NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

ISBN 0-315-59173-0

The Bootstrap and its Adaptation to Klein's Model I

Pierre Vaillancourt

A Thesis

in

The Department

of

Mathematics

Presented in Partial Fulfillment of the Requirements
of the degree of Master of Science in Mathematics
Concordia University
Montreal, Quebec, Canada

January, 1990

ABSTRACT

The Bootstrap and its Adaptation to Klein's Model I

Pierre Vaillancourt

The bootstrap methodology described by Efron (1979, 1982) is being extensively used to study finite sample properties. In this thesis, we present some empirical results on a well-known econometric model using bootstrap methodology.

ACKNOWLEDGEMENT

I would like to take this opportunity to express my gratitude to Dr. T. D. Dwivedi, whose guidance and assistance were invaluable in the completion of this thesis, and whose creative comments may inspire further possible work.

Table of Contents

Tables and Figures

Tables

Figures

Introduction

After an introduction to the simultaneous linear statistical model which is generally appropriate to econometric phenomena, the thesis will proceed in Chapter I with a brief review of some widely-known estimators now available to estimate the coefficients of these models, as well as asymptotic estimators of their second moments; it will be shown, notably, that the structure of the simultaneous equation model renders the Classical Least Squares (CLS) estimators inconsistent, necessitating the derivation of estimators which tend to eliminate or circumvent the relationship that exists between the endogenous variables and the error terms of the system.

The Monte Carlo approximations of the distribution of small sample statistics, given by resampling methodologies, are often made necessary by the difficulties associated in deriving exactly the distribution and moments of the statistics theoretically. In Chapter II, after describing the general rationale behind resampling schemes, we will describe the bootstrap method in general, and as applied to a very general regression problem, as well as to more specific linear regression problems, including simultaneous linear equation models. On the way, we shall compare the bootstrap to another well-known and much-used resampling method, the (delete-one) Jackknife, as well as other variants of the Jackknife (e.g. delete-r Jackknife).

Finally, in Chapter III, we will describe an application of the bootstrap to a well-known econometric model, Klein's Model I. After briefly describing the model, we will describe our specific choice of bootstrap resampling and estimation techniques, . This will be followed with a description and discussion of our results.

CHAPTER I -
THE LINEAR SIMULTANEOUS EQUATION MODEL AND ESTIMATORS

1. INTRODUCTION

The statistical study of social and econometric phenomena has involved the imposition of a variety of general models on multivariate data sets; the social and econometric milieux which humanity has created and perceives as interrelated necessitates the use of mathematical statistical constructions to express this interrelatedness. One wide variety of such models are the linear regression models; classical linear regression involves a least squares estimate of regression coefficients which assumes the constancy of the independent variables along with other assumptions, which as we shall see, yield inconsistent estimators if applied to models appropriate to econometric phenomena. In effect, the simultaneity of econometric equations, as well as their dynamic nature and the random nature of observations in any given application, make a whole new approach necessary.

The usual estimation technique used in linear regression problems is Classical Least Squares (CLS). In this chapter, we will introduce the simultaneous equation model, showing through a simple example that straightforward application of CLS leads to inconsistent estimates of the regression coefficients, due to the simultaneous nature of the equations in the model. This will be followed by a description of some of the available estimation techniques which deal with the simultaneity problem directly or somehow circumvent it: two -stage least squares

(2SLS), limited information maximum likelyhood (LIML); k-class and double k-class, and h-class estimators as well as a few others.

2. A GENERAL SIMULTANEOUS EQUATION MODEL

Consider a system of linear relationships of the following structural form

$$a_{1m}y_1(t)+..+a_{Mm}y_M(t)+..+b_{1m}x_1(t)+..+b_{Km}x_K(t) + u_m(t) = 0 \qquad (I.1)$$
$$(t=1,2,\ldots,n)$$

where

$$y(t)=\begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_M(t) \end{bmatrix}$$

is the tth joint observation vector of endogenous (dependent) variables; (t=1,2,...,n)

$$\begin{bmatrix} a_{1m} \\ a_{2m} \\ \vdots \\ a_{Mm} \end{bmatrix}$$

is the set of coefficients attributed to the joint observation of the dependent variables, for the mth equation of the system (m=1,2,...,M)

and where

$$x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_k(t) \end{bmatrix}$$

is the tth joint observation vector of the predetermined set of variables. The predetermined variables include the exogenous variables, i.e. those whose values are not to be accounted for within the current system, the assumption being that they are determined by external factors, and the lagged endogenous variables, if any. We shall see later that, depending on whether we consider the model to be dynamic or not, different statistical assumptions will be made, which entail different bootstrapping procedures;

and where

$$\begin{bmatrix} b_{1m} \\ b_{2m} \\ \\ b_{Km} \end{bmatrix}$$

is the set of coefficients attributed to the predetermined variables for the mth equation of the system;

and finally, where $u_m(t)$ is the unobserved value of the disturbance for the tth observation of the mth equation.

Assembling all n observations for the M equations under one single

expression, we can write (I.1) as

$$YA + XB + U = 0 \quad (I.2)$$

where

$$\underset{(nXM)}{Y} = \begin{bmatrix} y'(1) \\ y'(2) \\ \vdots \\ y'(n) \end{bmatrix} \qquad \underset{(MXM)}{A} = \begin{bmatrix} a_{11} & \cdots & a_{1M} \\ a_{21} & & a_{2M} \\ \vdots & & \vdots \\ a_{M1} & \cdots & a_{MM} \end{bmatrix}$$

and where

$$\underset{(nXK)}{X} = \begin{bmatrix} x'(1) \\ x'(2) \\ \vdots \\ x'(n) \end{bmatrix} \qquad \underset{(KXM)}{B} = \begin{bmatrix} b_{11} & \cdots & b_{1M} \\ b_{21} & & b_{2M} \\ \vdots & & \\ b_{K1} & \cdots & b_{KM} \end{bmatrix}$$

and finally where

$$\underset{(nXM)}{U} = \begin{bmatrix} u_1(1) & \cdots & u_M(1) \\ u_1(2) & & u_M(2) \\ \vdots & & \vdots \\ u_1(n) & \cdots & u_M(n) \end{bmatrix} = \begin{bmatrix} u'(1) \\ u'(2) \\ \vdots \\ u'(n) \end{bmatrix}$$

There are as many jointly dependent variables as there are equations: M equations with M unknown dependent variables; hence we may use the predetermined variable realizations (ie. X = x) and the observed errors to evaluate Y. Indeed, for a dynamic model, it is possible to recursively generate successive values of y'(t) from initial values y'(0), the predetermined variable realizations, and the errors.

To actually solve for Y, one can express (I.2) in its reduced form. Assuming A nonsingular, we have

$$\underset{(nXM)}{Y} = \underset{(nXK)}{-X}\ \underset{(KXM)}{BA^{-1}} - \underset{(nXM)}{UA^{-1}} = XC + V \qquad (I.3)$$

Note that

$$C = -BA^{-1}$$

and so, even if we can evaluate (I.3) consistently for C (see I.5), the coefficients in B and A may not be uniquely determined, even given the imposition of numerous constraints on them.

This general model is complemented by a set of assumptions

$$1)\quad E\{u(t)\} = 0 \qquad (t=1,2,\ldots,n)$$

$$2)\quad E\{u(t)u'(t')\} = \underset{(MXM)}{D} \neq 0 \qquad (\text{if } t = t')$$
$$= 0 \qquad (\text{if } t \neq t')$$

where D is a matrix of constants. We note therefore that, across all equations, for a given set of observations, disturbances from different equations may exhibit non-zero correlation; but that disturbances are uncorrelated, across different observations.

It can be shown, in addition, that, if the u(t) are generated by a stationary multivariate stochastic process (SMSP) with dependence between vectors sufficiently weak, we will obtain weak convergence of

the sample disturbance variance-covariance matrix to the population value; i.e.

$$\text{plim } n^{-1}\sum_{t=1}^{n} u(t)u'(t') = D$$

which is assumed to be non-stochastic.

3) Assume also that X is generated by an SMSP and that any dependence in the process again is sufficiently weak so that

a) $\text{plim } \bar{x} = \underset{(kX1)}{\mu}$

$\bar{x} = n^{-1}X'\iota \quad (\iota = \underset{(nX1)}{(1,1,1,\ldots,1)})$

b) defining $S = n^{-1}(X'X - \bar{x}\,\bar{x}')$,
we have $\text{plim } S = D_d$

Note that this will imply that

$$\text{plim } n^{-1} \sum_{t=1}^{n} x(t)x'(t) = D_{XX}$$

this is a matrix of constants, since D_d and μ are constant. We will assume D_{XX} is nonsingular.

4) Assume that

$$E\, x(t)u'(t') = E\, x(t)\; E\, u'(t') = 0$$

for all t and t', and that

$$\underset{}{\text{plim}}\ n^{-1}\ X'U = \underset{t=1}{\text{plim}}\ n^{-1}\sum^{n} x(t)u'(t) = 0$$

The lagged endogenous variables in X are correlated to $u_{t'}$, for t' less than t, but it is tenable to assume that they are not correlated to contemporary disturbances (at t'= t) or succeeding disturbances.

5) We need, however, to add an additional assumption to this set, if there are lagged endogenous variables present. In this case, the reduced form of (I.3) can be rewritten as

$$\underset{(1XM)}{Y_t} = \underset{(1XK)}{X_t}\ \underset{(KXM)}{C} + \underset{(1XM)}{V_t}$$

for time t. Let us now define

$$X_t = \left[\ \underset{(1XM)}{Y_{t-1}}\ \vdots\ \underset{(1X(K-M))}{Z_t}\ \right]$$

where Z_t is the vector of all exogenous variables in the system; then we may write

$$\underset{(1XM)}{Y_t} = \underset{(1XM)}{Y_{t-1}}\ \underset{(MXM)}{C_1} + \underset{(1X(K-M))}{Z_t}\ \underset{((K-M)XM)}{C_2} + \underset{(1XM)}{V_t} \qquad (I.4)$$

Now, using

$$Y_{t-1} = Y_{t-2}C_1 + Z_{t-1}C_2 + V_{t-1}$$

we then have

$$Y_t = (Y_{t-2}C_1 + Z_{t-1}C_2 + V_{t-1})C_1 + Z_tC_2 + V_t$$

$$= Y_{t-2}C_1^2 + Z_{t-1}C_2C_1 + Z_tC_2 + V_{t-1}C_1 + V_t \qquad (I.5)$$

Now, lagging (I.4) repeatedly and substituting back into (I.4), and using $C_1^0 = I$ we will have, after s times,

$$Y_t = Y_{t-s-1}C_1^{s+1} + \sum_{\tau=0}^{s} Z_{t-\tau}C_2C_1^{\tau} + \sum_{\tau=0}^{s} V_{t-\tau}C_1^{\tau} \qquad (I.6)$$

Now, for Y_t to be finite, that is, for stability of the system, we will need

$$\lim_{\tau\to\infty} C_1^{\tau} = 0 \qquad (I.7)$$

In this case,

$$Y_t = \sum_{\tau=0}^{\infty} Z_{t-\tau}\, C_2\, C_1^{\tau} + \sum_{\tau=0}^{\infty} V_{t-\tau}\, C_1^{\tau} \qquad (I.8)$$

a form that will be needed in later derivations. Therefore (I.7) is the additional assumption sought: it ensures the stability of the simultaneous system.

But we may further elaborate on this assumption, defining conditions under which (I.7) applies. In effect, this can be done by studying the inherent dynamic properties of the model. Define

$$E = \sum_{\tau=0}^{\infty} C_2\, C_1^{\tau}$$

which will converge, under condition (I.7). Under conditions of equilibrium, where the exogenous variable vector Z_t is sustained at some constant value $\bar{Z}$, the endogenous variable vector approaches the equilibrium value

$$\bar{Y} = \bar{Z}\, E$$

(this may be seen to be approximately true, if one assumes

$$\sum_{\tau=0}^{\infty} V_{t-\tau} C_1^{\tau} = 0 \;)$$

This can be rewritten as

$$\bar{Y} = \sum_{\tau=0}^{t-1} \bar{Z} C_2 C_1^{\tau} + (\sum_{\tau=0}^{\infty} \bar{Z} C_2 C_1^{\tau}) C_1^{t}$$

$$= \sum_{\tau=0}^{t-1} \bar{Z} C_2 C_1^{\tau} + \bar{Y} C_1^{t} \qquad (I.9)$$

using s=t-1 in (I.6), we may subtract (I.6) from the latter to obtain

$$Y_t^{*} = Y_0^{*} C_1^{t} + \sum_{\tau=0}^{t-1} Z_{t-\tau}^{*} C_2 C_1^{\tau} + \sum_{\tau=0}^{t-1} V_{t-\tau} C_1^{\tau} \qquad (I.10)$$

where $Z_{t-\tau}^{*} = (Z_{t-\tau} - \bar{Z})$ is the deviation from the constant vector $\bar{Z}$ and $Y^{*} = (Y - \bar{Y})$ is the deviation from the corresponding equilibrium value $\bar{Y}$.

Equation (I.10) expresses the time path of the endogenous variables in terms of three components: initial conditions (Y_0^{*}), the time path of exogenous variable deviations from equilibrium $(Z_{t-\tau}^{*})$, and the time path of the disturbances. To study the inherent time path of the endogenous variables, one may reasonably set $Z_{t-\tau}^{*} = 0$ and $V_{t-\tau} = 0$ for all $\tau \geq 0$, which leaves

$$Y_t^{*} = Y_0^{*} C_1^{\tau} \qquad (I.11)$$

The (M X M) matrix C_1 may be expressed as

$$C_1 = P \Lambda Q$$

where $Q = P^{-1}$ and $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_M)$, the diagonal matrix of eigenvalues of C_1 (this is so as long asthere are no multiple eigenvalues); hence

$$C_1^t = P\Lambda P^{-1}P\Lambda P^{-1}P\Lambda P^{-1}\ldots P\Lambda P^{-1} \quad (t \text{ times})$$

$$= P\,\Lambda^t\,P^{-1}$$

where $\Lambda^t = \text{diag}(\lambda_1^t, \lambda_2^t, \ldots, \lambda_M^t)$. We may rewrite C_1^t as

$$C_1^t = \begin{bmatrix} P_1 & P_2 & \ldots & P_M \end{bmatrix} \begin{bmatrix} \lambda_1^t & \cdots & 0 \\ 0 & \lambda_2^t & \\ \vdots & & \vdots \\ 0 & & \lambda_M^t \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_M \end{bmatrix}$$

$$= \sum_{m=1}^{M} \lambda_m^t P_m Q_m = \sum_{m=1}^{M} \lambda_m^t R_{m\,(M\times M)}$$

where P_m is the mth column of P and Q_m is the mth row of Q. Hence (I.11) may finally be written as

$$Y_t^* = Y_0^* \sum_{m=1}^{M} \lambda_m^t R_m$$

From this it is clear that the inherent dynamic properties of the model are determined by the eigenvalues of C_1. The stability condition given in (I.7)

$$\lim_{\tau\to\infty} C_1^\tau = \lim_{\tau\to\infty} \sum_{m=1}^{M} \lambda_m^\tau R_m = 0$$

implies that $|\lambda_m| < 1$, for all m. Hence the assumption under (I.7) necessitates that all eigenvalues of C_1, the matrix of all reduced coefficients for the lagged endogenous variables, be less than 1 in absolute value, for the system to be stable.

