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 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. We are grateful to two anonymous referees for comments on earlier versions.

June, 2007.

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 method based on the use of polar coordinates to study the existence, or non-existence, of moments of IV estimators. This approach resembles that of Forchini and Hillier (2003) and is considerably easier than the methods used prior to that paper; see also Hillier (2006).[1]

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 some standard results about the existence of moments for IV and certain 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:

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

[1] We are grateful to an anonymous referee for bringing these two papers to our attention.

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 for 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) solves the estimating equation

$$\boldsymbol{w}_1^\top(\boldsymbol{y} - \boldsymbol{x}\hat{\beta}_{\text{IV}}) = 0, \tag{2}$$

where the n–vectors $\boldsymbol{y}$ and $\boldsymbol{x}$ have typical elements y_t and x_t, respectively. It is well known that the estimator defined by (2) has no moments, since there are no overidentifying restrictions.

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 solves the estimating equation

$$\boldsymbol{x}^\top\boldsymbol{P_W}(\boldsymbol{y} - \boldsymbol{x}\hat{\beta}_{\text{IV}}) = 0, \tag{3}$$

where $\boldsymbol{P_W} \equiv \boldsymbol{W}(\boldsymbol{W}^\top\boldsymbol{W})^{-1}\boldsymbol{W}^\top$ is the orthogonal projection on to the span of the columns of $\boldsymbol{W}$. The degree of overidentification is $l-1$, which is also the number of moments that the estimator possesses. If $l = 1$, the estimating equations (2) and (3) are equivalent. 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 $\boldsymbol{W}$. Since $\hat{\beta}_{\text{IV}}$ depends on $\boldsymbol{W}$ only through the linear span of its columns, there is also no loss of generality in supposing that $\boldsymbol{W}^\top\boldsymbol{W} = \mathbf{I}$, the $l \times l$ identity matrix.

Replacing the matrix $\boldsymbol{P_W}$ in (3) by other matrices $\boldsymbol{A}^\top$ with the property that $\boldsymbol{AW} = \boldsymbol{W}$ 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 $\boldsymbol{P_W}$, then we obtain a K–class estimator. With $K = 1$, of course, we recover the generalized IV estimator $\hat{\beta}_{\text{IV}}$.

Quite generally, consider the estimator $\hat{\beta}$ defined by the estimating equation

$$\boldsymbol{x}^\top\boldsymbol{A}^\top(\boldsymbol{y} - \boldsymbol{x}\hat{\beta}) = 0, \tag{4}$$

where the $n \times n$ matrix $\boldsymbol{A}$ depends somehow on the linear span of the columns of an $n \times l$ matrix $\boldsymbol{W}$ of exogenous instruments, and is such that $\boldsymbol{AW} = \boldsymbol{W}$. If we denote by β^0 the true parameter of the DGP (1), we see that

$$\begin{aligned}\boldsymbol{x}^\top\boldsymbol{A}^\top(\boldsymbol{y} - \boldsymbol{x}\beta^0) &= \sigma_u\boldsymbol{x}^\top\boldsymbol{A}^\top\boldsymbol{u} = \sigma_u\sigma_v(a\boldsymbol{w}_1^\top + \boldsymbol{v}^\top)\boldsymbol{A}^\top(\rho\boldsymbol{v} + \boldsymbol{u}_1)\\ &= \sigma_u\sigma_v(a\rho\boldsymbol{w}_1^\top\boldsymbol{A}^\top\boldsymbol{v} + \rho\boldsymbol{v}^\top\boldsymbol{A}^\top\boldsymbol{v} + a\boldsymbol{w}_1^\top\boldsymbol{A}^\top\boldsymbol{u}_1 + \boldsymbol{v}^\top\boldsymbol{A}^\top\boldsymbol{u}_1),\end{aligned} \tag{5}$$

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

$$\mathrm{E}\big(\boldsymbol{x}^\top\boldsymbol{A}^\top(\boldsymbol{y}-\boldsymbol{x}\beta^0)\,|\,\boldsymbol{v}\big) = \rho\sigma_u\sigma_v(a{\boldsymbol{w}_1}^\top\boldsymbol{v}+\boldsymbol{v}^\top\boldsymbol{A}\boldsymbol{v}). \tag{6}$$

Similarly,

$$\boldsymbol{x}^\top\boldsymbol{A}^\top\boldsymbol{x} = \sigma_v^2(a^2+a\boldsymbol{v}^\top\boldsymbol{A}^\top\boldsymbol{w}_1+a{\boldsymbol{w}_1}^\top\boldsymbol{v}+\boldsymbol{v}^\top\boldsymbol{A}\boldsymbol{v}). \tag{7}$$

