Econometric Theory, **25**, 2009, 793–805. Printed in the United States of America.
doi:10.1017/S0266466608090300

OPTIMAL INVARIANT INFERENCE WHEN THE NUMBER OF INSTRUMENTS IS LARGE

LAURA CHIODA
Princeton University

MICHAEL JANSSON
University of California, Berkeley

This paper studies the asymptotic behavior of a Gaussian linear instrumental variables model in which the number of instruments diverges with the sample size. Asymptotic efficiency bounds are obtained for rotation invariant inference procedures and are shown to be attainable by procedures based on the limited information maximum likelihood estimator. The bounds are obtained by characterizing the limiting experiment associated with the model induced by the rotation invariance restriction.

1. INTRODUCTION

This paper is concerned with asymptotic efficiency in a linear instrumental variable (IV) regression model with Gaussian errors and a "large" number of IVs.[1] A noteworthy feature, which distinguishes models with a large number of IVs from models with a fixed number of (strong) IVs, is that models of the former type exhibit the nonstandard property that the two stage least squares (2SLS) and limited information maximum likelihood (LIML) estimators of structural coefficients are not asymptotically equivalent, even to first order. On the other hand, the fact that the LIML estimator is asymptotically normally distributed (albeit with a nonstandard variance) suggests that the complications faced in models with a large number of IVs are much less severe than those encountered in models with a fixed number of weak IVs.[2]

The overall objective of the present paper is to investigate the extent to which models with a large numbers of IVs are fundamentally nonstandard (i.e., exhibit properties qualitatively similar to the properties of models with a fixed number of weak IVs) and/or amenable to statistical analysis using essentially standard tools (i.e., exhibit properties qualitatively similar to the properties of models with a fixed number of strong IVs). More specifically, the purpose of the paper is

The authors thank Whitney Newey, Jim Powell, a co-editor, and a referee for helpful comments. Address correspondence to Michael Jansson, Department of Economics, UC Berkeley, 549 Evans Hall #3880, Berkeley, CA 94720-3880, USA; e-mail: mjansson@econ.berkeley.edu.

 0266-4666/09 $15.00

twofold. First, the paper seeks to obtain a characterization of the limiting experiment associated with a sequence of models in which the number of IVs tends to infinity with the sample size. Second, the paper will attempt to use such a characterization of the limiting experiment to obtain attainable asymptotic efficiency bounds for inference procedures involving the structural parameter of our model and explore whether these bounds are attained by procedures currently available in the literature.

Upon imposing a rotation invariance restriction (motivated by an invariance property of the model under consideration) and employing assumptions that are standard in the literature, we find that the likelihood ratios of the resulting model are locally asymptotically normal (LAN). In other words, the limiting experiment is a Gaussian shift experiment, the statistical properties of which depend solely on its information matrix. This result constitutes our first main finding. The information matrix of the limiting experiment is found to depend smoothly on the parameter that characterizes the asymptotic behavior of the number of IVs. In relation to the existing literature, an attractive feature of the smooth dependence on the parameter characterizing the number of IVs is that it enables us to provide a unified treatment of the limiting experiments obtained by employing many weak instrument (MWI) asymptotics (e.g., Chao and Swanson, 2005; Stock and Yogo, 2005; Hansen, Hausman, and Newey, 2005) and many instrument (MI) asymptotics (e.g., Kunitomo, 1980; Morimune, 1983; Bekker, 1994; Hahn, 2002; Hansen, Hawman, and Newey, 2005), respectively.

Our second main finding is that the efficiency bounds for estimation and testing implied by the structure of the limiting experiment(s) are sharp, being attained by procedures based on the LIML estimator and its asymptotic equivalents.[3] In particular, the LIML estimator is asymptotically efficient among regular, rotation invariant estimators. Moreover, because of the LAN structure of the model, estimation and testing are dual problems, and the efficiency result for LIML furthermore implies that one- and two-sided Wald tests based on LIML enjoy asymptotic optimality properties.[4]

In combination, our main findings therefore show that there is a precise sense in which the problem of doing inference on a structural coefficient is “standard” in an MI/MWI setting, the only modifications (relative to the conventional strong IV setting) required being that the class of available asymptotically efficient estimators is smaller and that more care must be exercised when obtaining standard errors for these estimators. This property is not shared by a model with a fixed number of weak IVs, and our main findings therefore shed further light on the difference between the nature of the problems caused by weak IVs and MIs/MWIs, respectively.

The paper most closely related to the present paper is probably Hahn (2002). That paper studies a model virtually identical to the model studied herein, obtains a characterization of the limiting experiment associated with a sequence of models that do not impose rotation invariance, and uses the limiting experiment to derive an efficiency bound for regular estimators of the structural parameter.

