Specification Tests in Econometrics

J. A. Hausman

Econometrica, Volume 46, Issue 6 (Nov., 1978), 1251-1271.

Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28197811%2946%3A6%3C1251%3ASTIE%3E2.0.CO%3B2-X

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

Econometrica is published by The Econometric Society. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/econosoc.html.

Econometrica
©1978 The Econometric Society

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact jstor-info@umich.edu.

©2002 JSTOR

http://www.jstor.org/
Sun Nov 3 10:31:39 2002

ECONOMETRICA

VOLUME 46 NOVEMBER, 1978 NUMBER 6

SPECIFICATION TESTS IN ECONOMETRICS

BY J. A. HAUSMAN[1]

Using the result that under the null hypothesis of no misspecification an asymptotically efficient estimator must have zero asymptotic covariance with its difference from a consistent but asymptotically inefficient estimator, specification tests are devised for a number of model specifications in econometrics. Local power is calculated for small departures from the null hypothesis. An instrumental variable test as well as tests for a time series cross section model and the simultaneous equation model are presented. An empirical model provides evidence that unobserved individual factors are present which are not orthogonal to the included right-hand-side variable in a common econometric specification of an individual wage equation.

1. INTRODUCTION

SPECIFICATION TESTS FORM ONE of the most important areas for research in econometrics. In the standard regression framework, $y = X\beta + \varepsilon$, the two stochastic specifications are first that the conditional expectation of ε given X be zero (or for fixed X, ε have expectation zero) and that ε have a spherical covariance matrix

$$(1.1a) \qquad E(\varepsilon|X) = 0 \quad \text{or in large samples} \quad \operatorname{plim} \frac{1}{T} X'\varepsilon = 0,$$

$$(1.1b) \qquad V(\varepsilon|X) = \sigma^2 I.$$

Failure of the first assumption, sometimes called the orthogonality assumption, leads to biased estimates while failure of the second assumption, sometimes called the sphericality assumption, leads to loss of efficiency although the central tendency of the estimator is still correct. While in many problems the payoff to detecting failure of assumption (1.1a) is presumably greater than detecting failure of assumption (1.1b), most of the attention in the econometric literature has been paid to devising tests for the latter assumption. Ramsey [**16**] and Wu [**25**] are among the few references to specification tests. Yet, the problem is so important that increased attention should be paid, especially since efficient estimators under assumption (1.1) are now available in almost all situations; and these estimators are often quite sensitive to failures of the first assumption.

In this paper a general form of specification test is proposed which attempts to provide powerful tests of assumption (1.1a) and presents a unified approach to specification error tests. Thus, an ad hoc test would not need to be devised for each specific situation, but the general scheme presented here could be applied

[1] I would like to thank T. Amemiya, D. W. Carlton, G. Chamberlain, G. Chow, F. M. Fisher, Z. Griliches, R. H. Gordon, R. E. Hall, T. J. Rothenberg, H. L. White, and A. Zellner for helpful discussions. A. S. Kelso and E. R. Rosenthal provided research assistance. Research support has been provided by the NSF. An editor and referee of *Econometrica* provided very helpful comments. The views expressed in this paper are the author's sole responsibility and do not reflect those of the Department of Economics, the Massachusetts Institute of Technology, or the National Science Foundation.

to specific situations. A main stumbling block to specification tests has been a lack of precisely specified alternative hypotheses. Here, I point out that in many situations, including time series-cross section specifications, errors in variables specifications, and simultaneous equation specifications, the alternative hypothesis that assumption (1.1a) fails may be tested in an expanded regression framework. The basic idea follows from the existence of an alternative estimator which is consistent under both null and alternative hypotheses. By comparing the estimates from this estimator with the efficient estimator (under assumption 1.1a) and noting that their difference is uncorrelated with the efficient estimator when the null hypothesis is true, easily used tests may be devised from the regression

$$y = X\beta + \tilde{X}\alpha + v \tag{1.2}$$

where $\tilde{X}$ is a suitably transformed version of X. The specification tests are performed by constructing a test of the hypothesis H_0: $\alpha = 0$. Also local power considerations are discussed, and the distribution of the power function under the alternative hypothesis is derived.

In Section 2 the basic lemma regarding these types of specification tests is proven. The test is applied to an errors in variables problem and equation (1.2) is derived. The following two sections discuss two new specification tests for the time series-cross section model and for the simultaneous equation model. Both tests are always available (unlike the errors in variables test which requires an instrumental variable) and should be used for these two important model specifications. Lastly, an example is provided. The example is interesting since a widely used time series-cross section specification, the random effects model, is found not to be consistent with the alternative specification. The general principle of this paper can be applied in additional problems not considered here. Therefore the tests should be useful to the applied econometrician.

2. THEORY AND A TEST OF ERRORS IN VARIABLES

The theory underlying the proposed specification tests rests on one fundamental idea. Under the (null) hypothesis of no misspecification, there will exist a consistent, asymptotically normal and asymptotically efficient estimator, where efficiency means attaining the asymptotic Cramer–Rao bound.[2] Under the alternative hypothesis of misspecification, however, this estimator will be biased and inconsistent. To construct a test of misspecification, it is necessary to find another estimator which is not adversely affected by the misspecification; but this estimator will not be asymptotically efficient under the null hypothesis. A consideration of the difference between the two estimates, $\hat{q} = \hat{\beta}_1 - \hat{\beta}_0$ where $\hat{\beta}_0$

[2] This paper will concentrate on the large sample case since in each test one or both of the estimators has a normal distribution only asymptotically. Most econometric estimators, except for least squares, have this property. A discussion of the notion of asymptotic efficiency may be found in Rothenberg [18, Ch. 2]. Henceforth, efficient will stand for asymptotically efficient and likewise for bias, while variance means variance of the asymptotic distribution. Analogous finite sample results hold true under appropriate conditions.

is the efficient estimate under H_0 and $\hat{\beta}_1$ is a consistent estimator under H_1, will then lead to a specification test. If no misspecification is present, the probability limit of $\hat{q}$ is zero. With misspecification plim $\hat{q}$ will differ from zero; and if the power of the test is high, $\hat{q}$ will be large in absolute value relative to its asymptotic standard error. Hopefully, this procedure will lead to powerful tests in important cases because the misspecification is apt to be serious only when the two estimates differ substantially.

In constructing tests based on $\hat{q}$, an immediate problem comes to mind. To develop tests not only is the probability limit of $\hat{q}$ required, but the variance of the asymptotic distribution of $\sqrt{T}\hat{q}$, $V(\hat{q})$, must also be determined. Since $\hat{\beta}_0$ and $\hat{\beta}_1$ use the same data, they will be correlated which could lead to a messy calculation for the variance of $\sqrt{T}\hat{q}$. Luckily, this problem is resolved easily and, in fact, $V(\hat{q}) = V(\hat{\beta}_1) - V(\hat{\beta}_0) = V_1 - V_0$ under the null hypothesis of no misspecification. Thus, the construction of specification error tests is simplified, since the estimators may be considered separately because the variance of the difference $\sqrt{T}\hat{q} = \sqrt{T}(\hat{\beta}_1 - \hat{\beta}_0)$ is the difference of the respective variances. The intuitive reasoning behind this result is simple although it appears to have remained generally unnoticed in constructing tests in econometrics. The idea rests on the fact that the efficient estimator, $\hat{\beta}_0$, must have zero asymptotic covariance with $\hat{q}$ under the null hypothesis for any other consistent, asymptotically normal estimator $\hat{\beta}_1$. If this were not the case, a linear combination of $\hat{\beta}_0$ and $\hat{q}$ could be taken which would lead to a consistent estimator $\hat{\beta}_*$ which would have smaller asymptotic variance than $\hat{\beta}_0$ which is assumed asymptotically efficient. To prove the result formally, consider the following lemma:

LEMMA 2.1: *Consider two estimators $\hat{\beta}_0$, $\hat{\beta}_1$ which are both consistent and asymptotically normally distributed with $\hat{\beta}_0$ attaining the asymptotic Cramer–Rao bound so $\sqrt{T}(\hat{\beta}_0 - \beta) \stackrel{A}{\sim} N(0, V_0)$ and $\sqrt{T}(\hat{\beta}_1 - \beta) \stackrel{A}{\sim} N(0, V_1)$ where V_0 is the inverse of Fisher's information matrix. Consider $\hat{q} = \hat{\beta}_1 - \hat{\beta}_0$. Then the limiting distributions of $\sqrt{T}(\hat{\beta}_0 - \beta)$ and $\sqrt{T}\hat{q}$ have zero covariance, $C(\hat{\beta}_0, \hat{q}) = 0$, a zero matrix.*[3,4]

