



Department of Economics
University of Southampton
Southampton SO17 1BJ
UK

**Discussion Papers in
Economics and Econometrics**

1999

This paper is available on our website
<http://www.soton.ac.uk/~econweb/dp/dp99.html>

Conditional Inference for Possibly Unidentified Structural Equations

Giovanni Forchini and Grant Hillier
University of York and University of Southampton

May, 1999

Abstract

The possibility that a structural equation may not be identified casts doubt on the measures of estimator precision that are normally used. We argue that the observed identification test statistic is directly relevant to the precision with which the structural parameters can be estimated, and hence argue that inference in such models should be conditioned on the observed value of that statistic (or statistics).

We examine in detail the effects of conditioning on the properties of the ordinary least squares (OLS) and two-stage least squares (TSLS) estimators for the coefficients of the endogenous variables in a single structural equation. We show that: (a) conditioning has very little impact on the properties of the OLS estimator, but a substantial impact on those of the TSLS estimator; (b) the conditional variance of the TSLS estimator can be very much larger than its unconditional variance (when the identification statistic is small), or very much smaller (when the identification statistic is large); and (c) conditional mean-square-error comparisons of the two estimators favour the OLS estimator when the sample evidence only weakly supports the identification hypothesis, can favour TSLS slightly when that evidence is moderately favourable, but there is nothing to choose between the two estimators when the data strongly supports the identification hypothesis.

1 Introduction

A nagging concern for econometricians for many years has been the possibility that structural models may not be identified (Sims (1980)), or may be only partially identified (Phillips (1989)), or may involve "weak instruments" (Staiger and Stock (1997)). Recent literature on both exact distribution theory and asymptotics for this model suggests that this concern is fully justified.

The work of Phillips (1983), (1989) and Hillier (1985), (1990) has made it clear that if the exclusion restrictions imposed on (the exogenous variables in) the structural equation are spurious (i.e., the structural equation is totally unidentified), then the densities of the ordinary least squares (OLS), two-stage least squares (TSLS), and limited information maximum likelihood (LIML) estimators of the coefficients of the endogenous variables depend only on the error covariance matrix, and not on the structural parameters, so that none of these statistics contains information about those parameters. Since also standard asymptotic theory breaks down when the model is unidentified, identification must be presumed if these statistics are to be interpreted as useful estimators.

Motivated by such concerns, recent literature on the distributions of the TSLS and LIML estimators has started to focus on intermediate situations where the equation of interest is neither formally identified, nor totally unidentified. To bridge the gap, Phillips (1989) has introduced the idea of partially identified models. These are models for which some, but not all, of the parameters are identified after a rotation of coordinates in the space of both the endogenous and exogenous variables. Choi and Phillips (1992) have found that in such models the distributions of the usual estimators of both the identified and the unidentified coefficients of the endogenous variables are mean- and covariance matrix- mixed-normal. Although the density of the estimator of the unidentified parameters does not depend on those parameters, it does depend on the identified parameters. As far as asymptotics are concerned, the estimator of the identified parameters is consistent, but conventional asymptotics does not apply. Moreover, the estimator of the unidentified parameters converges in law to a non-degenerate distribution, but this is different from the small-sample distribution.

In a different attempt to study intermediate cases Staiger and Stock (1997) define a notion of weakly identified models. Such models have two characteristics: they are formally identified for all finite sample sizes, but the covariance

between the instruments and the included endogenous variables is assumed to be in a $O(T^{-1/2})$ neighbourhood of zero, where T is the sample size. Staiger and Stock (1997) show that standard asymptotics fails for such models, and that the asymptotic distributions of the usual estimators have the same structure as their exact distributions under normality (i.e., mixed-normal). These results lead Staiger and Stock to stress the importance of reporting the test statistic for identification, but they do not indicate how the value of that statistic should modify inference, if at all.

A related recent development is based on the observation that inference on structural parameters is essentially a generalisation of the Fieller-Creasy problem (see Wallace (1980)). Scheffé (1970) defines a confidence set to be improper if it has positive probability of being the entire parameter space. In the Fieller-Creasy problem, this probability corresponds to the probability that what is essentially an identification test statistic is small. Generalising Koschat (1987) and Gleser and Hwang (1987), Dufour (1997) has shown that in a potentially unidentified structural model valid confidence sets must be improper in Scheffé's sense, and argues in favour of Anderson-Rubin-type confidence sets (a generalisation of the Fieller-Creasy solution). It is important to note that, in all three of these papers, "potentially unidentified" really means that, although the model may be formally identified, it is impossible to rule out a priori models that are arbitrarily close to being unidentified.

In this paper we argue that the correct way to take account of the possibility that such a model may be formally unidentified, or arbitrarily close to being so, is to explicitly take account of the sample evidence on the identification of the model by conditioning on an identification test statistic (or statistics). In particular, we argue that the reported precision of the parameter estimates should be that in the conditional distribution of the estimator given the observed value of an identification test statistic, and not the unconditional precision. The analysis is exact rather than asymptotic, and supplements the work of Staiger and Stock (1997) by suggesting exactly how the identification test statistic should affect inference procedures for these models.

The paper is organised as follows. The crux of the paper will be the argument that, in inference problems of this type, conditional measures of estimator precision are more relevant than unconditional measures. For expositional purposes we use the Fieller-Creasy problem as a model for this argument, and only later

(in Section 3) apply it to the more complex problem of structural estimation. The main argument is presented in Section 2. In Section 3 we introduce the structural model, and generalise the key parts of the argument from Section 2 to this more complicated case. In Section 4 we present explicit conditional results for the OLS and TSLS estimators of the coefficients of the endogenous variables, and give new measures of the precision of these estimators that properly reflect the possibility that the model may not be identified. Among our conclusions in this section is the fact that the conditional density of the OLS estimator is quite insensitive to the identification statistic, while that of the TSLS estimator is sensitive to it. Thus, OLS can dominate TSLS, or vice versa, depending on the data actually obtained. Clearly, this conclusion, and its implications for applied work, are quite at odds with received opinion on inference procedures for the structural model.

2 The Argument for Conditioning

2.1 Conditioning: Post-Data Precision

Reporting the results of an inference procedure, whether it be inference about the values of unknown parameters, or a decision about whether certain propositions of interest about the model under study are correct or false, entails two inter-related components: reporting the results of the procedure as applied to the data at hand, and reporting some measure of the likely precision that can be attributed to those results. For parametric estimation problems - our concern here - precision is usually indicated by reporting a confidence set for the parameter(s) of interest, together with its confidence level, or by reporting the (typically, estimated) variance or mean squared error (or approximations thereto) of the estimator chosen.

Now, the standard frequentist measures of the precision of an inference have, until recently, relied entirely on pre-data assessments of the procedure used, obtained by averaging (some relevant property of the procedure) over the entire sample space. But, as Lindsay and Li (1997) remark: "After the data is observed, however, the actual "postexperimental" error associated with that particular sample become more relevant". Thus, for an estimation problem: "...the average error is an attribute of an estimator,...., whereas the postexperimental error is an attribute of an estimate,...." (See also Goutis and Casella (1995) for

a recent discussion). The suggestion implicit here is that certain features of the sample actually obtained may be pertinent to the assessment of the precision of the estimate, but of course are immaterial to the pre-data (average) properties of the estimator.

Any attempt to assess the likely precision of an estimate - that is, to assess the likely precision of the procedure in the sample actually available - clearly (except to a committed Bayesian) must involve some sort of averaging process, but, equally, must hold certain aspects of the sample actually used fixed. In other words, whatever precision measure is used, it must, if it is to measure a property of the estimate (rather than the estimator) be conditional on certain aspects of the sample remaining fixed at their observed values. Thus, we take it as self-evident that any post-data measure of the precision of an estimate must be conditional. The problem, both in practice and from a philosophical point of view, is precisely what to condition on. That is, how can one identify events that are pertinent to the precision achieved by an inference procedure?

The only situations in which there seems to be widespread agreement among (frequentist) statisticians - both that conditioning is sensible, and upon what to condition - are those in which there is an exact ancillary statistic (cf. Cox (1958), Efron and Hinkley (1978), Barndorff-Nielsen (1980)). This agreement is almost certainly attributable to the fact that, under suitable assumptions, it is straightforward to show that the Fisher information based on the conditional distribution of (say) the maximum likelihood estimator given the ancillary is identical to that of the full sufficient statistic, while the marginal density of the estimator must yield smaller Fisher information. That is, conditioning on the ancillary recovers information that would otherwise be lost. Even here the problem is difficult (see the collected papers by D. Basu edited by Ghosh (1988)), and outside this fairly narrow class of problems study of the problem has barely begun. It is important to notice, though, that the converse of this advice is not implied, or even suggested, by its adherents. The motivation for conditioning at all, whether on an ancillary statistic or not, is undeniably a concern - similar to that expressed by Lindsay and Li - to obtain a more relevant measure of the inferential precision actually achieved in the sample (see also Barndorff-Nielsen's comments on the paper by Efron and Hinkley (1978), and the discussion in Chapter 2 of Cox and Hinkley (1974)).

We do not intend to enter into the wider debate on these issues here. However,

we shall advocate, for a particular class of problems to be described shortly, conditional measures of precision, and for precisely the reason that applies in the wider problem: to obtain a more relevant measure of the precision achieved in the sample actually available. As noted above, the problem is to identify aspects of the data that are pertinent to the precision of the estimate, and we shall argue that, in the types of models we are considering, the structure of the problem clearly identifies which events affect precision. We begin by discussing the simplest possible example of the type of problem we shall be concerned with - the Fieller-Creasy problem.

2.2 The Fieller-Creasy Problem

The Fieller-Creasy problem is a celebrated, apparently straightforward, inference problem that produces "paradoxical" assessments of precision, namely improper confidence sets (see Wallace (1980) for an historical survey of the problem). The problem contains virtually all the essential features of the structural model that is our main concern, and is also closely related to the linear calibration (inverse regression) problem (Hoadley (1970), Dobrigal, Fraser, and Gebotys (1987), Gleser and Hwang (1987)), errors-in-variables regression, some principal component inference problems, and inference on ratios of regression parameters.

Assume that the 2×1 vectors x_i ($i = 1, \dots, n$) are independent $N(1, \frac{1}{2}I_2)$, and that we are interested in either the ratio of means $\bar{A} = \bar{x}_1/\bar{x}_2$; or in the direction of the vector 1 , parameterised by the angle, \hat{A} , when 1 is expressed in polar coordinates in the usual way. The vector of sample means, \bar{x} , and

$$s^2 = \sum_{i=1}^n [(x_{1i} - \bar{x}_1)^2 + (x_{2i} - \bar{x}_2)^2];$$

are jointly sufficient for $(1; \frac{1}{2})$, $\bar{x} \gg N(1; (\frac{1}{2}=n) I_2)$; \bar{x} is independent of s^2 ; and $s^2 = \frac{1}{2} \gg \hat{A}^2 (2(n-1))$. The maximum likelihood estimator for \bar{A} is $\hat{\bar{A}} = \bar{x}_1/\bar{x}_2$, but the (unconditional) variance of $\hat{\bar{A}}$ does not exist. Fieller (1954), and Creasy (1954) considered the problem of constructing a confidence interval for \bar{A} , and the standard "Fieller solution" is based on the observation that

$$n(\bar{x}_1/\bar{x}_2 - \bar{A})^2 = \frac{1}{2} [1 + \bar{A}^2] \gg F(1; 2(n-1)); \quad (1)$$

where $\frac{1}{2} = s^2/2(n-1)$. However, this solution entails the "paradox" that the confidence interval it produces is:

- (a) the entire real line if $k\hat{x}k^2 = (\hat{x}_1^2 + \hat{x}_2^2) < c = \frac{1}{2}F_{\alpha}(1; 2(n-1))$, where $F_{\alpha}(\nu_1; \nu_2)$ is such that $\Pr\{F(\nu_1; \nu_2) < F_{\alpha}(\nu_1; \nu_2)\} = 1 - \alpha$;
- (b) the interior of a finite interval if $\hat{x}_2^2 > c$, and
- (c) the exterior of a finite interval if $k\hat{x}k^2 > c$ and $\hat{x}_2^2 < c$.

Thus, it is entirely possible for, say, the 90% confidence interval for \tilde{A} to consist of the whole real line, and the (unconditional) expected length of the interval is infinite.

Shapiro (1970), James, Wilkinson, and Venables (1974), Dobrigal, Fraser, and Gebotys (1987), and Koschat (1987) all treat the problem in terms of the interest parameter \tilde{A} , rather than \tilde{A} . Essentially the same paradox arises although, as we shall see, there is a subtle difference between the two versions of the problem.

Koschat (1987), for this problem, and Gleser and Hwang (1987) for a wider class of closely related problems, have shown that problems of this type have the property that every confidence set with non-zero confidence coefficient must have positive probability of being the entire parameter space, and hence must be improper in Shapiro's (1970) sense. Hence, the difficulty is not peculiar to the Fieller solution. Dufour (1997) has extended these results to the structural model, and other models of interest in econometrics.