The efficiency bound derived by Hahn (2002) is more ambitious than that obtained herein in the sense that it is obtained without imposing rotation invariance. On the other hand, it remains an open question whether the Hahn (2002) bound is attainable in the absence of fairly strong prior knowledge about the relative strength of the available IVs. In contrast, the efficiency bound obtained in this paper is demonstrably sharp. Another related paper is Anderson, Kunitomo, and Matsushita (2006), which establishes an efficiency result for LIML that is complementary to ours in the sense that it pertains to a smaller class of estimators than those considered herein, but is obtained under less restrictive distributional assumptions (i.e., without assuming normality). Finally, the large-sample distributional properties of procedures based on LIML and the Fuller (1977) modification thereof (whose asymptotic optimality properties are established herein under the assumption of normality) have been shown by Hansen et al. (2005) to be invariant to the underlying error distribution under relatively mild assumptions on the errors.

Section 2 introduces the model. The main results of the paper are collected in Section 3 and proved in Section 4.

2. THE MODEL

We consider a model with one endogenous variable, no exogenous variables, and K nonrandom IVs. The model consists of a structural equation and a reduced form equation given by

$$y_1 = y_2\beta + u,$$
$$y_2 = Z\Pi + v_2, \tag{1}$$

where y_1 and y_2 are n-dimensional vectors of endogenous variables, Z is an $n \times K$ full column rank matrix of nonrandom IVs, the scalar β and the K-dimensional vector Π are parameters, and each row of the $n \times 2$ matrix of disturbances (u, v_2) is assumed to be independent and identically distributed (i.i.d). Gaussian with mean zero and a nonsingular covariance matrix.[5]

The reduced form equations, relating the endogenous variables to the IVs, can be written as

$$y_1 = Z\Pi\beta + v_1,$$
$$y_2 = Z\Pi + v_2, \tag{2}$$

where $v_1 = u + v_2\beta$. We denote the reduced form covariance matrix by Ω; that is, Ω is (implicitly) defined by the assumption that

$$\mathrm{vec}\,(v_1, v_2) \sim \mathcal{N}\,(0, \Omega \otimes I_n)\,. \tag{3}$$

Our goal is to obtain attainable asymptotic efficiency bounds for inference procedures concerning β in the presence of the unknown nuisance parameters Ω and

Π in a sequence of models (indexed by n, the sample size) in which the number of IVs is "large" in the sense that K, the dimension of Π, tends to infinity with n.

Inference in this setting is potentially problematic for two reasons. First, although the assumption of Gaussianity implies that the probability model for $Y = (y_1, y_2)$ constitutes an exponential family of distributions, this exponential family is curved (in the terminology of Efron, 1975) whenever $K > 1$. This is so because the parameter space is of smaller dimension than the pair of minimal sufficient statistics $\left(Z'Z\right)^{-1/2} Z'Y$ and $S^{\perp} = \mathrm{vech}\left[Y'\left(I_n - P_Z\right)Y\right]$, where $(\cdot)^{1/2}$ denotes the symmetric square root of the argument and $P_Z = Z\left(Z'Z\right)^{-1} Z'$ is the projection matrix onto the column space of Z. (The dimension of the parameter space is $K+4$, and the minimal sufficient statistic is of dimension $2K+3$.) Second, under the asymptotics considered herein the dimensions of the sufficient statistic and of the nuisance parameter tend to infinity as the sample size increases without bound, potentially introducing an incidental parameter-type problem.

To avoid the latter potential difficulty we follow Andrews, Moreira, and Stock (2006) and focus on those inference procedures that (a) depend on Y through the sufficient statistics $\left(Z'Z\right)^{-1/2} Z'Y$ and $S^{\perp}$ and (b) are invariant under transformations of the form $\left[\left(Z'Z\right)^{-1/2} Z'Y, S^{\perp}\right] \to \left[\mathcal{O}\left(Z'Z\right)^{-1/2} Z'Y, S^{\perp}\right]$, for every orthogonal $K \times K$ matrix $\mathcal{O}$.[6] The (rotation) invariance restriction (b) and the sufficiency argument (a) allow us to reduce the data to a six-dimensional statistic given by the pair of random vectors $S = (S_{11}, S_{12}, S_{22})'$ and $S^{\perp} = (S_{11}^{\perp}, S_{12}^{\perp}, S_{22}^{\perp})'$, whose components are of the form $S_{ij} = y_i' P_Z y_j$ and $S_{ij}^{\perp} = y_i'\left(I_n - P_Z\right)y_j$, respectively.

The restrictions we impose are satisfied by any procedure based on a member of the familiar k-class of estimators, provided that the (possibly sample dependent) parameter k is rotation invariant. In particular, inference procedures based on 2SLS, bias adjusted 2SLS (Nagar, 1959), LIML, and the Fuller (1977) modification of LIML are covered by our results, as are procedures based on the random effects quasi–maximum likelihood estimators of Chamberlain and Imbens (2004). Furthermore, the restrictions are satisfied by many testing procedures robust to weak IVs, such as those of Anderson and Rubin (1949), Kleibergen (2002), Moreira (2003), and Andrews et al. (2006). Excluded from consideration, on the other hand, are inference procedures that utilize only a subset of the available IVs (e.g., Donald and Newey, 2001).