We may, from the detailed assumptions (1) through (5) that we have just described, derive properties for the reduced form (eq.I.3), which can be written as

$$Y = XC + V = - XBA^{-1} - UA^{-1}. \qquad (I.12)$$

We may also write

$$EV = - EU\ EA^{-1} = 0$$

and

$$E\ n^{-1}V'V = n^{-1}(A^{-1})'(E\ U'U)A^{-1} = n^{-1}(A^{-1})'D\ A^{-1}$$

since

$$E\ U'U = E \sum_{t=1}^{n} u(t)u(t)' = n\ D$$

This implies that

$$\text{plim}\ n^{-1}V'V = (A^{-1})'D\ A^{-1} = F \quad (F \text{ constant})$$

and

$$\text{plim}\ V = 0$$

also that

$$\text{plim}\ n^{-1}X'V = 0$$

We may hence estimate the matrix C consistently. For the mth equation, we have, from (I.12)

$$\underset{nX1)}{y_m} = \underset{(nXK)}{X}\ \underset{(KX1)}{C_m} + \underset{(nX1)}{v_m}$$

where y_m, C_m and v_m represent the mth columns of thematrices Y, C and V, respectively. The CLS estimate of C is given by

$$\hat{C}_m = (X'X)^{-1}X'y_m$$

and so

$$\hat{C}_m - C_m = (X'X)^{-1}X'v_m$$

and

$$\text{plim } (\hat{C}_m - C_m) = \text{plim } (X'X)^{-1}X'v_m$$

$$= \text{plim } (n^{-1}X'X)^{-1} \text{ plim } n^{-1}X'v_m$$

$$= D_{xx}^{-1} \cdot 0 = 0$$

and so finally

$$\text{plim } \hat{C}_m = C_m \qquad (I.13)$$

with asymptotic covariance matrix

$$\bar{D}_{\hat{C}_m \hat{C}_m} = \bar{E}\,(\hat{C}_m - C_m)(\hat{C}_m - C_m)'$$

$$= n^{-1} \text{ plim } \left[(X'X)^{-1} X' v_m v_m' X (X'X)^{-1}\right]$$

$$= n^{-1} \text{ plim } (n^{-1} X'X)^{-1} \text{ plim } (X'v_m v_m'X) \text{ plim } (n^{-1}X'X)^{-1}$$

$$= n^{-1} D_{xx}^{-1} D_{xx} D_{xx}^{-1} \sigma_{mm} = n^{-1} D_{xx}^{-1} \sigma_{mm}$$

where

$$I_n \sigma_{mm} = \text{plim } n^{-1} (v_m v_m')$$

is the (m,m)th entry of F.

Collecting the results for all M equations, we have

$$\hat{C} = (X'X)^{-1} X' Y$$

which is a consistent estimate of C.

We also note that if one attempts to apply CLS to individual equations of the structural equation system (eq. I.2), one obtains inconsistent results. Taking the mth columns out of A, B and U, we have the structural form for the mth equation.

$$Y A_m + X B_m + u_m = 0$$

Noticing that

$$\underset{(nXM)}{Y} = \left[y_1 \vdots y_2 \vdots \ldots \vdots y_M \right]$$

we may take out say, the first column of Y, and, assigning the value -1 to the topmost entry of A_m, normalize all other entries as a consequence of this; also, we remove from A_m and B_m all coefficients whose values are a priori constrained to 0, at the same time removing from X and the remainder of Y all their associated variables; we then have

$$\underset{(nX1)}{y_1} = \underset{(nXM_1-1)}{Y_1}\ \underset{(M_1-1X1)}{a_1} + \underset{(nXK_1)}{X_1}\ \underset{(K_1X1)}{b_1} + \underset{(nX1)}{u_1} \qquad (I.14)$$

where a_1, b_1 and u_1 are the first columns of A and B respectively, with the zeros removed; and where u_1 is the first column of U. Note also, that Y_1 and X_1 contain only the predetermined and the endogenous variables actually included in the first equation.

Applying CLS to (I.14) directly would yield inconsistent results. To see this, we have only to note that

$$\text{plim } (n^{-1} Y' U) = \text{plim } (n^{-1} C' X' U - (A')^{-1} U' U)$$

using (I.3). This may be written as

$$\text{plim } (n^{-1} C' X' U) - \text{plim } (n^{-1} (A')^{-1} U' U)$$

$$= C' \text{ plim } (n^{-1} X' U) - (A')^{-1} \text{plim } (n^{-1} U' U)$$

$$= 0 - A'^{-1} D \neq 0$$

To see this more clearly, let us examine a simple example. Consider the simultaneous system

$$z_t = a + b\, y_t + u_t \qquad (a)$$

$$y_t = z_t + i_t \qquad (b) \qquad (I.15)$$
$$(1X1) \qquad (t=1,2,\ldots,n)$$

these designate the linear relations at time t; u_t is the unobserved error at t. One could assign econometric definitions to the variables z_t, y_t and i_t but this is immaterial here. One need only note that whereas equation I.15(a) includes an error term, I.15(b) is an exact relation; also, we define i_t as exogenous or predetermined, and z_t and y_t as endogenous.

We assume, applying the assumptions (1) - (4) in the Linear Simultaneous Equation Model, that

1) $E\, u_t = 0$

$$\text{plim} \sum_{t=1}^{n} u_t u_{t'} = E\, u_t u_{t'} = \sigma^2 \quad (t=t')$$
$$= 0 \quad (t \neq t')$$

2) $$\text{plim}\ n^{-1}X'X = D_{xx}$$
$$(2X2)$$

where $X'X = \begin{bmatrix} \iota' \\ i' \end{bmatrix} [\iota' \quad i']$

where $\iota' = (1,1,\ldots,1)$
$(nX1)$

$i' = (i_1, i_2, \ldots, i_n)$

Solving I.15 explicitly for y_t and z_t, the two endogenous variables, we have

$$z_t = a/(1-b) + b/(1-b)\, i_t + 1/(1-b)\, u_t \qquad (a)$$
$$y_t = a/(1-b) + 1/(1-b)\, i_t + 1/(1-b)\, u_t \qquad (b) \qquad (I.16)$$

Note that, since

$$E\, i_t u_{t'} = 0$$

by assumption (3) of the Simultaneous Equation Model,

$$E\, y_t u_{t'} = 1/(1-b)\; E\, u_t u_{t'} = \sigma^2/(1-b), \quad (t=t') \qquad (I.17)$$

Now, if y_t is part of a stationary stochastic process with rapidly dying off dependence, it can be shown (Goldberger 1964) that

$$\text{plim}\; n^{-1} \sum_{t=1}^{n} y_t u_t = E\, y_t u_t = \sigma^2/1-b \qquad (I.18)$$

Using CLS, we may calculate, directly from I.15(a),

$$\hat{b} = \sum_{t=1}^{n} (z_t - \bar{z})(y_t - \bar{y}) \;/\; \sum_{t=1}^{n} (y_t - \bar{y})^2$$

where $\bar{z}$ and $\bar{y}$ are the sample means of z_t and y_t respectively. This can be written as

$$\hat{b} = b + \sum_{t=1}^{n} u_t (y_t - \bar{y}) \;/\; \sum_{t=1}^{n} (y_t - \bar{y})^2$$

since

$$\sum_{t=1}^{n} (z_t - \bar{z})(y_t - \bar{y}) = \sum_{t=1}^{n} (z_t)(y_t - \bar{y})$$

$$= \sum_{t=1}^{n} (a + b\, y_t + u_t)(y_t - \bar{y})$$

and so

$$\text{plim}\; \hat{b} = b + \text{plim} \sum_{t=1}^{n} u_t (y_t - \bar{y}) \;/\; \sum_{t=1}^{n} (y_t - \bar{y})^2$$

Now, by an application of Slutzky's theorem,

$$\text{plim} \sum_{t=1}^{n} u_t (y_t - \bar{y}) \; / \; \sum_{t=1}^{n} (y_t - \bar{y})^2 \quad = \quad (1-b)^{-1}\sigma^2 \; / \; \sigma_y^{\,2}$$

using I.18 and the fact that

$$\text{plim} \sum_{t=1}^{n} (y_t - \bar{y})^2 \; = \; \sigma_y^{\,2}$$

the latter being true under conditions of stationarity of the stochastic process of y_t.

Hence $\text{plim } \hat{b} \neq b$, and so the estimator is not consistent for b.

Hence, in a simultaneous equation model, the non-zero correlation of the endogenous variables with the errors means that the straightforward application of CLS to each equation will yield inconsistent results. We can apply CLS, with resulting consistent estimation, to the reduced form of the system, but as we can see through equation (I.3), the structural coefficients are not necessarily uniquely determined this way. How, then, are we to estimate the latter uniquely and consistently?

In the next section, we will present some widely-known estimation methods which accomplish this.

3. CONSISTENT ESTIMATION TECHNIQUES FOR THE SIMULTANEOUS EQUATION MODELS.

Consistent methods of estimation of the structural coefficients for our simultaneous model break down into two (2) categories; into what Goldberger (1964) calls single equation methods, and into systems methods, respective of whether equations are estimated one at a time, or

whether coefficient estimates are made simultaneously for the full system.

a. Single equation methods

i) The most widely used and computationally simple of the single equation methods is two-stage least squares (2SLS), derived originally and independently by Theil (1953) and Basmann (1957). We provide here a description given to us by Goldberger (1964).

The first equation of the system can be written as in (I.14)

$$\underset{(nX1)}{y_1} = \underset{(nXM_1-1)}{Y_1}\ \underset{(M_1-1X1)}{a_1} + \underset{(nXK_1)}{X_1}\ \underset{(K_1X1)}{b_1} + \underset{(nX1)}{u_1} \qquad (I.19)$$

where we recall that Y_1 and X_1 contain only the variables included in the first equation, and correspondingly, a_1 and b_1 include only the non-zero coefficients.

Before we can proceed any further, we note that the reduced form of the system

$$Y = X C + V$$

can be written as

$$\underset{(nX1)(nXM_1-1)(nXM_2)}{\left[\, y_1 \,:\, Y_1 \,:\, Y_2 \,\right]} = \underset{(nXK_1)(nXK_2)}{\left[\, X_1 \,:\, X_2 \,\right]} \underset{(KX1)(KXM_1-1)(KXM-M_1+1)}{\begin{bmatrix} C_{10} & C_{11} & C_{12} \\ C_{20} & C_{21} & C_{22} \end{bmatrix}} + \left[\, v_1 \,:\, V_1 \,:\, V_2 \,\right]$$

$$\text{where } M_1 + M_2 = M \qquad K_1 + K_2 = K \qquad (I.20)$$

where the subscript "2" in X_2 and Y_2 designates the submatrix of variables excluded from equation 1.

From (I.20), one can extract

$$\underset{(nXM_1-1)}{Y_1} = \underset{(nXK_1)}{X_1}\ \underset{(K_1XM_1-1)}{C_{11}} + \underset{(nXK_2)}{X_2}\ \underset{(K_2XM_1-1)}{C_{12}} + \underset{(nXM_1-1)}{V_1}$$

which can be summarized as

$$Y_1 = X C_1 + V_1 \qquad (I.21)$$

Substituting the above into (I.19), we have

$$y_1 = X C_1 a_1 + X_1 b_1 + (u_1 + V_1 a_1)$$

where the residual term is now $(u_1 + V_1 a_1)$. Hence we obtain an equation containing only predetermined variables on the right-hand side. If we can obtain consistent estimates of C_1, we may hence attempt to estimate consistently this structural equation. Such can be obtained, as demonstrated in (I.13).

Here, using $\hat{C}_1$, the consistent CLS estimate of C_1, we obtain

$$\hat{Y}_1 = X \hat{C}_1$$

This is the first stage of 2SLS. We then apply CLS again, in a second stage, to (I.19), which we write as

$$y_1 = \hat{Y}_1 a_1 + X_1 b_1 + (u_1 + (\hat{Y}_1 - Y_1) a_1)$$

the term in brackets being unobserved errors, since

$$(\hat{Y}_1 - Y_1)$$

is a matrix of residuals, a_1 and u_1 being as yet unknown.

Thus the set of normal equations for the estimation of a_1 and b_1 is

$$\begin{bmatrix} \hat{Y}_1'\hat{Y}_1 & \hat{Y}_1'X_1 \\ X_1'\hat{Y}_1 & X_1'X_1 \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{b}_1 \end{bmatrix} = \begin{bmatrix} \hat{Y}_1'y_1 \\ X_1'y_1 \end{bmatrix} \qquad (I.22)$$

This is equivalent to

$$\begin{bmatrix} \hat{Y}_1'Y_1 & \hat{Y}_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{b}_1 \end{bmatrix} = \begin{bmatrix} \hat{Y}_1'y_1 \\ X_1'y_1 \end{bmatrix} \qquad (I.23)$$

One need only consider

$$\hat{Y}_1'\hat{Y}_1 = Y_1'X_1(X'X)^{-1}X'X(X'X)^{-1}X'Y_1 = \hat{Y}_1'Y_1 \qquad (I.24)$$

and

$$X_1'\hat{Y}_1 = \underset{K_1 \times K_2 \ (n \times K)}{\begin{bmatrix} I_{K_1} & 0 \end{bmatrix}} \begin{bmatrix} X_1' \\ X_2' \end{bmatrix} \begin{bmatrix} X_1 \vdots X_2 \end{bmatrix} \begin{bmatrix} X_1' & X_1 \vdots X_2 \\ X_2' & \end{bmatrix}^{-1} \begin{bmatrix} X_1' \\ X_2' \end{bmatrix} Y_1$$

$$= \begin{bmatrix} I_{K_1} \vdots & 0 \end{bmatrix} \begin{bmatrix} X_1' \\ X_2' \end{bmatrix} Y_1 = X_1'Y_1 \qquad (I.25)$$

to see this.

