Testing Overidentifying Restrictions
in Unidentified Models

John G. Cragg*
Department of Economics
University of British Columbia
and
Stephen G. Donald
Department of Economics
Boston University

July 28, 1996

Abstract

This paper considers the properties of the Anderson-Rubin, Basmann and Byron tests of overidentifying restrictions in instrumental-variables models when the instruments fail to satisfy the rank condition for identification. Under the null, the tests have an asymptotic size that is less than the nominal size. When the null is false, the last two tests are shown to be consistent while the Anderson-Rubin test is inconsistent and has power no larger than the nominal size of the test. Local power considerations are developed to examine the properties of the tests when the identification is weak (in the sense that we consider a sequence of models that goes to an unidentified model) and when the misspecification of excluding the instrumental variables is weak in the same sense. A small Monte Carlo study indicates that the asymptotic results are helpful in evaluating likely small-sample performance.

Keywords: Overidentification tests; under identification; weak instrumental variables; rank of a submatrix; two-stage least-squares.

*Research supported by the Social Sciences and Humanities Research Council of Canada under grant 94-0737. We are indebted to Yanjun Liu, Alice Nakamura and participants at a seminar at UBC for helpful comments.
1. Introduction

Considerable interest has been expressed recently in the problems of instrumental-variables estimation when the instruments are poor in the sense of failing or almost failing the rank condition for identification. Phillips (1989) derived the properties of the IV estimator when the instruments are completely unrelated to the endogenous variables. Staiger and Stock (1993) have derived the properties when the relationship is weak. Nelson and Startz (1990a) studied the small-sample distributions in a special case, although Maddala and Jeong (1992) have pointed out difficulties with the representativeness of the results. It is apparent that the sampling properties are quite poor, suggesting that tests for lack of identification, such as the one proposed by Koopmans and Hood (1953) and generalized by Cragg and Donald (1993) may be useful indicators of this problem.

Possible failure of overidentification conditions is logically distinct from the question of identification, being concerned with whether there is correlation of the instruments with the dependent variable of interest in a particular equation beyond that accounted for by the specification. By contrast, the identification question is whether the correlations with the variables needing instrumentation are rich enough to allow the instruments to serve their function.

Since over-identification tests are often used in model formulation it is useful, given the potential problem of having poor instruments, to know what properties such tests have when the instruments are weak. Some progress has been made by Cragg and Donald (1993). They show that overidentification tests which are framed as tests of the rank of a matrix, in the spirit of the Anderson and Rubin (1949) test, have no power when the parameters are not identified. Two other tests have been suggested in the literature. They are asymptotically equivalent to the Anderson and Rubin (A-R) test when the parameters are identified, but the tests are rather different when the parameters are not identified. Basmann (1960) suggested the first of these tests. It may be considered to be a GMM test in the spirit of the overidentification tests due to Hansen (1982); Byron (1974) suggested the second test, developing it in the spirit of the Wald test. Little is know
about the behavior of these tests in finite samples although Nelson and Startz (1990b) have investigated Basmann's test in a small Monte Carlo study. They found that the test may have a size much larger than the nominal size, at least for the particular parameters chosen in their Monte Carlo investigation. This finding is partly elucidated in the Monte Carlo investigation that Steiger and Stock (1993) made of the tests which throws further light on their small-sample properties.

In this paper we conduct a systematic study of the three tests of overidentification when the instruments are poor. We consider the limiting properties of the tests when the rank condition for identification fails. The main result is that none of the usual tests is of correct size when the parameters are not identified, but that the Basmann and Byron tests are consistent while the A-R test is not. The problems associated with these standard tests lead us to suggest a two-step procedure. In it we first estimate the extent to which (if at all) the parameters are not identified in the sense of the order by which the necessary and sufficient condition for identifiability fails. We then use this information to produce a test which is of the right size asymptotically, provided that the estimate of the rank of the matrix is consistent. Thus we propose a method that allows one to check consistently whether the exclusion restrictions giving rise to the overidentification conditions on the structural equation are valid.

Local-power investigations suggest more about the differences in the behavior of the tests when identification is weak. We consider cases where the parameters are not identified and the misspecification that renders the null hypothesis incorrect is treated as a sequence of misspecifications $O(N^{-1/2})$, where $N$ is the number of observations. We also investigate a sequence of matrices which differ from one of lesser rank by virtue of coefficients of the same order of magnitude, a procedure already used by Steiger and Stock (1993).

These investigations produce some ambiguous findings and give only asymptotic results. Furthermore, in the relevant case, the results do not in general seem to agree with the finding of Nelson and Startz (1990b) that the size of the Basmann test seems to be too large; our results suggest that it is too small. A Monte Carlo study is used to
investigate some aspects of the tests when identifiability conditions fail or almost fail.

Having thus established these various aspects of the asymptotic behavior of the tests, we consider the implications of the results for practice and provide a summary of our main findings in the concluding section.

2. Model and Tests

The model to be considered is the standard linear model,

\[ y_{2i} = y'_{2i} \gamma + x'_{2i} \beta_1 + x'_{2i} \beta_2 + u_i, \quad i = 1, \ldots, N; \]  

(1)

\(N\) being the sample size. Here \(y_{2i}\) is a scalar, \(y_{2i}\) is a \(G\)-vector of other endogenous variables, \(x_{1i}\) is a \(K_1\)-vector of predetermined variables and \(x_{2i}\) is a \(K_2\)-vector of other predetermined variables. We assume that \(E(u_i|x_{1i}, x_{2i}) = 0\). Define \(K = K_1 + K_2\).

Denote the reduced form relationship between \(y_{2i}\) and the predetermined variables as

\[ y'_{2i} = x'_{1i} \Pi_{12} + x'_{2i} \Pi_{22} + v'_{2i}; \]  

(2)

where \(E(v_{2i}|x_{1i}, x_{2i}) = 0\). Together equations (1) and (2) imply a reduced-form equation for \(y_{1i}\):

\[ y_{2i} = x'_{1i}(\Pi_{12} \gamma + \beta_1) + x'_{2i}(\Pi_{22} \gamma + \beta_2) + u_i + v_{2i} \gamma, \]

or

\[ y_{1i} = x'_{1i} \Pi_{11} + x'_{2i} \Pi_{21} + v_{1i}. \]

Here

\[ \Pi_{11} = \Pi_{12} \gamma + \beta_1; \]

\[ \Pi_{21} = \Pi_{22} \gamma + \beta_2. \]

We let \(\Sigma = E[(v_{1i}, \nu_{i2})(v_{1i}, \nu_{i2})']\), which is assumed to be the same for all \(i\). The restriction \(\beta_2 = 0\) gives rise to the possibility of using the \(x_{2i}\) variables as instruments for \(y_{2i}\). Using this constraint makes the parameters \(\gamma\) and \(\beta_1\) identified if and only if \(\rho(\Pi_{22}) = G\). (The notation \(\rho(\cdot)\) denotes the rank of the argument.)
2.1. Tests of Overidentification

Let

\[ \Pi_2 = [\Pi_{21} \quad \Pi_{22}] . \]

The A-R test arises from noting that if \( \rho(\Pi_{22}) = G \) and \( \beta_2 = 0 \); then \( \rho(\Pi_2) = G \); if \( \rho(\Pi_{22}) = G \) and \( \beta_2 \neq 0 \) then \( \rho(\Pi_2) = G + 1 \) provided that

\[ K_2 > G. \] (3)

Therefore tests of the null hypothesis\(^1\)

\[ H_0 : \beta_2 = 0 \] (4)

can be conducted by testing the hypothesis

\[ H_0 : \rho(\Pi_2) = G \] (5)

against the alternative

\[ H_a : \rho(\Pi_2) = G + 1. \]

Given condition (3), hypothesis (5) can be tested under standard assumptions by considering the smallest root, \( \hat{\lambda} \), of the determinantal equation

\[ |\hat{\Pi}_2'X_2'M_1X_2\hat{\Pi}_2 - \lambda\hat{\Sigma}| = 0; \]

that is, the smallest eigenvalue of \( \hat{\Pi}_2'X_2'M_1X_2\hat{\Pi}_2 \) in the metric of \( \hat{\Sigma} \). Here \( X = [X_1 \quad X_2] \)
is the \( N \times K \) matrix of exogenous variables with typical row \( \{x_n 1 \quad x_n 2\} \) partitioned into \( K_1 \) and \( K_2 \) columns; \( M_1 = I - X_1(X_1'X_1)^{-1}X_1' \); \( M \) will denote the corresponding matrix constructed using \( X \). \( \hat{\Pi}_2 \) is the unrestricted reduced-form estimator of \( \Pi_2 \) given by

\[ \hat{\Pi}_2 = (X_2'M_1X_2)^{-1}X_2'M_1Y \] (6)

\(^1\)Strictly speaking, all the tests we consider test the null hypothesis that \( \beta_2 = \Pi_{22} a \) for some \( a \) (which may be zero). However, for non-zero \( a \), the parameters \( \gamma \) and \( \beta \) are not well defined and specifying that \( a = 0 \) so that \( \beta_2 = 0 \) is simply an identifying restriction imposing no further constraint on the model.
and

\[ \hat{\Sigma} = Y'MY/(N - K) \]  

(7)

is the OLS estimate\(^2\) of \(\Sigma\). Finally \(Y = [Y_1 \quad Y_2]\) is the \(N \times (G + 1)\) matrix of endogenous variables with typical row \(\{y_{1i} \quad y_{2i}\}\).

The following assumption states that the data are such that the standard asymptotic properties hold for these estimates:

**Assumption 1:** The variables \(X\) and \(Y\) obey the conditions:

1. \(K_2 > G\);

2. \(\beta[X] = K\);

3. \(\text{plim}(\hat{\Pi}_2) = \Pi_2\);

4. \(\text{plim} \hat{\Sigma} = \Sigma\);

5. \([X_2' M_1 X_2]^{1/2}(\hat{\Pi}_2 - \Pi_2) \xrightarrow{d} (z_1, Z_2)\) with \(z = \text{vec}(z_1, Z_2) \sim N(0, \Sigma \otimes I_{K_2})\).

6. \(X_2' M_1 X_2 / N \xrightarrow{p} \bar{C}\), a positive definite matrix (adopted for notational simplicity) and that \(X_2' X_2 / N \xrightarrow{p} \bar{E}\) a positive definite matrix.

Under assumption 1 and the null hypothesis (4), as is well known,

\[ \hat{\lambda} \xrightarrow{d} \chi^2(K_2 - G) \]

when the parameters are identified. With identification and assumption 1, \(\hat{\lambda}\) diverges when \(\beta_2 \neq 0\).

We shall also be considering local alternatives through a sequences of models of the form

\[ \beta_2(N) = \frac{b}{\sqrt{N}}, \]

(8)

\(^2\)It is not unusual to use a different divisor in (7), for example \(N\) rather than \((N - K)\), a variation which will not affect our consideration of asymptotic properties. We shall use (7) when actually calculating the test statistics.
where $b \neq 0$, and the other parameters are constant. Under such a local alternative and assumption 1,

$$\hat{\lambda} \xrightarrow{d} \chi^2((K_2 - G), \tilde{\lambda})$$

where the non-centrality parameter $\tilde{\lambda}$ is the limit as $N \to \infty$ of the smallest eigenvalue of $N\Sigma^{-1/2}\Pi_0\tilde{C}\Pi_2\Sigma^{-1/2}$, and is displayed in expression (11) below.

Although Anderson and Rubin developed their test in the LIML context\(^3\), use of $\hat{\lambda}$ as a test statistic also arises from noting that

$$\hat{\lambda} = \min_{\pi} N\{((\hat{\pi} - \pi)^' \Sigma^{-1} \Sigma^{-1} (\hat{\pi} - \pi) \quad \text{s.t.} \quad \rho[\Pi_2] = G\}$$

where $\pi = \text{vec}(\Pi_2)$ and $\hat{\pi}$, and $\pi$, the variables with respect to which the minimization is taken, are defined correspondingly. Cf. eg. Cragg and Donald (1993).

The test due to Basmann (1960) uses the test statistic

$$\hat{\theta} = \frac{\hat{u}'X(X'X)^{-1}X'\hat{u}}{\hat{s}_1^2}$$

where

$$\hat{u} = Y_1 - Y_2\hat{\gamma} - X_1\hat{\beta}_1$$

and\(^4\)

$$\hat{s}_1^2 = \frac{\hat{u}'\hat{u}}{N}.$$ 

$(\hat{\gamma}',\hat{\beta}_1')'$ denotes the 2SLS estimates of the coefficients in (1). Since this statistic is explicitly testing the orthogonality of $u$ and $X$, it is a special case of the Hansen (1982)

\(^3\)As presented here, this is not exactly the Anderson and Rubin test statistic (which is based on the usual specification and approach used for LIML estimation) but this statistic is easier to analyze and, moreover, is asymptotically equivalent for all the cases of interest in this paper. This follows since the statistic used by Anderson and Rubin, $N\log(1 + \hat{\lambda}/N) = \hat{\lambda} + O_p(N^{-1})$ under $H_0$ so that without loss asymptotically we can analyze $\hat{\lambda}$. The same is true when we consider local power of the test statistic. Under $H_0$, both diverge, but then $\hat{\lambda}/N$ (which is of $O_p(1)$) may be poorly approximated by $\log(1 + \hat{\lambda}/N)$.