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 $\boldsymbol{A}^\top\boldsymbol{w}_1=\boldsymbol{w}_1$. Subtracting the estimating equation (4) from (6), dividing by $\boldsymbol{x}^\top\boldsymbol{A}^\top\boldsymbol{x}$, and using (7) gives for the estimator $\hat\beta$ that

$$\mathrm{E}(\hat\beta-\beta^0\,|\,\boldsymbol{v}) = \frac{\rho\sigma_u}{\sigma_v}\left(\frac{a{\boldsymbol{w}_1}^\top\boldsymbol{v}+\boldsymbol{v}^\top\boldsymbol{A}\boldsymbol{v}}{a^2+a\boldsymbol{v}^\top\boldsymbol{A}^\top\boldsymbol{w}_1+a{\boldsymbol{w}_1}^\top\boldsymbol{v}+\boldsymbol{v}^\top\boldsymbol{A}\boldsymbol{v}}\right). \tag{8}$$

Notice that both the numerator and the denominator of this conditional expectation vanish if $\boldsymbol{v}=-a\boldsymbol{w}_1$. We therefore make the change of variables

$$\boldsymbol{\xi}=\boldsymbol{v}+a\boldsymbol{w}_1,$$

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 (8) 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 (8) therefore simplifies to

$$\frac{\boldsymbol{\xi}^\top\boldsymbol{A}\boldsymbol{\xi}-a{\boldsymbol{w}_1}^\top\boldsymbol{A}\boldsymbol{\xi}}{\boldsymbol{\xi}^\top\boldsymbol{A}\boldsymbol{\xi}} = 1-a\frac{{\boldsymbol{w}_1}^\top\boldsymbol{A}\boldsymbol{\xi}}{\boldsymbol{\xi}^\top\boldsymbol{A}\boldsymbol{\xi}}. \tag{9}$$

Clearly, the estimator $\hat\beta$ has just as many moments as the second term on the right-hand side of (9), and we will come back to this point in the next section.

First, however, 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).[2] 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 $\boldsymbol{W}$:

$$\boldsymbol{x}=\boldsymbol{W}\boldsymbol{\gamma}+\sigma_v\boldsymbol{v}. \tag{10}$$

[2] 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.

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 (10) without observation t.

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

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

where the $\hat{u}_t$ are the residuals from regression (10) run on the full sample, and h_t is the t^{th} diagonal element of $\boldsymbol{P_W}$; see, for instance, equation (2.63) in Davidson and 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

$$\tilde{x}_t = \boldsymbol{W}_t\hat{\gamma}^{(t)} = (\boldsymbol{P_W x})_t - \frac{1}{1-\|\boldsymbol{W}_t\|^2}(\boldsymbol{P_W})_{tt}(\boldsymbol{M_W x})_t = x_t - \frac{(\boldsymbol{M_W x})_t}{1-\|\boldsymbol{W}_t\|^2}, \tag{11}$$

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 (11). 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

$$a_{ts} = \frac{1}{1-\|\boldsymbol{W}_t\|^2}\big((\boldsymbol{P_W})_{ts} - \delta_{ts}\|\boldsymbol{W}_t\|^2\big), \tag{12}$$

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. It is also clear that $\boldsymbol{A}$ depends on $\boldsymbol{W}$ only through the projection $\boldsymbol{P_W}$.

With $\tilde{\boldsymbol{x}}$ defined as above, the JIVE estimator $\hat{\beta}_{\text{JIV}}$ satisfies the estimating equation

$$\tilde{\boldsymbol{x}}^\top(\boldsymbol{y} - \boldsymbol{x}\hat{\beta}_{\text{JIV}}) = \boldsymbol{x}^\top\boldsymbol{A}^\top(\boldsymbol{y} - \boldsymbol{x}\hat{\beta}_{\text{JIV}}) = 0, \tag{13}$$

and so it falls into the class of estimators defined by equation (4).

