Faculty Working Papers

DECISION RULES FOR THE CHOICE OF STRUCTURAL EQUATIONS

Kimio Morimune and Takamitsu Sawa

#443

College of Commerce and Business Administration
University of Illinois at Urbana-Champaign

FACULTY WORKING PAPERS

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

October 20, 1977

DECISION RULES FOR THE CHOICE OF STRUCTURAL EQUATIONS

Kimio Morimune and Takamitsu Sawa

#443

Decision Rules for the Choice of Structural Equations

Kimio Morimune and Takamitsu Sawa

October 11, 1977

Prepared Under

National Science Foundation

GRANT SOC 76-22232

Abstract

In practical econometric analysis we are very often faced with the problem of how to specify structural equations. The conventional t-test of coefficients is apparently inappropriate. The largest root, say λ, of a certain determinantal equation provides us with a basis for the test of over-identifying restrictions. The preliminary test, based on λ, may give us a possible decision rule for the choice of the most adequate structural equation from given nested alternatives. However, ambiguity remains about how to choose the significance level. As an alternative procedure, we apply the minimum Akaike Information Criterion to our problem. This gives us a quite simple decision rule based on the comparison of λ's. Moreover, we propose another decision rule called the unbiased decision rule; unbiased in the sense that we reach a correct decision with more than a half probability. Applications of these newly developed procedures are exemplified by Klein's Model I.

1. Introduction

In recent years, much emphasis has been laid on the problem of statistical model identification: how to identify a model statistically when it cannot be completely specified from *a priori* ground. In fact, a considerable number of works have been done in the last decade with regard to the choice of the most adequate regression model. The purpose of the present paper is to extend the statistical procedures developed for the choice of regression models to a simultaneous equations system. When we discuss the model identification, we must fix the idea about the *adequacy* of a model. That is, we need to introduce a suitable measure of the discrepancy or the distance of a model from the unknown true structure. Different measures lead us to different procedures, of course.

It is ordinarily expected that the more complicated model will provide the better approximation to reality. However, on the contrary, the less complicated model would be preferred if we wish to pursue accuracy of estimation. In general, closeness to the truth is quite likely to be incompatible with parsimony of parameters. That is, if one pursues one of the criteria, the other must be necessarily sacrificed.

Akaike [1] has proposed a widely applicable statistic that incorporates these two criteria ingeniously. As it is based on Kullback-Leibler's information measure for discrimination of two probability distributions, Akaike's statistic is called the *Akaike Information Criterion* and is abbreviated as the AIC. It is defined as minus twice the *maximized* likelihood function plus twice the number of parameters

in a model. (See equation (3.1).) Given a set of alternative models, we choose the one which gives the smallest AIC. The procedure is called the Minimum AIC (MAIC). The advantage of this procedure is its applicability to any statistical problem so long as each of the alternative models well defines the likelihood function.

Following Akaike, Sawa [8] has recently developed another information criterion aimed specifically for the choice of linear regression models. This criterion is also based on Kullback-Leibler's information criterion.

Mallows [7] proposed a criterion called the C_p statistic which defines another procedure for selecting the optimal linear regression model. The C_p statistic is defined to be the residual sum of squares (RSS) plus twice the number of parameters (p) multiplied by an unbiased estimate $\hat{\omega}^2$ of the true variance of error terms:

$$(1.1) \qquad C_p = RSS + 2p\hat{\omega}^2 .$$

Obviously, the first term measures the accuracy of a model, and the second term stands for the penalty paid for increasing the number of parameters. We note that application of the MAIC to linear regression yields an asymptotically equivalent decision rule as Mallows' C_p.

Sawa and Takeuchi [9] proposed another criterion for choosing an optimal regression equation. The decision rule defined by this criterion is called the unbiased decision rule: <u>unbiased</u> in the sense that it leads us to the choice of the most adequate model with probability greater than one-half, when we compare two alternatives.

In Section 2, models and notations are described. In Section 3, we develop the MAIC procedure for selecting the most adequate structural equation. This gives us a quite simple criterion, so long as we define the AIC in terms of concentrated likelihood function for the limited information maximum likelihood estimate (Anderson and Rubin [2]). Moreover, the implication of the MAIC procedure will be discussed in the context of conventional hypotheses testing. In Section 4, we define a specification error of a structural equation in terms of identification conditions. We examine the distribution of the AIC criterion when both of the structural equations, being compared, are incorrectly specified. In Section 5, we propose Mallows' type risk function of postulating a particular structural equation as a model. Based on this risk function and the distribution theory developed in Section 4, the unbiased decision rule is derived. Critical points of unbiased decision rule are numerically evaluated and tabulated. In Section 6, a numerical example will be given.

2. Models and Notations

Suppose N alternative structural equations are given, and we are facing a problem: how to identify the most adequate one therefrom. The i-th equation is written as

$$(2.1) \qquad \underset{\sim}{y} = \underset{\sim}{Y}_i \underset{\sim}{\beta}_i + \underset{\sim}{Z}_i \underset{\sim}{\gamma}_i + \sigma \underset{\sim}{u}_i \ , \qquad i = 1, \ldots, N \ ,$$

where $\underset{\sim}{y}$ and $\underset{\sim}{Y}_i$ are T x 1 and T x G_i matrices, respectively, of observations on the endogenous variables; $\underset{\sim}{Z}_i$ is a T x K_i matrix of observations on the K_i exogenous variables; $\underset{\sim}{\beta}_i$ and $\underset{\sim}{\gamma}_i$ are, respectively, G_i-dimensional and K_i-dimensional column vectors of unknown parameters; $\underset{\sim}{u}_i$ is a

T-dimensional column vector of disturbances. Note that every alternative equation shares a common explained endogenous variable. The components of $\underset{\sim}{u}_i$ are independently normally distributed with mean 0 and unit variance, and σ is a (small) positive number. The reduced form of the complete system of equations includes

$$(2.2) \qquad \underset{\sim}{y} = \underset{\sim}{Z}\underset{\sim}{\pi}^* + \sigma\underset{\sim}{v} = \underset{\sim}{Z}_i\underset{\sim}{\pi}_i + \underset{\sim}{\bar{Z}}_i\underset{\sim}{\bar{\pi}}_i + \sigma\underset{\sim}{v} \ ,$$

$$(2.3) \qquad \underset{\sim}{Y}_i = \underset{\sim}{Z}\underset{\sim}{\Pi}_i^* + \sigma\underset{\sim}{V}_i = \underset{\sim}{Z}_i\underset{\sim}{\Pi}_i + \underset{\sim}{\bar{Z}}_i\underset{\sim}{\bar{\Pi}}_i + \sigma\underset{\sim}{V}_i \ ,$$

where $\underset{\sim}{Z}$ is a T x K matrix of observations on all the predetermined variables in the system; $\underset{\sim}{Z}_i$ and $\underset{\sim}{\bar{Z}}_i$ are T x K_i and T x $(K-K_i)$ matrices of observations, respectively, on the included and excluded predetermined variables in the i-th equation (2.1); $\underset{\sim}{\pi}_i^*$ is a K-dimensional vector of reduced form coefficients subdivided conformably with Z; $\underset{\sim}{\Pi}_i^*$ is K x G_i matrix of reduced form coefficients subdivided conformably with Z; $\underset{\sim}{v}$ is a T-dimensional vector and V_i is a T x G_i matrix of disturbances. Without losing any generality, we may assume

$$(2.4) \qquad \underset{\sim}{Z}_i' \underset{\sim}{\bar{Z}}_i = \underset{\sim}{0} \ .$$