\(^4\)Again, a different divisor might be used, say $(N - K)$, which is the one we actually used in the calculations reported below. In the event that different denominators are used in calculating $\hat{s}_1^2$ and $\tilde{\Sigma}$, the relationship among test statistics developed below in lemma 3(a) may not hold in finite samples.
test for overidentifying restrictions in GMM models, as indeed are the other tests. Again
under assumption 1, if the model is identified so that $\rho(\Pi_{22}) = G$ and if $\beta_2 = 0$, then

$$\hat{\theta} \xrightarrow{d} \chi^2(K_2 - G).$$

$\hat{\theta}$ diverges when $\beta_2 \neq 0$. Furthermore, under local alternative (8), $\hat{\theta} \xrightarrow{d} \chi^2((K_2 - G), \hat{\theta})$
where

$$\tilde{\theta} = \tilde{b}'[\tilde{C} - \tilde{C}\Pi_{22}\tilde{C}]^{-1}\Pi_{22}\tilde{C}\tilde{b}/(1 - \gamma')\Sigma(1 - \gamma')' = \lambda. \quad (11)$$

The statistic suggested by Byron (1974) arises from the same considerations as the
A-R test, but imposes a particular normalization on the linear combination that arises
from $\rho(\Pi_{22}) = \rho(\Pi_{22})$. This rank is presumed to be $G$ in the original formulation of the
test. The procedure regresses $C_{1/2}\tilde{\Pi_{21}}$ on $C_{1/2}\tilde{\Pi_{22}}$, where $C_{1/2}$ is a matrix such that
$C_{1/2}C_{1/2} = C$, and then tests whether there is only as much variance in the residuals
from the regression as can be inferred from the specified model under hypothesis (4).
More precisely, let

$$\hat{\gamma} = \tilde{C}\tilde{\Pi}_{22}\tilde{C}\tilde{\Pi}_{22}^{-1}\hat{\Pi}_{22}\tilde{C}\hat{\Pi}_{21}$$

and

$$\tilde{\beta}_2 = (1, -\hat{\gamma}')\tilde{\Sigma}(1, -\hat{\gamma})'.$$

Under assumption 1 and $H_0 : \beta_2 = 0$, the test statistic

$$\hat{\nu} = N\hat{\Pi}_{21}'[C - C\Pi_{22}\hat{\Pi}_{22}'C\Pi_{22}^{-1}\hat{\Pi}_{22}C]\hat{\Pi}_{21}/\tilde{\beta}_2^2 \xrightarrow{d} \chi^2(K_2 - G). \quad (12)$$

The numerator of this statistic is the same as that in (10) defining $\hat{\theta}$; that is, letting $\nu$
be the numerator,

$$\nu = N\hat{\Pi}_{21}'[C - C\hat{\Pi}_{22}\hat{\Pi}_{22}'C\hat{\Pi}_{22}^{-1}\hat{\Pi}_{22}C]\hat{\Pi}_{21} = \hat{\nu}'X(X'X)^{-1}X'\hat{\nu}. \quad (13)$$

The denominators of the two statistics are different. Again, when $\beta_2 \neq 0$ and $\rho(\Pi_{22}) = G$,
$\hat{\nu}$ diverges under assumption 1 while its limiting distribution under the local alternative
(8) is $\chi^2((K_2 - G), \hat{\theta})$. 
The three statistics in general produce different values for the test statistic. It is well known, however, that, under assumption 1 when \( \rho(\Pi_{22}) = G \) and \( \beta_2 = 0 \),

\[
\text{plim}(\hat{\theta} - \lambda) = \text{plim}(\hat{\tau} - \lambda) = 0.
\]

This is not true when \( \rho(\Pi_{22}) < G \).

2.2. Testing with Inadequate Instruments

What is not well known about the tests we are considering is their properties when the instruments are weak. This is true either when weakness is meant in the sense of Phillips (1989) where the structural parameters \( \{\gamma, \beta_1\} \) are not identified because \( \rho(\Pi_{22}) < G \), or in the sense of Staiger and Stock (1993) where \( \Pi_{22} \) converges with increases in the sample size to a matrix of smaller rank.

The different tests are affected in different ways by the following assumption:

Assumption 2: \( \rho(\Pi_{22}) = G_0 < G \).

With assumption 2, the model is not identified although the instruments are orthogonal to the residuals in equation 1.

The first result summarizes findings of Schott (1984) and Cragg and Donald (1993).

**Lemma 1:** Given assumptions 1 and 2,

(a) under \( H_0 : \beta_2 = 0 \),

\[
\lim_{n \to \infty} P[\hat{\lambda} \leq c] > \int_0^c d\chi^2(K_2 - G);
\]

(b) under \( H_a : \beta_2 \neq 0 \),

\[
\lim_{n \to \infty} P[\hat{\lambda} \leq c] \geq \int_0^c d\chi^2(K_2 - G)
\]

with equality only when \( G_0 = G - 1 \).

The proof of part (a) is found in Schott (1984) or Cragg and Donald (1993). Part (b) is an immediate corollary of the proof of theorem 8 below.

Thus when the structural parameters are not identified, the test statistic \( \hat{\lambda} \) is stochastically dominated by the supposed \( \chi^2 \) distribution\(^5\) under \( H_0 \) and under \( H_a \) the power

---

\(^5\)In fact, we can go farther in characterizing the asymptotic distribution under \( H_0 \) and assumption
is at most equal to the nominal size. The reason for the latter finding is not surprising: under assumption 2, the null hypothesis being tested, if interpreted to be $\rho[\Pi_2] \leq G$, is actually correct when $\beta_2 \neq 0$ and the test statistic retains its properties when the suggested rank is less than or equal to the maximum possible rank minus one even though its rank is larger than the rank of the critical submatrix.

To investigate the behavior of the Basmann and Byron tests when the parameters are not identified, we first show that under $H_0$ the numerator of the statistics, when suitably normalized, does converge in distribution to $\chi^2(K_2 - G)$. Let $H$ be the eigenvectors of $\Pi_{22}^\top \Pi_{22}$ arranged so that the corresponding eigenvalues are decreasing, and partition this matrix as $H = [H_1 \quad H_2]$ of $G_0$ and $(G - G_0)$ columns respectively. Let

$$
\delta_1 = H_1^\top \gamma.
$$

Define the $(G + 1) \times (G + 1)$ matrix $H^*$ as

$$
H^* = \begin{bmatrix}
1 & 0 \\
0 & H
\end{bmatrix}
$$

and the $(G + 1) \times (G + 1)$ matrix $\Omega$ as

$$
\Omega = H^* \Sigma H^* = \begin{bmatrix}
\omega_{11} & \omega_{12} & \omega_{13} \\
\omega_{12}^\top & \Omega_{22} & \Omega_{23} \\
\omega_{13}^\top & \Omega_{32} & \Omega_{33}
\end{bmatrix},
$$

where the partitioning is into 1, $G_0$ and $(G - G_0)$ rows and columns. With these definitions, we obtain the following results about the numerator of the statistics, $\nu$:

**Lemma 2:** Given assumptions 1 and 2,

(a) under $H_0$ : $\beta_2 = 0$,

$$
\nu / \sigma^2 \xrightarrow{\Delta} \chi^2(K_2 - G)
$$

2. Let $\lambda_{\min}(\cdot)$ denote the smallest eigenvalue of the argument and let $W_q(p, I)$ denote a Wishart matrix of dimension $q$ with degrees of freedom $p$ and variance parameter $I$. Then as shown by Schott (1984) using the Poincarè Separation theorem, under assumptions 1 and 2 and under $H_0$

$$
\lambda \overset{d}{=} \lambda_{\min}(W_{G - G_0}(K_2, I))
$$

while under $H_0$ : $\beta_2 \neq 0$,

$$
\lambda \overset{d}{=} \lambda_{\min}(W_{G - G_0}(K_2, I)).
$$
where
\[
\sigma^2 = (1, -\delta_1^T) \left[ \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \Omega_{22} \end{pmatrix} - \begin{pmatrix} \omega_{13} \\ \omega_{12} \end{pmatrix} \Omega_{23}^{-1} \begin{pmatrix} \omega_{13} & \Omega_{23} \end{pmatrix} \right] \begin{pmatrix} 1 \\ -\delta_1 \end{pmatrix};
\]  
(17)

(b) under \( H_0 : \beta_2 = 0 \),
\[
\nu/N \xrightarrow{p} h > 0 \quad a.s.;
\]

(c) under \( H_0 : \beta_2 = \bar{\beta}/\sqrt{N} \),
\[
\lim_{N \to \infty} P[\nu/\sigma^2 \leq c] < \int_{0}^{c} d\chi^2(K_2 - G).\]

The problems with the test statistics arise from the denominators, which the following result shows are too large for the test statistic to be distributed as \( \chi^2(K_2 - G) \):

Lemma 3: Given assumptions 1 and 2,

(a) \( \delta_1^2 = \delta_2^2 + a_1 \) where \( a_1 \) is a random variable which is positive with probability 1 when \( \Pi_{22} \neq 0 \) and is zero when \( \Pi_{22} = 0 \); and

(b) \( \delta_2^2 \xrightarrow{p} \sigma^2 + a_2 \), where \( a_2 \) is a random variable which is positive with probability 1.

We can summarize these results as follows:

Theorem 4: Given assumptions 1 and 2,

(a) \( \hat{\delta} \leq \hat{\tau} \); and

(b) under \( H_0 : \beta_2 = 0 \),
\[
\lim_{N \to \infty} P[\hat{\tau} \leq c] > \int_{0}^{c} d\chi^2(K_2 - G); \text{ while}
\]

(c) under \( H_0 : \beta_2 \neq 0 \),
\[
\lim_{N \to \infty} P[\hat{\tau} < c] = 0
\]

for any \( c < \infty \).

The results under local alternative (8) are ambiguous as to whether the power is greater asymptotically than the nominal size. The power is, however, greater than the actual asymptotic size under the null since the numerator's distribution stochastically dominates its distribution under the null hypothesis, as lemma 3 shows, while the denominators of the statistics are the same asymptotically under the null and the local
alternative. The ambiguity arises because this feature may not be enough to overcome the size problem produced by the denominators' already being too large.

The intuition behind theorem 4 is that in estimating the parameters in unidentified models, the minimization problem to which 2SLS is the solution exploits not only the dependence of \( Y_1 \) on \( Y_2 \) through the structural relationship but also its dependence through the correlation of its residuals with those of \( Y_2 \). The formula for estimating the variance of the resulting sum of squares does not recognize this aspect of the minimization and as a result gives too high a value. That \( \hat{\theta} \) is smaller than \( \hat{\gamma} \), which is a feature of the statistics in finite samples, as the derivation of the result in the appendix shows, arises because the Basmann approach implicitly uses estimates of \( \Sigma \) derived from the constrained reduced-form based on the 2SLS estimates rather than the unconstrained OLS estimate. Therefore, it implicitly uses a larger covariance matrix for the reduced-form residuals.

The result that the Byron and Basmann tests are consistent while the A-R test is not arises because these tests impose a normalization which requires that the first column of the submatrix be a linear function of the other columns while the A-R test only requires that some column of the submatrix be a linear function of the remaining ones.

As a by-product of obtaining the result in theorem 4, we derived an expression for the limit of the two-stage least-squares estimator of \( \gamma \) which is displayed in equation (A.8) in the appendix. It shows that the 2SLS estimator is not consistent and instead converges in probability to a random vector. This feature of 2SLS estimates of unidentified parameters has been discussed at more length in Phillips (1989) and in Staiger and Stock (1993). The mean of the random vector is not \( \gamma \); instead, when \( \Pi_{22} = 0 \) its mean does not even depend on \( \gamma \), but is simply the regression coefficients relating the first reduced-form residual to the others. When \( \rho[\Pi_{22}] \neq 0 \), certain linear combinations of \( \gamma \) are identifiable, just as Phillips (1989) showed that some transformations of \( \beta_1 \) are still identified when \( \Pi_{22} = 0 \). We shall exploit this fact in the next section.
2.3. Devising Tests of the Correct Size Asymptotically

The key problem revealed in subsection 2.2 is that the overidentification tests are not of appropriate size when the instruments are inadequate. For the Basmann and Byron tests, the problem stems primarily from not calculating an appropriate scale-factor to normalize the numerator of the test statistic. The problem with the A-R test arises from the facts that the size of the test is too small when the rank is actually less than under the null hypothesis and that the null hypothesis of the test is still correct under the alternative.

One way to proceed is to use a two-step procedure, estimating in the first step the rank of $\Pi_{22}$ and then using this estimate in the second step to test the over-identification hypothesis. As shown in Cragg and Donald (1996), several consistent procedures are available for estimating the rank of a matrix, based either on model-selection procedures or else on sequential up-testing of the rank-deficiency with a significance level that declines with the sample size at an appropriate rate.

Given a consistent estimate $\hat{G}_0$ of $G_0 = \rho[\Pi_{22}]$, we can proceed as follows for the A-R test. First calculate $\hat{H}_1$, the $\hat{G}_0$ eigenvectors of $\hat{\Pi}_{22} C \hat{\Pi}_{22}$ corresponding to the $\hat{G}_0$ largest eigenvalues. Define $\hat{\Pi}_{22}^R = \hat{\Pi}_{22} \hat{H}_1$, $\hat{\Pi}_R = [\hat{\Pi}_{21} \hat{\Pi}_{22}^R]$, $\hat{\Pi}_{22}^R = \Pi_{22} H_1$, $\Pi_R = [\Pi_{21} \Pi_{22}^R]$ and

$$\Omega_R = \left( \begin{array}{cc} 1 & 0 \\ 0 & \hat{H}_1' \end{array} \right) \Sigma \left( \begin{array}{cc} 1 & 0 \\ 0 & \hat{H}_1 \end{array} \right).$$

Define $\Omega_R$ correspondingly using $H_1$ and $\Sigma$. Let $\hat{\lambda}_R$ be the smallest eigenvalue of $\hat{\Omega}_R^{-1/2} \hat{\Pi}_R C \hat{\Pi}_R \hat{\Omega}_R^{-1/2}$. Using these definitions produces the following results:

**Theorem 5:** Given assumptions 1 and 2, and given that $\hat{G}_0$ estimates $G_0$ consistently,

(a) under $H_0 : \beta_2 = 0$,

$$\hat{\lambda}_R \sim \chi^2(K_2 - G_0);$$

(b) under $H_1 : \beta_2 \neq 0$,

$$\lim_{N \to \infty} P(\hat{\lambda}_R < c) = 0$$
for any \( c < \infty \); and

(c) under \( H_0 : \beta_2 = \bar{b}/\sqrt{N} \),

\[
\hat{\lambda}_R \xrightarrow{d} \chi^2((K_2 - G_0), \lambda_0)
\]

where \( \lambda_0 \) is the smallest eigenvalue of \( \Omega_R^{-1/2} \hat{\Pi}^R \hat{C} \hat{\Pi}^R \Omega_R^{-1/2} \).

