SPECIFICATION TESTING IN PANEL DATA WITH INSTRUMENTAL VARIABLES

Gilbert E. Metcalf
Princeton University

Econometric Research Program
Research Memorandum No. 358

January 1991

JEL Classification: 210,211

My thanks to Whitney Newey and Gregory Chow for helpful advice. Any errors, of course, remain my own.

Econometric Research Program
Princeton University
203 Fisher Hall
Princeton, NJ 08544-1021, U.S.A.

Abstract

This paper shows a convenient way to test whether instrumental variables are correlated with individual effects in a panel data set. It shows that the correlated fixed effects specification tests developed by Hausman and Taylor (1981) extend in an analogous way to panel data sets with endogenous right hand side variables. In the panel data context, there are different sets of instrumental variables that can be used to construct the test. Asymptotically, we show that the test typically is more efficient if an incomplete set of instruments is used. However, in small samples one is likely to do better using the complete set of instruments. Simulation results demonstrate the likely gains for different assumptions about the degree of variance in the data across observations relative to across time.

KEYWORDS: Econometrics, Panel Data, Instrumental Variables, Specification Testing.

I. Introduction

The use of panel data sets has increased dramatically since the pioneering research of Mundlak (1961), Nerlove (1971) and Maddala (1971), among others. An important benefit of pooled cross section time series data is the possibility of controlling for unobservable individual specific effects. If these unobserved variables are correlated with right hand side variables in the regression, ordinary least squares (OLS) estimates of the coefficients will be biased and inconsistent. In the presence of correlated individual effects, first difference or fixed effects (within) estimators yield consistent estimates of the regression parameters. However consistency comes at a cost: the between groups information is ignored which may substantially reduce the efficiency of the estimates. As a result, a good deal of research has been undertaken to derive tests to detect this possible correlation (e.g. Hausman (1978), Hausman and Taylor (1981), Chamberlain (1983), and Holtz-Eakin (1988)).

However, none of the tests allow for the possibility that some of the right hand side variables are correlated with the random error (aside from the individual effect). This is perhaps not surprising. Little work has been done on estimation of panel data models in a simultaneous system. Schmidt and Wyhowski (1988) review the limited literature and provide results which extend the results from the single equation literature in a limited information context (2SLS) and a full information context (3SLS). They do not discuss the issue of specification testing in the context of instrumental variable estimation.

Below, I extend the results of Hausman (1978) and Hausman and Taylor (1981) to the case where right hand side variables are assumed to be endogenous (specifically, correlated with the time varying component of the error structure). It turns out that the IV analogous specification tests for correlated fixed effects given in Hausman and Taylor (1981) are applicable in this context. However, it is important to specify the instrument set appropriately for the specification test. I then consider the small sample properties of the test statistic under different assumptions about the quality of the instrument and the degree of correlation of the fixed effects and the instrument. I show that asymptotically the quality of the instrumental variable (in the sense of its correlation with the endogenous right hand side variables) does not affect the power of the test. The power is a function of the degree to which the null hypothesis is violated and the information contained in the instrument (its variance). However, the small sample properties of the test will depend on the quality of the instrument. This will be true even in relatively large data sets. Perhaps surprisingly, the more powerful test statistic uses an inefficient estimator. Asymptotically, while the variance used to construct the test statistic will be greater than the variance associated with using the efficient estimator, its asymptotic bias will also be greater as the null hypothesis of no correlation is violated. The increase in bias more than offsets the increase in variance thereby leading to a more powerful test statistic.

The degree to which the test statistic using an inefficient estimator is an improvement over the statistic using the efficient estimator depends on the relative amounts of the variance of the explanatory variables and the instruments which is due to variation across individuals versus across time (the "between" versus the "within" variation). If the ratio of the variance components is the same for the explanatory variables and the instruments, then the two test statistics are equally powerful. However, in small samples the test statistic using the more efficient estimator performs better as we show below.

The next section shows that the test statistic as suggested by Hausman and Taylor (1981) carries over to the 2SLS case. We discuss the appropriate construction of the instrument set given various assumptions about the type of correlation between the instruments and the individual effects The following section presents results from a simple Monte Carlo experiment. Finally there is a brief conclusion which considers other issues specific to specification testing and estimation in a limited information context.

II. The Model and Test

The model under consideration is

$$Y = X\beta + \alpha \otimes e_T + \epsilon \tag{1}$$

where Y is an NT x 1 vector, X an NT x k matrix, α an N x 1 vector of individual effects (α_i iid with mean 0 and variance σ^2_α) and ϵ an iid random vector with mean 0 and covariance matrix $\sigma^2_\epsilon I_{NT}$. The vector e_T is a T x 1 vector of ones. The data are stacked by individuals over time. That is, $Y' = [Y'_1 \; Y'_2 \; \ldots \; Y'_N]$ where Y_i is a T x 1 vector of observations on the i^{th} individual. This equation is part of a simultaneous system and by assumption some columns of X are correlated with ϵ. It is assumed that some (possibly all) columns of X are also correlated with the individual effects. There is a set of instruments Z, a matrix NT x L, $L \geq k$, valid in the sense that Z is correlated with X but uncorrelated with ϵ. It is assumed that columns of X which are uncorrelated with ϵ are contained in Z. The present purpose is to test whether Z is correlated with the fixed effects. In the literature, the correlation has been assumed to be of two types and we construct two sets of hypotheses to test. Below we discuss the assumptions underlying the two sets of hypotheses and their relationship. Specifically, we test:

$$H_1 \quad \begin{array}{l} H_o: \underset{N\to\infty}{\text{Plim}}\{\sum_{i=1}^{N} Z'_{i.}\alpha_i/N\} = \\ H_A: \underset{N\to\infty}{\text{Plim}}\{\sum_{i=1}^{N} Z'_{i.}\alpha_i/N\} \neq 0 \end{array}$$

or

$$H_2 \quad \begin{array}{l} H_o: \operatorname{Plim}_{N\to\infty}\{\sum_{i=1}^{N} Z'_{it}\alpha_i/N\} \neq 0 \\ H_A: \operatorname{Plim}_{N\to\infty}\{\sum_{i=1}^{N} Z'_{it}\alpha_i/N\} \neq 0 \end{array} \quad \forall\, t$$

holding T fixed, where $Z_{i.}$ is the average over time of the observations of Z_{it}.[1] Given the loss of information resulting from the use of the within or first difference estimators to eliminate correlated fixed effects, there is a large gain possible if we can assume the null hypothesis. In this case, the GLS-IV estimator will be an improvement.

