Econometrica, Vol. 70, No. 5 (September, 2002), 1781–1803

PIVOTAL STATISTICS FOR TESTING STRUCTURAL PARAMETERS IN INSTRUMENTAL VARIABLES REGRESSION

BY FRANK KLEIBERGEN[1]

We propose a novel statistic for conducting joint tests on all the structural parameters in instrumental variables regression. The statistic is straightforward to compute and equals a quadratic form of the score of the concentrated log-likelihood. It therefore attains its minimal value equal to zero at the maximum likelihood estimator. The statistic has a χ^2 limiting distribution with a degrees of freedom parameter equal to the number of structural parameters. The limiting distribution does not depend on nuisance parameters. The statistic overcomes the deficiencies of the Anderson-Rubin statistic, whose limiting distribution has a degrees of freedom parameter equal to the number of instruments, and the likelihood based, Wald, likelihood ratio, and Lagrange multiplier statistics, whose limiting distributions depend on nuisance parameters. Size and power comparisons reveal that the statistic is a (asymptotic) size-corrected likelihood ratio statistic. We apply the statistic to the Angrist-Krueger (1991) data and find similar results as in Staiger and Stock (1997).

KEYWORDS: Instrumental variables regression, weak instruments, pivotal statistics.

1. INTRODUCTION

STATISTICAL INFERENCE IN INSTRUMENTAL VARIABLES (IV) regression models crucially depends on the quality of the instruments. When the quality of the instruments is poor, the limiting distributions of the likelihood based statistics, Wald, likelihood ratio, and Lagrange multiplier, that test joint hypotheses on all structural parameters of the IV regression model, depend on nuisance parameters; see e.g. Bekker (1994), Dufour (1997), Phillips (1989), Staiger and Stock (1997), and Wang and Zivot (1998). Bad instruments therefore not only lead to imprecise estimates of the structural parameters but also imply that the standard statistics, that we use to assess these estimates, are unreliable. A statistic whose limiting distribution is robust to instrument quality is the Anderson-Rubin (AR) statistic; see Anderson and Rubin (1949). A deficiency of the AR statistic is, however, that the degrees of freedom parameter of its limiting distribution is equal to the number of instruments. Hence, when the number of instruments strongly exceeds the number of structural parameters, which commonly occurs in applied work (see e.g. Angrist and Krueger (1991)), the AR statistic has low power.

[1] I thank Paul Bekker, Peter Boswijk, Jean-Marie Dufour, Lynda Khalaf, Geert Ridder, Jim Stock, Eric Zivot, participants of the Weak Instruments session of the Eighth World Conference of the Econometric Society, the editor and especially the referees for many helpful comments and suggestions that significantly improved the manuscript. The research documented in this article has been funded by the NWO Vernieuwingsimpuls research grant "Empirical Comparison of Economic Models."

We propose a statistic, which we refer to as the K-statistic, that remedies the drawback of the AR statistic without the loss of the nuisance parameter-free limiting distribution. Instead of projecting the disturbances of the structural equation on all instruments, as in the AR statistic, the K-statistic projects these disturbances on an IV estimate of the endogenous variables that belong to the structural parameters. Under the hypothesis of interest, this estimate is asymptotically independent of the disturbances of the structural equation. Irrespective of the quality of the instruments, the K-statistic has therefore a χ^2 limiting distribution with a degrees of freedom parameter that is equal to the total number of structural parameters. This limiting distribution is robust to various kinds of asymptotic theory that we use to analyze it. Hence, the same limiting distribution is obtained under regular asymptotics and the weak instrument asymptotics that is used in the literature to analyze the sensitivity to the specification of the instruments; see e.g. Staiger and Stock (1997). The limiting distributions of the likelihood-based statistics depend on nuisance parameters when we analyze them using the latter kind of asymptotic theory.

The K-statistic is a quadratic form of the score of the concentrated log-likelihood. The limited information maximum likelihood (LIML) estimator sets the score to zero. The minimal value of the K-statistic is therefore attained at the LIML estimate and equals zero. Hence, confidence sets that result from the K-statistic are never empty and contain the LIML estimate at every significance level. The K-statistic thus leads to a more robust kind of statistical inference in IV regression models that is centered around the LIML estimator.

Sensitivity to instrument quality is not only a property of the limiting distributions of likelihood based statistics but also of their finite sample distributions. This is revealed by a comparison of the sizes of these statistics, and others, at their asymptotic critical values in finite samples. For the data generating processes that we consider, the likelihood ratio and 2SLS t-statistic become size distorted when the number of instruments increases and/or the instrument quality is poor. The AR and K-statistics do not have these size distortions. The size comparison allows us to construct size-corrected critical values that we use in a power comparison of the different statistics. The power comparison reveals that we have (implicitly) proposed a way of size-correcting the likelihood ratio statistic. This results as the power curves of the size-corrected likelihood ratio statistic and the K-statistic are almost identical. We therefore consider the K-statistic as a size-corrected likelihood ratio statistic.

The paper is organized as follows. Section 2 contains a brief introduction of the IV regression model. In Section 3, we propose the K-statistic and show that its limiting distribution is robust to various kinds of asymptotic theory. We reveal that the K-statistic is a quadratic form of the score of the concentrated likelihood so it attains its minimal value at the LIML estimator. Section 4 mentions the different shapes of the confidence sets that can result when we use the K-statistic to construct them. Section 5 discusses the size and power comparison. Section 6 contains an application to the Angrist-Krueger (1991) data. Finally, the seventh section concludes.

We use the following notation throughout the paper: vec(A) stands for the (column) vectorization of the $T \times n$ matrix A, vec$(A) = (a_1' \dots a_n')'$, when $A = (a_1 \dots a_n)$. $P_A = A(A'A)^{-1}A'$ is a projection on the columns of the full rank matrix A and $M_A = I_T - P_A$ is a projection on the space orthogonal to A. Convergence in probability is denoted by "$\underset{P}{\rightarrow}$" and convergence in distribution by "$\underset{d}{\rightarrow}$".

2. INSTRUMENTAL VARIABLES REGRESSION MODEL

The IV regression model in structural form can be represented as a limited information simultaneous equation model (see e.g. Hausman (1983)),

$$
\begin{aligned}
y &= Y\beta + Z\gamma + \varepsilon, \\
Y &= X\Pi + Z\Gamma + V,
\end{aligned} \tag{1}
$$

where y and Y are a $T \times 1$ and $T \times m$ matrix of endogenous variables, respectively, Z is a $T \times k_Z$ matrix of included exogenous variables, X is a $T \times k_X$ matrix of excluded exogenous variables (or instruments), ε is a $T \times 1$ vector of structural errors, and V is a $T \times m$ matrix of reduced form errors. The $m \times 1$ and $k_Z \times 1$ parameter vectors β and γ contain the structural parameters. The $k_Z \times m$ and $k_X \times m$ parameter matrices Γ and Π contain the parameters of the second set of equations, which are in reduced form. The matrix $(X\ Z)$ is assumed to be of full column rank. In order to be able to construct limiting distributions, we make assumptions with respect to the disturbances, ε and V, and the exogenous variables, X and Z; see e.g. Staiger and Stock (1997).

ASSUMPTION 1: *When the sample size T converges to infinity, the following convergence results hold jointly*:

a. $(1/T)\,(\varepsilon\ V)'\,(\varepsilon\ V) \underset{P}{\rightarrow} \Sigma$, *with Σ a positive definite $(m+1) \times (m+1)$ matrix and*

$$
\Sigma = \begin{pmatrix} \sigma_{\varepsilon\varepsilon} & \sigma_{\varepsilon V} \\ \sigma_{V\varepsilon} & \Sigma_{VV} \end{pmatrix}, \quad \sigma_{\varepsilon\varepsilon} : 1 \times 1, \quad \sigma_{\varepsilon V} = \sigma_{V\varepsilon}' : 1 \times m, \quad \Sigma_{VV} : m \times m.
$$

b. $(1/T)\,(Z\ X)'\,(Z\ X) \underset{P}{\rightarrow} Q$, *with Q a positive definite $k \times k$ matrix*,

$$
Q = \begin{pmatrix} Q_{ZZ} & Q_{ZX} \\ Q_{XZ} & Q_{XX} \end{pmatrix},
$$

$$
Q_{ZZ} : k_Z \times k_Z, \quad Q_{ZX} = Q_{XZ}' : k_Z \times k_X, \quad Q_{XX} : k_X \times k_X.
$$

c. $$\frac{1}{\sqrt{T}}\,(Z\ X)'\,(\varepsilon\ V) \underset{d}{\rightarrow} \begin{pmatrix} \psi_{Z\varepsilon} & \psi_{ZV} \\ \psi_{X\varepsilon} & \psi_{XV} \end{pmatrix}, \quad \textit{with}$$

$$
\psi_{Z\varepsilon} : k_Z \times 1, \quad \psi_{X\varepsilon} : k_X \times 1, \quad \psi_{ZV} : k_Z \times m, \quad \psi_{XV} : k_X \times m,
$$

and

$$\operatorname{vec}\begin{pmatrix} \psi_{Z\varepsilon} & \psi_{ZV} \\ \psi_{X\varepsilon} & \psi_{XV} \end{pmatrix} \sim N(0, \Sigma \otimes Q).$$

We can adapt the moment matrices in Assumption 1 to account for serially correlated errors. Assumption 1 can also straightforwardly be extended to allow for deterministically trending variables. For expository purposes, we refrain from these extensions. Assumption 1(c) implies that the variables in X and Z are uncorrelated with ε and V, and are therefore weakly exogenous for β and Π; see Engle, Hendry, and Richard (1983).

We assume that k_X is larger than or equal to m such that the structural parameter vector β is "apparently" identified by the order condition. We call the model just-identified when k_X is equal to m and the model over-identified when k_X exceeds m. The degree of over-identification is then equal to $k_X - m$. The parameter β is identified if and only if $\operatorname{rank}(\Pi) = m$.