That is, the two-step procedure restores the usual properties of the A-R test except that the degrees of freedom are \((K_2 - G_0)\) rather than \((K_2 - G)\).

When considering the Baslam and Byron tests two possibilities emerge, both relying on the Byron form of the test. The first version, like the adaptation of the A-R test just considered, applies the logic of the Byron test to \( \hat{\Pi}_{21} \) using \( \hat{\Pi}^R_{22} = \hat{\Pi}_{22} \hat{H}_1 \), the \( K_2 \times G_0 \) matrix derived from \( \hat{\Pi}_{22} \) by linear transformation whose probability limit has rank \( G_0 \).

Let \( \hat{\delta}_1 = (\hat{\Pi}^R_{22} \hat{C} \hat{\Pi}^R_{22})^{-1} \hat{\Pi}^R_{22} \hat{C} \hat{\Pi}_{21} \) and \( s_R^2 = (1, -\hat{\delta}_1)^T \hat{\Omega}_R (1, -\hat{\delta}_1)' \). We suggest the test statistic

\[
\hat{\tau}_R = N \hat{\Pi}_{21} [C - \hat{C} \hat{\Pi}^R_{22} (\hat{\Pi}^R_{22} \hat{C} \hat{\Pi}^R_{22})^{-1} \hat{\Pi}^R_{22} C] \hat{\Pi}_{21} / s_R^2
\]

based on the following result:

**Theorem 6:** Given the conditions of theorem 5,

(a) under \( H_0 : \beta_2 = 0 \),

\[
\hat{\tau}_R \xrightarrow{d} \chi^2(K_2 - G_0);
\]

(b) under \( H_0 : \beta_2 \neq 0 \),

\[
\lim_{N \to \infty} P(\hat{\tau}_R < c) = 0
\]

for any \( c < \infty \); and

(c) under \( H_0 : \beta_2 = \bar{b}/\sqrt{N} \),

\[
\hat{\tau}_R \xrightarrow{d} \chi^2((K_2 - G_0), \tau_R)
\]

where \( \tau_R = \bar{b}'(\hat{C} - \hat{C} \hat{\Pi}^R_{22} [\hat{\Pi}^R_{22} \hat{C} \hat{\Pi}^R_{22}]^{-1} \hat{\Pi}^R_{22} \hat{C}) \hat{b}/(1, -\hat{\delta}_1)^T \hat{\Omega}_R (1, -\hat{\delta}_1)' \).
Like the modified A-R test, the test using $\hat{\tau}_A$ has the properties that the Byron and Basmann tests have when $\rho(\Pi_{22}) = G$, though with a larger number of degrees of freedom.

The second possibility for adapting the Byron and Basmann tests to not having a $\Pi_{22}$ matrix of full rank involves using the same numerator for the test statistic as those tests employ, but calculating the denominator in such a way that it goes in probability to the appropriate normalization factor. Defining

$$\hat{\Omega} = \begin{pmatrix} 1 & 0 \\ 0 & \hat{H} \end{pmatrix} \hat{\Sigma} \begin{pmatrix} 1 & 0 \\ 0 & \hat{H} \end{pmatrix} = \begin{bmatrix} \hat{\omega}_{11} & \hat{\omega}_{12} & \hat{\omega}_{13} \\ \hat{\omega}_{12}' & \hat{\Omega}_{22} & \hat{\Omega}_{23} \\ \hat{\omega}_{13}' & \hat{\Omega}_{32} & \hat{\Omega}_{33} \end{bmatrix}$$ (18)

and

$$\hat{\sigma}^2 = (1 - \hat{\delta}_1) \left[ \begin{pmatrix} \hat{\omega}_{11} & \hat{\omega}_{12} \\ \hat{\omega}_{12}' & \hat{\Omega}_{22} \end{pmatrix} - \begin{pmatrix} \hat{\omega}_{13} \\ \hat{\Omega}_{23} \end{pmatrix} \hat{\Omega}_{33}^{-1} \begin{pmatrix} \hat{\omega}_{13}' \\ \hat{\Omega}_{32} \end{pmatrix} \right] \left( \begin{pmatrix} 1 \\ -\hat{\delta}_1 \end{pmatrix} \right),$$

where the partitioning of $\hat{\Omega}$ is the same as that of $\Omega$ used in (17), we can use the test statistic $\hat{\tau}_A = \nu / \hat{\sigma}^2$. Its relevant properties, as shown in the appendix, are summarized in the next proposition:

**Theorem 7:** Given the conditions of theorem 5,

(a) under $H_0 : \beta_2 = 0$,

$$\hat{\tau}_A \xrightarrow{d} \chi^2(K_2 - G);$$

(b) under $H_0 : \beta_2 \neq 0$,

$$\lim_{N \to \infty} P[\hat{\tau}_A < c] = 0$$

for any $c < \infty$; and

(c) under $H_0 : \beta_2 = \hat{b}/\sqrt{N}$,

$$\lim_{N \to \infty} P[\hat{\tau}_A \leq c] < \int_0^c d\chi^2(K_2 - G)$$

for all $0 < c < \infty$.

Thus, the major original properties of the Basmann and Byron tests are restored by using the adjustment to the denominator to allow for $\Pi_{22}$'s not being of full rank, including having the same degrees of freedom. Unfortunately, there is no clear comparison
of the two adjusted versions of the test in terms of asymptotic local power. \( \hat{\tau}_A \) has a different and much more complicated asymptotic distribution than \( \hat{\tau}_R \). Both tests are equivalent asymptotically to the Basmann and Byron tests when \( \Pi_{22} \) is of full rank since then \( \lim_{N \to \infty} P[\hat{G}_0 = G] = 1 \).

2.4. Testing with Asymptotically Inadequate Instruments

To see what happens when the instruments are weak, but the model is still identified though only barely, we consider a sequence of models in which the reduced-form submatrix \( \Pi_{22} \) of rank \( G \) converges to one of rank \( G_0 \). Steiger and Stock (1993) have investigated this question when \( G_0 = 0 \). In particular, we make the following assumption:

Assumption 3: Let \( \bar{\Pi} \) be a \( K \times (G + 1) \) reduced-form matrix partitioned into \( K_1 \) and \( K_2 \) rows and 1 and \( G \) columns with \( \rho[\bar{\Pi}_{22}] = G \) and \( \bar{\Pi}_{21} = \bar{\Pi}_{22} \gamma \). Let \( H \) be the eigenvectors of \( \bar{\Pi}_{22} \bar{C} \bar{\Pi}_{22} \) arranged so that the corresponding eigenvalues are in descending order. Let \( [\bar{\Phi}_1, \bar{\Phi}_2] = \bar{C}^{1/2} \bar{\Pi}_{22} [H_1 \ H_2] \), where the partitioning of \( H \) is into \( G_0 \) and \( (G - G_0) \) columns. Assume that the data sequences under consideration are generated by a sequence of reduced-form matrices indexed by sample size \( N \), \( \Pi(N) \), such that \( \Pi_{22}(N) = \bar{C}^{-1/2} [\bar{\Phi}_1 \ \bar{\Phi}_2 / \sqrt{N}] H \), \( \Pi_{21}(N) = \Pi_{22}(N) \gamma + \beta_2 \), \( \Pi_{11}(N) = \bar{\Pi}_{11} \) and \( \Pi_{12}(N) = \bar{\Pi}_{12} \).

This change in the nature of \( \Pi_{22} \) is not enough to restore the size or the consistency of the test using \( \hat{\lambda} \). Instead, the relevant proposition for the A-R test can be stated as the following theorem:

Theorem 8: Given assumptions 1 and 3.

(a) under \( H_0 : \beta_2 = 0 \),

\[
\lim_{N \to \infty} P[\hat{\lambda} \leq c] > \int_0^c d\chi^2(K_2 - G);
\]

and (b) under \( H_0 : \beta_2 \neq 0 \),

\[
\lim_{N \to \infty} P[\hat{\lambda} < c] > 0
\]

for any \( c > 0 \).

Under the local alternative (8), nothing very useful can be said. \( \hat{\lambda} \) then converges in probability to the smallest eigenvalue of a non-central Wishart Distribution, all of whose
non-centrality parameters are non-zero. What its relationship is to the \( \chi^2 \) distribution is unknown, though it depends undoubtedly on the magnitudes of the misspecification and of \( \Phi_2 \). In other words, all that having weak instruments in this sense, rather than completely inadequate ones, buys us asymptotically is that the asymptotic power of the A-R test is not necessarily less than its nominal size.

The situation for the other two tests is somewhat more complicated, and clear results do not emerge. Relative to the developments in lemmas 2 and 3, we can characterize the behavior of the parts of the test statistics by the following proposition:

Proposition 9: Under assumptions 1, and 2 and \( H_0: \beta_2 = 0 \)

\[
\frac{\nu}{\sigma^2} \overset{d}{\rightarrow} \int_0^{\infty} \chi^2((K_2 - C),q)f(q)dq
\]

where the non-centrality parameter \( q \) (displayed in the appendix) is a random variable with density \( f(q) \) derived from that of \( \Pi_{22} \), and \( \sigma^2 \) is the quantity defined in equation (17). Furthermore, (a) \( s_1^2 = s_2^2 + a_3 \) and (b) \( s_3^2 \overset{p}{\rightarrow} \sigma^2 + a_4 \) where \( a_3 \) and \( a_4 \) are nonnegative random variables, except that \( a_3 = 0 \) when \( \Pi_{22} = 0 \).

The problem with this result is that it neither relates the test statistic to the distribution that the statistic has when \( \rho[\Pi_{22}] = C \), nor does it allow us usefully to compare the extent to which the numerator's being too large relative to the \( \chi^2 \) distribution is offset by the denominators' also being too large. Furthermore, the denominators considered here cannot be compared in useful fashion with those obtained earlier, as shown by considering their values displayed in the appendix. Thus we still lack a useful characterization of the distributions of the test statistics when the instruments are weak in the sense considered in this section.

Having asymptotically inadequate instruments does little to make the asymptotic distribution of the estimates of \( \gamma \) resemble the result for identified parameters. An expression to which these estimates converge in probability, which contains a random element that remains \( O_p(1) \) rather than \( O_p(N^{-1/2}) \), is given in equation (A.14) in the appendix. It suggests that trying to obtain useful inferences about parameter magnitudes when the instruments are inadequate is extremely hazardous and makes one suspect that
OLS estimates, though possibly seriously misleading, may be preferable.

Since the problem of inference about the overidentifying restrictions when the instruments are asymptotically inadequate remains severe, it is worth considering what happens to the tests suggested in subsection 2.3. First, note that the test procedures and test statistic suggested in Cragg and Donald (1996), being based on the smallest root of a non-central Wishart matrix when examining $G_0$ as a possible estimate, will retain the property that $\hat{G}_0 \stackrel{p}{\rightarrow} G_0$, even though this is not the rank of $\Pi_{22}$ but instead is the rank of the matrix to which $\Pi_{22}$ converges in the limit.

The results for the modified tests are fairly straightforward, if not entirely happy:

**Proposition 10:** Given assumptions 1 and 3 and that $\hat{G}_0 \stackrel{p}{\rightarrow} G_0$,

(a) under $H_0 : \beta_2 = 0$,

\[
\hat{\lambda}_R \overset{d}{\rightarrow} \chi^2((K_2 - G_0), \bar{\lambda}_1),
\]

\[
\hat{\tau}_R \overset{d}{\rightarrow} \chi^2((K_2 - G_0), \bar{\lambda}_1),
\]

where $\bar{\lambda}_1$ is displayed in the appendix;

(b) under $H_0 : \beta_2 \neq 0$;

\[
\lim_{N \rightarrow \infty} P[\hat{\lambda}_R < c] = 0,
\]

\[
\lim_{N \rightarrow \infty} P[\hat{\tau}_R < c] = 0
\]

for any $c < \infty$; and

(c) under $H_a : \beta_2 = \overline{b}/\sqrt{N}$,

\[
\hat{\lambda}_R \overset{d}{\rightarrow} \chi^2((K_2 - G_0), \bar{\lambda}_2),
\]

\[
\hat{\tau}_R \overset{d}{\rightarrow} \chi^2((K_2 - G_0), \bar{\lambda}_2),
\]

where $\bar{\lambda}_2$ is also displayed in the appendix.

Unfortunately, it is not necessarily the case that $\bar{\lambda}_1 < \bar{\lambda}_2$ so that the test is not only of the wrong size asymptotically, but it may be biased for particular local alternatives. Similarly, while $\hat{\tau}_A$ does appropriately adjust the denominator, the numerator now is not appropriate, so that any gain to be made is problematic and the same problem arises under local alternatives.
3. A Sampling Experiment

The results of section 2 contain a number of ambiguities about to the best way to interpret them as approximations in finite samples. In particular, one may wonder whether — or when — it is appropriate to consider performance under asymptotically inadequate instruments as being more representative of finite-sample performance than the more usual asymptotic results where possible problems of the inadequacy of instruments disappear. Anyway, the results are asymptotic and their applicability to finite samples is problematic. We conduct a small Monte Carlo experiment to investigate their applicability.

The need for sampling-experiment investigation is particularly compelling since theorem 4 seems to give the opposite conclusion to the Monte Carlo experiment in Nelson and Startz (1990b) where it was found that there was substantial over rejection under the null when the instruments are weak. Of course the result of Nelson and Startz (1990b) is a finite-sample finding and ours is asymptotic. Furthermore, Maddala and Jeong (1992) have noted that in the simulations of Nelson and Startz (1990a) (which are based on the same type of model as in Nelson and Startz (1990b)) the normalizations implied a singular covariance matrix for the residuals. Moreover, in Nelson and Startz (1990b) the simulations concerning the overidentifiability tests were done with parameterizations that implied that the instruments were almost perfectly collinear. Thus the results of Nelson and Startz (1990b) could have more to do with the effect of these features of their design than with the lack of identification.

