Econometric Theory, **22**, 2006, 913–931. Printed in the United States of America.
DOI: 10.1017/S0266466606060415

YET MORE ON THE EXACT PROPERTIES OF IV ESTIMATORS

GRANT HILLIER
University of Southampton

We revisit the exact properties of two-stage least squares and limited information maximum likelihood estimators in a structural equation/instrumental variables regression under Gaussian assumptions. Simple derivations based on conditioning serve both to demystify the apparently complicated formulas, and to isolate the key quantities that determine the properties of the estimators. Some recent results obtained under weak-instrument asymptotics are sharpened and clarified by the exact analysis.

1. INTRODUCTION

The literature studying the statistical properties, both asymptotic and exact, of instrumental variables (IV) estimators in structural models is very large and still growing. Asymptotic results, under the assumption of full identification, are quite standard, but recent work, beginning with the Phillips (1989) paper on partially identified models, suggests that asymptotic results can, in weakly identified models, be a poor guide to the actual properties of the estimators of interest (see also the discussion of the role of identification in Phillips, 1985; and the recent survey paper by Stock, Wright, and Yogo, 2002). Thus, interest has been reawakened in the exact properties of estimators and their relation to asymptotic properties.

The topic has gained impetus from the weak-instrument asymptotics introduced by Staiger and Stock (1997). Under "weak-instrument asymptotics" assumptions the asymptotic properties of various estimators and test procedures for quite general models have the same formal structure as Gaussian exact results in the more stylized models that have been the context of most finite-sample research. That is, the "classical" finite-sample results (that appear, e.g., in Phillips, 1983) have much broader relevance than was previously realized. Phillips (1983) provides the definitive summary of results prior to the mid-1980s, and Stock et al. (2002) covers much of the work since then.

Even in simplified models, however, exact results have always seemed difficult both to derive and to interpret. One purpose of this paper will be to pro-

Thanks to Peter Phillips and several anonymous referees for helpful comments that improved the paper considerably. Address correspondence to Grant Hillier, Economics Division, School of Social Sciences, University of Southampton, Highfield, Southampton, SO17 1BJ, United Kingdom; e-mail: ghh@soton.ac.uk.

 0266-4666/06 $12.00

vide simple derivations of, and greater insight into the properties of, the exact densities—in particular, the features of the problem that determine the location and concentration of the distributions of the two-stage least squares (TSLS) and limited information maximum likelihood (LIML) estimators. And, because recent attention has focused on the question of how the properties of the instruments affect the properties (i.e., density) of the estimator for the structural coefficients (Nelson and Startz, 1990; Maddala and Jeong, 1992; Woglom, 2001), our second purpose will be to clarify that relationship.

Both of these objectives are, to a large extent, achieved by the same device—our method of deriving the exact results: we express the marginal density of the estimator as the average of its conditional density given the first-stage regression estimator. This shows directly that the properties of the estimator are induced by those of the first-stage regression. In addition, because our formula for the density is so simple and easily interpreted—at least in the exactly identified case—we are able to isolate the two key parameters that determine its location and concentration. Finally, the roles of these key parameters are brought into sharper focus by interpreting the density in a normalization-invariant form based on the argument in Hillier (1990).

In a recent, closely related, paper, Phillips (2006) uses an even simpler model than that used here for similar purposes and to discuss various aspects of the limit theory for these estimators under variations on the weak-instrument asymptotics assumption.

1.1. Model, Assumptions

To begin with we, like others who have dealt with these issues recently, consider the stripped-down exactly identified model

$$y_t = \beta x_t + u_t, \tag{1}$$

$$x_t = \gamma z_t + v_t, \tag{2}$$

with the vectors $(u_t, v_t)'$, $t = 1, \ldots, T$ *i.i.d.* $N(0, \Sigma)$. The interest parameter is β, and we refer to (2) as the "first-stage regression." Later, in Section 3, we briefly consider the overidentified version of this simplified model.

We write the covariance matrix of (u_t, v_t), Σ, as

$$\Sigma = \begin{pmatrix} \sigma_u^2 & \sigma_u \sigma_v \rho \\ \sigma_u \sigma_v \rho & \sigma_v^2 \end{pmatrix}$$

and the covariance matrix of (y_t, x_t) as

$$\Omega = \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix} \Sigma \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix}' = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{pmatrix}.$$

Thus, the correlation between u_t and v_t is ρ. The model (reduced form) assumed to have generated the data is of the form

$$Y|z \sim N(z\alpha', I_T \otimes \Omega), \tag{3}$$

where $Y = (y, x)$ is the $T \times 2$ matrix with rows (y_t, x_t), $t = 1, \ldots, T$, and z is $T \times 1$ with tth element z_t. Equations (1) and (2) imply the parameterization $\alpha' = (\alpha_1, \alpha_2) = (\gamma\beta, \gamma)$.

Remark 1. It follows from (3) that the conditional density of y given x is

$$y|x \sim N(x\beta + \rho(\sigma_u/\sigma_v)(x - z\gamma), \sigma_u^2(1 - \rho^2)I_T). \tag{4}$$

Thus, unless $\rho = 0$, the interest parameter β has no simple interpretation as a property of this conditional distribution. On the other hand, β does have a simple interpretation in terms of the model family (3): $\beta = \alpha_1/\alpha_2$ is the ratio of the elements of α and, when defined, is therefore one possible measure of the "direction" of α. This interpretation will be discussed further in Section 2.4.

In the multivariate normal model (reduced form) (3), the parameter space for α would usually be all of $\mathbb{R}^2$, including the origin ($\alpha = 0$) and the axis $\alpha_2 = 0$. But, when $\alpha = 0$ neither β nor the "direction" of α is determined, whereas when $\alpha \neq 0$ but $\alpha_2 = 0$, the direction of α is determined but β remains undefined. This, of course, is the familiar identification problem, but, as discussed by Phillips (1985) 20 years ago, the fact that β is undefined does not prevent one from computing "estimators" for it or tests of hypotheses about it. However, we expect to find that the properties of such statistics reflect the unidentifiability of the parameter, and they do: the exact distributions of both the LIML and TSLS estimators do not depend on the interest parameter, whatever the sample size, and test statistics (such as the Anderson–Rubin statistic; Anderson and Rubin, 1949) have power equal to size (i.e., the statistic also has a distribution free of the interest parameter).

Far from being a deficiency of such inferential procedures, however, these characteristics are precisely what one would hope for: inferential precision for β *should* be low when the model is unable to clearly identify it (for more discussion and references on this, see Forchini and Hillier, 2005). Interestingly, as others have shown using asymptotic arguments (see, e.g., Morimune, 1983; Newey, 2004; Chao and Swanson, 2006), our exact analysis adds further weight to the growing view that ad hoc procedures like OLS and TSLS do not have the same desirable characteristics as likelihood-based procedures like LIML, at least in circumstances of practical relevance where the sample size may be small or instruments weak—that is, in circumstances where the "information" on the interest parameter is in some sense poor.