Now, the form in (I.23) is the set of normal equations if we used

$$\begin{bmatrix} \hat{Y}_1' \\ X_1' \end{bmatrix}$$

as a set of instrumental variables in equation (I.19). Note that

$$\text{plim } n^{-1} \begin{bmatrix} \hat{Y}_1' \\ X_1' \end{bmatrix} u_1 = 0$$

by assumption (4) of the model, and since

$$\text{plim } \hat{Y}_1'u_1 = \hat{C}_1 \text{ plim } X'u_1 = 0$$

Hence the coefficient vectors $\hat{a}_1$ and $\hat{b}_1$ are consistent for a_1 and b

We note in passing that a computationally efficient form of (I.24) is given by

$$\begin{bmatrix} Y_1'X(X'X)^{-1}X'Y_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{b}_1 \end{bmatrix} = \begin{bmatrix} Y_1'X(X'X)^{-1}X'y_1 \\ X_1'y_1 \end{bmatrix} \quad (I.26)$$

The asymptotic variance-covariance matrix of the 2SLS estimate is given by

$$\bar{D}_{\hat{d}\,\hat{d}} = n^{-1} \text{ plim } \left[n^{0.5}(\hat{d}_1 - d_1)\, n^{0.5}(\hat{d}_1 - d_1) \right]$$

where

$$\hat{d}_1 = \begin{bmatrix} \hat{a}_1 \\ \hat{b}_1 \end{bmatrix}$$

This becomes

$$n^{-1} \text{ plim } n \left\{ \begin{bmatrix} \hat{Y}_1'\hat{Y}_1 & \hat{Y}_1'X_1 \\ X_1'\hat{Y}_1 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{Y}_1' \\ X_1' \end{bmatrix} u_1u_1' \begin{bmatrix} \hat{Y}_1 \vdots X_1 \end{bmatrix} \right.$$

$$\left. X \begin{bmatrix} \hat{Y}_1'\hat{Y}_1 & \hat{Y}_1'X_1 \\ X_1'\hat{Y}_1 & X_1'X_1 \end{bmatrix}^{-1} \right\}$$

and by Slutzky's theorem and its consequences, we may write

$$n^{-1} \left\{ \text{plim } n^{-1} \begin{bmatrix} \hat{Y}_1'Y_1 & \hat{Y}_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix} \right\}^{-1} \left\{ \text{plim } n^{-1} \begin{bmatrix} \hat{Y}_1'\hat{Y}_1 & \hat{Y}_1'X_1 \\ X_1'\hat{Y}_1 & X_1'X_1 \end{bmatrix} \right\} X$$

$$\left\{ \text{plim } n^{-1} \begin{bmatrix} \hat{Y}_1'Y_1 & \hat{Y}_1'X_1 \\ X_1'\hat{Y}_1 & X_1'X_1 \end{bmatrix}^{-1} \right\} \sigma^2 \quad (I.27)$$

since

$$\text{plim}\left\{ n^{-1}\begin{bmatrix}\hat{Y}_1' \\ X_1'\end{bmatrix} u_1 u_1' \begin{bmatrix}\hat{Y}_1 \vdots X_1\end{bmatrix}\right\} = \sigma^2 \; D_{Z_1 Z_1}$$

where

$$D_{Z_1 Z_1} = \text{plim}\; n^{-1} Z_1' Z_1$$

$$Z_1 = \begin{bmatrix}\hat{Y}_1 \vdots X_1\end{bmatrix}$$

by the stationarity of the stochastic process for u_t and by the independence of

$$\begin{bmatrix}\hat{Y}_1 \vdots X_1\end{bmatrix} \quad \text{and} \quad u_1$$

The matrix (I.27) is consistently estimated by

$$s^2 \begin{bmatrix}\hat{Y}_1'Y_1 & \hat{Y}_1'X_1 \\ X_1'\hat{Y}_1 & X_1'X_1\end{bmatrix}^{-1} \begin{bmatrix}\hat{Y}_1'\hat{Y}_1 & \hat{Y}_1'X_1 \\ X_1'\hat{Y}_1 & X_1'X_1\end{bmatrix} \begin{bmatrix}\hat{Y}_1'Y_1 & \hat{Y}_1'X_1 \\ X_1'\hat{Y}_1 & X_1'X_1\end{bmatrix}^{-1}$$

$$= s^2 \begin{bmatrix}\hat{Y}_1'Y_1 & \hat{Y}_1'X_1 \\ X_1'\hat{Y}_1 & X_1'X_1\end{bmatrix}^{-1}, \text{ where } s^2 = n^{-1}\left[\sum_{t=1}^{n} u_1(t)^2\right] \qquad \text{(I.28)}$$

One need only take plim of (I.28), noting that

$$\text{plim}\; s^2 = \sigma^2$$

(provided we have a stationary stochastic process with dependence between the $u_1(t)$) rapidly dying off, to see this.

ii) (k)-class, double k-class and and h-estimators:

The CLS estimators, as well as the 2SLS estimators, are part of a family of estimators called the (k)-class of estimators, which we owe to

Theil (1961).

We note first that the matrix of observed errors obtained in the reduced form (I.21) can be obtained from

$$Y_1 = X\hat{C}_1 + \hat{V}_1 = \hat{Y}_1 + \hat{V}_1$$

or

$$\hat{V}_1 = (Y_1 - \hat{Y}_1)$$

We note also that

$$\hat{V}_1'\hat{V}_1 = (Y_1 - \hat{Y}_1)'(Y_1 - \hat{Y}_1)$$
$$= Y_1'Y_1 - \hat{Y}_1'\hat{Y}_1$$

since

$$\hat{Y}_1'Y_1 = \hat{Y}_1'\hat{Y}_1 = Y_1'\hat{Y}_1$$

and so we may write (I.23) as

$$\begin{bmatrix} Y_1'Y_1 - \hat{V}_1'\hat{V}_1 & \hat{Y}_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{b}_1 \end{bmatrix} = \begin{bmatrix} (Y_1' - \hat{V}_1')\, y_1 \\ X_1'y_1 \end{bmatrix} \quad (I.29)$$

using (I.25).

We note finally that a straightforward CLS estimation of the structural equation (I.19) would yield

$$\begin{bmatrix} Y_1'Y_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix} \begin{bmatrix} \hat{a}_1^* \\ \hat{b}_1^* \end{bmatrix} = \begin{bmatrix} Y_1'y_1 \\ X_1'y_1 \end{bmatrix} \quad (I.30)$$

The only difference between (I.29) and (I.30), which allows consistent estimation of the structural coefficients, is the inclusion of the functions of the residual matrix $\hat{V}_1$.

Equations (I.29) and (I.30) may be written as

$$\begin{bmatrix} Y_1'Y_1 - k\hat{V}_1'\hat{V}_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{b}_1 \end{bmatrix} = \begin{bmatrix} (Y_1' - k\hat{V}_1')\, y_1 \\ X_1'y_1 \end{bmatrix} \quad (I.31)$$

with $k = 0$ for CLS, and $k = 1$ for 2SLS. Having thus defined a family of estimators, called the k-class of estimators, can we determine what conditions are needed for the possible members of this family to be consistent estimators of coefficients?

We may study some asymptotic properties of (I.31), to illustrate the conditions under which its distribution is the same as that of the 2SLS estimator in (I.29). To derive the sampling error of the estimator $\begin{bmatrix} \hat{a}_1 \\ \hat{b}_1 \end{bmatrix}$ in (I.31), let us first rewrite the RHS of that equation as Theil (1971) did

$$\begin{bmatrix} (Y_1' - k\hat{V}_1') \\ X_1' \end{bmatrix} \left\{ \begin{bmatrix} Y_1 \vdots X_1 \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} + u_1 \right\}$$

$$= \begin{bmatrix} Y_1'Y_1 - k\hat{V}_1'Y_1 & Y_1'X_1 - k\hat{V}_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} + \begin{bmatrix} (Y_1' - k\hat{V}_1') \\ X_1' \end{bmatrix} u_1$$

$$= \begin{bmatrix} Y_1'Y_1 - k\hat{V}_1'\hat{V}_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} + \begin{bmatrix} (Y_1' - k\hat{V}_1') \\ X_1' \end{bmatrix} u_1 \quad (I.32)$$

This is so, since

$$\hat{V}_1'Y_1 = \hat{V}_1'(\hat{Y}_1 + \hat{V}_1)$$

$$= \hat{V}_1'(X(X'X)^{-1}X'Y_1 + \hat{V}_1)$$

$$= \hat{V}_1'\hat{V}_1$$

and since

$$\underset{(M_1-1Xn)(nXk)}{\hat{V}_1'\ X} = \underset{(M_1-1Xk)}{0}$$

Now we may write (I.31) as

$$\begin{bmatrix}\hat{a}_1\\ \hat{b}_1\end{bmatrix} = \begin{bmatrix} Y_1'Y_1 - k\hat{V}_1'\hat{V}_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1\end{bmatrix}^{-1}\begin{bmatrix}(Y_1' - k\hat{V}_1')y_1 \\ X_1'y_1\end{bmatrix} \quad (I.33)$$

We may then write the sampling error of d_k , combining (I.32) and (I.33)

$$\begin{bmatrix}\hat{a}_1\\ \hat{b}_1\end{bmatrix} - \begin{bmatrix}a_1\\ b_1\end{bmatrix} = \begin{bmatrix} Y_1'Y_1 - k\hat{V}_1'\hat{V}_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1\end{bmatrix}^{-1}\begin{bmatrix}(Y_1' - k\hat{V}_1') \\ X_1'\end{bmatrix}u_1 \quad (I.34)$$

In this form, it is possible to derive some asymptotic properties of the k-class estimators, by means of a comparison to the 2SLS estimator. First, we write

$$n^{-1}\begin{bmatrix} Y_1'Y_1 - k\hat{V}_1'\hat{V}_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1\end{bmatrix}$$

$$= n^{-1}\begin{bmatrix} Y_1'Y_1 - \hat{V}_1'\hat{V}_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1\end{bmatrix} - (k-1/n)\begin{bmatrix}\hat{V}_1'\hat{V}_1 & 0\\ 0 & 0\end{bmatrix} \quad (I.35)$$

Now, the term

$$n^{-1}\hat{V}_1'\hat{V}_1 = n^{-1}(Y_1 - \hat{Y}_1)'(Y_1 - \hat{Y}_1)$$

$$= Y_1'M\,M\,Y_1\,n^{-1} = Y_1'M\,Y_1\,n^{-1}$$

$$= Y_1'Y_1\,n^{-1} - (n^{-1}Y_1'X)(n^{-1}X'X)(n^{-1}X'Y_1)$$

converges to a finite probability limit, by the assumptions of the

Linear Simultaneous Equation Model. And so, if plim(k - 1) = 0, we will have

$$\text{plim } n^{-1} \begin{bmatrix} Y_1'Y_1 - k\,\hat{V}_1'\hat{V}_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix}^{-1}$$

$$= n^{-1} \begin{bmatrix} Y_1'Y_1 - \hat{V}_1'\hat{V}_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix}^{-1}$$

Again using the condition plim (k-1) = 0 and the above result, it can finally be shown that the entire expression (I.34) has the same probability limit as that of the sampling error for the 2SLS estimator, which has value 0. Hence, under the condition

$$\text{plim}(k - 1) = 0 \quad \text{(I.36)}$$

The k-class estimators are shown to have the sameasymptotic properties as the 2SLS estimator, and hence to be consistent.

Another currently used modification of the 2SLS method is the following. Recall that consistent estimates of structural coefficients were obtained in the 2SLS procedure by substituting $\hat{Y}_1$ for Y_1 in

$$y_1 = Y_1 a_1 + X_1 b_1 + u_1$$

(equation I.19), and applying CLS to the new equation; note also that

$$\hat{Y}_1 = Y_1 - \hat{V}_1 .$$

We may create a new family of estimators, using

$$\hat{Y}_1 = Y_1' - h\,\hat{V}_1'$$

instead; (note here also that, if k = 0, the CLS case results, whereas if k = 1 , we return to the 2SLS estimator); the estimator here will be

$$\begin{bmatrix} (Y_1'-h\hat{V}_1')(Y_1-h\hat{V}_1) & (Y_1'-h\hat{V}_1')X_1 \\ X_1'(Y_1-h\hat{V}_1) & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} (Y_1'-h\hat{V}_1')y_1 \\ X_1'y_1 \end{bmatrix}$$

$$= \begin{bmatrix} Y_1'Y_1 - 2h\, Y_1'M\, Y_1 + h^2\, Y_1'M\, Y_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} (Y_1' - h\,\hat{V}_1')y_1 \\ X_1'\, y_1 \end{bmatrix}$$

where

$$M = I_n - X\,(X'X)^{-1}X'$$

This becomes

$$\begin{bmatrix} Y_1'Y_1 - (2h-h)^2\,\hat{V}_1'\hat{V}_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} (Y_1' - h\,\hat{V}_1')y_1 \\ X_1'\, y_1 \end{bmatrix} \quad (I.37)$$

since

$$Y_1'M\,Y_1 = Y_1'MM\,Y_1 = (Y_1 - \hat{Y}_1)'(Y_1 - \hat{Y}_1)$$

Consider now (I.33) and (I.37): Theil (1971) suggests that the k-and h-class estimators are special cases of what he and Nagar call the double-k-class of estimators, which can be written as

$$\begin{bmatrix} Y_1'Y_1 - k_1\,\hat{V}_1'\hat{V}_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix} \begin{bmatrix} \bar{a}_1 \\ \bar{b}_1 \end{bmatrix} = \begin{bmatrix} (Y_1' - k_2\,\hat{V}_1') \\ X_1' \end{bmatrix} y_1 \quad (I.38)$$

Note that if $k_1 = k_2 = 0$, the CLS estimator results; and we have 2SLS if $k_1 = k_2 = 1$. Also, if $k_1 = k_2 = k$, we have the k-class, and the h-class if $k_1 = 2h - h^2$, $k_2 = h$.

There exists a relation which links the double-k-class to k-class, CLS, 2SLS estimators, given to us by Dwivedi. We give here the

description made by Donatos (1984).

Note first that we may write

$$y_1 = Y_1 a_1 + X_1 b_1 + u_1 = Z_1 d_1 + u_1$$

where

$$Z_1 = \left[Y_1 \vdots X_1 \right] \qquad d_1 = \begin{bmatrix} a_1 \\ b_1 \end{bmatrix}$$

(I.38) may then be expressed as

$$\begin{bmatrix} \bar{a}_1 \\ \bar{b}_1 \end{bmatrix} = \begin{bmatrix} Y_1'Y_1 - k_1 Y_1'MY_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} (Y_1' - k_2 (MY_1)')y_1 \\ X_1'y_1 \end{bmatrix}$$

$$= \begin{bmatrix} Y_1'(I_n - k_1M)Y_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix}^{1} \begin{bmatrix} Y_1'(I_n - k_2 M)y_1 \\ X_1'y_1 \end{bmatrix}$$

$$= \left\{ \begin{bmatrix} Y_1' \\ X_1' \end{bmatrix} (I_n - k_1M) \left[Y_1 \vdots X_1 \right] \right\}^{-1} \left\{ \begin{bmatrix} Y_1' \\ X_1' \end{bmatrix} (I_n - k_2 M) \right\} y_1$$

$$\left[\text{since } X_1'(I - k_1M) = X_1' - k_1 \left[I \vdots 0 \right] X'(I - X(X'X)^{-1}X') = X_1' \right]$$

$i=1,2$

This last expression is equal to

$$\left[Z_1'(I_n - k_1M)Z_1 \right]^{-1} Z_1'(I_n - k_2M)y_1 \qquad (I.39)$$

Now, we may write

$$\left[Z_1'(I_n - k_1M)Z_1 \right]^{-1} = \left[Z_1'Z_1 - k_1 Z_1'M Z_1 \right]$$

$$= \left[Z_1'Z_1 - k_1 Z_1'Z_1 + Z_1'M_s Z_1 \right] \qquad (I.40)$$

equating

$$M = I - X(X'X)^{1} X' = I - M_s$$

Now, (I.40) can then be written

$$\left[(1 - k_1) Z_1'Z_1 + k_1 Z_1'M_s Z_1 \right]^{-1}$$

This latest expression is equivalent to

$$k_1^{-1}\left\{ I - (1 - k_1)\left[Z_1'(I_n - k_1M)Z_1\right]^{-1}Z_1'Z_1\right\}(Z_1'M_sZ_1)^{-1}$$

$$= k_1^{-1}(Z_1'M_sZ_1)^{-1} - (1 - k_1)k_1^{-1}\left[Z_1'(I_n - k_1M)Z_1\right]^{-1}Z_1'Z_1(Z_1'M_sZ_1)^{-1} \qquad (I.41)$$

(Note that we used the general relation

$$(Q_1 + Q_2)^{-1} = \left[I - (Q_1 + Q_2)^{-1}Q_2\right]Q_1^{-1}$$

which is valid whenever $(Q_1 + Q_2)^{-1}$ and Q_1^{-1} exist.)

Taking the second matricial component of (I.39), we may transform it also

$$Z_1'(I_n - k_2M)y_1 = k_1Z_1'M_sy_1 - (1 - k_1)Z_1'M_sy_1 + (1 - k_2)Z_1'M\,y_1$$

$$= k_1Z_1'M_sy_1 - (1 - k_1)Z_1'M_sZ_1\hat{d}_{1(2sls)} + (1 - k_2)Z_1'M\,y_1 \qquad (I.42)$$

In the above, we have used

$$\hat{d}_{1(2sls)} = (Z_1'M_sZ_1)^{-1}Z_1'M_sy_1$$

which is the 2SLS estimator derived from (I.39), with $k_2 = k_1 = 1$.