2.3 Interest Parameters and Critical Sets

Now, notice that in the above Fieller problem, there is certainly no difficulty in obtaining an always-bounded confidence set for the underlying parameter vector θ . The region (sphere) defined by the acceptance region for a likelihood ratio test:

$$F = n(\hat{x} - \theta)'(\hat{x} - \theta) = 2\frac{1}{2} < F_{\alpha}(2; 2(n-1)) \tag{2}$$

is the most obvious candidate. Apart from being the likelihood ratio statistic, F is the (unique) maximal invariant (under the group of transformations $G = \{aH; a > 0; H(2 \times 2) \text{ orthogonal}\}$ acting on the statistics $(\hat{x}; s^2)$ by:

$$((\hat{x} - \theta); s^2) \rightarrow (aH(\hat{x} - \theta); a^2s^2):$$

Thus, confidence regions for θ based on F characterise the entire class of invariant regions.

The important point, for us, about the regions (2) is that the Fieller problem is naturally embedded in a model for which no paradox arises. In view of this

observation, it seems clear that the “Fieller paradox” derives from the properties of the mapping [underlying parameters] β ! [interest parameter], and not from any intrinsic property of the embedding model. We shall see shortly that this observation also applies to the structural model, but before doing so we elaborate on its implications for the Fieller problem itself, where the situation is simpler.

If $\theta = \theta(\beta)$ is an everywhere continuous one-to-one function of β , the image of any closed bounded set of β -values is a closed bounded set of θ -values, so that the (proper) confidence set (2) for β naturally induces a proper confidence set for θ (with the same confidence level). In this sense, inferential results for β are sufficient for inference about any everywhere continuous and 1-1 function of β ; and it seems reasonable to assert that, in these circumstances, there is no essential difference between the inference problem for β and that for (any such function) θ . Of course, actual measures of precision, like estimator variances, or Fisher information, change under reparameterisation, but, although seldom stated explicitly, estimator loss functions would usually embody exactly this invariance under smooth reparameterisations, and so reflect the main idea.

On the other hand, the key characteristic of (both versions of) the Fieller problem is that the mapping from the underlying parameter β to the interest parameter \tilde{A} (or \hat{A}) is not everywhere continuous and one-to-one. In the case of the interest parameter \tilde{A} , the mapping $\beta \mapsto (\tilde{A}; \beta_2)$ (from \mathbb{R}^2 to \mathbb{R}^2) is continuous and 1-1 everywhere except along the β_1 axis, where \tilde{A} can take any value. In the case of the interest parameter \hat{A} , the mapping $\beta \mapsto (\hat{A}; \beta_2)$ (where $\hat{A} = k\beta$) is continuous and 1-1 everywhere except at the origin ($\beta = 0$), where \hat{A} can take any value. Dufour (1997) calls such subsets of the parameter space non-identification subsets, but we shall refer to them as critical sets. Notice that the critical set depends on the definition of the interest parameter(s) in terms of the parameters of the embedding model.

The Fieller solution (1) is easily obtained from (2). The line $\beta_1 \pm \tilde{A}\beta_2 = 0$ intersects the sphere $k\beta \pm \beta_1 k^2 < f$ just if $(\beta_1 \pm \tilde{A}\beta_2)^2 = (1 + \tilde{A}^2) < f$: The induced (Fieller) confidence set for \tilde{A} is thus the set for which this line intersects the sphere, and is a finite interval if the sphere does not cross the β_1 -axis (i.e., $\beta_2^2 > f$), is the exterior of a finite interval if the sphere crosses the β_1 axis but does not include the origin, (i.e., $\beta_2^2 < f$ but $k\beta k^2 > f$), and is the entire real line if the sphere contains the origin (i.e. $k\beta k^2 < f$) - see Figure 1 below, where these three cases are labelled (a), (b), and (c) respectively. It is clear, too,

that, for either version of the Fieller problem, the region (2) must intersect the critical set for some values of $(\hat{x}; s^2)$, unless it has confidence level zero. That is, every confidence set for \tilde{A} (or \hat{A}) induced by (2) must be improper unless it has confidence level zero.

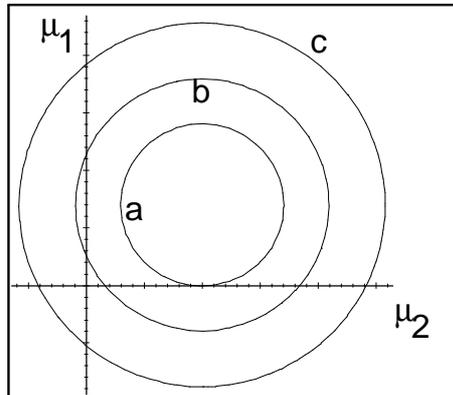


Figure 1

We now generalise these observations on the Fieller problem to an arbitrary model in which the interest parameter can take any value on a subset of the full parameter space for the embedding model.

An Index of Precision

Generalising the Fieller problem, let $p(x; \mu)$ denote the model density for either the sample data or the minimal sufficient statistic, with $\mu \in M \subseteq \mathbb{R}^k$, and consider a reparameterisation $\phi : M \rightarrow \mathbb{R}^k$. We assume that the interest parameter ϕ_1 is a subvector of ϕ , and, as above, that there is a non-empty subset $C \subseteq M$ such that, for $\mu \in C$, ϕ_1 can take any value in some set \mathbb{R}^1 of dimension greater than one. That is, for $\mu \in C$; the mapping ϕ becomes a correspondence, not a function. We call C the critical subset of M (for the interest parameter ϕ_1). Note that membership of C is an attribute of μ ; not of ϕ_1 . Let $\hat{\mu}$ denote the MLE for μ ; and let

$$LR_c(x) = \frac{n}{\mu; p(x; \mu)} = \frac{o}{\hat{\mu}} > c \quad ; 0 < c < 1; \quad (3)$$

denote the confidence region for μ based on the acceptance region for a likelihood ratio test. The confidence level is determined by the choice of c , with smaller values of c corresponding to larger confidence levels.

In practice the interest parameter may be defined in terms of some subset of the full parameter vector μ . If so, $LR_c(x)$ must be modified by replacing the

numerator by its maximum value over parameters not involved in the definition of θ_1 , and the denominator by its maximum over the values of all parameters. For expository purposes we work with the simplest case. We assume that the regions (3) are bounded with probability one. Obviously, the acceptance region based on any testing principle could be used in place of (3), with obvious modifications to the argument that follows.

For fixed data x , and a fixed critical set C , there are two possible outcomes for the region (3):

(a) $LR_c(x)$ intersects C for all, or all but very large, values of c . That is, $LR_c(x)$ is "close to" C for all but very large values of c . In this case the induced confidence region for θ_1 will be the entire set \mathcal{I}_1 for all but very small confidence levels, and it seems reasonable to conclude that, with this particular data x , θ_1 is determined with "low precision";

(b) $LR_c(x)$ and C may be disjoint for all but very small values of c (of course, they cannot be disjoint for all c if C is non-empty). That is, $LR_c(x)$ is "remote from" C for all but very large confidence levels. In this case it seems reasonable to conclude that, with this particular data x , θ_1 is determined with "high precision".

If we take as given the standard (frequentist) position that confidence sets are an adequate, if somewhat philosophically problematic, representation of the precision of an inference, these remarks suggest that the location of $LR_c(x)$ in relation to the critical set C has, for problems of this type, a direct bearing on the precision with which θ_1 can be learnt from the data available. Moreover, they immediately suggest an intuitively reasonable "post-data index of precision" for inference about θ_1 in the type of problem we are considering, namely:

$$k(\theta_1; x) = 1 - c_0; 0 < k < 1; \quad (4)$$

with c_0 the smallest value of c in (3) for which $LR_c(x)$ does not intersect C . For, if c_0 is near 1 only very "small" regions (i.e., low confidence levels) for μ would be compatible with a non-trivial region for θ_1 (with the given data), and this indicates low precision in the ability of the data to inform about θ_1 , while c_0 near zero would mean that very large likelihood ratio confidence regions (with high confidence levels) would be (with this data) compatible with a non-trivial interval for θ_1 .

Remarks

1. If C is empty, so that the mapping $\mu \mapsto k$ is everywhere 1-1, $c_0 = 0$, because $LR_c(x)$ can be made as large as we like without intersecting C . Thus, in this case, $k = 1$ for any x , and the precision with which any everywhere 1-1 function of μ can be located is identical to that with which μ itself can be located, as seems natural.
2. If, with the data available, $LR_c(x)$ intersects C for all c , $c_0 = 1$; and $k = 0$, indicating that, with this data, there is no prospect of locating θ_1 .
3. Although the acceptance region for any test on μ could be used as the underlying confidence set for μ , that based on the likelihood ratio in (3) has the advantage that different choices for the mapping $\mu \mapsto k$ that leave the interest parameter θ_1 invariant will leave k invariant.
4. k is evidently a monotonic increasing function of the maximum confidence level (for likelihood ratio based confidence sets for μ) that (with the given data) is compatible with a non-trivial confidence set for the interest parameter θ_1 . This maximum compatible level (mcl) could itself be used as an index of precision (see Table 1 below).
5. The precision for inference about the interest parameter, θ_1 , has been defined so as to depend on the entire vector μ , because membership of the critical set C is an attribute of μ , not θ_1 . In this respect k reflects information in the data about both "nuisance parameters" and the "interest parameter". This, it seems to us, is natural when the true source of the inference problem has been recognised.
6. Since, under regularity conditions, the power of the likelihood ratio test goes to unity as $n \rightarrow \infty$, it is clear from (5) that, provided $\mu \notin C$, $k \rightarrow 1$ in probability as $n \rightarrow \infty$. That is, the precision with which θ_1 can be located will, if $\mu \notin C$, approach 1 as the sample size increases. Just as importantly, the converse is also easily seen to be true: if $\mu \in C$, the precision, k , goes to zero in probability as $n \rightarrow \infty$.

In both versions of the Fieller problem above, the second component of the transformation $\mu \mapsto k$ (in the first case, 1_2 , in the second, $\frac{1}{2}$) is constant (zero) on the critical set C . In the general case, the transformation $\mu \mapsto k = (\theta_1; \theta_2)$ can usually be chosen so that θ_2 vanishes on C . The implication of this remark, then, is that the "non-interest" parameter θ_2 is, far from being a "nuisance parameter", actually the key to the precision with which θ_1 can be determined: it would be quite misleading to report precision measures for θ_1 that did not depend on θ_2 ,

or on the sample evidence on the value of θ_2 . In fact - as we shall see shortly - this evidence is encapsulated in the precision index, k .

Now, as initially defined, k measures how remote (or otherwise) likelihood ratio based confidence regions for μ are from the critical set C . It is easy to see, however, that k is completely determined by the likelihood ratio test statistic for testing the hypothesis $H_0 : \mu \in C$ against an unrestricted alternative. For, denoting this test statistic by $LR(C)$, we have,

$$k = 1 - \sup_{\mu \in C} \frac{p(x; \mu)}{p(x; \hat{\mu})} = 1 - LR(C) \quad (5)$$

Hence, the post-data precision for θ_1 , as measured by $k(\theta_1; x)$, will be large just if the likelihood ratio test statistic for testing $\mu \in C$ is small (suggesting rejection), and vice versa.

Examples: For the two versions of the Fieller-Creasy problem we have:

1. Interest parameter \tilde{A} : In this case the critical set $C = \{f_1 : \theta_2 = 0\}$, and the region $F < F_{\alpha}$ in (2) cuts the θ_1 -axis unless $F_{\alpha} = f_2 = 2 = n k^2 = 2 \frac{1}{4}$. Hence, for inference about the ratio of means, \tilde{A} :

$$k(\tilde{A}; x; s^2) = 1 - \frac{f_2}{2(n-1)}$$

2. Interest parameter \hat{A} : The largest value of F_{α} such that the region $F < F_{\alpha}$ in (2) does not intersect $C = \{f_0\}$ is clearly $f_1 = n k^2 = 2 \frac{1}{4}$. Hence, for inference about the direction of θ_1 :

$$k(\hat{A}; x; s^2) = 1 - \frac{f_1}{n-1}$$

Note that, as one would expect, k depends on the sample size, n . Clearly, k differs in these two versions of the Fieller problem precisely because the interest parameter, and hence the critical set C , differs.

Some values of the precision index k for the Fieller problem with interest parameter \tilde{A} are given in Table 1 below. As an example, if the observed value of f_2 is 1, and $n = 20$, the maximum confidence level (mcl) for which the confidence region (2) just intersects the θ_1 -axis is .389, and the corresponding index of precision for \tilde{A} , k , is .405. Both numbers indicate that \tilde{A} is poorly determined with this data.

