A family of tests of significance is developed for coefficients in a single equation of a simultaneous system. Different members of this family are distinguished by the $k$-class estimator on which they are based, and on the alternative hypothesis against which they test. The size of the test is found when the disturbances are small, and the test is shown to be consistent if $\operatorname{plim} k = 1$ or $k = \lambda$, the limited information maximum likelihood value.

1. INTRODUCTION

Several kinds of tests have been proposed for the estimated coefficients of an equation in a system of simultaneous linear equations.

Anderson and Rubin [1] proposed a test that the restrictions on the coefficients of the predetermined variables are correct against the alternative hypothesis that one or more of the restrictions is false. Hood and Koopmans [9] showed that the same test statistic can be used to test the null hypothesis that the restrictions on the coefficients of both predetermined and endogenous variables are correct against the alternative that one or more of the restrictions is false. This test statistic is closely associated with the limited information maximum likelihood estimator of the structural coefficients. More recently Basmann [3] proposed a test with the same null and alternative hypotheses as Koopmans and Hood, but with two-stage least squares estimators used in place of limited information maximum likelihood estimators. The consistency of these tests has been debated by Liu and Breen [14 and 15], and Fisher and Kadane [6].

A second kind of test has a much more restricted alternative, namely that a single specified restriction is false. Such tests are suggested in Dhrymes [4, 5, pp. 272–277].

Dhrymes [4, pp. 222–223] also proposed a third kind of test which generalizes the two kinds above by allowing an arbitrary alternative hypothesis containing the null hypothesis. He also proposed a new test statistic based on two-stage least squares. This paper proposes a new family of statistics to test the coefficients of any $k$-class estimator, including limited information maximum likelihood against this more general alternative.

The tests here proposed have an especially close relationship to the first kind of test above. When the alternative is that one or more of the restrictions on the coefficients of either endogenous or exogenous variables is wrong, the limited information test statistic is that of Anderson and Rubin, and Hood and Koopmans, and the two-stage least squares test statistic is that of Basmann.

The distribution of these test statistics is also analyzed, using the small-$\sigma$ method (Kadane [10 and 11]). The lowest order term, $\sigma^0$, does not depend on which $k$-class method was used to estimate, and yields a simple $F$ distribution under the null hypothesis, generalizing the results of Kadane on the Anderson-
Rubin-Hood-Koopmans statistic and on the Basmann statistic. When the alternative allows only a single specified restriction, an intuitively appealing test would compare some function of the observations to a $t$ distribution with degrees of freedom equal to the number of time periods ($T$) minus the number of parameters estimated. Dhrymes [4] was not able to justify such a test with his theory, but the theory here does give a test with the above degrees of freedom.

Finally, the consistency of these new tests is investigated. They are shown to be large-sample consistent, generalizing results of Fisher and Kadane [6] on the consistency of the Anderson-Rubin-Hood-Koopmans test.

An interesting research question would be to compare the two-stage least squares test proposed here to the Dhrymes test and the asymptotic test. Is the Dhrymes test consistent? Is one test universally more powerful than the others? Initial explorations in this direction have been conducted by Morgan and Vandaele [17], and by Maddala [16].

2. THE GENERAL DECISION PROBLEM

The development of a simultaneous equation econometric model is often shrouded in uncertainty. The specification of restrictions on various equations is never as easy as a text would lead one to believe. Hence a working econometrician may desire ways of testing models to see which specification best fits the data and whatever prior opinions he or she may have. There are several approaches to this problem currently advocated.

The most classical approach asks the user to specify a strict null hypothesis and an alternative hypothesis which is more general. Thus certain coefficients can be tested against zero. One contribution of this paper is to allow a more general assortment of alternative hypotheses than had been available. However, tests of significance suffer from certain criticisms, which apply here. First, the null hypothesis is not usually really believed. Why, then, should it be a subject of special study? Second, as the sample size grows, any consistent test of significance will reject the null hypothesis unless it is exactly, literally true. Thus even if the restrictions are nearly correct, sufficient data will permit rejection of the null hypothesis (see Kadane, Lewis, and Ramage [12] for a case where this consideration was important).

A second approach, the Bayesian method, asks for the econometrician's opinion in great detail. Since typical simultaneous equation models involve many parameters, eliciting such an opinion would be very difficult (Savage [21]). Phrased in Bayesian terms, the problem is to choose the right model. Research in this area has thus far been concentrated on single equation regression models. So far, reasonable situations have not been found in which the simpler model is chosen, since the more complicated model always has more information and thus fits better. Therefore to date most applications require the two models being compared to have the same number of parameters. (See [7 and 23].) Further research is needed before adequate Bayesian procedures can be found, even with a solution to the elicitation problem.
A third approach is called "data analysis" [22]. It has much less formal structure, and consists mainly in trying various different models, estimators, and ways of looking at the data, and choosing a few to report. While some variant of this method, possibly tempered with some tests of significance, is probably the most widely used, it is hard to analyze and justify.

Thus although the results of this paper are in the mold of classical tests of significance, that should not be taken as evidence that I favor the first approach above. No doubt all three are legitimate areas for modern research in econometric methods.