We focus on the parameter β in our analysis. We can therefore simplify the presentation of the results without changing their implications by setting $\gamma = 0$ and $\Gamma = 0$ such that Z drops out of the model. In what follows, we let $k = k_X$ denote the number of instruments. The analytical results for β in this simplified case carry over to the more general case where $\gamma \neq 0$ and $\Gamma \neq 0$ by interpreting all data matrices as residuals from the projection on Z.

3. K-STATISTIC FOR TESTING THE STRUCTURAL PARAMETERS

3.1. *The Asymptotically Pivotal AR and K-statistics*

A well-known statistic for testing a hypothesis $\mathrm{H}_0 : \beta = \beta_0$, that is specified on all the structural parameters of the endogenous variables, is the Anderson-Rubin (AR) statistic (see Anderson and Rubin (1949)),

$$\mathrm{AR}(\beta_0) = \frac{\frac{1}{k}(y - Y\beta_0)' P_X (y - Y\beta_0)}{\frac{1}{T-k}(y - Y\beta_0)' M_X (y - Y\beta_0)}. \tag{2}$$

The AR statistic converges, under H_0 and Assumptions 1(a)–(c), in distribution to $1/k$ times a $\chi^2(k)$ distributed random variable. This limiting distribution does not depend on nuisance parameters so the AR statistic is asymptotically pivotal. Under independently and identically distributed normal errors, the AR statistic is pivotal and has an exact $F(k, T-k)$ distribution. A deficiency of the AR statistic is that the degrees of freedom parameter of its limiting distribution is equal to the number of instruments, k. This number is larger than or equal to the number of elements of β_0, m. The AR statistic has therefore low power when the degree of over-identification is large. We propose a statistic, which we refer to as the K-statistic, that overcomes this deficiency,

$$K(\beta_0) = \frac{(y - Y\beta_0)' P_{\tilde{Y}(\beta_0)} (y - Y\beta_0)}{\frac{1}{T-k}(y - Y\beta_0)' M_X (y - Y\beta_0)}, \tag{3}$$

where

$$\widetilde{Y}(\beta_0) = X\widetilde{\Pi}(\beta_0),$$
$$\widetilde{\Pi}(\beta_0) = (X'X)^{-1}X'\left[Y - (y - Y\beta_0)\frac{s_{\varepsilon V}(\beta_0)}{s_{\varepsilon\varepsilon}(\beta_0)}\right],$$
$$s_{\varepsilon\varepsilon}(\beta_0) = \frac{1}{T-k}(y - Y\beta_0)'M_X(y - Y\beta_0),$$
$$s_{\varepsilon V}(\beta_0) = \frac{1}{T-k}(y - Y\beta_0)'M_X Y.$$

Instead of projecting $(y - Y\beta_0)$ on the k columns of X, such as in the AR statistic, the K-statistic projects $(y - Y\beta_0)$ on the m columns of $X\widetilde{\Pi}(\beta_0)$. Under $\mathrm{H}_0 : \beta = \beta_0$, $\widetilde{\Pi}(\beta_0)$ is both a consistent estimator of Π and asymptotically independent of $(y - Y\beta_0)'X$. Because of this asymptotic independence, the K-statistic converges under H_0 to a $\chi^2(m)$ random variable regardless of whether $\widetilde{\Pi}(\beta_0)$ is consistent or $\sqrt{T}\widetilde{\Pi}(\beta_0)$ converges to a random variable. Since $X\widetilde{\Pi}(\beta_0)$ has only m columns, the degrees of freedom parameter of the limiting distribution of the K-statistic is, unlike the degrees of freedom parameter of the limiting distribution of the AR statistic, equal to the number of elements of β_0, m. Theorem 1 states the limiting behavior of the K-statistic for different limiting sequences of the reduced form parameter Π; see e.g. Staiger and Stock (1997).

THEOREM 1: *Under $H_0 : \beta = \beta_0$ and Assumptions* 1(*a*)–(*c*), *the K-statistic* (3) *converges in distribution to a $\chi^2(m)$ random variable when*:

(i) *the instruments are valid such that Π has a fixed full rank value*;

(ii) *the instruments are weak such that $\Pi = \Pi_T = (1/\sqrt{T})C$ with C a fixed full rank $k \times m$ matrix*;

(iii) *the instruments are invalid such that $\Pi = 0$.*

PROOF: See the Appendix.

Theorem 1(i)–(ii) assume fixed full rank values of Π and C respectively. Theorem 1(iii) shows that the same limiting distribution of the K-statistic is obtained for a zero value of Π. The full rank Assumptions in Theorem 1(i)–(ii) are therefore not necessary and it can be shown that the same limiting distribution is obtained for any fixed value of Π and C respectively. For reasons of brevity, we refrain from explicitly showing these results.

The K-statistic differs from the likelihood based statistics, Wald, likelihood ratio, and Lagrange multiplier, to test H_0 that do not have $\chi^2(m)$ limiting distributions under the assumptions in Theorem 1(ii)–(iii); see e.g. Dufour (1997), Phillips (1989), Staiger and Stock (1997), and Wang and Zivot (1998). The assumptions in Theorem 1(ii)–(iii) concern the specification of the parameter Π that is associated with the instruments. The limiting distribution of the K-statistic is robust to the specification of Π. This results because the estimator of Π that is involved in the K-statistic, $\widetilde{\Pi}(\beta_0)$, is such that the K-statistic converges under H_0 to a $\chi^2(m)$ random variable regardless of whether $\widetilde{\Pi}(\beta_0)$ is consistent or $\sqrt{T}\widetilde{\Pi}(\beta_0)$ converges to a random variable. The likelihood based statistics involve estimators of Π, say $\widehat{\Pi}$, that are such that their limiting distributions depend on nuisance parameters when $\sqrt{T}\widehat{\Pi}$ converges to a random variable. This occurs under the assumptions in Theorem 1(ii)–(iii).

Theorem 1(ii) considers a limiting sequence of Π where Π is modelled as local-to-zero. We refer to this limiting sequence as weak instrument asymptotics; see Staiger and Stock (1997). Staiger and Stock (1997) show that weak instrument asymptotics leads to limiting distributions of estimators for β that often provide a more accurate approximation of the finite sample distribution than the limiting distributions that result from the standard asymptotics, which assume a fixed full rank value for Π. Another kind of limiting sequence, to which we refer as many instrument asymptotics, is considered in Bekker (1994). Under the many instrument asymptotics, the number of instruments and observations jointly converge to infinity in such a manner that their ratio, k/T, converges to a constant fraction τ that lies between zero and one. The many instrument asymptotics also maintains a fixed value of $(1/T)\Sigma_{VV}^{-\frac{1}{2}\prime}\Pi'X'X\Pi\Sigma_{VV}^{-\frac{1}{2}}$, which equals the concentration parameter divided by T, that should be considerably large. Bekker (1994) constructs the limiting distributions of estimators of β that result from the many instrument asymptotics. These limiting distributions also often provide a more accurate approximation of the finite sample distribution than the limiting distributions that result from standard asymptotics. Bekker and Kleibergen (2001) obtain the limiting distribution of the K-statistic under the many instrument asymptotics and show that it depends on the value of $(1/T)\Sigma_{VV}^{-\frac{1}{2}\prime}\Pi'X'X\Pi\Sigma_{VV}^{-\frac{1}{2}}$. For considerable values of $(1/T)\Sigma_{VV}^{-\frac{1}{2}\prime}\Pi'X'X\Pi\Sigma_{VV}^{-\frac{1}{2}}$, the K-statistic still converges to a $\chi^2(m)$ random variable but for zero values of $(1/T)\Sigma_{VV}^{-\frac{1}{2}\prime}\Pi'X'X\Pi\Sigma_{VV}^{-\frac{1}{2}}$, $(1-\tau)$ times the K-statistic converges to a $\chi^2(m)$ random variable, where τ is the constant fraction between zero and one to which k/T converges. These two separate cases constitute upper and lower bounds on the limiting distribution of the K-statistic for unknown values of $(1/T)\Sigma_{VV}^{-\frac{1}{2}\prime}\Pi'X'X\Pi\Sigma_{VV}^{-\frac{1}{2}}$ under many instrument asymptotics; see Bekker and Kleibergen (2001).

The K-statistic is identical to the AR statistic in case of just identification when k is equal to m. As with the AR statistic, the K-statistic is invariant to transformations of y and Y that preserve the specification of the null-hypothesis, $H_0: \beta = \beta_0$. For example, when $y^* = ya$ and $Y^* = YD + yb$, with $a: 1\times 1$, $b: 1\times m$, and $D: m\times m$ and nonsingular, and we use y^* and Y^* instead of y and Y to calculate $K(\delta_0)$ with

$$\delta_0 = \frac{a}{1 + bD^{-1}\beta_0} D^{-1}\beta_0,$$

$K(\delta_0)$ is identical to $K(\beta_0)$ that is calculated using y and Y.

3.2. *The K-statistic, Lagrange Multiplier Statistic, and the LIML Estimator*

We analyze the relationship between the K-statistic and the score or Lagrange multiplier (LM) statistic to test $\mathrm{H}_0 : \beta = \beta_0$ that is based on the concentrated log-likelihood of β under independent normal disturbances with a fixed covariance matrix,

$$\log \mathscr{L}(\beta_0|X, Y) \propto -\frac{1}{2}T \log\left(1 + \frac{k}{T-k}\mathrm{AR}(\beta_0)\right); \tag{4}$$

see e.g. Anderson and Rubin (1949) and Hausman (1983). The LM statistic equals the quadratic form of the score, or first order derivative, with respect to the information matrix; see e.g. Engle (1984). To obtain the expression of the LM statistic, we therefore construct the score and the information matrix.