Table 1: Values of k and maximum confidence levels

n = 10								
f_2	6	5	4	3	2	1	.5	.2
k	.944	.914	.865	.786	.651	.417	.240	.104
mcl	.925	.890	.835	.750	.612	.385	.220	.094

5% critical value for LR(C) (f_2): 4.41

n = 20							
f_2	8	6	4	2	1	.5	.2
k	.978	.947	.865	.641	.405	.230	.099
mcl	.973	.938	.851	.623	.389	.220	.095

5% critical value for LR(C) (f_2): 4.41

The argument above establishes that, for any problem of the type we are interested in, LR(C) is clearly relevant to the precision with which, with the given data, the interest parameter θ_1 can be located, from a frequentist point of view. Various other treatments of the problem can also be invoked to the same end. First, for the Fieller-Creasy problem (and hence the inverse linear regression/calibration problems), Hoadley (1970) has shown that LR(C) emerges naturally as an index of precision when the problem is viewed from a Bayesian perspective. In particular, under Hoadley's assumptions, the posterior density is more sharply peaked the larger is LR(C). Second, using a ...ducial argument, Dobrigal, Fraser, and Gebotys (1987) suggest confidence sets (for the angle, \hat{A} , in the polar coordinate version of the Fieller problem) conditioned on the value of LR(C). The length of the interval suggested decreases with LR(C), but the interval is always bounded. Finally, it is not difficult to show that, for the Fieller problem in either form, the shape of the likelihood function near its maximum depends heavily on LR(C), so that adherents to the strict likelihood principle (including Edwards (1970)) would use LR(C) to indicate how accurately the MLE is determined. Thus, whatever one's statistical persuasions, for problems of this type the observed value of the likelihood ratio test statistic for testing $H_0 : \mu \in C$ is clearly relevant to the precision with which the data is revealing about the interest parameter θ_1 .

To summarise, we have argued that:

- (a) Inference problems, like the Fieller problem, and the structural equation model, can usually be embedded in a model for which no paradox arises;
- (b) The problem arises because the reparameterisation involved in the definition of the interest parameter entails a critical set in the embedding parameter space on which the interest parameter can take any value in a set of dimension greater than one;
- (c) The likelihood ratio test statistic for testing whether, in the embedding model, and with the data actually available, the parameter lies in the critical set, is a natural measure of the ability of the sample to inform about the interest parameter;
- (d) Far from being a “nuisance”, the sample evidence on the non-interest parameter component of the reparameterised model is directly relevant to the precision with which the interest parameter itself can be located.

Our conclusion is that the observed likelihood ratio test statistic for testing membership of the critical set is directly relevant to the precision with which the interest parameter can be estimated, and that it therefore makes no sense to average over values of that statistic that have not occurred. That is, we conclude that any sensible assessment of the properties of estimates of the interest parameter(s) must be made conditional on such a test statistic (or other relevant indicators of critical set membership).

Before applying this argument to the structural equation model, we end this section with some further remarks on the “Fieller solution” that are also pertinent in the structural equation context.

2.4 Further Remarks on Fieller’s solution

In seeking a basis for making confidence statements about an interest parameter, say θ_1 , it is common practice to seek a pivotal function, $Q(x; \theta_1)$ say, that is a function of the data and the interest parameter, but not a function of nuisance parameters. That is, a function whose (unconditional) distribution (induced by the density of x , $p(x; \mu)$) does not depend on unknown parameters (cf. Basu (1981), reprinted in Ghosh (1988)). If Ω_1 is the parameter space for θ_1 , and X the sample space for x , the inequality $Q(x; \theta_1) < q$ (say) defines a subset E_q , say, of $X \in \Omega_1$, and the confidence region for θ_1 induced by Q and the observed

value of x is the x -section of E_q , $E_q^x = \{x; \theta_1 \in E_q\}$. The confidence level $\hat{\alpha}(q)$ corresponding to q is $\hat{\alpha}(q) = \Pr F(Q; x; \theta_1) < q$, which, if Q is pivotal, does not depend on μ . The set E_q^x is the confidence set estimator for θ_1 , and the pair $\{E_q^x; \hat{\alpha}(q)\}$ together are usually interpreted as indicating the precision with which θ_1 is determined by the data. The motivation for using a pivot in the construction of $\{E_q^x; \hat{\alpha}(q)\}$ is, of course, that $\hat{\alpha}(q)$ can, in principle, be known exactly.

As Example 6 in Basu (1981) makes clear, the interpretation of E_q^x as a confidence set estimator for θ_1 (and hence an indicator of the precision with which θ_1 is determined by the data) depends on the knowledge that x contains information on θ_1 , but certainly does not imply that it does. In other words, a minimal necessary condition for the set E_q^x (induced by the pivotal quantity Q) to reveal anything about θ_1 is that it is known that the observed x contains (in some other sense) information about θ_1 . At the very least, the density $p(x; \mu)$ must vary with θ_1 under the reparameterisation $\mu \rightarrow \theta_1$.

Now, in the cases we are concerned with here, it is known that, in the underlying parameterisation of the model $p(x; \mu)$, there is a critical set C for which θ_1 can take any value. Hence it is known that (whatever definition of "information" is adopted) the data may be uninformative about θ_1 , but it is not known whether $\mu \in C$ or not. We now consider the implication of this for the Fieller solution itself.

The interpretation of the Fieller-Creasy problem above throws the Fieller solution into doubt, because, as we shall now show, the statistic upon which it is based is either pivotal and uninformative about the interest parameter, or not pivotal with respect to the family of underlying models $p(x; \mu)$, and $\hat{\alpha}(q)$ can be arbitrarily small. We discuss only the case of the interest parameter \bar{A} ; similar remarks apply in the case of the interest parameter \bar{A} .

The quantity $Q(x; s^2; \bar{A}) = n(x_1 - \bar{A}x_2)^2 / (s^2(1 + \bar{A}^2))$ on which the Fieller solution is based is pivotal for the interest parameter \bar{A} , provided $1 \notin C$, in the sense that, for $1 \notin C$ satisfying $x_1 - \bar{A}x_2 = 0$ (but not otherwise), $Q \gg F(1; 2(n-1))$. What can be said about the distribution Q when $\mu \in C$? It seems to us that there are two possibilities:

(1) If it were true that $x_1 - \bar{A}x_2 = 0$ implied $x_1 = 0$, as is the case under the reparameterisation $\bar{A} = (x_1/x_2)$ used above, then it would remain true that $Q \gg F(1; 2(n-1))$ for $\mu \in C$. This point of view could be sustained if we

could claim to know that, for every member of the family of densities $p(x; \mu; \sigma^2)$, $\mu = 0$ implies $\sigma = 0$; and this in fact is the maintained hypothesis in both the calibration problem and the structural equation model. Note, however, that this interpretation of the problem is clearly not implied by merely declaring the interest parameter to be $\tilde{\mu} = \mu = \sigma$.

Taking this point of view, though, implies that, when $\mu \in C$, (a) $p(x; \mu; \sigma^2)$, when reparameterised in terms of $(\tilde{\mu}; \sigma^2)$, does not depend on $\tilde{\mu}$; (b) Q is ancillary for all real $\tilde{\mu}$; (c) $\tilde{\mu}$ is arbitrary in this parameterisation. These are all simply different manifestations of the same phenomenon: that the transformation $\mu \rightarrow (\tilde{\mu}; \sigma^2)$ is not one-to-one when $\mu = 0$. Clearly, since it is not known whether $\mu \in C$ or not, the "confidence set" induced by Q has nothing whatever to say about either the value of $\tilde{\mu}$, or the precision with which it has been located by the data.

(2) If, instead, one takes the view that, when $\mu \in C$, μ is arbitrary, then, for $\mu \in C$, Q has the non-central $F^0(1; 2(n-1); \lambda)$ distribution, with non-centrality parameter $\lambda = n\mu^2/\sigma^2 = \tilde{\mu}^2/\sigma^2$, for every real $\tilde{\mu}$. Hence, under this interpretation, Q , is not pivotal with respect to the entire family of distributions $p(x; \mu)$: its distribution depends on whether $\mu \in C$ or not. In fact, when $\mu \in C$, the "confidence level" $\hat{\alpha}(q) = \Pr\{FQ < q\}$ can be arbitrarily small, because the non-centrality parameter λ can be arbitrarily large. It seems clear that, under this interpretation of the problem, it would be quite unreasonable to claim that the "confidence set" induced by Q , and its "confidence level" (calculated assuming $\mu = 0$), have, by themselves, anything to (unambiguously) say about the interest parameter $\tilde{\mu}$.

Thus, both interpretations of the Fieller problem leave the "Fieller solution" in doubt. Exactly analogous difficulties arise in the interpretation of Anderson-Rubin confidence regions for the coefficients of the right-hand-side endogenous variables in the structural equation model - the "solution" favoured by Dufour (1997). The point, of course, is that the sample does contain information on whether or not $\mu \in C$, and this information is ignored when the inference is based on $Q(x; s^2; \mu)$ (and its "confidence level") alone (although it is reflected in the "confidence set" that this produces).

Scheffé's (1970) advice is to report the Fieller interval if it is bounded, and to declare that nothing has been learnt about $\tilde{\mu}$ if not - a kind of informal conditioning argument. However, it is not difficult to show that the conditional confidence

level achieved by this procedure is strictly less than the nominal level implied in (1). A second possibility would simply be to report the value of the maximum likelihood estimator, \hat{A} , together with our suggested measure of precision, k , or, equivalently, the likelihood ratio test statistic for testing $H_0 : \mu \in C$. This is essentially the advice offered by Staiger and Stock (1997) for the structural equation model, although without the interpretation we have given here. For precision measures based on confidence sets, therefore, the suggestion would be to report the value of k , or, equivalently, the largest confidence level for which the confidence region for μ in the underlying model just intersects the critical set C .

For point estimation problems our suggestion is to report the conditional variance of the estimator of interest, or, more completely, its conditional density, conditional on the observed value(s) of the relevant identification test statistic(s). We turn now to an investigation of the implications that this advice has for inference in the single structural equation model.

3 The structural equation model

We consider a single structural equation written without explicit normalization:

$$Y - \mu = Z_1 \alpha + u; \quad (6)$$

where Y is a $T \times (n + 1)$ matrix of endogenous variables, Z_1 is a $T \times k_1$ matrix of exogenous variables, and μ and α are, respectively, $(n + 1) \times 1$ and $k_1 \times 1$ vectors of parameters. The reduced form corresponding to (6) is

$$Y = Z_1 \beta + Z_2 \gamma + V; \quad (7)$$

where Z_2 is a $T \times k_2$ matrix of exogenous variables not included in the structural equation, β and γ are matrices of parameters of dimension $k_1 \times (n + 1)$, $k_2 \times (n + 1)$ respectively. We assume throughout that $k_2 \leq n$. The rows of V are assumed to be independent normal vectors with mean zero and common $(n + 1 \times n + 1)$ covariance matrix

$$\Sigma = \begin{pmatrix} \mu & & \\ & \Gamma_{11} & \Gamma_{21} \\ & \Gamma_{21}' & \Gamma_{22} \end{pmatrix};$$

where Γ_{11} , Γ_{21} and Γ_{22} are respectively (1×1) , $(n \times 1)$ and $(n \times n)$ matrices of

parameters.

Compatibility of the structural equation (6) and reduced form (7) requires the existence of some $\beta_M \neq 0$, such that

$$\beta_M \gamma = 0; \quad (8)$$

where $\beta_M = \beta$, $V \beta_M = u$. Note that (8) implies $\text{rank}(\gamma) \leq n$. If $\text{rank}(\gamma) = n$, equation (8) uniquely determines the direction of β_M (i.e., β_M is restricted to lie in a one-dimensional space): If $\text{rank}(\gamma) < n$, β_M can be written as a linear combination of the $n - \text{rank}(\gamma)$ basis vectors spanning the space orthogonal to the space spanned by the rows of γ . Thus, in this case β_M lies in a space of dimension greater than one. Our assumptions about $\text{rank}(\gamma)$ will be discussed shortly.

The structural equation is usually normalized by setting $\beta_M = (1; \gamma_0)'$, and $\gamma = (\gamma_1; \gamma_2)'$, so that (8) becomes

$$\gamma_1 + \gamma_2 \gamma_0 = 0; \quad (9)$$

and the structural equation (6) has the form

$$y_1 = Y_2 \gamma_0 + Z_1 \alpha + u; \quad (10)$$

Note that this normalisation implies $\text{rank}(\gamma) < n + 1$; and γ_0 is uniquely determined if the rank of γ_2 is n . Equation (10) is identified if $\text{rank}(\gamma_2) = n$; and is totally unidentified if $\gamma_2 = 0$. In all other cases we have a partially identified structural equation, where some of the parameters are identified after a rotation of coordinates in the space of the endogenous variables (Phillips (1989)). It is well known (Phillips (1983), Hillier (1985)) that the densities of the OLS, the TSLS, and the LIML estimator of γ_0 (and β_M) are free of any of the parameters in (10) and the estimators are therefore uninformative about them, when the structural equation is totally unidentified. Moreover, in this case, the densities of the TSLS and the LIML "estimators" do not depend on the sample size, so that conventional asymptotics for these "estimators" break down.