PROOF: Suppose $\hat{\beta}_0$ and $\hat{q}$ are not orthogonal. Since plim $\hat{q} = 0$ define a new estimator $\hat{\beta}_2 = \hat{\beta}_0 + rA\hat{q}$ where r is a scalar and A is an arbitrary matrix to be chosen. The new estimator is consistent and asymptotically normal with asymptotic variance

$$V(\hat{\beta}_2) = V(\hat{\beta}_0) + rAC(\hat{\beta}_0, \hat{q}) + rC'(\hat{\beta}_0, \hat{q})A' + r^2AV(\hat{q})A'. \tag{2.2}$$

Now consider the difference between the asymptotic variance of the new

[3] Besides consistency and asymptotic normality, uniform convergence is also required to rule out superefficiency. However, it is not difficult to demonstrate that standard econometric estimators converge uniformly. A sufficient condition which leads to a straightforward proof is to assume that the parameter space is compact. T. Amemiya and T. Rothenberg have helped in resolving this issue.

[4] A statement of this lemma in the finite sample case for one parameter is contained in a paper by R. A. Fisher [8], a reference supplied by W. Taylor. It is clearly related to an asymptotic version of the Rao-Blackwell theorem (Rao [17]).

estimator and the old asymptotically efficient estimator

$$F(r) = V(\hat{\beta}_2) - V(\hat{\beta}_0) = rAC + rC'A' + r^2 AV(\hat{q})A'. \tag{2.3}$$

Taking derivatives with respect to r yields

$$F'(r) = AC + C'A' + 2rAV(\hat{q})A'. \tag{2.4}$$

Then choose $A = -C'$ and note that C is symmetric, which leads to

$$F'(r) = -2C'C + 2rC'V(\hat{q})C. \tag{2.5}$$

Therefore at $r = 0$, $F'(0) = -2C'C \leqslant 0$ in the sense of being nonpositive definite. But $F(0) = 0$ so for r small $F(r) < 0$ and there is a contradiction unless $C = C(\hat{\beta}_0, \hat{q}) = 0$ since $\hat{\beta}_0$ is asymptotically efficient implies $F(r) \geqslant 0$.

Once it has been shown that the efficient estimator is uncorrelated with $\hat{q}$, the asymptotic variance of $\hat{q}$ is easily calculated.

COROLLARY 2.6: $V(\hat{q}) = V(\hat{\beta}_1) - V(\hat{\beta}_0) \geqslant 0$ *in the sense of being nonnegative definite.*

PROOF: Since $\hat{q} + \hat{\beta}_0 = \hat{\beta}_1$, $V(\hat{q}) + V(\hat{\beta}_0) = V(\hat{\beta}_1)$. Furthermore, $\hat{\beta}_0$ attains the asymptotic CR bound. Given the above result a general misspecification test can be specified by considering the statistic

$$m = T\hat{q}'\hat{V}(\hat{q})^{-1}\hat{q} \tag{2.7}$$

where $\hat{V}(\hat{q})$ is a consistent estimate of $V(\hat{q})$. This statistic will be shown to be distributed asymptotically as central χ^2_K under the null hypothesis where K is the number of unknown parameters in β when no misspecification is present. In what follows, it is sometimes easier to work with $\hat{q}$ rather than $\sqrt{T}\hat{q}$ so define $M_0 = (1/T)V_0$, $M_1 = (1/T)V_1$, and $M(\hat{q}) = (1/T)V(\hat{q})$. In terms of the M's, the statistic is $m = \hat{q}'\hat{M}(\hat{q})^{-1}\hat{q}$.

The statistic m in equation (2.7) specifies the distribution of the difference of the two estimators when no misspecification is present. The other operating characteristic of a test is its power. Unfortunately, power considerations have not been paid much attention in econometrics, probably due to the impreciseness of alternative hypotheses and the complexity of deriving distributions of power functions. The power of our specification test depends on the nonnull distribution of the statistic in equation (2.7). In most applications I will show that the power can be approximated in large samples, for alternatives close to the null hypothesis, by the noncentral χ^2 distribution with noncentrality parameter

$$\delta^2 = \bar{q}'M(\hat{q})^{-1}\bar{q}, \tag{2.8}$$

where $\bar{q} = \operatorname{plim}(\hat{\beta}_1 - \hat{\beta}_0)$ the probability limit of the difference between the two estimates.[5]

[5] The discussion of local power which follows is due to the extremely helpful guidance of T. J. Rothenberg. A good reference is Cox and Hinkley [6, Ch. 9].

Power considerations are important because they give the probability of rejecting the null hypothesis when it is false. In many empirical investigations $\hat{\beta}_0$ and $\hat{\beta}_1$ seem to be far apart yet the null hypothesis that $q = 0$ is not rejected. If the probability of rejection is small for a difference of say q_A where q_A is large enough to be important, then not much information has been provided by the test. Now deriving the large sample distributions of test statistics under the alternative hypothesis is a difficult matter especially under a wide range of alternative hypotheses which are being considered here. Therefore, I will only be able to derive the asymptotic distributions of the power function of a sequence of models under local conditions where the sequence of alternatives $\bar{q}$ is of order $a/\sqrt{T}$ where a is a constant vector. Only alternatives close to the null hypothesis can be investigated in this manner but they hopefully provide a guide to a broader range of cases. The necessity of this limitation can be best shown by a simple example. Consider a two equation triangular system

$$y_1 = x_1\gamma + u_1, \tag{2.9a}$$

$$y_2 = \beta y_1 + u_2. \tag{2.9b}$$

If u_1 and u_2 have zero covariance, least squares on equation (2.9b) is the (asymptotically) efficient estimator of β while with nonzero covariance it is inconsistent. Then, an instrumental variable estimator (say two stage least squares) is consistent. The test statistic m from equation (2.7) is asymptotically equivalent to a test that $\sigma_{12} = 0$ where the estimated covariance is formed from the residuals of the 2SLS estimate of equation (2.9b), $\hat{u}_2$, and the residuals of the OLS estimate of equation (2.9a), $\hat{u}_1$. Under the alternative hypothesis assume that the true covariance is σ_{12}, and I want to construct a test based on the fact that $\sqrt{T}(\hat{\sigma}_{12} - \sigma_{12}) \overset{A}{\sim} N(0, v_{12})$. Assume a consistent estimate $\hat{v}_{12}$ is used for v_{12} and let $v_{12}^{\frac{1}{2}} = w$ while $\hat{v}_{12}^{\frac{1}{2}} = \hat{w}$. Tests will be usually formed from the statistic $\sqrt{T}[(\hat{\sigma}_{12} - \sigma_{12}^0)/\hat{w}]$ where σ_{12}^0 is the hypothesized value of σ_{12}; here $\sigma_{12}^0 = 0$. Adding and subtracting the true σ_{12} leads to the expression

$$\sqrt{T}\left(\frac{\hat{\sigma}_{12} - \sigma_{12}}{\hat{w}}\right) - \sqrt{T}\left(\frac{\sigma_{12}^0 - \sigma_{12}}{\hat{w}}\right). \tag{2.10}$$

Under the null hypothesis only the first term is present since $\sigma_{12} = \sigma_{12}^0 = 0$ so asymptotically normal or central χ^2 distributions are derived for tests of $\hat{\sigma}_{12} = \sigma_{12}^0$. When $\sigma_{12} \neq \sigma_{12}^0$ under the alternative hypothesis, the second term remains finite only when a sequence of models is considered so that $\sqrt{T}(\sigma_{12} - \sigma_{12}^0)$ converges to a finite constant since $\hat{w}$ is a consistent estimate of w. Otherwise the second term explodes and large sample power functions cannot be derived unless further approximations are made. However, the explosion of this term insures a consistent test. The analysis of the case where σ_{12} converges to σ_{12}^0 at rate $\sqrt{T}$ corresponds to the idea of local power: the distribution of the test statistic under the alternative hypothesis is considered for cases close to the null hypothesis.