Letting $u = \alpha \otimes e_T + \epsilon$, then

(2) $$E(uu') \equiv \Omega = T\sigma_\alpha^2 P_v + \sigma_\epsilon^2 I_{NT}$$

or

(2') $$E(uu') = \sigma_1^2 P_v + \sigma_\epsilon^2 Q_v$$

where $\sigma_1^2 = T\sigma_\alpha^2 + \sigma_\epsilon^2$, $P_v = (I_N \otimes e_T e_T')/T$ and $Q_v = I - P_v$. For future reference, we use the fact that $\Omega^{-1/2} = \sigma_1^{-1} P_v + \sigma_\epsilon^{-1} Q_v$ and denote $\Omega^{-1/2}$ by H. $P_v X$ replaces the observations for each column of X by the average of the observations for each individual over time. $Q_v X$ replaces the observations by the deviations from the time averages.

We wish to show that the Hausman type specification test comparing the GLS-IV estimator with the fixed effects (within) estimator can be constructed using the within and the between estimators. The essential step is to show that the GLS-IV estimator of β can be written as a matrix weighted average of within and between estimators. Define the operator A^π as the projection operator: $A^\pi = A(A'A)^{-1}A'$. $A^\pi X$ is the projection of the columns of X onto the space spanned by the columns of A. Given the panel nature of the data, the GLS estimator of β accounts for the non-spherical error structure of u. If Z

---

[1] In Schmidt and Wyhowski's (1988) terminology, Z is a set of doubly exogenous variables under the null hypothesis and singly exogenous under the alternative.

is a set of variables uncorrelated with ϵ, there are different possible instrument sets that we can use. Following the general approach of Schmidt and Wyhowski, we consider instrument sets of the form $\tilde{Z} = [Q_v Z, P_v B]$ where B is defined according to the appropriate null hypothesis. The GLS-IV estimator is given by

(3) $$\hat{\beta}^{IV}_{GLS} = (X' H' \tilde{Z}^{\pi} HX)^{-1} X' H' \tilde{Z}^{\pi} HY.$$

It can easily be shown that $H' \tilde{Z}^{\pi} H = \sigma_{\epsilon}^{-2}(Q_v Z)^{\pi} + \sigma_1^{-2}(P_v B)^{\pi}$ and so

(4) $$\hat{\beta}^{IV}_{GLS} = \left[\sigma_{\epsilon}^{-2} X' (Q_v Z)^{\pi} X + \sigma_1^{-2} X' (P_v B)^{\pi} X\right]^{-1} \cdot \left[\sigma_{\epsilon}^{-2} X' (Q_v Z)^{\pi} Y + \sigma_1^{-2} X' (P_v B)^{\pi} Y\right].$$

If we define the Nxk matrix $\bar{X} = [\bar{X}'_{1.} \bar{X}'_{2.} \ldots \bar{X}'_{N.}]'$ where $\bar{X}_{i.}$ is the mean of the T observations on X for the i^{th} individual (and similarly for B, Y, and Z), we can rewrite (4) using the fact that $P_v B = \bar{B} \otimes e_T$. Making this substitution and some simple algebra leads to

(4') $$\hat{\beta}^{IV}_{GLS} = \left[\sigma_{\epsilon}^{-2} X' (Q_v Z)^{\pi} X + T\sigma_1^{-2} \bar{X}' \bar{B}^{\pi} \bar{X}\right]^{-1} \cdot \left[\sigma_{\epsilon}^{-2} X' (Q_v Z)^{\pi} Y + T\sigma_1^{-2} \bar{X}' \bar{B}^{\pi} \bar{Y}\right].$$

Now the within estimator $(\hat{\beta}^{IV}_{W})$ follows first from eliminating α from equation 1:

(5) $$Q_v Y = Q_v X\beta + Q_v \epsilon$$

We premultiply this by $\tilde{Z}^{\pi}$ and, noting that $\tilde{Z}^{\pi} = (Q_v Z)^{\pi} + (P_v B)^{\pi}$, obtain

(6) $$(Q_v Z)^{\pi} Y = (Q_v Z)^{\pi} X\beta + (Q_v Z)^{\pi} \epsilon$$

and therefore

(7) $$\hat{\beta}^{IV}_{W} = \left[X' (Q_v Z)^{\pi} X\right]^{-1} \left[X' (Q_v Z)^{\pi} Y\right].$$

This is simply the 2SLS regression of $Q_v Y$ on $Q_v X$ using $Q_v Z$ as instruments. The between estimator $(\hat{\beta}^{IV}_{B})$ is similarly derived and is given by

(8) $$\hat{\beta}^{IV}_{B} = \left[\bar{X}' \bar{B}^{\pi} \bar{X}\right]^{-1} \left[\bar{X}' \bar{B}^{\pi} \bar{Y}\right].$$

Again, this is simply the 2SLS regression of $\bar{Y}$ on $\bar{X}$ using $\bar{B}$ as instruments.

Defining

(9) $$\Lambda = \left[\sigma_\epsilon^{-2}X'(Q_vZ)^{\pi}X + T\sigma_1^{-2}\bar{X}'\bar{B}^{\pi}\bar{X}\right]^{-1}\sigma_\epsilon^{-2}X'(Q_vZ)^{\pi}X,$$

then it follows immediately that

(10) $$\hat{\beta}_{GLS}^{IV} = \Lambda\,\hat{\beta}_W^{IV} + (I - \Lambda)\,\hat{\beta}_B^{IV}\ .$$

Equation (10) shows that the GLS-IV estimator can be written as a matrix weighted average of the within IV and the between IV estimators and is the IV analogy to the result for OLS estimators presented in Maddala (1971).