The results to follow make use of the standard hypergeometric function notation:

$$ {}_pF_q(a_1,..,a_p;c_1,..,c_q;z) = \sum_{j=0}^{\infty} \frac{(a_1)_j \ldots (a_p)_j}{(c_1)_j \ldots (c_q)_j} \frac{z^j}{j!}, $$

where $(c)_j$ denotes the forward factorial: $(c)_j = c(c+1)\ldots(c+j-1)$. We use only the cases ${}_0F_1$ (Bessel) and ${}_1F_1$ (confluent), which converge for all z. See Muirhead (1982, Ch. 1) for more details.

2. THE IV ESTIMATOR AND ITS PROPERTIES

2.1. Estimator, Conditional Density

The maximum likelihood estimator (MLE) for α is easily seen to be $\hat{\alpha} = (y,x)'z(z'z)^{-1}$, and $\hat{\alpha} \sim N(\alpha,(z'z)^{-1}\Omega)$. The MLE for β is therefore $b = \hat{\alpha}_1/\hat{\alpha}_2 = (y'z)/(x'z)$, which is both the LIML and TSLS estimator in this simplified setup. Note that $\hat{\alpha}_2 = \hat{\gamma} = (z'z)^{-1}z'x$ is the ordinary least squares (OLS) estimator for γ in the first-stage regression (2).

Remark 2. Just as $\beta = \alpha_1/\alpha_2$ is one possible measure of the direction of α, so $b = \hat{\alpha}_1/\hat{\alpha}_2$ is one possible measure of the direction of $\hat{\alpha}$. As we shall see, its distribution properties reflect that fact.

Remark 3. It is clear that the results and discussion that follow apply to any statistic that can be written as the ratio of correlated, not necessarily zero-mean, normal variates. Thus, what follows can be viewed as revisiting Marsaglia (1965).

The first step in the derivation is to note that the *conditional* distribution of b given $\hat{\gamma}$ is, for any value $\hat{\gamma} \neq 0$,

$$ b|\hat{\gamma} \sim N\left(\beta + \frac{\sigma_u \rho(\hat{\gamma}-\gamma)}{\sigma_v \hat{\gamma}}, \frac{\sigma_u^2(1-\rho^2)}{s_{ZZ}\hat{\gamma}^2}\right), \tag{5} $$

where we have put $s_{ZZ} = z'z$. Marginally, of course, $\hat{\gamma} \sim N(\gamma, \sigma_v^2/s_{ZZ})$. The following result is clear at once.

PROPOSITION 1. *For each value $\hat{\gamma} \neq 0$ of the first-stage regression estimator for γ, the conditional distribution of b given $\hat{\gamma}$ is unimodal (normal), with conditional mean $E(b|\hat{\gamma}) = \beta + \sigma_u\rho(\hat{\gamma}-\gamma)/(\sigma_v\hat{\gamma})$ and conditional variance* $\mathrm{var}(b|\hat{\gamma}) = \sigma_u^2(1-\rho^2)/(s_{ZZ}\hat{\gamma}^2)$.

This conditional distribution is of interest in its own right (for discussion, Forchini and Hillier, 2003), and the reader may easily confirm that the conditional mean squared error $MSE(b|\hat{\gamma})$ depends on the observed value of the first-

stage regression estimator $\hat{\gamma}$. This is not surprising, because it is precisely $\hat{\gamma}$ that carries the sample information about the identifiability of β, and in Forchini and Hillier (2003) we argue (in a more general setting) that it is the conditional distribution that provides the proper measure of the precision of b: it makes no sense to average over sample outcomes for $\hat{\gamma}$ that have not occurred.

However, it is the unconditional properties of b that are our interest here. What is clear already is that the possible bimodality of the marginal density—one of its notable features—arises entirely from the process of averaging the conditional density with respect to the density of $\hat{\gamma}$. To simplify this averaging we first standardize $\hat{\gamma}$ to have unit variance by setting $g = \sqrt{s_{ZZ}}\hat{\gamma}/\sigma_v$ and $\bar{\gamma} = \sqrt{s_{ZZ}}\gamma/\sigma_v$, so that $g \sim N(\bar{\gamma},1)$. Then define the scaled estimation error in b by

$$e = \frac{\sigma_v(b-\beta)}{\sigma_u\sqrt{(1-\rho^2)}}, \tag{6}$$

so that, from (5), $e|g \sim N(\eta(1-\bar{\gamma}/g),1/g^2)$, where $\eta = \rho/\sqrt{1-\rho^2}$, $-\infty \leq \eta \leq \infty$, is a measure of the degree of endogeneity in the model. Thus, setting $w = e - \eta$, $w|g \sim N(-\eta\bar{\gamma}/g,1/g^2)$. Recall that, for the OLS estimator $b_0 = (x'x)^{-1}x'y$, $e_0 \to_p \eta$ when $\gamma = 0$.

Because scaling and recentering have no impact on the bimodality of the density, and these densities depend only upon $\bar{\gamma}$ and η, we can state the following proposition.

PROPOSITION 2. *The bimodality or otherwise of the unconditional density of w is determined entirely by the values of $\bar{\gamma} = \sqrt{s_{ZZ}}\gamma/\sigma_v$ and $\eta = \rho/\sqrt{(1-\rho^2)}$.*

2.2. Marginal Density

It is a simple matter to derive the marginal density of w from the facts that $w|g \sim N(-\eta\bar{\gamma}/g,1/g^2)$ and $g \sim N(\bar{\gamma},1)$ because the joint density of (w,g) is

$$\begin{aligned} pdf(w,g) &= pdf(w|g)pdf(g) \\ &= (2\pi)^{-1}|g|\exp\left\{-\frac{1}{2}[(gw+\eta\bar{\gamma})^2 + (g-\bar{\gamma})^2]\right\}. \end{aligned} \tag{7}$$

Integrating out g, the marginal density of w is

$$\begin{aligned} pdf(w) &= (2\pi)^{-1}\exp\left\{-\frac{1}{2}\bar{\gamma}^2(1+\eta^2)\right\} \\ &\quad \times \int_{-\infty}^{\infty}|g|\exp\left\{-\frac{1}{2}g^2(1+w^2)\right\}\exp\{g\bar{\gamma}(1-\eta w)\}(dg). \end{aligned} \tag{8}$$

But, the integral in (8) is

$$\int_0^\infty g \exp\left\{-\frac{1}{2}\, g^2(1+w^2)\right\}[\exp\{g\bar{\gamma}(1-\eta w)\} + \exp\{-g\bar{\gamma}(1-\eta w)\}](dg),$$

and it is easy to see that $(e^x + e^{-x})/2 = {}_0F_1(\frac{1}{2}; \frac{1}{4}x^2)$, so that the integral becomes

$$= 2\int_0^\infty g \exp\left\{-\frac{1}{2}\, g^2(1+w^2)\right\} {}_0F_1\left(\frac{1}{2};\, \frac{1}{4}\, g^2\bar{\gamma}^2(1-\eta w)^2\right)(dg)$$

$$= 2(1+w^2)^{-1}\, {}_1F_1\left(1, \frac{1}{2};\, \frac{1}{2}\,\bar{\gamma}^2\,\frac{(1-\eta w)^2}{(1+w^2)}\right), \tag{9}$$

on transforming from g to $q = g^2$ and integrating over $q > 0$.