To return from the simple example to our more general case I will consider a sequence of models corresponding to the concept of local power. Thus as before under H_0, I assume both estimates are consistent, that $\sqrt{T}(\hat{\beta}_0-\beta) \xrightarrow{A} N(0, V_0)$, and that $\sqrt{T}(\hat{\beta}_1-\beta) \xrightarrow{A} N(0, V_1)$. I assume that under H_1, $\sqrt{T}(\hat{\beta}_0-\text{plim}\,\hat{\beta}_0)$ and $\sqrt{T}(\hat{\beta}_1-\beta)$ are asymptotically normal with covariance matrices that are continuous functions of the true β.

THEOREM 2.1: *Under H_0, the test statistic $m = T\hat{q}'\hat{V}(\hat{q})^{-1}\hat{q} \xrightarrow{A} \chi^2_K$ where $\hat{V}(\hat{q})$ is a consistent estimate (under H_0), of $V(\hat{q})$ using $\hat{\beta}_1$ and $\hat{\beta}_0$.*[6]

PROOF: Let $\sqrt{T}\hat{q} = \sqrt{T}(\hat{\beta}_1-\hat{\beta}_0) \xrightarrow{A} N(0, V(\hat{q}))$ using the corollary. Then $T\hat{q}'\hat{V}(\hat{q})^{-1}q$ is distributed asymptotically as central χ^2_K since it has the same asymptotic distribution as $T\hat{q}'V(\hat{q})^{-1}\hat{q}$.

As an approximation for practical use, the statistic $\hat{q}'\hat{M}(\hat{q})^{-1}\hat{q}$ can be used in place of m.

To derive the asymptotic distribution of the test statistic under the alternative hypothesis, consider local alternatives. That is, consider a sequence of models such that the sequence of alternatives $\bar{q}$ is of order $(1/\sqrt{T})$. Then I can show that as long as $\hat{V}(\hat{q})$ approaches $V(\hat{q})$ the asymptotic distribution of the test statistic is non-central χ^2.

THEOREM 2.2: *Under H_1 consider a sequence of models represented by a sequence of parameters $q/\sqrt{T}(q \neq 0)$ so that $g_T = \text{plim}\,\hat{\beta}_{0T}-\beta = \check{\beta}_T-\beta$ such that $\lim_{T\to\infty}\sqrt{T}g_T = a < \infty$. Then as $T\to\infty$ along the chosen path $m_T = T\hat{q}_T'\hat{V}_T(\hat{q})^{-1}\hat{q}_T$ is distributed asymptotically as noncentral χ^2 with k degrees of freedom and noncentrality parameter $\delta^2 = \lim_{T\to\infty} Tg_T'V(\hat{q})^{-1}g_T$ which is approximately $\bar{q}'M(\hat{q})^{-1}\bar{q}$ so long as $\hat{V}_T(\hat{q})$ is a consistent estimate of $V(\hat{q})$ under H_1.*

PROOF: Because the asymptotic covariance matrices of $\hat{\beta}_0$ and $\hat{\beta}_1$ are continuous functions of β, along the sequence of local departures of the model as $T\to\infty$, their covariance matrices approach V_0 and V_1, respectively. For each local departure from the null hypothesis in the sequence, $\{g_T\}$, $\hat{\beta}_{0T}$ is inconsistent. However, since the departures are only local, it can be shown (Cox and Hinkley [**6**, pp. 317–18]) that asymptotically $\sqrt{T}(\hat{\beta}_{0T}-\check{\beta}_T) \xrightarrow{A} N(0, V_0)$. Thus although the mean of the asymptotic distribution of $\hat{\beta}_0$ has changed from the true β to $\check{\beta}_T$, the asymptotic variance remains the same. Furthermore $\hat{V}_T(\hat{q})$, the estimate of $V(\hat{q})$ is still consistent. Therefore, since asymptotically $\sqrt{T}\hat{q}_T \xrightarrow{A} N(a, V(\hat{q}))$ the test statistic m_T is distributed approximately as noncentral χ^2 with degrees of freedom k and noncentrality parameter δ^2.

To make this argument more concrete, return to the example of equations (2.9). Define $K_T = (1/T)\gamma'x_{1T}'x_{1T}\gamma$ and assume it approaches a finite limit K.

[6] Any consistent estimate of $V(\hat{q})$ under H_0 is sufficient to cause Theorem 2.1 to hold. Power considerations under H_1 may lead to a specific choice of an estimate. These considerations are discussed for a specific example following equation (2.11).

Now under H_1 let $\sigma_{12} \neq 0$ and then the inconsistency in $\hat{\beta}_0$ is $\operatorname{plim} \hat{\beta}_0 - \beta = \sigma_{12}/(K+\sigma_{22})$. To determine the limiting distribution of $\hat{\beta}_0$ it is convenient to assume that u_1 and u_2 have a bivariate normal distribution. Then Rothenberg [19] has shown that

$$(2.11) \quad \sqrt{T}\left[\hat{\beta} - \beta - \frac{\sigma_{12}}{K_T + \sigma_{22}}\right] \stackrel{A}{\sim} N\left[0, \frac{\sigma_{11}}{K+\sigma_{22}} - \frac{\sigma_{12}^2}{(K+\sigma_{22})^2} - \frac{2\sigma_{12}^2 K^2}{(K+\sigma_{22})^4}\right]$$

where $\lim \sqrt{T}[\sigma_{12}/(K+\sigma_{22})]$ is the a of Theorem 2.2. However $V_0 = \sigma_{11}/(K+\sigma_{22})$ so it needs to be shown that for local departures from the null hypothesis the last two terms in the asymptotic variance disappear as $T \to \infty$. But since by assumption $\sqrt{T}\sigma_{12}$ converges to a (finite) constant the terms involving σ_{12}^2 converge to zero so long as $(K+\sigma_{22})$ is nonzero. Thus, for local departures in large samples V_0 gives a correct approximation and a noncentral χ^2 distribution may be used.[7,8]

For a given size of test the power increases with δ^2 which in turn depends on how far the plim of the biased and inconsistent estimator $\hat{\beta}_0$ is from the plim of the consistent estimator $\hat{\beta}_1$ when misspecification is present. Thus, the comparison estimator $\hat{\beta}_1$ should be chosen so that if a certain type of misspecification is feared to be present, $\hat{q}$, which is the difference of the estimates, is expected to be large. The other consideration in equation (2.8) is to keep $V(\hat{q})$ small so that a large departure between $\hat{\beta}_0$ and $\hat{\beta}$ will not arise by chance. This requirement means that $\hat{\beta}_1$ should be relatively efficient but at the same time sensitive to departures from the model specification. To highlight the power considerations the specification test of equation (2.7) will be reformulated in a statistically equivalent form. Also, the reformulated test may be easier to use with available econometrics computer programs. To demonstrate this reformulated test, an errors in variables example is considered.

An errors in variables test attempts to determine if stochastic regressors and the disturbances in a regression are independent. In the simplest case consider the model

$$(2.12) \quad y_i = \beta x_i + \varepsilon_{1i} \qquad (i = 1, \ldots, T)$$

where ε_{1i} is iid with mean zero and distributed normally. Under the null

[7] Wu's [25] derivation of the (nonlocal) limiting distribution of the test statistic under the alternative hypothesis in equation (3.12) of his paper seems incorrect since application of the central limit theorem on p. 748 requires the sum of random variables with zero mean. Thus his variable e_1 does not have a limiting distribution. Interpreted locally Wu's results seem valid since only the usual least squares variance term V_0 is needed.

[8] A referee points out that in general there may exist many estimates of $V(\hat{q})$ which are consistent under H_0 so that Theorem 2.1 holds. However, the estimate which provides the greatest power under H_1 is preferable to use. If one considers a class of estimators such that $\operatorname{plim} \hat{V}(\hat{q}) = cV(\hat{q})$ where c is a constant, Wu's [25] results (interpreted locally) lead to the conclusion that one should use the estimator associated with the smallest c. Therefore for the instrumental variable test, s_0^2, the least squares estimate of σ^2, seems appropriate to use for this example. For Theorem 2.2 to hold in general, consistent estimates of the nuisance parameters are required.

hypothesis x_i and ε_1 are orthogonal in large samples:

$$(2.13) \quad \operatorname{plim} \frac{1}{T} x'\varepsilon_1 = 0,$$

while under the alternative hypothesis the plim is nonzero.

