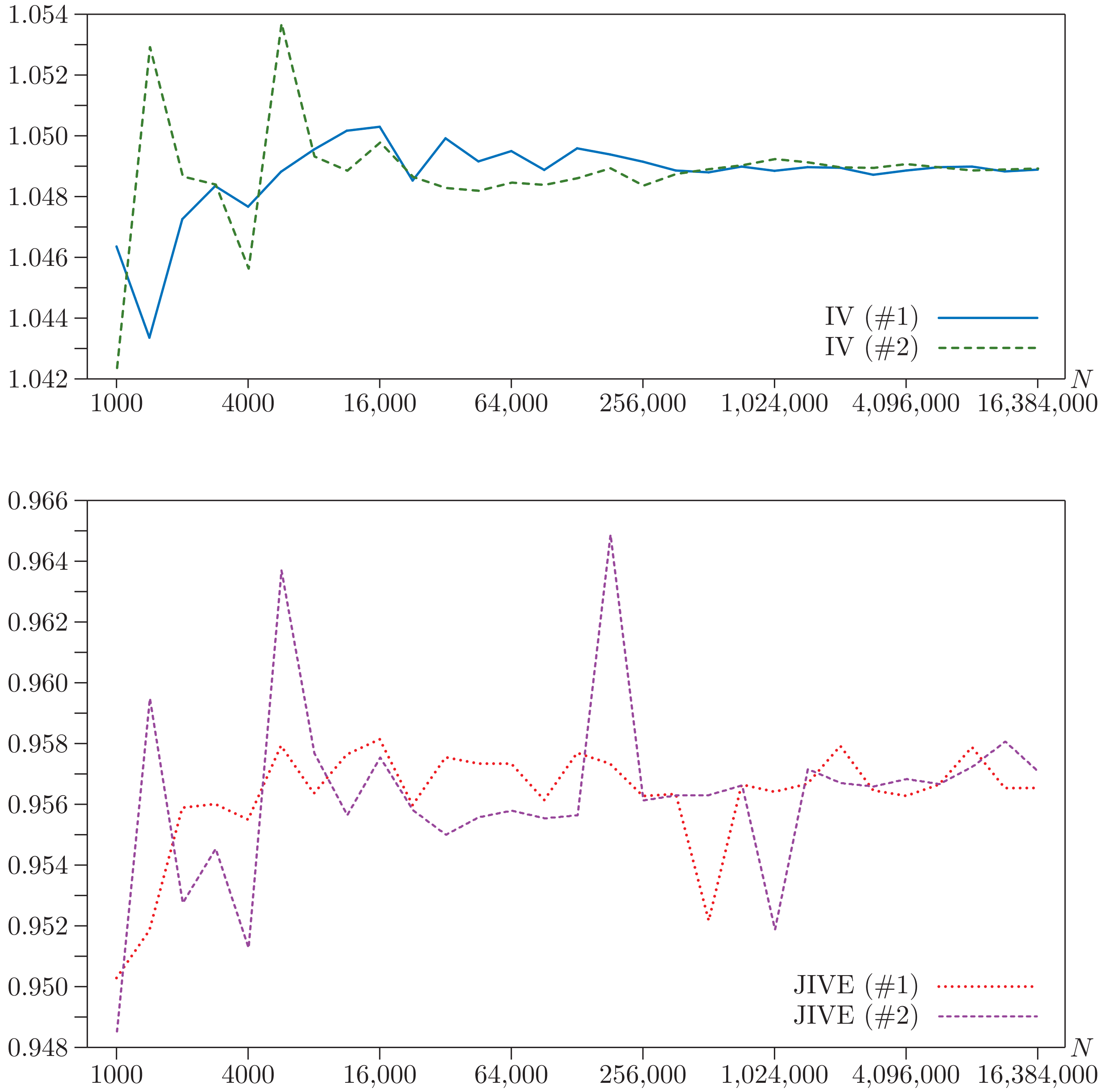

Figure 2. Means of two estimates of β: n = 50, a = 1.0