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 by constructing an $n \times n$ orthogonal matrix $\boldsymbol{U}$ with its first l columns identical to those of the instrument matrix $\boldsymbol{W}$, and the remaining $n-l$ columns constituting an orthonormal basis for $\mathcal{S}^{\perp}(\boldsymbol{W})$, the orthogonal complement of the span of the columns of $\boldsymbol{W}$. If we denote column j of $\boldsymbol{U}$ by $\boldsymbol{u}_j$, we can define $z_j \equiv \boldsymbol{u}_j^{\top}\boldsymbol{\xi}$, for $j = 1, \ldots, n$. Then the z_i are mutually independent, and they are distributed as standard normal except for z_1, which is distributed as $\mathrm{N}(a, 1)$. Because $\boldsymbol{U}$ is an orthogonal matrix,

$$\boldsymbol{\xi}^{\top}\boldsymbol{A}\boldsymbol{\xi} = \boldsymbol{\xi}^{\top}\boldsymbol{U}\boldsymbol{U}^{\top}\boldsymbol{A}\boldsymbol{U}\boldsymbol{U}^{\top}\boldsymbol{\xi} = \boldsymbol{z}^{\top}\boldsymbol{U}^{\top}\boldsymbol{A}\boldsymbol{U}\boldsymbol{z}, \tag{14}$$

where $\boldsymbol{z}$ is the n–vector with typical element z_j. Similarly, since $\boldsymbol{w}_1 = \boldsymbol{u}_1$, we see that

$$\boldsymbol{w}_1^{\top}\boldsymbol{A}\boldsymbol{\xi} = \boldsymbol{u}_1^{\top}\boldsymbol{A}\boldsymbol{U}\boldsymbol{z}. \tag{15}$$

We now once more change variables, so as to use the polar coordinates that correspond to the Cartesian coordinates z_i, $i = 1, \ldots, n$.[3] The first polar coordinate is denoted r, and it is the positive square root of $\sum_{i=1}^{n} z_i^2$. The other polar coordinates, denoted $\theta_1, \ldots, \theta_{n-1}$, are all angles. They are defined as follows:

$$\begin{aligned} z_1 &= r\cos\theta_1, \\ z_i &= r\cos\theta_i \prod_{j=1}^{i-1}\sin\theta_j, \quad i = 2, \ldots, n-1, \text{ and} \\ z_n &= r\prod_{j=1}^{n-1}\sin\theta_j, \end{aligned} \tag{16}$$

where $0 \le \theta_i < \pi$ for $i = 1, \ldots, n-2$, and $0 \le \theta_{n-1} < 2\pi$. It can be shown easily that for $1 \le j < n$,

$$\sum_{i=j+1}^{n} z_i^2 = r^2 \prod_{i=1}^{j}\sin^2\theta_i. \tag{17}$$

The joint density of the z_i is

$$\frac{1}{(2\pi)^{n/2}}\exp\Big(-\frac{1}{2}\Big((z_1-a)^2 + \sum_{i=2}^{n} z_i^2\Big)\Big) = \frac{e^{-a^2/2}}{(2\pi)^{n/2}}\exp\Big(-\frac{1}{2}(r^2 - 2ar\cos\theta_1)\Big). \tag{18}$$

[3] The use of polar coordinates in more than two dimensions is relatively uncommon. Anderson (2003, p. 285) uses somewhat different conventions from ours, but with similar results.

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

$$\frac{\partial(z_1, \ldots, z_n)}{\partial(r, \theta_1, \ldots, \theta_{n-1})} = r^{n-1} \sin^{n-2} \theta_1 \sin^{n-3} \theta_2 \ldots \sin \theta_{n-2}\,; \tag{19}$$

see James (1954) and Anderson (2003, p. 286).

We first consider the generalized IV estimator defined by (3), for which $\boldsymbol{A} = \boldsymbol{A}^\top = \boldsymbol{P_W}$. Partition the matrix $\boldsymbol{U}$ as $[\boldsymbol{W} \quad \boldsymbol{Z}]$, where $\boldsymbol{Z}$ is an $n \times (n-l)$ matrix. Then

$$\boldsymbol{U}^\top \boldsymbol{A} \boldsymbol{U} = \begin{bmatrix} \boldsymbol{W}^\top \\ \boldsymbol{Z}^\top \end{bmatrix} \boldsymbol{P_W} \, [\boldsymbol{W} \quad \boldsymbol{Z}\,] = \begin{bmatrix} \mathbf{I}_l & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}. \tag{20}$$

For this estimator, therefore, the expression $\boldsymbol{w}_1^\top \boldsymbol{A\xi} / \boldsymbol{\xi}^\top \boldsymbol{A\xi}$ that occurs in (9) and determines how many moments the estimator has, becomes, with the help of (14) and (15),

$$\frac{\boldsymbol{w}_1^\top \boldsymbol{\xi}}{\boldsymbol{\xi}^\top \boldsymbol{P_W} \boldsymbol{\xi}} = \frac{z_1}{\sum_{i=1}^{l} z_i^2}. \tag{21}$$