3. DERIVATION OF THE LIKELIHOOD RATIO TEST AND ITS ANALOGS

Consider a complete system

$$(1) \quad YB + Z\Gamma + U = 0$$

where $Y$ is a $T \times G$ matrix of endogenous variables and $Z$ is a $T \times K$ matrix of exogenous variables. $B$ is a $G \times G$ non-singular matrix of parameters, $\Gamma$ is a $K \times G$ matrix of parameters, and $U$ is a $T \times G$ matrix of jointly normal residuals with zero means and covariances $E u_t u_{t'} = \sigma_{ij}\delta_{tt'}$ where $\delta_{tt'}$ is 1 if $t = t'$ and zero otherwise. Note that no lagged endogenous variables are permitted in the system (1).

The null hypothesis specifies certain restrictions on the coefficients of the first equation, so that it may be written

$$(2) \quad y = Y_1\beta_1 + Z_1\gamma_1 + u$$

where $Y$ is partitioned $Y = (y, Y_1, Y_2, Y_3): T \times 1 + G_1 + G_2 + G_3$; $Z$ is partitioned $Z = (Z_1, Z_2, Z_3): T \times K_1 + K_2 + K_3$; and $u$ is partitioned $u = (u, U'): T \times 1 + (G - 1)$. The vectors $\beta_1$ and $\gamma_1$ are conformable vectors of parameters.

The alternative hypothesis specifies a subset of the restrictions specified under the null hypothesis. The remaining restrictions—those specified by the null hypothesis but not by the alternative, are the hypotheses being tested.

With only the assumptions of the alternative hypothesis, the first equation may be written

$$(3) \quad y = Y_1\beta_1 + Y_2\beta_2 + Z_1\gamma_1 + Z_2\gamma_2 + u;$$

again the $\beta$'s and $\gamma$'s are conformable.

A test of the null hypothesis (2) against the alternative (3) may thus be thought of as a test of $\beta_2 = 0$ and $\gamma_2 = 0$. As examples, if a single $\beta$ is being tested, then $K_2 = 0$ and $G_2 = 1$. Similarly if a single $\gamma$ is being tested, $G_2 = 0$ and $K_2 = 1$. If all the $\beta$'s and $\gamma$'s are being tested, $K_3 = 0$, and $G_3 = 0$.

Under each hypothesis $i$, the concentrated log-likelihood function is well-known to be of the form

$$(4) \quad L^i = k' + \frac{T}{2} \log l_i$$
where $k$ is a constant and $l_1$ and $l_2$ are given by

\begin{align}
(5) \quad l_1 &= \min_{\beta_1} \frac{\beta_1^* Y_1^* P_{z_1} Y_1^* \beta_1^*}{\beta_1^* Y_1^* P_z Y_1^* \beta_1^*} \\
&= \frac{\beta_1^* Y_1^* P_{z_1} Y_1^* \beta_1^*}{\beta_1^* Y_1^* P_z Y_1^* \beta_1^*}
\end{align}

and

\begin{align}
(6) \quad l_2 &= \min_{\beta_2} \frac{\beta_2^* Y_2^* P_{z_2} Y_2^* \beta_2^*}{\beta_2^* Y_2^* P_z Y_2^* \beta_2^*} \\
&= \frac{\beta_2^* Y_2^* P_{z_2} Y_2^* \beta_2^*}{\beta_2^* Y_2^* P_z Y_2^* \beta_2^*}
\end{align}

where $Y_1^* = (y', Y_1'), \quad Y_2^* = (y', Y_1', Y_2'), \quad \text{and} \quad P_F = I - F(F'F)^{-1}F'$ for any matrix $F$.

The minimizing choice in (5) and (6), when normalized, can be written as $(-1, 1')$ ($i = 1, 2$), where $\beta^1$ is the limited information maximum likelihood estimator of $\beta_1$ in equation (2), and $\beta^2$ is the limited information maximum likelihood estimator of $(\beta_1, \beta_2)$ in equation (3). Note that when the alternative hypothesis allows (3) to be underidentified, $\beta^2$ will not be unique. Nevertheless, any minimizing choice yields $l_2 = 1$. This causes no problems in the subsequent analysis.

The log-likelihood ratio is

\begin{equation}
(7) \quad L_1 - L_2 = \frac{T}{2} \log(l_1/l_2),
\end{equation}

which is equivalent to $l_1/l_2$. Hence the likelihood ratio statistic is equivalent to $l_1/l_2$.

\textbf{Theorem 1}: The likelihood ratio statistic for testing the null hypothesis (2) against the alternative hypothesis (3) is equivalent to $l_1/l_2$, where $l_1$ is given by (5) and $l_2$ is given by (6).

In the special case where the alternative hypothesis specifies no restrictions on the equation, it is well-known that $l_2 = 1$. Hence the statistic in (7) simplifies to $l_1$, which is the Anderson-Rubin-Hood-Koopmans statistic.

Possibly some other estimator might be used instead of limited information maximum likelihood in (5) and (6). Let $l_1(k)$ be the expression in (5) with the $k$-class estimator with parameter $k$ for equation (2) substituted in place of limited information maximum likelihood. Similarly let $l_2(k)$ be the expression in (6) with the $k$-class estimator with parameter $k$ for equation (3) substituted in place of limited information maximum likelihood there.