The score of the concentrated log-likelihood reads

$$\begin{aligned} \left.\frac{\partial \log \mathscr{L}(\beta|X, Y)}{\partial \beta'}\right|_{\beta_0} &= -\frac{1}{2}\frac{Tk}{T-k}\left(1+\frac{k}{T-k}\mathrm{AR}(\beta_0)\right)^{-1}\left(\left.\frac{\partial \mathrm{AR}(\beta)}{\partial \beta'}\right|_{\beta_0}\right) \\ &= \frac{1}{\frac{1}{T}(y-Y\beta_0)'(y-Y\beta_0)}(y-Y\beta_0)' \\ &\quad \times P_X\left[Y-(y-Y\beta_0)\frac{s_{\varepsilon V}(\beta_0)}{s_{\varepsilon\varepsilon}(\beta_0)}\right], \end{aligned} \tag{5}$$

where $s_{\varepsilon\varepsilon}(\beta_0)$ and $s_{\varepsilon V}(\beta_0)$ are defined below equation (3), $|_{\beta_0}$ stands for evaluated in β_0, and

$$\begin{aligned} \left.\frac{\partial \mathrm{AR}(\beta)}{\partial \beta'}\right|_{\beta_0} &= -2\frac{T-k}{k}\left(\frac{(y-Y\beta_0)'P_X Y}{(y-Y\beta_0)'M_X(y-Y\beta_0)}\right. \\ &\qquad \left. - \frac{(y-Y\beta_0)'P_X(y-Y\beta_0)(y-Y\beta_0)'M_X Y}{[(y-Y\beta_0)'M_X(y-Y\beta_0)]^2}\right) \\ &= -\frac{2}{k}\frac{1}{s_{\varepsilon\varepsilon}(\beta_0)}(y-Y\beta_0)'P_X\left[Y-(y-Y\beta_0)\frac{s_{\varepsilon V}(\beta_0)}{s_{\varepsilon\varepsilon}(\beta_0)}\right]. \end{aligned} \tag{6}$$

The score (5) shows that the K-statistic is a quadratic form of it. The LM statistic is also a quadratic form of the score but with respect to the information matrix. The expression of the information matrix that would make the LM statistic identical to the K-statistic is[2]

$$\mathcal{I}(\beta_0) = \frac{1}{\frac{1}{T}(y - Y\beta_0)'(y - Y\beta_0)} \left(Y - (y - Y\beta_0) \frac{s_{\varepsilon V}(\beta_0)}{s_{\varepsilon\varepsilon}(\beta_0)} \right)' P_X \left(Y - (y - Y\beta_0) \frac{s_{\varepsilon V}(\beta_0)}{s_{\varepsilon\varepsilon}(\beta_0)} \right). \tag{7}$$

This expression of the information matrix is, however, not identical to the information matrix for β that results from the concentrated log-likelihood; see Zivot, Startz, and Nelson (1998) for an expression of the information matrix that results from the concentrated log-likelihood. It is also not equal to the elements of the inverse of the joint information matrix of $(\beta, \Pi, \sigma_{\varepsilon\varepsilon}, \sigma_{\varepsilon V}, \Sigma_{VV})$ that belong to β. The K-statistic is therefore, just like the LM statistic, a quadratic form of the score but not with respect to the information matrix. This difference explains why the limiting distribution of the K-statistic applies more generally than the limiting distribution of the LM statistic.

Since the K-statistic is a quadratic form of the score of the concentrated log-likelihood, it is equal to zero at the minimum and maximum of the concentrated log-likelihood. The minimum of the concentrated log-likelihood is uniquely attained at the LIML estimator of β. The K-statistic is therefore equal to zero at the LIML estimator. The zero value of the K-statistic that is attained at the maximum of the concentrated log-likelihood leads to the maximal value of the AR statistic. When we apply the K-statistic, we check whether small values of it are caused by the minimum or maximum of the concentrated log-likelihood.

4. CONFIDENCE SETS

Confidence sets for β can be constructed with the K-statistic by usage of sequences of strictly increasing values for each element of β. For example, for the ith element of β, β_i, a sequence could be $\beta_{i,1} < \cdots < \beta_{i,n}$, $i = 1, \ldots, m$. These m sequences specify an m dimensional grid that contains n^m different values of β_0. For each of these n^m different values of β_0, we test the hypothesis $\mathrm{H}_0 : \beta = \beta_0$ with the K-statistic, $K(\beta_0)$. A specific value of β_0 is within the $(1-\alpha)100\%$ asymptotic confidence set when the asymptotic p-value of $K(\beta_0)$ exceeds α. For

[2] We note that this expression of the information matrix leads to an expression of the LM statistic that is equal to

$$\frac{s_{\varepsilon\varepsilon(\beta_0)}}{\frac{1}{T}(y - Y\beta_0)'(y - Y\beta_0)} K(\beta_0).$$

Under $\mathrm{H}_0 : \beta = \beta_0$,

$$\frac{s_{\varepsilon\varepsilon(\beta_0)}}{\frac{1}{T}(y - Y\beta_0)'(y - Y\beta_0)} \underset{p}{\to} 1$$

such that both statistics converge to the same limit. We use this expression of the information matrix because it is more straightforward to interpret.

an appropriate choice of the m sequences, we obtain the $(1-\alpha)100\%$ asymptotic confidence set for β that results from the asymptotically pivotal K-statistic. For the construction of these $(1-\alpha)100\%$ asymptotic confidence sets, we discard values of β_0 with asymptotic p-values that exceed α that are caused by the zero value of the K-statistic that results from the maximum of the concentrated log-likelihood. We achieve this by discarding finite subsets that do not contain the LIML estimator.

Confidence sets that result from asymptotically pivotal statistics are known to have peculiar shapes in some cases that differ from the standard convex asymptotic $(1-\alpha)100\%$ confidence sets; see e.g. Dufour (1997) and Zivot, Startz, and Nelson (1998). As well as finite convex confidence sets, possible shapes that can occur are infinite and nonconvex confidence sets.

The K-statistic attains its minimal value, which equals zero, at the LIML estimate of β. The $(1-\alpha)100\%$ asymptotic confidence sets that are constructed using the K-statistic therefore always contain the LIML estimate of β. The confidence sets that result from the K-statistic can thus not be empty. Under weak instrument asymptotics, the limiting distribution of the LIML estimator is similar to its finite sample distribution; see Staiger and Stock (1997). Since the finite sample distribution of the LIML estimator depends on unknown nuisance parameters (see e.g. Mariano and Sawa (1972) and Kleibergen (2000)), also the limiting distribution of the LIML estimator under the weak instrument asymptotics depends on nuisance parameters. This hampers the construction of valid asymptotic confidence sets based on the LIML estimator. The limiting distribution of the K-statistic is, however, robust to weak instrument asymptotics. The K-statistic therefore offers a convenient approach to construct valid confidence sets that always contain the LIML estimate of β. The AR statistic also attains its minimum at the LIML estimate of β. The minimum of the AR statistic is, however, strictly larger than zero so there is a significance level at which the AR statistic rejects all possible values of β. Confidence sets that result from the AR statistic can therefore be empty and do not always contain the LIML estimate of β; see e.g. Dufour (1997) and Staiger and Stock (1997).

5. SIZE AND POWER COMPARISON

We conduct size and power comparisons of the K-statistic with other statistics that are frequently used in practice. We compare the K-statistic to the AR statistic, the 2SLS t-statistic, the likelihood ratio statistic under independent normal errors with a fixed covariance matrix, a split-sample statistic, and Bekkers LIML statistic; see Bekker (1994). Before we discuss the size and power properties of these statistics, we first briefly mention the split-sample statistic and Bekkers LIML statistic and specify the data generating process (DGP) that we use to analyze size and power.

5.1. *Statistics*

Split-Sample Statistic. Dufour and Jasiak (2001) propose a split-sample statistic that divides the sample into two parts. Part one,

$$
\begin{aligned}
y_1 &= Y_1\beta + \varepsilon_1, \\
Y_1 &= X_1\Pi + V_1,
\end{aligned}
\tag{8}
$$

where $y_1 : T_1 \times 1$, $T_1 < T$, $y = (y_1' \ y_2')'$, $\varepsilon_1 : T_1 \times 1$, $\varepsilon = (\varepsilon_1' \ \varepsilon_2')'$, $Y_1 : T_1 \times m$, $Y = (Y_1' \ Y_2')'$, $V_1 : T_1 \times m$, $V = (V_1' \ V_2')'$, and $X_1 : T_1 \times k$, $X = (X_1' \ X_2')'$; is used to construct the least squares estimator of Π, $\check{\Pi} = (X_1'X_1)^{-1}X_1'Y_1$. Part two,

$$
\begin{aligned}
y_2 &= Y_2\beta + \varepsilon_2, \\
Y_2 &= X_2\Pi + V_2,
\end{aligned}
\tag{9}
$$

is used to test $\mathrm{H}_0 : \beta = \beta_0$. By letting T_1 and T jointly converge to infinity, $\check{\Pi}$ can be shown to be a consistent estimator of Π. Furthermore, when ε_2 is not correlated with V_1, $(\sqrt{T_1})(\check{\Pi} - \Pi)$ is asymptotically independent from $(1/\sqrt{T - T_1})X_2'\varepsilon_2$. The split-sample statistic (see Dufour and Jasiak (2001))

$$
\mathrm{SS}(\beta_0) = \frac{(y_2 - Y_2\beta_0)'P_{X_2\check{\Pi}}(y_2 - Y_2\beta_0)}{\frac{1}{T - T_1 - k}(y_2 - Y_2\beta_0)'M_{X_2\check{\Pi}}(y_2 - Y_2\beta_0)}
\tag{10}
$$

has then, under $\mathrm{H}_0 : \beta = \beta_0$ and Assumptions 1a–c, a $\chi^2(m)$ limiting distribution regardless of the value of Π.