Since (21) depends only on the z_i for $i = 1, \ldots, l$, we can use polar coordinates based on these only. This means that we replace n by l in (16), (18), and (19). With these polar coordinates, we find that $z_1 = r \cos \theta_1$ and $\sum_{i=1}^{l} z_i^2 = r^2$. Consequently, the m^{th} moment of (21), if it exists, is given by the integral

$$\delta \int_0^\pi \cos^m \theta_1 \sin^{l-2} \theta_1 \int_0^\infty r^{l-m-1} \exp\Big(-\tfrac{1}{2}(r^2 - 2ar \cos \theta_1)\Big)\, dr\, d\theta_1, \tag{22}$$

where

$$\delta \equiv \frac{e^{-a^2/2}}{(2\pi)^{(l-2)/2}} \prod_{j=2}^{l-2} \int_0^\pi \sin^{l-j-1} \theta_j \, d\theta_j.$$

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 (22) 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, for the special case of just one endogenous regressor, some of the more general results on the existence of moments of various estimators which were first demonstrated in Kinal (1980). Forchini and Hillier (2003) also proved the above results by arguments similar to ours.

As an example, and because the analysis is quite similar to that of the JIVE estimator in the next section, we consider another class of estimators examined in Kinal (1980), namely, K–class estimators with fixed $K < 1$. For these estimators, as we saw earlier, the matrix $\boldsymbol{A}$ is $\mathbf{I} - K\boldsymbol{M_W}$.

Observe that

$$\boldsymbol{AU} = (\mathbf{I} - K\boldsymbol{M_W})[\boldsymbol{W} \quad \boldsymbol{Z}] = [\boldsymbol{W} \quad (1-K)\boldsymbol{Z}],$$

from which it is easy to see that

$$\boldsymbol{U}^\top\boldsymbol{AU} = \begin{bmatrix} \mathbf{I}_l & \mathbf{O} \\ \mathbf{O} & (1-K)\mathbf{I}_{n-l} \end{bmatrix}. \tag{23}$$

From (14), it follows that

$$\boldsymbol{\xi}^\top\boldsymbol{A\xi} = \boldsymbol{z}^\top\boldsymbol{U}^\top\boldsymbol{AUz} = \boldsymbol{z}_1^\top\boldsymbol{z}_1 + (1-K)\boldsymbol{z}_2^\top\boldsymbol{z}_2, \tag{24}$$

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. Similarly, from (15),

$$\boldsymbol{w}_1^\top\boldsymbol{A\xi} = \boldsymbol{u}_1^\top\boldsymbol{AUz} = z_1,$$

where z_1 is the first element of $\boldsymbol{z}$. We now use the full transformation to polar coordinates, rather than just the one for the l components of $\boldsymbol{z}_1$. We have that $z_1 = r\cos\theta_1$, as before. Then, from the result (17) 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

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

The right-hand side of (24) can therefore be written as

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

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

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

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 specified in (12) by its typical element. By (12), the t^{th} element of $\boldsymbol{A}\boldsymbol{u}_j$, for $j = l+1, \ldots, n$, is

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

Since $\boldsymbol{A}\boldsymbol{W} = \boldsymbol{W}$, it follows that

$$\boldsymbol{U}^\top\boldsymbol{A}\boldsymbol{U} = \mathbf{I} - \boldsymbol{D}, \tag{25}$$

where the first l columns of $\boldsymbol{D}$ are zero, and the other elements are given by

$$d_{ij} = \sum_{t=1}^{n} \frac{u_{ti}u_{tj}}{1-h_t}, \quad i = 1, \ldots, n, \;\; j = l+1, \ldots, n. \tag{26}$$

As in the previous section, we partition $\boldsymbol{z}^\top$ as $[\boldsymbol{z}_1^\top \;\; \boldsymbol{z}_2^\top]$, where $\boldsymbol{z}_1$ contains the first l elements. We define $\boldsymbol{D}_{12}$ and $\boldsymbol{D}_{22}$ as submatrices of $\boldsymbol{D}$, the former with rows 1 through l and columns $l+1$ through n, and the latter with rows and columns $l+1$ through n. Then, from (14),

$$\boldsymbol{\xi}^\top\boldsymbol{A}\boldsymbol{\xi} = \boldsymbol{z}_1^\top(\boldsymbol{z}_1 - \boldsymbol{D}_{12}\boldsymbol{z}_2) + \boldsymbol{z}_2^\top(\mathbf{I} - \boldsymbol{D}_{22})\boldsymbol{z}_2. \tag{27}$$

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 $\boldsymbol{d}^\top$ is the top row of the matrix $\boldsymbol{D}_{12}$. The last equality in (28) follows from the fact that $\boldsymbol{u}_1^\top\boldsymbol{A}\boldsymbol{U}$ is just the top row of the matrix on the left-hand side of equation (25).