The efficient estimator under the null hypothesis is, of course, least squares. Under the alternative hypothesis least squares is biased and inconsistent with H_1: $\operatorname{plim} \hat{\beta}_0 = \beta(m_x^2 - \sigma_{\varepsilon_2}^2)/m_x^2$ where the observed $x_i = x_i^* + \varepsilon_{2i}$ is the sum of the "true" regressor and a normal random variable with mean zero which is assumed independent of ε_{1i} while $m_x^2 = \operatorname{plim} (1/T)x'x$. The comparison estimator $\hat{\beta}_1$ will be an instrumental variable (IV) estimator based on the instrument z with properties

$$(2.14) \quad \operatorname{plim} \frac{1}{T} z'\eta = 0, \qquad \operatorname{plim} \frac{1}{T} z'x \neq 0, \qquad \text{for } \eta_i = \varepsilon_{1i} - \beta\varepsilon_{2i}.$$

Then the IV estimator is

$$(2.15) \quad \hat{\beta}_1 = (z'x)^{-1} z'y.$$

To form the test statistic under the null hypothesis using Corollary 2.6

$$(2.16) \quad \sqrt{T}\hat{q} = \sqrt{T}(\hat{\beta}_1 - \hat{\beta}_0) \xrightarrow{A} N(0, D)$$

with $D = V(\hat{q}) = \sigma^2[\operatorname{plim} ((1/T)\hat{x}'\hat{x})^{-1} - \operatorname{plim} ((1/T)x'x)^{-1}]$ where $\hat{x} = z(z'z)^{-1}z'x$. Again using the corollary, $T\hat{q}'D^{-1}q$ is distributed as χ_1^2 under the null hypothesis. Then the test of misspecification using s_1^2, the IV estimator of σ^2, to form $\hat{B}$, becomes:

$$(2.17) \quad m = \hat{q}'\hat{B}^{-1}\hat{q} \xrightarrow{A} \chi_1^2$$

where $(1/T)B$ is our finite sample approximation to D, $B = \sigma^2[(\hat{x}'\hat{x})^{-1} - (x'x)^{-1}]$. Under H_1, the probability limit of q, $\bar{q} = \beta \cdot \sigma_{\varepsilon_2}^2/m_x^2$ so the asymptotic distribution of m for local departures depends on the magnitude of the two coefficients and the correlation of the right hand side variable with the disturbance. To compute the power as a function of β, equation (2.8) can be used. The IV estimates, $\hat{\beta}_{IV}$ and s_1^2, are consistent under both the null and alternative hypotheses.

A consistent estimate of m_x^2 follows from the data, and an estimate of $\sigma_{\varepsilon_2}^2$ is derived from the equation $\hat{\sigma}_{\varepsilon_2}^2 = (1 - \hat{\beta}_{OLS}/\hat{\beta}_{IV})\hat{m}_x^2$. Then an estimate of $\bar{q}$ may be calculated for any choice of β and the non-centrality parameter δ^2 is a quadratic function around $\beta = 0$, $\delta^2 = (\beta^2\sigma_{\varepsilon_2}^4/m_x^4(\hat{q}))$. Note that the asymptotic variance of the IV estimator enters the denominator as expected, so that IV estimates with large variance decrease the power of the test. The tables of the noncentral χ^2 test in Scheffé [**21**] can be consulted to find the probability of the null hypothesis being rejected for a given value of β if the alternative hypothesis is true conditional on the estimates of the incidental parameters of the problem. This type of IV (instrumental variable) test for errors in variables was first

proposed by Liviatan [12]. Wu [25] considers tests with different estimates of the nuisance parameter σ^2 to derive a finite sample F test under a stronger hypothesis about the stochastic properties of x.[9]

The IV test for errors in variables is known in the literature, but an alternative formulation of the test leads to easier implementation.[10] Partition the vector x into two orthogonal components, $x = \hat{x} + v$, which is the sum of the instrument and that part of x orthogonal to z. Then the least squares regression specification of equation (2.12) can be rewritten as

$$y = \beta x + \varepsilon_1 = \beta \hat{x} + \beta v + \varepsilon_1. \tag{2.18}$$

Now consider running this regression to compare the two estimates of β.

The variable $\hat{x}$ is asymptotically orthogonal to ε_1 under both the null and alternative hypothesis and is orthogonal to v by construction. Therefore the least squares regression coefficient of $\hat{x}$ is consistent under both hypotheses, being the IV estimate $\hat{\beta}_1$. The estimate of β corresponding to the variable v, however, should only have the same plim as $\hat{\beta}_1$ under the null hypothesis when v is orthogonal to ε_1. Thus, we might test whether the two estimates are equal. Since under the alternative hypothesis the plim of the second coefficient is no longer β, I will refer to it as γ and then rewrite equation (2.15) after adding and subtracting βv to make the test of equality easier:

$$y = \beta \hat{x} + \gamma v + \varepsilon_1 = \beta(\hat{x} + v) + (\gamma - \beta)v + \varepsilon_1 = \beta x + \alpha v + \varepsilon_1. \tag{2.19}$$

Thus for $\alpha = \gamma - \beta$, the proposed test is a large sample test on the hypothesis that $\alpha = 0$. One last minor simplification can be made by noting that an equivalent regression to equation (2.19) is

$$y = \beta x + \alpha \hat{x} + \varepsilon_1 \tag{2.20}$$

since $\hat{\alpha} = (v'Q_x v)^{-1} v'Q_x y = -(\hat{x}'Q_x\hat{x})^{-1}\hat{x}'Q_x y$ where $Q_x = I - x(x'x)^{-1}x'$. A test of $\alpha = 0$ from equation (2.20) under the null hypothesis is then based on the statistic $\sigma^2\chi^2 = \hat{\alpha}'(\hat{x}'Q_x\hat{x})\hat{\alpha}$. But $(1/\sigma^2)(\hat{x}'Q_x\hat{x})^{-1} = (\hat{x}'\hat{x})^{-1}B^{-1}(\hat{x}'\hat{x})^{-1}$ and $\hat{\alpha} = (\hat{x}'Q_x\hat{x})^{-1}(\hat{x}'\hat{x})\hat{q}$. Thus, this formulation is equivalent to the IV test of equation (2.17) since

$$\begin{aligned} \frac{1}{\sigma^2}\hat{\alpha}'(\hat{x}'Q_x\hat{x})\hat{\alpha} &= \frac{1}{\sigma^2}\hat{q}'(\hat{x}'\hat{x})(\hat{x}'Q_x\hat{x})^{-1}(\hat{x}'\hat{x})\hat{q} \\ &= \hat{q}'B^{-1}\hat{q}. \end{aligned} \tag{2.21}$$

[9] The instrumental variable test can also be considered a formalization and an improvement of a suggestion by Sargan [20] who recommended checking whether the least squares estimates lie outside the confidence regions of the IV estimates. For individual coefficients the procedure used here is to see whether the least squares estimate lies outside the confidence regions centered at the IV estimate and with length formed from the square root of the *difference* of the IV variance minus the OLS variance. Thus shorter confidence intervals follow from the current procedure than from Sargan's suggestion. The joint χ^2 test on all the coefficients in equation (2.14) if there are more than one, however, is the preferred test of the null hypothesis rather than separate consideration of each confidence interval.

[10] Presentation of this alternative method of testing has been improved from an earlier version of the paper using a suggestion of Z. Griliches.

A simple large sample normal test of $\alpha = 0$ based on the OLS estimate $\hat{\alpha}$ from equation (2.15) yields a test on whether errors in variables is present and is equivalent asymptotically to the test of equation (2.17) using s_0^2, the least squares estimate of σ^2, under the null hypothesis.[11] Besides ease of computation another advantage may be present. Three outcomes of the test will be encountered leading to simple approximate power interpretations which may not be as evident using the previous formulation of the test. First, $\hat{\alpha}$ may be large relative to its standard error. This result points to rejection of the hypothesis of no misspecification. The other clear cut case is a small $\hat{\alpha}$ with a small standard error which presents little evidence against H_0. The last result is a large standard error relative to the size of $\hat{\alpha}$. This finding indicates a lack of power which will be very evident to the user since he will not have a precise estimate of α.