In the remainder of the paper we will consider the classical normalization given in equation (10). This is the analogue of the Fieller problem with interest parameter \tilde{A} . Similar results can be derived for the alternative normalisation $\beta_M \gamma_0 = 1$; corresponding to the interest parameter \hat{A} in the Fieller problem. It will be assumed throughout that the structural equation (10) is

formally identified, but that points in β_2 -space arbitrarily close to the critical set $C = \{\beta_2 : \text{rank}(\beta_2) < ng\}$ cannot be ruled out a priori. As shown by Gleser and Hwang (1987) and Dufour (1997), the model can be totally uninformative about the interest parameter if there exists a sequence of points in the β_2 -space converging to some point in the critical set. Exactly as for the Fieller-Creasy problem discussed in the previous section, the sample evidence on the distance of β_2 from the critical set reflects how well (or how poorly) β_0 can be located with the data actually available. And, as argued above, this suggests that the relevant post-data properties of estimates of β_0 (including measures of precision) are those conditional on that evidence.

We consider the OLS and the TSLS estimators of β_0 in (10), which can both be written in the form:

$$b = (Y_2^0 P Y_2)^{-1} Y_2^0 P y_1$$

with, in the OLS case, $P = P_{Z_1}$, where $P_A = I - A(A^0 A)^{-1} A^0$ for any matrix A , and in the case of TSLS, $P = P_{Z_1} - P_{Z_1} P_{Z_2}$, where $Z = (Z_1; Z_2)$. Joint minimal sufficient statistics for β_0 and β_2 in (7) are:

$$\begin{aligned} \hat{\beta}_0 &= (Z_1^0 Z_1)^{-1} Z_1^0 Y; \\ \hat{\beta}_2 &= (Z_2^0 P_{Z_1} Z_2)^{-1} Z_2^0 P_{Z_1} Y; \\ S &= Y^0 P_Z Y; \end{aligned}$$

and these remain minimal sufficient for any β_0 with $\text{rank}(\beta_0) > 0$. These statistics are independent of each other, and

$$\begin{aligned} \hat{\beta}_0 &\gg N_{k_1}(\beta_0) + (Z_1^0 Z_1)^{-1} Z_1^0 Z_2 \beta_2; (Z_1^0 Z_1)^{-1} - - -; \\ \hat{\beta}_2 &\gg N_{k_2}(\beta_2); (Z_2^0 P_{Z_1} Z_2)^{-1} - - -; \\ S &\gg W_{n+1}(\hat{A} - k_2; -); \end{aligned}$$

where $\hat{A} = T - k_1$.

Partitioning $\hat{\beta}_2 = \begin{pmatrix} \hat{\beta}_{21} \\ \hat{\beta}_{22} \end{pmatrix}$, and S conformably with $\hat{\beta}_2$,

$$S = \begin{pmatrix} S_{11} & S_{21}^0 \\ S_{21} & S_{22} \end{pmatrix};$$

inference about β_2 can be based on the matrix pivot

$$F_{\beta_2} = S_{22}^{-1} \hat{\beta}_{22} \hat{\beta}_{22}^0 Z_2^0 P_{Z_1} Z_2 \hat{\beta}_{22} S_{22}^{-1}; \quad (11)$$

which has the matrix-variate F-distribution (Muirhead (1982) Theorem 10.4.1.).

Using the acceptance region for the likelihood ratio test, a confidence region for β_2 can be constructed by finding all values of β_2 for which

$$j|I_n + F_{\beta_2}j < c: \tag{12}$$

This is the analogue of (2) for the Fieller problem. The confidence region for β_2 based on (12), $P = \{ \beta_2 : j|I_n + F_{\beta_2}j < cg \}$; intersects the critical set $C = \{ \beta_2 : \text{rank}(\beta_2) < ng \}$ only if c is larger than $\min_{\beta_2 \in C} j|I_n + F_{\beta_2}j$. Let $f_1 \geq f_2 \geq \dots \geq f_n$ be the ordered eigenvalues of

$$F_0 = S_{22}^{-1/2} \hat{\beta}_2^0 Z_2^0 P_{Z_1} Z_2 \hat{\beta}_2^0 S_{22}^{-1/2}: \tag{13}$$

It is straightforward to check that the region (12) intersects the critical set C just if $f_1 < c$, and that the post data index of precision for the interest parameter β_0 (defined in (4)) is, in this case:

$$k_{\beta_0; \hat{\beta}_2; S_{22}} = 1 + j (1 + f_1)^{i/2}: \tag{14}$$

The quantity $(1 + f_1)^{i/2}$ is the likelihood ratio statistic for testing the null hypothesis that $\text{rank}(\beta_2) = n - j$ against the alternative that $\text{rank}(\beta_2) = n$.

In the case $n = 1$ the argument in Section 2 suggests that inference on β_0 should be made conditional on the observed value of f_1 . In particular, the precision (variance) reported for an estimate of β_0 should be that in the conditional density of the estimator given the observed value of f_1 . In the case $n > 1$ the same argument (based on the confidence region for β_2 induced by the acceptance region for the likelihood ratio test) would also suggest conditioning on f_1 alone. However, the likelihood ratio test principle is, in the case $n > 1$, only one of a number of plausible candidates for constructing confidence regions for β_2 : there is, in this case, no unique optimal invariant test. In fact, we now argue that, for the case $n > 1$, a case can be made for conditioning on all n eigenvalues of F_0 .

The problem of testing $H_0 : \beta_2 = \beta_2^0$ (a particular value of β_2) can be reduced by sufficiency and the fact that the hypothesis does not involve β_0 to tests based on $(\hat{\beta}_2; S_{22})$; where $\hat{\beta}_2 = (Z_2^0 P_{Z_1} Z_2)^{-1} \hat{\beta}_2^0$. But, for tests based on $(\hat{\beta}_2; S_{22})$, the problem is invariant (in the sense discussed in Muirhead (1982), Chapter 6) under the group of transformations $G = \{ f(j; E) : j \geq 2 O(k_2); E \in GL(n) \}$ (where $O(n)$ denotes the group of $n \times n$ orthogonal matrices, and $GL(n)$ the general linear

group of $n \times n$ non-singular matrices), with group operation $(j_1; E_1)(j_2; E_2) = (j_1 j_2; E_1 E_2)$, acting on the space of statistics $(j_2; S_{22})$ by $(j; E)(j_2; S_{22}) = (j j_2 E^0; E S_{22} E^0)$, and with induced group of transformations on the parameter space given by $(j; E)(j_2; -22) = (j j_2 E^0; E -22 E^0)$, where $j_2 = (Z_2^0 P_{Z_1} Z_2)^{\frac{1}{2}} j_2$. Under the group of transformations G , a maximal invariant is $(f_1; \dots; f_n)$, where f_1, \dots, f_n are the eigenvalues of F_0 in (13). Moreover, the distribution of $(f_1; \dots; f_n)$ depends only on the eigenvalues of $\alpha = -\frac{i}{22} j_2^0 j_2 - \frac{i}{22}$; the maximal invariant under the induced group of transformations on the parameter space. (Muirhead (1982), Theorem 6.1.12).

Since every invariant test depends only on the maximal invariant, and the likelihood ratio test is just one member of this class, it seems preferable in the case $n > 1$ to condition on the full maximal invariant rather than on any particular scalar function of $(f_1; \dots; f_n)$. Thus, in the case $n > 1$, we shall condition on all n roots $(f_1; \dots; f_n)$ of F_0 . In Appendix A we derive the conditional densities of the OLS and TSLS estimators for β_0 in (10), conditional on the full maximal invariant $(f_1; \dots; f_n)$. Since the exact results for the general case of $n + 1$ endogenous variables are difficult to interpret, in the next Section we analyse in detail the results for the case $n = 1$; for which the identification test statistic is simply f_1 .

4 Conditional Results for the case $n = 1$

4.1 Conditional Densities and Moments

In this section we analyse the consequences of conditioning on f_1 for the OLS and TSLS estimators of β_0 in (10), assuming $n = 1$. Note that, when $n = 1$, $f_1 = j_2^0 Z_2^0 P_{Z_1} Z_2 j_2 = S_{22}$ is a scalar, and we shall denote this simply by f . Ideally we would want to obtain the analogous results for the LIML estimator as well, but these have so far proved intractable. To simplify the notation, but without loss of generality, we employ the standardizing transformations described in Phillips (1983). For either estimator, b , for β_0 , define the transformed statistic

$$r = (-\frac{1=2}{22} b j_2 - \frac{i}{22} j_2^0 j_2) =! ; \tag{15}$$

and the transformed parameter:

$$\beta = (-\frac{1=2}{22} \beta_0 j_2 - \frac{i}{22} j_2^0 j_2) =! ; \tag{16}$$

while, for the TSLS estimator:

$$\text{pdf}_{\text{TSLS}}(r|f) = [B(\frac{1}{2}; \frac{\lambda}{2})]^{-1} (f=(1+f))^{\frac{1}{2}} (1 + \frac{fr^2}{1+f})^{-\frac{\lambda+1}{2}}; \quad (23)$$

Theorem 2 Conditional Moments: For non-negative integers p and q, let

$${}_p H_q(z) = \frac{{}_1F_1(\frac{\lambda}{2}; \frac{\lambda}{2}; \frac{k_2}{2} + q; z)}{{}_1F_1(\frac{\lambda}{2}; \frac{\lambda}{2}; z)}; \quad (24)$$

and

$${}_p H_q^{\alpha}(z) = \frac{{}_2F_2(\frac{\lambda}{2}; \frac{\lambda}{2}; \frac{3}{2}; \frac{1}{2}; \frac{k_2}{2} + q; z)}{{}_1F_1(\frac{\lambda}{2}; \frac{\lambda}{2}; z)}; \quad (25)$$

In the following expressions, $z = \frac{f}{1+f}$:

(i) the conditional means are:

$$E_{\text{TSLS}}(r|f) = \frac{2}{k_2} {}_0H_1(z); \quad (26)$$

$$E_{\text{OLS}}(r|f) = \frac{f}{(1+f)} E_{\text{TSLS}}(r|f) = \frac{2z}{k_2} {}_0H_1(z); \quad (27)$$

(ii) the conditional variances are:

$$\text{Var}_{\text{OLS}}(r|f) = (\frac{\lambda}{2})^{-1} [{}_1H_0(z) + \frac{2z^{-2}}{k_2} {}_1H_1^{\alpha}(z)] [\frac{2z}{k_2} {}_0H_1(z)]^2; \quad (28)$$

$$\text{Var}_{\text{TSLS}}(r|f) = (f(\frac{\lambda}{2})=(1+f))^{-1} [{}_1H_0(z) + \frac{2z^{-2}}{k_2} {}_1H_1^{\alpha}(z)] [\frac{2z}{k_2} {}_0H_1(z)]^2; \quad (29)$$

(iii) the conditional mean-square-errors are:

$$\text{MSE}_{\text{OLS}}(r|f) = z^{-2} + (\frac{\lambda}{2})^{-1} [{}_1H_0(z) + \frac{2z^{-2}}{k_2} {}_1H_1^{\alpha}(z)] [\frac{4z^{-2}}{k_2} {}_0H_1(z)]; \quad (30)$$

$$\text{MSE}_{\text{TSLS}}(r|f) = z^{-2} + (f(\frac{\lambda}{2})=(1+f))^{-1} [{}_1H_0(z) + \frac{2z^{-2}}{k_2} {}_1H_1^{\alpha}(z)] [\frac{4z^{-2}}{k_2} {}_0H_1(z)]; \quad (31)$$

Remarks

1. When the structural equation (10) is totally unidentified the OLS estimator is independent of f , but this is not the case for the TSLS estimator. In general, conditioning affects the properties of both the OLS and the TSLS estimators. More precisely, conditioning makes the functional forms of the densities of the OLS and TSLS estimators different (even though their unconditional densities have the same functional form), and thus has an impact on the choice between these estimators (see below).
2. The leading terms in the densities of the OLS and TSLS estimators (i.e., equations (22) and (23) respectively) are both proportional to a Student-t density with $\hat{\lambda}$ degrees of freedom, so that integer (conditional) moments exist up to order $\hat{\lambda} - 1$, and the variances are $(\hat{\lambda} - 2)^{-1}$ and $(f(\hat{\lambda} - 2) = (1 + f))^{-1}$ respectively. For the TSLS estimator unconditional integer moments exist only up to order $k_2 - 1$.
3. As in the case of the unconditional densities (see Phillips (1983)), the conditional densities $\text{pdf}_{\text{OLS}}(r|f)$ and $\text{pdf}_{\text{TSLS}}(r|f)$ are not symmetric about \bar{r} , except when $\bar{r} = 0$.
4. From equation (27) it is clear that $|jE_{\text{OLS}}(r|f)| < |jE_{\text{TSLS}}(r|f)|$, that is, the conditional mean of the OLS estimator is always closer to the origin than that of the TSLS estimator, by a factor that depends on f . We shall see shortly that, as this result suggests, the OLS estimator can certainly conditionally dominate the TSLS estimator for small values of both \bar{r} and f : Mean-squared-error comparisons of the two estimators are given in the next subsection.
5. As f goes to infinity, for fixed parameter values, the densities and the MSE's of the OLS and TSLS estimators tend to the same function. This is evidence that if there is a clear sign that the model is identified, then the OLS and TSLS estimators are equivalent in small samples. This, again, stresses the importance of using the identification test statistic to give a feeling for the robustness of the results to the choice of the estimator.