The matrices $Y'P_ZY$ and $Y'\left(I_n - P_Z\right)Y$ are independent with (noncentral Wishart) distributions depending only on K and the parameters β, Q, and Ω, where

$$Q = K^{-1}\Pi' Z' Z \Pi. \tag{4}$$

Therefore, the K-dimensional nuisance parameter Π affects the distribution of $\left(S, S^{\perp}\right)$ only through the scalar nuisance parameter Q, which itself is proportional to the so-called concentration parameter, a unitless measure of the strength of the IVs, which plays an important role in the IV literature (e.g., Rothenberg, 1984; Stock, Wright, and Yogo, 2002).

Although considerable simplification is achieved by imposing rotation invariance, the first potential difficulty remains a possible cause for concern because the dimension of $\left(S, S^{\perp}\right)$ exceeds that of the associated parameter space. One of the main conclusions of this paper is that this discrepancy vanishes asymptotically in an MI/MWI framework where K and $\Pi' Z' Z \Pi$ tend to infinity at the same rate. Indeed, our results show that there is a precise sense in which the problem of doing inference on β is "standard" in an MI/MWI setting.

3. RESULTS

This section characterizes the limiting experiment(s) associated with the sequence of probability models for $\left(S_n, S_n^{\perp}\right)$ under MI and MWI asymptotics. The characterization of the limiting experiment(s) will be obtained by studying a sequence of models in which the dimension of the IVs is a sequence $\{K_n\}$ satisfying

$$\lim_{n\to\infty} K_n = \infty \quad \text{and} \quad \lim_{n\to\infty} \frac{K_n}{n - K_n} = \alpha, \tag{5}$$

where the scalar α satisfies $\alpha \in [0, \infty)$. In other words, the ratio $K_n/(n - K_n)$ of the degrees of freedom parameters associated with S_n and $S_n^{\perp}$, respectively, is assumed to be convergent (with limit α). The parameters β and Ω are modeled as convergent sequences $\{\beta_n\}$ and $\{\Omega_n\}$ with limits denoted by β_0 and Ω_0, respectively. Finally, in the spirit of Hansen et al. (2005) (and many others), we model the degree of identification Q as a convergent sequence $\{Q_n\}$ whose limit is denoted by Q_0. Specifically, we employ reparameterizations of the form

$$\beta_n = \beta_0 + K_n^{-1/2} h_\beta, \tag{6}$$

$$Q_n = Q_0 + Q_0 K_n^{-1/2} h_Q, \tag{7}$$

and

$$\operatorname{vech}(\Omega_n) = \operatorname{vech}(\Omega_0) + \begin{pmatrix} 1 & 2\beta_0 & \beta_0^2 \\ 0 & 1 & \beta_0 \\ 0 & 0 & 1 \end{pmatrix} (n - K_n)^{-1/2} h_\Omega, \tag{8}$$

where β_0, Ω_0, and Q_0 are treated as known, whereas $h = \left(h_\beta, h_Q, h_\Omega'\right)' \in \mathbb{R}^5$ is a vector of unknown parameters.[7] Accordingly, we index probability distributions by h and denote the density of $\left(S_n, S_n^{\perp}\right)$ by $f_n(\cdot|h)$. Among the elements of h, h_β is the parameter of interest, whereas h_Q and h_Ω act as nuisance parameters.

Most papers in the existing literature distinguish between MI and MWI asymptotics (e.g., Hansen et al., 2005), the key difference between the two arguably being that the number of IVs is modeled as a vanishing fraction of the sample size under MWI asymptotics. In models with possibly non-Gaussian errors, a further difference between MI and MWI asymptotics is that more stringent regularity conditions appear to be required when developing distribution theory under the former type of asymptotics. Because of the fact that Gaussianity is assumed herein, there is no need to distinguish sharply between MI and MWI asymptotics, and it is sufficient to regard the latter as being the special case of the former in which $\alpha = 0$.

Theorem 1, our main result, shows that the likelihood ratios of the sequence of local experiments for $\left(S_n, S_n^{\perp}\right)$ are LAN. The formulation of the theorem deliberately highlights only those features of the limiting experiment that will prove important when obtaining asymptotic efficiency bounds, relegating to the proof a complete description of the limiting experiment. Let $o_{p_0}(1)$ be shorthand for "$o_p(1)$ under the distributions associated with $h = 0$" and let $\rightarrow_{d_0}$ abbreviate "$\rightarrow_d$ under the distributions associated with $h = 0$."