Finally, we may multiply (I.41) and (I.42) to obtain the double k-class estimator in its new form.

$$\hat{d}_{1(k_1,k_2)} = \hat{d}_{1(2sls)} + \left[Z_1'(I_n - k_1M)Z_1\right]^{-1}\left[(1 - k_2)Z_1'M\,y_1 - (1 - k_1)Z_1'M_sZ_1\hat{d}_{1(2sls)}\right] \qquad (I.43)$$

We may also write the k-class estimator from (I.39) using $k_2 = k_1 = k$ Hence

$$d_{1(k)} = \left[Z_1'(I_n - k\,M)Z_1\right]^{-1}Z_1'(I_n - k\,M)y_1$$

With manipulations analogous to those leading to (I.41), it can be

shown that

$$\hat{d}_{1(k_1,k_2)} = d_{1(k)} - (k_2 - k_1)\left[Z_1'(I_n - k_1M)Z_1\right]^{-1} Z_1'My_1 \quad (I.44)$$

We can derive, finally, a relation given to us by Dwivedi (1981), which links the double (k)-class, (k)-class, 2SLS and CLS estimators.

First, we observe that

$$\left[Z_1'(I_n - k_1M)Z_1\right]^{-1} Z_1' My_1 = \left[Z_1'(I_n - k_1M)Z_1\right]^{-1} Z_1'y_1$$

$$- k_1^{-1}\left\{I_n - (1 - k_1)\left[Z_1'(I_n - k_1M)Z_1\right]^{-1} Z_1'Z_1\right\}(Z_1'M_sZ_1)^{-1}Z_1'M_sy_1$$

the latter equality is obtained through (I.41). In turn, this can be written as

$$\left[Z_1'(I_n - k_1M)Z_1\right]^{-1} Z_1'y_1 - k_1^{-1}\left\{I_n - (1 - k_1)\left[Z_1'(I_n - k_1M)Z_1\right]^{-1} Z_1'Z_1\right\} \hat{d}_{1(2sls)} \quad (I.45)$$

We may also rewrite the first term on the RHS of (I.45) as

$$\left[Z_1'(I_n - k_1M)Z_1\right]^{-1} Z_1'y_1 =$$

$$= (1 - k_1)^{-1}\left[I_n - \left[Z_1'(I_n - k_1M)Z_1\right]^{-1} k_1Z_1'M_sZ_1\right] (Z_1'Z_1)^{-1}Z_1'y_1 \quad (I.46)$$

$$= (1 - k_1)^{-1}\left[I_n - \left[Z_1'(I_n - k_1M)Z_1\right]^{-1} k_1Z_1'M_sZ_1\right] \hat{d}_{1(CLS)}$$

where $\hat{d}_{1(CLS)}$ is the CLS estimator.

Finally, we may assemble (I.44)and (I.46) into the final form.

$$\hat{d}_{1(k_1,k_2)} = d_{1(k)} - (1-k_1)^{-1}(k_2-k_1)\Big\{ I_n - \Big[Z_1'(I_n - k_1M)Z_1\Big]^{-1} X \; k_1 Z_1' M_s Z_1 \Big\} \hat{d}_{1(CLS)} + k_1^{-1}(k_2-k_1)\Big\{ I - (1-k_1) X \; \Big[Z_1'(I_n - k_1M)Z_1\Big]^{-1} Z_1'Z_1 \Big\} \hat{d}_{1(2SLS)} \qquad (I.47)$$

(iii) Limited information maximum likelyhood

The methods in (i) and (ii) are non parametric. We will now describe a parametric method, which was developed prior to the 2SLS and k-class methods by Anderson and Rubin (1949,1950); we summarize here the description given to us by Goldberger (1964) of this method. In addition to the assumptions made on the structural disturbances, which the 2SLS and k-class estimators utilise, the LIML method assumes that the structural disturbances are normally distributed.

Let us first rewrite equation (I.19) as

$$Y_1^* a_1^* + X_1 b_1 + u_1 = 0$$

where

$$Y_1^* = \left[y_1 \vdots Y_1 \right]$$

and

$$a_1^* = \begin{bmatrix} -1 \\ a_1 \end{bmatrix} \begin{matrix} (1X1) \\ (M_1 - 1 \; X \; 1) \end{matrix}$$

We also note that

$$Y_1^* = X C_1^* + V_1^*$$

is the reduced form which includes all the endogenous variables in equation 1. It is obtained from (I.20), with

$$C_1^* = \underset{(KXM_1)}{\begin{bmatrix} C_{10} & C_{11} \\ C_{20} & C_{21} \end{bmatrix}} \qquad V_1^* = \underset{(nXM_1)}{\begin{bmatrix} y_1 & \vdots & V_1 \end{bmatrix}}$$

We seek, in LIML, to minimize the generalized variance of the residuals V_1^* (analogous to CLS), but with restrictions applied to the structural coefficients. In fact, we minimize

$$z = 0.5 \log | W | - \mu' C_{21}^* a_1^* \qquad (I.48)$$

where

$$C_{21}^* = \begin{bmatrix} C_{20} & \vdots & C_{21} \end{bmatrix}$$

and

$$W = n^{-1} (V_1^*)' V_1^*$$

and finally where μ is a $(k_2 \text{ X } 1)$ set of Lagrange multipliers.

We can show that the resulting estimator can be written as

$$\hat{d}_{1(LIML)} = \begin{bmatrix} Y_1'Y_1 - \hat{l}\,\hat{V}_1'\hat{V}_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} (Y_1' - \hat{l}\,\hat{V}_1') \\ X_1' \end{bmatrix} y_1 \qquad (I.49)$$

which is a (k)-class estimator with $k = \hat{l}$; it is computed through a very involved series of calculations, from (I.48) (see Goldberger (1964)).

It can also be shown that plim $(\hat{l} - 1) = 0$ which insures that (I.49) is consistent.

b) System methods

Single equation methods omit part of all the information available in the simultaneous system and so are expected to yield more variable results. A review of the methods given in part a) reveals that, in estimating one equation, we only draw on part of the information available in the remainder of the set of equations: in effect, we only draw on them to tell us which of the variables are the excluded predetermined variables. We have only to consider equation (I.20) to see this in 2SLS, for example: we see clearly in this method that there is no use made of the fact that the matrix of excluded endogenous variables Y_2 exists and that its components are used in other parts of the system. One can show, in fact, that asymptotically, full system methods such as 3SLS (Three Stage Least Squares) and FIML (Full Information Maximum Likelyhood) display less theoretical variability. But, as they are computationally very involved, their use is limited, for the moment, and so we have not attempted to adapt them to the resampling and estimation method used in arriving at the main results of this paper, the bootstrap.

We shall not therefore develop the concept of full system methods for the consistent estimation of simultaneous linear systems, here. We refer the reader to Goldberger (1964), Theil (1971), Johnston (1984) for further information on this topic

We thus end our review of the consistent simultaneous equation structural coefficient estimators which are available. We will now proceed to a brief description of the second major set of statistical tools needed for the computation of the results of this thesis, which are the experimental methods designed to study finite sample properties of

chosen estimators. These include the traditional Monte Carlo methods, as well as methods involving resampling of the data. Chapter II will develop this topic, specifically focusing on resampling methods, in particular the bootstrap and the jackknife, and even more specifically, as these apply to the study of finite sample properties of regression coefficients.

CHAPTER II

THE STUDY OF FINITE SAMPLE PROPERTIES

THE BOOTSTRAP AND THE JACKKNIFE

1. INTRODUCTION

Estimators obtained using the existing methods for simultaneous equation models are only consistent. It is well-known that consistency is a large sample property of estimators, and does not apply to small sample behaviour. Dwivedi and Srivastava (1984) have tried to study the finite sample properties of these estimators; several Monte Carlo studies have also been executed towards this end; Johnston (1972) and Donatos (1984) provide some review of Monte Carlo results; see in particular Johnston (1972, pp. 408-420) for a clear, alhough somewhat dated, review. Recently, computer-based methodologies such as the Jackknife and the Bootstrap have been developed, which can be used in this respect. In this Chapter, we describe the two above-mentionned methodologies in some detail. We shall describe the basic concepts with some examples, and show some applications to regression problems.

2. THE BOOTSTRAP METHOD

The bootstrap as described by Efron (1979a, 1982a, 1983, 1986) is basically used as a method to estimate the distribution of any estimate $\hat{\theta}$ of a parameter θ by resampling from the original sample. The principle is as follows.

Let a random sample of size n be chosen from a completely unspecified probability distribution F, i.e. choose $X_i = x_i$, X_i is iid F-distributed (i=1,2,...,n).

In any application F is a distribution on R^k (k=1,2,...m). We let

$$X = (X_1, X_2, \ldots, X_n)$$

and

$$x = (x_1, x_2, \ldots, x_n)$$

denote the random sample and its observed realization, respectively.

The problem to be solved can be described as follows: given a specified random variable R(X,F) which may depend on both X and F, we must estimate the sampling distribution of R = R(X,F), using the realization X = x.

In practice, we proceed in the following manner:

1) Construct the sample probability distribution $\hat{F}$, by assigning the probability mass 1/n at each point $(x_1, x_2, \ldots, x_n)$. (II.1a)

2) With $\hat{F}$ fixed, draw a random sample of size n from $\hat{F}$, say $X_i^* = x_i^*$, where X_i^* is iid $\hat{F}$-distributed (i=1,2,...,n). (II.1b)

This is called the bootstrap sample; it is of the same size n as the original sample, permitting comparison between R and R^* (see below). Note also that x_i^* is selected with replacement from the set $(x_1, x_2, \ldots, x_n)$

3) Finally, we may calculate

$$R^* = R(X^*, \hat{F}) \qquad \text{(II.1c)}$$

for each resample realization $X^* = x^*$. The values of R^* may then be

plotted into the bootstrap distribution, which is meant to approximate the sampling distribution of R. R^* may be computed an arbitrary number of times, to suit our needs. We may also compute the bootstrap estimate of R's standard deviation, SD.

The basic premises and hopes of the procedure in (1) - (3) are the following. The distribution of R^* which in theory can be evaluated exactly once we have X = x equals the desired distribution of R if $F = \hat{F}$. Any non-parametric estimator of R's distribution (that is, one that provides a good estimation with no prior restrictions on the form of F), must give an approximately right answer when $F = \hat{F}$, since $\hat{F}$ is in Efron's words (1979a)

"...a central point amongst the class of likely F's."

In fact, it can be shown that $\hat{F}$ is the maximum likelyhood estimate of F.

A simple example, drawn from Efron (1982), is appropriate here. Consider

$$F = \begin{cases} Pr_F(X = 1) = p = \theta(F) \\ \\ Pr_F(X = 0) = 1 - p \end{cases}$$

We may seek to approximate the sampling distribution of

$$R(X,F) = \bar{X} - \theta(F)$$

where

$$\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$$

Having observed X = x, the bootstrap sample

$$X^* = (X_1^*, X_2^*, \ldots, X_n^*)$$

has each component independently equal to 1 with probability $\bar{x} = \theta(\hat{F})$ and equal to 0 with probability $(1 - \bar{x})$.

We may then approximate the sampling distribution of R by its bootstrap equivalent

$$R(X^*, \hat{F}) = \bar{X}^* - \bar{x}$$

where

$$\bar{X}^* = n^{-1} \sum_{i=1}^{n} X_i^*$$

Representing the value of the population mean and variance under $\hat{F}$ by E_* and Var_*, we may ascertain the distribution of $\bar{X}^*$ to be

$$\binom{n}{k} (1 - \bar{x})^{n-k} (\bar{x})^k$$

the distribution of $(\bar{X}^* - \bar{x})$ will be the same, but with mean

$$E_*(\bar{X}^* - \bar{x}) = \bar{x} - \bar{x} = 0$$

and variance

$$Var_*(\bar{X}^* - \bar{x}) = n^{-1} \bar{x} (1 - \bar{x}) \qquad \text{(II.2)}$$

Hence one may, applying the above information on R^* to R, conclude that $\bar{X}$ is unbiased for $\theta(F)$, with variance approximately equal to (II.2).

A second more elaborate example from Efron (1979a) will now be given which will clearly illustrate the rationale for the bootstrap. Let the sample space Θ (from which we draw $X = x$) be itself a finite set $\theta = \{1,2,\ldots,L\}$ and let $\{X_i = l\}$; that is, the sample random variables X_i are equal to any l, $(1 \leq l \leq L)$.

The multinomial representation becomes appropriate here. Hence the distribution F can be represented by the vector of probabilities

$$f = (f_1, f_2, \ldots, f_L)$$

$$f_l = \text{Prob}_F(X_i = l) \qquad \text{(II.3)}$$

For a given random sample X, let

$$\hat{f}_l = \#\{X_i = l\}/\, n$$

and

$$\hat{f} = (\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_L)$$

Now if R(X,F) is defined such that it is invariant under permutations of the components of X, it is easy to see that we can simply write it as a function of $\hat{f}$ and f : i.e., define

$$R(X,F) = Q(\hat{f}, f) \qquad \text{(II.4)}$$

Obviously, since we are using $\hat{F}$ as population distribution in the resampling, we will have

$$R(X^*, \hat{F}) = Q(\hat{f}^*, f) \qquad \text{(II.5)}$$

where

$$\hat{f}_l^* = \#\{X_i^* = l\}/\, n$$

and

$$\hat{f}^* = (\hat{f}_1^*, \hat{f}_2^*, \ldots, \hat{f}_L^*) \qquad \text{(II.6)}$$

Bootstrap methods will therefore be used to estimate the sampling distribution of $Q(\hat{f}, f)$, given the true distribution of f, by the conditional distribution of $Q(\hat{f}^*, \hat{f})$, given the observed value of $\hat{f}$. This is plausible; the respective conditional distributions of $\hat{f}$ and $\hat{f}^*$ are, respectively

$$\hat{f}/f \sim \text{Multinomial}_L(n, f) \qquad \text{(II.7a)}$$

i.e. $\hat{f}$ has an L-category multinomial distribution, and

$$\hat{f}^*/\hat{f} \sim \text{Multinomial}_L(n, \hat{f}) \qquad \text{(II.7b)}$$

In large samples, we would expect $\hat{f}$ to converge to f, so that (II.7a) and (II.7b) should imply the approximate validity of the bootstrap in its attempt to estimate the sampling distribution of $Q(\hat{f},f)$.

The actual asymptotic validity of the bootstrap is easy to verify within this set up, but we have to first assume the following. Let

$$Q(f,f) = 0$$

$$u(\hat{f}^*,\hat{f}) = \begin{bmatrix} \partial Q(\hat{f}^*,\hat{f})/\partial \hat{f}^*_1 \\ \partial Q(\hat{f}^*,\hat{f})/\partial \hat{f}^*_2 \\ \vdots \\ \partial Q(\hat{f}^*,\hat{f})/\partial \hat{f}^*_L \end{bmatrix}$$

if these partial derivatives exist in an open neighbourhood of (f,f). Also, let $u(f,f) \neq 0$.

We may express (II.4) and (II.5) as

$$\begin{aligned} Q(\hat{f},f) &= Q(\hat{f},f)\Big|_{\hat{f}=f} + (\hat{f} - f)(\mu + e_n) \\ &= (\hat{f} - f)(\mu + e_n) \end{aligned} \qquad \text{(II.8a)}$$

and analogously

$$Q(\hat{f}^*,\hat{f}) = (\hat{f}^* - \hat{f})(\mu + \hat{e}_n) \qquad \text{(II.8b)}$$

this was obtained by expanding by a Taylor polynomial, with corresponding remainders e_n and $\hat{e}_n$; and by showing, through a multinomial version of Borel's Law of Large Numbers, that both $\hat{f}$ and $\hat{f}^*$ converge strongly, that is, with probability one, to f. We do not display this proof here.