4.2 Numerical properties

The conditional densities of the OLS and TSLS estimators are characterized by the three known quantities $(\hat{A} = T^{-1} \sum_{i=1}^T k_i, k_2, f)$, and the two unknown parameters (σ^2, τ^2) : Figures 2 and 3 illustrate the effects of conditioning. Plots of the densities given in Theorem 1 have been drawn for $\tau^2 = 0.35$, $\sigma^2 = 0.5$ and $T = 20$, $k_1 = 3$, $k_2 = 4$ (this choice of values for the parameters is based on the results by Anderson, Morimune and Sawa (1983)). To avoid values of f which are unlikely (with this value of σ^2), we choose four values $(f_i, i = 1; \dots; 4)$, $f_0 = 0; f_5 = 1$, such that $\Pr\{f_i < f < f_{i+1}\} = 0.2$ for $i = 0; \dots; 4$: The figures look very similar for other values of σ^2 , and for other values of τ^2 .

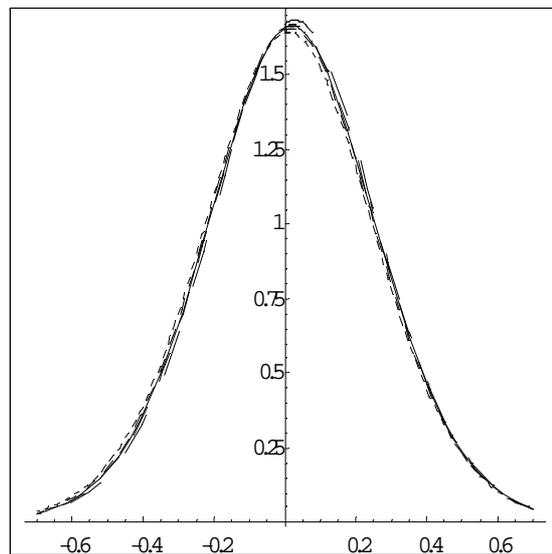


Figure 2: Marginal and Conditional Densities: OLS Estimator

The figures show the marginal densities with a solid line. The dashed lines represent the conditional densities, with the finest line corresponding to the density conditional on f_1 , the coarsest line to the density conditional on f_4 . Figure 2 displays the conditional and marginal densities for the OLS estimator. Clearly, the properties of the OLS estimator are not significantly affected by conditioning on the identification test statistic. Figure 3 shows the densities for the TSLS estimator. In this case the impact of conditioning is appreciable. It is clear that conditioning on the identification test statistic has an imperceptible effect on the mean of the TSLS estimator, but quite a dramatic effect on its variance. This

again supports the interpretation of the identification test statistic as index of precision.

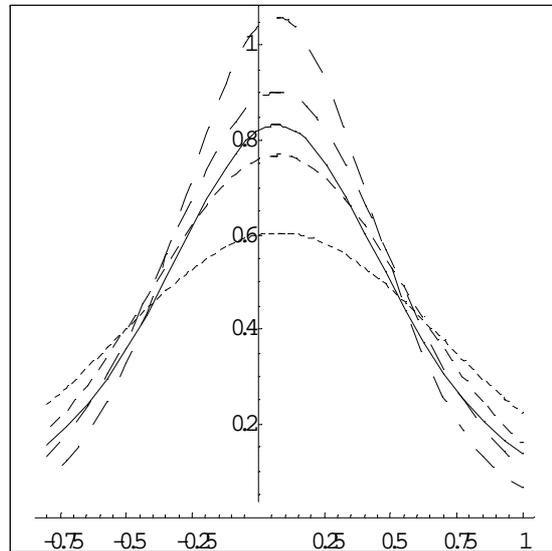


Figure 3: Marginal and Conditional Densities: TSLS Estimator

To quantify the difference between the conditional and the unconditional variances we report in Tables 2 and 3 the ratio of the conditional to the unconditional variances, i.e. $\frac{\text{Var}(r|f)}{\text{Var}(r)}$, for both estimators and for various values of ρ and f . Specifically, for f_i , $i = 0; \dots; 10$ such that $\Pr(f_i < f < f_{i+1}) = 0.1$; $i = 0; \dots; 9$; $f_0 = 0$, $f_{10} = 1$. The other parameters have been set at the same values as before, i.e. $T = 20$, $k_1 = 3$, $k_2 = 4$. Table 2 shows that the difference between $\text{Var}_{OLS}(r|f)$ and $\text{Var}_{OLS}(r)$ is negligible. On the other hand, Table 3 reinforces the impression gained from Figure 3 that for TSLS the difference between $\text{Var}_{TSLS}(r|f)$ and $\text{Var}_{TSLS}(r)$ can be very large, especially for small values of the noncentrality parameter. Notice that small observed values of f imply, as expected, lower conditional precision for the estimator (relative to the unconditional), but also that larger observed values of f imply higher conditional precision. It is also clear that for small values of f the conditional precision of the TSLS estimator is poor, but for large f it is quite good.

Table 2: Ratio of Conditional to unconditional variance: OLS

	$\lambda = 0:05$	$\lambda = :5$	$\lambda = 5$	$\lambda = 10$	$\lambda = 25$
f_1	1.004	1.032	1.115	1.112	1.080
f_2	1.003	1.023	1.059	1.057	1.044
f_3	1.003	1.015	1.024	1.024	1.020
f_4	1.00	1.008	0.997	0.998	1.002
f_5	1.00	1.000	0.973	0.976	0.987
f_6	0.999	0.993	0.951	0.956	0.974
f_7	0.997	0.985	0.931	0.938	0.960
f_8	0.997	0.975	0.910	0.918	0.946
f_9	0.996	0.963	0.885	0.894	0.930

Notes: each entry is calculated as $\frac{\text{Var}_{\text{OLS}}(r|f)}{\text{Var}_{\text{OLS}}(r)}$. $\Pr(f_i < f < f_{i+1}) = 0:1$.
 $T = 20, k_1 = 3, k_2 = 4, \rho = 0:35$.

Table 3: Ratio of Conditional to unconditional variance: TSLS

	$\lambda = 0:05$	$\lambda = :5$	$\lambda = 5$	$\lambda = 10$	$\lambda = 25$
f_1	1.821	1.810	1.533	1.365	1.185
f_2	1.193	1.196	1.159	1.131	1.078
f_3	0.919	0.915	0.974	1.006	1.013
f_4	0.746	0.744	0.854	0.919	0.966
f_5	0.623	0.625	0.765	0.851	0.927
f_6	0.529	0.531	0.691	0.794	0.893
f_7	0.451	0.455	0.626	0.742	0.861
f_8	0.381	0.385	0.567	0.691	0.828
f_9	0.312	0.316	0.503	0.634	0.790

Notes: each entry is calculated as $\frac{\text{Var}_{\text{TSLS}}(r|f)}{\text{Var}_{\text{TSLS}}(r)}$. $\Pr(f_i < f < f_{i+1}) = 0:1$.
 $T = 20, k_1 = 3, k_2 = 4, \rho = 0:35$.

In practice the variance, whether conditional or unconditional, depends on unknown parameters (in the cases of interest here, σ_u^2 and σ_v^2), and the reported variance is typically calculated by simply replacing these by estimates. The natural (unrestricted maximum likelihood) estimator for σ_u^2 is simply $f=2$: Replacing σ_u^2 by $f=2$ in the variance ratios $\text{Var}_{\text{OLS}}(r_j | f) = \text{Var}_{\text{OLS}}(r)$ and $\text{Var}_{\text{TSLs}}(r_j | f) = \text{Var}_{\text{TSLs}}(r)$ means that these are both functions of σ_u^2 alone, and σ_v^2 here can be interpreted as the value of the estimator, rather than the unknown parameter. Thus, one can easily examine the effect of conditioning on the estimated, as distinct from the actual, variances of the two estimators. The results of doing so are in broad agreement with the message conveyed in Tables 2 and 3: for the OLS estimator, the ratio of estimated variances always remains close to unity for all values of f and σ_v^2 ; while for the TSLs estimator the ratio can be either much greater than one (when f is small), or much less than one (when f is large). Thus, the estimated conditional variance continues to reflect, more adequately than the estimated unconditional variance, the precision of the TSLs estimator for the sample actually available.

Finally, since both estimators are (conditionally and unconditionally) biased, it is of interest to compare their conditional mean-square-errors, given f . From equations (30) and (31), the difference between the conditional mean-square-errors is:

$$\begin{aligned} \Phi \text{MSE}(r_j | f) &= \text{MSE}_{\text{OLS}}(r_j | f) - \text{MSE}_{\text{TSLs}}(r_j | f) \\ &= (1 - w) \left[\frac{4 \sigma_u^{-2}}{k_2} {}_0H_1(z) - [w(\bar{A}_i - 2)]^i {}_1H_0(z) - \frac{2 \sigma_u (1+w)^{-2}}{w k_2 (\bar{A}_i - 2)} {}_1H_1^a(z) \right] \end{aligned} \quad (32)$$

where $w = f/(1+f)$: Using the contiguity relations for the confluent hypergeometric function (Abramowitz and Stegun (1965), p.506), this can be expressed entirely in terms of the function ${}_1H_0$. We find, after some tedious algebra,

$$\begin{aligned} \Phi \text{MSE}(r_j | f) &= [(1 - w)(1 - \rho^2) = w] \left[\frac{1}{2} \left(2 + \frac{(1+w)(k_2 - 1)}{w(\bar{A}_i - k_2 - 2)} \right) \right. \\ &\quad \left. {}_i {}_1H_0(z) - \frac{1}{2} \left(2 + \frac{(1+w)(k_2 - 1)}{w(\bar{A}_i - k_2 - 2)} \right) + \frac{(1 - \rho^2)}{(\bar{A}_i - 2)} + \frac{2 \sigma_u \frac{1}{2} (1+w)(\bar{A}_i - 3)^{3/4}}{(\bar{A}_i - 2)(\bar{A}_i - k_2 - 2)} \right] \end{aligned} \quad (33)$$

where $\frac{1}{2} \rho^2 = \sigma_v^2 / (1 + \sigma_v^2)$ is the squared correlation between the right-hand-side endogenous variable and the error term in (10). Clearly, ΦMSE is negative when

$\frac{1}{2}^2 = 0$, i.e., $\bar{\epsilon} = 0$, so that OLS conditionally dominates TSLS when there is no correlation between the right hand side endogenous variable and the error term in the structural equation. Also, it is clear that $\Phi \text{MSE} \neq 0$ as $w \neq 1$ (i.e.; $f \neq 1$), so that there is, conditionally, nothing to choose between these estimators when the sample identifiability test statistic is large. In Figures 4a - 4f we plot conditional MSE indifference curves (along which $\Phi \text{MSE} = 0$) in the $(w; \frac{1}{2}^2)$ plane, for various values of $(\bar{A}; k_2)$, and, on each graph, several values of λ . OLS dominates to the south west of each curve, TSLS to the north east.

Figure 4

Conditional Mean-square-error Indifference Curves: OLS/TSLS

Vertical axis: $\frac{1}{2}^2 = \bar{\epsilon}^2 = (1 + \bar{\epsilon}^2)$; Horizontal axis: $w = f = (1 + f)$

Plotted for $\lambda = .05$ (top), $\lambda = .5$; $\lambda = 1.0$, and $\lambda = 10$ (bottom)

OLS dominates to the south west of each curve.