Bekkers LIML Statistic. Bekker (1994) constructs an estimate of the asymptotic variance of the LIML estimator, $\hat{\beta}_{\mathrm{LIML}}$, that is consistent under the many instrument asymptotics. Under the many instrument asymptotics and $\mathrm{H}_0 : \beta = \beta_0$, $\sqrt{T - m}(\hat{\beta}_{\mathrm{LIML}} - \beta_0)$ has a normal limiting distribution with mean zero and an asymptotic variance that is consistently estimated by

$$
\mathrm{avar}(\hat{\beta}_{\mathrm{LIML}}) = \tilde{\sigma}_{\varepsilon\varepsilon}^2 B_{YY}^{-1}(C_{YY} + D_{YY})B_{YY}^{-1}
\tag{11}
$$

(see e.g. Bekker (1994) and Hahn and Inoue (1999)), where

$$
\begin{aligned}
\tilde{\sigma}_{\varepsilon\varepsilon}^2 &= \frac{1}{T - m}(y - Y\hat{\beta}_{\mathrm{LIML}})'(y - Y\hat{\beta}_{\mathrm{LIML}}), \\
\lambda(\hat{\beta}_{\mathrm{LIML}}) &= \frac{(y - Y\hat{\beta}_{\mathrm{LIML}})'P_X(y - Y\hat{\beta}_{\mathrm{LIML}})}{(y - Y\hat{\beta}_{\mathrm{LIML}})'M_X(y - Y\hat{\beta}_{\mathrm{LIML}})},
\end{aligned}
$$

and B_{YY}, C_{YY}, and D_{YY} are $m \times m$ matrices that are specified as

$$
\begin{aligned}
B_{YY} &= \frac{1}{T-m}[Y'P_X Y - \lambda(\hat{\beta}_{\mathrm{LIML}})Y'M_X Y], \\
C_{YY} &= \frac{1}{T-m}\left[Y'P_X Y - \frac{\lambda(\hat{\beta}_{\mathrm{LIML}})}{(y - Y\hat{\beta}_{\mathrm{LIML}})'M_X(y - Y\hat{\beta}_{\mathrm{LIML}})}\right. \\
&\qquad \left. \times Y'M_X(y - Y\hat{\beta}_{\mathrm{LIML}})(y - Y\hat{\beta}_{\mathrm{LIML}})'M_X Y\right], \\
D_{YY} &= \lambda(\hat{\beta}_{\mathrm{LIML}})[C_{YY} - B_{YY}].
\end{aligned} \tag{12}
$$

Bekkers LIML statistic to test the hypothesis $\mathrm{H}_0 : \beta = \beta_0$ reads:

$$
\mathrm{BLIML}(\beta_0) = (\hat{\beta}_{\mathrm{LIML}} - \beta_0)'\left(\frac{\mathrm{avar}(\hat{\beta}_{\mathrm{LIML}})}{T-m}\right)^{-1}(\hat{\beta}_{\mathrm{LIML}} - \beta_0), \tag{13}
$$

and has a standard $\chi^2(m)$ limiting distribution under the many instrument asymptotics.

5.2. *Data Generating Process*

We use the DGP,

$$
\begin{aligned}
y &= Y\beta + \varepsilon \\
Y &= X\pi + V,
\end{aligned} \tag{14}
$$

with $m = 1$, $T = 100$, y, $Y : T \times 1$, $X : T \times k$, $(\varepsilon\ V) \sim N(0, \Sigma \otimes I_T)$; $X \sim N(0, I_k \otimes I_T)$, $\pi : k \times 1$, $\pi = (\pi_1 \cdots \pi_k)'$, $\pi_2 = \cdots = \pi_k = 0$, $\beta = 0$,

$$
\Sigma = \begin{pmatrix} 1 & 0.99 \\ 0.99 & 1 \end{pmatrix};
$$

for values of π_1 equal to 0.1, i.e. a weak instrument, and 1, i.e. a valid instrument, and values of k equal to 1, 5, and 20. The data that we generate from DGP (14) therefore only differ over the value of π_1 and are the same for the different values of k. Hence, for a fixed π_1, we only add superfluous instruments to the model when we increase k and the concentration parameter $\sigma_{VV}^{-1}\pi'X'X\pi$ is kept constant. Also the generated endogenous variables remain the same ones. In this manner, we visualize the robustness of the statistics to adding superfluous instruments. This increase of k is intended to capture the effect of the many instrument asymptotics. The specification of the covariance matrix Σ implies that y and Y are strongly endogenous.

5.3. *Size Comparison*

Tables I and II contain the observed sizes of the statistics at the 10, 5, and 1% asymptotic critical values. The sizes are computed using 10000 simulations from DGP (14). We use three different versions of the split-sample statistic (10), one with $T_1 = 25$ (SS1), one with $T_1 = 50$ (SS2), and one with $T_1 = 75$ (SS3). The other

TABLE I

SIZES (IN PERCENTAGES) OF DIFFERENT STATISTICS THAT TEST $H_0 : \beta = \beta_0$ FOR DGP (12) WITH $\pi_1 = 1$ USING 10, 5, AND 1% ASYMPTOTIC CRITICAL VALUES (ACV)

	$\pi_1 = 1$								
	$k = 1$			$k = 5$			$k = 20$		
ACV	10	5	1	10	5	1	10	5	1
SS1	10.0	5.3	1.1	10.1	5.4	1.2	10.1	5.2	1.2
SS2	11.0	6.0	1.4	10.2	5.8	1.4	10.7	5.8	1.4
SS3	12.5	7.2	2.0	12.3	7.2	1.9	12.0	7.0	1.9
AR	9.8	5.3	1.1	10.1	4.9	0.9	10.1	4.6	1.4
K	9.8	5.3	1.1	9.9	5.4	1.1	9.9	5.3	1.0
LR	10.2	5.6	1.2	11.8	6.8	1.2	20.5	13.7	5.6
BLIML	9.9	5.2	1.6	9.8	5.2	1.6	9.7	5.2	1.6
2SLS	9.9	5.2	1.6	13.5	8.6	3.2	50.7	40.4	23.1

statistics considered are the AR (AR), K (K), likelihood ratio (LR), Bekkers LIML statistic (BLIML), and the 2SLS t-statistic (2SLS). Table I contains the sizes for the case of valid instruments, $\pi_1 = 1$, while the instruments are weak in Table II, $\pi_1 = 0.1$.

Table I shows that the likelihood ratio and 2SLS t-statistic become size distorted when the number of instruments increases even in the case of valid instruments. This occurs because the limiting distributions of these statistics are not robust under the many instrument asymptotics; see Bekker (1994). For all the other statistics, the asymptotic and observed sizes more or less coincide, which is in line with their robustness under the many instrument asymptotics with a considerable value of $(1/T)\sigma_{VV}^{-1}\pi' X' X\pi$.

Table II shows that the likelihood ratio, Bekker's LIML statistic, and the 2SLS t-statistic are size distorted when the instruments are weak. For Bekker's LIML

TABLE II

SIZES (IN PERCENTAGES) OF DIFFERENT STATISTICS THAT TEST $H_0 : \beta = \beta_0$ FOR DGP (12) WITH $\pi_1 = 0.1$ USING 10, 5, AND 1% ASYMPTOTIC CRITICAL VALUES (ACV)

	$\pi_1 = 0.1$								
	$k = 1$			$k = 5$			$k = 20$		
ACV	10	5	1	10	5	1	10	5	1
SS1	10.0	5.3	1.1	10.5	4.5	1.1	10.5	5.2	1.1
SS2	10.9	6.0	1.4	10.9	6.0	1.4	11.4	6.0	1.5
SS3	12.5	7.3	2.0	11.7	6.4	1.6	11.8	6.7	1.6
AR	9.8	5.4	1.0	10.5	4.9	0.9	10.7	5.3	1.2
K	9.8	5.4	1.0	9.9	5.2	1.0	10.5	5.7	1.2
LR	10.2	5.6	1.2	13.0	7.5	2.0	28.0	20.0	12.8
BLIML	19.6	16.3	11.3	19.9	16.5	11.5	21.8	19.1	12.8
2SLS	19.6	16.3	11.3	91.5	89.1	83.6	100	100	100

statistic and the 2SLS t-statistic this even holds in case of just identification when $k = m = 1$ in which case they are numerically identical. This shows that the assumption of a considerable value of $(1/T)\sigma_{VV}^{-1}\pi' X' X\pi$ is important for the limiting distribution of Bekker's LIML statistic. The likelihood ratio statistic becomes size distorted when the number of instruments increases. The size distortion of the 2SLS t-statistic in Tables I and II is so dramatic that we do not use it for the power comparison.

5.4. *Power Comparison*

To conduct a power comparison of the different statistics, we generate 10000 datasets from DGP (14) for various values of β and compute the frequency of rejecting $H_0 : \beta = 0$ using the 5% size-corrected critical value of the statistic under H_0. These size-corrected critical values result from the size comparison that we conducted previously. For the statistics that revealed no size distortion, i.e. the split-sample, AR, and K-statistics, we use the asymptotic critical values.

Figures 1–3 contain the power curves in the case of a valid instrument, $\pi_1 = 1$, and values of k equal to 1, 5, and 20. Figures 4–6 contain the power curves in the case of a weak instrument, $\pi_1 = 0.1$, and values of k equal to 1, 5, and 20.

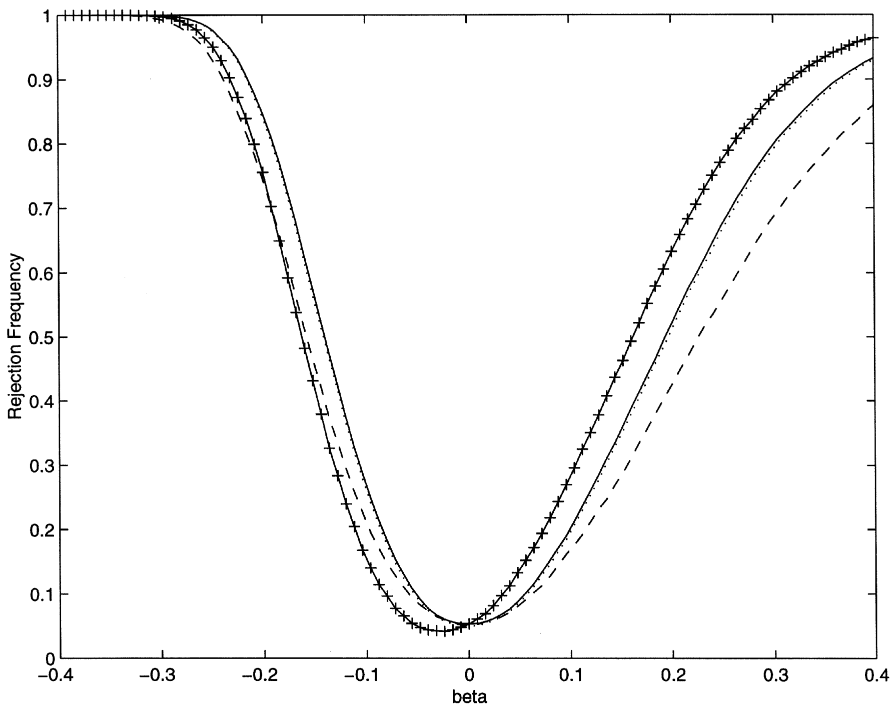

FIGURE 1.— Power curves of K (solid), AR (dotted-dashed), LR (dotted), SS1 (dashed), and BLIML (solid with plusses) statistics that test $H_0 : \beta = 0$ for various values of β in DGP (14) with $\pi_1 = 1$ and $k = 1$.