Under the null hypothesis, $\hat{\beta}_{GLS}^{IV}$ and $\hat{\beta}_W^{IV}$ are consistent estimators of β with $\hat{\beta}_{GLS}^{IV}$ the more efficient estimator while under either alternative from H_1 or H_2, $\hat{\beta}_W^{IV}$ is consistent and $\hat{\beta}_{GLS}^{IV}$ is inconsistent. A Hausman test statistic of the form

(11) $$q_1 = (\hat{\beta}_{GLS}^{IV} - \hat{\beta}_W^{IV})'\,V(\hat{\beta}_{GLS}^{IV} - \hat{\beta}_W^{IV})^{-1}(\hat{\beta}_{GLS}^{IV} - \hat{\beta}_W^{IV})$$

can be constructed. Under the null, q_1 is distributed as a chi-square statistic with k degrees of freedom. Simple algebra using equation (10) shows that q_1 is equal to to q_2 where

(12) $$q_2 = (\hat{\beta}_W^{IV} - \hat{\beta}_B^{IV})'(V_W + V_B)^{-1}(\hat{\beta}_W^{IV} - \hat{\beta}_B^{IV}).$$

And we are done. One advantage of q_2 over q_1 is that the covariance matrix of the difference between the between and within estimators is easier to compute. While the $\mathrm{Cov}(\hat{\beta}_{GLS}^{IV} - \hat{\beta}_W^{IV})$ is equal to $V_W - V_{GLS}$ if $\hat{\beta}_{GLS}^{IV}$ is asymptotically efficient, the estimated difference of the covariance matrices may not be positive definite in small samples. Moreover, the method used in Hausman (1978) to show that the covariance of the differences equals the differences of the covariances cannot necessarily be used as there is no guarantee that the GLS-IV estimator is asymptotically efficient[2]. Besides being easier to

---

[2] Moreover, Amemiya and MaCurdy (1986) (hereafter A-M) show that the Hausman-Taylor GLS estimator is not asymptotically efficient for a particular class of instrument sets considered in the A-M paper.

compute than q_1, q_2 has the advantage that the covariance matrix of $\hat{\beta}_W^{IV} - \hat{\beta}_B^{IV}$ is guaranteed to be positive definite in small samples.

The result on the equality of the two specification tests is quite general given instrument sets of the form $[Q_vZ, P_vB]$. Under the null hypothesis that the means of the instruments are uncorrelated with the individual effects (H_1), an obvious choice for B is Z itself. Then $\tilde{Z} = [Q_vZ, P_vZ]$. In other words, the instruments Z are used twice: first as deviations from their time means and then as the time means themselves. This is essentially the Hausman-Taylor (HT) estimator discussed in Breusch, Mizon, and Schmidt (1989). With no additional assumptions, Schmidt and Wyhowski (1988) show that the estimator in equation (3) is the GLS-IV estimator and additionally can be derived as a generalized method of moments type estimator.

With the additional assumptions embodied in H_2, a more efficient estimator is available. If the values of Z_{it} are uncorrelated with the individual effects for each t, then each of the T NxL matrices Z_t, where $Z_t = [Z'_{1t}, \ldots, Z'_{Nt}]'$, can be used as instruments for X_t. As a result, more instruments are available which cannot decrease the efficiency of the GLS-IV estimator. The estimator in equation 3 only requires that the <u>means</u> of Z be uncorrelated with the individual effects under the null hypothesis. Under the null hypothesis that individual values of Z are uncorrelated with the individual effects for all values of t, then the instrument set $B = Z^*$ provides more efficient estimates of β where

$$Z^* = [Z_1', Z_2' \ldots Z_N']'$$

is an N x TL matrix formed by combining the Z_is. Note that $Q_vZ^* = 0$ and $P_vZ^* = Z^*$. Hence, $\tilde{Z} = [Q_vZ, Z^*]$.

Amemiya and MaCurdy (1986) discuss the implications of the first null hypothesis in H_1 for the null hypothesis in H_2 described above. Strictly speaking, the first assumption is a weaker assumption and explains why the GLS-IV estimator using Z^* rather than P_vZ is a more efficient estimator.

However, they note the following fact. In addition to the assumptions underlying H_1, impose the assumption that our estimator for β continue to be consistent if we carry out the estimation using any T-1 of the T time periods. That is, assume that $\text{Plim}_{N\to\infty}(\sum_{i=1}^{N} Z_{i.}^{s\prime}\alpha_i/N) = 0$, for s = 1,...,T where $Z_{i.}^{s} = \frac{1}{T-1}\sum_{t\neq s} Z_{it}$. This set of T assumptions along with the assumption that $\text{Plim}_{N\to\infty}(\sum_{i=1}^{N} Z'_{i.}\alpha_i/N) = 0$ implies the set of T assumptions in the second null hypothesis. While there may be circumstances in which this second set of T-1 assumptions fails to hold while the assumption under the null hypothesis in H_1 holds, it seems more plausible to believe that this is an unusual case. Therefore from hereon we assume that the assumption underlying H_2 are relevant when considering the null hypothesis. We assume that Z^* is the appropriate set of instruments if the null hypothesis is correct.

This suggests that the appropriate specification test should employ the more efficient GLS estimator to obtain greater power. However, this will not turn out to be the case. The next section considers the asymptotic efficiency for a particular class of data processes to illustrate the issue.

III. Asymptotic Efficiency

We consider data of the following form for the model described in equation 1:

(13a) $$X_{it} = \gamma_1 X_{it-1} + \nu_{it}$$

(13b) $$Z_{it} = \gamma_2 Z_{it-1} + \eta_{it}$$

(13c) $$\text{plim}_{N\to\infty}(\frac{1}{NT}\nu'\eta) = \Sigma_{\nu\eta}$$

(13d) $$\text{plim}_{N\to\infty}(\frac{1}{NT}\nu'\nu) = \Sigma_{\nu}$$

(13e) $$\text{plim}_{N\to\infty}(\frac{1}{NT}\eta'\eta) = \Sigma_{\eta}$$

where Σ_ν and Σ_η are positive definite matrices and $\Sigma_{\nu\eta}$ is a non-zero matrix. Σ_ν is a k×k matrix, Σ_η is an L×L matrix and $\Sigma_{\nu\eta}$ is a k×L matrix. We will occasionally refer to Σ_x and Σ_z. They equal $\frac{1}{1-\gamma_1^2}\Sigma_\nu$ and $\frac{1}{1-\gamma_2^2}\Sigma_\eta$ respectively. We assume that η is uncorrelated with ϵ. However, it may be correlated with α, this correlation noted by

$$(14) \qquad \underset{N\to\infty}{\text{plim}}\left(\frac{1}{N} Z_t'\alpha\right) = \Sigma_{z\alpha}$$

where $\Sigma_{z\alpha}$ will be an L×1 zero vector under the null hypothesis and non-zero otherwise.