Each row of $(\underset{\sim}{v}\ \underset{\sim}{V}_i)$ is independently normally distributed with mean $\underset{\sim}{0}$ and (nonsingular) covariance matrix

$$(2.5) \qquad \begin{pmatrix} \omega & \underset{\sim}{\omega}_i' \\ \underset{\sim}{\omega}_i & \underset{\sim}{\Omega}_{ii} \end{pmatrix} .$$

If we post-multiply (2.3) by $-\underset{\sim}{\beta}_i$ and add it to (2.2), we have

$$(2.6) \qquad \underset{\sim}{u}_i = \underset{\sim}{v} - \underset{\sim}{V}_i\underset{\sim}{\beta}_i \ .$$

In order that (2.1) be properly written with $\bar{Z}_i$ omitted,

$$\begin{pmatrix} \pi_i & \Pi_i \\ \bar{\pi}_i & \bar{\Pi}_i \end{pmatrix} \begin{pmatrix} 1 \\ -\beta_i \end{pmatrix} = \begin{pmatrix} \gamma_i \\ 0 \end{pmatrix} . \tag{2.7}$$

If $\bar{\pi}_i = \bar{\Pi}_i \beta_i$ permits a unique solution for β_i, then (2.1) is said to be identifiable.

We define the minimum variance ratio for the i-th equation as

$$\lambda_i = \frac{(y - Y_i \hat{\beta}_i)' \bar{P}_{Z_i} (y - Y_i \hat{\beta}_i)}{(y - Y_i \hat{\beta}_i)' \bar{P}_Z (y - Y_i \hat{\beta}_i)} , \tag{2.8}$$

where $\bar{P}_F = I - P_F = I - F(F'F)^{-1}F'$ and $\hat{\beta}_i$ is the LIML estimator of β_i. Note that λ_i never falls below unity.

3. Decision Rule by the Akaike Information Criterion (AIC)

In this section, we first derive the AIC for a structural equation, which provides us with a decision rule to identify the most adequate structural equation from a given set of alternatives. Then we consider about the implication of the MAIC procedure in the context of conventional hypotheses testing. For this purpose an extensive use is made of the small-σ asymptotic expansion originated by Kadane [4,5].

The AIC is generally defined for a particular model with well-defined likelihood function as follows:

$$\text{AIC} = -2 \log \text{(the maximized likelihood)} + 2 \text{ (number of parameters)} \tag{3.1}$$

The <u>concentrated</u> likelihood function for the i-th structural equation (2.1) is

(3.2) $$\text{constant} - \frac{T}{2} \log \lambda_i \ ,$$

where λ_i is the minimum variance ratio for the i-th model. (See Koopmans and Hood [3], pp. 166-8]. Hence we have the following propositions.

Proposition 3.1: The AIC for the i-th structural equation is

(3.3) $$AIC(i) = T \log \lambda_i + 2(K_i + G_i) \ .$$

The first term is interpreted to measure the degree of goodness-of-fit; it decreases along with the augumentation of the model. More precisely, if we augument the right-hand-side variables in (2.1), λ_i approaches one, and it is exactly equal to one whenever (2.1) is just-identified. The second term stands for the penalty for losing degrees-of-freedom by increasing the number of unknown parameters. Hence the AIC is said to be a statistic that takes into account the trade-off between the two desirable properties of statistical models; i.e., goodness-of-fit and parsimonious use of parameters.

The MAIC procedure is described as follows:

Proposition 3.2 (Decision Rule): We choose the j-th structural equation if and only if

$$AIC(j) \leq AIC(i) \quad \text{for } i = 1, 2, \ldots, N \ .$$

Now we consider about the statistical implication of the MAIC procedure. Let us confine ourselves to the case when N = 2; i.e., two alternative equations, say M1 and M2, are given. We assume that M1 is nested in M2. We note that in conventional hypotheses testing M1 is

taken as a null-hypothesis and M2 as an alternative hypothesis. According to Proposition 3.2, we choose M1 over M2 if

$$\text{(3.4)} \qquad T \log (\lambda_1/\lambda_2) < 2 P_{12} \text{ with } P_{12} = K_2 + G_2 - K_1 - G_1 ,$$

and vice versa. The statistic $T \log (\lambda_1/\lambda_2)$ is asymptotically distributed as $\chi^2 (P_{12})$ when M1 is true (Anderson and Rubin [2]). Then the decision rule defined by (3.4) is asymptotically equivalent to the classical pre-test procedure with significance levels given in Table 3.1.

Table 3.1: Significance Levels Implied by the MAIC Procedure

P_{12}	1	2	3	4	5
%	17	16	15	8	7

The significance level is fixed at, for example, 5% or 10% in a conventional pre-test regardless of the value of P_{12}. However, the MAIC procedure adapts it to the degree of freedom.

More precise finite-sample distribution of the relevant statistic was given by Kadane [6]. Theorem 2 of Kadane [6] is worth citing as a lemma:

Lemma 3.1 (Kadane): As σ goes to zero

$$\text{(3.5)} \qquad \frac{T-K_2-G_2}{P_{12}} \left(\frac{\lambda_1}{\lambda_2} - 1\right) \sim F(P_{12}, T-K_2-G_2) ,$$

if M1 is true.

Combining (3.4) and (3.5) yields a decision rule such that if

(3.6) $$F_{12} < \frac{T-K_1-G_1}{P_{12}} [\exp (\frac{2P_{12}}{T}) - 1]$$

or approximately

(3.7) $$F_{12} < 2 \frac{T-K_1-G_2}{T} ,$$

we choose M1, where

(3.8) $$F_{12} = \frac{T-K_2-G_2}{P_{12}} (\frac{\lambda_1}{\lambda_2} - 1) .$$

If (3.6) does not hold, we choose M2.

It may be of some interest to compute the critical points of the MAIC procedure and examine the implied significance levels on the basis of the approximate F-distribution. However, this will lead us to virtually the same results that Sawa [8] has obtained with regard to linear regression. As usual, Kadane's small-σ asymptotics justify in dealing with a structural equation as if it were a linear regression if the disturbance variance is relatively small.

4. Specification Error and Non-Central F-Distributions

In this section we give a definition of a specification error occurring in a structural equation. In most practical situations it is quite likely that all of the alternative equations are incorrectly specified. Therefore, it would be worthwhile to derive the distribution of the AIC statistic when every alternative is more or less misspecified.