Our experiments employed a system with five endogenous variables and eight exogenous variables, seven of which (all except the constant) are eligible to serve as instrumental variables since they have zero $\beta$ coefficients under the null hypothesis being tested. There is no instrumental-variable model that is generally considered to be prototypical, so we set the parameters somewhat arbitrarily. The $[\Pi_{12} \quad \Pi_{22}]$ matrix was chosen to be zero except for the principal diagonal of $\Pi_{22}$ which took on the values$^6$ 0.5, 1.0, 2.0

---

$^6$These values may seem arbitrary but it must be remembered that the model is invariant to linear row transformations of the $\Pi$ matrix, balanced by corresponding transformations of the $X$ variables.
Table 1

Size of Tests (in %'s) using a nominal 5% level.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Test statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Def</td>
<td>(\lambda)</td>
</tr>
<tr>
<td>------</td>
<td>----------------</td>
</tr>
<tr>
<td></td>
<td>(N = 25)</td>
</tr>
<tr>
<td>0</td>
<td>2.1</td>
</tr>
<tr>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>2</td>
<td>0.0</td>
</tr>
<tr>
<td>3</td>
<td>0.0</td>
</tr>
<tr>
<td>4</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>(N = 50)</td>
</tr>
<tr>
<td>0</td>
<td>3.9</td>
</tr>
<tr>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>2</td>
<td>0.1</td>
</tr>
<tr>
<td>3</td>
<td>0.0</td>
</tr>
<tr>
<td>4</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>(N = 100)</td>
</tr>
<tr>
<td>0</td>
<td>5.3</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>0.1</td>
</tr>
<tr>
<td>3</td>
<td>0.0</td>
</tr>
<tr>
<td>4</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>(N = 400)</td>
</tr>
<tr>
<td>0</td>
<td>3.9</td>
</tr>
<tr>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td>2</td>
<td>0.0</td>
</tr>
<tr>
<td>3</td>
<td>0.0</td>
</tr>
<tr>
<td>4</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>(N = 1600)</td>
</tr>
<tr>
<td>0</td>
<td>4.8</td>
</tr>
<tr>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td>2</td>
<td>0.0</td>
</tr>
<tr>
<td>3</td>
<td>0.0</td>
</tr>
<tr>
<td>4</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>(N = 6400)</td>
</tr>
<tr>
<td>0</td>
<td>4.7</td>
</tr>
<tr>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td>2</td>
<td>0.0</td>
</tr>
<tr>
<td>3</td>
<td>0.0</td>
</tr>
<tr>
<td>4</td>
<td>0.0</td>
</tr>
</tbody>
</table>
and 3.0. The $X$ variables were selected from $N(0, 1)$, except for the first one which was a vector of units corresponding to the constant terms in the equations. Different values of $X$ were used in different replications of the experiments. Finally each $\gamma$ coefficient was set at unity while $\Sigma$ was chosen as

$$
\Sigma = \begin{bmatrix}
1.0 & 0.447 & 0.447 & 0.447 & 0.447 \\
0.447 & 1 & 0 & 0 & 0 \\
0.447 & 0 & 1 & 0 & 0 \\
0.447 & 0 & 0 & 1 & 0 \\
0.447 & 0 & 0 & 0 & 1
\end{bmatrix}.
$$

This means that the conditional variance of $v_{21}$, given $v_{21}$, is twenty per cent of the marginal variance. Each experiment consisted of 1825 replications, and experiments were conducted for sample sizes of 25, 50, 100, 400, 1600, and 6400 observations.

Inadequate instruments were produced by setting the non-zero elements of $\Pi$ equal to zero starting with the smallest. This produced rank deficiencies going from 0 to 4. In the latter case the whole $\Pi$ matrix is zero so that the instruments are totally irrelevant.

The sizes of the tests as estimated in the experiments are shown in table 1. Only values in the range 4.0% - 6.0% are not significantly different from 5.0% at the 0.05 level. Several things stand out in this table. First is the extent to which the Byron and Basman tests have too small size when inadequate instruments are used. This is particularly noticeable in the largest samples. This weakness, however, is even more pronounced with the A-R test. Second, in contrast to this pattern, the three tests which first estimate the rank of the matrix and then test the null hypothesis do have pretty much the right size for samples greater than 100. For these sample sizes, the correct rank was always estimated, which accounts for the results of the two-step tests being identical to the corresponding one-step tests when there was no rank deficiency. With small samples the sizes of these tests are far too large when the instruments are adequate and their apparently superior performance with inadequate instruments here arises from

and linear combinations of the columns other than the first balanced by corresponding changes to $\Sigma$. These parameter values were chosen to give clearly ascending roots to $\Pi^{22}C\Pi^{22}$ with the experiments investigating rank deficiency removing the roots in order of their importance and replacing them with zero. The arbitrary part of the specification is the magnitudes of the parameters, which were chosen to give values that might not be considered to be extreme and the joint diagonal structure of the covariance matrix of the $X$ variables, of $\Pi$ and of $\Sigma^{22}$. 
balancing the effects when the rank is estimated to be too large with those obtained when the correct rank is estimated. Third, there is a dramatically large difference between the Basmann and Byron tests with adequate instruments when a sample size of 25 is used. By contrast, the difference in their denominators, which persists asymptotically, was not large enough to affect the estimated sizes of the two tests when the sample size was 6400. In no case is there evidence of the Basmann test having too large a size.

Investigation of the power of the tests was done by having a non-zero value of \( \pi_{61} \), a variable which has zero coefficients in \( \Pi_{22} \). The coefficient took on the value \( \sqrt{25}/\sqrt{N} \). Estimates of the power of the tests when the nominal 0.05 level of significance\(^7\) was used are shown in table 2, which can be compared with table 1. On the encouraging side, the Basmann and the Byron tests have power greater than the nominal size even with rank deficiency. When the rank deficiency is complete, the power is not great (being roughly 15\% in large samples when \( \Pi_{22} = 0 \)) but this should also be compared with the very small actual sizes recorded in table 1. The adjusted tests all have power in the rank deficient cases roughly comparable to what was observed when there was no rank deficiency, though it must be recalled that with the smallest samples their sizes are too large. It is notable, especially with the larger samples where the sizes of the tests are not significantly different from their nominal levels, that the power of the \( \hat{\pi}_4 \) test is less than the powers of the others though the differences are not large. Though not reported here explicitly in tabular form, it may be noted that basically similar patterns emerged when the values of \( \pi_{61} \) were set at smaller values than used in table 2, though of course the power estimates then were all smaller.

Asymptotically inadequate instruments were produced by multiplying the coefficients in the basic \( \Pi_{22} \) matrix giving an identified system by \( \zeta \sqrt{25}/\sqrt{N} \). The resulting coefficients were substituted for the coefficients set equal to zero when we investigated rank deficiency in the previous experiments. Values of \( \zeta \) of 0.1, 0.25 and 1.0 were used. The

---

\(^7\)Estimation of power when the critical region of the test was adjusted to give a de facto size of 0.05 was not attempted because the sample size means that the estimate of the critical size would itself be subject to substantial error. In any case, it is of interest to see the performance of the tests using the nominal size (which would tend to be used in practice.)
Table 2

Local Power of Tests (in %'s) using a nominal 5% level.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Test statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Def</td>
<td>$\hat{\lambda}$</td>
</tr>
<tr>
<td></td>
<td>$\hat{\lambda}$</td>
</tr>
<tr>
<td></td>
<td>$\hat{\lambda}_R$</td>
</tr>
<tr>
<td></td>
<td>$\hat{\lambda}_A$</td>
</tr>
</tbody>
</table>

$N = 25$

<table>
<thead>
<tr>
<th>Def</th>
<th>$\hat{\lambda}$</th>
<th>$\hat{\theta}$</th>
<th>$\hat{\lambda}_R$</th>
<th>$\hat{\theta}_R$</th>
<th>$\hat{\theta}_A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>30.8</td>
<td>37.3</td>
<td>68.2</td>
<td>81.9</td>
<td>90.8</td>
</tr>
<tr>
<td>1</td>
<td>41</td>
<td>25.6</td>
<td>81.8</td>
<td>80.3</td>
<td>77.5</td>
</tr>
<tr>
<td>2</td>
<td>31.2</td>
<td>14.4</td>
<td>81.0</td>
<td>84.3</td>
<td>81.6</td>
</tr>
<tr>
<td>3</td>
<td>23.8</td>
<td>10.9</td>
<td>86.2</td>
<td>88.6</td>
<td>89.2</td>
</tr>
<tr>
<td>4</td>
<td>17.5</td>
<td>8.2</td>
<td>89.8</td>
<td>90.9</td>
<td>95.0</td>
</tr>
</tbody>
</table>

$N = 50$

<table>
<thead>
<tr>
<th>Def</th>
<th>$\hat{\lambda}$</th>
<th>$\hat{\theta}$</th>
<th>$\hat{\lambda}_R$</th>
<th>$\hat{\theta}_R$</th>
<th>$\hat{\theta}_A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>56.0</td>
<td>72.5</td>
<td>77.1</td>
<td>86.4</td>
<td>85.3</td>
</tr>
<tr>
<td>1</td>
<td>38.8</td>
<td>42.4</td>
<td>82.1</td>
<td>86.8</td>
<td>81.3</td>
</tr>
<tr>
<td>2</td>
<td>30.0</td>
<td>24.3</td>
<td>90.2</td>
<td>91.1</td>
<td>84.5</td>
</tr>
<tr>
<td>3</td>
<td>22.0</td>
<td>16.5</td>
<td>92.7</td>
<td>92.8</td>
<td>89.9</td>
</tr>
<tr>
<td>4</td>
<td>15.7</td>
<td>12.3</td>
<td>94.2</td>
<td>94.4</td>
<td>96.2</td>
</tr>
</tbody>
</table>

$N = 100$

<table>
<thead>
<tr>
<th>Def</th>
<th>$\hat{\lambda}$</th>
<th>$\hat{\theta}$</th>
<th>$\hat{\lambda}_R$</th>
<th>$\hat{\theta}_R$</th>
<th>$\hat{\theta}_A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>79.8</td>
<td>84.3</td>
<td>81.5</td>
<td>88.2</td>
<td>88.1</td>
</tr>
<tr>
<td>1</td>
<td>42.2</td>
<td>50.9</td>
<td>90.0</td>
<td>92.0</td>
<td>84.4</td>
</tr>
<tr>
<td>2</td>
<td>31.2</td>
<td>28.6</td>
<td>93.5</td>
<td>93.8</td>
<td>86.3</td>
</tr>
<tr>
<td>3</td>
<td>22.2</td>
<td>20.1</td>
<td>94.4</td>
<td>94.5</td>
<td>89.8</td>
</tr>
<tr>
<td>4</td>
<td>18.4</td>
<td>15.3</td>
<td>95.4</td>
<td>95.4</td>
<td>95.9</td>
</tr>
</tbody>
</table>

$N = 400$

<table>
<thead>
<tr>
<th>Def</th>
<th>$\hat{\lambda}$</th>
<th>$\hat{\theta}$</th>
<th>$\hat{\lambda}_R$</th>
<th>$\hat{\theta}_R$</th>
<th>$\hat{\theta}_A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>92.9</td>
<td>93.4</td>
<td>92.9</td>
<td>93.4</td>
<td>93.4</td>
</tr>
<tr>
<td>1</td>
<td>40</td>
<td>51.8</td>
<td>93.7</td>
<td>94.0</td>
<td>88.3</td>
</tr>
<tr>
<td>2</td>
<td>31.7</td>
<td>31.1</td>
<td>96.0</td>
<td>96.1</td>
<td>88.7</td>
</tr>
<tr>
<td>3</td>
<td>22.6</td>
<td>22.3</td>
<td>96.6</td>
<td>96.6</td>
<td>90.0</td>
</tr>
<tr>
<td>4</td>
<td>16.7</td>
<td>16.0</td>
<td>97.5</td>
<td>97.5</td>
<td>96.2</td>
</tr>
</tbody>
</table>

$N = 1600$

<table>
<thead>
<tr>
<th>Def</th>
<th>$\hat{\lambda}$</th>
<th>$\hat{\theta}$</th>
<th>$\hat{\lambda}_R$</th>
<th>$\hat{\theta}_R$</th>
<th>$\hat{\theta}_A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>95.5</td>
<td>95.5</td>
<td>96.5</td>
<td>95.5</td>
<td>95.6</td>
</tr>
<tr>
<td>1</td>
<td>3.6</td>
<td>53.3</td>
<td>94.6</td>
<td>94.8</td>
<td>88.5</td>
</tr>
<tr>
<td>2</td>
<td>31.9</td>
<td>31.8</td>
<td>96.5</td>
<td>96.5</td>
<td>86.8</td>
</tr>
<tr>
<td>3</td>
<td>21.3</td>
<td>21.2</td>
<td>96.1</td>
<td>96.1</td>
<td>89.6</td>
</tr>
<tr>
<td>4</td>
<td>17.4</td>
<td>17.3</td>
<td>97.5</td>
<td>97.5</td>
<td>97.0</td>
</tr>
</tbody>
</table>

$N = 6400$

<table>
<thead>
<tr>
<th>Def</th>
<th>$\hat{\lambda}$</th>
<th>$\hat{\theta}$</th>
<th>$\hat{\lambda}_R$</th>
<th>$\hat{\theta}_R$</th>
<th>$\hat{\theta}_A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>95.3</td>
<td>95.3</td>
<td>95.3</td>
<td>95.3</td>
<td>95.3</td>
</tr>
<tr>
<td>1</td>
<td>3.9</td>
<td>53.3</td>
<td>95.5</td>
<td>95.7</td>
<td>88.3</td>
</tr>
<tr>
<td>2</td>
<td>32.4</td>
<td>32.3</td>
<td>96.1</td>
<td>96.1</td>
<td>86.5</td>
</tr>
<tr>
<td>3</td>
<td>21.7</td>
<td>21.7</td>
<td>96.2</td>
<td>96.2</td>
<td>89.8</td>
</tr>
<tr>
<td>4</td>
<td>15.6</td>
<td>15.5</td>
<td>97.3</td>
<td>97.3</td>
<td>95.8</td>
</tr>
</tbody>
</table>
sizes of the tests when $\zeta$ was 0.25 are shown in table 3. Again the entries can be compared with those of table 1, in which the value of $\zeta$ is zero.