Also, using (II.7a) and (II.7b) and by a multinomial version of the De Moivre-Laplace law, and also by the above-stated fact that $\hat{f}$ converges to f with probability one, we will finally be able to adopt the following convergence result. In effect, we will have that

$$n^{0.5}(\hat{f} - f)/f \xrightarrow{p} N_L(0,\Sigma_f) \qquad (II.9a)$$

and that

$$n^{0.5}(\hat{f}^* - \hat{f})/\hat{f} \xrightarrow{p} N_L(0,\Sigma_f) \qquad (II.9b)$$

i.e. weak convergence exists to an identical multivariate normal.

Here, the variance matrix equals

$$\Sigma_f = \begin{bmatrix} f_1(1-f_1) & -f_1f_2 & \cdots & -f_1f_L \\ -f_2f_1 & f_2(1-f_2) & & \vdots \\ & & & f_L(1-f_L) \end{bmatrix}$$

Finally, using (II.8a and b) as well as (II.9a and b), and utilising the fact that e_n and $\hat{e}_n$ both converge to 0 with probability 1, we conclude that

$$n^{0.5}\, Q(\hat{f},f)/f = n^{0.5}(\hat{f} - f)\,(\mu + e_n)/f \xrightarrow{a,s} N_L(0,\ \mu\Sigma_f\mu)$$

and that

$$n^{0.5}\, Q(\hat{f}^*,\hat{f})/\hat{f} = n^{0.5}(\hat{f}^* - \hat{f})\,(\mu + \hat{e}_n)/\hat{f} \xrightarrow{a,s} N_L(0,\ \mu\Sigma_f\mu)$$

that is, that the conditional bootstrap distribution and the sampling distribution of Q converge to the same normal distribution.

The example just described depicted an ideal utilization of the bootstrap. Although such an ideal situation rarely exists in practice, it clearly points to a basic premise needed by the bootstrap, that the sample from which resampling is done be drawn from the full range of

population values, i.e. that it be on average, truly representative.

We can relate further asymptotic results on the validity of the bootstrap as a tool for the estimation of a R(X,F)'s distribution. We first report the theoretical results of Singh (1981).

The latter showed that for some basic statistics based on the sample mean, and without assuming anything about the structure of F (the underlying distribution), and finally if $Ex^2 < \infty$, then

$$P\ [n^{0.5}(\bar{X}_n - \mu) \leq \varepsilon\ ] \ - \ P^*[n^{0.5}(\bar{X}_n^* - \bar{x}_n) \leq \varepsilon\ ] \ \underset{\infty}{\xrightarrow{a,s}} \ 0$$

where P is the actual distribution of the sample mean $\bar{X}_n$ minus the population mean μ and where P^* is the bootstrap distribution of the bootstrap sample mean minus the realized sample mean, i.e.

$$\bar{X}^* - \bar{x}$$

He showed, in other words, that the bootstrap distribution of

$$R(X^*, \hat{F}) \ = \ \bar{X}^* - \bar{x}$$

converges with probability one towards the actual distribution, regardless of P, the underlying distribution. He also showed that the same strong convergence applies to

$$Pr\left[n^{0.5}(\bar{X}_n - \mu)/S_n \leq \varepsilon\right]$$

and to

$$Pr\left[n^{0.5}(\bar{X}_n^* - \bar{x})/s_n \leq \varepsilon\right]$$

where S_n and s_n are the population value and sample estimate respectively of the standard deviation of X .

Further, if the underlying distribution P is non-lattice, he showed

that, in the case of the standardized mean, the bootstrap is more accurate than the approximation by the limiting normal distribution given to us by the Central Limit Theorem.

We have so far presented the general concept of the bootstrap, and applied it to some theoretical examples: we have seen it to be fairly simple in principle and to be ideally applicable to circumstances where one is certain of having obtained a representative initial sample $X = x$. In addition, we have seen the bootstrap to be asymptotically valid, for certain basic statistics, in estimating their underlying sampling distributions.

The bootstrap has been applied with success to a wide variety of estimation problems. It will have been abundantly clear, by the wide generality of the random variable $R(X,F)$, that this method will apply potentially to many different circumstances, inasmuch as the X part of the argument of R can be permuted, without changing the value of R. It will also be clear that potentially very complex forms of R will lend themselves to this method, where often theoretical computations of the sampling distribution of R are currently inaccessible. A limit to the applicability of the bootstrap may appear to be the computational costs involved in computing R^*, B times, from each bootstrap sample realization $X^* = x^*$; this in practice does not appear, in at least some applications, to pose any real problem.

We will end this section with a series of examples of the application of the bootstrap. The applications cited in this section illustrate the general applicability of the method to complex situations. We will leave discussion of applications to regression problems for

section 4, where also comparisons will be made to the jackknife.

Our first example is drawn from Efron (1982a) and concerns the evaluation of the sampling distribution of the trimmed means. Let

$$x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$$

be the order statistics of a sample on the real line; if we remove a proportion a = k/n (k constant) of points from each end of the ordered sample, we may write the trimmed means as the average of the remaining n(1-2a) central order statistics. For example, if n = 21 and a = 0.1, then

$$\hat{\theta} = 17^{-1}\left[\sum_{i=4}^{18} x_{(i)} + (0.9)(x_{(3)} + x_{(19)}) \right]$$

note that n X a = 2.1; this entails the procedure of counting (.9) of the 3rd and 19th order statistics in the sum. The variance of $\hat{\theta}$ is generally computed as

$$\hat{VAR} = 1/(n)(n-1)(1-2a)^2 \left[\sum_{i=1}^{n} (x_{(wi)} - \bar{x}_{(w)}) \right]$$

where

$$wi = \begin{cases} g + 1, & \text{for all } i \leq g + 1 \\ i, & \text{if } g + 1 < i < n - g \\ n - g, & \text{if } i \geq n - g \end{cases}$$

where

$$g = (n - 1)\, a$$

Note that the end points $x_{(g)}$ and $x_{(n-g)}$ are counted g times in this computation.

Applied to our example where n = 21, we obtain g=2, and so

$$\hat{VAR} = 1/(21)(20)(.64)^2 \left[\sum_{i=1}^{21} (x_{(wi)} - \bar{x}_{(w)}) \right]$$

where

$$w_i = \begin{cases} 3, & \text{for all } i \leq 3 \\ i, & \text{if } 3 < i < 19 \\ 19, & \text{if } i \geq 19 \end{cases}$$

The procedure in (II.1a,b, and c) was applied to the trimmed means (a =.25) of an initial sample of n = 20 drawn from $F \sim N(0,1)$ and $G \sim e^{-x}$; 200 bootstrap samples were drawn using this initial sample, and $\hat{\theta}$ calculated; but, to be able to generalize these results, and assure that they were not due to a particularity of the specific initial sample drawn, 1,000 such initial samples were drawn from F and 3,000 from G, from each of which 200 bootstrap samples were drawn and the calculations described above performed. Each of these constituted a "trial". The results are given in Table 2.1, where the average value of the bootstrap standard deviation SD and the value of its standard deviation over the 200 trials are displayed. The results for the jackknife are also shown for later reference.

For purposes of comparison, the "true" Standard deviation, along with the minimum possible CV, were computed. The bootstrap is seen to perform moderately well. We note the greater variability of the jacknife in this case.

A further example, again drawn from Efron (1982a) is needed, to better illustrate the possible applications of the bootstrap. In a Monte Carlo study, a series of 200 samples (n = 14) were drawn from the bivariate normal distribution

$$Z_i \sim N_2(\mu, \Sigma), \quad i = 1,2,\ldots 14$$

$$Z_i = (X_i, Y_i)$$

where the correlation coefficient between X and Y was assigned as

Table 2.1

Jackknifing and bootstrapping the trimmed means

(from Efron (1982a))

	Average of $\hat{SD}$	Std. dev. of $\hat{SD}$	Coeff. of variation
Using $F \sim N(0,1)$			
1) Bootstrap (1000 trials; B=200/trial)	.236	.047	.200
2) Jackknife (delete-one; 1000 trials)	.236	.070	.300
3) true SD (minimum)	.224		(.160)
Using $F \sim e^{-x}$			
1) Bootstrap (3000 trials; B=200/trial)	.222	.072	.700
2) Jackknife (delete-one; 3000 trials)	.234	.143	.610
3) true SD (minimum)	.222		(.270)

$r = .5$ (note: the author did not specify the parameters μ and Σ); now, each of these 200 samples served as a basis for the generation of B = 128 and B = 572 bootstrap samples, from which values for r^* were calculated, following procedure (II.1a,b, and c). The results are in Table 2.2; here again, as in Table 2.1, the average of SD, its standard deviation, the coefficient of variation, and also the root MSE (of estimated minus true standard deviation) are presented.

Note that the results were computed also for

$$\tanh^{-1}(\hat{r}) = 0.5 \log \left[(1 + \hat{r})/(1 - \hat{r})\right]$$

Again for comparison purposes, the theoretical true values, given the parameters of the bivariate normal used, were calculated for the population value of the standard deviation. Also, estimates for the average values of the standard deviation, its standard deviation and CV, as well the root MSE are computed; theoretical values are given based on a normal approximation for the distribution of $\tanh^{-1}(\hat{r})$.

We note first the general similarity of results achieved with the bootstrap to those derived through the theoretical calculations made using the normal theory approximation. We also note that there seems to be little improvement, whether one resamples B = 128 times or B = 512 times. Note also that the resemblance to the normal theory and true results improves if one first transforms $\hat{r}$ to $\tanh^{-1}(\hat{r})$.

Many more examples can be given, showing the successful application of the bootstrap for a wide variety of R(X,F). For example, in the sampling distribution of the median (Efron, 1979a). But, as the work of this thesis focuses on applications in regression, we shall not give any

Table 2.2

Jackknifing and bootstrapping the correlation coefficient

(from Efron (1982a))

	Average of $\hat{SD}$	Std. dev. of $\hat{SD}$	Coeff. of variation	$(MSE)^{0.5}$
results for $\hat{r}$				
1) Bootstrap (200 trials; B=128/trial)	.206	.066	.320	.067
2) Bootstrap (200 trials; B=512/trial)	.206	.063	.310	.064
2) Jackknife (delete-one; 200 trials)	.223	.085	.380	.085
4) Normal Theory	.217	.056	.260	.056
5) True value	.218			
results for $\tanh^{-1}(\hat{r})$				
1) Bootstrap (200 trials; B=128/trial)	.301	.065	.220	.065
2) Bootstrap (200 trials; B=512/trial)	.301	.062	.210	.062
3) Jackknife (delete-one; 200 trials)	.314	.090	.290	.091
4) Normal Theory	.302	0	0	.003
5) True value	.299			

further such examples here. Rather, let us describe the general concept of the jackknife, comparing it and some general examples to the above given description and examples of the bootstrap.

B. THE JACKKNIFE

The jackknife was originally given to us by Quenouille (1949) and Tukey (1958). Like the bootstrap, it is used to estimate the sampling distribution of some statistic $\hat{\theta}$ which estimates θ, a parameter of some distribution F. We may describe it in general as follows.

Let a random sample of size n be drawn from a completely unspecified probability distribution F; F is a distribution over R^k (k=1,2,...m).

We observe X = x, and compute $\hat{\theta} = \hat{\theta}(x_1, x_2, \ldots, x_n)$. But whereas the bootstrap estimates the sampling distribution of a general random variable R(X,F), the jackknife focuses traditionally on the sampling distribution of

$$R(X,F) = E_F\theta(\hat{F}) - \theta(F) \equiv \text{Bias}$$

(or a standardized form of the above) to estimate the sampling distribution of $\hat{\theta}$. E_F is the expectation under

$$X_1, X_2, \ldots, X_n \sim_{iid} F$$

However, the resampling procedure under which Bias is estimated differs from that of the bootstrap. One does the following:

1) Removing, sequentially, points x_i from X = x, we assign

$$\hat{F}_{(i)} : \text{mass } 1/n-1 \qquad \text{(II.10a)}$$

to

$$x_1 , x_2 , \ldots , x_{i-1}, x_{i+1}, \ldots , x_n$$

2) Then, compute the estimate of Bias

$$\overline{BIAS} = (n-1)(\hat{\theta}_{(.)} - \hat{\theta}) \qquad \text{(II.10b)}$$

where

$$\hat{\theta}_{(.)} = n^{-1}\sum_{i=1}^{n} \hat{\theta}_{(i)} \qquad \text{(II.10c)}$$

and where

$$\hat{\theta}_{(i)} = \theta(\hat{F}_{(i)})$$

We may then compute the so-called bias-corrected jackknife estimate of $\hat{\theta}$

$$\tilde{\theta} = \hat{\theta} - \overline{BIAS} = n\hat{\theta} - (n-1)\hat{\theta}_{(.)} \qquad \text{(II.10.d)}$$

The theory also provides for an estimate of the standard deviation of $\hat{\theta}$. It is

$$(\hat{VAR})^{0.5} = \left[(n-1)/n \sum_{i=1}^{n} (\hat{\theta}_{(i)} - \hat{\theta}_{(.)})^2\right]^{0.5} \qquad \text{(II.10e)}$$

One can also provide a basic rationale, as we did for the bootstrap, to motivate our use of such a resampling method.

Let

$$E_n = E_F \hat{\theta}(X_1 , X_2 , \ldots , X_n)$$

denote the expectation under F for a sample size of n; then, for many common statistics, we may write

$$E_n = \theta + a_1(F)/n + a_2(F)/(n^2) + \ldots \qquad \text{(II.11)}$$

where the functions $a_i(F)$ do not depend on n.

Noting that

$$E_F(\hat{\theta}_{(.)}) = E_{n-1} = \theta + a_1(F)/(n-1) + a_2(F)/(n-1)^2 + \ldots \quad (II.12)$$

we may then write the expectation of the bias-corrected estimate $\tilde{\theta}$ of θ as

$$E_F\tilde{\theta} = nE_n - (n-1)E_{n-1} = \theta + a_2(F)/n(n-1) + a_3(F)\left[1/n^2 - 1/(n-1)^2\right] + \ldots$$

Hence $\tilde{\theta}$ is biased $O(n^{-2})$ compared to $O(n^{-1})$ for $\hat{\theta}$.

Let us now provide a simple example: the sample average. Having observed $X = x$, we may calculate

$$\bar{x} = n^{-1}\sum_{i=1}^{n} x_i \quad (II.13)$$

Following the procedure (II.10), we obtain

$$\hat{\theta}_{(i)} = \bar{x}_{(i)} = (n-1)\sum_{j\neq i} x_j \quad (i=1,2,\ldots,n)$$

also

$$\overline{BIAS} = (n-1)(\bar{x}_{(.)} - \bar{x})$$

note that $\bar{x}_{(.)} = \bar{x}$, and so $\overline{BIAS} = 0$.

This is not surprising, because of the unbiasedness of the sample mean. Note also that

$$(\hat{VAR}) = (n-1)/n\sum_{i=1}^{n}(\bar{x}_{(i)} - \bar{x}_{(.)})^2 \quad (II.14)$$

This is equivalent to the variance of $\bar{X}$

$$1/n(n-1)\sum_{i=1}^{n}(x_i - \bar{x})^2$$

One need only substitute

$$\bar{x}_{(i)} = (n\bar{x} - x_i)/(n-1)$$

and use (II.13), into (II.14) to see this.