Hence, from (8) and (9):

$$pdf(w) = \pi^{-1}(1+w^2)^{-1} \exp\left\{-\frac{1}{2}\,\lambda\right\} {}_1F_1\left(1, \frac{1}{2};\, \frac{1}{2}\,\lambda\left\{\frac{(1-\eta w)^2}{(1+w^2)(1+\eta^2)}\right\}\right), \tag{10}$$

where

$$\lambda = \bar{\gamma}^2(1+\eta^2) = \frac{\gamma^2 s_{ZZ}}{\sigma_v^2(1-\rho^2)} = s_{ZZ}\alpha'\Omega^{-1}\alpha. \tag{11}$$

Remark 4. Equation (10) evidently reduces to the Cauchy distribution when $\lambda = 0$, as is well known (Phillips, 1983; Hillier, 1990). When $\lambda = 0$ the density does not depend on T, which is enough to show that standard asymptotics fails when $\lambda = 0$—as is also now well understood. When $\lambda \neq 0$ the density (10) depends on the sample size T only through the term s_{ZZ}. Much more general versions of equation (10) appear in Sargan (1976, App. B), Phillips (1980, eqn. (14)), and Hillier (1985, eqn. (20)).

Remark 5. The key parameter λ can be expressed as $\lambda = i_{\tilde{\beta}\tilde{\beta}}/(1-\rho^2)$, where $i_{\tilde{\beta}\tilde{\beta}} = s_{ZZ}\gamma^2/\sigma_v^2$ is the Fisher information on the scaled parameter $\tilde{\beta} = \sigma_v\beta/\sigma_u$. Thus, λ increases with both $i_{\tilde{\beta}\tilde{\beta}}$ and ρ^2. Phillips (2006) and Forchini (2006) discuss limiting forms of the density under different assumptions on the joint behavior of $i_{\tilde{\beta}\tilde{\beta}}$ and ρ^2.

2.3. The Density of *e*

Because $w = e - \eta$, with e the scaled estimation error in b, we hope to find the density of w centered on, and concentrated near, the point $w = \eta$. However, it is easy to see that the moments of w (hence also of e and b) do not exist, so the factors that determine the location and concentration of the density are not imme-

diately apparent. Obviously, the terms γ, s_{ZZ}, and σ_v^2 influence the density only through λ, and (10) shows that the density of w depends only upon (λ, η). We seek to discover the roles played by these two parameters in determining the properties of the density.

The function ${}_1F_1(1,\frac{1}{2};x)$ is monotonic on $x > 0$, so the properties of the hypergeometric function in (10) are inherited from those of the term

$$c(w)^2 = \frac{(1 - \eta w)^2}{(1 + w^2)(1 + \eta^2)}. \tag{12}$$

If $\eta = 0$ (i.e., $\rho = 0$, so x_t in (2) is weakly exogenous), this has a single mode at $w = 0$ (the correct point) and is well behaved. But if $\eta \neq 0$ it has two turning points, a maximum at $w = -\eta$ (as hoped for) and a minimum at $w = \eta^{-1}$. The hypergeometric function with argument $\frac{1}{2}\lambda c(w)^2$ retains these characteristics, but damped or amplified depending on λ. We may state the following result.

PROPOSITION 3. *The locations of the turning points of the confluent hypergeometric function in (10) are determined entirely by the degree of endogeneity, as measured by* $\eta = \rho/\sqrt{1 - \rho^2}$. *The parameter* λ, *which we will see subsequently is the concentration parameter, depends on the properties of the instrument (through* s_{ZZ}*), the parameters of the first-stage regression (through* γ^2/σ_v^2*), and the degree of endogeneity (through* η*).*

The other term in the density (10)—$(1 + w^2)^{-1}$—is of course symmetric around *zero*, not $-\eta$, so its effect is to shift the mode of the density away from the correct point. In Figure 1 we display both the hypergeometric component of the density (10), that is, the term $\exp\{-\frac{1}{2}\lambda\}{}_1F_1(1,\frac{1}{2};\frac{1}{2}\lambda c(w)^2)$ (on the left), and the density itself (on the right), for several values of ρ ($\rho = 0.2$, 0.9, and 0.99) and, for each of these, for several values of the information quantity $i_{\tilde{\beta}\tilde{\beta}}$ ($i_{\tilde{\beta}\tilde{\beta}} = 0.2$ [solid line], $i_{\tilde{\beta}\tilde{\beta}} = 2.0$ [dotted line], and $i_{\tilde{\beta}\tilde{\beta}} = 20$ [dashed line]). Thus, when $\rho = 0.2$, $\lambda = 0.208$, 1.05, and 22.2 (Figure 1a), and these values are simply multiplied by 10 and 100, respectively, in Figures 1b and 1c. The functions are displayed as functions of the estimation error e, so the density should ideally be centered at, and concentrated near, zero.

As Figure 1 illustrates, the parameter $i_{\tilde{\beta}\tilde{\beta}}$ determines which of the two turning points dominates the behavior of the hypergeometric component and hence the degree to which the density is concentrated near zero. For small values of $i_{\tilde{\beta}\tilde{\beta}}$ the hypergeometric component of the density is quite flat, and in this case the position and shape of the density are determined almost entirely by those of the function $(1 + w^2)^{-1}$, whose mode (as a function of e) is at $e = \eta$ (*not* zero). As $i_{\tilde{\beta}\tilde{\beta}}$ increases the hypergeometric component has more impact, pulling the mode toward the correct point (zero) and increasing the concentration near the mode. As the figures show, for small to moderate values of $i_{\tilde{\beta}\tilde{\beta}}$, and values of ρ near 1, the density can certainly be bimodal.