<u>Definition 4.1</u>: <u>The structural equation (2.1) is said to be incorrectly specified if</u>

$$(4.1)\qquad \begin{pmatrix} \pi_i & \Pi_i \\ \bar{\pi}_i & \bar{\Pi}_i \end{pmatrix} \begin{pmatrix} 1 \\ -\beta_i \end{pmatrix} = \begin{pmatrix} \gamma_i \\ 0 \end{pmatrix} + \sigma\eta_i$$

<u>where η_i is a column vector with at least one nonzero element among the last K_2 elements.</u>

We note that (2.7) is an <u>a priori</u> restriction on the reduced-form coefficients, which must be taken into account when we maximize the likelihood function to obtain the limited information maximum likelihood estimate (Anderson and Rubin [2]). In order to identify a structural equation, we need to impose these <u>a priori</u> restrictions, even if we are uncertain about the validity of them. In any case, our <u>a priori</u> knowledges about the economy are described in terms of restrictions such as (2.7). Therefore, it would be reasonable to define specification errors of a structural equation in such a way as Definition 4.1. The specification error term η_i is multiplied by σ. This amounts to assuming that the specification error is in its magnitude of comparable order with disturbance terms in the equation.

Using (4.1) and post-multiplying 1 and $-\beta_i$ to (2.2) and (2.3), we can write the true structural equation as

$$(4.2)\qquad y = Y_i\beta_i + Z_i\gamma_i + \sigma Z\eta_i + \sigma u .$$

To illuminate the implication of our defining specification errors as such let us suppose that the true structural equation includes some extra endogenous and exogenous variables, say Y_s and Z_s; i.e.,

$$(4.3)\qquad y = Y_i\beta_i + Z_i\gamma_i + \sigma\, Y_s\beta_s + \sigma\, Z_s\gamma_s + \sigma u .$$

The neglected terms in mis-specified equations (2.1) are assumed to be of comparable order with disturbances. Substituting

$$(4.4) \qquad Y_s = Z\Pi_s + \sigma V_s ,$$

into (4.3) yields

$$(4.5) \qquad y = Y_i\beta_i + Z_i\gamma_i + \sigma Z(\Pi_s\beta_s + I_s\gamma_s) + \sigma u$$

where terms of $O(\sigma^2)$ are neglected, and I_s is a K x s matrix such that $Z_s = Z\, I_s$. Comparing (4.5) with (4.2), we see that

$$(4.6) \qquad \eta_i = \Pi_s\beta_s + I_s\gamma_s .$$

Further we see that

$$(4.7) \qquad \sigma u = \sigma v - \sigma V_i\beta_i - \sigma^2 V_s\beta_s ,$$

where σu is the disturbance of the true structural equation (4.3). Combining this with (2.6), we have

$$(4.8) \qquad \sigma u_i = \sigma v_i - \sigma V_i\beta_i = \sigma u + O(\sigma^2) .$$

Hence (2.1) and (4.3) have the same disturbance term up to order $O(\sigma)$ in small-σ asymptotics sense.

Lemma 3.1 was obtained assuming that the null model M1 is true. However, if a true structural model is (4.2) or equivalently (4.3) in small-σ sense, noncentral parameters must be included in the F-distribution.

<u>Theorem 4.1</u>: <u>As σ goes to zero</u>

$$(4.9) \qquad \frac{T-K_2-G_2}{P_{12}} \left(\frac{\lambda_1}{\lambda_2} - 1\right) \sim F(P_{12}, T-K_2-G_2 \,|\, \delta_1^2-\delta_2^2, \delta_2^2) ,$$

where

$$\delta_k^2 = \eta_k' Z' \bar{P}_{X_k} Z \eta_k \ , \tag{4.10}$$

and

$$X_k = \begin{pmatrix} Z\Pi_k^* \\ Z_k \end{pmatrix} , \quad k = 1,\ 2\ . \tag{4.11}$$

This theorem will be proved in the Appendix. The distribution in (4.9) is a doubly noncentral F with noncentral parameters $\delta_i^2 - \delta_j^2$ and δ_j^2. A version of the above theorm is as follows:

Lemma 4.2. As σ goes to zero

$$\frac{T-K}{P_{12}} (\lambda_1 - \lambda_2) \sim F(P_{12},\ T-K | \delta_1^2 - \delta_2^2,\ 0)\ . \tag{4.12}$$

The proof will also be given in the Appendix. This distribution includes only one noncentral parameter $\delta_1^2 - \delta_2^2$. On the other hand, we lose some degrees of freedom in denominator since $K \geq K_2 + G_2$.

Noncentral F distributions derived in this section will be used in the next section to obtain unbiased decision rule for choosing one from two alternative equations.

5. Mallows' Risk and Unbiased Critical Points (UCP)

Following Mallows [7], we choose

$$W_i = E\ ||y^0 - \hat{y}_i||^2 \tag{5.1}$$

as a risk of postulating a structural equation (2.1), where

$$y^0 = Z\pi^* + \sigma v^0 \tag{5.2}$$

is a vector of new independent observations on $\underset{\sim}{y}$ for the same set of predetermined variables $\underset{\sim}{Z}$; $\hat{\underset{\sim}{y}}_i$ is a vector of predicted values for $\underset{\sim}{y}$ based on the limited information maximum likelihood estimation of the equation (2.1): i.e.,

$$\hat{\underset{\sim}{y}}_i = \underset{\sim}{Z}_i \hat{\underset{\sim}{\pi}}_i + \overline{\underset{\sim}{Z}}_i \hat{\hat{\underset{\sim}{\pi}}}_i = \underset{\sim}{P}_{Z_i} y + \underset{\sim}{P}_{\overline{Z}_i} (\underset{\sim}{y} - \underset{\sim}{Y}_i \hat{\underset{\sim}{\beta}}_i) \hat{\rho}_i \tag{5.3}$$

where

$$\hat{\rho}_i = \frac{(\underset{\sim}{y} - \underset{\sim}{Y}_i \hat{\underset{\sim}{\beta}}_i)' \overline{\underset{\sim}{P}}_Z \underset{\sim}{y}}{(\underset{\sim}{y} - \underset{\sim}{Y}_i \hat{\underset{\sim}{\beta}}_i)' \overline{\underset{\sim}{P}}_Z (\underset{\sim}{y} - \underset{\sim}{Y}_i \hat{\underset{\sim}{\beta}}_i)} , \tag{5.4}$$

$\hat{\pi}_i$ and $\hat{\hat{\pi}}_i$ are the limited-information maximum likelihood (LIML) estimators of $\underset{\sim}{\pi}_i$ and $\overline{\underset{\sim}{\pi}}_i$ (Anderson and Rubin [2]).

It was proposed by Takeuchi [10] to make use of the LIML estimators of the reduced form coefficients to make predictions. The method is adequately called the <u>single equation method of prediction</u> in analogy with the single equation method of estimation.

We now evaluate W_i asymptotically as σ goes to zero. The proof of this theorem will be given in the Appendix.

<u>Theorem 5.1</u>: <u>As σ goes to zero</u>

$$W_i = \sigma^2 \omega \{T + (1 - r^2) \frac{T-K-1}{T-K-2} K + [r^2 - \frac{1-r^2}{T-K-2}](K_i + G_i) + [r^2 + \frac{1-r^2}{T-K-2}] \delta_i^2 \} + O(\sigma^3) , \tag{5.5}$$

<u>where δ_i^2 is defined in (4.10), ω is defined in (2.5) and</u>

$$r^2 = \frac{E(\underset{\sim}{u}'\underset{\sim}{v})^2}{E(\underset{\sim}{u}'\underset{\sim}{u})E(\underset{\sim}{v}'\underset{\sim}{v})} \tag{5.6}$$

is the square of the correlation coefficient between the structural disturbance $\underset{\sim}{u}$ and the reduced-form disturbance $\underset{\sim}{v}$ for $\underset{\sim}{y}$.