Another simple example is in the estimation of the Bias, and the calculation of $\tilde{\theta}$ for the sample variance; here, $\hat{\theta}$ can be written as

$$\hat{\theta} = n^{-1}\sum_{i=1}^{n} (x_i - \bar{x})^2$$

Here,

$$\overline{BIAS} = -1/n(n-1)\sum_{i=1}^{n} (x_i - \bar{x})^2$$

and so

$$\tilde{\theta} = 1/n \sum_{i=1}^{n} (x_i - \bar{x})^2 + 1/n(n-1)\sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$= 1/(n-1)\sum_{i=1}^{n} (x_i - \bar{x})^2$$

which is the usual unbiased estimate of θ.

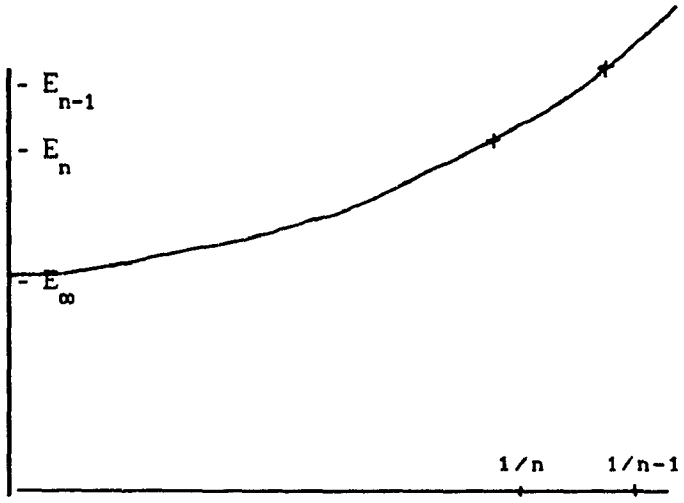
We next discuss some asymptotic properties of the jackknife, as supplied to us by Miller (1974a) and Efron (1982a). Here, we describe some theory on the asymptotics of $\overline{BIAS}$.

Assume the graph of (FIG 2.1) to be that of E_n versus $1/n$, from (II.11). One observes that it is nearly linear. The near linearity is due to the rapidly diminishing effects of the terms beyond the order n. Note that we denote θ, the true value, by E_∞ which again follows from (II.11), if the terms in $a_1(F)$ are finite.

If we now approximate the graph in Fig. 2.1 to be linear (equivalently, this means that all terms of order n^{-2} are dropped), we

Fig 2.1

Asymptotics of BIAS; graph of E_n vs 1/n



have

$$(E_n - E_\infty)/(E_{n-1} - E_n) = \left[\theta + (a_1(F)/n) - \theta\right]/\left[\theta + a_1(F)/(n-1) - \theta - a_1(F)/n\right]$$

$$= (1/n)\ /(1/n-1 - 1/n)$$

which will yield the population value Bias

$$\text{Bias} = E_n - E_\infty = (n-1)(E_{n-1} - E_n) \qquad (II.15)$$

and so

$$\theta = E_\infty = n\,E_n - (n-1)\,E_{n-1} \qquad (II.16)$$

The jackknife formulas (II.10b) and (II.10d) simply approximate this by replacing E_n and E_{n-1} in (II.15) and (II.16) by their unbiased estimates $\hat{\theta}$ and $\hat{\theta}_{(.)}$ respectively.

The extrapolation method discussed above has a solid foundation in numerical analysis. An example of this is Aitken acceleration: an interesting parallel can be made between the above approximation method and this numerical analysis method. Denoting the bias for sample size n as

$$\text{Bias}_n = E_n - E_\infty$$

we may then write E_∞ as

$$E_\infty = (E_n - r\,E_{n-1})/(1-r)$$

where

$$r = \text{Bias}_n\ /\ \text{Bias}_{n-1}$$

If again Bias_n is approximated as linear in 1/n, we will have

$$r = (n-1)/n$$

This in turn, used in (II.17), yields (II.16), which as we have said

can be approximated by (II.10d). Further, suppose we wish to approximate the infinite sum (see II.11)

$$S_\infty = \sum_{k=0}^{\infty} b_k$$

on the basis of finite sums S_n, where

$$S_\infty = \sum_{k=0}^{n} b_k + \sum_{k=n}^{\infty} b_k = S_n + B_n$$

(Note B_n would represent the "bias" terms in this non-random problem), we may then write, similarly to (II.17)

$$S_\infty = (S_n - r\, S_{n-1})/(1 - r)$$

where

$$r = B_n / B_{n-1}$$

Aitken acceleration approximates r using

$$\hat{r} = (S_n - S_{n-1})/(S_{n-1} - S_{n-2})$$

(If S_∞ is the geometric series, $\hat{r} = r$, since

$$\hat{r} = c\, r^n / c\, r^{n-1} = r \quad (c \text{ constant})$$

The asymptotic theory given so far for the Jackknife tends to model itself on exact methods in numerical analysis, and works on certain parameters of the distribution of R(X,F), such as Bias; we have not, at this writing, located general asymptotic theoretical material concerning the parameters and/or the form of the sampling distribution of Jackknife estimates of R(X,F), as we found for the Bootstrap method. In addition, we will now supply proof that the Jackknife variance estimate given earlier

$$\hat{VAR} = (n-1)/n \left[\sum_{i=1}^{n} (\hat{\theta}_{(i)} - \hat{\theta}_{(.)})^2 \right]$$

is biased in many applications.

$\hat{VAR}$ may be written as

$$\hat{VAR} = (n-1)/n \; \tilde{VAR} \qquad (II.18)$$

We will show in the following, summarising a proof by Efron and Stein (1981) that

$$E_F \; \tilde{VAR} \geq Var_{n-1} \qquad (II.19)$$

where Var_{n-1} is the true sampling variance of $\hat{\theta}$. The proof of (II.19) proceeds using an ANOVA n-way decomposition of $\hat{\theta}$. Note first that

$$\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$$

which makes explicit the point that $\hat{\theta}$ is a function of the i.i.d. X_i.

Now, define

$$\mu = E_F(\hat{\theta})$$

$$a_i = a(X_i) = n \left[E_F(\hat{\theta}/X_i) - \mu \right]$$

$$b_{ii'} = b(X_i, X_{i'}) = n^2 \left[E_F(\hat{\theta}/X_i X_{i'}) - E_F(\hat{\theta}/X_i) - E_F(\hat{\theta}/X_{i'}) + \mu \right]_{(i \neq i')}$$

etc.

the last definition being

$$m_{123\ldots n} = m(X_1, X_2, \ldots, X_n) = n^n \left[\hat{\theta} - E_F(\hat{\theta}/X_1, X_2, \ldots, X_{n-1}) - E_F(\hat{\theta}/X_1, X_2, \ldots, X_{n-2}) - \ldots + (-1)^n \mu \right] \qquad (II.20)$$

Here μ is the "grand mean", $(n\, a_i)$ is the "main effect", $(n^2\, b_{ii'})$ is the ii' th interaction, etc..

Let us explain for instance the term $E_F(\hat{\theta}/X_i)$ It is the expected value under F of $\hat{\theta}$ given X_i; under a specific realization $X = x$, this

expected value would be different from μ. Distinct values are expected, for any given realization $X = x$, for each of the"effects" noted above in (II.20); these in turn contribute to the sum

$$\hat{\theta}(X_1, X_2, \ldots, X_n) = \mu + n^{-1} \sum^{n} a_i + n^{-2} \sum^{n} b_{ii'} + n^{-n} m_{1,2,..,n} \tag{II.21}$$

where each of the little sums on the R.H.S. of (II.21) designates summation over all combinations for the designated effect. The relation (II.21) is proven formally in Efron and Stein (1981). For each givenrealization $X = x$, the sum will define a specific value of $\hat{\theta}$.

Three (3) properties of the terms defined in (II.20) must be mentioned before the main proof is given:

1) Each term is a function only of the X_i indicated. For example b_{21} is a function of X_2 and X_1.

2) Each term, when conditioned upon all but one of its X_i ,will have conditional expectation 0. An example of this is

$$E_F (b_{ii'}/X_i) = n^2 \left[E_F (\hat{\theta}/X_i) - E_F (\hat{\theta}/X_{i'}) - E_F (\hat{\theta}) + \mu \right] = 0$$

3) The terms are mutually uncorrelated. (To see this, one need only consider that the terms are themselves functions independent X_i). Hence, for example

$$E_F (a_1 a_2) = 0$$

Main Proof:

First, define

$$\text{Var}(a_1) = \sigma_a^2$$

$$\text{Var}(b_{11},) = \sigma_b^2$$

etc

Notice also that the main result (II.19) concerns a sample of size (n-1) (since $\tilde{\text{VAR}}$, written as

$$\tilde{\text{VAR}} = n/(n-1)\ \hat{\text{VAR}}$$

can be considered as an estimate of Var_{n-1} obtained through a sample adjustment of (n/n-1) on $\hat{\text{VAR}}$). Hence, (II.19) is really a statement about $\hat{\theta}(X_1, X_2, \ldots, X_{n-1})$ and there is no need to even define $\hat{\theta}(X_1, X_2, \ldots, X_n)$.

We therefore can directly apply the expansion (II.21) to calculate

$$\text{Var}\left[\hat{\theta}\ (X_1, X_2, \ldots, X_n)\right] = \text{Var}_{n-1}$$

This last equality comes from dropping all terms a_n , b_{in}, etc. in II.21 which are normally conditioned on X_n , which is not included here, and applying property (2).We may then simply derive

$$\begin{aligned}\text{Var}_{n-1} &= (n-1)/(n-1)^2\ \sigma_a^2 + \binom{(n-1)}{2} \sigma_b^2/(n-1)^4 \\ &\quad + \binom{(n-1)}{3} \sigma_c^2/(n-1)^6 + \ldots \\ &= \sigma_a^2/(n-1) + \binom{(n-2)}{1} \sigma_b^2/2(n-1)^3 \\ &\quad + \binom{(n-2)}{2} \sigma_c^2/3(n-1)^5 + \ldots \end{aligned} \qquad \text{(II.23)}$$

where, for example

$$\binom{(n-1)}{3} (n-1)^{-6} = (n-2)!/(n-4)!.3.2!\,(n-1)^5$$

$$= (n-2)!/(n-4)!.2!3(n-1)^5$$

$$= \binom{(n-2)}{2} \, 1/3(n-1)^5$$

Now, note that we can write

$$\hat{\theta}(X_1, X_2, \ldots, X_n) = \hat{\theta}_{(n)}$$

in the Jackknife notation.

Applying (II.21) to

$$\hat{\theta}(X_1, X_2, \ldots, X_{n-2}, X_n) = \hat{\theta}_{(n-1)}$$

we may then calculate

$$E_F\left[\hat{\theta}_{(n)} - \hat{\theta}_{(n-1)}\right]^2 = E_F\left[(n-1)^{-1}\left(\sum_{i \neq n} a_i - \sum_{j \neq n-1} a_j\right)\right.$$

$$\left. + (n-1)^{-2}\left(\sum_{i,i' \neq n} b_{ii'} - \sum_{j,j' \neq n} b_{jj'}\right) + \ldots\right]^2 \quad \text{(II.24)}$$

$$= E_F\left[(n-1)^{-1}\left[\sum_{i \neq n}(a_i - Ea_i) - \sum_{j \neq n-1}(a_j \quad Ea_j)\right]\right.$$

$$\left. + (n-1)^{-2}\left[\sum_{i,i' \neq n}(b_{ii'} - E\,b_{ii'}) - \sum_{j,j' \neq n-1}(b_{jj'} - E\,b_{jj'})\right] + \ldots\right]^2$$

where $Ea_i = E_F\left[E_F(\hat{\theta}/X_i) - \mu\right] = 0$; similarly $E\,b_{ii'} = 0$, etc..

The form (II.24) can be written finally as

$$E_F\left[\hat{\theta}_{(n)} - \hat{\theta}_{(n-1)}\right]^2 = 2/(n-1)^2\,\sigma_a^2 + 2\binom{(n-2)}{1}\sigma_b^2/(n-1)^4 \quad \text{(II.25)}$$

Now, using (II.18) and (II.10e), we may write $\tilde{VAR}$ as

$$\tilde{VAR} = \sum_{i=1}^{n}(\hat{\theta}_{(i)} - \hat{\theta}_{(.)})^2 \quad \text{(II.26)}$$

This may be rewritten as

$$n^{-1} \sum_{i=1}^{n} (\hat{\theta}_{(i)} - \hat{\theta}_{(i')})^2 \quad (i \neq i') \qquad (II.27)$$

The equivalence between (II.26) and (II.27) can be observed if one uses the relation

$$\hat{\theta}_{(.)} = n^{-1} \sum_{i=1}^{n} \hat{\theta}_{(i)}$$

then, multiplying out the crossproducts in (II.25), one may regroup the resulting terms in $\hat{\theta}_{(i)}\hat{\theta}_{(i')}$ and factor out to form the new crossproducts

$$n^{-1} (\hat{\theta}_{(i)} - \hat{\theta}_{(i')})(\hat{\theta}_{(i)} - \hat{\theta}_{(i')})$$

which, summed, yield (II.27).

Now all the terms in (II.27) have expectation (II.25) and so

$$E_F \tilde{VAR} = \sigma_a^2/(n-1) + \binom{(n-2)}{1} \sigma_b^2/(n-1)^3 + \binom{(n-2)}{2} \sigma_c^2/(n-1)^5 + \ldots \quad (II.28)$$

(the "2"s in the terms of (II.25) disappear by the symmetry of the terms in brackets).

Finally, we may substract (II.23) from (II.28) to obtain

$$E_F \tilde{VAR} - Var_{n-1} = \binom{(n-2)}{1} \sigma_b^2/2(n-1)^3 + 2\binom{(n-2)}{2} \sigma_c^2/3(n-1)^5 + \ldots$$

Note that there are (n-2) terms on the RHS. Since all these terms are positive, we have proven our result.

A comment: if we can write $\hat{\theta}$ as

$$\hat{\theta} = \mu + n^{-1} \sum_{i=1}^{n} a_i$$

(the sample mean can be written this way), then we will have

$$E_F \tilde{VAR} = Var_{n-1}$$

Otherwise, the inequality holds: this is the case for instance, for $\hat{b}$, the estimator of the classical least squares regression coefficient vector b , since $\hat{b}$ is a non-linear function of the sample X.

We end this section by referring back to the tables of results given for the two applications of the Bootstrap method, the trimmed means and the correlation coefficient, Tables 2.1 and 2.2 respectively. We note that in both examples, the Jackknife showed more variability than the Bootstrap. To be fair, however, one can also find counterexamples, where the Bootstrap performs more poorly, in terms of bias or standard deviation of the estimator of interest, than the Jackknife.

We will now discuss the specific application of both of these methods to regression problems.

4 THE BOOTSTRAP THE JACKKNIFE AND SOME APPLICATIONS IN REGRESSION

The bootstrap and the jackknife may be applied to a variety of regression models. This section will give some applications to single equation models, as well as to multiple and simultaneous equation systems.

a. Single equations:

In this subsection, both the bootstrap and the jackknife will be described in their handling of some general regression models. Also, some specific applications will be briefly described. An attempt will be made, throughout this subsection, to compare the two resampling methods in their success at depicting sampling distribution characteristics of the regression statistics of interest.

i) Bootstrap applications, single equations

Efron (1979a ,1982a) described an application to a general regression problem. His ideas are summarized here. Let

$$Y_i = f_i(b) + e_i \quad (i=1,2,\ldots,n)$$

be such a model, where the $f_i(.)$ are known functions of the as yet unknown (p x 1) vector b of parameters; the $f_i(.)$ depend usually on some vector of covariates X_i. The e_i are such that $e_i \sim_{iid} F$.