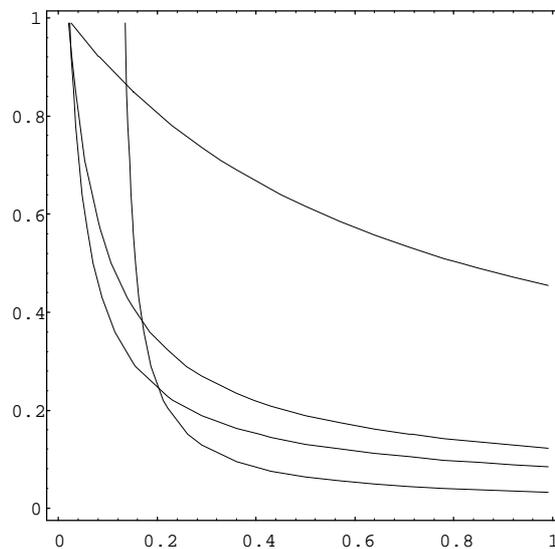


Figure 4a: $\bar{A} = 30; k_2 = 4:$

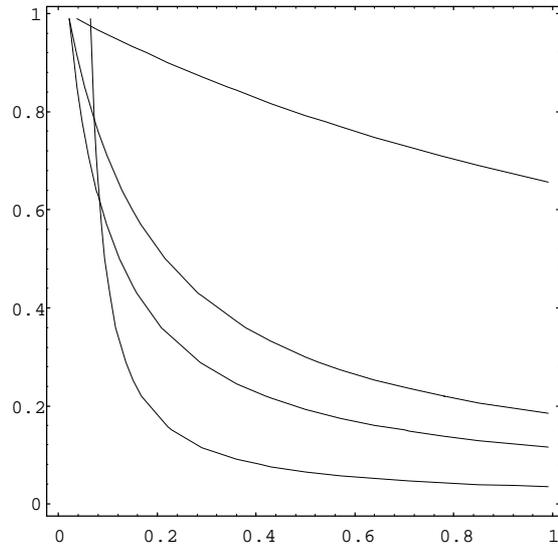


Figure 4b: $\dot{\Lambda} = 30; k_2 = 10:$

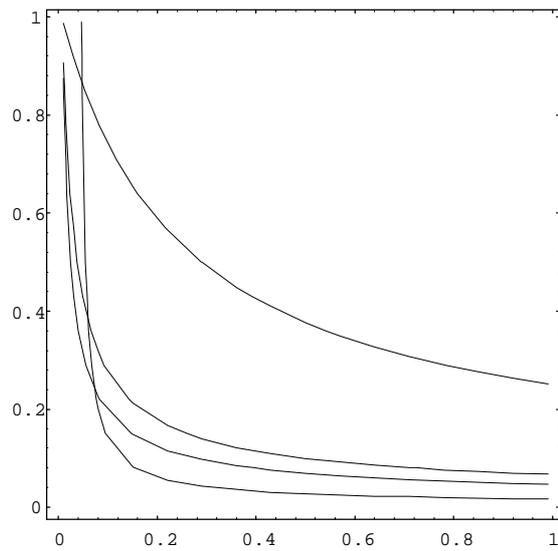


Figure 4c: $\dot{\Lambda} = 80; k_2 = 4:$

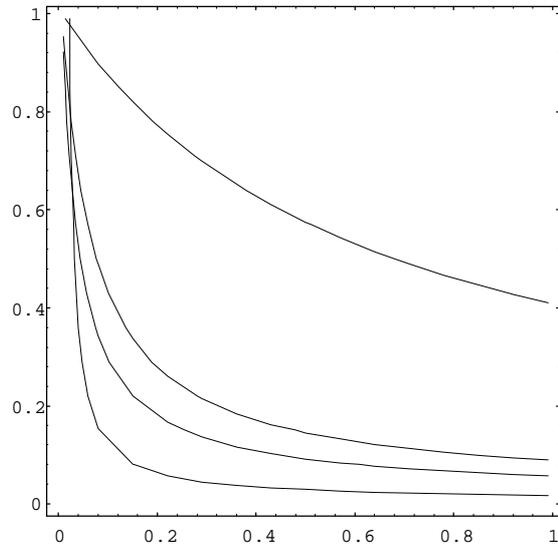


Figure 4d: $\dot{\Lambda} = 80; k_2 = 10$:

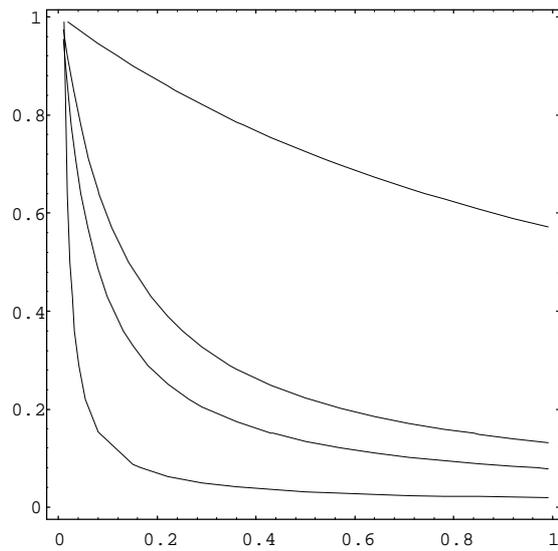


Figure 4e: $\dot{\Lambda} = 80; k_2 = 20$:

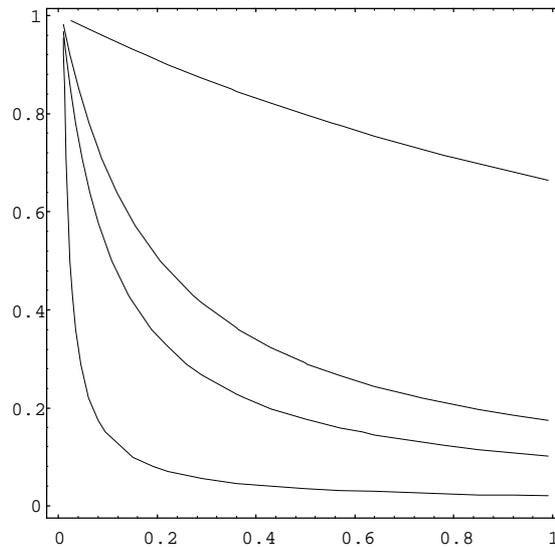


Figure 4f: $\lambda = 80; k_2 = 30$:

Several important conclusions can be drawn from Figure 4:

- (i) In an actually weakly identified model ($\alpha = .05$) OLS dominates TSLS over a large range of $(w; \frac{1}{2})$ values. Even quite strong sample evidence of identification (i.e.; a moderate value of f) does not imply superiority of the TSLS estimator.
- (ii) If the sample evidence (f) indicates that the model is only weakly identified, OLS is the preferred estimator.
- (iii) The greater the degree of overidentification (k_2), the larger the region over which OLS dominates TSLS, but, ceteris paribus, a larger sample size increases the extent of the region over which TSLS dominates.

Although not immediately obvious from (33), it is also true that, while TSLS can conditionally dominate OLS, the conditional MSE advantage of TSLS is never great, but, particularly when the identification test statistic f is small, OLS can be dramatically better than TSLS. We conclude that, when the data only weakly indicate that the model is identified, OLS is certainly the preferred estimator of these two. If the data are moderately favourable to the identification of the model, TSLS may be slightly superior, but, generally, if the data strongly support the identification of the model there is nothing to choose between the two estimators.

5 Concluding Remarks

In a single structural equation model the possibility that the model is unidentified, or is arbitrarily close to being unidentified, cannot be ruled out a priori. Even though it is well known that both the small sample and asymptotic properties of the usual estimators break down in such a case, no practical procedure has been developed to take account of this possibility. In this paper we have argued that, in possibly unidentified models, (i) inference should be made conditional on an identification test statistic, and (ii) the identification test statistic should be considered as an index of precision for inference.

We have derived the exact distributions of the OLS and TSLS estimators conditional on an identification test statistic, and shown that conditioning makes the functional forms of the densities of the OLS and TSLS estimators different. In the extreme case of a totally unidentified model the OLS estimator is independent of the identification test statistic, but that is not the case for the TSLS estimator. The variances of the OLS and TSLS estimators are affected by conditioning more than their means, and, for the TSLS estimator, conditioning can have a substantial effect on the apparent precision of the estimator, in both directions. Conditional mean-square-error comparisons of the two estimators imply that, particularly when the data suggests only weak identification, OLS is the preferred estimator, and that, when properly conditioned, there is little to choose between these two estimators. It would obviously be of interest to include the LIML estimator in these comparisons, but technical problems have so far prevented us from doing so.

The practical implications of these results are important. They show that it is very important to report the identification test statistic, both because it tells us directly how confident we can be that the model is identified, and because it indicates the precision of the estimator of the structural parameters of interest (especially for the TSLS estimator). We believe that it should become a standard practice to report these statistics in empirical work, and to use the (estimated) conditional variance as the appropriate measure of precision. Moreover, the exact results obtained here (under our admittedly very special assumptions) can be expected to hold as approximations in much more general models, such as time series models, in much the same way that Phillips (1989) and Staiger and Stock (1997) extend their analyses to broader classes of models.

Finally, although we have focused on point estimation, the argument that inference in this model should be conditioned on the identification test statistic can be extended to hypothesis testing, including both tests on the structural parameters, and tests for exogeneity. These problems remain for future work.

6 References

Abramowitz, M., and I.A. Stegun (1965), "Handbook of Mathematical Functions", Dover, N.Y.

Anderson, T.W., K. Morimune and T. Sawa (1983), "The Numerical Values of Some Key Parameters in Econometric Models", *Journal of Econometrics*, 21, 229-243.

Barndorff-Nielsen, O.E. (1980), "Conditionality Resolutions", *Biometrika*, 67, 2, 293-310.

Basu, D. (1981), "On Ancillary Statistics, Pivotal Quantities and Confidence Statements", in Y.P. Chaubey, and T.D. Dwivedi (eds.), *Topics in Applied Statistics*, Concordia University, Montreal, Reprinted in Gosh (1988).

Cox, D.R. (1958), "Some Problems Connected with Statistical Inference", *The Annals of Mathematical Statistics*, Vol. 29, 357-37.

Chikuse, Y. and A.W. Davis (1986), "A Survey on the Invariant Polynomials with Matrix Arguments in Relation to Econometric Distribution Theory", *Econometric Theory*, 2, 232-248.

Choi, I. and P.C.B. Phillips (1992), "Asymptotic and Finite Sample Distribution Theory of IV Estimators and Tests in Partially Identified Structural Equations", *Journal of Econometrics*, 51, p. 113-150.

Cox, D.R., and D.V. Hinkley (1974), *Theoretical Statistics*, Chapman and Hall, London.

Creasy, M.A. (1954), "Limits for the Ratio of Means", *Journal of the Royal Statistical Society, Series B*, Vol. XVI, 186-194.

Davis, A.W. (1979), "Invariant Polynomials with Two Matrix Arguments Extending the Zonal Polynomials: Applications to Multivariate Distribution Theory", *Annals of the Institute of Statistical Mathematics*, 31, Part A, 465-485.

- Dobrigal, A., D.A.S. Fraser and R. Gebotys (1987), "Linear Calibration and Conditional Inference", *Communications in Statistics, Theory and Methods*, 16, 4, 1037-1048.
- Dufour, J.-M. (1997), "Some Impossibility Theorems in Econometrics with Applications to Instrumental Variables and Dynamic Models", *Econometrica*, Vol. 65, No. 6, 1365-1388.
- Efron, B. and D.V. Hinkley (1978), "Assessing the Accuracy of the Maximum Likelihood Estimator: Observed versus Expected Fisher Information", *Biometrika*, 65, 3, 457-487.
- Fieller, E.C. (1954), "Some Problems in Interval Estimation", *Journal of the Royal Statistical Society, Series B*, Vol. XVI, 175-185.
- Forchini, G. (1998), "Exact Distribution Theory for some Econometric Problems", Unpublished Ph.D. Dissertation, University of Southampton, January 1998.
- Ghosh, J.K. (1988), *Statistical Information and Likelihood. A Collection of Critical Essay by Dr. D. Basu*, Lecture Notes in Statistics, Springer Verlag.
- Gleser, L.J. and J.T. Hwang (1987), "The Nonexistence of $100(1-\alpha)\%$ Confidence Sets of Finite Expected Diameter in Errors in Variables and Related Models", *The Annals of Statistics*, 15, 1351-1362
- Goutis, C. and G. Casella (1995), "Frequentist Post-Data Inference", *International Statistical Review*, 63, 3, 325-344.
- 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. (1990), "On the Normalization of Structural Equations: Properties of Direction Estimators", *Econometrica*, Vol. 58, No. 5, 1181-1194.
- Hillier, G.H. and C.L. Skeels (1993), "Some Further Exact Results for Structural Equation Estimators", in P.C.B. Phillips (ed.), *Models, Methods, and Applications of Econometrics. Essays in Honor of A.R. Bergstrom*, Basil Blackwell, New York.

Hoadley, B. (1970), "A Bayesian Look at Inverse Linear Regression", *Journal of the American Statistical Association*, Vol. 65, No 329, 356-369.

James, A.T. (1954), "Normal Multivariate Analysis and the Orthogonal Group", *The Annals of Mathematical Statistics*, 25, 40-75.

James, A.T., G.N. Wilkinson and W.N. Venables (1974), "Interval Estimates for a Ratio of Means", *Sankhyā*, Vol. 36, Series A, Pt. 2, 177-183.

Koschat, M.A. (1987), "A Characterization of the Fieller Solution", *The Annals of Statistics*, Vol. 15, No. 1, 462-468.

Lindsay, B.G. and B. Li (1997), "On Second-Order Optimality of the Observed Fisher Information", *The Annals of Statistics*, Vol. 25, No. 5, 2172-2199.

Muirhead, R.J. (1982), *Aspects of Multivariate Statistical Theory*, John Wiley & Sons, Inc., New York.

Phillips, P.C.B. (1983), "Exact Small Sample Theory in the Simultaneous Equation Model", in M.D. Intriligator and Z. Griliches, eds., *Handbook of Econometrics*, Amsterdam: North Holland.

Phillips, P.C.B. (1989), "Partially Identified Econometric Models", *Econometric Theory*, 5, 181-240.

Scheffé, H. (1970), "Multiple Testing versus Multiple Estimation. Improper Confidence Sets. Estimation of Directions and Ratios", *The Annals of Mathematical Statistics*, Vol. 41, No.1, 1-29.