The analogous statistic to (7) is then $l_1(k)/l_2(k)$. For any $k$-class estimator, when the model under the alternative hypothesis is just or underidentified, $l_2(k) \equiv 1$, and so $l_1(k)/l_2(k) = l_1(k)$. The Basmann test [3] is $l_1(1)$ with this alternative hypothesis.
4. DISTRIBUTION OF THE LIKELIHOOD RATIO STATISTIC AND ITS ANALOGS 
WHEN THE DISTURBANCES ARE SMALL

In [10] I introduced an approach to the distribution of econometric quantities based on asymptotic expansion in \( \sigma \), the variance of the disturbances in the system. As \( \sigma \to 0 \), the system becomes less and less variable around the true regression. Applied to the distribution of the Anderson-Rubin-Hood-Koopmans statistic, the first order term, which does not depend on \( \sigma \), gave a distribution which approaches, as \( T \to \infty \), the large-sample results of Anderson and Rubin. Additionally, the first order term of the distribution of Basmann’s test statistic is the same \( F \) distribution found in a Monte Carlo study [2]. Applied to the statistic found in Section 3, we obtain:

**Theorem 2:** Asymptotically as \( \sigma \to 0 \), when the null hypothesis is true,

\[
\frac{T - K + L_2}{L_1 - L_2}[(l_1(k)/l_2(k)) - 1]
\]

has the Fisher variance ratio distribution \( F_{L_1 - L_2, T - K + L_2} \) where \( L_1 = K_2 + K_3 - G_1 \) and \( L_2 = [K_3 - G_1 - G_2]^+ \) are the degrees of overidentification of the two models, and \( K \) is the number of exogenous variables in the system. This holds for every member of the \( k \) class with \( k \) fixed and for limited information maximum likelihood.

**Proof of Theorem 2:** From Kadane [10], when the null hypothesis is true,

\[
l_1(k) = \frac{u' P_{X1} u}{u' P_{Z2} u} + 0(\sigma),
\]

where \( X^1 = [Z \Pi_1, Z_1] \).

Similarly \( l_2(k) \) can be thought of as an analogous equation in a model with more included endogenous and exogenous variables (but some “true” coefficients are zero, even though they are estimated). Then, by exactly the same argument as before,

\[
l_2(k) = \frac{u' P_{X2} u}{u' P_{Z2} u} + 0(\sigma),
\]

where \( X^2 = [Z \Pi_1, Z \Pi_2, Z_1, Z_2] \).

Putting (9) and (10) together,

\[
\frac{l_1(k)}{l_2(k)} = \frac{u' P_{X1} u}{u' P_{X2} u} + 0(\sigma)
\]

\[
= \frac{u' P_{X1} u}{u' P_{X2} u} + 0(\sigma).
\]
Since the columns of $X^1$ are included among those in $X^2$, $X^1\bar{P}_{X^2} = 0$. Hence $\bar{P}_{X^1}\bar{P}_{X^2} = \bar{P}_{X^2}$, so

(11) $\bar{P}_{X^2}[\bar{P}_{X^1} - \bar{P}_{X^2}] = 0$.

Then

(12) $l_1(k)/l_2(k) = 1 + \frac{u' [\bar{P}_{X^1} - \bar{P}_{X^2}] u}{u' \bar{P}_{X^2} u} + o_p(\sigma)$.

Here $u' [\bar{P}_{X^1} - \bar{P}_{X^2}] u$ and $u' \bar{P}_{X^2} u$ have independent $\chi^2$ distributions, using (11). Their degrees of freedom are:

- $\text{tr} \bar{P}_{X^2} = T - K + L_2$
- $\text{tr} [\bar{P}_{X^1} - \bar{P}_{X^2}] = T - K + L_1 - (T - K + L_2) = L_1 - L_2$.

Therefore

$$[(l_1(k)/l_2(k)) - 1] \frac{T - K + L_2}{L_1 - L_2}$$

has an asymptotic (as $\sigma \to 0$) $F$ distribution with degrees of freedom $L_1 - L_2$ and $T - K + L_2$. Q.E.D.

Less transformation is required in the following equivalent statement:

**Corollary 1:** $l_2(k)/l_1(k)$ has an asymptotic (as $\sigma \to 0$) Beta distribution with parameters $(T - K + L_2)/2$ and $(L_1 - L_2)/2$. (See, for instance, Rao [20, p. 135]).

In the special case in which the model is just or underidentified under the alternative hypothesis, $L_2 = 0$, $l_2 = 1$, and Theorem 2 reduces to the Theorem of Kadane [10]. In the special case in which only a single restriction is being tested, $L_1 - L_2 = 1$, and Theorem 2 reduces to

$$(T - K + L_2)[l_1/l_2 - 1] \sim F_{1,T-K+L_2}.$$ 

Notice that $T - K + L_2 = T - K_1 - K_2 - G_1 - G_2$ is a natural degrees of freedom parameter, namely the number of time periods minus the number of estimated parameters.

5. **Consistency of the Tests**

One interesting property a test may have is to be consistent against certain alternative hypotheses, which means that under each of those alternative hypotheses, the power of the test approaches one as the sample size increases [13, p. 305]. This section reports on investigations into the consistency of the tests proposed in Sections 3 and 4.