Also, F is centered at 0 in one way or another; examples of this are

$$E_F e = 0 \quad \text{or} \quad \text{Prob}_F\{e<0\} = 0.5$$

If we now put actual observations $Y_i = y_i$ and $X_i = x_i$ into the model,

we may now estimate b by minimizing a distance D(y,f) between y and the predictors of y, $f = [f(a), f(b),\ldots,f(b)]'$

$$\hat{b} = \min_{b} D\left[y, f(b)\right]$$

The above model may involve very complex forms of f which are beyond current analytical methods. We shall see some examples of this later in this subsection.

The bootstrap algorithm may be applied in two distinct manners;

Algorithm (a):

1) Construct the mass function $\hat{F}$ as follows

$$\hat{F} : 1/n \quad \text{over} \quad \hat{e}_i = y_i - f_i(\hat{b})$$

2) Draw $\hat{e}_i^*$ and build

$$\hat{e}^* = \left[\hat{e}_1^*, \hat{e}_2^*,\ldots,\hat{e}_n^* \right]$$

using $\hat{F}$.

3) Construct

$$y_i^* = f_i(\hat{b}) + \hat{e}_i^*$$
$$y^* = \left[y_1^*, y_2^*,\ldots,y_n^* \right]$$

((2) and (3) together form the bootstrap sample described in the basic algorithm (see II.1b))

4) Calculate,

$$\hat{b}^* = \min_{b} D\left[y^*, f(\hat{b})\right]$$

from results in (3);

5) Repeat steps (2) to (4) B times, obtaining

$$\hat{b}_1^*, \hat{b}_2^*, \hat{b}_3^*,\ldots,\hat{b}_B^*$$

We may estimate $\hat{b}$'s covariance matrix from

$$\hat{COV} = (B-1)^{-1} \sum_{b=1}^{B} (\hat{b}_b^* - \hat{b}_\cdot^*)(\hat{b}_b^* - \hat{b}_\cdot^*)'$$

where

$$\hat{b}^{*} = (B)^{-1} \sum_{b=1}^{B} \hat{b}_{b}^{*}$$

We may apply this algorithm to a linear situation, where

$$f_i(b) = \underset{(1Xp)}{X_i} \underset{(pX1)}{b}$$

and where

$$D\left[y, f(b) \right] = \sum_{i=1}^{n} (y_i - X_i b)^2$$

Also, let

$$X = \underset{(nXp)}{\left[\iota \vdots X_2 \vdots \ldots \vdots X_p \right]} \qquad \underset{(nX1)}{\iota} = (1, 1, \ldots, 1)' \qquad \text{(II.29a)}$$

In this context, if we assume

$$1) \quad E_F e = 0$$

$$2) \quad r(X'X) = p \quad (p \leq n) \qquad \text{(II.29b)}$$

we may then generate the CLS estimate of b

$$\hat{b} = (X'X)^{-1}X'y$$

Applying the general algorithm (1) - (5) above to this situation, and assuming $\underset{\hat{F}}{E}(e^{*}) = 0$, (e^{*} is the random variable whose realization is $\hat{e}^{*}$) and also assuming

$$\text{Var } e_i^{*} = \hat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} (\hat{e}_i)^2$$

we may then calculate

$$\hat{b}^{*} = (X'X)^{-1}X' \, y^{*}$$

which will have as true covariance

$$\begin{aligned} \text{Cov} &= E\,(\hat{b}^{*} - \hat{b})(\hat{b}^{*} - \hat{b})' \\ &= E\left[(X'X)^{-1}X' (e^{*} e^{*\prime}) X' (X'X)^{-1} \right] \\ &= \hat{\sigma}^2 (X'X)^{-1} \end{aligned} \qquad \text{(II.30)}$$

Note that the bootstrap gives the standard estimate of covariance in this case, except for the use of n instead of (n-p) in the denominator of $\hat{\sigma}^2$.

Algorithm b:

A second algorithm may be stated as follows

1) Let

$$Z_i = \underset{(1X1)}{Y_i} \; : \; \underset{(1Xp)}{X_i}$$

2) Use (II.1a,b,c) substituting Z_i for X_i in the model

$$Y_i^* = X_i^* b + e_i$$

estimating b with $\hat{b}^*$ as before

$$\hat{b}^* = (X^{*\prime}X^*)^{-1}X^{*\prime}y^*$$

The choice of either Algorithm (a) or (b) poses a dilemma. On the one hand (a) yields results quite parallel to classical results. On the other hand, (b) could be best suited to preserving the link betwee Y_i and X_i; but the rigid assignation of residuals to specific pairs of Y_i^* and X_i^* would bring us into an estimator of the bootstrap covariance Cov of the form

$$\hat{COV} = (X'X)^{-1}\left[\sum_{i=1}^{n} X_i'X_i\,\hat{e}_i^{\,2}\right](X'X)^{-1} \qquad (II.31)$$

(One need only consider (II.30), substituting $\hat{e}$ for e^* to see this). If the $\hat{e}_i$ are equal, we have the traditional formula. If not, the different weights could substantially change the estimate (see also derivations for the jackknife, section 3.a.ii to follow).

Freedman (1981) has shown, with algorithm (a) and for the linear model, that the bootstrap approximation to the distribution of $\hat{b}$ is quite valid asymptotically. The linear model assumed is as in (II.29a,b, and c). In addition, if we assume that

$$n^{-1}X'X \xrightarrow{p} W \qquad \text{(II.32a)}$$

W being positive definite and constant, and if the elements of X are uniformly small with respect to $n^{0.5}$, then

$$n^{0.5}(\hat{b} - b) \overset{n\to\infty}{\sim} N_p(0,\sigma^2W^{-1}) \qquad \text{(II.32b)}$$

Freedman cautions to center the errors, by computing

$$\tilde{e}_i = \hat{e}_i - n^{-1}\sum_{i=1}^{n}\hat{e}_i \qquad \text{(II.32c)}$$

before applying the bootstrap procedure before algorithm (a). Hence, he resamples from $\tilde{e}_i$, drawing

$$\tilde{e}_i^* \overset{iid}{\sim} \hat{F}$$

This yields the bootstrap vector of centered errors $\tilde{e}_i^*$ and so we may compute

$$y^* = X\hat{b} + \tilde{e}^*$$

and calculate

$$\hat{b}^* = (X'X)^{-1}X'y^*$$

The rationale for centering of the error terms is provided in the following. The basic assumption of the bootstrap is that

$$n^{0.5}(\hat{b}^* - \hat{b})$$

(which can be approximated by $n^{0.5}(\hat{b}_b^* - \hat{b}^*)$) approximates the distribution of (II.32b). But in order for this to occur, the errors to be resampled must first be centered. Consider

$$n^{0.5}(\hat{b}^* - \hat{b}) = (X'X)^{-1}X'e^* \qquad \text{(II.32d)}$$

In this form, we see that the distribution of $n^{0.5}(\hat{b}^* - \hat{b})$ will include a bias term which is random; hence it will not degenerate asymptotically, which means that it will not converge to an asymptotic constant. This, despite the fact that the distribution of the uncentered error terms converges to F, the reason for this is that we depend wholly on a specific unbalanced sample vector $\hat{e}$ from which the resampling occurs.

Freedman proves that, along almost all sample sequences e (centered), given $Y = [Y_1, Y_2, \ldots, Y_n]$, the distribution of (II.32d) converges weakly to

$$N_p(0, \sigma^2 W^{-1}) \qquad \text{(II.33)}$$

To prove this, we need some definitions:

1) Let d_1^p be the Mallows metric for probabilities in R^p relative to the Euclidean norm $||.||$. Thus, if μ and υ are probabilities in R^p,

$$d_1^p(\mu, \upsilon) = \inf E(||U - V||^1)^{1/1}$$

Hence the Mallows metric establishes a relation between a measure of the distance between two vectors U and V in p-space and their respective probability laws. One may consult Bickel and Freedman (1981, Section 8) for the proofs of some basic relations concerning Mallows metrics, including the establishment of this measure as a metric on G_p, the probability space over vectors in R^p.

2) Assign the distribution P(F) such that

$$n^{0.5}(\hat{b} - b) \sim P(F)$$

This uses the fact that

$$e_i \sim \text{iid } F$$

We need only consult (II.32d) to understand the notation P(F).

Main proof:

1. To prove this theorem, we need the following result, which is an application of lemma 8.9 in Bickel and Freedman (1981). In effect, if H is an alternative law for e_i (other than F); that is, if $e_i \sim_{iid} H$ and if

$$E_H(e_i) = 0, \text{ and}$$
$$Var_H(e_i) < \infty$$

then it can be shown that

$$d_2^P \left[P(F), P(H) \right]^2 \leq n \text{ trace } \left[(X'X)^{-1} \right] d_2^1(F,H)^2$$

This relation will allow us to simplify relations in P(F) and P(H) to known relations in F and H.

2. Accepting this result, and substituting $\hat{F}$ for H (note that $\hat{F}$ is the empirical distribution from which we collect the resampled $\hat{e}^*$), we have that

$$d_2^P \left[P(F), P(\hat{F}) \right]^2 \leq n \text{ trace } \left[(X'X)^{-1} \right] d_2^1(F,\hat{F})^2 \qquad (II.34)$$

3. Freedman also shows that the Mallows metric for F and $\hat{F}$ converges almost everywhere (that is, strongly, in probability theory language) to 0. That is,

$$d_2^P(F,\hat{F})^2 = \inf E(||e - \hat{e}||^2)^{0.5} \xrightarrow{a.e} 0 \quad (II.35)$$

4. Since also, we have, by assumption in the regression model

$$n^{-1}X'X \xrightarrow{P} W \quad (W \text{ a constant}) \qquad (\text{see II.32a})$$

we have finally, bringing (II.34) and (II.35) together

$$d_2^p \left[P(F),\ P(\hat{F}) \right]^2 \xrightarrow{a.e} 0$$

Hence, the strong convergence to 0 of the distance between the vectors e and $\hat{e}$ brings about the convergence of the distribution of

$$n^{0.5}(\hat{b}^* - \hat{b})$$

and

$$n^{0.5}(\hat{b} - b)$$

Since the latter converges weakly to $N_p(0, \sigma^2 W^{-1})$ then so does the former. QED

Note that Freedman goes on to prove also that the distribution of the pivot

$$(X'X)^{0.5}(\hat{b}^* - \hat{b})\ (\hat{\sigma}^*)^{-1} / X = x \quad (II.36)$$

where

$$\hat{\sigma}^* = \left[n^{-1} \left[\sum_{i=1}^{n} (\hat{e}_i^*)^2 - \left(n^{-1} \sum_{i=1}^{n} \hat{e}_i^* \right)^2 \right] \right]^{0.5}$$

converges weakly to $N_p(0, I)$. The results in (II.33) and (II.36) apply to X being non-random. If X is a simple random sample from R^p, Freedman has shown that results are parallel, but convergence occurs to a centered multivariate normal with a different variance-covariance matrix.

Hence, at least for some modes of bootstrapping (algorithm a), the asymptotics prove the procedure to be valid. We now turn to some actual applications in single equation regression problems.

Bootstrap methods have been applied with varying degrees of success

to a wide variety of regression problems, many of which used estimators having a complex form. In particular, it has had increased use in contexts where analytical techniques are not currently available for the derivation of useful results. To briefly mention a few such applications, Wahrendorf, Becker and Brown (1987) have applied the bootstrap to enable the comparison of the degree of fit of two non-nested generalized linear models to sets of data. To cite another example, Gu (1987), applying the smoothing spline (a complex method which compensates for otherwise poor inference achieved in small sample situations) to a regression problem, has modified the standard smoothing spline procedure by including a bootstrapping step in it, rendering it a more useful inferential tool. Quenneville (1986) has used the bootstrap procedure, determining conditions under which it can better be used for testing linear hypotheses without the normality assumption.

We cite in more detail here another application, to give a clear example of the use of the bootstrap in a regression situation whose analysis is very difficult if not impossible if one were to try theoretical means. Droge (1987) has considered the problem of estimating the MSEP (Mean Squared Error of Production) in the non-linear regression case, using the bootstrap.

The model considered was

$$Y_i = f(X_i) + e_i \quad (i=1,2,\ldots,n)$$

$$E(e_i) = 0$$

$$Var(e_i) = \sigma^2$$

where f(.) is unknown and is to be approximated by a real-valued function

g_b (eg: $g_b = x^b$) where b is a (px1) vector of coefficients.

Letting

$$Y = \left[Y_1, Y_2, \ldots, Y_n \right] \quad (n \times 1)$$

$$X = \left[X_1, X_2, \ldots, X_n \right] \quad (n \times p)$$

$$f_X = \left[f(X_1), f(X_2), \ldots, f(X_n) \right] \quad (n \times 1)$$

Droge estimated b by a function of the weighted least-squares regression method, where the quadratic to be minimized appeared as an argument to the function g. The MSEP was then evaluated as

$$\text{MSEP of predictor } \hat{y} = \sum_{i=1}^{n} v_i \, E \, (z_i - \hat{y}_i)^2$$

where the z_i were themselves predictors of $f(X_i)$

Evaluating

$$\hat{e}_i = s^{-1}\left[y_i - \hat{f}(X_i) \right]$$

where

$$s^2 = n^{-1}\left[\sum_{i=1}^{n} (y_i - \hat{f}(X_i))^2 \right]$$

he bootstrapped the errors $\hat{e}_i$ using the procedure already described in (II.1), generating the pseudo-values

$$\hat{y}^* = \hat{f}(X=x) + \hat{\sigma}\, \hat{e}^*$$

recalculating $\hat{b}^*$ and $\hat{y}^*$ which enabled the calculation of $MSEP^*$, going through the quadratic argument procedure with the pseudo-data, and using some further adjustments which will not be mentioned here

This approach yielded some success, yielding lower variability than with other non-bootstrap measurements.

This subsection has explained the general algorithms available for the application of the bootstrap in single equations, showing its flexible and straightforward nature. It has also shown the asymptotic validity of at least one algorithm. Finally, we have shown it to be adaptable and useful in some complex regression problems, where pure analysis would be at the very least difficult, if not sometimes probably impossible. The next subsection will briefly develop applications of the jackknife to single equation regression problems; it will also be seen as an adaptable tool in studying the distribution of complex regression statistics; finally, some comparisons will be made with the bootstrap.

(ii) The jackknife and regression models, applications

We may apply the jackknife strictly as defined in (II. 10a, b, and c) to the linear regression model used in (II. 29 a and b) for the bootstrap. That is,

$$Y_i = X_i b + e_i \quad (i=1,2,\ldots,n)$$

$$e_i \sim \text{iid} F$$

$$E(e_i) = 0 \qquad Var(e_i) = \sigma^2$$

$$r(X'X) = p$$

and b estimated by

$$\hat{b} = (X'X)^{-1}X'y$$

To apply (II.10a, b, and c), we have but one recourse: using the realizations

$$z_i = \underset{(1\times(p+1))}{\left[\, y_i \,:\, x_i \,\right]}$$

of the random vectors z_i (i = 1, 2,..., n) we remove one point at a time from $\left[\, y \,:\, x \,\right]$ (n X (p+1)), recomputing

$$\hat{b}_{(i)} = (X_{(i)}'X_{(i)})^{-1}X_{(i)}'y_{(i)}$$

where the subscript (i) designates the removal of the ith row from X or y; and so the notation $\hat{b}_{(i)}$ designates the estimate of b made with point i, $\left[\, y_i \,:\, x_i \,\right]$ removed.