THEOREM 1. *If* (5)–(8) *hold, then*

$$\log \frac{f_n\left(S_n, S_n^{\perp} | h\right)}{f_n\left(S_n, S_n^{\perp} | 0\right)} = h' V_\alpha^{-1} \Delta_n - \frac{1}{2} h' V_\alpha^{-1} h + o_{p_0}(1) \qquad \forall h \in \mathbb{R}^5,$$

where $\Delta_n \rightarrow_{d_0} \mathcal{N}(0, V_\alpha)$ *and* $V_\alpha \in \mathbb{R}^{5\times 5}$ *has first diagonal element equal to*

$$V_{\alpha,\beta\beta} = Q_0^{-1}(1, -\beta_0)\,\Omega_0\,(1, -\beta_0)' + Q_0^{-2}(1+\alpha)\,|\Omega_0|. \tag{9}$$

It follows from Theorem 1 that the limiting experiment depends on K_n only through α. In fact, the limiting experiment depends smoothly on α, implying that the MWI and MI cases can be treated in a unified way, the former (which has $\alpha = 0$) being a limiting case of the latter (which has $0 < \alpha < \infty$).

Let

$$\tilde{S}_n = \frac{S_n - \mathrm{E}_0(S_n)}{Q_0\sqrt{K_n}}, \qquad \tilde{S}_n^{\perp} = \frac{S_n^{\perp} - \mathrm{E}_0\left(S_n^{\perp}\right)}{\sqrt{n - K_n}}, \tag{10}$$

where E_0 denotes expectation computed under the distribution associated with $h = 0$. The proof of Theorem 1 utilizes the fact that for any h, we have a convergence result of the form

$$\begin{pmatrix} \tilde{S}_n \\ \tilde{S}_n^{\perp} \end{pmatrix} \rightarrow_{d_h} \mathcal{N}\left[\begin{pmatrix} h_S + \sqrt{\alpha}\, Q_0^{-1} h_\Omega \\ h_\Omega \end{pmatrix}, \begin{pmatrix} \Sigma & 0 \\ 0 & \Sigma^{\perp} \end{pmatrix}\right], \tag{11}$$

where $h_S = \left(0, h_\beta, h_Q\right)'$ and $\rightarrow_{d_h}$ abbreviates "$\rightarrow_d$ under the distributions associated with h." (The matrices Σ and $\Sigma^{\perp}$ are defined in the proof of Theorem 1.) The pointwise convergence (11) is not sufficient to prove Theorem 1, though. Certain uniformity is required.[8] We use normality of (v_1, v_2) to establish such uniformity.

An important implication of Theorem 1 is that the asymptotic optimality problem concerning estimation of β is isomorphic to the optimality problem concerning estimation of h_β in the experiment based on one observation from an $\mathcal{N}(h, V_\alpha)$ distribution. In particular, proceeding as in van der Vaart (1998) it is possible to conclude that a regular, rotation invariant estimator $\hat{\beta}_n$ of β is asymptotically efficient (among regular, rotation invariant estimators) if

$$\sqrt{K_n}\left(\hat{\beta}_n - \beta_0\right) \to_{d_0} \mathcal{N}\left(0, V_{\alpha,\beta\beta}\right). \tag{12}$$

(A rotation invariant estimator $\hat{\beta}_n$ is regular if there exists a random variable $\mathcal{B}$ such that $\sqrt{K_n}\left(\hat{\beta}_n - \beta_n\right) \to_{d_h} \mathcal{B}$ for every $h \in \mathbb{R}^5$.)

This is not the first efficiency bound derived in the MI literature. Employing distributional assumptions very similar to ours, Hahn (2002) elegantly obtains an efficiency bound for the structural parameter in a linear simultaneous equations model under MI asymptotics without imposing a rotation invariance restriction. The bound derived by Hahn (2002) is strictly smaller than that obtained here, but we are not aware of any estimators that attain that bound. In contrast, it turns out that the bound reported in (12) is attainable. Indeed, it follows from Theorem 1 of Anderson et al. (2006) that the LIML estimator satisfies (12). As a consequence, we have the following result.

COROLLARY 2. *Under the assumptions of Theorem 1, the LIML estimator is asymptotically efficient among regular, rotation invariant estimators.*

This efficiency property of LIML is not shared by the 2SLS estimator (which is inconsistent in general), but it follows from Hansen et al. (2005) that the Fuller (1977) modification of LIML is asymptotically equivalent to LIML (and hence asymptotically efficient) under the assumptions made herein.

It follows from Bekker (1994) that the LIMLK (i.e., LIML with known Ω, the known Ω being Ω_0) estimator $\hat{\beta}_{LIMLK,n}$ satisfies

$$\sqrt{K_n}\left(\hat{\beta}_{LIMLK,n} - \beta_0\right) \to_{d_0} \mathcal{N}\left(0, V_{0,\beta\beta}\right). \tag{13}$$

The LIMLK estimator is therefore superior to the LIML estimator whenever $\alpha > 0$. In fact, Theorem 1 can be used to show that $\hat{\beta}_{LIMLK,n}$ is asymptotically efficient when Ω is known, and so the quantity $V_{0,\beta\beta}^{-1} - V_{\alpha,\beta\beta}^{-1}$ can be interpreted as the information content of Ω when $\alpha > 0$.