Three things stand out in table 3. First, instruments as inadequate as these ones do little to improve the size of the A-R test, which is again far smaller than the nominal 0.05 level. Second, with the larger values for the rank deficiencies, particularly with the larger sample sizes, the sizes of the Basman and Byron tests are quite close to the nominal size. This is less evident with the smaller rank deficiencies (where the adequacy of the inadequate instruments is less because the underlying parameters are smaller.) It is, of course, partly fortuitous. As the theoretical development indicated, asymptotically too large a numerator is being balanced by too large a denominator. Indeed, when $\zeta = 1.0$ the corresponding entries for rank deficiencies of 3 and 4 are significantly greater than 5% at the 0.05 level. Third, the sizes of the adjusted tests now are too large, again as predicted by the asymptotic theory. This tendency is evident throughout the table, but is most serious with the larger sample sizes when comparison is made with the sizes of the tests when the value of $\zeta = 1$. This tendency is less marked in the case of the $\tilde{r}_A$ test in which only the denominator is adjusted.

The extent to which the distribution of the estimates of $\gamma$ are affected by having asymptotically inadequate instruments is summarized in table 4. The experiment design means that the rank deficiency largely affects which parameters are not identified, since (asymptotically) the first $k \gamma$ coefficients are not identified when the rank-deficiency is $k$ while the others are. (With other designs all coefficients would usually be combinations of the identified and unidentified parameters.) In consequence we shall examine the case of full rank deficiency ($\Pi_{22} = 0$) and investigate the effect of having asymptotically inadequate instruments. In table 4 results are reported for the various values of $\zeta$ used, with inadequacy level 0 indicating no adjustment to the parameters – that is the identified standard model is used, while levels 1-4 correspond to values of $\zeta$ of 1.0, 0.25, 0.1 and 0.0; i.e. level 4 represents completely inadequate instruments.

The table reports the medians and the interquartile ranges of the estimates. As expected, because of the structure of $\Pi_{22}$, $\gamma_1$ tends to be the least precisely estimated
Table 3

Asymptotically Inadequate Instruments
Size of tests (in %) using nominal 5% level.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Test statistics</th>
<th>Def</th>
<th>( \lambda )</th>
<th>( \hat{\lambda} )</th>
<th>( \theta )</th>
<th>( \lambda_R )</th>
<th>( \tau_R )</th>
<th>( \tau_A )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>( N = 25 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td>2.1</td>
<td>9.5</td>
<td>1.6</td>
<td>21.3</td>
<td>26.5</td>
<td>16.6</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>0.9</td>
<td>6.5</td>
<td>0.9</td>
<td>17.2</td>
<td>20.7</td>
<td>14.7</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>0.1</td>
<td>5.2</td>
<td>0.9</td>
<td>14.8</td>
<td>16.5</td>
<td>15.6</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>0.0</td>
<td>5.3</td>
<td>0.8</td>
<td>34.5</td>
<td>36.2</td>
<td>24.8</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>0.0</td>
<td>6.2</td>
<td>1.2</td>
<td>53.9</td>
<td>58.0</td>
<td>44.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( N = 50 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td>3.9</td>
<td>8.2</td>
<td>4.7</td>
<td>14.6</td>
<td>17.9</td>
<td>10.5</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>0.4</td>
<td>3.7</td>
<td>1.8</td>
<td>6.4</td>
<td>7.8</td>
<td>7.3</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>0.1</td>
<td>3.2</td>
<td>1.4</td>
<td>13.0</td>
<td>13.8</td>
<td>10.1</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>0.0</td>
<td>4.0</td>
<td>2.3</td>
<td>42.5</td>
<td>43.3</td>
<td>18.6</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>0.0</td>
<td>5.0</td>
<td>3.3</td>
<td>74.0</td>
<td>76.4</td>
<td>46.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( N = 100 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td>5.3</td>
<td>6.2</td>
<td>5.2</td>
<td>6.7</td>
<td>7.7</td>
<td>6.5</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>0.8</td>
<td>4.3</td>
<td>3.5</td>
<td>6.5</td>
<td>7.2</td>
<td>7.4</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>0.1</td>
<td>3.7</td>
<td>2.6</td>
<td>13.9</td>
<td>14.2</td>
<td>10.0</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>0.1</td>
<td>3.9</td>
<td>3.0</td>
<td>43.5</td>
<td>43.7</td>
<td>18.4</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>0.0</td>
<td>5.6</td>
<td>4.5</td>
<td>79.5</td>
<td>81.2</td>
<td>48.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( N = 400 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td>3.9</td>
<td>4.3</td>
<td>4.9</td>
<td>3.9</td>
<td>4.3</td>
<td>4.3</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>0.4</td>
<td>3.5</td>
<td>3.2</td>
<td>6.1</td>
<td>6.4</td>
<td>6.4</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>0.0</td>
<td>2.0</td>
<td>1.9</td>
<td>12.3</td>
<td>12.5</td>
<td>7.3</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>0.1</td>
<td>3.3</td>
<td>3.1</td>
<td>49.1</td>
<td>49.2</td>
<td>17.1</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>0.0</td>
<td>5.3</td>
<td>5.2</td>
<td>86.4</td>
<td>87.6</td>
<td>48.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( N = 1600 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td>4.8</td>
<td>5.1</td>
<td>4.8</td>
<td>4.8</td>
<td>5.1</td>
<td>5.1</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>0.3</td>
<td>3.3</td>
<td>3.2</td>
<td>6.5</td>
<td>6.6</td>
<td>5.3</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>0.0</td>
<td>2.2</td>
<td>2.1</td>
<td>11.8</td>
<td>11.8</td>
<td>6.4</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>0.0</td>
<td>3.5</td>
<td>3.4</td>
<td>47.4</td>
<td>47.4</td>
<td>14.8</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>0.0</td>
<td>5.2</td>
<td>5.2</td>
<td>91.9</td>
<td>92.4</td>
<td>51.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( N = 6400 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td>4.7</td>
<td>4.7</td>
<td>4.7</td>
<td>4.7</td>
<td>4.7</td>
<td>4.7</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>0.1</td>
<td>3.0</td>
<td>3.0</td>
<td>6.5</td>
<td>6.5</td>
<td>5.4</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>0.0</td>
<td>2.5</td>
<td>2.5</td>
<td>13.4</td>
<td>13.4</td>
<td>6.6</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>0.0</td>
<td>3.7</td>
<td>3.7</td>
<td>48.9</td>
<td>48.9</td>
<td>17.2</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>0.0</td>
<td>5.0</td>
<td>5.0</td>
<td>93.9</td>
<td>94.1</td>
<td>50.7</td>
</tr>
</tbody>
</table>
while $\gamma_4$ is the most precise. Similarly, precision declines with the extent of inadequacy of the instruments. The most notable feature of using inadequate instruments is that indeed the dispersion of the estimates does not decline noticeably as the sample size increases except possibly initially. Second, the central tendency of the estimates largely depends on the size of the relevant elements of $\Pi_{22}$. For example, with inadequacy level 1 and the largest sample size, the median estimate of $\gamma_4$ is very close to the population value of 1.0. With inadequacy level 3, it is much closer to the value when $\Pi_{22} = 0$ — in which case the central tendency of the distribution is close to its population asymptotic central value of .447. By contrast, the medians of $\gamma_1$ for inadequacy levels of 1 and 3 are .81 and .44, in each case closer to the value expected with completely inadequate instruments. Third, for all sample sizes the distributions of the estimates with inadequate instruments appear to be roughly the same, and so the asymptotic theory appears to provide a reasonably adequate approximation to the distribution. By contrast, using the usual asymptotic approximation misses the important weakness of the central tendency. This is shown by the contrast in the medians for $\gamma_1$ with very small samples and with 6400 observations where with inadequate instruments the central tendencies are much closer to those of the distributions with $N = 25$ than they are to the asymptotic values of 1.0 which they would have if the coefficients producing inadequacy were kept constant at $\zeta = .25/80$ times their value in the base specification as the sample sized increases without bound.

4. Discussion

What should one do in practice? The asymptotic results imply that the Basmann-Byron style tests are still likely to be useful in practice, even if the model is unidentified. It would also seem to be the case that the adjusted tests are the most useful if inadequacy is a severe problem. The fly in this ointment is the behavior of the tests with asymptotically inadequate instruments where the predictions of the asymptotic theory were found not to be unrepresentative of finite sample behaviour with seriously inadequate instruments. Here, the adjusted tests tend to have too large a size, though retaining their power. Of
Table 4

Asymptotically Inadequate Instruments
Estimates of $\gamma$

<table>
<thead>
<tr>
<th>Inadequacy</th>
<th>$\gamma_1$ Median</th>
<th>I.Q.R.</th>
<th>$\gamma_2$ Median</th>
<th>I.Q.R.</th>
<th>$\gamma_3$ Median</th>
<th>I.Q.R.</th>
<th>$\gamma_4$ Median</th>
<th>I.Q.R.</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = 25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>.779</td>
<td>.305</td>
<td>.918</td>
<td>.308</td>
<td>.973</td>
<td>.158</td>
<td>.990</td>
<td>.084</td>
</tr>
<tr>
<td>1</td>
<td>.779</td>
<td>.505</td>
<td>.917</td>
<td>.308</td>
<td>.975</td>
<td>.158</td>
<td>.990</td>
<td>.084</td>
</tr>
<tr>
<td>2</td>
<td>.466</td>
<td>.613</td>
<td>.576</td>
<td>.549</td>
<td>.737</td>
<td>.419</td>
<td>.889</td>
<td>.247</td>
</tr>
<tr>
<td>3</td>
<td>.435</td>
<td>.455</td>
<td>.472</td>
<td>.468</td>
<td>.531</td>
<td>.438</td>
<td>.662</td>
<td>.346</td>
</tr>
<tr>
<td>4</td>
<td>.437</td>
<td>.336</td>
<td>.450</td>
<td>.340</td>
<td>.460</td>
<td>.331</td>
<td>.447</td>
<td>.335</td>
</tr>
<tr>
<td>N = 50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>.555</td>
<td>.408</td>
<td>.959</td>
<td>.221</td>
<td>.988</td>
<td>.112</td>
<td>.996</td>
<td>.059</td>
</tr>
<tr>
<td>1</td>
<td>.763</td>
<td>.514</td>
<td>.927</td>
<td>.294</td>
<td>.994</td>
<td>.155</td>
<td>.994</td>
<td>.082</td>
</tr>
<tr>
<td>2</td>
<td>.467</td>
<td>.605</td>
<td>.570</td>
<td>.533</td>
<td>.899</td>
<td>.404</td>
<td>.599</td>
<td>.254</td>
</tr>
<tr>
<td>3</td>
<td>.442</td>
<td>.448</td>
<td>.471</td>
<td>.427</td>
<td>.662</td>
<td>.424</td>
<td>.662</td>
<td>.346</td>
</tr>
<tr>
<td>4</td>
<td>.433</td>
<td>.307</td>
<td>.448</td>
<td>.326</td>
<td>.438</td>
<td>.324</td>
<td>.438</td>
<td>.343</td>
</tr>
<tr>
<td>N = 100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>.949</td>
<td>.288</td>
<td>.975</td>
<td>.166</td>
<td>.997</td>
<td>.082</td>
<td>.998</td>
<td>.039</td>
</tr>
<tr>
<td>1</td>
<td>.814</td>
<td>.298</td>
<td>.925</td>
<td>.298</td>
<td>.982</td>
<td>.156</td>
<td>.994</td>
<td>.078</td>
</tr>
<tr>
<td>2</td>
<td>.483</td>
<td>.572</td>
<td>.574</td>
<td>.572</td>
<td>.738</td>
<td>.412</td>
<td>.897</td>
<td>.245</td>
</tr>
<tr>
<td>3</td>
<td>.437</td>
<td>.461</td>
<td>.459</td>
<td>.461</td>
<td>.515</td>
<td>.434</td>
<td>.658</td>
<td>.337</td>
</tr>
<tr>
<td>N = 400</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>.983</td>
<td>.159</td>
<td>.996</td>
<td>.079</td>
<td>.999</td>
<td>.041</td>
<td>.999</td>
<td>.020</td>
</tr>
<tr>
<td>1</td>
<td>.810</td>
<td>.485</td>
<td>.930</td>
<td>.287</td>
<td>.981</td>
<td>.154</td>
<td>.993</td>
<td>.072</td>
</tr>
<tr>
<td>2</td>
<td>.494</td>
<td>.555</td>
<td>.598</td>
<td>.558</td>
<td>.750</td>
<td>.413</td>
<td>.899</td>
<td>.296</td>
</tr>
<tr>
<td>3</td>
<td>.455</td>
<td>.455</td>
<td>.480</td>
<td>.455</td>
<td>.533</td>
<td>.418</td>
<td>.669</td>
<td>.337</td>
</tr>
<tr>
<td>4</td>
<td>.463</td>
<td>.326</td>
<td>.447</td>
<td>.326</td>
<td>.440</td>
<td>.341</td>
<td>.455</td>
<td>.325</td>
</tr>
<tr>
<td>N = 1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>.999</td>
<td>.082</td>
<td>.999</td>
<td>.039</td>
<td>.999</td>
<td>.020</td>
<td>1.000</td>
<td>.100</td>
</tr>
<tr>
<td>1</td>
<td>.789</td>
<td>.507</td>
<td>.917</td>
<td>.288</td>
<td>.983</td>
<td>.152</td>
<td>.994</td>
<td>.081</td>
</tr>
<tr>
<td>2</td>
<td>.473</td>
<td>.604</td>
<td>.561</td>
<td>.553</td>
<td>.749</td>
<td>.383</td>
<td>.894</td>
<td>.230</td>
</tr>
<tr>
<td>3</td>
<td>.431</td>
<td>.457</td>
<td>.454</td>
<td>.446</td>
<td>.517</td>
<td>.410</td>
<td>.654</td>
<td>.322</td>
</tr>
<tr>
<td>4</td>
<td>.433</td>
<td>.341</td>
<td>.437</td>
<td>.319</td>
<td>.443</td>
<td>.309</td>
<td>.444</td>
<td>.316</td>
</tr>
<tr>
<td>N = 6400</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>.999</td>
<td>.039</td>
<td>1.000</td>
<td>.020</td>
<td>1.000</td>
<td>.011</td>
<td>1.000</td>
<td>.065</td>
</tr>
<tr>
<td>1</td>
<td>.306</td>
<td>.492</td>
<td>.926</td>
<td>.285</td>
<td>.978</td>
<td>.155</td>
<td>.997</td>
<td>.078</td>
</tr>
<tr>
<td>2</td>
<td>.458</td>
<td>.647</td>
<td>.578</td>
<td>.521</td>
<td>.741</td>
<td>.390</td>
<td>.913</td>
<td>.250</td>
</tr>
<tr>
<td>3</td>
<td>.438</td>
<td>.466</td>
<td>.456</td>
<td>.449</td>
<td>.515</td>
<td>.418</td>
<td>.675</td>
<td>.344</td>
</tr>
<tr>
<td>4</td>
<td>.441</td>
<td>.345</td>
<td>.449</td>
<td>.350</td>
<td>.449</td>
<td>.318</td>
<td>.454</td>
<td>.339</td>
</tr>
</tbody>
</table>
these, the test that only adjusts the denominator of the Byron and Basmann statistics rather than also adjusting the numerator of the statistic may be a somewhat safer bet.