Consider next the expectation of $\boldsymbol{w}_1^\top\boldsymbol{A}\boldsymbol{\xi}/\boldsymbol{\xi}^\top\boldsymbol{A}\boldsymbol{\xi}$ conditional on $\boldsymbol{z}_2$. Since the elements of $\boldsymbol{z}$ are mutually independent, this conditional expectation, should it exist, can 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 element, z_1, unchanged and rotates the remaining $l-1$ elements in such a way that, for the value of $\boldsymbol{z}_2$ on which we are conditioning, all elements 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

$$r^2 - \delta_1 r\cos\theta_1 - \delta_2 r\sin\theta_1\cos\theta_2 - \boldsymbol{z}_2^\top(\boldsymbol{D}_{22} - \mathbf{I})\boldsymbol{z}_2.$$

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 $\boldsymbol{z}_2^\top(\boldsymbol{D}_{22}-\mathbf{I})\boldsymbol{z}_2$, and it is indeed positive, since the matrix in its definition is positive definite. This can be seen by noting from (26) that element ij of the matrix $\boldsymbol{D}-\mathbf{I}$ is

$$\sum_{t=1}^{n}\frac{u_{ti}u_{tj}}{1-h_t}-\delta_{ij}=\sum_{t=1}^{n}u_{ti}u_{tj}\left(\frac{1}{1-h_t}-1\right)=\sum_{t=1}^{n}u_{ti}u_{tj}\frac{h_t}{1-h_t},$$

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 hence also the lower right-hand block, $\boldsymbol{D}_{22}-\mathbf{I}$, is positive definite as well.

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 may appear that this conclusion is not true if $\rho=0$; see (8). But it is not hard to see that the *unconditional* expectation of the JIVE estimator fails to exist even though the conditional expectation vanishes in this special case.

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 substantial simplification follows from the fact that the matrix $\boldsymbol{A}$ becomes $\boldsymbol{P_W}-\big(l/(n-l)\big)\boldsymbol{M_W}$, as can be seen from (12) 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 (29), 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:

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{x}\beta + \boldsymbol{W}_2\gamma + \sigma_u\boldsymbol{u}, \\ \boldsymbol{x} &= \sigma_v(\boldsymbol{W}_1\boldsymbol{a}_1 + \boldsymbol{W}_2\boldsymbol{a}_2 + \boldsymbol{v}). \end{aligned} \tag{30}$$

Here $\boldsymbol{W}_2$ has l' columns, and the full set of instruments is contained in the matrix $\boldsymbol{W} = [\boldsymbol{W}_1 \quad \boldsymbol{W}_2]$. Without loss of generality, we again assume that $\boldsymbol{W}^\top\boldsymbol{W} = \mathbf{I}$ and 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 (30), be expressed as solutions to the estimating equations (4), but with different $\boldsymbol{A}$ matrices. In fact, $\boldsymbol{A} = \boldsymbol{M}_2\boldsymbol{A}_0$, $\boldsymbol{A}_0$ being the matrix $\boldsymbol{A}$ for the original model (1) and $\boldsymbol{M}_2$ being the matrix that projects orthogonally on to $\mathcal{S}^\perp(\boldsymbol{W}_2)$. Since $\boldsymbol{A}_0\boldsymbol{W} = \boldsymbol{W}$, it follows that $\boldsymbol{A}\boldsymbol{W}_1 = \boldsymbol{W}_1$ and $\boldsymbol{A}\boldsymbol{W}_2 = \boldsymbol{A}^\top\boldsymbol{W}_2 = \mathbf{O}$. These conditions are enough for equations (5), (6), and (7) still to be 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 (9).