Note that consistency as defined above is a *large-sample*, not a *small-$\sigma$* concept. In this section all probability limits are as $T \to \infty$. For a sophisticated treatment of the power of the Anderson-Rubin-Hood-Koopmans test in the small-$\sigma$ context, see Ramage [19].
**Theorem 3:** Suppose: (i) (a) \(\text{plim } k = 1\) under both the null and alternative hypotheses, or (b) that \(k = \lambda\), the limited information maximum likelihood value; (ii) \(\text{plim } Z'Z/T = M\), a positive definite matrix; (iii) \(\text{plim } Z'U/T = 0\); and (iv) under the alternative hypothesis, \(\text{plim } T^{-1}Z[y_{6i} + Y_i + Z_i] \neq 0\) if \(\hat{\alpha}, \hat{\beta}, \text{ and } \hat{\gamma}\) are not all zero.

Then the test \(l_1(k)/l_2(k)\) is consistent.

The assumptions of Theorem 3 deserve some comment. Assumption (i) limits the \(k\)-class estimators being considered, but includes two-stage least squares, \(k = 1 + (L - 1)T^{-1}\) discussed by Nagar [18], \(k = 1 + (L - 1)(T - K)^{-1}\) discussed by Kadane [11, p. 727], and limited information maximum likelihood. It excludes ordinary least squares. Assumptions (ii) and (iii) are standard for large-sample theory [5]. Assumption (iv) says, roughly, that under the alternative hypothesis there is no way “to manipulate the model to obtain a linear expression in \(Y_i^*\) and \(Z_i\) which is equal to the disturbance term only” [6]. Thus the equation in question is overidentified under the alternative hypothesis.

**Proof of Theorem 3:** First, some notation is needed. Let us partition \(B^{-1} = (b, B_1, B_2, B_3)\) where \(B_i\) is \(G \times G_i\) \((i = 1, 2, 3)\) so that

\[
Y_i = -Z_i \Gamma B_i - UB_i = Z_i \Pi_i + V_i \quad (i = 1, 2, 3).
\]

Also partition \(M\) as follows:

\[
M = \text{plim } \frac{Z'Z}{T} = \text{plim } \frac{Z_i'Z_i}{T} = \begin{bmatrix}
Z_i'Z_i & Z_i'Z_2 & Z_i'Z_3 \\
T & T & T \\
Z_2'Z_1 & Z_2'Z_2 & Z_2'Z_3 \\
T & T & T \\
Z_3'Z_1 & Z_3'Z_2 & Z_3'Z_3 \\
T & T & T
\end{bmatrix}
= \begin{bmatrix}
M_{11} & M_{12} & M_{13} \\
M_{21} & M_{22} & M_{23} \\
M_{31} & M_{32} & M_{33}
\end{bmatrix}
\]

Finally let

\[
\text{plim } \frac{Z_i'Z_i}{T} = \begin{bmatrix}
M_{1i} \\
M_{2i} \\
M_{3i}
\end{bmatrix} = M_{i} \quad (i = 1, 2, 3)
\]

and let \(M_{i} = M'_{i}\). Next some consequences of assumption (iv) are recorded in Lemma 1.
LEMMA 1: If assumption (iv) holds, then under the alternative hypothesis, (a) \( \Pi_1 \) is of full rank \( G_1 \); (b) \( \{ MP_2 \beta_2 + M_2 \gamma_2, MP_1, M_{.1} \} \) are a set of \( 1 + G_1 + K_1 \) linearly independent \( K \)-vectors.

PROOF: Suppose \( \tilde{\alpha}, \tilde{\beta}, \tilde{\gamma} \) are not all zero. Then:

\[
0 \neq \text{plim} \; T^{-1} Z'[\{ (Y_1 \beta_1 + Y_2 \beta_2 + Z_1 \gamma_1 + Z_2 \gamma_2 + u) \tilde{\alpha} + Y_1 \tilde{\beta}_1 + Z_1 \tilde{\gamma} \] = \text{plim} \; T^{-1} Z'[(Z \Pi_1 \beta_1 + Z \Pi_2 \beta_2 + Z_1 \gamma_1 + Z_2 \gamma_2) \tilde{\alpha} + Z_1 \Pi_1 \tilde{\beta} + Z_1 \tilde{\gamma}] = (MP_1 \beta_1 + MP_2 \beta_2 + M_{.1} \gamma_1 + M_{.2} \gamma_2) \tilde{\alpha} + MP_1 \tilde{\beta} + M_{.1} \tilde{\gamma} = (MP_2 \beta_2 + M_{.2} \gamma_2) \tilde{\alpha} + MP_1 (\beta_1 \tilde{\alpha} + \tilde{\beta}) + M_{.1}(\gamma_1 \tilde{\alpha} + \tilde{\beta}).
\]

Let \( \tilde{\alpha}_1 = \tilde{\alpha}, \tilde{\beta}_1 = \beta_1 \tilde{\alpha} + \tilde{\beta}, \) and \( \tilde{\gamma}_1 = \gamma_1 \tilde{\alpha} + \tilde{\gamma}. \) Then \( (\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}) \) are not all zero if and only if \( (\tilde{\alpha}_1, \tilde{\beta}_1, \tilde{\gamma}_1) \) are not all zero. Thus \( \{ MP_2 \beta_2 + M_{.2} \gamma_2, MP_1, M_{.1} \} \) are a linearly independent set of vectors of length \( K. \) Finally, \( \rho(MP_1) = G_1 \) by linear independence. Since \( M \) is non-singular, \( \rho(\Pi_1) = G_1. \) Q.E.D.