This is a particularly simple structure for the data generation process but it has the appealing property that as γ increases from 0 to 1, an increasing fraction of the variance of the random variable is due to the variation across individuals.[3] Since panel data are often slow moving over time, the performance of the specification test at high levels of γ is of considerable interest. We define the between estimator using the means of the instruments as $\hat{\beta}_1$ and the between estimator using Z^* as $\hat{\beta}_2$. Under the null hypothesis, the asymptotic covariance of $\sqrt{N}\hat{\beta}_1$ is given by

$$(15) \qquad V_1 = \sigma_u^2 \frac{T^2(2\sum_t b_t - T)}{(\Sigma(a_t + b_t)-T)^2}\left[\sum_{xz}\sum_{zz}^{-1}\sum_{xz}'\right]^{-1}$$

where $a_t = \sum_s^t \gamma_1^{s-1}$ and $b_t = \sum_s^t \gamma_2^{s-1}$. The asymptotic covariance matrix for $\sqrt{N}\ \hat{\beta}_2$ is given by

$$(16) \qquad V_2 = \sigma_u^2 \left[\frac{FB^{-1}F'}{T^2}\sum_{xz}\sum_{zz}^{-1}\sum_{xz}'\right]^{-1}$$

where $F = [a_T+b_1-1 \quad a_{T-1}+b_2-1 \ \dots\ a_1+b_T-1]$ and B is the TxT matrix

---

[3] For example, the between variance for X as a fraction of the total variance equals $(2\sum_t a_t - T)/T^2$ where $a_t = \sum_{s=1}^{t}\gamma_1^{s-1}$. This fraction varies between 1/T and 1 as γ_1 increases from 0 to 1.

$$B = \begin{bmatrix} 1 & \gamma_2 & \gamma_2^2 & \gamma_2^3 \cdots & \gamma_2^{T-1} \\ \gamma_2 & 1 & \gamma_2 & \gamma_2^2 \cdots & \gamma_2^{T-2} \\ & & \vdots & & \\ \gamma_2^{T-1} & \gamma_2^{T-2} & \gamma_2^{T-3} & \gamma_2^{T-4} \cdots & 1 \end{bmatrix}$$

We consider local alternatives of the form $\sum_{z\alpha} \neq 0$ and $\sqrt{N} \sum_{z\alpha} \rightarrow \psi < \infty$ as N approaches ∞. Under the null hypothesis, the probability limit (as $N \rightarrow \infty$) of $\hat{\beta}_w^{IV} - \hat{\beta}_b^{IV}$ is zero and q_2 is Chi square with k degrees of freedom. Under the alternative hypothesis, q_2 is distributed as a non-central chi square random variable with k degrees of freedom and non-centrality parameter δ where $\delta^2 = \bar{q}' M^{-1} \bar{q}$, $\bar{q}$ is the probability limit of $\sqrt{N}(\hat{\beta}_w^{IV} - \hat{\beta}_b^{IV})$ and M is the asymptotic covariance of $\sqrt{N} q_2$ (see Scheffe (1959)). We define $\bar{q}_i$ as $\bar{q}$ with $\hat{\beta}_i$ substituted for $\hat{\beta}_b^{IV}$ $(i=1,2)$[4]. The asymptotic biases for the two estimators using the different set of instruments are

(18) $$\bar{q}_1 = \frac{T^2}{\Sigma(a_t+b_t) - T} \left[\sum_{xz} \sum_{z}^{-1} \sum_{xz}' \right]^{-1} \sum_{xz} \sum_{z}^{-1} \sum_{z\alpha}$$

and

(19) $$\bar{q}_2 = \frac{T \cdot FB^{-1} e_T}{FB^{-1} F'} \left[\sum_{xz} \sum_{z}^{-1} \sum_{xz}' \right]^{-1} \sum_{xz} \sum_{z}^{-1} \sum_{z\alpha}.$$

If $\gamma_1 = \gamma_2$, it is straightforward to show that $V_1 = V_2$ and that $\bar{q}_1 = \bar{q}_2$ leading to the following proposition:

---

[4] Do not confuse q_i, a chi square test statistic with $\bar{q}_i$, the asymptotic bias of a chi square statistic.

Proposition 1: Given the model in equations 13 - 14 and the assumption that $\gamma_1 = \gamma_2$, the two estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ are equally efficient asymptotically and the power of the specification test of the hypothesis that the instrumental variables are uncorrelated with the individual effects is unaffected by the choice of instrument set.

As $\gamma_1 - \gamma_2$ diverges from zero, the variances and power of the specification test begins to differ. Since $V_1 - V_2$, is positive definite, it would appear that the power of the specification test should increase using Z^* as the set of instruments. However, it will turn out that $\bar{q}_1$ will be also be greater than $\bar{q}_2$ which will increase the power of the test using the means of Z as instruments. For $\gamma_1 > 0$, this result is formalized in the following proposition.

Proposition 2: If $\gamma_1 > 0$ and $\gamma_2 = 0$, then the power of the test statistic using $\bar{Z}$ as instruments is greater than the power using Z^* as instruments.

Proof: see appendix.

Figure 1 graphs the increase in power of the test statistic using $\bar{Z}$ as instruments over Z^* as instruments for various values of T and γ_1 for k=L= 1. The covariance of Z and α is set equal to half the variance of Z. The graph depicts the ratio of the power of the two test statistics. A fraction greater than one indicates that the test statistic using $\bar{Z}$ as instruments correctly rejects the null hypothesis more frequently. The graph shows that the two tests perform significantly differently at high values of γ_1. Furthermore, the tests do not quickly converge as T increases. For $\gamma_1 = .8$, the test using $\bar{Z}$ correctly rejects 19.4% more frequently at T=8 (power is .24 versus .20); the ratio declines to 12.9% at T=16 (power is .62 versus .55). Even at modest degrees of correlation across time (e.g. $\gamma_1 = .5$), the increase in power is on the order of 4 to 6% for T between 4 and 8.

We now turn our attention to the case where $\gamma_1 = 0$ and $\gamma_2 > 0$. Some tedious algebra shows that the asymptotic bias for the test statistic under the alternative hypothesis is greater when $\bar{Z}$ is used as the set of instruments. Again, the variance of the between estimator is greater when $\bar{Z}$ is used. In this case it is difficult however to show that the power of the test is greater when $\bar{Z}$ is used as the instrument set rather than Z^*. In the simple case where k=L=1 a grid search shows that the test statistic using $\bar{Z}$ is more powerful than when Z^* is used.

The increase in power is more dramatic now as illustrated in figure 2. Again, K=L=1 and the covariance of Z and α is set equal to half the variance of Z. as illustrated in figure 2. At $\gamma_2 = .8$ and T = 7, the increase in the number of rejections is 41% (power equals .14 versus .10) and declines to 26% at T=16 (power is .33 versus .26). At $\gamma_2 = .9$, the test using the mean of the instruments rejects nearly twice as often as when the instruments for each time period are used separately. Note that at $\gamma = .9$ and T = 7, 80% of the variation in the data occurs across individuals rather than for individuals across time. It is quite typical for many panel data applications to lose 80% of the variance in the data when using the fixed effects estimator.