Corollary 2 is related to Theorem 2 of Anderson et al. (2006). The latter gives conditions under which LIML is efficient among consistent estimators that can be represented as $\phi\left[S_n/K_n, S_n^{\perp}/(n-K_n)\right]$ for some smooth function ϕ (not depending on n). (In the context of MI asymptotics, a similar result can also be found in van der Ploeg and Bekker, 1995, and Hahn, 2002.) The class of estimators covered by Theorem 2 of Anderson et al. (2006) is a strict subset of the class of regular estimators.[9] As a consequence, Corollary 2 is a stronger result than Theorem 2 of

Anderson et al. (2006) under the assumptions made herein.[10] On the other hand, because Anderson et al. (2006) do not assume normality, their Theorem 2 is not a special case of Corollary 2.[11]

In addition to enabling us to generalize the Anderson et al. (2006) efficiency result to the class of all regular estimators, Theorem 1 and Corollary 2 make it possible to show that one-sided (two-sided) Wald tests centered around LIML (and employing Bekker, 1994, standard errors, as in Hansen et al., 2005) are asymptotically uniformly most powerful (unbiased) among rotation invariant testing procedures. Precise statements can be found in Choi, Hall, and Schick (1996, Sects. 3 and 4); to conserve space, we do not replicate their results here. In other words, the LAN property implies that estimation and testing are dual problems that do not require separate treatment.[12] In this precise sense, the statistical properties of the MWI and WI frameworks are "standard" and closely resemble the properties of the conventional (strong IV) asymptotics framework associated with fixed K and fixed Π.

4. PROOFS

Proof of Theorem 1. The proof proceeds under the assumption that $\beta_0 = 0$. This assumption simplifies the algebra but entails no loss of generality, as it simply corresponds to a model in which y_1 and Ω have been replaced by $y_1 - \beta_0 y_2$ and

$$\begin{pmatrix} 1 & -\beta_0 \\ 0 & 1 \end{pmatrix} \Omega \begin{pmatrix} 1 & -\beta_0 \\ 0 & 1 \end{pmatrix}',$$

respectively.

We start by establishing (11) for

$$\Sigma^{\perp} = \begin{pmatrix} 2\omega_{11,0}^2 & 2\omega_{12,0}\omega_{11,0} & 2\omega_{12,0}^2 \\ 2\omega_{12,0}\omega_{11,0} & \omega_{12,0}^2 + \omega_{11,0}\omega_{22,0} & 2\omega_{12,0}\omega_{22,0} \\ 2\omega_{12,0}^2 & 2\omega_{12,0}\omega_{22,0} & 2\omega_{22,0}^2 \end{pmatrix}$$

and

$$\Sigma = Q_0^{-1} \begin{pmatrix} 0 & 0 & 0 \\ 0 & \omega_{11,0} & 2\omega_{12,0} \\ 0 & 2\omega_{12,0} & 4\omega_{22,0} \end{pmatrix} + Q_0^{-2} \Sigma^{\perp},$$

where $\omega_{ij,0}$ denotes element (i, j) of Ω_0. Proceeding as in Hahn (2002), it can be shown that

$$\begin{pmatrix} \tilde{S}_n - \mu_n(h) \\ \tilde{S}_n^{\perp} - \mu_n^{\perp}(h) \end{pmatrix} \to_{d_h} \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma & 0 \\ 0 & \Sigma^{\perp} \end{pmatrix} \right] \qquad \forall h, \tag{14}$$

where $\mu_n(h) = \mathrm{E}_h\left(\tilde{S}_n\right) - \mathrm{E}_0\left(\tilde{S}_n\right)$, $\mu_n^{\perp}(h) = \mathrm{E}_h\left(\tilde{S}_n^{\perp}\right) - \mathrm{E}_0\left(\tilde{S}_n^{\perp}\right)$, and E_h denotes expectation computed under the distribution associated with h. Moreover, it follows from simple calculations that

$$\mu_n(h) = h_S + \sqrt{\alpha}\, Q_0^{-1} h_\Omega + o(1) \tag{15}$$

and

$$\mu_n^{\perp}(h) = h_\Omega. \tag{16}$$

The result (11) follows from (14)–(16).