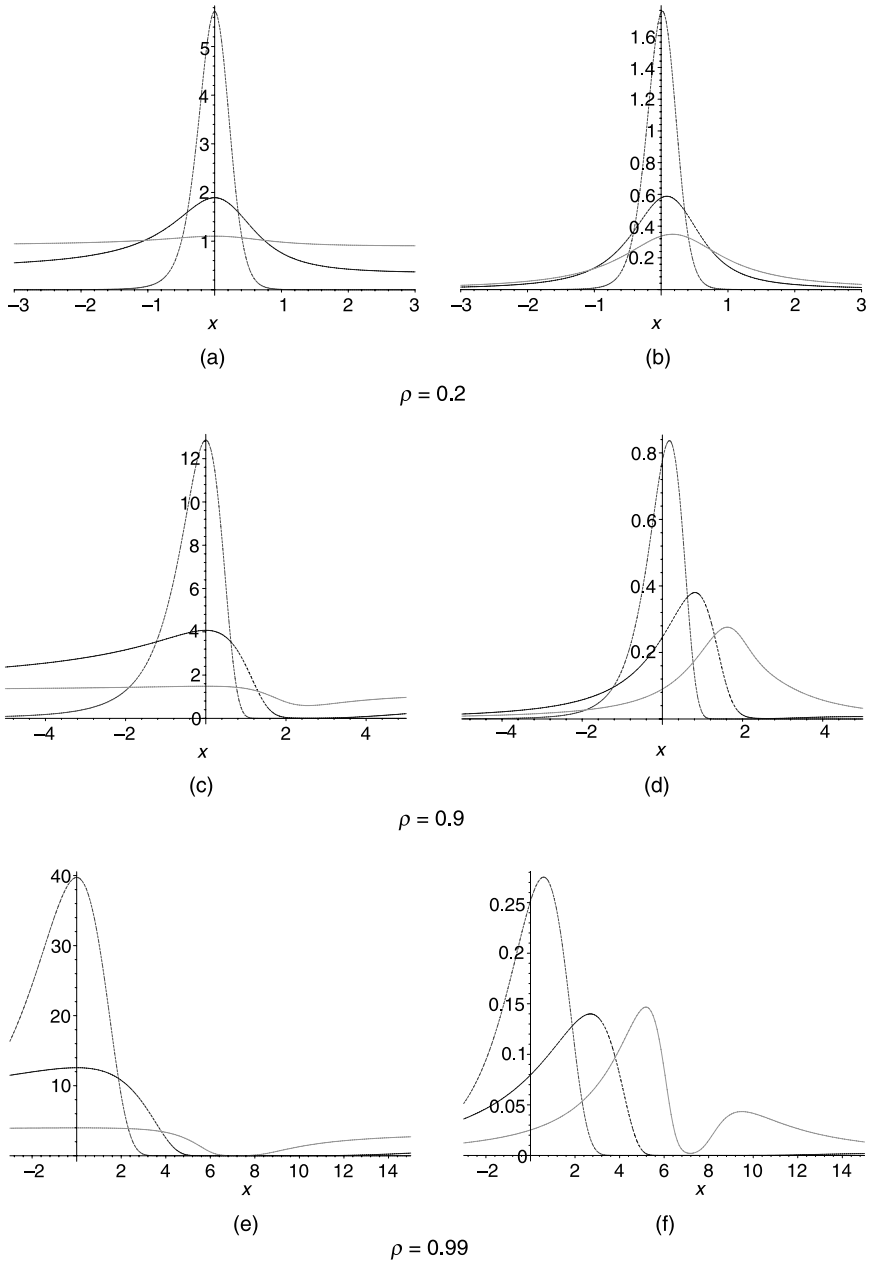

FIGURE 1. Hypergeometric component and density of e for various values of ρ and the information parameter $i_{\tilde{\beta}\tilde{\beta}}$.

2.4. The Direction Estimator

As noted in the Introduction, in two dimensions to declare interest in the ratio of the coordinates of an unknown point (α) is to declare interest in the *direction* of that point from the origin: the coordinate ratio is constant along lines through the origin. Thus, in the model (1) and (2), interest in β is equivalent to interest in the direction of the vector α in (3). But there are a variety of ways to parameterize direction: by the ratio of coordinates, by points on the circumference of the unit circle (or an ellipse), by the angle the line makes with one axis (i.e., its polar coordinate), and so on. These parameterizations are equivalent, so (at least locally) there are 1–1 functions relating the parameters that define them, and the MLEs for those parameters are automatically related to each other by those same functions.

Thus, for maximum likelihood, it is irrelevant which parameterization is used, but that is not necessarily so for other estimation methods. I argued in Hillier (1990) that the properties of the TSLS estimator (for instance) are distorted by the fact that it is not normalization invariant. In the present (highly simplified) setting the two methods are identical, so that problem does not arise. Nevertheless, the other point made in Hillier (1990)—that the exact properties of estimators are easier to interpret with one parameterization than another—certainly does apply. Indeed, this is no accident: the differential-geometric representation of the density of a statistic explained in Hillier and Armstrong (1999) makes it clear that the density of a given function of some random variables depends crucially on the geometric properties of the manifold (*statistic = constant*) but *not on how that manifold is parameterized.* That is, the induced properties of the statistic (its density) will reflect its intrinsic geometric structure but not the manner in which that structure is realized.

So, we can expect to see the "direction" quality of the estimator for β manifest in the distribution of b or, less directly, of w. And indeed it is: the term $c(w)^2$ in equation (12) is in fact $\cos^2(\psi)$, where ψ is the angle between the vectors

$$h = \begin{pmatrix} 1 \\ -w \end{pmatrix}(1 + w^2)^{-1/2} \quad \text{and} \quad \varphi = \begin{pmatrix} 1 \\ \eta \end{pmatrix}(1 + \eta^2)^{-1/2}, \tag{13}$$

which are both points on the unit circle ($h'h = \varphi'\varphi = 1$), so $\cos(\psi) = h'\varphi$. (Strictly, both h and φ lie on the unit semicircle with positive first coordinate—which is all that is needed to determine β—but it is convenient to think of h as being distributed on the whole circumference of the unit circle by *defining* $pdf(-h) = pdf(h)$, at the same time identifying $-\varphi$ with φ.) The random vector h is the "estimator" for φ induced by w, and we may transform $w \to h$ as explained in Phillips (1985) and Hillier (1990, Sec. 3). The term $(1 + w^2)^{-1}$ in $pdf(w)$—the term that in a sense distorts the properties of b—is the Jacobian of the inverse transformation $h \to w$; it arises entirely because of the particular parameterization chosen for the circle.

With this change of attitude (10) becomes the much nicer result

$$pdf(h) = \pi^{-1} \exp\left\{-\frac{1}{2}\lambda\right\}{}_1F_1\left(1, \frac{1}{2}; \frac{1}{2}\lambda\cos^2(\psi)\right), \tag{14}$$

where the density is evaluated with respect to the invariant measure on the unit circle (for details, see Muirhead, 1982, Ch. 2). The properties of h as an estimator of φ are nice: provided $\lambda \neq 0$, $pdf(h)$ is symmetric about the true points $\pm\varphi$, with modes at $\pm\varphi$. The concentration of the density of h about these modes—however that is measured—depends only on λ. Thus we have the following result.