LEMMA 2: Under the alternative hypothesis, and assuming (i) (a) of Theorem 3, \( \text{plim} \; (\beta_k - \beta) = (\Pi'_1 C \Pi_1)^{-1} \Pi'_1 C (\Pi_2 \beta_2 + I_2 \gamma_2) \)

where

\[
C = M - M_{.1} M^{-1}_{.1} M_{.1}, \quad \text{and} \quad I_2 = \begin{pmatrix} 0 & \cdots & 0 \\ I & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & K_2 \end{pmatrix}, K_2 \times K_2.
\]

PROOF: The general \( k \)-class estimator is given by

\[
\begin{pmatrix} \beta_k \\ \gamma_k \end{pmatrix} = \begin{pmatrix} Y'_1[I - k \bar{P}_Z] Y_1 & Y'_1 Z_1 \\ Z'_1 Y_1 & Z'_1 Z_1 \end{pmatrix}^{-1} \begin{pmatrix} Y'_1[I - k \bar{P}] Y_1 & Y'_1 Z_1 \\ Z'_1 Y_1 & Z'_1 Z_1 \end{pmatrix}^{-1} \begin{pmatrix} Y'_1[I - k \bar{P}_Z] Y_1 & Y'_1 Z_1 \\ Z'_1 Y_1 & Z'_1 Z_1 \end{pmatrix}^{-1} Y'_1[I - k \bar{P}_Z] Y_1 \begin{pmatrix} \beta \\ \gamma \end{pmatrix} + \begin{pmatrix} Y'_1[I - k \bar{P}_Z] Y_1 & Y'_1 Z_1 \\ Z'_1 Y_1 & Z'_1 Z_1 \end{pmatrix}^{-1} \begin{pmatrix} Y'_1[I - k \bar{P}] Y_1 & Y'_1 Z_1 \\ Z'_1 Y_1 & Z'_1 Z_1 \end{pmatrix}^{-1} \begin{pmatrix} Y'_1[I - k \bar{P}_Z] Y_1 & Y'_1 Z_1 \\ Z'_1 Y_1 & Z'_1 Z_1 \end{pmatrix}^{-1} Y'_1[I - k \bar{P}_Z] Y_1 \begin{pmatrix} \beta \\ \gamma \end{pmatrix} \times (Y_2 \beta_2 + Z_2 \gamma_2 + u).
\]

Therefore

\[
\text{plim} \; \begin{pmatrix} \beta_k - \beta \\ \gamma_k - \gamma \end{pmatrix} = \text{plim} \begin{pmatrix} Y'_1[I - k \bar{P}_Z] Y_1 & Y'_1 Z_1 \\ T & T \\ Z'_1 Y_1 & Z'_1 Z_1 \\ T & T \end{pmatrix}^{-1} \begin{pmatrix} Y'_1[I - k \bar{P}_Z](Y_2 \beta_2 + Z_2 \gamma_2 + u) \\ T \\ Z'_1(Y_2 \beta_2 + Z_2 \gamma_2 + u) \\ T \end{pmatrix}
\]
The probability limit of the inverse is already available in the literature (Dhrymes [5, p. 179]):

\[
\text{plim} \begin{pmatrix}
\frac{Y_1'[I - k\bar{P}_Z]Y_1}{T} & \frac{Y_1'Z_1}{T} \\
\frac{Z_1'Y_1}{T} & \frac{Z_1'Z_1}{T}
\end{pmatrix}^{-1} = \begin{bmatrix} \Pi_1'M\Pi_1 & \Pi_1'M_{11} \\ M_{11}\Pi_1 & M_{11} \end{bmatrix}^{-1}.
\]

Now

\[
\text{plim} \frac{Y_1'[I - k\bar{P}_Z](Y_2\beta_2 + Z_2\gamma_2 + u)}{T} = \text{plim} \frac{(\Pi_1'Z' + \sigma V_1')(I - k\bar{P}_Z)(Z\Pi_2\beta_2 + V_2\beta_2 + Z_2\gamma_2 + u)}{T}
\]

\[
= \Pi_1'M\Pi_2\beta_2 + \Pi_1'M_{22}\gamma_2 + \text{plim} \frac{B_1'U'[I - k\bar{P}_Z]UB_2\beta_2}{T} - \text{plim} \frac{B_1U[I - k\bar{P}_Z]u}{T}
\]

\[
= \Pi_1'M\Pi_2\beta_2 + \Pi_1'M_{22}\gamma_2 + B_1'\Sigma B_2\beta_2 \text{plim} \left( \frac{T - k(T - K)}{T} \right) - B_1'\sigma_{11} \text{plim} \left( \frac{T - k(T - K)}{T} \right)
\]

where \( \sigma_{11} = EU'u/T \), the first row of \( \Sigma \). Now since

\[
\text{plim} \left( \frac{T - k(T - K)}{T} \right) = \text{plim} \left( 1 - k + \frac{K}{T} \right) = 0,
\]

\[
\text{plim} \frac{Y_1'[I - k\bar{P}_Z](Y_2\beta_2 + Z_2\gamma_2 + u)}{T} = \Pi_1'M\Pi_2\beta_2 + \Pi_1'M_{22}\gamma_2.
\]