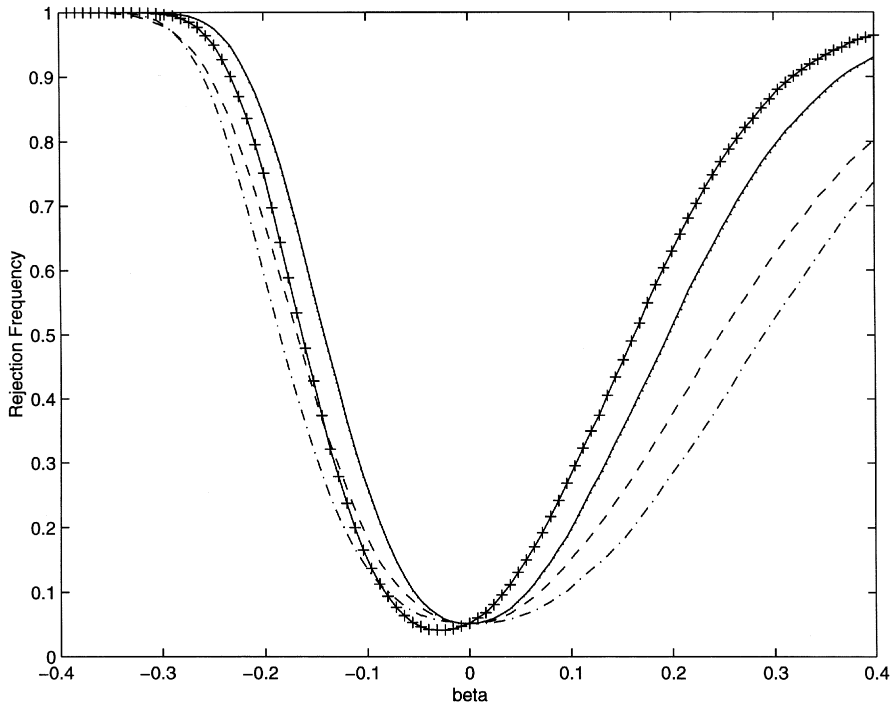

FIGURE 2.— Power curves of K (solid), AR (dotted-dashed), LR (dotted), SS1 (dashed), and BLIML (solid with plusses) statistics that test $H_0 : \beta = 0$ for various values of β in DGP (14) with $\pi_1 = 1$ and $k = 5$.

The figures contain the power curves of the AR, K, BLIML, LR and the most powerful split-sample statistic, which is SS1 in Figures 1, 2, 4, 5, and SS2 in Figures 3 and 6. Because of the enormous size-distortion of the 2SLS t-statistic in Tables I and II, the 2SLS t-statistic is not reported.

We conclude from the figures that: (i) The power curve of the K statistic is nearly identical to the power curve of the LR statistic when we use its size-corrected critical value. (ii) The power of the AR statistic decreases when the number of instruments increases (for $k = 1$, Figures 1 and 4, the AR and K-statistics are identical). (iii) The power of the split-sample statistic is always (considerably) dominated by the power of the K-statistic. (iv) The BLIML statistic has typically good power properties but involves a size-corrected critical value. Because of the close resemblance of the power-curves of the LR, with the size-corrected critical value, and K-statistics, we consider the K-statistic as a (asymptotic) size-corrected LR statistic.

6. APPLICATION TO THE ANGRIST-KRUEGER DATA

Angrist and Krueger (1991) analyze the return of education on earnings. They use quarter of birth or quarter of birth interacted with other (dummy) variables

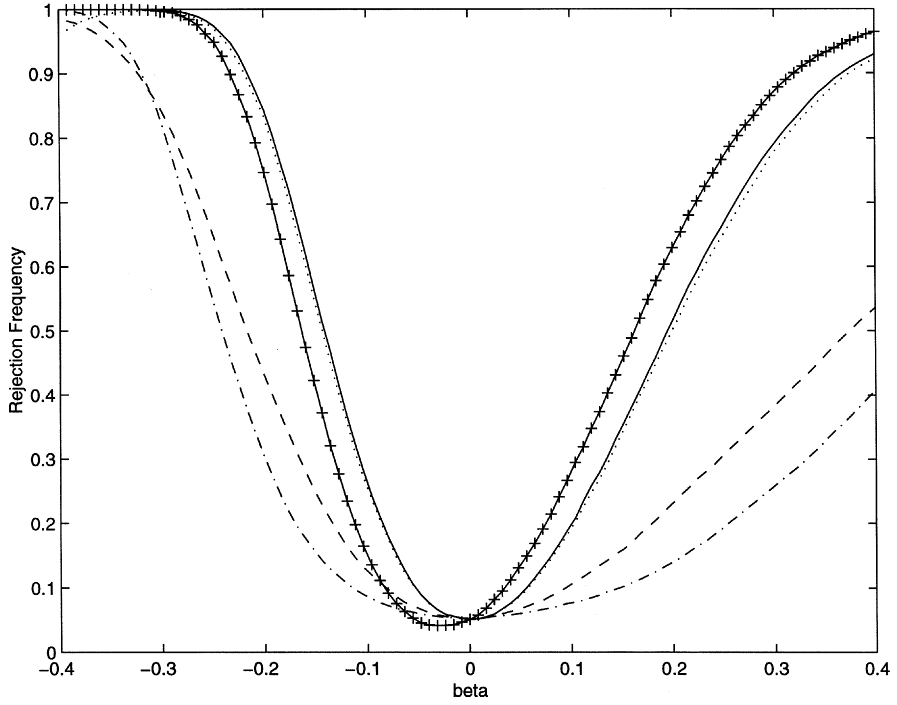

FIGURE 3.— Power curves of K (solid), AR (dotted-dashed), LR (dotted), SS1 (dashed), and BLIML (solid with plusses) statistics that test $H_0 : \beta = 0$ for various values of β in DGP (14) with $\pi_1 = 1$ and $k = 20$.

as instruments in the earnings equation. These quarter of birth related variables can serve as instruments since the quarter of birth is randomly distributed over the population. Because of the age at which a person enters school and state-dependent compulsory school attendance laws, the quarter of birth does, however, affect the educational attainment. We use the "men born between 1930–39" part of the Angrist and Krueger (1991) dataset, which gives us 329.509 observations. This part of the dataset is also analyzed by Staiger and Stock (1997). Our dataset contains five variables, i.e. year of birth, state of birth, quarter of birth, years of education, and log-earnings; we lack observations on the variables: race, standard metropolitan statistical area, region and married, that are also used by Angrist and Krueger (1991) and Staiger and Stock (1997).

The model that is used by Angrist and Krueger (1991) reads:

$$(15)\qquad \begin{aligned} w_i &= c_w + \beta e_i + \gamma_w' z_i \phantom{{}+\pi' x_i} + u_i,\\ e_i &= c_e \phantom{{}+\beta e_i} + \gamma_e' z_i + \pi' x_i + v_i, \end{aligned}$$

where w_i are the log-earnings of individual i, e_i the number of years of education, z_i contains the included exogenous variables, and x_i contains the instruments. The variables u_i and v_i contain the disturbances. We estimated the parameters

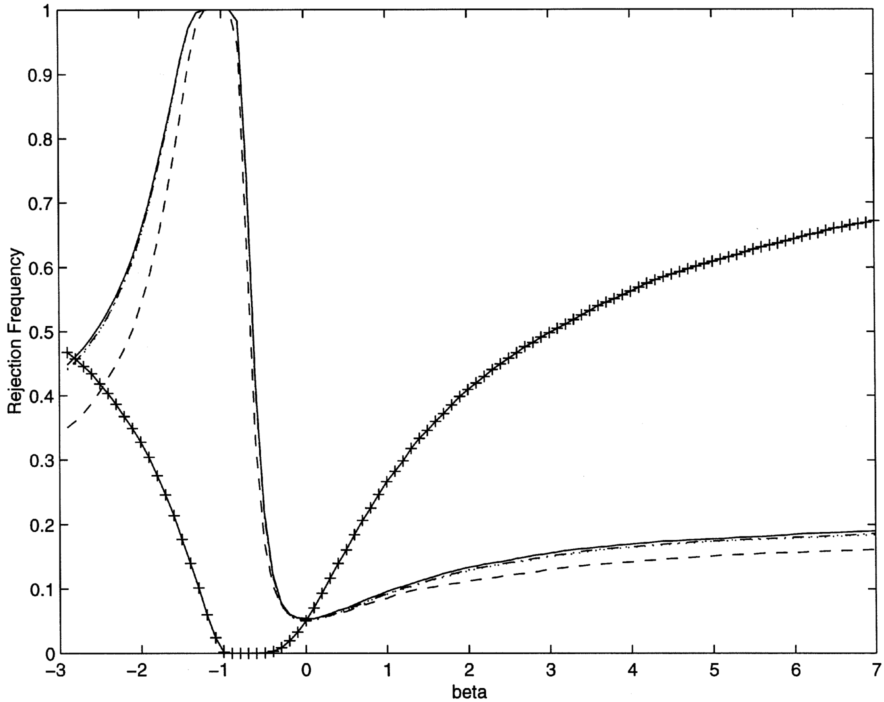

FIGURE 4.— Power curves of K (solid), AR (dotted-dashed), LR (dotted), SS1 (dashed), and BLIML (solid with plusses) statistics that test $H_0 : \beta = 0$ for various values of β in DGP (14) with $\pi_1 = 0.1$ and $k = 1$.

of (15) for a specification where z_i contains year of birth dummies, state of birth dummies, age, age^2, and x_i contains quarter of birth interacted with year of birth dummies and quarter of birth interacted with state of birth dummies. Except for the controls for race, standard metropolitan statistical area, region, and married, this specification is also used by Angrist and Krueger (Table 7) and Staiger and Stock (specification 4).