Sims, C.A. (1980), "Macroeconomics and Reality", *Econometrica*, Vol. 48, No. 1, 1-48.

Staiger, D. and J.H. Stock (1997), "Instrumental Variables Regression with Weak Instruments". *Econometrica*, Vol. 65, No. 3, 557-586.

Wallace, D.L. (1980), "The Behrens-Fisher and the Fieller-Creasy Problems" in S.E. Fienberg and D.V. Hinkley (eds.), *R.A. Fisher: an Appreciation*, Lecture Notes in Statistics I, Springer-Verlag, New York.

APPENDIX A: Conditional Densities

In this Appendix we make extensive use of the notation and multivariate integration techniques explained in detail in Muirhead (1982), especially Chapters 2 and 7. To keep the derivations as brief as possible, we ask the reader to refer to these sources for details of the notation and main results that are used. See also Hillier (1985), and Hillier and Skeels (1993).

Let

$$\begin{aligned} X &= (Z_2^0 P_{Z_1} Z_2)^{\frac{1}{2}} \hat{\Gamma}_{22}^{-\frac{1}{2}}; \\ M &= E(X) = (Z_2^0 P_{Z_1} Z_2)^{\frac{1}{2}} \Gamma_{22}^{-\frac{1}{2}}; \end{aligned}$$

and $R = -\frac{1}{22} S_{22} - \frac{1}{22}$: The maximal invariant for the testing problem of interest consists of the n characteristic roots of F_0 in equation (14) in the text, or, equivalently, of $F = R^{\frac{1}{2}} X^0 X R^{\frac{1}{2}}$. Since $X \gg N(M; I_{k_2} - I_n)$; and $R \gg W_n(\hat{\Lambda}_j, k_2; I_n)$, and X and R are independent, the marginal joint density of the roots of F is given by Muirhead (1982), Theorem 10.4.2 as:

$$\text{pdf}(F) = C_2 \text{etr} \{ -\alpha g J F J^{\frac{k_2}{2}} \} |F|_n \prod_{i=1}^n (1 + F_j)^{\frac{\hat{\Lambda}_j}{2} - 1} F_1^{(n)} \left(\frac{\hat{\Lambda}_j}{2}, \frac{k_2}{2}; \alpha; F(I_n + F)^{-1} \right) \prod_{i < j} (f_i - f_j) \quad (A1)$$

where $\alpha = \frac{1}{2} M^0 M$, $F = \text{diag} \{ f_1, f_2, \dots, f_n \}$, and

$$C_2 = \frac{\frac{1}{4} \frac{n^2}{2} |I_n(\frac{\hat{\Lambda}}{2})|}{|I_n(\frac{n}{2})| |I_n(\frac{\hat{\Lambda}_j k_2}{2})| |I_n(\frac{k_2}{2})|};$$

Note that here and below we use F to denote both the original matrix variate, and the diagonal matrix containing the characteristic roots of that matrix. This, we hope, will economise on notation without confusing the reader too much.

Using the definitions of r and \bar{r} in (5) and (6) in the text, and the assumptions made in Section 3, it is straightforward to check that:

$$r_{OLS} \{ X; R \} \sim N((R + X^0 X)^{-1} X^0 M^{-}; (R + X^0 X)^{-1}); \quad (A2)$$

and

$$r_{TSL} \{ X \} \sim N((X^0 X)^{-1} X^0 M^{-}; (X^0 X)^{-1}); \quad (A3)$$

Notice that in the totally unidentified case $M = 0$ and neither the conditional densities (A2) and (A3), nor those of the conditioning variates $(X; R)$, depend on

any parameters of the model. Hence, unconditionally, the densities of r_{OLS} and r_{TSLS} are independent of the model parameters, and those of the unstandardised b_{OLS} and b_{TSLS} depend only on α ; as noted in the introduction.

From (A2) and (A3), it is straightforward to obtain the joint density of $(r; X; R)$ for both of these estimators. To obtain the joint density of $(r; F)$ in both cases we need to transform to a set of new variates including $(r; F)$, and then integrate out the redundant variables. We have, as the starting points:

$$\text{pdf}_{OLS}(r; X; R) = C_1 \text{etr}f_i \alpha g j R j^{\frac{A_i k_2}{2} i^p} j R + X^0 X j^{\frac{1}{2}} \text{etr}f_i \frac{1}{2} (R + X^0 X) (I_n + r r^0) g \text{etr}f_i X^0 M (I_n + r r^0) g \exp f_i \frac{1}{2}^{-0} M^0 X (R + X^0 X) i^{-1} X^0 M^{-} g; \quad (A4)$$

and

$$\text{pdf}_{TSLS}(r; X; R) = C_1 \text{etr}f_i \alpha g j R j^{\frac{A_i k_2}{2} i^p} j X^0 X j^{\frac{1}{2}} \text{etr}f_i \frac{1}{2} R g \text{etr}f_i \frac{1}{2} X^0 X (I_n + r r^0) g \text{etr}f_i X^0 M (I_n + r r^0) g \exp f_i \frac{1}{2}^{-0} M^0 X (X^0 X) i^{-1} X^0 M^{-} g; \quad (A5)$$

where $p = (n + 1)/2$ and

$$C_1 = [(2^{1/4})^{\frac{n(k_2+1)}{2}} 2^{\frac{n(A_i k_2)}{2}} i^{-n} (\frac{A_i k_2}{2})^{i-1}];$$

Conditional density of the OLS estimator

We deal with the OLS case, equation (A4), first. The characteristic roots of F are invariant under the transformations $X \rightarrow K X P H; R \rightarrow H^0 P^0 R P H;$ with $K \in O(k_2); P \in GL(n);$ and $H \in O(n).$ Choosing $P = (I_n + r r^0)^{-1/2},$ making these transformations in (A4) (the Jacobian is $(1 + r^0 r)^{i \frac{k_2+n+1}{2}}$), and averaging over $O(k_2)$ and $O(n),$ the joint density in (A4) is replaced by:

$$C_1 \text{etr}f_i \alpha g (1 + r^0 r)^{i \frac{(A+1)}{2}} j R j^{\frac{(A_i k_2)}{2} i^p} j R + X^0 X j^{\frac{1}{2}} \text{etr}f_i \frac{1}{2} (R + X^0 X) g \frac{C_A^{0:[k]} (\frac{1}{2} M G G^0 M^0; i \frac{1}{2} M^{-0} M^0) C_A^{0:[k]} (\frac{1}{2} X X^0; X (R + X^0 X) i^{-1} X^0)}{j! k! (\frac{n}{2})^{(k)} C_A(I_{k_2})}; \quad (A6)$$

where we have temporarily put $G = (I_n + r r^0) (I_n + r r^0)^{-1/2};$ and we have made extensive use of results from Davis (1979) and Chikuse and Davis (1986). The integral over $O(n)$ is a direct application of Muirhead (1982), Theorem 7.4.1,

while the integral over $O(k_2)$ uses Davis (1979), equation (1.2). The notation used in (A6) is explained in detail in the papers referred to, and in Hillier (1984).

Next, set $W = XR^{i \frac{1}{2}}$, so that $F = W^0W$; and $(dX) = jRj^{\frac{k_2}{2}}(dW)$; and evaluate the integral over $R > 0$ using Davis (1979), equations (2.12) and (2.1), to obtain:

$$C_{11} \text{etr} f_i \alpha g(1+r^0 r)^i \frac{(A+1)}{2} j I_n + W^0 W j^{\frac{A}{2}} \prod_{j;k=0}^{\otimes; [k]; A} \frac{\mu_A^{\otimes; [k]} (\frac{A+1}{2})^{\otimes} (\frac{n}{2})_A}{j! k! (\frac{n}{2})^{\otimes} (\frac{k_2}{2})_A C_A(I_n)}$$

$$\in C_A^{\otimes; [k]} (\alpha (I_n + r^0) (I_n + r r^0)^i (I_n + r^{-0}); i \alpha^{-0}) C_A (W^0 W (I_n + W^0 W)^i)^{-1}; \quad (A7)$$

where $C_{11} = 2^{\frac{n(A+1)}{2}} i_n (\frac{A+1}{2}) C_1$, and we have used the special structure of the arguments of the invariant polynomials in (A6) to simplify the expressions. We now transform $W \rightarrow V; F$; with $F = W^0W$ and $V = W(W^0W)^i \frac{1}{2}$; so that $V^0V = I_n$, and $(dW) = 2^i n j F j^{\frac{k_2}{2} i p} (dF) (V^0 dV)$; where $(V^0 dV)$ denotes Haar measure on the Stieffel manifold (Muirhead (1982), Chapter 2). Since (A7) does not depend on V we may integrate it out using Muirhead (1982), Theorem 2.1.15, leaving a function only of F :

Finally, we transform F to the diagonal matrix containing its characteristic roots (which we continue to denote by F), and an orthogonal matrix, L say, containing its characteristic vectors - see Muirhead (1982) Section 3.2.5. Integrating over $L \in O(n)$ is straightforward since (A7) is not a function of L . We thus obtain:

$$\text{pdf}_{OLS}(r; F) = C_1^\alpha \text{etr} f_i \alpha g(1+r^0 r)^i \frac{(A+1)}{2} j I_n + F j^{\frac{A}{2}} j F j^{\frac{k_2}{2} i p} \prod_{i < j} (f_i - f_j)$$

$$\prod_{j;k=0}^{\otimes; [k]; A} \frac{\mu_A^{\otimes; [k]} (\frac{A+1}{2})^{\otimes} (\frac{n}{2})_A}{j! k! (\frac{n}{2})^{\otimes} (\frac{k_2}{2})_A C_A(I_n)}$$

$$C_A^{\otimes; [k]} (\alpha (I_n + r^0) (I_n + r r^0)^i (I_n + r^{-0}); i \alpha^{-0}) C_A (F (I_n + F)^i)^{-1}; \quad (A8)$$

where

$$C_1^\alpha = \frac{i_n (\frac{A+1}{2})}{\frac{1}{4} \frac{n}{2} i_n (\frac{A}{2})} C_2;$$

The conditional density $\text{pdf}(r | F)$ is obtained by simply dividing (A8) by (A1), giving:

$$\text{pdf}_{OLS}(r | F) = \frac{i_n (\frac{A+1}{2})}{\frac{1}{4} \frac{n}{2} i_n (\frac{A}{2})} (1+r^0 r)^i \frac{(A+1)}{2} \prod_{j;k=0}^{\otimes; [k]; A} \frac{\mu_A^{\otimes; [k]} (\frac{A+1}{2})^{\otimes} (\frac{n}{2})_A}{j! k! (\frac{n}{2})^{\otimes} (\frac{k_2}{2})_A C_A(I_n)}$$

$$g_A(F; \alpha) C_A^{\otimes [k]}(\alpha(I_n + r^0)(I_n + rr^0)^{-1}(I_n + r^{-0}); \alpha^{-0}) \quad (A9)$$

where

$$g_A(F; \alpha) = \frac{C_A(F(I_n + F)^{-1})}{{}_1F_1^{(n)}\left(\frac{A}{2}; \frac{k_2}{2}; \alpha; F(I_n + F)^{-1}\right)}; \quad (A10)$$

Note that, in the totally unidentifiable case $M = 0$, only the first line of (A8) remains, so that, when $M = 0$; r and F are independent. Equation (22) in the text gives this result for the case $n = 1$. Also, when $n = 1$, it is straightforward to check that (A9) reduces to equation (18) in the text: both arguments of the polynomials $C_A^{\otimes [k]}(\zeta; \zeta)$ that occur in (A9) are scalars, and, in the case $n = 1$,

$$g_A(F; \alpha) = \frac{\left(\frac{f_1}{1+f_1}\right)^{j+k}}{{}_1F_1\left(\frac{A}{2}; \frac{k_2}{2}; \frac{f}{1+f}\right)}$$

Conditional density of the TSLS estimator

Starting now from (A5), since (A5) does not depend on R , and R is independent of X , we first transform $X \mid (V; Q)$, with $V = X(X^0X)^{-\frac{1}{2}}$ and $Q = X^0X$; so that $(dX) = 2^{i-n} j Q^{\frac{k_2}{2} i p} (dQ)(V^0 dV)$. F depends on X only through Q , so we need only the joint density $\text{pdf}_{\text{TSLS}}(r; Q)$; obtained by integrating over $V^0 V = I_n$. But, the integral over the Stiefel manifold is invariant under the transformations $M(I_n + r^0)Q^{\frac{1}{2}} \mid M(I_n + r^0)Q^{\frac{1}{2}}H; H \in O(n)$: Hence, averaging over $O(n)$, the term $\text{etr} F V^0 M(I_n + r^0)Q^{\frac{1}{2}}g$ that occurs in (A5) may be replaced by

$${}_0F_1\left(\frac{n}{2}; \frac{1}{4}V^0 M(I_n + r^0)Q(I_n + r^{-0})M^0 V\right):$$