Finally,

\[
\text{plim} \frac{Z_1'(Y_2\beta_2 + Z_2\gamma_2 + u)}{T} = \text{plim} \frac{Z_1'[Z\Pi_2\beta_2 + V_2\beta_2 + Z_2\gamma_2 + \sigma u]}{T} = M_1\Pi_2\beta_2 + M_{12}\gamma_2.
\]
Then
\[
\text{plim} \left( \frac{\beta_k - \beta}{\gamma_k - \gamma} \right) = \left[ \frac{\Pi'_1 M \Pi_1}{M_1 \Pi_1} \quad \frac{\Pi'_1 M_{11}}{M_1 \Pi_1} \right]^{-1} \left[ \frac{\Pi'_1 M \Pi_2 \beta_2 + \Pi'_1 M_{21}}{M_1 \Pi_2 \beta_2 + M_1 \beta_2} \right] \\
= \left[ \frac{(\Pi'_1 CP_1)^{-1} - (\Pi'_1 CP_1)^{-1} \Pi'_1 M_{11}}{-M^T_{11} M_1. \Pi_1 (\Pi'_1 CP_1)^{-1}} \right] \\
M^T_{11} + M^{-1}_{11} M_1. \Pi_1 (\Pi'_1 CP_1)^{-1} \Pi'_1 M_{11}^{-1} \\
\left[ \frac{\Pi'_1 M \Pi_2 \beta_2 + \Pi'_1 M_{21}}{M_1 \Pi_2 \beta_2 + M_1 \beta_2} \right] \\
= \left[ (\Pi'_1 CP_1)^{-1} \Pi'_1 C (\Pi_2 \beta_2 + I_2 \gamma_2) \right] \\
M^{-1}_{11} M_1. (I - \Pi'_1 (\Pi'_1 CP_1)^{-1} \Pi'_1 C) (\Pi_2 \beta_2 + I_2 \gamma_2) \\
\text{Q.E.D. Lemma 2.}
\]

Then
\[
\overline{P}_Z Y' \beta_k = \overline{P}_Z (y, Y_1) \left( \begin{array}{c} -1 \\ \widehat{\beta}_k \end{array} \right) \\
= \overline{P}_Z (-y + Y_1 \beta_k) \\
= \overline{P}_Z (Y_1 \beta_k - Y_1 \beta - Z_1 \gamma_1 - Z_2 \gamma_2 - Y_2 \beta_2 - u) \\
= \overline{P}_Z (Y_1 (\beta_k - \beta) - Z_2 \gamma_2 - Y_2 \beta_2 - u) \\
= \overline{P}_Z ((Z \Pi_1 (\beta_k - \beta) - Z_2 \gamma_2 - Z \Pi_2 \beta_2) \\
+ (-UB_1(\beta_k - \beta) + UB_2 \beta_2 - u)) \\
= \overline{P}_Z (N + S),
\]
where
\[
N = Z \Pi_1 (\beta_k - \beta) - Z_2 \gamma_2 - Z \Pi_2 \beta_2
\]
and
\[
S = -UB_1(\beta_k - \beta) + UB_2 \beta_2 - u.
\]

Now
\[
\text{plim} l_i(k) = \text{plim} \frac{(N' + S') \overline{P}_Z (N + S)}{(N' + S') \overline{P}_Z (N + S)} \\
= \frac{\text{plim} (N' + S') \overline{P}_Z (N + S)}{T} \\
= \frac{\text{plim} N' \overline{P}_Z N}{T} + \frac{\text{plim} S' \overline{P}_Z S}{T}.
\]
Now $N'\bar{P}_Z = 0$ by inspection, so that $\text{plim} \, N'\bar{P}_Z N/T = 0$. Let

$$e = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

a $G$-dimensional unit vector. Then

$$S = -U[B_1(\beta_k - \beta_1) + B_2 \beta_2 - e]$$

$$= -UD,$$

where $D = B_1(\beta_k - \beta_1) + B_2 \beta_2 - e$, and let

$$\bar{D} = \text{plim} \, D$$

$$= \text{plim} \, B_1(\beta_k - \beta_1) + B_2 \beta_2 - e$$

$$= B_1(\Pi_1'C\Pi_1)^{-1}\Pi_1'C(\Pi_2\beta_2 + \gamma_2) + B_2 \beta_2 - e,$$

$$e = B^{-1} \begin{pmatrix} -1 \\ \beta_1 \\ \beta_2 \\ 0 \end{pmatrix} = (b, B_1, B_2, B_3) \begin{pmatrix} -1 \\ \beta_1 \\ \beta_2 \\ 0 \end{pmatrix} = -b + B_1 \beta_1 + B_2 \beta_2.$$