PROPOSITION 4. *For the direction estimator h, η is a pure location parameter, in the sense that the modes of the density occur at the correct points $\pm\varphi$, which are determined entirely by η, and λ is a pure concentration parameter.*

Remark 6. Equation (14) is easily derived directly from the density of $\hat{\alpha}$. First set $\tilde{\alpha} = L'\hat{\alpha} \sim N(L'\alpha, (s_{ZZ})^{-1}I_2)$, with L as defined in equation (23), which follows, and note that $\bar{\alpha} = L'\alpha = (\gamma/\sigma_v)\binom{-\eta}{1}$. Then transform $\tilde{\alpha} \to (h, q)$, with $h = \tilde{\alpha}(\tilde{\alpha}'\tilde{\alpha})^{-1/2}$ and $q = \tilde{\alpha}'\tilde{\alpha}$ (the Jacobian is $\frac{1}{2}$), taking account of the sign convention discussed previously. Integration over $q > 0$ then yields equation (14). The points h indicate the direction of $\tilde{\alpha}$ by projecting onto the unit circle or equivalently that of $\hat{\alpha}$ by projecting onto the ellipse $\hat{\alpha}'\Omega^{-1}\hat{\alpha} = 1$.

In the totally unidentified case when $\lambda = 0$, h is uniformly distributed on the circle, reflecting the fact that h contains no information about φ, as is clearly correct (for more discussion of this, see Hillier, 1990). Figure 2 displays the density of h—supported on the unit circle—for several values of λ. For very small values of λ the density is almost uniform, whereas for large values of λ it is very sharply peaked at its modes $\pm\varphi$.

Remark 7. In the model (3) parameterized by (α, Ω), the Fisher information matrix is block diagonal with (α, α) block $s_{ZZ}\Omega^{-1}$. When the model is parameterized by (β, γ, Σ), as suggested by (1) and (2), the (partial) information on β is $i_{\beta\beta} = s_{ZZ}\gamma^2/\sigma_u^2$. When parameterized by first transforming α to $\bar{\alpha} = L'\alpha$ and then expressing $\bar{\alpha}$ in polar coordinates (δ, θ) ($\bar{\alpha}' = (\delta\sin(\theta)), \delta\cos(\theta))$, with θ the angle between $\bar{\alpha}$ and the horizontal axis, the partial information on θ is $i_{\theta\theta} = s_{ZZ}\alpha'\Omega^{-1}\alpha = \lambda$. It is clear from (14) that, however the points on the circle are parameterized (i.e., whatever statistic is used to indicate the direction of h), the induced density of the corresponding statistic will depend fundamentally on the information about the direction of α, $\lambda = i_{\theta\theta}$.

3. THE OVERIDENTIFIED CASE

If there are $k > 1$ instruments, equation (2) is replaced by

$$x_t = z_t'\gamma + v_t, \tag{15}$$

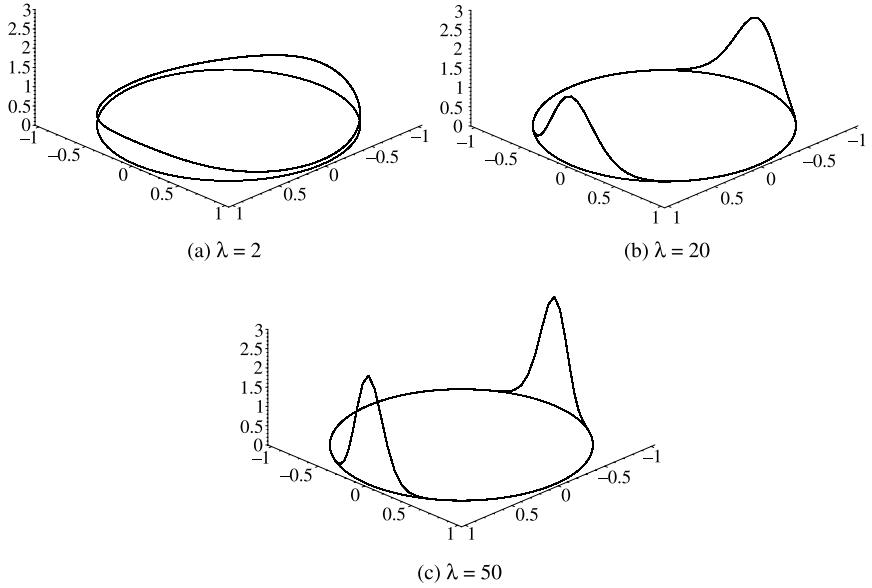

FIGURE 2. Densities of the direction estimator.

with z_t and γ $k \times 1$, and equation (3) is replaced by

$$Y|Z \sim N(Z(\gamma\beta, \gamma), I_T \otimes \Omega) = N(Z(\pi, \gamma), I_T \otimes \Omega), \tag{16}$$

say, where Z is the $T \times k$ matrix with tth row z_t'. Defining $(\hat{\pi}, \hat{\gamma}) = (Z'Z)^{-1} Z'(y, x)$ and $(a, g) = (Z'Z)^{1/2}(\hat{\pi}, \hat{\gamma})/\sigma_v$, the TSLS estimator is

$$b = (\hat{\gamma}'Z'Z\hat{\gamma})^{-1}\hat{\gamma}'Z'Z\hat{\pi} = (g'g)^{-1}g'a. \tag{17}$$

In this case this is *not* the LIML estimator for β, which must be treated separately.

3.1. TSLS

Defining e, w, and η as before, the conditional density of e given $g \neq 0$ is

$$e|g \sim N(\eta(1 - (g'g)^{-1}g'\bar{\gamma}), (g'g)^{-1}), \tag{18}$$

and $g \sim N(\bar{\gamma}, I_k)$, where $\bar{\gamma} = (Z'Z)^{1/2}\gamma/\sigma_v$. Hence,

$$pdf(w,g) = (2\pi)^{-(k+1)/2}(g'g)^{1/2} \times \exp\left\{-\frac{1}{2}\left[(g'g)(w + \eta(g'g)^{-1}g'\bar{\gamma})^2 + (g-\bar{\gamma})'(g-\bar{\gamma})\right]\right\}$$

$$= (2\pi)^{-(k+1)/2}(g'g)^{1/2}\exp\left\{-\frac{1}{2}\bar{\gamma}'\bar{\gamma}\right\}\exp\left\{-\frac{1}{2}g'g(1+w^2)\right\} \times \exp\{g'\bar{\gamma}(1-\eta w)\}\exp\left\{-\frac{1}{2}\eta^2(g'\bar{\gamma})^2/(g'g)\right\}, \quad \textbf{(19)}$$

and so

$$pdf(w) = (2\pi)^{-(k+1)/2}\exp\left\{-\frac{1}{2}\bar{\gamma}'\bar{\gamma}\right\}\int_{R^k}(g'g)^{1/2}\exp\left\{-\frac{1}{2}g'g(1+w^2)\right\} \times \exp\{g'\bar{\gamma}(1-\eta w)\}\exp\left\{-\frac{1}{2}\eta^2(g'\bar{\gamma})^2/(g'g)\right\}(dg). \quad \textbf{(20)}$$

These are the exact analogues of the previous results for the case $k = 1$. As before, the conditional density of w given g is unimodal. When $\gamma = 0$ it is clear from (20) that $pdf(w) \propto (1+w^2)^{-(k+1)/2}$ and is therefore concentrated near the origin (the *wrong* point), the more so the larger is k.