Suppose that we must choose one from two alternative structural equations, say M1 and M2, the former of which is nested in the latter. Let W_1 and W_2 be the risks of postulating the models M1 and M2, respectively. Our decision is <u>correct</u> if we choose M1 when $W_1 \leq W_2$ and M2 otherwise. Approximating W_j ($j = 1, 2$) by their small-σ asymptotic expansion given by (5.5), we can easily show that the inequality $W_1 \leq W_2$ is equivalent to:

$$(5.7) \qquad \delta_1^2 - \delta_2^2 \leq s\, P_{12}$$

where

$$(5.8) \qquad s = \frac{(T-K-1)r^2 - 1}{(T-K-3)r^2 + 1} \; ;$$

We note that $0 \leq s \leq 1$ and $s = 1$ only when $r^2 = 1$, which is the case when no endogenous variables are included in a structural equation (4.3).

For simplicity let us confine ourselves to a class of decision rules based on a ratio or difference of λ_1 and λ_2. That is, we decide to choose M1 if λ_1/λ_2 (or $\lambda_1 - \lambda_2$) is less than some preassigned constant c and choose M2 otherwise. Each decision rule is simply characterized by a constant c, which we call the <u>critical point</u>. The MAIC decision rule is a member of this class with c equalling the right-hand-side of (3.6). A decision based on Kadane's [6] preliminary test is also a member of this class, the critical point of which is determined depending on a preassigned significance level.

In what follows we will derive another member of the class which has a desirable property of unbiasedness. The definition of unbiasedness is as follows:

Definition 5.1: A decision rule with a critical point c^* is said to be unbiased, if

$$(5.9) \qquad P(F_{12} \leq c^* | W_1 \leq W_2) \leq .5$$

and

$$(5.10) \qquad P(F_{12} > c^* | W_1 > W_2) \geq .5 ,$$

where F_{12} is a test statistic found in either (4.9) or (4.12).

In words, if a decision rule leads us to the correct choice with probability greater than one-half, then it is said to be unbiased.

Since F_{12} is continuously distributed, the conditions (5.9) and (5.10) are equivalent to an equality:

$$(5.11) \qquad P(F_{12} \leq c^* | W_1 = W_2) = .5 .$$

From (5.7) we see that $W_1 = W_2$ if and only if

$$(5.12) \qquad \delta_1^2 - \delta_2^2 = s\, P_{12}$$

We note that the left-hand-side of (5.12) is one of the noncentrality parameters in the noncentral F distribution of F_{12} (see Theorem 4.1 and Lemma 4.2). The coefficient s depends on the unknown correlation coefficient r^2 given by (5.6), which must be estimated from sample observations.

Now we propose two decision rules, which are based on small-σ asymptotic distributions in Theorem 4.1 and Lemma 4.2, respectively.

Decision Rule I: We choose M1 if

$$(5.13) \qquad \frac{T-K_2-G_2}{P_{12}} \left(\frac{\lambda_1}{\lambda_2} - 1\right) \leq c_1 \ ;$$

otherwise we choose M2, where c_1 is the median of the noncentral F distribution $F(P_{12}, T-K_2-G_2 | \hat{s} P_{12}, 0)$ where $\hat{s}$ is the right-hand-side of (5.6) with r^2 substituted by its maximum likelihood estimate.

The noncentrality parameter in the denominator is equated to zero. This is justifiable when $\delta_2^2 = 0(\sigma^2)$. This simplifying assumption must be inevitably made, because there is no way of estimating δ_2^2, which measures the distance of the postulated model M2 from the true equation (4.3). It should be noted that equating δ_2^2 to zero implies that the augumented model M2 is virtually true in small-σ sense.

Decision Rule II: We choose M1 if

$$(5.14) \qquad \frac{T-K}{P_{12}} (\lambda_1 - \lambda_2) \leq c_2 \ ,$$

where c_2 is the median of F $(P_{12}, T-K | \hat{s} P_{12}, 0)$; we choose M2 if (5.14) is not satisfied.

The small-σ asymptotic distribution of the statistic on the left-hand-side of (5.14) is a singly noncentral F as was shown in Lemma 4.2. Therefore, in order to justify the decision rule II, we need not assume that the augumented model is true. In this sense the decision rule II might be preferred to the decision rule I which is based on a strong assumption that the augumented model is true in small-σ sense. However, it would be fair to note that in large econometric models K is

far greater than $K_2 + G_2$ and hence the degree of freedom in the denominator is drastically reduced by switching from the decision rule I to the rule II.

6. Numerical Example

The unbiased critical points (UCP) are computed and tabulated in Tables 2 and 3 for various values of P_{12}, $n = T-K_2-G_2$ (or T-K), and $s = 0.2(0.2)0.8$. We observe that these UCP's are smaller than Sawa and Takeuchi's [8] UCP's for linear regression models. Significance levels implied by the unbiased decision rule are also tabulated in Tables 4 and 5.

As an example of application, we compare two alternative structural wage functions in Klein's model I (T = 21, K = 8):

$$\text{(6.1)} \quad \begin{aligned} &\text{M1:} \quad W = 1.37 + 0.58X, \ \lambda_1 = 2.47 \\ &\text{M2:} \quad W = 1.50 + 0.44X + 0.13t + 0.146X_{-1}, \ \lambda_2 = 3.25 \end{aligned}$$

where W is the private wage bill, X is the private total production, and t is the time trend. The estimates of s are 0.87 for M1 and 0.68 for M2. Klein chose M2 as his wage function.

We base our decision on either

$$\text{(6.2)} \quad F_{12} = \frac{T-K_2-G_2}{P_{12}} \left(\frac{\lambda_1}{\lambda_2} - 1\right) = 2.69$$

or

$$\text{(6.3)} \quad F^*_{12} = \frac{T-K}{P_{12}} (\lambda_1 - \lambda_2) = 5.08 \ .$$

The MAIC critical point is 1.784; the unbiased critical points when $s = 0.7$ are 1.303 for F_{12} and 1.320 for F^*_{12}; the critical point of

Kadane's [5] 5% level pre-test is 3.59. Therefore, our decision rules developed in this paper strongly support Klein's choice of the wage function, while the conventional pre-test procedure leads us to the choice of the null-model M1.