Hence $B_2 \beta_2 - e = b - B_1 \beta_1$. Then

$$\bar{D} = b + B_1[(\Pi_1'C\Pi_1)^{-1}\Pi_1'C(\Pi_2\beta_2 + \gamma_2) - \beta_1]$$

$$= (b, B_1, B_2, B_3) \begin{pmatrix} 1 \\ (\Pi_1'C\Pi_1)^{-1}\Pi_1'C(\Pi_2\beta_2 + \gamma_2) - \beta_1 \\ 0 \\ 0 \end{pmatrix}.$$

Now if $\bar{D}$ were zero, then because of the invertibility of $B^{-1} = (b, B_1, B_2, B_3)$, as a consequence, it would be true that

$$\begin{pmatrix} 1 \\ (\Pi_1'C\Pi_1)^{-1}\Pi_1'C(\Pi_2\beta_2 + \gamma_2) - \beta_1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$
which is impossible. Hence $\bar{D} \neq 0$. Then
\[
\text{plim} \frac{S'\bar{P}_z S}{T} = \bar{D}' \text{plim} \left( \frac{U'\bar{P}_z U}{T} \right) \bar{D} = \bar{D}' \Sigma \bar{D} \text{plim} \left( \frac{T - K}{T} \right) = \bar{D}' \Sigma \bar{D} > 0.
\]
Also, similarly,
\[
\text{plim} \frac{S'\bar{P}_z S}{T} = \bar{D}' \Sigma \bar{D} \text{plim} \left( \frac{T - K_1}{T} \right) = \bar{D}' \Sigma \bar{D} > 0.
\]
Finally it remains for us to analyze the term $\text{plim} \bar{N}'\bar{P}_z, N/T$.
\[
\bar{N} = Z\Pi_1(\Pi'_1C\Pi_1)^{-1}\Pi'_1C(\Pi_2\beta_2 + I_2\gamma_2) - Z_2\gamma_2 - Z\Pi_2\beta_2
\]
\[
= Z[I - \Pi_1(\Pi'_1C\Pi_1) - 1\Pi'_1C](\Pi_2\beta_2 + I_2\gamma_2).
\]
Then
\[
\text{plim} \frac{\bar{N}'\bar{P}_z, N}{T} = (\gamma'_2I'_2 + \beta'_2\Pi_2)[1 - C\Pi_1(\Pi'_1C\Pi_1)^{-1}\Pi'_1]
\]
\[
\times C[I - \Pi_1(\Pi'_1C\Pi_1)^{-1}\Pi'_1C](\Pi_2\beta_2 + I_2\gamma_2)
\]
\[
= (\gamma'_2I'_2 + \beta'_2\Pi_2)[C - C\Pi_1(\Pi'_1C\Pi_1)^{-1}\Pi'_1C](\Pi_2\beta_2 + I_2\gamma_2).
\]
Now $M$ is a square, $K$-dimensional positive definite matrix. Therefore it has a square-root matrix $R$ such that $M = RR'$ and $M^{-1} = R'^{-1}R^{-1}$. Consider
\[
C = M - M_{11}M_{11}^{-1}M_1.
\]
\[
= R[I - R^{-1}M_{11}M_{11}^{-1}M_1, R'^{-1}]R'.
\]
Because
\[
M_{11}M_{11}^{-1}M_1 = M_{11}M_{11}^{-1}M_1, \begin{bmatrix} I \\ 0 \end{bmatrix} = M_{11}, \begin{bmatrix} I \\ 0 \end{bmatrix} = M_{11},
\]
the quantity in brackets is idempotent. Since it is also symmetric, it is a projection. Also
\[
\text{tr} [I - R^{-1}M_{11}M_{11}^{-1}M_1, R'^{-1}] = K - K_1
\]
and
\[
[I - R^{-1}M_{11}M_{11}^{-1}M_1, R'^{-1}]R^{-1}M_{11} = 0.
\]
Finally \( \rho(R^{-1}M_1) = \rho(M_1) = K_1 \) (using Lemma 1), so
\[
I - R^{-1}M_1M_1^{-1}R^{-1} = \bar{P}_{R^{-1}M_1},
\]
and
\[
C = R\bar{P}_{R^{-1}M_1}R'.
\]
Consider now
\[
C - C\Pi_1(P_i^1C\Pi_1)^{-1}\Pi_1' = R\{P_{R^{-1}M_1} - \bar{P}_{R^{-1}M_1}R\Pi_1(P_i^1R\bar{P}_{R^{-1}M_1}R\Pi_1)^{-1}\Pi_1'R\bar{P}_{R^{-1}M_1}\}R'.
\]
The quantity in curly brackets is again easily seen to be symmetric and idempotent, and hence a projection. Its trace is \( K - K_1 - G_1 \), and it annihilates \( (R^1\Pi_1, R^{-1}M_1) \).
Since \( \rho(R^1\Pi_1, R^{-1}M_1) = \rho(M\Pi_1, M_1) = G_1 + K_1 \), again using Lemma 1, we have
\[
C - C\Pi_1(P_i^1C\Pi_1)^{-1}\Pi_1' = R\bar{P}_{R^1\Pi_1, R^{-1}M_1}R'.
\]
Finally
\[
\text{plim } \frac{N'\bar{P}_zN}{T} = (\gamma_2P_2' + \beta_2'\Pi_2)R\bar{P}_{R^1\Pi_1, R^{-1}M_1}R'(\Pi_2\beta_2 + I_2\gamma_2).
\]
Clearly this is a quadratic form which is non-negative, and which is positive if and only if \( R^1\Pi_1, R^{-1}M_1, \) and \( R'(\Pi_2\beta_2 + I_2\gamma_2) \) are linearly independent vectors. But this is true if and only if \( M_{\Pi_1}, M_1, \) and \( M(\Pi_2\beta_2 + I_2\gamma_2) \) are linearly independent, which they are by Lemma 1. Hence
\[
\text{plim } \frac{N'\bar{P}_zN}{T} = p > 0.
\]
In conclusion, under the alternative hypothesis,
\[
\text{plim } l_1(k) = \frac{p + D^2D'}{D^2D'} > 1
\]
for all \( k \) satisfying assumption (i)(a) of the theorem.