Finally, figure 3 graphs the efficiency gains from using the means of the instruments for k=L=1 when T=5 and γ_1 and γ_2 vary between .1 and .9. As pointed out above, the tests perform equivalently when $\gamma_1 = \gamma_2$ and the test using the means of the instruments performs better as the two autocorrelations move apart. However, the improvement is not dramatic with a maximum improvement of less than 14%. This raises the issue of the performance of the tests in small samples. We turn our attention to this issue in the next section.

IV. Small Sample Characteristics of the Test

Specification test statistics in general have been criticized for having low power (e.g. Holly [1982], Newey [1985]). One might expect that the power of the test would deteriorate further as a result of the additional noise from the instrumenting of variables in X. To consider how well the test works in practice, we present results from a Monte Carlo experiment. We consider a simple model with k=L=1, set β equal to 0 in equation 1 and take draws from a normal distribution for X_{it}, Z_{it}, ϵ_{it}, and α_i, each with mean 0. The first three variables have variance 1 and α_i has variance 1/T. The covariance of Z and ϵ is zero while the other covariances vary from experiment to experiment. After generating the data, we compute the within and between estimates of β, their variances, and the chi-square test (which has 1 degree of freedom).[5] We repeat the process 1000 times for each model.

For the first set of results, we assume that $\gamma_1=\gamma_2=0$. With these assumptions, δ^2 is given by the formula

$$\delta^2 = \psi^2 T^2 \left(\frac{T-1}{2T-1} \right) \tag{19}$$

The asymptotic power of the test increases with more time periods, and with a higher correlation between the time means of the instruments and the individual effects. Note that the tests should perform equally well based on the results from the last section.

The first results are from a simulation with N=200 and T=5. $Cov(X_{it}, \epsilon_{it}) = 0.4$ and $\sigma_{z\alpha}$ and σ_{xz} vary from 0 to 0.06 and 0.1 to 0.7 respectively. Table 1 shows the fraction of times the null hypothesis is rejected due to q_2 in equation 12 exceeding the 5% critical value for a Chi-square random variable with 1 degree of freedom. The top panel of table 1 presents results using the mean of Z as the instrument set while the bottom panel uses the set Z^*. For future reference, call the first test statistic c_1 and the second statistic c_2. The first column in each panel shows the computed size of the test. For each of the four simulations ranging σ_{xz} from 0.1 to 0.7, the size of the test using c_2 is higher and closer to the theoretical level of 5%. However,

---

[5] Equivalently, we could take the square root of the statistic and use the standard normal distribution. Constructing the experiment with one degree of freedom allows us to avoid issues of direction in defining the local alternatives which affect the power of the test.

neither of these tests has a computed size near 5% at very low levels of correlation between Z and X. This is suggestive of the results of Nelson and Startz (1990a, 1990b) who have shown that the distribution of IV estimators diverges dramatically from the asymptotic distribution in the presence of poor instruments.

The remaining columns of the two panels in table 1 show the power of the test in the face of increasing correlation of Z with α. In every instance, the power of c_2 is higher than that of c_1. The increase in power can be significant, particularly with poor instruments. These results are striking given the number of individuals in the data set (N=200) as well as the fact that c_2 has the same distribution asymptotically as c_1. Clearly, the main advantage of c_2 over c_1 lies in its performance in the presence of poor instruments.

Whether this is of any practical significance depends on the degree to which the two test statistics give conflicting answers about the null hypothesis. Table 2 presents this information. For each simulation this table indicates the fraction of times either c_1 or c_2 (but not both) is significantly different from zero (bottom number) and the fraction of times c_2 is significantly different from zero conditional on c_1 not being significant at the 5% level (top number). For example, when $\sigma_{xz} = 0.3$ and $\sigma_{z\alpha} = 0.04$, the tests give conflicting results 13.5% of the time. Conditional on the tests disagreeing, c_2 gave the correct result 72.6% of the time.

This table shows that when the test statistics disagree, it is much more likely that c_2 gives the correct result, assuming $\sigma_{z\alpha} \neq 0$. It is true that in the case when the null hypothesis is true, c_2 is more likely to reject this hypothesis; however, this has the effect of making the size closer to the theoretical size of the test. Furthermore, the tests can conflict in a great many cases - as much as 22% of the time. Conditional on $\sigma_{z\alpha} \neq 0$, the average fraction of times the test statistics give conflicting answers is slightly more than 8% of the time. In those cases, c_2 will give the correct answer over 80% of the time.

Why does c_1 perform so badly compared to c_2? There are three possible reasons: the distributions of the estimator themselves (both within and between), the moment matrices required to construct the variance of the estimators and the estimated variance of the error term constructed from the residuals. Table 3 presents results for varying qualities of instruments (σ_{xz}) for the case where the null hypothesis is correct. The first panel of the table presents information on the between estimators using $\bar{Z}$ and Z^* as the instrument sets respectively. A subscript of 1 indicates the between estimator using $\bar{Z}$ as the instrument set while a subscript of 2 indicates the between estimator using Z^*. The second column shows the theoretical variance of $\hat{\beta}$ for different values of σ_{xz}. The theoretical variance of the error term (σ_1^2) is 0.40. Columns 3 and 4 show the sample variance of $\hat{\beta}$ from the simulation's 1000 replications. Except for the case where σ_{xz}=.1, the sample variances are close to the true variance. However $\hat{\beta}_1$ has an extremely dispersed distribution in the presence of a poor IV. While not reported, the sample means of the estimators correspond fairly well to their theoretical value of zero. The estimated variance of $\hat{\beta}_2$ and of the residual variance is less sensitive to the quality of the instrument than in the case of estimators using $\bar{Z}$ as the instrument set. For all values of σ_{xz} the estimated variance of $\hat{\beta}_1$ tends to be too large, thereby contributing to the poor power characteristics of c_1.

This is confirmed in the second part of the table where results for the fixed effects estimator are presented. Column 2 gives the asymptotic variance of $\hat{\beta}_W$; the variance of the error term equals 1 in the limit. The sample variances of $\hat{\beta}$ are close to the true variance for reasonably good instruments though there is greater dispersion for σ_{xz} = 0.1. The estimated variance of $\hat{\beta}$ as well as the estimated variance of the error term are very close to the true variances for $\sigma_{xz} > 0.1$, suggesting that the problem with c_1 in cases where $\sigma_{xz} > 0.1$ is with the between estimator rather than the within.