Next, to define the objects Δ_n and V_α, let

$$\begin{pmatrix} \sigma_{AA} & \sigma_{AB} \\ \sigma_{BA} & \Sigma_{BB} \end{pmatrix} = \begin{pmatrix} \Sigma + \alpha Q_0^{-2} \Sigma^{\perp} & -\sqrt{\alpha}\, Q_0^{-1} \Sigma^{\perp} \\ -\sqrt{\alpha}\, Q_0^{-1} \Sigma^{\perp} & \Sigma^{\perp} \end{pmatrix},$$

where the matrix on the left is partitioned after the first row and column. Define

$$\begin{pmatrix} \delta_n \\ \Delta_n \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\sigma_{BA}\sigma_{AA}^{-1} & I_5 \end{pmatrix} \begin{pmatrix} \tilde{S}_n - \sqrt{\alpha}\, Q_0^{-1} \tilde{S}_n^{\perp} \\ \tilde{S}_n^{\perp} \end{pmatrix}$$

and

$$\begin{pmatrix} V_{\alpha,\delta\delta} & 0 \\ 0 & V_\alpha \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\sigma_{BA}\sigma_{AA}^{-1} & I_5 \end{pmatrix} \begin{pmatrix} \sigma_{AA} & \sigma_{AB} \\ \sigma_{BA} & \Sigma_{BB} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\sigma_{BA}\sigma_{AA}^{-1} & I_5 \end{pmatrix}.$$

(The objects δ_n and $V_{\alpha,\delta\delta}$ are scalars.) It follows from (11) that

$$\begin{pmatrix} \delta_n \\ \Delta_n \end{pmatrix} \to_{d_h} \mathcal{N}\left[\begin{pmatrix} 0 \\ h \end{pmatrix}, \begin{pmatrix} V_{\alpha,\delta\delta} & 0 \\ 0 & V_\alpha \end{pmatrix} \right]. \tag{17}$$

In particular, $\Delta_n \to_{d_h} \mathcal{N}(h, V_\alpha)$. Moreover, because the upper left 2×2 block of $\Sigma + \alpha Q_0^{-2} \Sigma^{\perp}$ is

$$Q_0^{-1} \begin{pmatrix} 0 & 0 \\ 0 & \omega_{11,0} \end{pmatrix} + (1+\alpha)\, Q_0^{-2} \begin{pmatrix} 2\omega_{11,0}^2 & 2\omega_{12,0}\omega_{11,0} \\ 2\omega_{12,0}\omega_{11,0} & \omega_{12,0}^2 + \omega_{11,0}\omega_{22,0} \end{pmatrix},$$

the first diagonal element of V_α is given by $V_{\alpha,\beta\beta}$.

Because (δ_n, Δ_n) is obtained from $\left(S_n, S_n^{\perp}\right)$ by a bijective affine transformation,

$$\log \frac{f_n\left(S_n, S_n^{\perp} | h\right)}{f_n\left(S_n, S_n^{\perp} | 0\right)} = \log \frac{g_n(\delta_n, \Delta_n | h)}{g_n(\delta_n, \Delta_n | 0)},$$

where $g_n(\cdot|h)$ denotes the density of (δ_n, Δ_n). In addition, it follows from (17) that the (continuous and strictly positive) densities g_∞ of the limiting distribution of (δ_n, Δ_n) satisfy

$$\log \frac{g_\infty(\delta, \Delta|h)}{g_\infty(\delta, \Delta|0)} = h' V_\alpha^{-1} \Delta - \frac{1}{2} h' V_\alpha^{-1} h.$$

The proof of Theorem 1 can therefore be completed by showing that $g_n(\cdot|h)$ converges to $g_\infty(\cdot|h)$ in the topology of uniform convergence on compacta (for every $h \in \mathbb{R}^5$).

To do so, fix $h \in \mathbb{R}^5$ and let $\varphi_n(\cdot|h)$ denote the characteristic function associated with $g_n(\cdot|h)$; that is, let

$$\varphi_n(s, t|h) = \int_{\mathbb{R}} \int_{\mathbb{R}^5} \exp\left[i\left(s\delta + t'\Delta\right)\right] g_n(\delta, \Delta|h)\, d\Delta\, d\delta.$$

Similarly, let $\varphi_\infty(\cdot|h)$ denote the characteristic function associated with $g_\infty(\cdot|h)$. Using Muirhead (1982, Thm. 10.3.3) and the proof of that theorem, simple bounding arguments can be used to show that there exists a finite integer N_h (the value of which depends on h) for which

$$\int_{\mathbb{R}} \int_{\mathbb{R}^5} \sup_{n \geq N_h} |\varphi_n(s, t|h)|\, dt\, ds < \infty. \tag{18}$$

This dominance result implies that

$$g_n(\delta, \Delta|h) = \frac{1}{(2\pi)^6} \int_{\mathbb{R}} \int_{\mathbb{R}^5} \exp\left[-i\left(s\delta + t'\Delta\right)\right] \varphi_n(s, t|h)\, dt\, ds \tag{19}$$

for every $n \geq N_h$ (by the inversion formula for characteristic functions) and furthermore justifies our subsequent applications of the dominated convergence theorem.