Table 2

UNBIASED CRITICAL POINTS FOR THE CHOICE OF THE MOST ADEQUATE MODEL

	S = 0.2						S = 0.4				
n\P	1	2	3	4	5	n\P	1	2	3	4	5
5	0.642	0.938	1.093	1.163	1.205	5	0.774	1.141	1.293	1.370	1.417
6	0.626	0.935	1.067	1.135	1.176	6	0.755	1.113	1.262	1.338	1.384
7	0.615	0.918	1.049	1.116	1.157	7	0.742	1.093	1.241	1.316	1.361
8	0.607	0.905	1.036	1.102	1.142	8	0.732	1.079	1.226	1.300	1.344
9	0.601	0.896	1.026	1.091	1.131	9	0.725	1.068	1.214	1.287	1.331
10	0.596	0.889	1.013	1.083	1.123	10	0.719	1.059	1.204	1.277	1.321
11	0.592	0.883	1.011	1.076	1.115	11	0.714	1.052	1.197	1.269	1.313
12	0.589	0.878	1.006	1.070	1.110	12	0.710	1.048	1.190	1.263	1.306
13	0.586	0.873	1.001	1.066	1.105	13	0.707	1.041	1.185	1.257	1.300
14	0.583	0.870	0.997	1.062	1.101	14	0.704	1.037	1.181	1.253	1.296
15	0.581	0.867	0.994	1.058	1.097	15	0.702	1.033	1.177	1.248	1.291
16	0.580	0.864	0.991	1.055	1.094	16	0.699	1.030	1.173	1.245	1.288
17	0.578	0.862	0.988	1.052	1.091	17	0.698	1.027	1.170	1.242	1.285
18	0.577	0.860	0.986	1.050	1.089	18	0.696	1.024	1.168	1.239	1.282
19	0.575	0.858	0.984	1.048	1.086	19	0.694	1.022	1.165	1.237	1.279
20	0.574	0.856	0.982	1.046	1.084	20	0.693	1.020	1.163	1.234	1.277
21	0.573	0.854	0.981	1.044	1.083	21	0.692	1.016	1.161	1.232	1.275
22	0.572	0.853	0.979	1.043	1.081	22	0.691	1.017	1.160	1.231	1.273
23	0.572	0.852	0.978	1.041	1.080	23	0.690	1.015	1.158	1.229	1.271
24	0.571	0.851	0.976	1.040	1.078	24	0.689	1.014	1.157	1.227	1.270
25	0.570	0.849	0.975	1.039	1.077	25	0.688	1.013	1.155	1.226	1.268
30	0.567	0.545	0.971	1.034	1.072	30	0.685	1.008	1.150	1.221	1.263
40	0.564	0.840	0.965	1.028	1.066	40	0.681	1.001	1.143	1.214	1.256
50	0.562	0.837	0.962	1.024	1.062	50	0.678	0.998	1.139	1.210	1.252
100	0.558	0.830	0.955	1.017	1.055	100	0.674	0.990	1.132	1.202	1.243
1000	0.554	0.825	0.949	1.011	1.049	1000	0.669	0.984	1.125	1.194	1.236

Table 3

UNBIASED CRITICAL POINTS FOR THE CHOICE OF THE MOST ADEQUATE MODEL

	S = 0.6						S = 0.8				
n\p	1	2	3	4	5	n\p	1	2	3	4	5
5	0.922	1.338	1.501	1.584	1.634	5	1.085	1.543	1.714	1.801	1.854
6	0.900	1.306	1.466	1.547	1.596	6	1.060	1.507	1.675	1.760	1.811
7	0.884	1.284	1.442	1.522	1.570	7	1.042	1.483	1.648	1.731	1.782
8	0.873	1.287	1.424	1.503	1.551	8	1.029	1.484	1.628	1.710	1.766
9	0.865	1.255	1.411	1.489	1.536	9	1.019	1.450	1.613	1.695	1.743
10	0.858	1.245	1.400	1.478	1.524	10	1.011	1.439	1.601	1.682	1.730
11	0.852	1.237	1.391	1.469	1.515	11	1.005	1.431	1.591	1.672	1.720
12	0.848	1.230	1.384	1.461	1.507	12	1.000	1.423	1.583	1.663	1.711
13	0.844	1.224	1.378	1.455	1.501	13	0.996	1.417	1.577	1.656	1.704
14	0.840	1.220	1.373	1.450	1.495	14	0.992	1.412	1.571	1.650	1.698
15	0.838	1.216	1.369	1.445	1.491	15	0.989	1.407	1.566	1.645	1.693
16	0.835	1.212	1.365	1.441	1.486	16	0.986	1.403	1.562	1.641	1.688
17	0.833	1.209	1.361	1.437	1.483	17	0.983	1.400	1.556	1.637	1.684
18	0.831	1.206	1.358	1.434	1.480	18	0.981	1.396	1.555	1.633	1.680
19	0.829	1.203	1.356	1.432	1.477	19	0.979	1.394	1.552	1.630	1.677
20	0.828	1.201	1.353	1.429	1.474	20	0.977	1.391	1.549	1.627	1.674
21	0.827	1.199	1.351	1.427	1.472	21	0.976	1.389	1.546	1.625	1.672
22	0.825	1.197	1.349	1.425	1.470	22	0.974	1.387	1.544	1.623	1.669
23	0.824	1.196	1.347	1.423	1.468	23	0.973	1.385	1.542	1.621	1.667
24	0.823	1.194	1.346	1.421	1.466	24	0.972	1.383	1.540	1.619	1.665
25	0.822	1.193	1.344	1.420	1.465	25	0.971	1.382	1.539	1.617	1.664
30	0.818	1.187	1.338	1.413	1.458	30	0.967	1.370	1.532	1.610	1.656
40	0.814	1.180	1.331	1.406	1.450	40	0.961	1.368	1.524	1.601	1.648
50	0.811	1.176	1.326	1.401	1.446	50	0.958	1.363	1.519	1.596	1.642
100	0.805	1.168	1.318	1.392	1.436	100	0.951	1.354	1.509	1.586	1.632
1000	0.800	1.160	1.310	1.384	1.428	1000	0.946	1.346	1.501	1.577	1.623

Table 4

SIGNIFICANCE LEVEL IMPLIED BY THE UNBIASED CRITICAL POINT

	S = 0.2						S = 0.4				
n\P	1	2	3	4	5	n\P	1	2	3	4	5
5	0.459	0.443	0.432	0.426	0.422	5	0.419	0.390	0.373	0.363	0.356
6	0.459	0.441	0.430	0.423	0.418	6	0.418	0.387	0.368	0.356	0.348
7	0.459	0.440	0.428	0.420	0.414	7	0.418	0.385	0.364	0.351	0.342
8	0.458	0.440	0.427	0.418	0.412	8	0.417	0.384	0.361	0.347	0.337
9	0.458	0.439	0.425	0.416	0.410	9	0.417	0.382	0.359	0.344	0.333
10	0.458	0.439	0.424	0.415	0.408	10	0.416	0.381	0.357	0.342	0.330
11	0.458	0.438	0.424	0.414	0.406	11	0.416	0.381	0.356	0.339	0.327
12	0.458	0.438	0.423	0.413	0.405	12	0.416	0.380	0.354	0.337	0.325
13	0.458	0.438	0.422	0.412	0.404	13	0.416	0.379	0.353	0.336	0.322
14	0.458	0.438	0.422	0.411	0.403	14	0.416	0.379	0.352	0.334	0.320
15	0.458	0.437	0.421	0.410	0.402	15	0.415	0.378	0.351	0.333	0.319
16	0.458	0.437	0.421	0.410	0.401	16	0.415	0.378	0.350	0.331	0.317
17	0.457	0.437	0.421	0.409	0.400	17	0.415	0.378	0.350	0.330	0.316
18	0.457	0.437	0.420	0.409	0.400	18	0.415	0.377	0.349	0.329	0.315
19	0.457	0.437	0.420	0.408	0.399	19	0.415	0.377	0.348	0.329	0.313
20	0.457	0.437	0.420	0.408	0.399	20	0.415	0.377	0.348	0.328	0.312
21	0.457	0.437	0.420	0.408	0.398	21	0.415	0.377	0.347	0.327	0.311
25	0.457	0.436	0.419	0.406	0.396	25	0.415	0.376	0.346	0.325	0.308
30	0.457	0.436	0.418	0.405	0.395	30	0.414	0.375	0.344	0.323	0.305
40	0.457	0.436	0.417	0.404	0.393	40	0.414	0.375	0.343	0.320	0.302
50	0.457	0.436	0.417	0.403	0.392	50	0.414	0.374	0.342	0.318	0.299
100	0.457	0.435	0.416	0.401	0.389	100	0.414	0.373	0.339	0.315	0.295
1000	0.457	0.435	0.415	0.400	0.387	1000	0.414	0.372	0.337	0.311	0.290