To give a sense of how quickly the asymptotic results hold, we repeated the experiment with N=1000 and T=5. Tables 4 and 5 give the essential results. In table 4, we report the fraction of null hypothesis rejections as σ_{xz} increases from .05 to .20 and as $\sigma_{z\alpha}$ ranges from 0 to .03. With moderately good instruments ($\sigma_{xz} \geq .20$) the distributions of c_1 and c_2 look very similar. The actual size of the tests corresponds closely to the theoretical size and the computed powers of the two tests are quite similar. It continues to be the case though that c_2 has greater power in all cases than c_1 although the differences become quite small as σ_{xz} increases. Both test statistics perform badly with very poor instruments as would be expected. The statistic c_1 continues to be somewhat too conservative a test at low levels of σ_{xz} (i.e. size too small).

Table 5 shows that the two tests, c_1 and c_2, continue to give conflicting results in roughly 10% of the cases (conditional on the alternative hypothesis). In those cases, c_2 correctly identifies the positive correlation between Z and α while c_1 does not anywhere from 53 to 85% of the time. Clearly one should not assume in panel data with instrumental variables that 1000 equals infinity.

The next set of results assumes that $\gamma_1 = 0.9$ and $\gamma_2 = 0$. Recall that this implies that 80% of the variance in X is lost when the fixed effects estimator is used. In all other respects, the model is the same as in the first experiment. Table 6 shows the power of the two test statistics. The clear advantage of c_1 over c_2 is evident here. While both test statistics have the correct size at moderate levels of correlation between Z and X, the power of c_1 is greater than the power of c_2 in every case conditional on the alternative. More significantly, table 7 shows that c_1 is more likely to detect correctly correlations between the instruments and the individual effects when the two test statistics give conflicting results for a 95% level significance test. While the two statistics are like likely to give conflicting results on average 9% of the time, c_1 correctly detects the correlation in the cases where it does in fact occur some 89% of the time. There is a clear advantage to using the more powerful test statistic in this case.

In most data applications, it is likely to be the case that both the explanatory variables (Xs) and the instruments (Zs) are likely to be slow moving data over time. The asymptotic theory says that the tests are equivalent (or nearly so) in this case. The final set of results tests whether this is in fact the case for small samples. The last experiment is repeated with the additional assumption that $\gamma_2 = \gamma_1 = 0.9$. Results are presented in tables 8 and 9. As is the case in the experiment when γ_1 and γ_2 both equal zero, there is a clear advantage to using the instrument set Z^* rather than the means of the instruments. The actual power of the test is greater and the fraction of cases in which c_2 correctly rejects the null hypothesis when c_1 fails to reject is .85. The two tests give conflicting results 16% of the time.

The results from this set of simulations can be summarized as follows. If the data exhibit the same relative degree of variation within individuals over time (in the sense of the value of γ) then the two test statistics c_1 and c_2 are asymptotically equivalent but in small samples c_2 can outperform c_1. Where the variation between and within is different for the instruments than the endogenous right hand side variables, than c_1 should be employed. In large samples c_1 should always be employed since it can never be less efficient than c_2. However, note that $N = 1000$ may not be large enough.

V. Conclusion

Testing for correlated individual effects has become increasingly important with the greater use of panel data sets. This paper shows that the type of specification test often employed in models where all the explanatory variables are considered exogenous carries over in a straight forward manner to models with endogenous explanatory variables. However greater attention must be paid to the quality of the instruments used for the explanatory variables if the actual size and power of the test statistic is to correspond to the theoretical size and power. Even in relatively large samples (N = 1000) the true size of the test may be quite a bit smaller than the theoretical size even if the instruments seem reasonable. Particular caution should be exercised with small data sets.

While the asymptotic theory suggests comparing the fixed effects estimator with the between estimator using the means of the instruments, there may be cases when it is preferable to use the full set of instruments Z^*. This is particularly so when the "within" variation in the data and instruments is a large fraction of the total variance. Conversely, if the Xs are slow moving over time but the instruments are not (or vice versa), then there is a distinct advantage to constructing the specification test using the less efficient estimator to take advantage of its greater asymptotic bias.

References

Amemiya, T. and T.E. MaCurdy, "Instrumental Variable Estimation of an Error Components Model," Econometrica, 54(1986): 869-881.

Breusch, T. G. Mizon and P. Schmidt, "Efficient Estimation Using Panel Data," Econometrica, 57(1989): 695-700.

Chamberlain, G.C. "Panel Data," in Handbook of Econometrics, Vol. II, ed. Z. Griliches and M. Intriligator. Amsterdam: Elsevier, 1984.

Hausman, J.A. "Specification Tests in Econometrics," Econometrica, 46(1978): 1251-1271.

Hausman, J.A. and W.E. Taylor, "Panel Data and Unobservable Individual Effects," Econometrica, 49(1981): 1377-1398.

Holly, A. "A Remark on Hausman's Specification Test," Econometrica, 50(1982): 749-760.

Holtz-Eakin, D. "Testing for Individual Effects in Autoregressive Models," Journal of Econometrics 39(1988): 297-307.

Maddala, G.S. "The Use of Variance Components Models in Pooling Cross Section and Time Series Data," Econometrica, 39(1971): 341-358.

Mundlak, Y. "Empirical Production Function Free of Management Bias," Journal of Farm Economics, 43(1961): 44-56.

Nelson, C. and R. Startz, "The Distribution of the Instrumental Variables Estimator and its t-Ratio When the Instrument is a Poor One," Journal of Business 63(1990a): S125-S140.

Nelson, C. and R. Startz, "Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator," Econometrica 58(1990b): 967-976.

Nerlove, M. "A Note on Error Components Models," Econometrica, 39(1971): 383-396.

Newey, W. "Generalized Method of Moments Specification Testing," Journal of Econometrics, 29(1985): 229-256.

Scheffe, H. The Analysis of Variance, New York: J. Wiley and Sons, 1959.

Schmidt, P. and D. Wyhowski, "Simultaneous Equations and Panel Data," mimeo, Department of Economics, Michigan State University, 1988.

Proof of Proposition 2: Define δ_1^2 as the non-centrality parameter for the test using $\bar{Z}$ as the instrument set. Similarly, define δ_2^2 for the instrument set Z^*. Let V_w be the variance of the within estimator. First, we note that V_1-V_2 is a positive definite matrix, assuming $\gamma_1 > 0$:

(A1) $$V_1 - V_2 = T^3\sigma_u^2 \left[\sum_{xz}\sum_{zz}^{-1}\sum_{xz}'\right]^{-1}\left[\frac{1}{(\Sigma a_t)^2} - \frac{1}{T\Sigma a_t^2}\right] > 0$$

by Chebyshev's Inequality.