It is possible also that a less stringent procedure for estimating the rank of the matrix than the one used here might give happier results—a possibility we have not explored. There is a range between the significance levels at which the hypothesis of a particular rank-inadequacy would be rejected in usual practice and the levels that it might be best to use in estimating the rank of the matrix before adjusting the test procedures.

It still appears to be prudent to check the relevance of the instruments before proceeding to make inferences about parameters themselves. One may legitimately be highly suspicious of results where the instruments appear to have little to do with the variables for which they are instruments or where the instruments are not clearly relevant in the reduced form equation for the main dependent variable. Unfortunately, when there is collinearity among instruments, it may well be the case that the set of instruments as a whole is highly relevant without this indicating that there are not serious problems of rank deficiency or instrument inadequacy. Thus to the advice of Staiger and Stock (1993) to check the relevance of instruments needs to be added the need to check the identifiability of the coefficients, since these could fail even with instruments that are highly significant both jointly and individually if, for example, they were proxies for a smaller number of relevant variables.

Our results indicate that of the usual tests possibly the Byron test is the least likely to be inadequate among the unadjusted tests and simply adjusting its denominator may be the safest procedure among those we considered. That seriously inadequate instruments can severely compromise instrumental-variable estimation is unfortunately an implication both of our theoretical results and our small sampling experiment.
Appendix of Proofs:

Proof of Lemma 2:

Part (2): Let

\[ \begin{align*}
\phi &= \bar{C}^{1/2} \Pi_{11}; \\
\Phi &= [\Phi_1, \Phi_2] = \bar{C}^{1/2} \Pi_{22} H; \\
\delta_1 &= H_1 \gamma;
\end{align*} \]

where the partitioning of \( \Phi \) is into \( G_0 \) and \( (G - G_0) \) columns. Note that \( \Phi_2 = 0 \) while, under \( H_0, \phi = \bar{C}^{1/2} \Pi_{22} \gamma = \Phi_1 \delta_1 \). Similarly define

\[ \begin{align*}
\hat{\phi} &= \bar{C}^{1/2} \hat{\Pi}_{11}; \\
\hat{\Phi} &= \bar{C}^{1/2} \hat{\Pi}_{22} H; \\
\Psi &= \begin{bmatrix} \psi & \psi_1 & \psi_2 \end{bmatrix} = \begin{bmatrix} (\hat{\phi} - \phi) & (\Phi_1 - \Phi_1) & (\Phi_2 - \Phi_2) \end{bmatrix}.
\end{align*} \]

(At present, \( \Psi_2 = \hat{\Phi}_2 \), but the wider definition will be useful later.) By construction and assumption 1,

\[ \sqrt{N} \Psi \overset{d}{\longrightarrow} \tilde{\Psi} = [\tilde{\psi}_1 \quad \tilde{\psi}_1 \quad \tilde{\psi}_2], \text{ say, where vec } \tilde{\Psi} \sim N(0, \Omega \otimes I_k). \]

We can write

\[ \begin{align*}
\hat{\phi} &= \hat{\Phi}_1 \delta_1 - \Psi_1 \delta_1 + \psi \\
&= \hat{\Phi}_1 \delta_1 - (\Psi_1 - \Psi_2 \Xi_2) \delta_1 - \Psi_2 \Xi_2 \delta_1 + \psi - \Psi_2 \Xi_1 + \Psi_2 \Xi_1 \\
&= \hat{\Phi}_1 \delta_1 + \Psi_2 \alpha_2 + \eta
\end{align*} \]  \hspace{1cm} (A.1)

where

\[ \Xi = \begin{bmatrix} \Xi_1 & \Xi_2 \end{bmatrix} = \Omega_{33}^{-1} (\omega_{31} \quad \Omega_{31}), \]

\[ \alpha_2 = \Xi_1 - \Xi_2 \delta_1, \]  \hspace{1cm} (A.2)

and

\[ \eta = \psi - \Psi_2 \Xi_1 - (\Psi_1 - \Psi_2 \Xi_2) \delta_1. \]  \hspace{1cm} (A.3)
By construction and assumption 1,
\[ \sqrt{N} \tilde{\eta} \xrightarrow{d} \tilde{\eta}, \text{ say, where } \tilde{\eta} \sim N(0, \sigma^2 I_{K_2}) \]  
(A.4) and \( \sigma^2 \) is defined in equation (17). Also by construction, \( \tilde{\eta} \) is independent of \( \tilde{\Psi}_2 \).

The numerator of the Byron and Basman test statistic can be expressed as
\[ \nu = N \tilde{\phi}' [I - \tilde{\phi}' \tilde{\phi}^{-1}] \tilde{\phi} \]
\[ = \sqrt{N} \tilde{\eta}' I_{K_2} - \]
\[ \begin{pmatrix} \tilde{\phi}' & \sqrt{N} \tilde{\phi}' \tilde{\Psi}_2 \end{pmatrix} \begin{pmatrix} \sqrt{N} \tilde{\phi}' \tilde{\phi}_1 \tilde{\Psi}_1 & \sqrt{N} \tilde{\phi}' \tilde{\phi}_2 \tilde{\Psi}_2 \end{pmatrix}^{-1} \begin{pmatrix} \tilde{\phi}' \tilde{\phi}_1 \tilde{\Psi}_1 \tilde{\Psi}_2 \tilde{\phi}' \tilde{\phi}_2 \end{pmatrix} \sqrt{N} \tilde{\eta} \]  
\[ \xrightarrow{d} \tilde{\eta}' [I_{K_2} - (\Phi_1 \tilde{\Psi}_2) [\Phi_1' \Phi_1 \tilde{\Psi}_2^{-1}]^{-1} (\Phi_1' \tilde{\Psi}_2) ] \tilde{\eta}. \]
Therefore expression (A.4) and the fact that (A.5) is an idempotent quadratic form together imply that, conditional on \( \tilde{\Psi}_2 \),
\[ \nu/\sigma^2 \xrightarrow{d} \chi^2(K_2 - G). \]  
(A.6)
Since this distribution does not involve \( \tilde{\Psi}_2 \), on which (A.6) was conditioned, (A.6) also gives the marginal asymptotic distribution of \( \nu/\sigma^2 \).

Part (b) Let
\[ \phi^* = \tilde{C}^{1/2} \beta_2 - \Phi_1 (\Phi_1' \Phi_1)^{-1} \Phi_1' \tilde{C}^{1/2} \beta_2 \]
\[ \delta^*_1 = (\Phi_1' \Phi_1)^{-1} \Phi_1' \phi = \delta_1 + (\Phi_1' \Phi_1)^{-1} \Phi_1' \tilde{C}^{1/2} \beta_2. \]
Then from equation (13),
\[ \nu/N = (\psi + \tilde{C}^{1/2} \beta_2 + \Phi_1 \delta_1)' [I - \tilde{\phi}' \tilde{\phi}^{-1}] \tilde{\phi}' (\psi + \tilde{C}^{1/2} \beta_2 + \Phi_1 \delta_1) \]
\[ \xrightarrow{d} \phi^*'[I_{K_2} - (\Phi_1 \tilde{\Psi}_2) [\Phi_1' \Phi_1 \tilde{\Psi}_2^{-1}]^{-1} (\Phi_1' \tilde{\Psi}_2)] \phi^* \]
which is a non-degenerate non-negative random variable so that asymptotically \( \nu/\sigma^2 \) is unbounded in probability.

Part (c): Equation (A.1) now becomes
\[ \tilde{\phi} = \tilde{\phi}_1 \delta_1 + \tilde{\Psi}_2 \alpha_2 + \eta + C^{1/2} \tilde{b}/\sqrt{N}. \]
If we define $\eta^* = \eta + C^{1/2}b/\sqrt{N}$, then

$$\sqrt{N}\eta^* \xrightarrow{d} \bar{\eta}^*, \text{ say, where } \bar{\eta}^* \sim N(C^{1/2}b, \sigma^2 I_{K_2}).$$

while (A.5) gives the limiting quadratic form when $\bar{\eta}^*$ is substituted for $\bar{\eta}$ in it. Therefore, conditional on $\bar{\Psi}$,

$$\nu/\sigma^2 \xrightarrow{d} \chi^2((K_2 - G), \bar{\sigma}/\sigma^2)$$

where

$$\tilde{\nu}^- = (C^{1/2}b)' \begin{bmatrix} I_{K_2} - (\Phi_1 \bar{\Psi}_2) \frac{\Phi_1'\Phi_1}{\bar{\Psi}_2'\Phi_1} \Phi_1'\bar{\Psi}_2 \end{bmatrix}^{-1} \begin{bmatrix} \Phi_1' \frac{\Phi_1'\Phi_1}{\bar{\Psi}_2'\Phi_1} \Phi_1'\bar{\Psi}_2 \end{bmatrix} C^{1/2}b.$$ 

Since $\lim_{N \to \infty} P\{\nu/\sigma^2 \leq c\} | \bar{\Psi}_2 < \int_0^c d\chi^2((K_2 - G), \bar{\sigma}/\sigma^2)$ for every value of $\bar{\Psi}_2$, the inequality holds in the marginal distribution. □

**The distribution of $\hat{\eta}$**

Let

$$D = \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix} = \begin{bmatrix} \Phi_1'\Phi_1 & \Phi_1'\bar{\Psi}_2 \\ \bar{\Psi}_2'\Phi_1 & \bar{\Psi}_2'\bar{\Psi}_2 \end{bmatrix}^{-1}.$$ 

Note that the (transformed) instrumental variable (2SLS) estimator of $\delta$,

$$\hat{\delta} = (\hat{\phi}'\hat{\phi})^{-1}\hat{\phi}'\hat{\psi}$$

$$= \begin{bmatrix} I_{G_0} & 0 \\ 0 & \sqrt{N}I_{(\sigma - G_0)} \end{bmatrix} \begin{bmatrix} \frac{\hat{\phi}'\hat{\psi}_1}{\sqrt{N}\hat{\psi}_2'\hat{\phi}_1} & \frac{\sqrt{N}\hat{\psi}_2'\hat{\phi}_1}{\sqrt{N}\hat{\psi}_2'\hat{\phi}_1} \end{bmatrix}^{-1} \times$$

$$\begin{bmatrix} I_{G_0} & 0 \\ 0 & \sqrt{N}I_{(\sigma - G_0)} \end{bmatrix} \begin{bmatrix} \hat{\phi}_1' \hat{\psi}_2' \\ \hat{\psi}_2' \hat{\psi}_2 \end{bmatrix}$$

$$= \begin{bmatrix} \hat{\delta}_1 \\ \alpha_2 \end{bmatrix} +$$

$$\begin{bmatrix} I_{G_0} & 0 \\ 0 & \sqrt{N}I_{(\sigma - G_0)} \end{bmatrix} \begin{bmatrix} \frac{\hat{\phi}'\hat{\psi}_1}{\sqrt{N}\hat{\psi}_2'\hat{\phi}_1} & \frac{\sqrt{N}\hat{\psi}_2'\hat{\phi}_1}{\sqrt{N}\hat{\psi}_2'\hat{\phi}_1} \end{bmatrix}^{-1} \begin{bmatrix} \hat{\phi}_1' \eta \\ \hat{\psi}_2' \eta \end{bmatrix},$$

where $\alpha_2$ is defined in equations (A.2), so that

$$\begin{bmatrix} \hat{\delta}_1 \\ \hat{\delta}_2 \end{bmatrix} - \begin{bmatrix} \delta_1 \\ \alpha_2 + \sqrt{N}D_{21}\phi_1'\eta + ND_{22}\hat{\psi}_2'\eta \end{bmatrix} \xrightarrow{p} 0.$$
while
\[
\begin{pmatrix}
\sqrt{N}(\hat{\delta}_1 - \delta_1)
\
(\hat{\delta}_2 - \alpha_2)
\end{pmatrix}
\xrightarrow{d}
\begin{pmatrix}
D_{11}\Phi_1'\bar{\eta} + D_{12}\Phi_2'\bar{\eta}
\
D_{21}\Phi_1'\bar{\eta} + D_{22}\Phi_2'\bar{\eta}
\end{pmatrix}.
\] (A.7)

From (A.7), we see that this vector has a straightforward (asymptotic) normal distribution conditional on \(\bar{\Psi}_2\) and the messier marginal distribution of regression coefficients when the regressors contain normally distributed and mathematical variables. It is therefore also the case that
\[
\hat{\gamma} = \hat{H}\hat{\delta}
\] (A.8)
\[
\xrightarrow{d} H_1H_1'\gamma + H_2\alpha_2 + H_2(D_{21}\Phi_1' + D_{22}\Phi_2')\tilde{\eta},
\]
where the last, random term is \(P_2(1)\) and, even conditional on \(\bar{\Psi}_2\), \(\hat{\gamma}\) does not have mean \(\gamma\).