Two immediate generalizations of the errors in variables specification test can be made. The test can be used to test any potential failure of assumption (1.1a) that additional right-hand-side variables are orthogonal to the error term so long as instrumental variables are available. First, additional right-hand-side variables can be present:

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon, \tag{2.22}$$

where the X_1 variables are possibly correlated with ε while the X_2 variables are known to be uncorrelated. Given a matrix of variables Z (which should include X_2), $\hat{q}$ will again be the difference between the IV estimator and the efficient OLS estimator. Letting $\hat{X}_1 = P_Z X_1$ where $P_z = Z(Z'Z)^{-1}Z'$ leads to the regression

$$y = X_1\beta_1 + X_2\beta_2 + \hat{X}_1\alpha + v \tag{2.23}$$

where a test of H_0: $\alpha = 0$ is a test for errors in variables.[12] The last orthogonality test involves a lagged endogenous variable which may be correlated with the disturbance. In this case, however, if the specification of the error process is known such as first order serial correlation, a more powerful test may be available.[13]

In this section the general nature of the misspecification problem has been discussed when there exists an alternative estimator which provides consistent estimates under misspecification. By demonstrating that the efficient estimator has zero asymptotic covariance with the difference between the consistent estimator $\hat{\beta}_1$ and the asymptotically efficient estimator (under H_0) $\hat{\beta}_0$, a simple expression for the variance of $(\hat{\beta}_0 - \hat{\beta}_1)$ test is found. Then by applying it to the

[11] Using s_0^2 to estimate σ^2 corresponds to the Lagrange multiplier form of the test while using s_1, the IV estimate, corresponds to using the Wald form of the test. The tests differ under the alternative hypothesis depending on the estimate of the nuisance parameter σ^2 which is used. Silvey [23] discusses the large sample relationship of the tests.

[12] For $V(\hat{q})$ to be nonsingular here, it is necessary that enough instruments be available to insure that $X_1 - \hat{X}_1$ has rank q.

[13] For the true regression problem (no lagged endogenous variables) under both the null hypothesis of no serial correlation and the alternative hypothesis $\hat{\beta}_0$, the OLS estimator, is unbiased and consistent since only assumption (1.1b) is violated. Therefore, if the null hypothesis of serial correlation is tested with an autoregressive estimator $\hat{\beta}_1$, plim $\hat{q} = \bar{q} = 0$ under *both* hypotheses. If $\hat{q}$ is large relative to its standard error, misspecification is likely to be present.

errors in variables problem, an easy method to apply the test is demonstrated which also makes power considerations clearer. The usefulness of this test is unfortunately decreased by the lack of a valid instrument in some situations. The next misspecification test, however, always can be done since the necessary data is available. It is a test of the random effects model which has been widely used in econometrics.

3. TIME SERIES-CROSS SECTION MODELS

Time series-cross section models have become increasingly important in econometrics. Many surveys, rather than being limited to a single cross section, now follow a panel of individuals over time. These surveys lead to a rich body of data given the wide variability between individuals coupled with much less variability for a given individual over time. Another important use of these models is to estimate demand across states over a period of time. Since for many goods (e.g., energy) considerable price variation exists across states while aggregate price indices move smoothly over time, time series-cross section models allow disentanglement of income and substitution effects which is often difficult to do with aggregate data.

The simplest time series-cross section model is specified as

$$y_{it} = X_{it}\beta + \mu_i + \varepsilon_{it} \qquad (i = 1, N;\ t = 1, T), \tag{3.1}$$

where μ_i is the individual effect. The two alternative specifications of the model differ in their treatment of the individual effect. The so-called fixed effects model treats μ_i as a fixed but unknown constant differing across individuals. Therefore, least squares on equation (3.1) is the correct estimator. To estimate the slope coefficients, deviation from means are used leading to the transformed observations $\tilde{y}_{it} = y_{it} - \bar{y}_i$, $\tilde{X}_{it} = X_{it} - \bar{X}_i$, $\tilde{\varepsilon}_{it} = \varepsilon_{it} - \bar{\varepsilon}_i$, and the regression specification,[14]

$$\tilde{y}_{it} = \tilde{X}_{it}\beta + \tilde{\varepsilon}_{it}. \tag{3.2}$$

An equivalent way of writing equation (3.2) is to let e be a T column vector of ones so that $e = (1, 1, \ldots, 1)'$ and to let $P_e = e(e'e)^{-1}e' = (1/T)ee' = (1/T)J_T$ with $Q_e = I \otimes (I - P_e)$. Then the fixed effects specification on the stacked model is

$$Q_e y = Q_e X\beta + Q_e\alpha + Q_e\varepsilon = \tilde{X}\beta + \tilde{\varepsilon} \tag{3.3}$$

which is identical to equation (3.2) since $Q_e\alpha = 0$.

The alternative specification for the time series-cross section model is known as the random effects or variance components model. Instead of treating μ_i as a fixed constant, this specification assumes that μ_i is drawn from an idd distribution, $\mu_i \sim N(0, \sigma_\mu^2)$, and is uncorrelated both with the ε_i and with the X_{it}. The specification then becomes

$$y_{it} = X_{it}\beta + \eta_{it}, \quad \eta_{it} = \mu_i + \varepsilon_{it}, \tag{3.4}$$

[14] Analysis of variance notation is being used, e.g., $\bar{y}_i = (1/T)\Sigma_{t=1}^{T} y_{it}$.

so that $E\eta = 0$ and the covariance matrix is block diagonal:

$$\Omega = V(\eta) = \begin{bmatrix} \sigma_\mu^2 J_T + \sigma_\varepsilon^2 I_T & & 0 \\ & \ddots & \\ 0 & & \sigma_\mu^2 J_T + \sigma_\varepsilon^2 I_T \end{bmatrix}. \tag{3.5}$$

Here the appropriate estimator is generalized least squares $\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$ which can be expressed in least squares form by transforming the variables by $\check{y}_{it} = y_{it} - \gamma\bar{y}_i$, $\check{X}_{it} = X_{it} - \gamma\bar{X}_i$ and then running ordinary least squares where

$$\gamma = 1 - \left(\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\mu^2}\right)^{\frac{1}{2}}.^{15} \tag{3.6}$$

Usually the variances, σ_μ^2 and σ_ε^2, are not known, so consistent estimates are derived from initial least squares estimates to form $\hat{\gamma}$ (see Wallace and Hussain [**24**]. This estimator is asymptotically efficient; and, if iterated to convergence, it yields the maximum likelihood estimates.

The choice of specification seems to rest on two considerations, one logical and the other statistical. The logical consideration is whether the μ_i can be considered random and drawn from an iid distribution. Both Scheffé [**21**] and Searle [**22**] contain excellent discussions of this question within an analysis of variance framework. Another way to consider the problem, suggested by Gary Chamberlain, is to decide whether the μ_i's satisfy di Finnetti's exchangeability criterion which is both necessary and sufficient for random sampling. Briefly, the criterion is to consider the sample $\mu = (\mu_1, \ldots, \mu_N)$ and to see whether we can exchange μ_i and μ_j (e.g., the constant for Rhode Island and California) while maintaining the same subjective distribution. If this logical criterion is satisfied, as it might well be for models of individuals like an earnings function, then the random effects specification seems appropriate. A statistical consideration is then to compare the bias and efficiency of the two estimators in estimating β, the slope coefficients.[16] Wallace and Hussain [**24**], Maddala [**13**], and Nerlove [**15**] have recently discussed this issue, all pointing out that the estimators become identical as T becomes large in the appropriate way as can be seen by the definition of γ in equation (3.6). Since the case in econometrics is usually that N is large relative to T, differences between the two estimators are an important problem.

[15] This method of estimating the random effects models seems to have gone unnoticed in the literature. It requires less computation than the usual GLS method or the matrix weighted average of two estimates.

[16] In other words, even if one decides that the random effects specification is appropriate on logical grounds, he may decide to use the fixed effects *estimator* which conditions on the particular sample of μ_i, thus treating them as fixed in the sample.