It is easily shown that $V_1^{-1}\bar{q}_1 = V_2^{-1}\bar{q}_2 = (\Sigma a_t/T)\sum_{xz}\sum_{zz}^{-1}\sum_{z\alpha} = \lambda$. Therefore,

(A2) $$\delta_1^2 - \delta_2^2 = \lambda'(R_1 - R_2)\lambda$$

and is greater than zero if R_1-R_2 is positive definite, where R_i equals $[V_i^{-1}V_wV_i^{-1} + V_i^{-1}]^{-1}$, $i = 1,2$. R_1-R_2 will be positive definite if R_2^{-1}- R_1^{-1} is positive definite. But

(A3) $$R_2^{-1} - R_1^{-1} = (V_2^{-1} - V_1^{-1})V_w(V_2^{-1} - V_1^{-1}) + (V_2^{-1} - V_1^{-1}).$$

Each of the bracketed terms in A3 is positive definite, so R_2^{-1}-R_1^{-1} is positive definite and $\delta_1^2 > \delta_2^2$. □

Table 1. Computed Power and Size

$\sigma_{x\epsilon} = .4$

N = 200 T = 5

c_1

		$\sigma_{z\alpha}$ = .0	.02	.04	.06
σ_{xz}	.1	.005	.036	.156	.304
	.3	.040	.154	.529	.863
	.5	.037	.210	.530	.855
	.7	.040	.171	.511	.831

c_2

		$\sigma_{z\alpha}$ = .0	.02	.04	.06
σ_{xz}	.1	.023	.078	.237	.444
	.3	.055	.208	.592	.881
	.5	.048	.243	.564	.881
	.7	.044	.200	.568	.859

These tables present the fraction of rejections of the null hypothesis that $\sigma_{z\alpha} = 0$ out of 1000 replications. The top panel (c_1) uses $\bar{Z}$ as an instrument for $\overline{X}$ while the bottom panel (c_2) uses Z^*.

Table 2. Quality of c_1 versus c_2

$\sigma_{x\epsilon} = .4$

N = 200 T = 5

σ_{xz} \ $\sigma_{z\alpha}$	.0	.02	.04	.06
.1	1.000 .016	.875 .056	.786 .140	.820 .217
.3	.714 .035	.810 .084	.726 .135	.667 .057
.5	.789 .019	.807 .057	.750 .064	.846 .039
.7	.778 .009	.921 .038	.951 .061	.938 .032

The top number in each cell is the fraction of times c_2 is significant and c_1 insignificant at the 5% level. The bottom number is the fraction of times the two test statistics give conflicting answers at the 5% level.

Table 3. Variance Estimates

$\sigma_{x\epsilon} = .4$

$N = 200 \quad T = 5$

Between Estimators

σ_{xz}	$V(\hat{\beta})$	Sample $V(\hat{\beta}_1)$	Sample $V(\hat{\beta}_2)$	Estimated $V(\hat{\beta}_1)$	Estimated $V(\hat{\beta}_2)$	Estimated $\hat{\sigma}^2_1$	Estimated $\hat{\sigma}^2_2$
.10	1.000	2700	.458	$2x10^9$	.705	498.8	.459
.30	.111	.127	.095	.148	.108	.430	.408
.50	.040	.038	.035	.044	.040	.407	.401
.70	.020	.019	.019	.021	.021	.402	.400

Fixed Effects Estimator

σ_{xz}	$V(\hat{\beta})$	Sample $V(\hat{\beta})$	Estimated $V(\hat{\beta})$	Estimated $\hat{\sigma}^2$
.10	.125	.638	15.58	1.73
.30	.014	.015	.015	1.02
.50	.005	.005	.005	1.01
.70	.003	.003	.003	1.01

Table 4. Computed Power and Size

$\sigma_{x\epsilon} = .4$

N = 1000 T = 5

c_1

		$\sigma_{z\alpha}$ .0	.01	.02	.03
σ_{xz}	.05	.006	.050	.170	.326
	.10	.026	.177	.537	.786
	.15	.034	.226	.578	.897
	.20	.043	.203	.605	.920

c_2

		$\sigma_{z\alpha}$ .0	.01	.02	.03
σ_{xz}	.05	.025	.091	.246	.444
	.10	.050	.219	.546	.832
	.15	.052	.255	.612	.907
	.20	.050	.224	.639	.923

These tables present the fraction of rejections of the null hypothesis that $\sigma_{z\alpha} = 0$ out of 1000 replications. The top panel (c_1) uses $\bar{Z}$ as an instrument for $\bar{X}$ while the bottom panel (c_2) uses Z^*.

Table 5. Quality of c_1 versus c_2

$\sigma_{x\epsilon} = .4$

N = 1000 T = 5

		$\sigma_{z\alpha}$ .0	.01	.02	.03
σ_{xz}	.05	.913 .023	.847 .059	.779 .136	.831 .178
	.10	.929 .028	.719 .096	.534 .131	.698 .116
	.15	.821 .028	.663 .089	.639 .122	.609 .046
	.20	.652 .023	.698 .053	.718 .078	.537 .041

The top number in each cell is the fraction of times c_2 is significant and c_1 insignificant at the 5% level. The bottom number is the fraction of times the two test statistics give conflicting answers at the 5% level.

Table 6. Computed Power and Size

$N = 200 \quad T = 5 \quad \gamma_1 = 0.9$

c_1

σ_{xz}	$\sigma_{z\alpha}$: .0	.06	.12	.18
1	.002	.007	.021	.023
3	.026	.124	.310	.362
.5	.042	.256	.514	.715
.7	.043	.328	.813	.942

c_2

σ_{xz}	$\sigma_{z\alpha}$: .0	.06	.12	.18
.1	.003	.005	.012	.011
.3	.040	.101	.242	.261
.5	.058	.191	.421	.639
.7	.046	.259	.608	.791

These tables present the fraction of rejections of the null hypothesis that $\sigma_{z\alpha} = 0$ out of 1000 replications. The top panel (c_1) uses $\bar{Z}$ as an instrument for $\bar{X}$ while the bottom panel (c_2) uses Z^*.

Table 7. Quality of c_1 versus c_2

$N = 200 \quad T = 5 \quad \gamma_1 = 0.9$

		$\sigma_{z\alpha}$ .0	.06	.12	.18
σ_{xz}	.1	.667 .003	.333 .006	.091 .011	.000 .012
	.3	.792 .024	.298 .057	.041 .074	.000 .061
	.5	.654 .026	.099 .081	.021 .097	.298 .188
	.7	.571 .021	.063 .079	.023 .215	.048 .167

The top number in each cell is the fraction of times c_2 is significant and c_1 insignificant at the 5% level. The bottom number is the fraction of times the two test statistics give conflicting answers at the 5% level.