It follows from (17) that $\varphi_n(\cdot|h)$ converges pointwise to $\varphi_\infty(\cdot|h)$. This fact, the dominated convergence theorem, and (18)–(19) can be used to show that (a) $g_n(\cdot|h)$ converges pointwise to $g_\infty(\cdot|h)$ and (b) the family $\{g_n(\cdot|h) : n \geq N_h\}$ is (uniformly) equicontinuous. By Rudin (1976, Prob. 7.16), these properties imply that $g_n(\cdot|h)$ converges to $g_\infty(\cdot|h)$ in the topology of uniform convergence on compacta. ■

Proof of Corollary 2. Because it follows from Theorem 1 of Anderson et al. (2006) that the LIML estimator satisfies (12), it suffices to show that a regular, invariant estimator is asymptotically efficient if it satisfies (12). The latter assertion can be proved in the same way as Theorem 8.8 of van der Vaart (1998). Specifically, an inspection of the proof of Theorem 8.8 of van der Vaart (1998) shows that the assumption of differentiability in quadratic mean is used only to justify the use of a simplified version of the asymptotic representation theorem (van der Vaart, 1998, Thm. 7.10). Moreover, it follows from van der Vaart (1998, Ch. 9) and Theorem 1 that the van der Vaart (1991) version of the asymptotic representation theorem is valid under the assumptions made herein. ■

NOTES

1. Here, and elsewhere in the paper, concepts such as "large," "strong," and "weak" are left undefined, but the usage of these adjectives follows the conventions of the existing IV literature.

2. The complications caused by weak IVs are well documented and have been studied intensively in recent years, important contributions being Dufour (1997), Staiger and Stock (1997), Kleibergen (2002), Moreira (2003), and Andrews et al. (2006). For reviews, see, for example, Stock et al. (2002), Dufour (2003), Hahn and Hausman (2003), and Andrews and Stock (2007).

3. As shown by Hansen et al. (2005), the class of estimators that are asymptotically equivalent to LIML (in the appropriate sense) includes the Fuller (1977) modification of LIML but does not include the 2SLS estimator.

4. It follows from Hansen et al. (2005) that such Wald tests should employ standard errors that account for the number of IVs, such as Bekker (1994) standard errors.

5. Throughout Section 2, we simplify the notation by suppressing the dependence of variables, parameters, and so on, on n.

6. The invariance restriction is motivated by the fact that inference problems concerning β are invariant under transformations of the form $\left[\left(Z'Z\right)^{-1/2}Z'Y, S^{\perp}\right] \rightarrow \left[\mathcal{O}\left(Z'Z\right)^{-1/2}Z'Y, S^{\perp}\right]$, as the induced transformations $[\beta,\pi,\Omega] \rightarrow \left[\beta, \left(Z'Z\right)^{-1/2}\mathcal{O}\left(Z'Z\right)^{1/2}\pi, \Omega\right]$ of the parameters leave β (and Ω) unchanged.

7. The reparameterizations employed are chosen in such a way that (a) the distributions associated with different values of h are mutually absolutely contiguous and (b) the resulting limiting experiment depends on h in a simple way. Requirement (a) gives rise to the rates $\sqrt{K_n}$, $\sqrt{K_n}$, and $\sqrt{n-K_n}$ in equations (6), (7), and (8), whereas (b) governs the choice of the multiplicative constants in the "drift" terms in these equations.

8. Theorem 1 is a statement about the limiting behavior of likelihood ratios. To show that this limiting behavior coincides with the behavior of the likelihood ratios obtained from the limiting distributions characterized in (11) we use the normality of (v_1, v_2) to show that the densities of $\left(\tilde{S}_n, \tilde{S}_n^{\perp}\right)$ converge to the densities of the associated limiting distributions in the topology of uniform convergence on compacta.

9. The estimators considered by Anderson et al. (2006) satisfy $\sqrt{K_n}\left(\phi\left[S_n/K_n, S_n^{\perp}/(n-K_n)\right] - \beta_n\right) \rightarrow_{d_h} \mathcal{N}\left(0, V_{\alpha,\beta\beta}^{\phi}\right) \forall h \in \mathbb{R}^5$ for some matrix $V_{\alpha,\beta\beta}^{\phi} \geq V_{\alpha,\beta\beta}$. In particular $\phi\left[S_n/K_n, S_n^{\perp}/(n-K_n)\right]$ is regular, but the class of regular estimators also contains estimators whose associated normalized estimation error has a limiting distribution that is non-Gaussian and/or has a nonzero mean.

10. A further difference between our results and those of Anderson et al. (2006) is that Theorem 1 of the present paper makes it possible to establish another optimality property (of the local asymptotic minimax variety) for LIML that avoids the assumption of regularity altogether. To conserve space, we omit a precise statement and refer the reader to van der Vaart (1998, Sect. 8.7) for details.

11. Theorem 2 of Anderson et al. (2006) is a corollary of a convergence in distribution result of the form (11), whose validity does not require normality. In contrast, the full force of Theorem 1 (a stronger result, the proof of which relies more heavily on normality) is utilized when obtaining Corollary 2.