The estimates of the return on education β that we obtained are (standard errors between parentheses): OLS: 0.0673 (0.0003), 2SLS: 0.0899 (0.0107), LIML: 0.108 (0.0147). We also computed the F-statistic for instrument relevance, which is the standard F-statistic that tests the hypothesis $H_0 : \pi = 0$ in the second equation of (15). This F-statistic is equal to 1.79 with a (asymptotic) p-value of 0.0001. The F-statistic for instrument relevance is significant at the 95% significance level but somewhat small (<2). Hence, the instruments are relatively weak. It implies that we have to interpret the 2SLS and LIML t-statistics with care. We therefore computed asymptotic confidence sets using the K, LR, AR, and 2SLS t-statistics. Figure 7 contains plots of one minus the (asymptotic) p-value of these statistics. Figure 7 also contains a straight line at 0.95 that enables us to straightforwardly construct the 95% asymptotic confidence set, which is the interval between the intersections of the graph of the statistic with the 95% line.

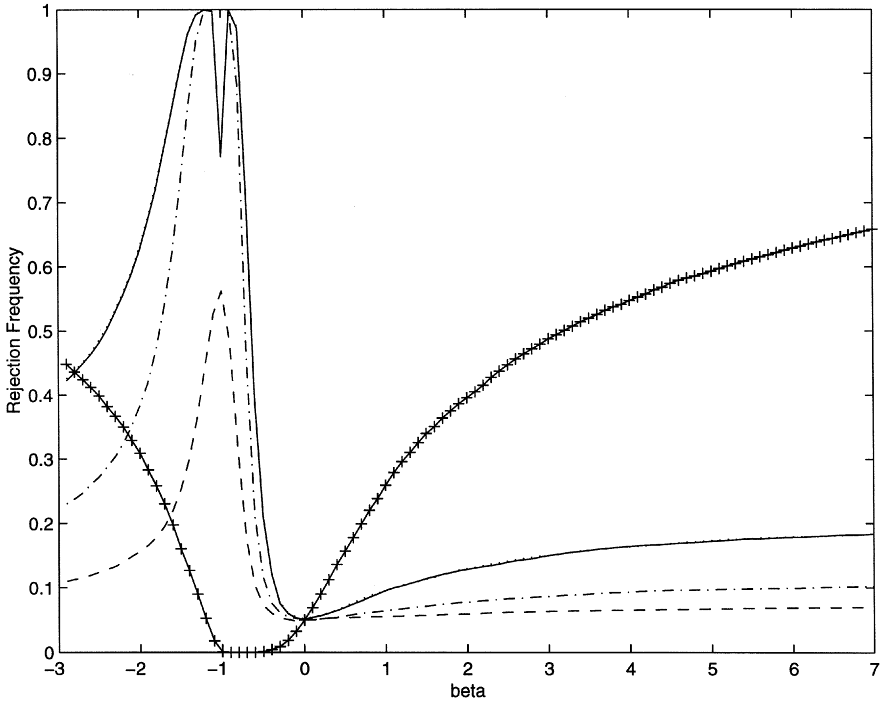

FIGURE 5.— Power curves of K (solid), AR (dotted-dashed), LR (dotted), SS1 (dashed), and BLIML (solid with plusses) statistics that test $H_0 : \beta = 0$ for various values of β in DGP (14) with $\pi_1 = 0.1$ and $k = 5$.

The plot of one minus the p-value of the K-statistic nicely shows that it attains its minimal value, that equals zero, at the LIML estimator. This of course also holds for the likelihood ratio statistic. The plot of one minus the p-value of the 2SLS t-statistic shows that the 2SLS estimator is biased in the direction of the OLS estimator; see e.g. Nelson and Startz (1990). For all values of α, the $(1-\alpha)100\%$ asymptotic confidence sets that result from the K-statistic contain the $(1-\alpha)100\%$ asymptotic confidence sets that result from the likelihood ratio statistic. Figure 7 shows that the AR statistic leads to a much larger 95% confidence set than the K-statistic, which results from the substantial degree of overidentification. The $(1-\alpha)100\%$ asymptotic confidence sets that result from the AR statistic are empty when α is less than 0.62. The asymptotic confidence sets that result from the K-statistic are never empty.

We have used the specification of the instruments and exogenous variables in model (15) that we considered the most plausible one. For some other specifications, that are used by Angrist and Krueger (1991), the quality of the instruments is so poor that the AR and K-statistics lead to infinite 95% confidence sets. Angrist and Krueger use confidence sets that are obtained from the 2SLS t-statistic which are, just like the sets based on the likelihood ratio statistic, always

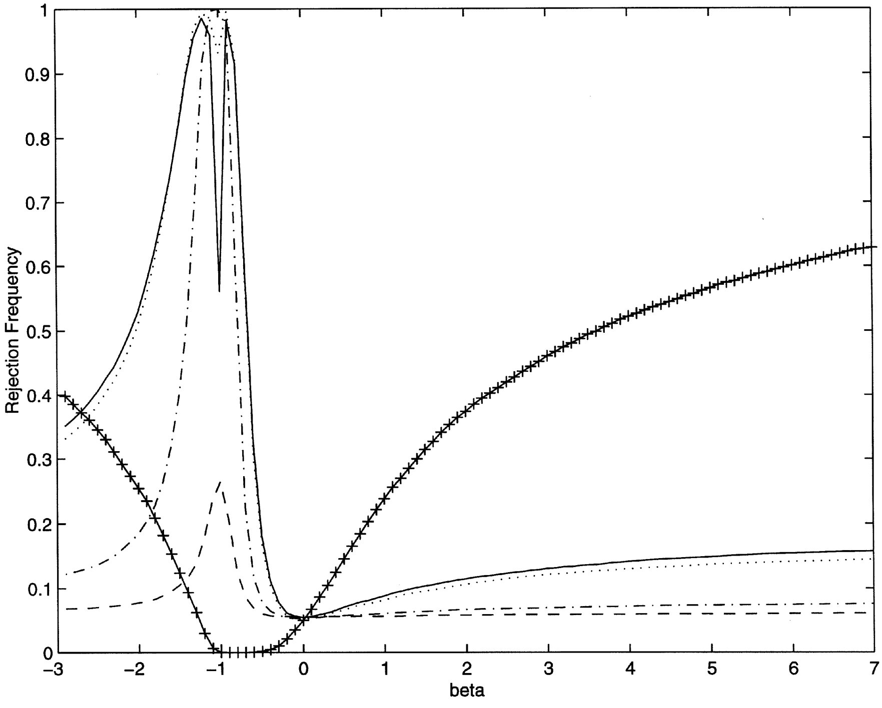

FIGURE 6.— Power curves of K (solid), AR (dotted-dashed), LR (dotted), SS1 (dashed), and BLIML (solid with plusses) statistics that test $H_0 : \beta = 0$ for various values of β in DGP (14) with $\pi_1 = 0.1$ and $k = 20$.

finite. These confidence sets then lead to wrong conclusions. This shows the importance of the use of asymptotically pivotal statistics.

The 95% confidence set that results from the K-statistic in Figure 7 is smaller than the 95% Bonferroni confidence set that is reported by Staiger and Stock (1997). The 95% Bonferroni confidence sets have an asymptotic confidence level of at least 95% so they are larger than or equal to the confidence sets that result from the K-statistic, which have an asymptotic coverage ratio of 95%.

7. CONCLUSIONS

We propose the K-statistic for conducting joint tests on all the structural parameters in IV regression. The K-statistic is asymptotically pivotal and its χ^2 limiting distribution has a degrees of freedom parameter that is equal to the number of structural parameters. The K-statistic is a quadratic form of the score of the concentrated likelihood. Hence, it is equal to zero at the LIML estimator. The K-statistic therefore allows us to conduct valid (asymptotic) statistical inference that is centered around the LIML estimator even in cases when the limiting distribution of the LIML estimator depends on nuisance parameters.

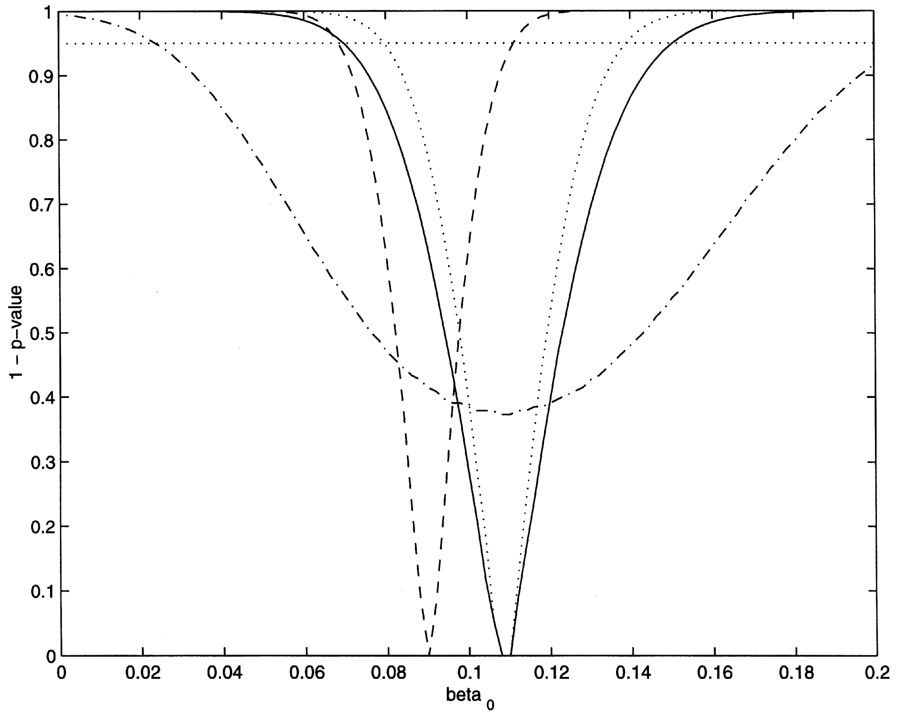

FIGURE 7.— $1-p$-value plots of statistics that test $H_0 : \beta = \beta_0$ using K (solid), AR (dashed-dotted), LR (dotted), and 2SLS t (dashed) statistics for the Angrist-Krueger model.