As in (14) and (15), $\boldsymbol{\xi}^\top\boldsymbol{A}\boldsymbol{\xi} = \boldsymbol{z}^\top\boldsymbol{U}^\top\boldsymbol{A}\boldsymbol{U}\boldsymbol{z}$, and $\boldsymbol{w}_1^\top\boldsymbol{A}\boldsymbol{\xi} = \boldsymbol{u}_1^\top\boldsymbol{A}\boldsymbol{U}\boldsymbol{z}$. We may write $\boldsymbol{U}$ in partitioned form as $[\boldsymbol{W}_1 \quad \boldsymbol{W}_2 \quad \boldsymbol{Z}]$. But since $\boldsymbol{A}\boldsymbol{W}_2 = \boldsymbol{W}_2^\top\boldsymbol{A} = \mathbf{O}$, if we define the $n \times (n-l')$ matrix $\boldsymbol{V}$ as $[\boldsymbol{W}_1 \quad \boldsymbol{Z}]$, we see that

$$\boldsymbol{z}^\top\boldsymbol{U}^\top\boldsymbol{A}\boldsymbol{U}\boldsymbol{z} = \boldsymbol{z}_\perp^\top\boldsymbol{V}^\top\boldsymbol{A}\boldsymbol{V}\boldsymbol{z}_\perp \quad \text{and} \quad \boldsymbol{u}_1^\top\boldsymbol{A}\boldsymbol{U}\boldsymbol{z} = \boldsymbol{u}_1^\top\boldsymbol{A}\boldsymbol{V}\boldsymbol{z}_\perp,$$

where $\boldsymbol{z}_\perp \equiv \boldsymbol{V}^\top\boldsymbol{\xi}$ is an $(n-l')$–vector of mutually independent elements, each distributed as N(0,1), except the first, which is distributed as $\mathrm{N}(a,1)$. This means that $\boldsymbol{z}_\perp$ is distributed just like the n–vector $\boldsymbol{z}$ of the previous two sections, except that the dimension n is replaced by $n-l'$. Thus we need to show that, for each of the estimators we consider, the matrix $\boldsymbol{V}^\top\boldsymbol{A}\boldsymbol{V}$ has exactly the same form as the matrix $\boldsymbol{U}^\top\boldsymbol{A}\boldsymbol{U}$ for the corresponding estimator without exogenous explanatory variables in the structural equation, but with n replaced by $n-l'$ and $\boldsymbol{W}$ by $\boldsymbol{W}_1$. If we can do that, then all of the results on existence of moments are unchanged, except for replacing n by $n-l'$.

Consider first the generalized IV estimator. It is easy to show that the estimating equation for $\hat\beta_{\text{IV}}$ is

$$\boldsymbol{x}^\top\boldsymbol{P}_1(\boldsymbol{y} - \boldsymbol{x}\hat\beta_{\text{IV}}) = \boldsymbol{x}^\top\boldsymbol{P}_{\boldsymbol{W}}\boldsymbol{M}_2(\boldsymbol{y} - \boldsymbol{x}\hat\beta_{\text{IV}}) = 0.$$

We see from (3) that $\boldsymbol{A} = \boldsymbol{M}_2\boldsymbol{A}_0$, as required. Moreover,

$$\boldsymbol{V}^\top\boldsymbol{A}\boldsymbol{V} = \begin{bmatrix} \mathbf{I}_l & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix},$$

which is the same as the $\boldsymbol{U}^\top\boldsymbol{A}\boldsymbol{U}$ of (20). Thus the presence of $\boldsymbol{W}_2$ does not change our analysis of the IV estimator. Moments of order m exist only for $m < l$.

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

$$\boldsymbol{x}^\top(\boldsymbol{M}_2 - K\boldsymbol{M}_{\boldsymbol{W}})(\boldsymbol{y} - \boldsymbol{x}\hat{\beta}) = \boldsymbol{x}^\top(\mathbf{I} - K\boldsymbol{M}_{\boldsymbol{W}})\boldsymbol{M}_2(\boldsymbol{y} - \boldsymbol{x}\hat{\beta}) = 0,$$

so that, once again, $\boldsymbol{A} = \boldsymbol{M}_2\boldsymbol{A}_0$. It is also immediate that $\boldsymbol{V}^\top\boldsymbol{A}\boldsymbol{V}$ is given by the right-hand side of (23), with n replaced by $n - l'$. We conclude 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. The estimating equation (4) for $\hat{\beta}_{\rm JIV}$ becomes

$$\tilde{\boldsymbol{x}}^\top\boldsymbol{M}_2(\boldsymbol{y} - \boldsymbol{x}\hat{\beta}_{\rm JIV}) = \boldsymbol{x}^\top\boldsymbol{A}_0^\top\boldsymbol{M}_2(\boldsymbol{y} - \boldsymbol{x}\hat{\beta}_{\rm JIV}) = 0,$$