Suppose that \(\rho(\Pi_{22}) = 0\), so that \(\Pi_{22} = 0\) and \(\sqrt{N}\Sigma^{1/2}\tilde{\Pi}_{22} \xrightarrow{d} \bar{\Psi}_2\) (while \(H_1\) is a null term and \(H_2 = H = I\) by conventional normalization). Then then \(\hat{\gamma} \xrightarrow{d} \Sigma_{22}^{-1}\sigma_{21} + (\bar{\Psi}_2'\bar{\Psi}_2)^{-1}\bar{\Psi}_2'\tilde{\eta}\), where the partitioning of \(\Sigma\) is into 1 and G rows and columns, so that now \(\hat{\gamma}\) does not depend on \(\gamma\) at all.

**Proof of lemma 3:**

**Part (a):** Note that
\[
\hat{u} = Y_1 - Y_2\hat{\gamma} - X_1\hat{\delta}_1
\]
\[
= [(Y_1 - X_1\hat{\Pi}_{11}) (Y_2 - X_1\hat{\Pi}_{12})]\begin{pmatrix}
1
-
\hat{\gamma}
\end{pmatrix}
\]
\[
= [(Y_1 - X_1\hat{\Pi}_{11} - X_2\hat{\Pi}_{21}) (Y_2 - X_1\hat{\Pi}_{12} - X_2\hat{\Pi}_{22})]\begin{pmatrix}
1
-
\hat{\gamma}
\end{pmatrix}
+ X_2(\hat{\Pi}_{21} - \hat{\Pi}_{22}\hat{\gamma}).
\]

Because \(\hat{\Pi}\) are regression coefficients, so that
\[
[(Y_1 - X_1\hat{\Pi}_{11} - X_2\hat{\Pi}_{21}) (Y_2 - X_1\hat{\Pi}_{12} - X_2\hat{\Pi}_{22})]'X_2 = 0,
\]
it follows that
\[
s_1^2 = (1, -\hat{\gamma}')\Sigma(1, -\hat{\gamma})' + (\bar{\Pi}_{21} - \hat{\Pi}_{22}\hat{\gamma})(X_2'X_2/N)(\hat{\Pi}_{21} - \hat{\Pi}_{22}\hat{\gamma})
\]
\[
= s_2^2 + (\bar{\Pi}_{21} - \hat{\Pi}_{22}\hat{\gamma})(X_2'X_2/N)(\hat{\Pi}_{21} - \hat{\Pi}_{22}\hat{\gamma}).
\]
Then, letting \( a_1 = (\hat{\Pi}_{11} - \gamma \hat{\Pi}_{12})(X_2'X_2/N)(\hat{\Pi}_{21} - \hat{\Pi}_{22} \gamma) \), note that
\[
\mathbf{1}_{a_1} \rightarrow (\gamma - \gamma)\Pi_{22}' \hat{E} \Pi_{22} (\gamma - \gamma). \tag{A.9}
\]

Unless \( \Pi_{22} = 0 \), \( \Pi_{22}' \hat{E} \Pi_{22} \) is positive semidefinite while from (A.8) \((\gamma - \gamma)\) is a random variable \( O_{p11} \) (which is the sum of a normally distributed vector and other terms) such that the values for which \( \Pi_{22}(\gamma - \gamma) = 0 \) are of measure zero and so we obtain part (a).

**Part (b):** Note that
\[
s_2^2 = \begin{pmatrix}
1 & -\gamma' \end{pmatrix} \Sigma^{-1} \Sigma^{1/2} \begin{pmatrix}
\Sigma^{1/2} \Sigma^{-1} \\
\Sigma^{-1} \Sigma^{1/2}
\end{pmatrix}
\begin{pmatrix}
1 & -\gamma'
\end{pmatrix}
\rightarrow
\begin{pmatrix}
1, -\gamma', -\gamma'
\end{pmatrix}
\begin{pmatrix}
\omega_{11} & \omega_{12} & \omega_{13} \\
\omega_{12} & \omega_{22} & \omega_{23} \\
\omega_{13} & \omega_{23} & \omega_{33}
\end{pmatrix}
\begin{pmatrix}
1 \\
-\delta_1 \\
-\delta_2
\end{pmatrix}
\]
\[
= (1, -\delta_1, -\delta_2)
\begin{pmatrix}
(\omega_{11} & \omega_{12} & \omega_{13})
\omega_{22} & \omega_{23} & \omega_{33}
\omega_{13} & \omega_{23} & \omega_{33}
\end{pmatrix}
\begin{pmatrix}
1 \\
-\delta_1 \\
-\delta_2
\end{pmatrix}
+ (1, -\delta_1, -\delta_2)
\begin{pmatrix}
\omega_{13} \Omega_{33}^{-1/2} \\
\Omega_{23} \Omega_{33}^{-1/2} \\
\Omega_{33}^{-1/2}
\end{pmatrix}
\begin{pmatrix}
\Omega_{33}^{1/2} \omega_{13} \\
\Omega_{33}^{1/2} \omega_{23} \\
\Omega_{33}^{1/2}
\end{pmatrix}
\begin{pmatrix}
1 \\
-\delta_1 \\
-\delta_2
\end{pmatrix}
\]
\[
= \sigma^2 + a_2,
\]
where
\[
a_2 = (1, -\delta_1, -\delta_2)
\begin{pmatrix}
\omega_{13} \Omega_{33}^{-1/2} \\
\Omega_{23} \Omega_{33}^{-1/2} \\
\Omega_{33}^{-1/2}
\end{pmatrix}
\begin{pmatrix}
\Omega_{33}^{1/2} \omega_{13} \\
\Omega_{33}^{1/2} \omega_{23} \\
\Omega_{33}^{1/2}
\end{pmatrix}
\begin{pmatrix}
1 \\
-\delta_1 \\
-\delta_2
\end{pmatrix}. \tag{A.10}
\]

Since \( \hat{\delta}_2 \) is a continuous random vector \( O_{p11} \), \( a_2 \) is non-zero with probability 1 even in the limit. □

**Proof of theorem 5:**

Since \( \hat{\Pi}_2 \) is consistent,
\[
\hat{\Pi}_2^0 C \hat{\Pi}_2 \rightarrow \Pi_2 \hat{E} \Pi_2
\]
which has \( G_0 \) non-zero eigenvalues. Since the corresponding \( G_0 \) eigenvectors are continuous functions of the elements of the matrix and since \( \hat{G}_0 \rightarrow G_0 \), \( \hat{H}_1 \rightarrow H_1 \). Similarly,
\[
\hat{\Pi}_H \rightarrow [\Pi_{21} \Pi_{22}]
\begin{pmatrix}
1 & 0 \\
0 & H_1
\end{pmatrix}
\]
while
\[
\tilde{C}^{1/2} \Pi_{21} = \tilde{C}^{1/2} \Pi_{22} \gamma = \tilde{C}^{1/2} \Pi_{22} H' H' \gamma = \tilde{C}^{1/2} \Pi_{22} H_1 H_1' \gamma = \Phi_1 \delta_1.
\]
Furthermore, since
\[
\hat{\Omega}_R \xrightarrow{p} \left( \begin{array}{cc} 1 & 0 \\ 0 & H_1' \end{array} \right) \Sigma \left( \begin{array}{cc} 1 & 0 \\ 0 & H_1 \end{array} \right),
\]
while \( \sqrt{N} \text{vec}(\hat{H}_R - \Pi_R) \xrightarrow{d} N(0, \Omega_R \otimes I_{K_2}) \), \( \hat{H}_R \) and \( \hat{\Omega}_R \) have the properties that give rise to the standard attributes of the A-R test summarized in section 2.1. \( \square \)

**Proof of theorem 6:**

**Part (a):** As noted above, the estimates are consistent so that
\[
\tilde{\delta}_1 \xrightarrow{p} \delta_1,
\]
and
\[
\sigma^2 \tilde{\tau}_R \xrightarrow{p} \tau^2 = (1 - \delta_1)' \Omega_R (1 - \delta_1)'.
\]
(A.11)
By assumption 1
\[
\sqrt{N}(C^{1/2} \hat{H}_{21} - \Phi_1 \delta_1) = \sqrt{N}(\psi - \Psi_1 \delta_1) \xrightarrow{d} N(0, \sigma^2 \tilde{\tau}_R I_{K_2}).
\]
It follows by the standard argument that
\[
\tilde{\tau}_R \xrightarrow{d} \left( \tilde{\psi} - \tilde{\Psi}_1 \delta_1 \right)' [I - \Phi_1 \Phi_1' \Sigma_1^{-1} \Phi_1'] (\tilde{\psi} - \tilde{\Psi}_1 \delta_1) / \sigma^2 \tilde{\tau}_R \sim \chi^2(K_2 - G_0).
\]
The proof of part (b) is also the standard one while part (c) arises from noting that now
\[
\sqrt{N} C^{1/2} (\hat{H}_{21} - \hat{H}_{22} \delta_1) = \sqrt{N}(\psi - \Psi_1 \delta_1) + C^{1/2} \tilde{\delta}_1. \quad \square
\]

**Proof of theorem 7:**

Note that while \( \hat{H}_1 \xrightarrow{p} H_1, \hat{H}_2 \) remains a random matrix even in the limit. (\( H_2 \) itself can only be made well-defined through an arbitrary normalization.) However, since
\[
\hat{H}_1 \xrightarrow{p} H_1 \quad \text{and} \quad \hat{H}_2 \hat{H}_1 = 0, \quad \hat{H}_2 \xrightarrow{p} H_2 A \quad \text{where} \ A \ \text{is an orthogonal} \ (G - G_0) \times (G - G_0)
\]
random matrix. To see this, suppose \( \hat{H}'_2 [H_1 \quad H_2] = [0 \quad A'] \) so that \( \hat{H}_2 = H_2 A \). Then
\[
H' \hat{H} = \left[ \begin{array}{cc} I & 0 \\ 0 & A \end{array} \right]
\]
so that

\[
\hat{H}' \hat{H}' \hat{H} = I = \begin{bmatrix}
1 & 0 \\
0 & A' A
\end{bmatrix}
\]

showing that \(A\) is orthogonal. Thus, using (18) and (16),

\[
\hat{\Omega} = \begin{bmatrix}
1 & 0 & 0 \\
0 & I_{G_0} & 0 \\
0 & 0 & A'
\end{bmatrix}
\begin{bmatrix}
\omega_{11} & \omega_{12} & \omega_{13} \\
\omega_{12}' & \Omega_{22} & \Omega_{23} \\
\omega_{13}' & \Omega_{23} & \Omega_{33}
\end{bmatrix}
\begin{bmatrix}
1 & 0 & 0 \\
0 & I_{G_0} & 0 \\
0 & 0 & A
\end{bmatrix}.
\]

From this it follows immediately that

\[
\begin{bmatrix}
\begin{pmatrix}
\omega_{11} \\
\omega_{12}' \\
\omega_{13}'
\end{pmatrix} \\
\begin{pmatrix}
\omega_{12} \\
\Omega_{22} \\
\omega_{13}'
\end{pmatrix}
\end{bmatrix}
\end{bmatrix}
\begin{bmatrix}
\Omega_{33}^{-1} \Omega_{33}' \\
\Omega_{33}^{-1} \Omega_{33}'
\end{bmatrix}
\]

\[
\begin{bmatrix}
\omega_{11} & \omega_{12} & \omega_{13} \\
\omega_{12}' & \Omega_{22} & \Omega_{23} \\
\omega_{13}' & \Omega_{23} & \Omega_{33}
\end{bmatrix}
\begin{bmatrix}
1 & 0 & 0 \\
0 & I_{G_0} & 0 \\
0 & 0 & A
\end{bmatrix}
\begin{bmatrix}
\begin{pmatrix}
\omega_{11} \\
\omega_{12}' \\
\omega_{13}'
\end{pmatrix} \\
\begin{pmatrix}
\omega_{12} \\
\Omega_{22} \\
\omega_{13}'
\end{pmatrix}
\end{bmatrix}
\begin{bmatrix}
\Omega_{33}^{-1} \Omega_{33}' \\
\Omega_{33}^{-1} \Omega_{33}'
\end{bmatrix}
\]

Since \(\hat{\sigma}_1 \sim \hat{\sigma}_1\),

\[
\hat{\sigma}^2 \sim \sigma^2.
\]

The theorem then follows from the properties of \(\nu\) developed in lemma 2. \(\square\)

Proof of theorem 8:

Part (a): again let

\[
\Phi = [\phi \begin{bmatrix}\Phi_1 \\
\Phi_2\end{bmatrix} = C^{1/2} \Pi_2 \begin{bmatrix}1 & 0 & 0 \\
0 & H_1 & H_2\end{bmatrix}
\]

where the partition of \(\Phi\) is into \(1\), \(G_0\) and \((G - G_0)\) columns. Let \(\hat{\Phi}\) be defined similarly using \(\hat{\Pi}_2\). As noted in (9), \(\hat{\lambda} = \min \{N \{v(c - \hat{\Phi})\} \Omega \otimes I_{K_2}^{-1} v(c - \hat{\Phi})\} \) s.t. \(\rho[\hat{\Phi}] = G\).

\(\hat{\lambda}\) is invariant to nonsingular linear transformations applied to \(\Phi\) when the corresponding transformation is applied to \(\Omega \otimes I_{K_2}\), a fact already used here in writing \(\hat{\lambda}\) in terms of \(\Phi\) and \(\Omega\) rather than of \(\Pi_2\) and \(\Sigma\).