This manner of conducting statistical inference uses the asymptotically pivotal K-statistic as a base and considers the LIML estimator as the value where the K-statistic attains its zero minimum. It is therefore not concerned about the limiting distribution of the LIML estimator nor its asymptotic variance. Asymptotic standard errors are not involved in the statistical analysis based on the K-statistic.

Unlike the K-statistic, the likelihood ratio statistic is not asymptotically pivotal. Power comparisons of the K-statistic with the likelihood ratio statistic using a size-corrected critical value reveal that the K-statistic is a (asymptotic) size-corrected likelihood ratio statistic.

We applied the K-statistic to the Angrist-Krueger (1991) data for which we obtained similar results as in Staiger and Stock (1997) albeit with a statistic that is less complicated to construct.

Dept. of Quantitative Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands; kleiberg@fee.uva.nl; http://www.fee.uva.nl/ke/kleibergen.htm.

Manuscript received October, 2000; final revision received November, 2001.

APPENDIX

Proof of Theorem 1

Under $\mathrm{H}_0 : \beta = \beta_0$ and Assumptions 1a–c, the joint limiting behavior of $(1/\sqrt{T})X'(y - Y\beta_0)$ and $(1/\sqrt{T})X'(Y - X\Pi)$ is characterized by

$$\left[\frac{1}{\sqrt{T}}X'(y-Y\beta_0) \quad \frac{1}{\sqrt{T}}X'(Y-X\Pi)\right] \underset{d}{\rightarrow} \left[\psi_{X\varepsilon} \quad \psi_{XV}\right], \tag{16}$$

where $(\psi_{X\varepsilon}\ \psi_{XV}) \sim N(0, \Sigma \otimes Q_{XX})$. Post-multiplying (16) by

$$R = \begin{pmatrix} 1 & -\dfrac{\sigma_{\varepsilon V}}{\sigma_{\varepsilon\varepsilon}} \\ 0 & I_m \end{pmatrix} \tag{17}$$

gives

$$\left[\frac{1}{\sqrt{T}}X'(y-Y\beta_0) \quad \frac{1}{\sqrt{T}}X'\left\{(Y-X\Pi)-(y-Y\beta_0)\frac{\sigma_{\varepsilon V}}{\sigma_{\varepsilon\varepsilon}}\right\}\right] \underset{d}{\rightarrow} \left[\psi_{X\varepsilon} \quad \psi_{XU}\right], \tag{18}$$

where

$$U = V - \varepsilon\frac{\sigma_{\varepsilon V}}{\sigma_{\varepsilon\varepsilon}} = (Y - X\Pi) - (y - Y\beta_0)\frac{\sigma_{\varepsilon v}}{\sigma_{\varepsilon\varepsilon}},$$

$$\mathrm{vec}(\psi_{X\varepsilon}\ \psi_{XU}) \sim N\left(0, \begin{pmatrix} \sigma_{\varepsilon\varepsilon} & 0 \\ 0 & \Sigma_{UU}\end{pmatrix} \otimes Q_{XX}\right),$$

and

$$\Sigma_{UU} = \Sigma_{VV} - \frac{\sigma_{V\varepsilon}\sigma_{\varepsilon V}}{\sigma_{\varepsilon\varepsilon}}.$$

Hence, $\psi_{X\varepsilon}$ and ψ_{XU} are independent random variables, which implies that

$$\begin{aligned} &\frac{1}{\sqrt{T}}X'(y-Y\beta_0) && \underset{d}{\rightarrow} \psi_{X\varepsilon}, \\ &\frac{1}{\sqrt{T}}X'\left\{(Y-X\Pi)-(y-Y\beta_0)\frac{\sigma_{\varepsilon V}}{\sigma_{\varepsilon\varepsilon}}\right\} && \underset{d}{\rightarrow} \psi_{XU}. \end{aligned} \tag{19}$$

The limiting expression (19) contains the unobserved parameters $\sigma_{\varepsilon V}$ and $\sigma_{\varepsilon\varepsilon}$. We can replace these parameters by the consistent estimators

$$s_{\varepsilon V} = \frac{1}{T-K}(y-Y\beta_0)'M_X Y \quad \text{and} \quad s_{\varepsilon\varepsilon} = \frac{1}{T-K}(y-Y\beta_0)'M_X(y-Y\beta_0).$$

The consistency of $s_{\varepsilon V}$ and $s_{\varepsilon\varepsilon}$ under H_0 results from Assumptions 1a and 1c. We note that $s_{\varepsilon V}$ is a consistent estimator of $\sigma_{\varepsilon V}$ that does not depend on Π since

$$\frac{1}{T-k}(y-Y\beta_0)'M_X Y = \frac{1}{T-K}(y-Y\beta_0)'M_X(Y-X\Pi).$$

When we replace $\sigma_{\varepsilon V}/\sigma_{\varepsilon\varepsilon}$ by $s_{\varepsilon V}/s_{\varepsilon\varepsilon}$ in the second expression of (19), it becomes

$$
\begin{aligned}
(20)\qquad & \frac{1}{\sqrt{T}}X'\left\{(Y-X\Pi)-(y-Y\beta_0)\frac{s_{\varepsilon V}}{s_{\varepsilon\varepsilon}}\right\}\\
&= \frac{1}{\sqrt{T}}X'\left\{(Y-X\Pi)-(y-Y\beta_0)\frac{\sigma_{\varepsilon V}}{\sigma_{\varepsilon\varepsilon}}\right\}\\
&\quad -\frac{1}{\sqrt{T}}X'(y-Y\beta_0)\left[\frac{s_{\varepsilon V}}{s_{\varepsilon\varepsilon}}-\frac{\sigma_{\varepsilon V}}{\sigma_{\varepsilon\varepsilon}}\right]\\
&= \frac{1}{\sqrt{T}}X'\left\{(Y-X\Pi)-(y-Y\beta_0)\frac{\sigma_{\varepsilon V}}{\sigma_{\varepsilon\varepsilon}}\right\}\\
&\quad -\frac{1}{\sqrt{T}}X'(y-Y\beta_0)\left\{[s_{\varepsilon V}-\sigma_{\varepsilon V}]\frac{1}{\sigma_{\varepsilon\varepsilon}}+s_{\varepsilon V}\left[\frac{1}{s_{\varepsilon\varepsilon}}-\frac{1}{\sigma_{\varepsilon\varepsilon}}\right]\right\}.
\end{aligned}
$$

The variance estimator $s_{\varepsilon V}$ is a ($\sqrt{T}$) consistent estimator of $\sigma_{\varepsilon V}$ and $1/s_{\varepsilon\varepsilon}$ is a ($\sqrt{T}$) consistent estimator of $1/\sigma_{\varepsilon\varepsilon}$. This implies that, because $(1/\sqrt{T})X'(y-Y\beta_0)\underset{d}{\to}\psi_{X\varepsilon}$, the last part of (20) converges in probability to zero,

$$
\frac{1}{\sqrt{T}}X'(y-Y\beta_0)\left\{[s_{\varepsilon V}-\sigma_{\varepsilon V}]\frac{1}{\sigma_{\varepsilon\varepsilon}}+s_{\varepsilon V}\left[\frac{1}{s_{\varepsilon\varepsilon}}-\frac{1}{\sigma_{\varepsilon\varepsilon}}\right]\right\}\underset{p}{\to}0,
$$

and thus

$$
\frac{1}{\sqrt{T}}X'\left\{(Y-X\Pi)-(y-Y\beta_0)\frac{s_{\varepsilon V}}{s_{\varepsilon\varepsilon}}\right\}\quad\text{and}\quad \frac{1}{\sqrt{T}}X'\left\{(Y-X\Pi)-(y-Y\beta_0)\frac{\sigma_{\varepsilon V}}{\sigma_{\varepsilon\varepsilon}}\right\}
$$

have the same limiting behavior,

$$
(21)\qquad
\begin{aligned}
&\frac{1}{\sqrt{T}}X'(y-Y\beta_0) && \underset{d}{\to}\psi_{X\varepsilon},\\
&\frac{1}{\sqrt{T}}X'\left\{(Y-X\Pi)-(y-Y\beta_0)\frac{s_{\varepsilon V}}{s_{\varepsilon\varepsilon}}\right\} && \underset{d}{\to}\psi_{XU},
\end{aligned}
$$

which expression we use to construct the limiting distribution of the K-statistic (3) for the different cases of Theorem 1.