When $\gamma \neq 0$, integrating out g in (20) is slightly more complicated, but there is a device that simplifies the problem greatly and that, because it does not seem to be widely known, is worth discussing—variations of it are often extremely useful.

The key observation is that the integral in (20), regarded as a function of $\bar{\gamma}$, is invariant under two types of transformation on $\bar{\gamma}$. First, it is invariant under $\bar{\gamma} \to -\bar{\gamma}$ (because, on making this change, we can replace g by $-g$, leaving the integral unchanged). This means that the term $\exp\{g'\bar{\gamma}(1-\eta w)\}$ can be replaced by its average over the two possible signs for $\bar{\gamma}$:

$$[\exp\{g'\bar{\gamma}(1-\eta w)\} + \exp\{-g'\bar{\gamma}(1-\eta w)\}]/2 = {}_0F_1\left(\frac{1}{2}, \frac{1}{4}(g'\bar{\gamma})^2(1-\eta w)^2\right)$$

(exactly as in the case $k = 1$). The final two terms in the integral (20) can therefore be replaced by the double series

$$\sum_{i,j=0}^{\infty} \frac{\left[\frac{1}{4}(1-\eta w)^2\right]^i\left[-\frac{1}{2}\eta^2\right]^j}{i!j!\left(\frac{1}{2}\right)_i(g'g)^j}(g'\bar{\gamma})^{2(i+j)}.$$

But now, by a more general version of the same argument, the resulting integral is also invariant under the transformations $\bar{\gamma} \to H\bar{\gamma}$, with $H \in O(k)$, the group of $k \times k$ orthogonal matrices. This is because, on replacing $\bar{\gamma}$ by $H\bar{\gamma}$, we can then transform $g \to H'g$ (a transformation with unit Jacobian), leaving the integral exactly as it was. One implication of this is that the integral (and hence the density) depends on $\bar{\gamma}$ only through $\bar{\gamma}'\bar{\gamma}$, a maximal invariant under this group of transformations. But it also implies that we can replace $\bar{\gamma}$ by $H\bar{\gamma}$, regard H as a *random* matrix uniformly distributed on $O(k)$, and integrate over $O(k)$ *before* integrating out g, and this makes the evaluation of the integral much easier.

The integral over $O(k)$ in this process is of the form $\int_{O(k)}(g'H\bar{\gamma})^{2r}(dH)$, where (dH) denotes the uniform distribution on $O(k)$. This is the coefficient of $t^{2r}/(2r)!$ in the expansion of the generating function

$$\int_{O(k)} \exp\{t(g'H\bar{\gamma})\}(dH) = {}_0F_1\left(\frac{k}{2}, \frac{1}{4}t^2(g'g)(\bar{\gamma}'\bar{\gamma})\right).$$

(This is a well-known integral—a much more general version of it is given as Theorem 7.4.1 in Muirhead, 1982.) Thus,

$$\int_{O(k)} (g'H\bar{\gamma})^{2r}(dH) = (g'g)^r(\bar{\gamma}'\bar{\gamma})^r\left(\frac{1}{2}\right)_r \Big/ \left(\frac{k}{2}\right)_r. \tag{21}$$

These two steps therefore reduce the integrand in equation (20) to a function that depends on g only through $q = g'g$, and the integral can now be completed by elementary methods (it is essentially a standard Laplace transform), yielding the known (cf. Phillips, 1983, eqn. (3.45)) formula:

$$pdf(w) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{k}{2}\right)}(1+w^2)^{-(k+1)/2}\exp\left\{-\frac{\lambda}{2}\right\}$$
$$\times \sum_{j=0}^{\infty}\left\{\frac{\left(\frac{k-1}{2}\right)_j}{j!\left(\frac{k}{2}\right)_j}\right\}\left(\frac{1}{2}\lambda\rho^2\right)^j {}_1F_1\left(\frac{k+1}{2}, j+\frac{k}{2}; \frac{1}{2}\lambda c(w)^2\right), \tag{22}$$

where

$$\lambda = \bar{\gamma}'\bar{\gamma}(1+\eta^2) = \gamma'Z'Z\gamma/\sigma_v^2(1-\rho^2).$$

Note that this can again be expressed in the form $\lambda = i_{\tilde{\beta}\tilde{\beta}}/(1-\rho^2)$. We may therefore state the following proposition.

PROPOSITION 5. *In the overidentified case, the leading term* $(1 + w^2)^{-(k+1)/2}$ *in the density of the TSLS estimator is no longer simply the Jacobian of the transformation* $h \to w$. *The hypergeometric component,*

$$\exp\left\{-\frac{\lambda}{2}\right\}\sum_{j=0}^{\infty}\left\{\frac{\left(\frac{k-1}{2}\right)_j}{j!\left(\frac{k}{2}\right)_j}\right\}\left(\frac{1}{2}\lambda\rho^2\right)^j {}_1F_1\left(\frac{k+1}{2}, j+\frac{k}{2}, \frac{1}{2}\lambda c(w)^2\right),$$

although more complicated, is, as before, a monotonic function of $c(w)^2$, *and, as before, the location of its turning points is determined entirely by* η. *Otherwise, the parameters* (λ, η) *play the same roles as before. The sample size affects the density of the TSLS estimator only through the matrix* $Z'Z$ *that appears in the definition of* λ.

Remark 8.

(i) For the OLS estimator $b_0 = (x'x)^{-1}x'y$, it is easy to see from (4) that the conditional distribution of the standardized estimator $e_0 = (b_0 - \beta)\sigma_v/(\sigma_u\sqrt{1-\rho^2})$, given x, has exactly the form (18) with g replaced by the $T \times 1$ vector $\tilde{x} = x/\sigma_v$, which is $N(Z\gamma/\sigma_v, I_T)$ (so that $\bar{\gamma}$ is replaced by $Z\gamma/\sigma_v$). Because $\bar{\gamma}'\bar{\gamma} = \gamma'Z'Z\gamma/\sigma_v^2$, it follows that the density of $w_0 = e_0 - \eta$ has exactly the form (22) with k replaced by the sample size T. This result is well known (cf. Phillips, 1983), but its implications are perhaps not as widely appreciated as they might be. However, for a recent discussion see Kiviet and Niemczyk (2005).

(ii) It is clear again that the density of the TSLS estimator depends on the sample size only through $Z'Z$. And, from point (i), it follows that the properties of the TSLS estimator as $k \to \infty$ will mimic those of the OLS estimator as $T \to \infty$ if λ remains bounded as $T \to \infty$, as it does under so-called weak-instrument asymptotics. This agrees with recent work on "many weak-instrument asymptotics" by Chao and Swanson (2005, 2006), who show that the TSLS bias converges to that of the OLS estimator as T and k go to infinity (while k/T remains bounded) under weak-instrument asymptotics. See also Phillips (2006) and Newey (2004).