Table 5

SIGNIFICANCE LEVEL IMPLIED BY THE UNBIASED CRITICAL POINT

	S = 0.6						S = 0.8				
n\p	1	2	3	4	5	n\p	1	2	3	4	5
5	0.381	0.342	0.322	0.310	0.302	5	0.345	0.301	0.279	0.266	0.257
6	0.379	0.338	0.315	0.301	0.291	6	0.343	0.295	0.270	0.255	0.245
7	0.378	0.335	0.310	0.294	0.283	7	0.341	0.291	0.264	0.247	0.235
8	0.377	0.332	0.305	0.289	0.277	8	0.340	0.287	0.258	0.240	0.227
9	0.377	0.330	0.302	0.284	0.271	9	0.339	0.285	0.254	0.235	0.221
10	0.376	0.329	0.299	0.280	0.266	10	0.338	0.283	0.251	0.230	0.215
11	0.376	0.327	0.297	0.277	0.262	11	0.338	0.281	0.248	0.226	0.211
12	0.375	0.326	0.295	0.274	0.259	12	0.337	0.280	0.245	0.223	0.207
13	0.375	0.325	0.293	0.272	0.256	13	0.337	0.278	0.243	0.220	0.203
14	0.375	0.325	0.292	0.270	0.253	14	0.336	0.277	0.241	0.217	0.200
15	0.375	0.324	0.290	0.268	0.251	15	0.336	0.276	0.239	0.215	0.197
16	0.374	0.323	0.289	0.266	0.249	16	0.336	0.275	0.238	0.213	0.195
17	0.374	0.323	0.288	0.265	0.247	17	0,335	0.275	0.237	0.211	0.192
18	0.374	0.322	0.287	0.263	0.245	18	0,335	0.274	0.235	0.209	0.190
19	0.374	0.322	0.286	0.262	0.244	19	0.335	0.273	0.234	0.208	0.189
20	0.374	0.321	0.286	0.261	0.242	20	0,335	0.273	0.233	0.207	0.187
25	0.373	0.320	0.283	0.257	0.237	25	0.334	0.271	0.230	0.201	0.180
30	0.373	0.319	0.281	0.254	0.233	30	0.333	0.269	0.227	0.198	0.176
40	0.372	0.318	0.278	0.250	0.228	40	0.333	0.268	0.224	0.193	0.170
50	0.372	0.317	0.276	0.247	0.225	50	0.332	0.267	0.222	0.190	0.166
100	0.372	0.316	0.273	0.242	0.218	100	0.332	0.265	0.218	0.184	0.159
1000	0.371	0.314	0.270	0.238	0.212	1000	0.331	0.263	0.214	0.179	0.152

Appendix

Kadane [4] is followed for proving Theorems 4.1 and 4.2. The subscript i in each lemma as well as theorem is for i = 1, ..., N.

Lemma A.1: $\lambda_i = 0_p(1)$ as $\sigma \to 0$.

Proof: $$1 \leq \lambda_i = \min_{\beta_i} \frac{(y - Y_i\beta_i)'\bar{P}_{Z_i}(y - Y_i\beta_i)}{(y - Y_i\beta_i)'\bar{P}_Z(y - Y_i\beta_i)}.$$

However, $y - Y_i\beta_i = Z_i\gamma_i + \sigma Z\eta_i + \sigma u$ from (4.2). Hence

$$1 \leq \lambda_i \leq \frac{(u + Z\eta_i)'\bar{P}_{Z_i}(u + Z\eta_i)}{u'\bar{P}_Z u}. \tag{A.1}$$

QED.

Lemma A.2: For any k-class estimator

$$\begin{pmatrix} \hat{\beta}_i \\ \hat{\gamma}_i \end{pmatrix}_k = \begin{pmatrix} \beta_i \\ \gamma_i \end{pmatrix} + \sigma(X_i' X_i) X_i' (u_i + Z\eta_i) + 0_p(\sigma^2)$$

if $k = 0_p(1)$. [In particular, k = 1 and k = λ]

The proof is straightforward from Lemma 2 of Kadane [4].

Lemma A.3

$$\lambda_i = \frac{(u + Z\eta_i)'\bar{P}_{Z_i}(u + Z\eta_i)}{u'\bar{P}_Z u} + 0_p(\sigma)$$

Proof: $\bar{P}_{Z_i}(y - Y_i\hat{\beta}_i) = \bar{P}_{Z_i}(y - Y_i\hat{\beta}_i - Z_i\hat{\gamma}_i)$

$$= \overline{P}_{Z_i} \{y - Y_i\beta_i - Z_i\gamma_i - \sigma P_{X_i} (u_i + Z\eta_i)\} + 0_p(\sigma^2) \quad \text{(from Lemma A.2)}$$

$$= \sigma\, \overline{P}_{Z_i}\, \overline{P}_{X_i} (u_i + Z\eta_i) + 0_p(\sigma^2) \quad \text{(from (A.1))}$$

$$= \sigma\, \overline{P}_{X_i} (u_i + Z\eta_i) + 0_p(\sigma^2) \ .$$

$$= \sigma\, \overline{P}_{X_i} (u + Z\eta_i) + 0_p(\sigma^2) \ ,$$

since $u_i = u + 0_p(\sigma)$ from (4.4) .

Similarly, $\overline{P}_Z (y - Y_i\hat{\beta}_i) = \sigma\, \overline{P}_Z\, \overline{P}_{X_i} (u + Z\eta_i) + 0_p(\sigma^2)$

$$= \sigma\, \overline{P}_Z u + 0_p(\sigma^2) \ .$$

<u>QED</u>.

<u>Proof of Theorem 4.1</u>:

By Lemma A.3, we have

$$\frac{\lambda_1}{\lambda_2} = \frac{(u + Z\eta_1)'\overline{P}_Z (u + Z\eta_1)}{(u + Z\eta_2)'\overline{P}_{Z_2} (u + Z\eta_2)} + 0_p(\sigma) \ .$$

However,

(A.2) $\quad (u + Z\eta_k)'\overline{P}_{Z_k} (u + Z\eta_k) \sim \chi^2 (T-K_k-G_k|\delta_k^2) \qquad k = 1, 2$,

and $(\overline{P}_{Z_1} - \overline{P}_{Z_2})$ is orthogonal to $\overline{P}_{Z_2}$.

<u>QED</u>.