where $\boldsymbol{A}_0$ is the $\boldsymbol{A}$ used in (13). Thus $\boldsymbol{A} = \boldsymbol{M}_2\boldsymbol{A}_0$, as required. Just as for (25), we see that $\boldsymbol{V}^\top\boldsymbol{A}\boldsymbol{V} = \mathbf{I} - \boldsymbol{D}$, except that the dimension of the square matrices $\mathbf{I}$ and $\boldsymbol{D}$ is $n - l'$ rather than n. Thus the analysis based on (27) and (28), leading to the conclusion that the JIVE estimator has no moments, 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 when, 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 variable. 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 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 and K–class estimators with fixed $K < 1$, we obtain standard results. 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**, 827–833.

Forchini, G., and G. Hillier (2003). "Conditional inference for possibly unidentified structural equations," *Econometric Theory*, **19**, 707–743.

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.

Hillier, G. (2006). "Yet more on the exact properties of IV estimators," *Econometric Theory*, **22**, 913–931.

James, A. (1954). "Normal multivariate analysis and the orthogonal group," *Annals of Mathematical Statistics*, **25**, 40–75.

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.

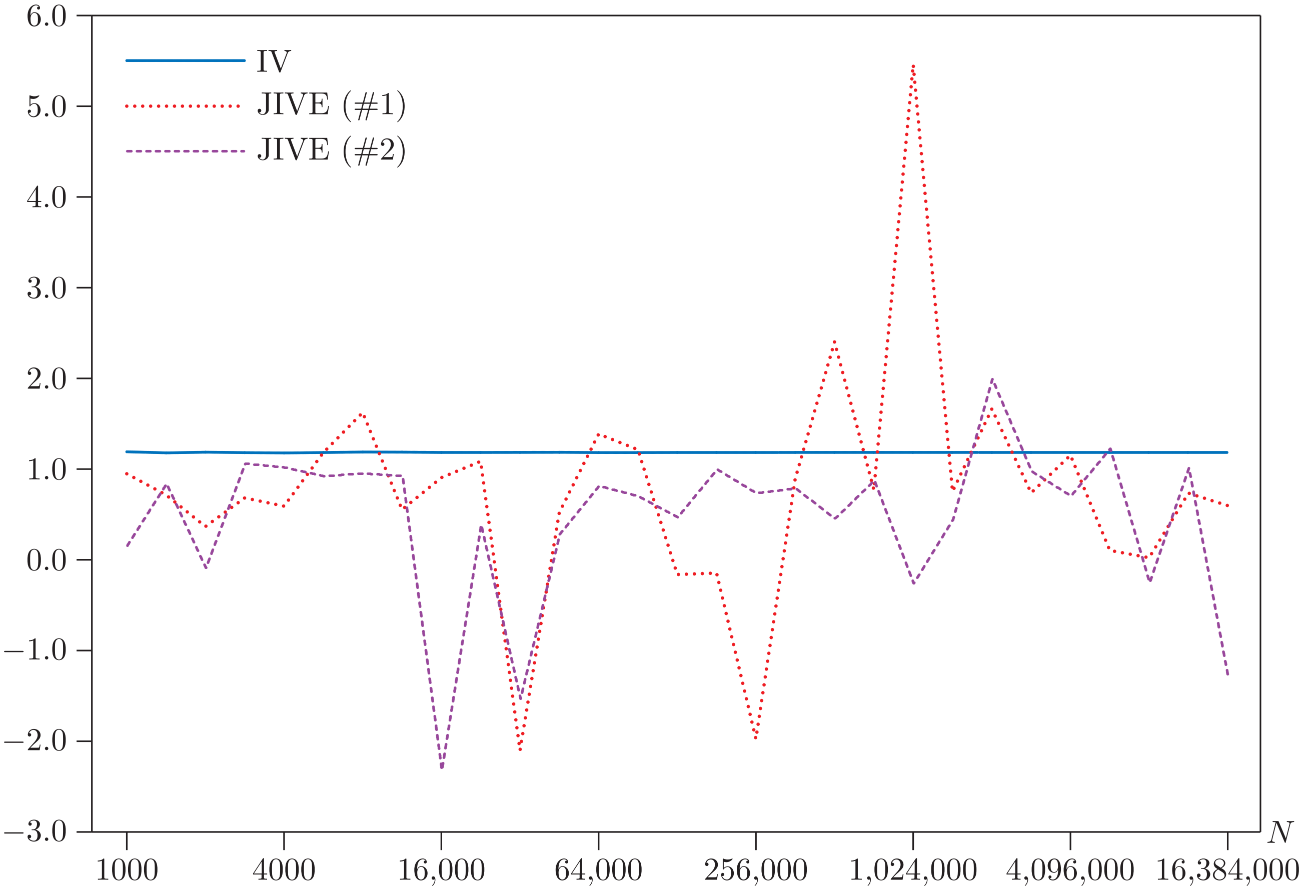

Figure 1. Means of two estimates of β: $n = 50$, $a = 0.5$

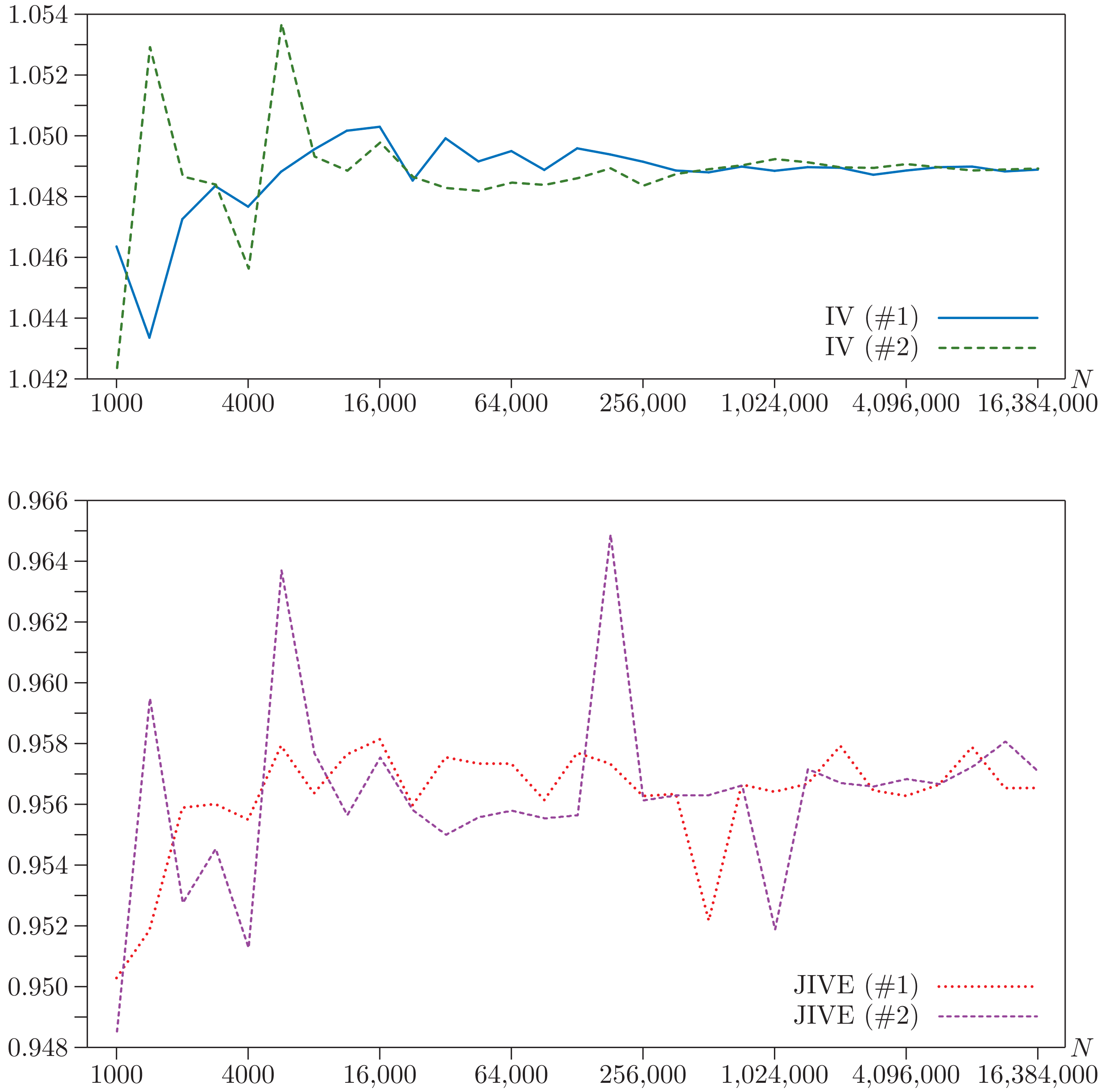

Figure 2. Means of two estimates of β: n = 50, a = 1.0