Under the random effects specification $\hat{\beta}_{GLS}$ is the asymptotically efficient estimator while the fixed effects estimator $\hat{\beta}_{FE}$ is unbiased and consistent but not efficient.[17] However, an important issue of specification arises which was pointed out by Maddala [**13**, p. 357] and has been further emphasized by Mundlak [**14**]. The specification issue is whether the conditional mean of the μ_i can be regarded as independent of the X_{it}'s, i.e., whether $E(\mu_i|X_{it}) = 0$.[18] If this assumption is violated, the random effects estimator is biased and inconsistent while the fixed effects estimator is not affected by this failure of orthogonality. Consider an individual earnings equation over time. If an unobserved variable, "spunk", affects education and has an additional effect on earnings, then the assumption of independent μ_i's will be violated. Thus, a natural test of the null hypothesis of independent μ_i's is to consider the difference between the two estimators, $\hat{q} = \hat{\beta}_{FE} - \hat{\beta}_{GLS}$. If no misspecification is present, then $\hat{q}$ should be near zero. Using the lemma, $V(\hat{q}) = V(\hat{\beta}_{FE}) - V(\hat{\beta}_{GLS})$ so a specification test follows from $m = \hat{q}'\hat{M}(\hat{q})^{-1}\hat{q}$ where $\hat{M}(\hat{q}) = (X'Q_eX)^{-1} - (X'\hat{\Omega}^{-1}X)^{-1}$. If the random effects specification is correct the two estimates should be near each other, rather than differing widely as has been reported sometimes in the literature as a virtue of the random effects specification. Therefore, while Maddala [**13**, p. 343] demonstrates that $\hat{\beta}_{GLS}$ is a matrix weighted average of $\hat{\beta}_{FE}$ (the within group estimator) and the between group estimator, if the specification is correct then plim $\hat{q} = 0$ so $\hat{\beta}_{GLS}$ and $\hat{\beta}_{FE}$ should be almost the same within sampling error. When the econometrician finds his estimates $\hat{\beta}_{FE}$ to be unsatisfactory, this evidence is a finding against his specification, not his choice of estimator. However, he should not necessarily accept the fixed effects estimates as correct but should reconsider the specification because errors in variables problems may be present which invalidate the fixed effects estimates.[19]

The equivalent test in the regression format is to test $\alpha = 0$ from doing least squares on

$$\check{y} = \check{X}\hat{\beta} + \tilde{X}\alpha + v \tag{3.7}$$

where $\check{y}$ and $\check{X}$ are the γ transformed random effects variables while $\tilde{X}$ are the deviations from means variables from the fixed effects specification. The tests can be shown to be equivalent using the methods of the previous section and the fact that $Q_e\check{y} = Q_e y$. This test is easy to perform since $\check{X}$ and $\tilde{X}$ differ only in the choice of γ from equation (4.6) while $\tilde{X}$ has $\gamma = 1$.

[17] A potentially important problem for the fixed effects estimator is its sensitivity to errors in variables. Since much variation is removed in forming deviations from individual means, the amount of inconsistency would be greater for the fixed effects estimates if errors in variables are present.

[18] If the regression specification of equation (3.1) is expanded to include a lagged endogenous variable, this variable is correlated by definition with the μ_i. Nerlove [**15**] discusses methods to estimate this specification. The test presented here would then be used to ascertain whether the μ_i are uncorrelated with the exogenous variables.

[19] Another possible test is to consider the difference of $\hat{\beta}_{FE}$, the within group estimator, from the between group estimator. Since the estimators are based on orthogonal projections, the variance of the difference equals the sum of the variances. However, this test seems less powerful than the test proposed here since our test statistic subtracts off the GLS variance from the fixed effects variance rather than adding on the between groups variance. The difference arises because our test uses the efficient estimator to form the comparison with the fixed effects estimator.

If $\hat{\gamma}$ is near unity the two estimators will give similar results and $\hat{q}$ will be near zero. It will often be the case in econometrics that $\hat{\gamma}$ will not be near unity. In many applications σ_μ^2 is small relative to σ_ε^2; and the problem sometimes arises that when σ_μ^2 is estimated from the data it may turn out to be negative. For a panel followed over time the X_{it} are often constant so that some of the parameters of interest will be absorbed into the individual constant when the fixed effects estimator is used. However, it seems preferable to have alternative estimates of the remaining slope coefficients to try to sort out possible interaction of the individual constants with the included right-hand-side variables. The misspecification test from equation (3.7) thus seems a desirable test of the random effects specification.[20]

In this section a test of the implicit assumption behind the random effects specification has been considered. This test should follow the logical specification of whether the μ_i are truly random. Thus, the situation is very similar to simultaneous equation estimation which follows the logical question of identification. In the next section, the specification of simultaneous equation systems is considered, and a test is developed for correct system specification.

4. SPECIFICATION OF SIMULTANEOUS EQUATION SYSTEMS

Most estimation associated with simultaneous equation models has used single equation, limited information estimators. Thus, two stage least sequares (2SLS) is by far the most widely used estimator. If a simultaneous equation system is estimated equation by equation, no check on the "internal consistency" of the entire specification is made. An important potential source of information on misspecification is thus neglected. This neglect is not total; one class of tests compares estimates of the unrestricted reduced form model with the derived reduced form estimates from the structural model as a test of the overidentifying restrictions.[21] Unfortunately, this type of test has not been widely used. Perhaps the reason has been the inconvenience of calculating the likelihood value or the nonlinear expansions which are required to perform the statistical comparison. In this section a test of system specification is proposed within a more simple framework. The test rests on a comparison of 2SLS to 3SLS estimates. Thus, the econometrician is comparing two different estimates of the structural parameters rather than the reduced form parameters. Usually, he has more knowledge about what comprises a "significant difference" with respect to the structural parameters. Under the null hypothesis of correct specification, 3SLS is efficient

[20] As previously mentioned, as T becomes large, γ in equation (3.6) approaches one and the two estimators approach each other. Thus both the numerator and denominator of the test statistic approach zero. The test appears to remain valid so long as γ does not exactly equal one and N increases faster than T; however, numerical problems of inverting a near singular matrix may arise.

[21] Within the single equation context this test has been proposed by Anderson and Rubin [1], Basmann [2], and Koopmans and Hood [11]. Within the full information context the likelihood ratio (LR) test has been used. Recently, Byron [4, 5] has simplified this test by advocating use of the Lagrange multiplier test or the Wald test both of which are asymptotically equivalent to the LR test under the null hypothesis. For further details see Silvey [23, Ch. 7].

but yields inconsistent estimates of all equations if any equation is misspecified. 2SLS is not as efficient as 3SLS, but only the incorrectly specified equation is inconsistently estimated if misspecification is present in the system. Thus, instead of comparing reduced form parameter estimates about which the econometrician often has little knowledge, the test compares estimates of the structural form parameters which he should have a better feeling for since they are derived from economic theory and are reported in estimates of other structural models.

Consider the standard linear simultaneous equation model

$$(4.1) \qquad YB + Z\Gamma = U$$

where Y is the $T \times M$ matrix of jointly dependent variables, Z is the $T \times K$ matrix of predetermined variables, and U is a $T \times M$ matrix of structural disturbances of the system. Full column rank of Z, nonsingularity of B, nonsingular probability limits of second order moment matrices, and the rank condition for identification are all assumed to hold. The structural disturbances are multivariate normal $U \sim N(0, \Sigma \otimes I_T)$. After a choice of normalization and imposition of zero restrictions each equation is written

$$(4.2) \qquad y_i = X_i\delta_i + U_i \quad \text{where} \quad X_i = [Y_i Z_i] \quad \text{and} \quad \delta_i = \begin{bmatrix} \beta_i \\ \gamma_i \end{bmatrix},$$

where β_i has r_i elements and γ_i has σ_i elements which correspond to the variables in X_i whose coefficients are not known a priori to be zero. It is convenient to stack the M equations into a system

$$(4.3) \qquad y = X\delta + U \quad \text{where}$$

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_M \end{bmatrix}, \quad X = \begin{bmatrix} X_1 & & 0 \\ & \ddots & \\ 0 & & X_M \end{bmatrix}, \quad \delta = \begin{bmatrix} \delta_1 \\ \vdots \\ \delta_M \end{bmatrix}, \quad U = \begin{bmatrix} U_1 \\ \vdots \\ U_M \end{bmatrix}.$$