The presence of the term $(1 + w^2)^{-(k+1)/2}$—which becomes much more concentrated near the origin (the *wrong* point) as k increases—severely distorts the properties of the TSLS estimator when k is large. We shall see shortly that this effect is not present for the LIML estimator, which helps to explain its apparent superiority over TSLS. Notice too that, if $\lambda = 0$, only this term remains, so the density of e becomes concentrated near η, the probability limit of the scaled OLS estimation error when $\lambda = 0$. These observations again accord with related work on many-instrument asymptotics for the TSLS and LIML estimators (e.g., Morimune, 1983; Chao and Swanson, 2005, 2006; Newey, 2004), which shows that the LIML estimator is much better centered than the TSLS estimator, and

also with important recent results of Chamberlain (2005), which provide a decision-theoretic analysis of this model.

3.2. LIML

The LIML estimator for β maximizes the familiar ratio

$$r(\beta) = \frac{(y - x\beta)' M_Z (y - x\beta)}{(y - x\beta)' Z(Z'Z)^{-1} Z'(y - x\beta)}.$$

Let

$$L = \begin{bmatrix} 1/(\sigma_u\sqrt{1-\rho^2}) & 0 \\ -[\eta/\sigma_v + \beta/(\sigma_u\sqrt{1-\rho^2})] & 1/\sigma_v \end{bmatrix}, \tag{23}$$

so that $L'\Omega L = I_2$. Then define $\tilde{Y} = YL \sim N(Z\gamma(-\eta,1)/\sigma_v, I_T \otimes I_2)$, $S = \tilde{Y}' M_Z \tilde{Y}$, a 2×2 central Wishart matrix with $m = T - k$ degrees of freedom and covariance matrix I_2, and $(a, g) = (Z'Z)^{-1/2} Z'\tilde{Y} \sim N(\bar{\gamma}(-\eta,1), I_k \otimes I_2)$. Finally, put $R = (a, g)'(a, g)$. The supremum of $r(\beta)$ is the largest root of the equation $\det[S - fR] = 0$, and the LIML estimator for β is defined in terms of the corresponding characteristic vector.

Exact distribution theory for the LIML estimator is, even in the present simplified setup, much more complex than that for the IV (TSLS) estimator. Phillips (1984, 1985) pioneered the theory, but for present purposes it is more convenient to work from Hillier (1987, eqn. (29)), where I obtained the conditional density of (w, f) given R, where w is now the LIML estimator standardized as before and f is the largest root of $\det[S - fR] = 0$. In our present notation, that result yields, after transforming to h, the conditional result

$$pdf(h|R) = c_m (h'Rh)^{-1} |R|^{(m+1)/2} \int_{f>0} \exp\left\{-\frac{1}{2} f(\mathrm{tr}[R])\right\} \times f^{m-1} {}_1F_1\left(2, \frac{m+3}{2}; \frac{1}{2} f|R|/h'Rh\right) df\, dt, \tag{24}$$

where $c_m = [2^{m+1}\Gamma((m+3)/2)\Gamma(m/2)\Gamma(\frac{1}{2})]^{-1}$.

Remark 9. The matrix (a, g) here is a standardized version of the OLS estimator for the complete reduced form coefficient matrix (π, γ) in (16), unconstrained by the requirement that $\pi = \gamma\beta$. The matrix R (a maximal invariant under the group of transformations $(a, g) \to H(a, g)$, $H \in O(k)$) is a natural candidate for testing the general identification condition $rank(\pi, \gamma) = 1$. The conditioning in (24) is thus a generalized version of the conditioning $pdf(w|g)$ used earlier.

Notice that $pdf(h|R)$ depends directly on the sample size (through m) but is parameter free. Thus, the parameters enter the unconditional distribution only through the density of R, and this depends only on k and λ (see equation (A.3) in the Appendix). The marginal density of h is evidently

$$pdf(h) = \int_{R>0} pdf(h|R)pdf(R)(dR).$$

By some elementary manipulations this expression may be reduced to a form (equation (A.4) in the Appendix) from which the following proposition readily follows.

PROPOSITION 6. *For the LIML estimator all of the properties of direction estimator h given in Proposition 4 for the case $k = 1$ continue to hold in the overidentified case. Transforming to w introduces only the Jacobian term $(1 + w^2)^{-1}$. The density depends on the sample size both directly through $m = T - k$ and indirectly through λ.*

Although the formula for the density is considerably more complicated, qualitatively the properties of the LIML estimator in the overidentified case are very similar to those discussed earlier for the case $k = 1$.

4. SUMMARY AND CONCLUSIONS

Simple derivations of the densities of the TSLS and LIML estimators in a simplified structural equation, obtained by first conditioning on the first-stage regression estimator, have been given and the two key factors that determine their properties—λ and η—identified. The main conclusions are as follows.

- In the exactly identified case, the conditional distribution of the TSLS/LIML structural coefficient estimator, given the estimator of the first-stage regression coefficient, is well behaved: normally distributed with precision that is increasing in the conditioning statistic. Forchini and Hillier (2003) argue that there are grounds for conditioning on that statistic.
- The properties of the estimators are determined by the parameters $\lambda = i_{\theta\theta} = i_{\tilde{\beta}\tilde{\beta}}/(1 - \rho^2)$—the Fisher information on the direction of α in (3)—and η, a measure of the degree of endogeneity. For the interest parameter β, these interact to determine the location and concentration of the density.
- The problem is equivalent to that of estimating the direction of a two-dimensional vector. The properties of the coefficient estimators reflect this, and, for the LIML "direction estimator," η is a pure location parameter and λ a pure concentration parameter. For the TSLS estimator this is so in the exactly identified case but not otherwise.
- The possible bimodality of the density of the estimator for β is an artifact of the way in which "direction" is parameterized: in the exactly identified

case it is the Jacobian of the transformation that induces the (possibility of) bimodality. For the LIML estimator this effect is independent of the degree of overidentification, but that is not so for the TSLS and OLS estimators: the densities of the TSLS and OLS estimators are more strongly pulled toward the origin the higher the degree of overidentification (resp. sample size).

Evidently, it is the information parameter $\lambda = i_{\theta\theta}$ that properly measures the weakness or otherwise of the instrument z_t, insofar as it affects inference on the structural parameter, and this is a function of—as intuition surely says it should be—both the parameters of the first-stage regression and the degree of endogeneity. In Phillips (2006) a limit theory for the case $\lambda = \lambda_T \to 0$ is discussed for a closely related, but special, inference problem. Finally, we remark that a very similar analysis can be carried out for this model from a Bayesian point of view. As the authors point out, the results in Chao and Phillips (1998, 2002) have much in common with the sampling theoretic results discussed here.

REFERENCES

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

Chamberlain, G. (2005) Decision Theory Applied to an Instrumental Variables Model. Mimeo, Harvard University.

Chao, J.C. & P.C.B. Phillips (1998) Posterior distributions in limited information analysis of the simultaneous equations model using the Jeffreys prior. *Journal of Econometrics* 87, 49–86.

Chao, J.C. & P.C.B. Phillips (2002) Jeffreys prior analysis of the simultaneous equations model in the case with $n + 1$ endogenous variables. *Journal of Econometrics* 111, 251–283.