To show that \( \text{plim } l_1(\lambda) = \lambda > 1 \) under the alternative hypothesis, recall (see for example, [8, p. 343, equation 6.88] and Hood and Koopmans [9, p. 181])
\[
|\Pi_\lambda^{*\beta}**M^{*\beta**}\Pi_\lambda^{*\beta**} - \text{plim} (\lambda - 1)\Omega_{\lambda\lambda} = 0,
\]
where \( *M^{*\beta**} \) and \( \Omega_{\lambda\lambda} \) are certain positive definite matrices, and \( \Pi_\lambda^{*\beta**} \) is a submatrix of \( \Pi_\lambda \) of order \( G_1 + 1 \).

Clearly \( \text{plim} (\lambda - 1) = 0 \) if \( \rho(\Pi_{\lambda}^{*\beta}) < G_1 \) and \( \text{plim} (\lambda - 1) > 0 \) if \( \rho(\Pi_{\lambda}^{*\beta}) = G_1 \). Since \( \Pi_1 \) is of full rank, (Lemma 1 of Theorem 3), so are each of its submatrices. Hence \( \rho(\Pi_{\lambda}^{*\beta}) = G_1 \), so \( \text{plim} (\lambda - 1) > 0 \) under the alternative hypothesis.

Thus \( \text{plim } l_1(k) > 1 \) for all the \( k \)-class estimators included under assumption (i).

The analysis of \( l_2(k) \) is simplified by the observation that \( \hat{\beta}_2^2 \) is a consistent estimate of \( \beta_2^2 \) for equation (3) (see, for example, [5, p. 201]). Thus no equivalent
to Lemma 2 is needed. Now
\[ P_{Z_1, Z_2} Y^2_{k*} = P_{Z_1, Z_2} (-y + Y_1 \hat{\beta}_{1k} + Y_2 \hat{\beta}_{2k}) \]
\[ = P_{Z_1, Z_2} (Y_1 (\hat{\beta}_{1k} - \beta_1) + Y_2 (\hat{\beta}_{2k} - \beta_2) - Z_1 \gamma_1 - Z_2 \gamma_2 - u) \]
\[ = P_{Z_1, Z_2} (Z \Pi_1 (\hat{\beta}_{1k} - \beta_1) + Z \Pi_2 (\hat{\beta}_{2k} - \beta_2) - u). \]

Then
\[ \lim \frac{\beta^2_{k*} Y^2_{k*} P_{Z_1, Z_2} Y^2_{k*}}{T} = \lim \frac{u P_{Z_1, Z_2} u}{T} \]
\[ = \sigma_{11} \lim \frac{T - K_1 - K_2}{T} = \sigma_{11}. \]

Similarly
\[ \lim \frac{\beta^2_{k*} Y^2_{k*} P_{Z_1, Z_2} Y^2_{k*}}{T} = \sigma_{11}; \]
thus \( \lim l_2(k) = 1 \), and hence
\[ \lim \frac{l_1(k)}{l_2(k)} = \lim \frac{l_1(k)}{l_2(k)} > 1. \]

As \( T \to \infty \), when the null hypothesis is true,
\[ \frac{T - K + L_2}{L_1 - L_2} (\frac{[l_1(k)/l_2(k)] - 1}{1}) \]
approaches a \( \chi^2 \) distribution with \( L_1 - L_2 \) degrees of freedom. Let \( \chi_a \) be the \( a \)th percentile of a \( \chi^2 \) distribution with \( L_1 - L_2 \) degree of freedom. Then the power of the test is the probability, under the alternative, that
\[ \frac{T - K + L_2}{L_1 - L_2} [l_1(k)/l_2(k) - 1] > \chi_a. \]

Under any alternative hypothesis of the type specified in the theorem,
\[ \lim (l_1(k)/l_2(k)) = \lim \frac{l_1}{l_2} > 1. \]

Then
\[ \lim \frac{T - K + L_2}{L_1 - L_2} [l_1(k)/l_2(k) - 1] \to \infty, \]
so the power approaches 1. Therefore the test is consistent against the specified class of alternatives. \( Q.E.D. \)

Carnegie-Mellon University

Manuscript received July, 1970; revision received December, 1972.
REFERENCES