The two stage least squares estimator when used on each equation of the system can conveniently be written in stacked form as $\hat{\delta}_2 = (X'\tilde{P}_Z X)^{-1} X'\tilde{P}_Z y$ where $\tilde{P}_Z = I_M \otimes Z(Z'Z)^{-1}Z'$. To simplify notation rewrite the estimator as $\hat{\delta}_2 = (\hat{X}'\hat{X})^{-1}\hat{X}'y$. Three stage least squares uses full information and links together all equations of the system through the estimate of the covariance matrix $\hat{\Sigma}$. Letting $\hat{P}_{\Sigma Z} = \hat{\Sigma}^{-1} \otimes Z(Z'Z)^{-1}Z'$, the 3SLS estimator is $\hat{\delta}_3 = (X'\hat{P}_{\Sigma Z}X)^{-1}X'\hat{P}_{\Sigma Z}y$ which is simplified to $\hat{\delta}_3 = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'y$.[22] Now 3SLS transmits misspecificiation throughout the entire system, affecting the estimates of all coefficients since $\hat{\delta}_3 - \delta = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'U$. Thus, if the jth equation is misspecified plim $(1/T)\hat{X}_j'U_j \neq 0$, and so assuming probability limits exist with $\hat{\Sigma}$

[22] If $T \leq K$ so 2SLS and 3SLS cannot be used, asymptotically equivalent instrumental variable estimators are discussed in Brundy and Jorgenson [3], Dhrymes [7], and Hausman [10]. Thus, the current misspecification test can be applied when the full information likelihood ratio test is not possible because unrestricted estimates of the reduced form cannot be made due to sample size limitations.

being the probability limit of the inconsistent estimate of Σ with $\tilde{\sigma}^{ij}$ the element of its inverse, the inconsistency is calculated from plim $(\hat{\delta}_3 - \hat{\delta}) = \text{plim}\ ((1/T)\tilde{X}'\tilde{X})^{-1} \cdot \text{plim}\ ((1/T)\tilde{X}'U)$. Looking at the crucial last term more closely, consider the unknown elements from the first equation δ_1. The last term takes the form

$$(4.4) \qquad \text{plim}\ \frac{1}{T} \sum_{m=1}^{M} \tilde{\sigma}^{1m} \hat{X}_1' U_m$$

so that the amount of inconsistency for the first equation due to misspecification in the jth equation depends both on the lack of orthogonality between $\hat{X}_1$ and U_j, and also on the size of $\tilde{\sigma}^{1j}$.

Lemma 2.1 leads us to consider the specification test based on the difference between the two estimators $\sqrt{T}\hat{q} = \sqrt{T}(\hat{\delta}_2 - \hat{\delta}_3)$ which has large sample variance $V(\hat{q}) = V(\hat{\delta}_2) - V(\hat{\delta}_3)$. However, an alternative procedure is to consider the regression on the stacked system

$$(4.5) \qquad y = \hat{X}\tilde{\delta} + \tilde{X}\alpha + V$$

and to test whether $\alpha = 0$. Since $\hat{X}$ and $\tilde{X}$ are computed by programs which have 2SLS and 3SLS estimators, the regression of equation (5.5) should not be difficult to perform.

The noncentrality parameter of the local noncentral χ^2 distribution will be proportional to plim $(1/T)\hat{X}_i'U_j$ for any equation which is misspecified and also the magnitude of the covariance elements $\tilde{\sigma}^{ij}$. If the inverse covariance elements are large, then $\hat{X}$ and $\tilde{X}$ will not be highly correlated so that the test will be powerful for a given size of inconsistency. As the $\tilde{\sigma}^{ij}$'s go to zero, then 3SLS approaches 2SLS and the test will have little power. Since the misspecification represented by the alternative hypothesis is not specific, the appropriate action to take in the case of rejection of H_0 is not clear. One only knows that misspecification is present somewhere in the system. If one is confident that one or more equations are correctly specified, then the specification of other equations could be checked by using them, say one at a time, to form a 3SLS type estimator. That is, if equation 1 is correct and equation 2 is to be tested, then 2SLS on equation 1 could be compared to 3SLS on equation 1 where $\hat{\sigma}_{ij}$ is set to zero for $i \neq j$ except for $i = 1, j = 2$ and vice-versa in the 3SLS estimator. Using this method the misspecification might be isolated; but, unfortunately, the size of the test is too complicated to calculate when done on a sequence of equations.[23]

The simultaneous equations specification test is the last to be presented although the same principle may be applied to further cases such as aggregation. I now turn to an empirical example of the specification test to demonstrate its potential usefulness.

[23] If one attempts to check the specification of the entire system by comparing the 2SLS and 3SLS estimates, the χ^2 test of Theorem 2.1 is appropriate under H_0. However, under H_1 the non-central χ^2 distribution is no longer appropriate since the 2SLS estimates are also inconsistent.

5. EMPIRICAL EXAMPLE

Comparing two alternative estimators as a means of constructing misspecification tests has been applied to a number of situations in the preceding sections. In this section an empirical example is presented. The example is the time series-cross section specification test discussion in Section 4. This type of data set is becoming increasingly common for econometric studies such as individuals' earnings, education, and labor supply. However, added interest in this test comes from the fact that it also implicitly tests much cross section analysis of similar specifications. Cross section analysis can allow for no individual constant but must assume, as does random effect analysis, that the right hand side variables are orthogonal to the residual: If the random effect specification is rejected serious doubt may be cast therefore on much similar cross section analysis.

For the time-series-cross-section specification test a wage equation is estimated for male high school graduates in the Michigan income dynamics study.[24] The sample consists of 629 individuals for whom all six years of observations are present. A wage equation has been chosen due to its importance in "human capital" analysis. The specification used follows from equation (3.1). The right hand side variables include a piecewise linear representation of age, the presence of unemployment or poor health in the previous year, and dummy variables for self-employment, living in the South, or in a rural area. The fixed effects estimates, $\hat{\beta}_{FE}$, are calculated from equation (3.3). They include an individual constant for each person and are consistent under both the null hypothesis of no misspecification and the alternative hypothesis. The random effects estimates, $\hat{\beta}_{GLS}$, are calculated from equations (3.4)–(3.6). The estimate of $\hat{\gamma}$ from equation (3.6) is .72736 which follows from least squares estimates of the individual variance $\hat{\sigma}_{\mu}^2 = .12594$ and the residual variance $\hat{\sigma}_{\varepsilon}^2 = .06068$. Under the null hypothesis the GLS estimate is asymptotically efficient, but under the alternative hypothesis it is inconsistent. The specification test consists of seeing how large the difference in estimates is, $\hat{q} = \hat{\beta}_{FE} - \hat{\beta}_{GLS}$, in relation to its variance $M(\hat{q}) = M(\hat{\beta}_{FE}) - M(\hat{\beta}_{GLS})$ which follows from Lemma (2.1). In comparing the estimates in column 1 and column 2 of Table I it is apparent that substantial differences are present in the two sets of estimates relative to their standard errors which are presented in column 3.[25] The effects of unemployment, self-employment, and geographical location differ widely in the two models. The geographical differences may be explained by the implicitly different way that migration is handled in the two specifications since the fixed effects coefficient specification coefficients only represent changes during the sample period. Unobserved individual characteristics might well be correlated with geographical location. Also, the effect of unemployment in the previous year is seen to be much less

[24] The specification used is based on research by Gordon [9] who also kindly helped me construct this example.

[25] Note that the elements of $\hat{q}$ and its standard errors are simply calculated given the estimates of $\hat{\beta}_{FE}$ and $\hat{\beta}_{GLS}$ and their standard errors, making sure to adjust to use the fixed effects estimate of σ_{ε}^2. The main computational burden involves forming and inverting $M(\hat{q})$.