12. In contrast, there is no such duality between estimation and testing in the weak IV case, where K_n is fixed and Π is $O\left(1/\sqrt{n}\right)$. As we have argued elsewhere, the weak IV scenario furthermore has the property that a case can be made for conducting conditional inference. (For details, see Chioda and Jansson, 2005.) Our argument for doing conditional inference in that setting hinges on the limiting

experiment not being a Gaussian shift experiment, and so a further implication of the main result of the present paper is that there does not appear to be any compelling reason to do conditional inference when the number of instruments is best modeled as a divergent sequence.

REFERENCES

Anderson, T.W., N. Kunitomo, & Y. Matsushita (2006) A New Light from Old Wisdoms: Alternative Estimation Methods of Simultaneous Equations with Possibly Many Instruments. Working paper, University of Tokyo.

Anderson, T.W. & H. Rubin (1949) Estimation of the parameters of a single equation in a complete set of stochastic equations. *Annals of Mathematical Statistics* 20, 46–63.

Andrews, D.W.K., M.J. Moreira, & J.H. Stock (2006) Optimal two-sided invariant similar tests for instrumental variables regression. *Econometrica* 74, 715–752.

Andrews, D.W.K. & J.H. Stock (2007) Inference with weak instruments. In R. Blundell, W.K. Newey, & T. Persson (eds.) *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress, vol. III,* pp. 122–173. Cambridge University Press.

Bekker, P.A. (1994) Alternative approximations to the distributions of instrumental variables estimators. *Econometrica* 62, 657–681.

Chamberlain, G. & G. Imbens (2004) Random effects estimators with many instrumental variables. *Econometrica* 72, 295–306.

Chao, J.C. & N.R. Swanson (2005) Consistent estimation with a large number of weak instruments. *Econometrica* 73, 1673–1692.

Chioda, L. & M. Jansson (2005) Optimal Conditional Inference for Instrumental Variables Regression. Working paper, Princeton University.

Choi, S., W.J. Hall, & A. Schick (1996) Asymptotically uniformly most powerful tests in parametric and semiparametric models. *Annals of Statistics* 24, 841–861.

Donald, S.G. & W.K. Newey (2001) Choosing the number of instruments. *Econometrica* 69, 1161–1191.

Dufour, J.-M. (1997) Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica* 65, 1365–1387.

Dufour, J.-M. (2003) Identification, weak instruments, and statistical inference in econometrics. *Canadian Journal of Economics* 36, 767–808.

Efron, B. (1975) Defining the curvature of a statistical problem (with applications to second order efficiency). *Annals of Statistics* 3, 1189–1242.

Fuller, W.A. (1977) Some properties of a modification of the limited information estimator. *Econometrica* 45, 939–953.

Hahn, J. (2002) Optimal inference with many instruments. *Econometric Theory* 18, 140–168.

Hahn, J. & J. Hausman (2003) Weak instruments: Diagnosis and cures in empirical econometrics. *American Economic Review* 93, 118–125.

Hansen, C., J. Hausman, & W. Newey (2005) Estimation with Many Instrumental Variables. Working paper, University of Chicago Graduate School of Business.

Kleibergen, F. (2002) Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica* 70, 1781–1803.

Kunitomo, N. (1980) Asymptotic expansions of the distributions of estimators in a linear functional relationship and simultaneous equations. *Journal of the American Statistical Association* 75, 693–700.

Moreira, M.J. (2003) A conditional likelihood ratio test for structural models. *Econometrica* 71, 1027–1048.

Morimune, K. (1983) Approximate distributions of k-class estimators when the degree of overidentifiability is large compared with the sample size. *Econometrica* 53, 821–841.

Muirhead, R.J. (1982) *Aspects of Multivariate Statistical Theory*. Wiley.

Nagar, A.L. (1959) The bias and moment matrix of general k-class estimators of the parameters in simultaneous equations. *Econometrica* 27, 575–595.

Rothenberg, T.J. (1984) Approximating the distributions of econometric estimators and test statistics. In Z. Griliches & M.D. Intriligator (eds.), *Handbook of Econometrics, vol. II.* pp. 881–935 North-Holland.

Rudin, W. (1976) *Principles of Mathematical Analysis,* 3rd ed. McGraw-Hill.

Staiger, D. & J.H. Stock (1997) Instrumental variables estimation with weak instruments. *Econometrica* 65, 557–586.

Stock, J.H., J.H. Wright, & M. Yogo (2002) A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics* 20, 518–529.

Stock, J.H. & M. Yogo (2005) Asymptotic distributions of instrumental variables statistics with many weak instruments. In D.W.K. Andrews & J.H. Stock (eds.) *Identification and Inference in Econometric Models: Essays in Honor of Thomas J. Rothenberg*, pp. 109–120. Cambridge University Press.

van der Ploeg, J. & P.A. Bekker (1995) Efficiency Bounds for Instrumental Variable Estimators under Group-Asymptotics. Working paper, University of Groningen.

van der Vaart, A.W. (1991) An asymptotic representation theorem. *International Statistical Review* 59, 97–121.

van der Vaart, A.W. (1998) *Asymptotic Statistics*. Cambridge University Press.