(i) The instruments are valid such that Π has a fixed full rank value. When Π has a fixed full rank value,

$$
(22)\qquad \frac{1}{\sqrt{T}}\left[X'\left\{Y-(y-Y\beta_0)\frac{s_{\varepsilon V}}{s_{\varepsilon\varepsilon}}\right\}-X'X\Pi\right]\underset{d}{\to}\psi_{XU},
$$

and therefore

$$
\begin{aligned}
(23)\qquad & \frac{1}{\sqrt{T}}\left\{Y-(y-Y\beta_0)\frac{s_{\varepsilon V}}{s_{\varepsilon\varepsilon}}\right\}'P_X(y-Y\beta_0)\\
&= \frac{1}{\sqrt{T}}\left\{(Y-X\Pi)-(y-Y\beta_0)\frac{s_{\varepsilon V}}{s_{\varepsilon\varepsilon}}\right\}'P_X(y-Y\beta_0)+\frac{1}{\sqrt{T}}\Pi'X'P_X(y-Y\beta_0)\underset{d}{\to}\Pi'\psi_{X\varepsilon}.
\end{aligned}
$$

Similarly,

$$
(24)\qquad \frac{1}{T}\left\{Y-(y-Y\beta_0)\frac{s_{\varepsilon V}}{s_{\varepsilon\varepsilon}}\right\}'P_X\left\{Y-(y-Y\beta_0)\frac{s_{\varepsilon V}}{S_{\varepsilon\varepsilon}}\right\}\underset{p}{\to}\Pi'Q_{XX}\Pi.
$$

The limiting behavior of $K(\beta_0)$ is then characterized by

$$
(25)\qquad K(\beta_0)\underset{d}{\to}\frac{1}{\sigma_{\varepsilon\varepsilon}}\psi'_{X\varepsilon}\Pi(\Pi'Q_{XX}\Pi)^{-1}\Pi'\psi_{X\varepsilon},
$$

since

$$\frac{1}{T-k}(y_1 - Y_2\beta_0)' M_X (y_1 - Y_2\beta_0) \underset{p}{\rightarrow} \sigma_{\varepsilon\varepsilon},$$

and $(1/\sigma_{\varepsilon\varepsilon})\psi'_{X\varepsilon}\Pi(\Pi' Q_{XX}\Pi)^{-1}\Pi'\psi_{X\varepsilon}$ is a standard $\chi^2(m)$ random variable.

(ii) The instruments are weak such that $\Pi = \Pi_T = (1/\sqrt{T})C$ with C a fixed full rank matrix. When $\Pi = \Pi_T = (1/\sqrt{T})C$ with C a fixed full rank $k \times m$ matrix,

$$\begin{aligned}\frac{1}{\sqrt{T}} X'\left\{Y - (y - Y\beta_0)\frac{s_{\varepsilon V}}{s_{\varepsilon\varepsilon}}\right\} &= \frac{1}{\sqrt{T}} X'\left\{(Y - X\Pi_T) - (y - Y\beta_0)\frac{s_{\varepsilon V}}{s_{\varepsilon\varepsilon}}\right\} + \frac{1}{\sqrt{T}} X' X \Pi_T \\ &\underset{d}{\rightarrow} \psi_{XU} + Q_{XX} C,\end{aligned} \tag{26}$$

such that

$$\left\{Y - (y - Y\beta_0)\frac{s_{\varepsilon V}}{s_{\varepsilon\varepsilon}}\right\}' P_X \left\{Y - (y - Y\beta_0)\frac{s_{\varepsilon V}}{s_{\varepsilon\varepsilon}}\right\} \underset{d}{\rightarrow} (Q_{XX}C + \psi_{XU})' Q_{XX}^{-1}(Q_{XX}C + \psi_{XU}) \tag{27}$$

and

$$\left\{Y - (y - Y\beta_0)\frac{s_{\varepsilon V}}{s_{\varepsilon\varepsilon}}\right\}' P_X (y - Y\beta_0) \underset{d}{\rightarrow} (Q_{XX}C + \psi_{XU})' Q_{XX}^{-1}\psi_{X\varepsilon}. \tag{28}$$

We then obtain the limiting behavior of $K(\beta_0)$ under weak instruments

$$K(\beta_0) \underset{d}{\rightarrow} \frac{1}{\sigma_{\varepsilon\varepsilon}} \psi'_{X\varepsilon} Q_{XX}^{-\frac{1}{2}} P_{Q_{XX}^{-\frac{1}{2}}(Q_{XX}C + \psi_{XU})} Q_{XX}^{-\frac{1}{2}} \psi_{X\varepsilon}, \tag{29}$$

where, because of the independence of ψ_{XU} and $\psi_{X\varepsilon}$,

$$\frac{1}{\sigma_{\varepsilon\varepsilon}} \psi'_{X\varepsilon} Q_{XX}^{-\frac{1}{2}} P_{Q_{XX}^{-\frac{1}{2}}(Q_{XX}C + \psi_{XU})} Q_{XX}^{-\frac{1}{2}} \psi_{X\varepsilon}$$

is a standard $\chi^2(m)$ random variable. The latter holds because the conditional distribution of $(Q_{XX}C + \psi_{XU})' Q_{XX}^{-1}\psi_{X\varepsilon}$ given ψ_{XU} is

$$N(0, \sigma_{\varepsilon\varepsilon}(Q_{XX}C + \psi_{XU})' Q_{XX}^{-1}(Q_{XX}C + \psi_{XU})).$$

Hence,

$$((Q_{XX}C + \psi_{XU})' Q_{XX}^{-1}(Q_{XX}C + \psi_{XU}))^{-\frac{1}{2}}(Q_{XX}C + \psi_{XU})' Q_{XX}^{-1}\psi_{X\varepsilon}$$

given ψ_{XU} is distributed as $N(0, \sigma_{\varepsilon\varepsilon} I_m)$ which does not depend on ψ_{XU} so this random variable is also unconditionally distributed as $N(0, \sigma_{\varepsilon\varepsilon} I_m)$.

(iii) The instruments are invalid such that $\Pi = 0$. When $\Pi = 0$,

$$\frac{1}{\sqrt{T}} X'\left\{Y - (y - Y\beta_0)\frac{s_{\varepsilon V}}{s_{\varepsilon\varepsilon}}\right\} \underset{d}{\rightarrow} \psi_{XU}, \tag{30}$$

such that

$$\left\{Y - (y - Y\beta_0)\frac{s_{\varepsilon V}}{s_{\varepsilon\varepsilon}}\right\}' P_X \left\{Y - (y - Y\beta_0)\frac{s_{\varepsilon V}}{s_{\varepsilon\varepsilon}}\right\} \underset{d}{\rightarrow} \psi'_{XU} Q_{XX}^{-1}\psi_{XU} \tag{31}$$

and

$$\left\{Y - (y - Y\beta_0)\frac{s_{\varepsilon V}}{s_{\varepsilon\varepsilon}}\right\}' P_X (y - Y\beta_0) \underset{d}{\rightarrow} \psi'_{XU} Q_{XX}^{-1}\psi_{X\varepsilon}. \tag{32}$$

Because of the independence of $\psi_{X\varepsilon}$ and ψ_{XU}, the limiting behavior of $K(\beta_0)$ under invalid instruments is characterized by

$$K(\beta_0) \underset{d}{\rightarrow} \frac{1}{\sigma_{\varepsilon\varepsilon}} \psi'_{X\varepsilon} Q_{XX}^{-\frac{1}{2}} P_{Q_{XX}^{-\frac{1}{2}}\psi_{XU}} Q_{XX}^{-\frac{1}{2}} \psi_{X\varepsilon}, \tag{33}$$

which is, for similar reasons as explained previously, a standard $\chi^2(m)$ random variable.

REFERENCES

ANDERSON, T. W., AND H. RUBIN (1949): "Estimators of the Parameters of a Single Equation in a Complete Set of Stochastic Equations," *The Annals of Mathematical Statistics*, 21, 570–582.

ANGRIST, J. D., AND A. B. KRUEGER (1991): "Does Compulsory School Attendence Affect Schooling and Earnings?" *Quarterly Journal of Economics*, 106, 979–1014.

BEKKER, P. A. (1994): "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica*, 62, 557–681.

BEKKER, P. A., AND F. KLEIBERGEN (2001): "Finite-Sample Instrumental Variables Inference using an Asymptotically Pivotal Statistic," Tinbergen Institute Discussion Paper, TI 01-055/4.

DUFOUR, J.-M. (1997): "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365–1388.

DUFOUR, J.-M., AND J. JASIAK (2001): "Finite Sample Inference for Simultaneous Equations and Models with Unobserved and Generated Regressors," *International Economic Review*, 42, 815–844.

ENGLE, R. F. (1984): "Wald, Likelihood Ratio and Lagrange Multiplier Tests in Econometrics," in *Handbook of Econometrics, Volume 2*, ed. by Z. Griliches and M. D. Intrilligator. Amsterdam: Elsevier Science.

ENGLE, R. F., D. F. HENDRY, AND J.-F. RICHARD (1983): "Exogeneity," *Econometrica*, 51, 277–304.

HAHN, J., AND A. INOUE (1999): "A Monte Carlo Comparison of Various Asymptotic Approximations to the Distribution of the Instrumental Variable Estimator," Unpublished Manuscript, North Carolina State University.

HAUSMAN, J. (1983): "Specification and Estimation of Simultaneous Equation Systems," in *Handbook of Econometrics, Volume 1*, ed. by Z. Griliches and M. D. Intrilligator. Amsterdam: Elsevier Science.

KLEIBERGEN, F. (2000): "On the Use of Orthogonal Statistics to Construct Exact Tests and the Density of the LIML Estimator in the IV Regression Model," Tinbergen Institute Discussion Paper, TI 2000-039/4.

MARIANO, R. S., AND T. SAWA (1972): "The Exact Finite-Sample Distribution of the Limited Information Maximum Likelihood Estimator in the Case of Two Included Endogenous Variables," *Journal of the American Statistical Association*, 67, 159–163.

NELSON, C. R., AND R. STARTZ (1990): "Some Further Results on the Exact Small Sample Properties of the Instrumental Variables Estimator," *Econometrica*, 58, 967–976.

PHILLIPS, P. C. B. (1989): "Partially Identified Econometric Models," *Econometric Theory*, 5, 181–240.

STAIGER, D., AND J. H. STOCK (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.

WANG, J., AND E. ZIVOT (1998): "Inference on Structural Parameters in Instrumental Variables Regression with Weak Instruments," *Econometrica*, 66, 1389–1405.

ZIVOT, E., R. STARTZ, AND C. R. NELSON (1998): "Valid Confidence Intervals and Inference in the Presence of Weak Instruments," *International Economic Review*, 39, 1119–1144.