TABLE I

DEPENDENT VARIABLE—LOG WAGE[a]

Variable	Fixed Effects	Random Effects	$\hat{q}$	$\hat{\epsilon}$
1. Age 1 (20–35)	.0557	.0393	.0164	.0291
	(.0042)	(.0033)	(.0030)	(.0060)
2. Age 2 (35–45)	.0351	.0092	.0259	.0015
	(.0051)	(.0036)	(.0039)	(.0070)
3. Age 3 (45–55)	.0209	−.0007	.0216	.0058
	(.0055)	(.0042)	(.0040)	(.0083)
4. Age 4 (55–65)	.0209	−.0097	.0306	−.0308
	(.0078)	(.0060)	(.0050)	(.0112)
5. Age 5 (65–)	−.0171	−.0423	.0252	−.0380
	(.0155)	(.0121)	(.0110)	(.0199)
6. Unemployed$_{-1}$	−.0042	−.0277	.0235	−.3290
	(.0153)	(.0151)	(.0069)	(.0914)
7. Poor Health$_{-1}$	−.0204	−.0250	.0046	−.1716
	(.0221)	(.0215)	(.0105)	(.0762)
8. Self-Employment	−.2190	−.2670	.0480	−.3110
	(.0297)	(.0263)	(.0178)	(.0558)
9. South	−.1569	−.0324	−.1245	.0001
	(.0656)	(.0333)	(.0583)	(.0382)
10. Rural	−.0101	−.1215	.1114	−.2531
	(.0317)	(.0237)	(.0234)	(.0352)
11. Constant	—	.8499	—	—
	—	(.0433)	—	—
s^2	.0567	.0694		.0669
degrees of freedom	3135	3763		3753

[a] 3774 observations. Standard errors are in parentheses.

important in affecting the wage in the fixed effects specification. Thus, unemployment has a more limited and transitory effect once permanent individual differences are accounted for. The test of misspecification which follows from Lemma 2.1 is

$$(5.1) \qquad m = \hat{q}'\hat{M}(\hat{q})^{-1}\hat{q} = 129.9.$$

Since m is distributed asymptotically as χ^2_{10} which has a critical value of 23.2 at the 1 per cent level, very strong evidence of misspecification in the random effects model is present. The right hand side variables X_{it} are not orthogonal to the individual constant μ_i so that the null hypothesis is decisively rejected. Considerable doubt about previous cross section work on wage equations may arise from this example.

The reason for this doubt about previous cross section estimation is that ordinary least squares on a cross section of one year will have the same expectation as $\hat{\beta}_{GLS}$, the random effects estimate, on the time series-cross section data. For example, cross section estimates of the wage equation have no individual constants and make assumption (1.1a) that the residual is uncorrelated with the right-hand-side variables. However, this example demonstrates that in the Michigan survey important individual effects are present which are *not* uncorrelated with the right-hand variables. Since the random effects estimates seem

significantly biased with high probability, it may well be important to take account of permanent unobserved differences across individuals. This problem can only be resolved within a time series-cross section framework using a specification which allows testing of an important maintained hypothesis of much cross section estimation in econometrics. Thus, the importance of this type of data is emphasized which permits us to test previously maintained hypotheses.

An equivalent formulation of the specification test is provided by the regression framework of equation (3.7). Instead of having to manipulate 10×10 matrices, $\tilde{y}$ is regressed on both $\tilde{X}$ and $\tilde{\tilde{X}}$. The test of the null hypothesis is then whether $\hat{\alpha} = 0$. As is apparent from column 4 of Table I many of the elements of $\hat{\alpha}$ are well over twice their standard error so that misspecification is clearly present. The misspecification test follows easily from comparing s^2, the estimated variance from the random effects specification, to s^2 from the augmented specification

$$(5.2) \qquad m = \frac{.06938 - .06689}{.06689} \cdot 3754 = 139.7.$$

Again m well exceeds the approximate critical χ^2 value of 23.2. Since this form of the test is so easy to implement when using a random effects specification as only one additional weighted least squares regression is required, hopefully applied econometricians will find it a useful device for testing specification.

The empirical example presented in this section illustrates use of the misspecification test. The example rejects an application of the random effects specification. I feel that this finding may well be quite general, and that the uncorrelated random effects model is not well suited to many econometric applications. The two requirements of exchangeability and orthogonality are not likely to be met in many applied problems. Certainly, the random effects estimates should be compared with the fixed effects estimates to see if significant differences occur. If they do, the specification of the equation should be reconsidered to either explain the difference or to try a different specification which corrects the problem.

6. EXTENSIONS AND CONCLUSION

Another possible application of the methodology presented here arises when one wants to test whether only a limited part of a model specification differs. For instance, consider two different model specifications, where the difference arises because the second specification has additional parameters which are restricted in the first specification, e.g., sample selection specifications. One could do maximum likelihood on each specification and then perform a likelihood ratio test thus comparing the different specifications. However, if interest of the model centers on a particular parameter which is unrestricted in both specifications, the traditional methodology yields no way to test for a significant difference only in that parameter. Lemma 2.1 applies so this paper provides a simple method of testing the hypothesis of a significant difference in that particular parameter

since the unrestricted model is inefficient under the null hypothesis while it is consistent under both the null and alternative hypotheses.

By using the result that under the null hypothesis of no misspecification, an asymptotically efficient estimator must have zero covariance with its difference from a consistent, but asymptotically inefficient estimator, specification tests are devised from a number of important model specifications in econometrics. New tests for the cross section-time series model and for the simultaneous equation model are presented. Lastly, an empirical example is provided. The example provides strong evidence that unobserved individual factors are present which are not orthogonal to included right-hand-side variables in a common econometric specification.

Massachusetts Institute of Technology

Manuscript received August, 1976; final revision received April, 1978.

REFERENCES

[1] ANDERSON, T. W., AND H. RUBIN: "Estimation of the Parameters of a Single Stochastic Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 20 (1949), 46–63.

[2] BASMANN, R. L.: "A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation," *Econometrica*, 25 (1957), 77–83.

[3] BRUNDY J., AND D. W. JORGENSON: "Efficient Estimation of Simultaneous Equation Systems by Instrumental Variables," *Review of Economics and Statistics*, 53 (1971), 207–224.

[4] BYRON, R. P.: "Testing for Misspecification in Econometric Systems Using Full Information," *International Economic Review*, 13 (1972), 745–756.

[5] ———: "Testing Structural Specification Using the Unrestricted Reduced Form," *Econometrica*, 42 (1974), 869–883.

[6] COX, D. R., AND D. V. HINKLEY: *Theoretical Statistics*. London: Chapman and Hall, 1974.

[7] DHRYMES, P. J.: "A Simplified Structural Estimator for Large-Scale Econometric Models," *Australian Journal of Statistics*, 13 (1971), 168–175.

[8] FISHER, R. A.: "Theory of Statistical Estimation," *Cambridge Philosophical Society Proceedings*, 22 (1925), 700–725.

[9] GORDON, R.: "Essays on the Causes and Equitable Treatment of Differences in Earnings and Ability," Massachusetts Institute of Technology Ph.D. Thesis, June, 1976.

[10] HAUSMAN, J.: "An Instrumental Variable Approach to Full-Information Estimators for Linear and Certain Nonlinear Econometric Models," *Econometrica*, 43 (1975), 727–738.

[11] KOOPMANS, T. C., AND W. HOOD: "The Estimation of Simultaneous Economic Relationships," in *Studies in Econometric Method*, ed. by W. Hood and T. C. Koopmans. New Haven: Yale University Press, 1953, pp. 113–199.

[12] LIVIATAN, N.: "Tests of the Permanent Income Hypothesis Based on a Re-interview Savings Study," in *Measurement in Economics*, ed. by C. Christ. Stanford: Stanford University Press, 1963, pp. 29–59.

[13] MADDALA, G. S.: "The Use of Variance Components Models in Pooling Cross Section and Time Series Data," *Econometrica*, 39 (1971), 341–358.

[14] MUNDLAK, Y.: "On the Pooling of Time Series and Cross Section Data," 1976, mimeo.

[15] NERLOVE, M.: "A Note on Error Component Models," *Econometrica*, 39 (1971), 383–396.

[16] RAMSEY, J. B.: "Classical Model Selection through Specification Error Tests," in *Frontiers of Econometrics*, ed. by P. Zarembka. New York: Academic Press, 1974.

[17] RAE, C. R.: *Linear Statistical Inference*. New York: Wiley, 1973.

[18] ROTHENBERG, T. J.: *Efficient Estimation With A Priori Information*. New Haven: Yale University Press, 1973.

[19] ROTHENBERG, T. J.: "The Asymptotic Distribution of the Least Squares Estimator in the Errors in Variables Model," mimeo, 1972.
[20] SARGAN, J. D.: "The Estimation of Economic Relationships Using Instrumental Variables," *Econometrica*, 26 (1958), 393–415.
[21] SCHEFFÉ, H.: *Analysis of Variance*. New York: Wiley, 1959.
[22] SEARLE, P.: *Linear Models*. New York: Wiley, 1971.
[23] SILVEY, S. D.: *Statistical Inference*. Harmondsworth: Penguin Press, 1970.
[24] WALLACE, T. D., AND A. HUSSAIN: "The Use of Error Components Models in Combining Cross Section with Time Series Data," *Econometrica*, 37 (1969), 57–72.
[25] WU, D.: "Alternative Tests of Independence Between Stochastic Regressors and Disturbances," *Econometrica*, 41 (1973), 733–750.