Table 8. Computed Power and Size

$N = 200 \quad T = 5 \quad \gamma_1 = \gamma_2 = 0.9$

c_1

		$\sigma_{z\alpha}$ = .0	.06	.12	.18
σ_{xz}	.1	.001	.008	.019	.022
	.3	.018	.071	.147	.203
	.5	.022	.187	.444	.489
	.7	.021	.302	.733	.799

c_2

		$\sigma_{z\alpha}$ = .0	.06	.12	.18
σ_{xz}	.1	.010	.013	.037	.053
	.3	.039	.163	.321	.369
	.5	.046	.306	.654	.718
	.7	.041	.422	.838	.906

These tables present the fraction of rejections of the null hypothesis that $\sigma_{z\alpha} = 0$ out of 1000 replications. The top panel (c_1) uses $\tilde{Z}$ as an instrument for $\bar{X}$ while the bottom panel (c_2) uses Z^*.

Table 9. Quality of c_1 versus c_2

$N = 200 \quad T = 5 \quad \gamma_1 = \gamma_2 = 0.9$

σ_{xz} \ $\sigma_{z\alpha}$	.0	.06	.12	.18
.1	1.000 .009	.778 .009	.909 .022	.833 .048
.3	.839 .031	.896 .115	.901 .212	.881 .218
.5	.750 .048	.837 .178	.879 .281	.893 .291
.7	.794 .034	.773 .216	.768 .194	.855 .152

The top number in each cell is the fraction of times c_2 is significant and c_1 insignificant at the 5% level. The bottom number is the fraction of times the two test statistics give conflicting answers at the 5% level.

Figure 1. Power Ratio of c_1 to c_2

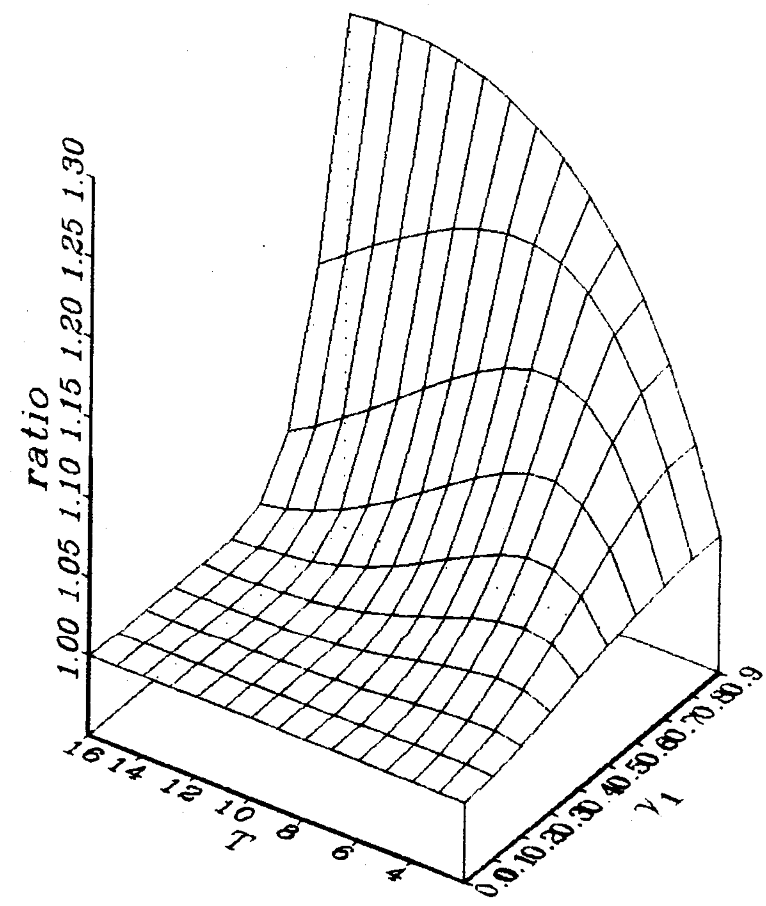

Figure 2. Power Ratio of c_1 to c_2

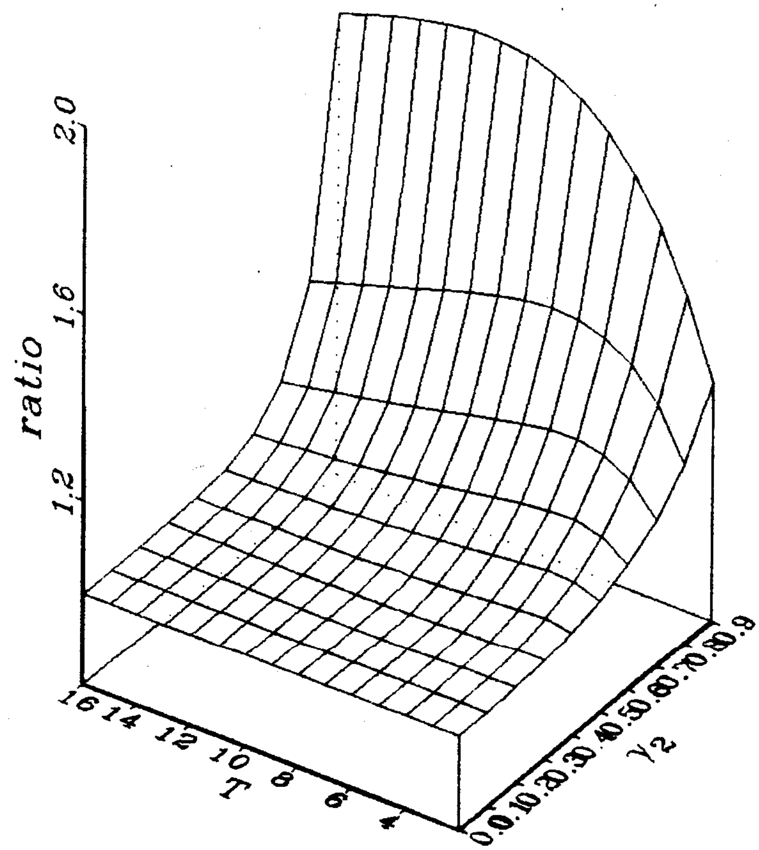

Figure 3. Power Ratio of c_1 to c_2