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

Chao, J.C. & N.R. Swanson (2006) Alternative approximations of the bias and MSE of the IV estimator under weak identification with an application to bias correction. *Journal of Econometrics*, forthcoming.

Forchini, G. (2006) On the bimodality of the exact distribution of the TSLS estimator. *Econometric Theory* 22, 932–946 (this issue).

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

Forchini, G. & G.H. Hillier (2005) Ill-Conditioned Problems, Fisher Information, and Weak Instruments. Cemmap Working paper CWP04/05.

Hillier, G.H. (1985) On the joint and marginal densities of instrumental variable estimators in a general structural equation. *Econometric Theory* 1, 53–72.

Hillier, G.H. (1987) Joint Distribution Theory for Some Statistics Based on LIML and TSLS. Cowles Foundation Discussion paper 840.

Hillier, G.H. (1990) On the normalization of structural equations: Properties of direction estimators. *Econometrica* 58, 1181–1194.

Hillier, G.H. & M. Armstrong (1999) The density of the maximum likelihood estimator. *Econometrica* 67, 1459–1470.

Kiviet, J.F. & J. Niemczyk (2005) The Asymptotic and Finite Sample Distributions of OLS and IV in Simultaneous Equations. UVA Econometrics Discussion paper 2005/01.

Maddala, G.S. & J. Jeong (1992) On the exact small sample distribution of the instrumental variable estimator. *Econometrica* 60, 181–184.
Marsaglia, G. (1965) Ratios of normal variables and the roots of sums of uniform variables. *Journal of the American Statistical Association* 60, 193–204.
Morimune, K. (1983) Approximate distributions of *k*-class estimators when the degree of overidentifiability is large compared to the sample size. *Econometrica* 51, 821–841.
Muirhead, R.J. (1982) *Aspects of Multivariate Statistical Theory*. Wiley.
Nelson, C. & R. Startz (1990) Some further results on the exact small sample properties of the instrumental variable estimator. *Econometrica* 58, 967–976.
Newey, Whitney K. (2004) Many Instrument Asymptotics. Mimeo, MIT.
Phillips, P.C.B. (1980) The exact distribution of instrumental variable estimators in an equation containing $n + 1$ endogenous variables. *Econometrica* 48, 861–878.
Phillips, P.C.B. (1983) Exact small sample theory in the simultaneous equation model. In M.D. Intriligator & Z. Griliches (eds.), *Handbook of Econometrics*, pp. 449–516. North-Holland.
Phillips, P.C.B. (1984) The exact distribution of LIML, part I. *International Economic Review* 25, 249–261.
Phillips, P.C.B. (1985) The exact distribution of LIML, part II. *International Economic Review* 26, 21–36.
Phillips, P.C.B. (1989) Partially identified econometric models. *Econometric Theory* 5, 181–240.
Phillips, P.C.B. (2006) A remark on bimodality and weak instrumentation in structural equation estimation. *Econometric Theory* 22, 947–960 (this issue).
Sargan, J.D. (1976) Econometric estimators and the Edgeworth approximation. *Econometrica* 44, 421–448.
Staiger, D. & J.H. Stock (1997) Instrumental variables regression 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.
Woglom, G. (2001) More results on the exact small sample properties of the instrumental variable estimator. *Econometrica* 69, 1381–1389.

APPENDIX: Details for the LIML Estimator

Let

$$H = (1+\eta^2)^{-1/2}\begin{bmatrix} -\eta & 1 \\ 1 & \eta \end{bmatrix} \in O(2), \tag{A.1}$$

so that, with $\tilde{R} = H'RH$, $\mathrm{tr}[R] = \mathrm{tr}[\tilde{R}]$ and $|R| = |\tilde{R}|$, whereas $h'Rh = h'H\tilde{R}H'h$. Defining $a = \tilde{R}_{11} > 0$, $d = |\tilde{R}|/\tilde{R}_{11}$, and $r = \tilde{R}_{11}^{-1}\tilde{R}_{12}$, we have that d is independent of a and r, $d \sim \chi^2(k-1)$, $r|a \sim N(0, a^{-1})$, and $a \sim \chi'^2(k,\lambda)$, whereas $\mathrm{tr}[\tilde{R}] = d + a(1+r^2)$, $|\tilde{R}| = ad$, and

$$\begin{aligned} pdf(h|a,d,r) = c_m(ad)^{(m+1)/2} &\int_{f>0}\int_{t>0} \exp\left\{-\frac{1}{2}f(d+a(1+r^2))\right\} \\ &\times \exp\left\{-\frac{1}{2}t(as^2 + 2scar + (d+ar^2)c^2)\right\} \\ &\times f^{m-1}\,{}_1F_2\left(2;1,\frac{m+3}{2};\frac{1}{4}da\,ft\right) df\,dt. \end{aligned} \tag{A.2}$$

Here we have put $h'H = (\sin(\theta), \cos(\theta)) = (s, c)$. Also,

$$pdf(a,d,r) = c_k \exp\left\{-\frac{1}{2}\lambda\right\} \exp\left\{-\frac{1}{2}(d + a(1+r^2))\right\}$$
$$\times a^{((k+1)/2)-1} d^{((k-1)/2)-1} {}_0F_1\left(\frac{k}{2}; \frac{1}{4}\lambda a\right) \tag{A.3}$$

with $c_k = [2^k \Gamma(k/2)\Gamma((k-1)/2)\Gamma(\frac{1}{2})]^{-1}$. As noted in the text, this depends only on k and λ. If $\lambda = 0$ it is clear at once that the distribution of h (and hence of all functions of it) is free of all parameters, so that h carries no information about those parameters.

Multiplying (A.2) by (A.3) and integrating out d and r is quite straightforward and leaves, after a trivial change of variable, the following expression for the density of h:

$$pdf(h) = c(m,k) \exp\left\{-\frac{1}{2}\lambda\right\} \int_{f>0}\int_{t>0}\int_{a>0} \exp\left\{-\frac{1}{2}a(1+f)(1+f+t)\right\}$$
$$\times a^{((m+k+1)/2)-1} f^{m-1} {}_2F_2\left(\frac{m+k}{2}, 2; \frac{m+3}{2}, 1; \frac{1}{2}aft\right)$$
$$\times {}_0F_1\left(\frac{k}{2}; \frac{1}{4}\lambda a(1+f+tc^2)\right) da\, df\, dt, \tag{A.4}$$

where $c(m,k) = 2^{(m+k+1)/2}\Gamma(\frac{1}{2})\Gamma((m+k)/2)c_m c_k$. The qualitative properties of the density (Proposition 6) follow easily from this expression. In particular, the density depends only on $c^2 = \cos^2(\theta)$, so it is symmetric about $\pm\varphi$, as before; λ is a pure concentration parameter, and the density is uniform on the sphere if $\lambda = 0$. The remaining steps needed to produce a formula for the density are elementary but messy. We resist the temptation to reproduce them here; they can be found in Hillier (1987).