Next, set $F^\dagger = Q^{\frac{1}{2}} R Q^{\frac{1}{2}}$, and note that the characteristic roots of F^\dagger are those of F^{-1} . The Jacobian of the transformation $(R; Q) \mid (F^\dagger; Q)$ is $j Q^p$, so the joint density of $(r; Q)$ may be expressed in the form:

$$2^{i-n} C_1 \text{etr} F_i^{-\alpha}(I_n + r^{-0}) g \left[F^\dagger^{-\frac{(A_i k_2)}{2} i p} j Q^{\frac{(A+1)}{2} i p} \text{etr} F_i \frac{1}{2} Q(I_n + rr^0 + F^\dagger) g \right. \\ \left. \int_{V^0 V = I_n} \exp\left\{-\frac{1}{2} M^0(I_n + V V^0) M^{-1} g\right\} \right. \\ \left. {}_0F_1\left(\frac{n}{2}; \frac{1}{4} V^0 M(I_n + r^0) Q(I_n + r^{-0}) M^0 V\right) (V^0 dV) \right] \quad (A11)$$

Evaluating the Laplace transform, and integrating over $V^0 V = I_n$ using results from Hillier (1985), we obtain:

$$C_{11}^\alpha \text{etr} F_i^{-\alpha}(I_n + r^{-0}) g \left[F^\dagger^{-\frac{(A_i k_2)}{2} i p} I_n + rr^0 + F^\dagger^{-\frac{(A+1)}{2} i p} \right]$$

$$\prod_{j:k=0}^{\infty} \prod_{k:[k];A} \frac{\binom{A+1}{2} \binom{k_2 i n}{2} k}{j! k! \binom{k_2}{2}^A} \mu_A^{\otimes [k]} \\ \in C_A^{\otimes [k]} (\alpha (I_n + r^0) (I_n + r r^0 + F^{\dagger})^{-1} (I_n + r^{-0}); \alpha^{-0}) \quad (A12)$$

where $C_{11}^{\alpha} = 2^{\frac{n(A+1)}{2}} \frac{1}{4} \frac{n k_2}{2} i n \binom{A+1}{2} C_{1=i n} \binom{k_2}{2}$:

Note, at this point, that in the totally unidentifiable case ($M = 0$); the second line of (A12) is redundant, so the joint density of $(r; F^{\dagger})$ becomes simply:

$$C_{11}^{\alpha} F^{\dagger \frac{-(A+1)k_2}{2}} i p^{-} I_n + r r^0 + F^{\dagger \frac{-(A+1)}{2}} : \quad (A13)$$

Transforming to $F = F^{\dagger i^{-1}}$ (which has the same characteristic roots as F), the Jacobian is $F^{-i(n+1)}$, and it is clear at once that, unlike the OLS case, r_{TSLS} and F are not independent when the model is totally unidentifiable. For the case $n = 1$; equation (A13), together with (A1), immediately yield equation (23) in the text.

In the general case ($n > 1$) (but still with $M = 0$) we need to transform F^{\dagger} to its characteristic roots and vectors in (A13), and then integrate over $O(n)$. To obtain a convergent expansion from this step we first write:

$$I_n + r r^0 + F^{\dagger} = (1 + r^0 r) I_n + F^{\dagger i^{-1}} \left(I_i - \frac{r^0 (I_n + F^{\dagger})^{-1} r}{(1 + r^0 r)} \right)$$

Using this in (A13), replacing F^{\dagger} by $L F L^0$, with $L \in O(n)$, and integrating over $O(n)$, we obtain, after a little simplification, for the case $M = 0$,

$$\text{pdf}_{\text{TSLS}}(r; F) = \frac{i n \binom{A+1}{2}}{\frac{1}{4} \frac{n}{2} i n \binom{A}{2}} C_2 (1 + r^0 r)^i \frac{A+1}{2} j F j^{\frac{k_2 i n}{2}} j I_n + F j^{\frac{A+1}{2}} \prod_{i < j} (f_i - f_j) \\ {}_2F_1 \left(\frac{1}{2}; \frac{A+1}{2}; \frac{n}{2}; \frac{r^0 r}{(1 + r^0 r)} (I_n + F)^{-1} \right) : \quad (A14)$$

Hence, for this totally unidentifiable case,

$$\text{pdf}_{\text{TSLS}}(r; F) = \frac{i n \binom{A+1}{2}}{\frac{1}{4} \frac{n}{2} i n \binom{A}{2}} (1 + r^0 r)^i \frac{A+1}{2} j F j^{\frac{1}{2}} j I_n + F j^{\frac{1}{2}} \prod_{i < j} (f_i - f_j) \\ {}_2F_1 \left(\frac{1}{2}; \frac{A+1}{2}; \frac{n}{2}; \frac{r^0 r}{(1 + r^0 r)} (I_n + F)^{-1} \right) : \quad (A15)$$

This reduces to equation (23) in the text in the special case $n = 1$.

In the case $n = 1$ (but $\alpha \neq 0$) the joint density of $(r; f_1)$ is easily obtained directly from (A12) by transforming $f_1^{-1} \rightarrow f_1 = f_1^{-1}$ and rearranging

slightly to give the result in Theorem 1 in the text. To complete the derivation in the general case we need to transform F^i to F^{i-1} in (A12), then from F to LFL^0 as above, and finally integrate over $O(n)$. This, unfortunately, is not straightforward, essentially because of manner in which F appears in equation (A12). We omit the details of this final step - they can be found in Forchini (1998).

APPENDIX B: Conditional Moments; $n = 1$.

Remark on notation: in this Appendix, which deals entirely with the case $n = 1$, we use lower case versions of the symbols used in Appendix A to denote the corresponding quantities when $n = 1$, except for the matrix R , whose (scalar) counterpart will here be denoted by t . Thus, in particular, X is x ; M is m ; V is v ; Q is q ; and R is t . We also, for the case $n = 1$, denote the identifiability test statistic simply by f (rather than f_1).

From (A2) and (A3):

$$E(r_{OLS} | x; t) = \frac{x^0 m}{x^0 x + t}; \quad (B1)$$

$$E(r_{TSLS} | x; t) = \frac{x^0 m}{x^0 x} \quad (B2)$$

$$\text{Var}(r_{OLS} | x; t) = (x^0 x + t)^{-1}; \quad (B3)$$

$$\text{Var}(r_{TSLS} | x; t) = (x^0 x)^{-1}; \quad (B4)$$

Transforming $(x; t)$ to $(v; q)$, with $v = x(x^0 x)^{-1/2}$ and $q = x^0 x$; then from $(q; t)$ to $(f; t)$; with $f = q/t$; these become:

$$E(r_{TSLS} | v; t; f) = \frac{f}{1+f} (v^0 m); \quad (B5)$$

$$E(r_{OLS} | v; t; f) = \frac{f}{1+f} E(r_{TSLS} | v; t; f); \quad (B6)$$

and

$$\text{Var}(r_{TSLS} | v; t; f) = (tf)^{-1}; \quad (B7)$$

$$\text{Var}(r_{OLS} | v; t; f) = [t(1+f)]^{-1} = \frac{f}{1+f} \text{Var}(r_{TSLS} | v; t; f) \quad (B8)$$

so that

$$E(r_{OLS}^2 | v; t; f) = [t(1+f)]^{i-1} \left[1 + \frac{-2f}{1+f} (v^0 m)^2\right]; \quad (B9)$$

and

$$E(r_{TSL}^2 | v; t; f) = (tf)^{i-1} \left[1 + \frac{-2f}{1+f} (v^0 m)^2\right]; \quad (B10)$$

It is straightforward to show that

$$\text{pdf}(v; t | f) = c_1 h(f; \lambda) t^{\frac{\lambda}{2} i - 1} \exp\left\{-\frac{1}{2} t(1+f)g\right\} \exp\left\{-\frac{f}{1+f} (v^0 m)g\right\} \quad (B11)$$

with $c_1 = \frac{1}{2} \left(\frac{k_2}{2}\right)^{-\frac{\lambda}{2} + 1} \frac{k_2}{2} \frac{1}{2} \left(\frac{\lambda}{2}\right)$ and

$$h(f; \lambda) = \frac{(1+f)^{\frac{\lambda}{2}}}{{}_1F_1\left(\frac{\lambda}{2}; \frac{k_2}{2}; \frac{-f}{1+f}\right)}; \quad (B12)$$

Hence, to obtain the conditional moments, given f , we need to multiply the expressions in (B5) and (B6) and, for the second moments, in (B9) and (B10), by $\text{pdf}(v; t | f)$ in (B11), and then integrate over $v^0 v = 1$ and $t > 0$: To evaluate the integrals of the terms $(v^0 m)$ and $(v^0 m)^2$ that occur, we use:

Lemma B1: If λ and m are $k_2 \in 1$;

$$\int_{v^0 v=1} (v^0 m) \exp\{f v^0 m g\} (v^0 dv) = V(k_2) \frac{2_{\lambda} a}{k_2} {}_0F_1\left(\frac{k_2}{2} + 1; \frac{1}{2}_{\lambda} a^2\right); \quad (B13)$$

$$\begin{aligned} \int_{v^0 v=1} (v^0 m)^2 \exp\{f v^0 m g\} (v^0 dv) &= V(k_2) \left[\frac{2_{\lambda} a}{k_2} {}_0F_1\left(\frac{k_2}{2} + 1; \frac{1}{2}_{\lambda} a^2\right) \right. \\ &\quad \left. + \frac{4_{\lambda} a^2}{k_2(k_2 + 2)} {}_0F_1\left(\frac{k_2}{2} + 2; \frac{1}{2}_{\lambda} a^2\right) \right] \\ &= \frac{2_{\lambda} V(k_2)}{k_2} {}_1F_2\left(\frac{3}{2}; \frac{1}{2}; \frac{k_2}{2} + 1; \frac{1}{2}_{\lambda} a^2\right); \end{aligned} \quad (B14)$$

where $\lambda = m^0 m = 2$ and $V(k_2) = 2^{\frac{k_2}{2}} \frac{1}{2} \left(\frac{k_2}{2}\right)$ is the surface content of the unit sphere in k_2 dimensions.

Proof: The integral

$$\int_{v^0 v=1} (v^0 m)^r \exp\{f v^0 m g\} (v^0 dv)$$

is the coefficient of $z^r=r!$ in the expansion of

$$\int_{v=1}^z \exp f(a+z)v^0 mg(v^0 dv) = V(k_2)_0 F_1\left(\frac{k_2}{2}; \frac{1}{2}; (a+z)^2\right):$$

Equations (B13) and (B14) follow by simply evaluating those coefficients for the cases $r = 1$ and $r = 2$: The second expression for the result in (B14) is easily obtained from the first on using the fact that $(c+1)_j = (c)_j + j(c)$.

From (B5), (B11), and (B13) (with $a = \frac{p}{f}$), we obtain:

$$E(r_{OLS} | j, f) = \frac{2^{-j} f^{-j} {}_1F_1\left(\frac{A}{2}; \frac{k_2}{2} + 1; \frac{f}{1+f}\right)}{k_2(1+f) {}_1F_1\left(\frac{A}{2}; \frac{k_2}{2}; \frac{f}{1+f}\right)} \quad (B15)$$

and the corresponding result for $E(r_{TSLS} | j, f)$ follows from (B6). To simplify the notation for subsequent results, define, for non-negative integers p and q ,

$${}_p H_q(z) = \frac{{}_1F_1\left(\frac{A}{2}; p; \frac{k_2}{2} + q; z\right)}{{}_1F_1\left(\frac{A}{2}; \frac{k_2}{2}; z\right)} \quad (B16)$$

so that, setting $z = \frac{f}{1+f}$,

$$E(r_{OLS} | j, f) = \frac{2z^{-j}}{k_2} {}_0 H_1(z): \quad (B17)$$

Similarly, using (B14) in (B9) and (B10) yields:

$$\begin{aligned} E(r_{OLS}^2 | j, f) &= (A - j - 2)^{j-1} [{}_1 H_0(z) + \frac{2z^{-2}}{k_2} {}_1 H_1(z)] + \frac{4z^{2-j}}{k_2(k_2 + 2)} {}_0 H_2(z) \\ &= (A - j - 2)^{j-1} [{}_1 H_0(z) + \frac{2z^{-2}}{k_2} {}_1 H_1^*(z)] \end{aligned} \quad (B18)$$

and

$$\begin{aligned} E(r_{TSLS}^2 | j, f) &= \frac{1+f}{f(A-j-2)} [{}_1 H_0(z) + \frac{2z^{-2}}{k_2} {}_1 H_1(z)] + \frac{4z^{2-j}}{k_2(k_2 + 2)} {}_0 H_2(z) \\ &= \frac{1+f}{f(A-j-2)} [{}_1 H_0(z) + \frac{2z^{-2}}{k_2} {}_1 H_1^*(z)]; \end{aligned} \quad (B19)$$

where

$${}_p H_q^*(z) = \frac{{}_2F_2\left(\frac{A}{2}; p; \frac{3}{2}; \frac{1}{2}; \frac{k_2}{2} + q; z\right)}{{}_1F_1\left(\frac{A}{2}; \frac{k_2}{2}; z\right)}. \quad (B20)$$

The conditional variances and mean-square-errors given in Theorem 2 in the text follow immediately from these results.