Let

\[
A_1 = \begin{bmatrix}1 & 0 \\
-1 & \hat{I}\end{bmatrix};
\]

\[
\Omega_1 = A_1' \Omega A_1;
\]

\[
A_2 = \begin{bmatrix}1 & -\omega_{21} / \omega_{11} \\
0 & \hat{I}\end{bmatrix};
\]

\[
\Omega_2 = A_2' \Omega A_2 = \begin{bmatrix}\omega_{21} & 0 \\
0 & \Omega_{22}\end{bmatrix}.
\]
where all partitionings are into 1 and \( G \) rows and columns. Define \( \hat{\Omega}^* \) correspondingly to \( \Omega^* \) in (A.12) using \( \hat{\Omega} \) instead of \( \Omega \). Furthermore, let

\[
\hat{\Phi}^* = \hat{\Phi}A_1A_2 = \begin{bmatrix} \hat{\Phi}^* & \hat{\Phi}_1 & \hat{\Phi}_2 \end{bmatrix}
\]

\[
\hat{\phi}^* = \hat{\phi}A_1A_2 = \begin{bmatrix} \hat{\phi}^* & \hat{\phi}_1 & \hat{\phi}_2 \end{bmatrix}
\]

and;

\[
\hat{\delta}^* = \begin{bmatrix} \hat{\Phi}_1'\hat{\phi}_1 & \hat{\Phi}_2'\hat{\phi}_2 \\
\hat{\phi}_1'\hat{\Phi}_1 & \hat{\phi}_2'\hat{\Phi}_2 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\phi}_1'\phi^* \\
\hat{\phi}_2'\phi^* \end{bmatrix}
\]

\[
= \begin{bmatrix} I_{G_0} & 0 \\
0 & \sqrt{NI_{G-C_0}} \end{bmatrix} \begin{bmatrix} \hat{\Phi}_1'\hat{\phi}_1 & \hat{\Phi}_2'\hat{\phi}_2 \\
\sqrt{N}\hat{\phi}_1'\hat{\Phi}_1 & N\hat{\phi}_2'\hat{\Phi}_2 \end{bmatrix}^{-1}
\]

\[
\begin{bmatrix} I_{G_0} & 0 \\
0 & \sqrt{NI_{G-C_0}} \end{bmatrix} \begin{bmatrix} \hat{\phi}_1^* \phi^* \\
\hat{\phi}_2^* \phi^* \end{bmatrix}
\]

\[
= \begin{bmatrix} 0 \\
B_{21}\sqrt{N}\hat{\phi}_1^*\phi^* + B_{22}N\hat{\phi}_2^*\phi^* \end{bmatrix} + o_p(1)
\]

\[
= \begin{bmatrix} o_p(1) \\
o_p(1) \end{bmatrix}.
\]

where

\[
B = \begin{bmatrix} B_{11} & B_{12} \\
B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} \hat{\Phi}_1'\hat{\phi}_1 & \hat{\Phi}_1'\sqrt{N}\hat{\phi}_2 \\
\sqrt{N}\hat{\phi}_1'\hat{\Phi}_1 & N\hat{\phi}_2'\hat{\Phi}_2 \end{bmatrix}^{-1}.
\]

Now consider the values

\[
\hat{\Phi}^* = (\hat{\Phi}^*_1 \hat{\Phi}^*_2)\delta^* \hat{\Phi}^*_1 \hat{\Phi}^*_2 = \begin{bmatrix} \hat{\phi}_1^* & \hat{\phi}_1^* \hat{\phi}^*_2 \end{bmatrix}.
\]

\( \hat{\Phi}^* \) are feasible for the problem defining \( \lambda \). Noting that \( \sqrt{N}\hat{\phi}^* \overset{d}{\rightarrow} N(0, \omega_1 I_{K_2}) \) and is independent asymptotically of \( \sqrt{N}(\hat{\phi}_1^* - \phi^*_1) = \hat{\phi}_2^* \), we obtain

\[
\lambda \leq \lambda
\]

\[
= N(\text{vec}(\hat{\Phi}^* - \phi^*))'(\hat{\Omega}^* \otimes I_{K_2})^{-1}(\text{vec}(\hat{\Phi}^* - \phi^*))
\]

\[
= N(\hat{\phi}^* - \phi^*)'(\hat{\phi}^* - \phi^*)/\omega_1
\]

\[
= N\hat{\phi}^* \left[ I - (\hat{\Phi}_1^* \sqrt{N}\hat{\phi}_2^*) \begin{bmatrix} \hat{\Phi}_1^* \hat{\phi}_1 & \sqrt{N}\hat{\phi}_1'\hat{\Phi}_2 \\
\sqrt{N}\hat{\phi}_1'\hat{\Phi}_1 & N\hat{\phi}_2'\hat{\Phi}_2 \end{bmatrix}^{-1} \begin{bmatrix} \phi_1' \\
\phi_2' \sqrt{N}\phi_2' \end{bmatrix} \right] \hat{\phi}^*/\omega_1
\]

\[
+ o_p(1)
\]

\[
\overset{d}{\rightarrow} \chi^2(K_2 - G).
\]
where again the distributional conclusion arises from noting that this is the asymptotic
distribution conditional on \([\bar{\Phi}_1, \sqrt{N}\bar{\Phi}_2^*}\).

Now, to ensure appropriate orders of magnitude, let

\[ h = \sqrt{N} \text{vec}(\bar{\Phi}_1^* - \bar{\Phi}_1^2) (\bar{\Phi}_2^* - \bar{\Phi}_2^2) \]

and

\[ \lambda = N \text{vec}(\langle \hat{\phi}^* - (\bar{\Phi}_1^* \bar{\Phi}_2^*) \delta^* \rangle) (\hat{\Phi}_1^* - \hat{\Phi}_1^2) (\hat{\Phi}_2^* - \hat{\Phi}_2^2))' \hat{\Omega}^{-1} \otimes I_{K_2} \]

\( = \text{vec}(\hat{\phi}^* - (\bar{\Phi}_1^* \bar{\Phi}_2^* \delta^*) (\hat{\Phi}_1^* - \hat{\Phi}_1^2) (\hat{\Phi}_2^* - \hat{\Phi}_2^2)). \]

Observe that

\[ \frac{\partial \lambda}{\partial \phi_{11}} = -2\sqrt{N} \hat{\phi}^* - (\bar{\Phi}_1^* \bar{\Phi}_2^*) \delta^* \hat{\phi}^* / \hat{\psi}_{11} \]

(A.13)

\[ = -2\sqrt{N} (\hat{\phi}^* - (\bar{\Phi}_1^* \bar{\Phi}_2^* \delta^*)' (0 \quad \delta^* \gamma / \omega_{11} + o_p(1)), \]

since the partial derivatives of \( \lambda \) with respect to \( \delta^* \) are zero identically. Since \( \delta^*_2 = O_p(1) \)
and is a continuous random variable, the derivative in (A.13) is non-zero with probability
1, which yields the conclusion that asymptotically \( \hat{\lambda} < \lambda \) a.s. \( \Box \)

Part (b): \( \hat{\lambda} \) now goes in distribution to the smallest eigenvalues of a non-central Wishart
matrix of dimension \((G - G_0) \) with \( K_2 \) degrees of freedom. If \( G - G_0 = 1 \), then this is the
non-central \( \chi^2 \) distribution and the power is greater than the nominal size asymptotically.
If \( (G - G_0) > 1 \), the relationship of distribution of the smallest root to the standard \( \chi^2(K_2) \)
distribution seems not to be known, but the root is \( O_p(1) \) so that in neither case does
the power go to unity.

Proof of Proposition 9:

Now \( \sqrt{N}\bar{\Phi}_2 \xrightarrow{d} \Phi_2 + \bar{\Psi}_2 = \bar{\Phi}_2 \), say, with \( \text{vec}\bar{\Phi}_2 \sim N(\text{vec}\bar{\Phi}_2, \Omega_2 \otimes I_{K_2}). \) The equivalent
of expression (A.1) is

\[ \hat{\phi} = \hat{\phi}_1 \delta_1 + \hat{\phi}_2 \delta_2 + \hat{\phi}_2 \alpha_2 - \Psi_2 \delta_2 - \Phi_2 \alpha_2 + \eta. \]

where \( \eta \) is defined in (A.3). It is still the case that \( \sqrt{N}\eta \xrightarrow{d} \eta \sim N(0, \sigma^2 I_{K_2}) \) which is
independent of \( \hat{\Phi}_2 \), being independent of \( \bar{\Psi}_2 \). Hence, conditional on \( \hat{\Phi}_2 \),

\[ \nu / \sigma^2 = (\sqrt{N}\eta - \sqrt{N}\Psi_2 \delta_2 - \bar{\Phi}_2 \alpha_2)' \]
\[
\begin{bmatrix}
I_{K_2} - (\Phi_1 \sqrt{N} \Phi_2) & \Phi_1' \sqrt{N} \Phi_2
\end{bmatrix}^{-1}
\begin{bmatrix}
\sqrt{N} \Phi_1' \Phi_2
\end{bmatrix}^{-1}
\begin{bmatrix}
(\Phi_1' \sqrt{N} \Phi_2)
\end{bmatrix}
\]
\[
(\sqrt{N} \eta - \sqrt{N} \Psi_2 \delta_2 - \Phi_2 \alpha_2)/\sigma^2
\]
\[
\xrightarrow{d}
(\tilde{\eta} - \tilde{\Psi}_2 \delta_2 - \tilde{\Phi}_2 \alpha_2)'
\begin{bmatrix}
I_{K_2} - (\Phi_1' \tilde{\Phi}_2)
\end{bmatrix}
\begin{bmatrix}
\Phi_1' \Phi_1 & \Phi_1' \Phi_2
\Phi_2' \Phi_1 & \Phi_2' \Phi_2
\end{bmatrix}^{-1}
\begin{bmatrix}
\Phi_1'
\Phi_2'
\end{bmatrix}
(\tilde{\eta} - \tilde{\Psi}_2 \delta_2 - \tilde{\Phi}_2 \alpha_2)/\sigma^2.
\]
\[
\sim \chi^2((K_2 - G), q)
\]

where
\[
q = (\tilde{\Psi}_2 \delta_2 + \tilde{\Phi}_2 \alpha_2)
\]
\[
\begin{bmatrix}
I_{K_2} - (\Phi_1' \tilde{\Phi}_2)
\end{bmatrix}
\begin{bmatrix}
\Phi_1' \Phi_1 & \Phi_1' \Phi_2
\Phi_2' \Phi_1 & \Phi_2' \Phi_2
\end{bmatrix}^{-1}
\begin{bmatrix}
\Phi_1'
\Phi_2'
\end{bmatrix}
(\tilde{\Psi}_2 \delta_2 + \tilde{\Phi}_2 \alpha_2)/\sigma^2.
\]

Needless to say, the marginal distribution is a great deal more complicated, since \( q \) is a highly non-linear transformation of \( \tilde{\Psi}_2 \) and the distribution is in principle obtained by integrating \( q \) out of the product of the conditional distribution of \( \nu/\sigma \) and the distribution of \( q \).

Formally, the expressions for the denominators of the Byron and the Basmann tests are the same as those presented in (A.9) and (A.10), the argument of lemma 2 going through without alteration due to replacement of assumption 2 with assumption 3. Thus formally \( a_2 = a_1 \) and \( a_3 = a_2 \). However, the expression for \( \hat{\gamma} \) is no longer that in equation (A.8). Instead, letting
\[
D^* = \begin{bmatrix}
D_{11}^* & D_{12}^*
D_{21}^* & D_{22}^*
\end{bmatrix}
= \begin{bmatrix}
\Phi_1' \Phi_1 & \Phi_1' \tilde{\Phi}_2
\tilde{\Phi}_2' \Phi_1 & \tilde{\Phi}_2' \tilde{\Phi}_2
\end{bmatrix}^{-1},
\]
equation (A.8) becomes, with assumption 3,
\[
\hat{\gamma} \xrightarrow{d} \gamma + H_2 \alpha_2 + H_2(D_{21}^* \Phi_1' + D_{22}^* \tilde{\Phi}_2')(\tilde{\eta} - \tilde{\Psi}_2 \delta_2 - \tilde{\Phi}_2 \alpha_2).
\]

Unfortunately, there seems to be no useful way of comparing the magnitude of \( a_2 \) and \( a_3 \), for which \( \hat{\gamma} \) is defined by (A.14) with \( a_1 \) and \( a_2 \) which use the expression in (A.8). □
Proof of Proposition 10

Since $G_0$ is consistent, we can again concentrate on the behavior of $N \hat{\Phi}(I - \hat{\Phi}_1[\hat{\Phi}_1^T\hat{\Phi}_1]^{-1}\hat{\Phi})$. We can now write

$$\sqrt{N}(\hat{\Phi} - \hat{\Phi}_1 \delta_1) = \sqrt{N} \psi - \sqrt{N} \Psi_1 \delta_1 + \hat{\Phi}_2 \delta_2 \xrightarrow{d} N(\hat{\Phi}_2 \delta_2, \sigma^2_R I_{K_1})$$

where $\sigma^2_R$ is defined in (A.11). Hence $\lambda_\tau$ now has the properties under the null that the restricted A-R test, developed in subsection 2.3, had under a local alternative, with $\bar{b} = \bar{b}^* = \hat{\Phi}_2 \delta_2$. Exactly the same considerations apply to $\tau_\tau$. Hence,

$$\lambda_1 = \bar{b}^*[(\hat{C} - \hat{C} \Pi_{22}^R[\Pi_{22}^R \hat{C} \Pi_{22}^R]^{-1}\Pi_{22}^R \hat{C})\bar{b}^*/\sigma^2_R]$$

If, in addition, $\beta_2 = \bar{b}/\sqrt{N}$, then the relevant vector becomes $\bar{b}^{**} = \bar{b}^* + \bar{b}$. The noncentrality parameter arising from it is

$$\lambda_2 = \bar{b}^{**}[(\hat{C} - \hat{C} \Pi_{22}^R[\Pi_{22}^R \hat{C} \Pi_{22}^R]^{-1}\Pi_{22}^R \hat{C})\bar{b}^{**}/\sigma^2_R]$$

Whether $\lambda_2$ is greater or smaller than $\lambda_1$ depends on what are the values of $\bar{b}$ and $\bar{b}^*$. \qed
References