<u>Proof of Theorem 4.2</u>:

By Lemma A.3,

$$\lambda_1 - \lambda_2 = \frac{(u + Z\eta_1)'\overline{P}_{Z_1} (u + Z\eta_2) - (u + Z\eta_2)'\overline{P}_{Z_2} (u + Z\eta_2)}{u'\overline{P}_Z u} + 0_p(\sigma) \ ,$$

where $\bar{P}_Z$ and $(\bar{P}_{Z_1} - \bar{P}_{Z_2})$ are orthogonal. Also (A.2) holds for each term in the numerator of the ratio. On the other hand,

$$u'\bar{P}_Z u \sim \chi^2(T-K) .$$

QED.

Proof of Theorem 5.1

From (5.2) and (5.3)

$$W_i = E||y^0 - \hat{y}_i||^2$$

$$= E||Z_i(\pi_i - \hat{\pi}_i) + \bar{Z}_i(\bar{\pi}_i - \hat{\bar{\pi}}_i) + \sigma v^0||^2 .$$

Expectations of cross products between any two of three terms in the above equation are zero since v^0 and v are independently distributed, and Z_i and $\bar{Z}_i$ are orthogonal. Then

$$\text{(A.3)} \quad W_i = \sigma^2 E||v^0||^2 + E||Z_i(\pi_i - \hat{\pi}_i)||^2 + E||\bar{Z}_i(\bar{\pi}_i - \hat{\bar{\pi}}_i)||^2 .$$

It is easy to show

$$\sigma^2 E||v^0||^2 = \sigma^2 T \omega .$$

From (2.2), (5.4), and the orthogonality between Z_i and $\bar{Z}_i$,

$$E||Z_i(\pi_i - \hat{\pi}_i)||^2 = \sigma^2 E||P_{Z_i} v||^2 = \sigma^2 K_i \omega .$$

Hereafter we derive the expectation of the third term in (A.3).

From (2.2), (5.3), (5.4), and the orthogonality between Z_i and $\bar{Z}_i$, we have

$$\text{(A.4)} \quad E||\bar{Z}_i(\bar{\pi}_i - \hat{\bar{\pi}}_i)||^2 = E||\sigma P_{\bar{Z}_i} v - P_{\bar{Z}_i}(y - Y_i\hat{\beta}_i)\hat{\rho}_i||^2 .$$

Following the proof of Lemma A.3,

$$(y - Y_i\beta_i)'\bar{P}_Z y = \sigma^2(u + Z\eta_i)'\bar{P}_Z v + 0_p(\sigma^3)$$

$$= \sigma^2 u'\bar{P}_Z v + 0_p(\sigma^3) ;$$

$$(\underset{\sim}{y} - \underset{\sim}{Y}_i\hat{\underset{\sim}{\beta}}_i)'\overline{\underset{\sim}{P}}_Z(\underset{\sim}{y} - \underset{\sim}{Y}_i\hat{\underset{\sim}{\beta}}_i) = \sigma^2\,\underset{\sim}{u}'\overline{\underset{\sim}{P}}_Z\underset{\sim}{u} + O_p(\sigma^3)\ .$$

Then

$$\text{(A.5)}\quad \hat{\rho}_i = \frac{\underset{\sim}{u}'\overline{\underset{\sim}{P}}_Z\underset{\sim}{v}}{\underset{\sim}{u}'\overline{\underset{\sim}{P}}_Z\underset{\sim}{u}} + O_p(\sigma)\ .$$

Similarly following the proof of Lemma A.3,

$$\text{(A.6)}\quad P_{\overline{Z}_i}(\underset{\sim}{y} - \underset{\sim}{Y}_i\hat{\underset{\sim}{\beta}}_i) = \sigma\,P_{\overline{\underset{\sim}{Z}}_i}\overline{\underset{\sim}{P}}_{x_i}(\underset{\sim}{u} + \underset{\sim}{Z}\underset{\sim}{\eta}_i) + O_p(\sigma^2)\ .$$

Using (A.5) and (A.6), (A.4) is

$$\sigma^2E||P_{\overline{\underset{\sim}{Z}}_i}\underset{\sim}{v} - P_{\overline{\underset{\sim}{Z}}_i}\overline{\underset{\sim}{P}}_{X_i}(\underset{\sim}{u} + Z\eta_i)\frac{\underset{\sim}{u}'\overline{\underset{\sim}{P}}_Z\underset{\sim}{v}}{\underset{\sim}{u}'\overline{\underset{\sim}{P}}_Z\underset{\sim}{u}}||^2 + O(\sigma^3)\ .$$

$$\text{(A.7)}\quad = \sigma^2E||P_{\overline{\underset{\sim}{Z}}_i}\{\underset{\sim}{v} - \overline{\underset{\sim}{P}}_{X_i}\underset{\sim}{u}\frac{\underset{\sim}{u}'\overline{\underset{\sim}{P}}_Z\underset{\sim}{v}}{\underset{\sim}{u}'\overline{\underset{\sim}{P}}_Z\underset{\sim}{u}}\} - \{P_{\overline{\underset{\sim}{Z}}_i}\overline{\underset{\sim}{P}}_{X_i}Z\underset{\sim}{\eta}_i\frac{\underset{\sim}{u}'\overline{\underset{\sim}{P}}_Z\underset{\sim}{v}}{\underset{\sim}{u}'\overline{\underset{\sim}{P}}_Z\underset{\sim}{u}}\}||^2 + O(\sigma^3)\ .$$

Expectation of the cross product between the first and the second brackets is zero since only odd moments are included therein. In order to take expectations of squares of the first and the second brackets, we introduce a vector random variable $\underset{\sim}{w}$ which is independent of $\underset{\sim}{u}$.

$$\text{(A.8)}\quad \underset{\sim}{w} = \underset{\sim}{v} - \rho\underset{\sim}{u}\ ,$$

where $\rho = E(\underset{\sim}{u}'\underset{\sim}{v})\,E(\underset{\sim}{u}'\underset{\sim}{u})^{-1}$.

The expectation of the squre of the first bracket in (A.7) is

$$\text{(A.9)}\quad \sigma^2\,E||P_{\overline{\underset{\sim}{Z}}_i}\underset{\sim}{v}||^2 + \sigma^2\,E||P_{\overline{\underset{\sim}{Z}}_i}\overline{\underset{\sim}{P}}_{X_i}\underset{\sim}{u}\frac{\underset{\sim}{u}'\overline{\underset{\sim}{P}}_Z\underset{\sim}{v}}{\underset{\sim}{u}'\overline{\underset{\sim}{P}}_Z\underset{\sim}{u}}||^2 - 2\sigma^2\,E||P_{\overline{\underset{\sim}{Z}}_i}\overline{\underset{\sim}{P}}_{X_i}\underset{\sim}{u}\underset{\sim}{v}'P_{\overline{Z}_i}\frac{\underset{\sim}{u}'\overline{\underset{\sim}{P}}_Z\underset{\sim}{v}}{\underset{\sim}{u}'\overline{\underset{\sim}{P}}_Z\underset{\sim}{u}}||$$

Then we have for the first term of (A.9)

$$E||P_{\overline{\underset{\sim}{Z}}_i}\underset{\sim}{v}|| = \omega(K-K_i)\ ,$$

and for the second term of (A.9),

$$\text{trace } E\,\{P_{\bar{Z}_i}\,\bar{P}_{X_i}\,u\,u'\bar{P}_{X_i}\,P_{\bar{Z}_i}\,\frac{(u'\bar{P}_Z\,v)^2}{(u'\bar{P}_Z\,u)^2}\}$$

$$= \text{trace } E\,\{P_{\bar{Z}_i}\,\bar{P}_{X_i}\,u\,u'\bar{P}_{X_i}\,P_{\bar{Z}_i}\,\frac{(u'\bar{P}_Z\,w)^2}{(u'\bar{P}_Z\,u)^2}\}$$

$$+\,\rho^2\,\text{trace } E\,\{P_{\bar{Z}_i}\,\bar{P}_{X_i}\,u\,u'\bar{P}_{X_i}\,P_{\bar{Z}_i}\} \qquad \text{(from (A.8))}$$

$$= (\omega - \rho^2)\,\text{trace } E\,\{P_{\bar{Z}_i}\,\bar{P}_{X_i}\,u\,u'\bar{P}_{X_i}\,P_{\bar{Z}_i}/u'\bar{P}_Z\,u\}$$

$$+\,\rho^2\,\text{trace }(P_{\bar{Z}_i}\,\bar{P}_{X_i}\,P_{\bar{Z}_i})$$

$$= (\omega - \rho^2)\,\text{trace }\{P_{\bar{Z}_i}\,\bar{P}_{X_i}\,[\frac{1}{T-K}\,\bar{P}_Z + \frac{1}{T-K-2}\,(I - \bar{P}_Z)]\,P_{X_i}\,P_{\bar{Z}_i}\}$$

$$+\,\rho^2\,(K-K_i-G_i)$$

$$= (\frac{\omega-\rho^2}{T-K-2} + \rho^2)\,(K-K_i-G_i)$$

since $\bar{P}_{X_i}\,P_Z = 0$, $E\,\dfrac{uu'}{u'\hat{\bar{P}}_Z\,u} = \dfrac{1}{T-K}\,\bar{P}_Z + \dfrac{1}{T-K-2}\,(I - \bar{P}_Z)$

(A.10) $P_{\bar{Z}_i}\,\bar{P}_{X_i} = P_{\bar{Z}_i} - P_{\bar{Z}_i\Pi_i}$, and $Eww' = (\omega - \rho^2)\,I$.

Finally for the third term of (A.9)

$$E\,||P_{\bar{Z}_i}\,\bar{P}_{X_i}\,u\,v'P_{\bar{Z}_i}\,\frac{u'\bar{P}_Z\,v}{u'\bar{P}_Z\,u}||$$

$$= E\,||P_{\bar{Z}_i}\,\bar{P}_{X_i}\,u\,u'\bar{P}_Z\,w\,w'P_{\bar{Z}_i}/u'\bar{P}_Z u|| + \rho^2\,E\,||P_{\bar{Z}_i}\,\bar{P}_{X_i}\,u\,u'P_{\bar{Z}_i}||$$

$$= \rho^2\,\text{trace }(P_{\bar{Z}_i}\,\bar{P}_{X_i}\,P_{\bar{Z}_i})$$

$$= \rho^2 \; (K-K_i-G_i) \; . \qquad \text{(from (A.10))}$$

Similarly the expectation of square of the second bracket in (A.7) is

$$\text{(A.11)} \; E\left|\left|P_{\bar{Z}_i} \bar{P}_{X_i} Z\eta_i \frac{u'\bar{P}_Z v}{u'\bar{P}_Z u}\right|\right|^2$$

$$= \eta_i' Z' \bar{P}_{X_i} P_{\bar{Z}_i} \bar{P}_{X_i} Z\eta_i \; E\left(\frac{u'\bar{P}_Z v}{u'\bar{P}_Z u}\right)^2$$

$$= \delta_i^2 \left\{\rho^2 + E\left(\frac{u'\bar{P}_Z w}{u'\bar{P}_Z u}\right)^2\right\} \qquad \text{(from (A.8))}$$

$$= \delta_i^2 \left\{\rho^2 + (\omega - \rho^2) \; E\,(u'\bar{P}_Z u)^{-1}\right\} \qquad \text{(from (A.10))}$$

$$= \delta_i^2 \left\{\rho^2 + \frac{\omega - \rho^2}{T-K-2}\right\}$$

since

$$\eta_i' Z' \bar{P}_{X_i} P_{\bar{Z}_i} \bar{P}_{X_i} Z\eta_i = \eta_i' Z' \bar{P}_{X_i} Z \eta_i = \eta_i' (P_{\bar{Z}_i} - P_{\bar{Z}_i\bar{\Pi}_i})\eta_i \qquad \text{(from (A.10))}$$

Combining the above terms, we have

$$W_i = \sigma^2 \left\{T\omega + K\omega + \left[\frac{\omega - \delta^2}{T-K-2} - \rho^2\right] (K-K_i-G_i) + \left[\rho^2 + \frac{\omega - \rho^2}{T-K-2}\right] \delta_i^2\right\} + 0(\sigma^3) \; .$$

$$= \sigma^2 \left\{T\omega + (\omega - \rho^2) \frac{T-K-1}{T-K-2} K + \left[\rho^2 - \frac{\omega - \rho^2}{T-K-2}\right] (K_i+G_i) + \left[\rho^2 + \frac{\omega - \rho^2}{T-K-2}\right] \delta_i^2\right\} + 0(\sigma^3) \; ,$$

Since $r^2 = \rho^2/\omega$, Theorem 5.1 is proved.

QED.

References

[1] Akaike, H., "Information Theory and an Extension of the Maximum Likelihood Principle," Proc. 2nd Info. Symp. on Information Theory, pp. 267-281, 1972.

[2] Anderson, T. W. and H. Rubin, "Estimation of the Parameters of a Single Equation on a Complete System of Stochastic Equations," Annals of Mathematical Statistics, Vol. 20, No. 1 (March, 1949), pp. 46-63.

[3] Hood, W. C. and T. C. Koopmans, Studies in Econometric Method, John Wiley and Sons, New York, 1953.

[4] Kadane, J. B., "Comparison of k-class Estimators when the Disturbances are Small," Econometrica, Vol. 39, No. 5 (Sept., 1971).

[5] Kadane, J. B., "Testing Overidentifying Restrictions when the Disturbances are Small," Journal of the American Statistical Association, (March, 1970), pp. 182-185.

[6] Kadane, J. B., "Testing a Subset of the Overidentifying Restrictions," Econometrica, Vol. 42, No. 5 (Sept., 1974).

[7] Mallows, C. L., "Some Comments on C_p," Technometrics, Vol. 15, No. 4, (November, 1973), pp. 661-675.

[8] Sawa, T., "Information Criteria for the Choice of Regression Models," Working Paper, College of Commerce and Business Administration, University of Illinois at Champaign-Urbana, 1977.

[9] Sawa, T. and K. Takeuchi, "Unbiased Decision Rules for the Choice of Regression Models, Working Paper, College of Commerce and Business Administration, University of Illinois at Champaign-Urbana, 1977.

[10] Takeuchi, K., "Single Equation Method of Prediction in a Simultaneous Equations System," (in Japanese), Economic Review Quarterly, May, 1972, pp. 48-55.