It can be shown that

$$\hat{b}_{(i)} = \hat{b} - (1 - x_i(X'X)^{-1}x_i')^{-1}(X'X)^{-1}x_i'\hat{e}_i \qquad (II.37)$$

where x_i and $\hat{e}_i$ are the ith rows of X (n X p) and the residual vector $\hat{e}$ (n X 1) respectively; we may also show that

$$1 - x_i(X'X)^{-1}x_i' = 1 - O(1/n)$$

that is, the expression in the denominator of (II.37) is equal to 1, but for a term of at least order n^{-1}. As the sample size increases, we may therefore neglect the denominator, approximating it to 1. We may then calculate, using the thus simplified version of (II.37)

$$\hat{b}_{(i)} - \hat{b}_{(.)} = -(X'X)^{-1}x_i'\hat{e}_i + (X'X)^{-1}\, n^{-1} \sum_{i=1}^{n} x_i'\hat{e}_i$$

yielding Tukey's estimate

$$\hat{COV} = (n-1)/n \sum_{i=1}^{n} \left[(X'X)^{-1} \left[n^{-1} \sum_{i=1}^{n} x_i' \hat{e}_i - x_i' \hat{e}_i \right] \right.$$

$$\left. X \left[n^{-1} \sum_{i=1}^{n} x_i \hat{e}_i - x_i \hat{e}_i \right] (X'X)^{-1} \right] \qquad (II.38)$$

But since, term by term

$$n^{-1} \sum_{i=1}^{n} x_i \hat{e}_i \; <<< \; x_i \hat{e}_i ,$$

we may approximate (II. 38) by

$$\hat{COV} = n^{-1} (n-1) \left[\sum_{i=1}^{n} (X'X)^{-1} x_i' x_i \hat{e}_i^2 (X'X)^{-1} \right]$$

$$= n^{-1} (n-1) (X'X)^{-1} \left[\sum_{i=1}^{n} x_i' x_i \hat{e}_i^2 \right] (X'X)^{-1}$$

If the errors $\hat{e}_i$ are all equal, then the above reduces to

$$\hat{COV} = n^{-1} (n-1) \left[\sum_{i=1}^{n} \hat{e}_i^2 (X'X)^{-1} \right] \qquad (II.39).$$

This is so, since

$$\sum_{i=1}^{n} x_i' x_i = X'X$$

In this case, the CLS estimate can be obtained through an adjustment $n(n-1)^{-1}(n-p)^{-1}$ made on (II.39)

If the residuals are not identical, the same comments' apply as for (II.31): the different squared residual weights $\hat{e}_i$ could make this estimate substantially different from the CLS estimate.

The jackknife has been used, as has the bootstrap, in numerous applications where analysis of the sampling destributions of statistics of interest is very difficult if not impossible. We will not give any examples of applications, here.

We will however, end this subsection with an expose of some pertinent results from a study by Wu (1986); the study attempted some comparisons between the two resampling methods developed here, the jackknife and the bootstrap, with some proposals for modifications.

Wu begins by pointing out some limitations of both the bootstrap, and the jackknife, such as we have described them so far. For the bootstrap, algorithm a, where the residuals are resampled, we will obtain the estimate in (II.30)

$$\hat{COV} = \hat{\sigma}^2 (X'X)^{-1}$$

inasmuch as we are reasonably sure that the errors are homoscedastic in the model; if sufficient testing reveals, however, that heteroskedasticity prevails, then we must express $\hat{COV}$ as

$$\hat{COV} = (X'X)^{-1}\left[\sum_{i=1}^{n} x_i' x_i \hat{e}_i^2\right](X'X)^{-1}$$

Application of bootstrap algorithm "a" would lead to inconsistent results, for reasons analogous to those given in our expose of some of the asymptotic theory on the bootstrap: the occurrence of a random bias throughout resampling renders the convergence to a constant impossible. Again, see (II.31).

And, for algorithm "b", the occurrence of very unbalanced data in the realizations of $Z_i=z_i$, $z_i = \left[\, y_i \,:\, x_i \,\right]$ $(1 \times (p+1))$, $(i=1,2,\ldots,n)$ that is, the occurrence of large contrasts in the values of the components of the vectors z_i, makes the resampling with or without replacement of these vectors yield inconsistent results, for a reason analogous to that given previously: the inclusion of randomly occurring, sharply contrasting z_i will not allow the convergence to constants of the estimators used. We

note the fact that this commentary, which applies equally well to sampling done with and without replacement is relevant to both bootstrap and jackknife methodologies.

We note in passing that Wu reported a finding by Miller (1974b) which corroborates, for the single equation linear model, the general finding we reported in detail earlier: (Chapter II, section 3) that is, that the jackknife estimate of the variance of $\hat{b}$ is generally biased upward. We add here, without further detail, that Wu's own results seem to confirm this.

As a remedy to some of the above shortcomings, especially the problem of unbalanced data, he proposes specific weighting schemes for the bootstrap, as well as for the jackknife, also proposing a more flexible choice of subset size for the latter. As the theoretical basis for these weighted estimators, he develops the following representation of the CLS estimator (given in II.29c)

$$\hat{b} = (X'X)^{-1}X'y$$

In the simplest regression model

$$\underset{(1X1)}{Y_i} = a + b\underset{1X1}{X_i} + e_i \qquad (i=1,2,\ldots,n)$$

where b and the X_i are scalars. The CLS estimator here can be expressed as

$$\hat{b} = \sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x}) / \sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$= \sum_{i<j}^{n}(y_i - y_j)(x_i - x_j) / \sum_{i<j}^{n}(x_i - x_j)^2 \qquad \text{(II.40)}$$

To see this, one uses

$$\bar{y} = (y_1 + y_2 + \ldots + y_n)\, n^{-1}$$

and

$$\bar{x} = (x_1 + x_2 + \ldots + x_n)\, n^{-1}$$

then, multiplying out the contents of the crossproducts

$$\left[y_i - n^{-1} \sum_{i=1}^{n} y_i \right] \left[x_i - n^{-1} \sum_{i=1}^{n} x_i \right]$$

we regroup the resulting factors in $x_i y_i$ and factor out to form the new crossproducts

$$(y_i - y_j)(x_i - x_j), \quad (i < j)$$

(II.40) can also be written as

$$\hat{b} = \sum_{i<j}^{n} (y_i - y_j)(x_i - x_j)^2 / \left[\sum_{i<j}^{n} (x_i - x_j)^2 \right] (x_i - x_j)$$

$$= \sum_{i<j}^{n} \mu_{ij}\, \hat{b}_{ij} \qquad \text{(II.41)}$$

where the $\hat{b}_{ij}$ are to be interpreted as being pairwise slopes for pairs of points (x_i, y_i) and (x_j, y_j).

The estimate $\hat{b}$ can be viewed, hence, as a weighted sum of pairwise slopes, with the weights proportionnal to

$$\mu_{ij} = (x_i - x_j)^2$$

Note also that we can write the latter as the following determinant

$$\left| (x_{(k)} - \bar{x}_{(k)})' (x_{(k)} - \bar{x}_{(k)}) \right| =$$

$$\left| \begin{bmatrix} (x_i - \bar{x}_{(k)}) & (x_j - \bar{x}_{(k)}) \end{bmatrix} \begin{bmatrix} (x_i - \bar{x}_{(k)}) \\ (x_j - \bar{x}_{(k)}) \end{bmatrix} \right| \qquad \text{(II.42)}$$

where $x_{(k)}$ is the kth subset of the (n X 2) matrix X which contains the values x_i and x_j and where $\bar{x}_{(k)}$ is the average

$$\bar{x}_{(k)} = (x_i + x_j)/2$$

(II.42) is also expressible as

$$2\left[(x_i - \bar{x}_{(k)})^2 + (x_j - \bar{x}_{(k)})^2\right]$$

Hence the weights μ_{ij} in (II.41) are proportionnal to the determinant of the product in (II.42). Note that if $x_i = x_j$, Wu defines $\mu_{ij}\hat{b}_{ij} = 0$; and so, for the subsets where $x_i = x_j$ (that is, all entries of the first column of X, which are all "1"), no contribution was made to the sum in (II.41).

This representation can be generalized. Let

$$s = \left[i_1, i_2, \ldots, i_r \right] \quad (r \leq n)$$

be a subset of $(1,2,\ldots,n)$. Also, we can express the regression coefficient for the (x_i, y_i) pairs in the subset "s" of the model in (II.29) as

$$\hat{b}_s = (X_s'X_s)^{-1}X_s'y_s$$

where

$$y_s = \underset{(r \times 1)}{\left[y_{i_1}, y_{i_2}, \ldots, y_{i_r} \right]'}$$

and

$$X_s = \underset{(r \times p)}{\left[X_{i_1} ; X_{i_2} ; \ldots ; X_{i_r} \right]'}$$

It can be shown generally that for any $k \leq r$, the full-data CLS estimator $\hat{b}$ can be written as

$$\hat{b} = \Sigma_r |X_s'X_s| \hat{b}_s / (\Sigma_r |X_s'X_s|) \quad \text{(II.43)}$$

where the summation is made over all the subsets "s" of size r, and where k denotes the number of components in $\hat{b}$. In the simple example given

earlier, r = 2, k = 2; for a sample of size n, there are $\binom{n}{2}$ subsets of size r = 2.

The result can be further generalized to cover the case of the weighted LSE. Let W = diag $(u_1, u_2, \ldots, u_n)$, $(u_i > 0)$; also, let W_s be the r-square submatrix corresponding to set "s". The full data weighted LSE is

$$\tilde{b} = (X'W\,X)^{-1}X'W\,y$$

For the subset "s", it is

$$\tilde{b}_s = (X_s'W_s\,X_s)^{-1}X_s'W_s\,y_s$$

Now, rewriting X_s as $(W_s^{0.5}X_s)$, and y_s as $(W_s^{0.5}y_s)$, we may now write (II.43)as

$$\tilde{b} = \Sigma_r\,|X_s'W_s\,X_s|\,\tilde{b}_s \,/\, (\Sigma_r|X_s'W_s\,X_s|) \qquad (II.44)$$

Finally, Wu transposes this result to resampling schemes. In effect, letting

$$1)\quad P^* = \left[P_1^*, P_2^*, \ldots, P_n^*\right]$$

$$(P_i^* > 0)\,,\ \sum_{i=1}^{n} P_i^* = 1 \qquad (II.45a)$$

be a resampling vector from X = x. For each $\hat{P}^*$, we may show that

$$\hat{b}^* = (X'D^*X)^{-1}X'D^*y \qquad (II.45b)$$

where

$$\hat{D}^* = \text{diag}\left[\hat{P}_1^*, \hat{P}_2^*, \ldots, \hat{P}_n^*\right]$$

The estimate $\hat{b}^*$ is hence a weighted LSE with weights proportionnal to the components $\hat{P}_i^*$. The full-data LSE $\hat{b}$ corresponds to

$$\hat{P}_i^* = 1/n \quad (i=1,2,\ldots,n)$$

2) Also, let us assume that certain conditions hold for the components of P^* (for example, that they are interchangeable).

Under conditions (1) and (2), then, he shows that for any resampling method (*)

$$b = E_*|X'D^*X|\, b^* / (E_*|X'D^*X|) \qquad (II.46)$$

Hence, that the LSE estimator is estimated in an unbiased way by a weighted function of the random variable $\hat{b}^*$, that is,

$$|X'D^*X|\, b^* / (E_*|X'D^*X|)$$

This result, according to Wu, explains the inadequacy of the bootstrap and the jackknife, such as we have described them so far, as satisfactory methods to describe the sampling distribution of the LSE estimator. This inadequacy is especially visible when $X = x$ is unbalanced.

He proposes a series of alternate estimators, which incorporate the above results. He proposes an estimate of the variance he calls the General Weighted jackknife. Rewriting (II.43) as

$$\Sigma_r |X_s'X_s|\, (\hat{b}_s - \hat{b}) = 0$$

he proposes a second order estimator

$$v_{j,r}(\hat{b}) = (n-r)^{-1}(r-k+1)\, \Sigma_r w_s (\hat{b}_s - \hat{b})(\hat{b}_s - \hat{b})' \qquad (II.47)$$

where the weights w_s are proportionnal to

$$|X_s'X_s|$$

and where the summation, again, is over all subsets of size "r". More intuitively, he proposes a weighted bootstrap estimator

$$v_w^* = E_*|X'D^*X|\,(b^* - \hat{b})(b^* - \hat{b})' \,/\, (E_*|X'D^*X|) \qquad (II.48)$$

(We note in passing that the calculation of (II.47) and (II.48) would involve significant computationnal difficulties; notably, Wu does not suggest any method for the selection of the subset size r in (II.47). Depending on the sample size, the selection of the optimal r could involve sizable computationnal problems or cost; one need only consider for example, the value of $\binom{n}{k}$ for even moderate values of n and k to see this.

We may take the results from Wu described above as a warning not to apply either the bootstrap or the jackknife without first considering the specifics of the data, whether it is homoscedastic or not, and whether it is balanced or not; and, we might generalize this to whether the data is such that the other assumptions of the CLS model are respected, whether it contains outliers, etc.. A cautious approach to the application of resampling plans seems advisable.

The bootstrap. (as well as the jackknife) is a very general, very adaptable method. In bootstrapping, we must, in generating the pseudo-data from which bootstrap estimates of the statistic of interest are calculated, make sure that we do not violate the assumptions of the

model used. We have already seen that, for a regression model as simple as the single equation linear one in II.29 a) and b), there are two distinct algorithms possible. In the following section, which also serves as introduction to applications of the bootstrap to multiple and simultaneous equation systems, we shall see examples of how basic model assumptions may be maintained in applying t he bootstrap.

b. Multiple and simultaneous equation models

By way of transition to applications of the bootstrap to simultaneous and dynamic linear equation applications, we offer this example given to us by Freedman and Peters (1984).

The above authors applied the bootstrap to a subset of the system of econometric equations called the Regional Demand Forecasting Model (RDFOR) designed to forecast demand for energy in the U.S through 1995. RDFOR forecasts what the demand for various fuel types will be in a future year, by consumption sector and geographical region; this, as a function of price and other exogenous variables. The specific subset of RDFOR which was the focus of study was that which concerned itself with the industrial sector's demand for fuel. The model may be represented as

follows:

$$y_{st} = a_s + b\,x_{st} + c\,z_{st} + d\,u_{st} + e\,y_{s,t-1} + f\,w_{st} + e_{st}$$

$$(s=1,2,\ldots,10)$$

$$(t= 1961, \ldots ,1978) \qquad \text{(II.49a)}$$

Where, in region "s" and year "t",

- y_{st} is the log of an index of overall fuel consumption
- x_{st} is the log of cooling degree-days;
- z_{st} the log of heating degree-days;
- u_{st} the log of an overall fuel price index;
- z_{st} the log of value added in manufacturing;
- and e_{st} a stochastic disturbance term.

Notice that we are in the presence, for any given t, of a set of ten (10) equations possessing a dynamic component, in that the lagged value of y is on the RHS. Also, whereas the a_s are left to vary by region, the other linear coefficients are constrained to be identical for all regions.

In addition we assume

1) $E(e_{st}) = 0$ for all (s, t) (II.49b)

2) the disturbances are stochastically independent of the exogenous variables, which are $(x_{st}, z_{st}, u_{st}, w_{st})$

3) the vectors $e_t = \left[e_{1t}, e_{2t}, \ldots, e_{10,t} \right]'$ are independent and identically distributed in time. That implies

$$E(e_t e_r') = \begin{cases} W & (r = t) \\ 0_{(10X10)} & (r \neq t) \end{cases} \qquad \text{(II.49c)}$$

W being constant and positive definite.

Note that

$$E\,(y_{st}e_{s,t-1}) \neq 0 \qquad \text{(II.49d)}$$

through the lagged y.

One may summarise the above information in the following equation set

$$\begin{bmatrix} y_{1961} \\ y_{1962} \\ \vdots \\ y_{1978} \end{bmatrix} = \left[\begin{array}{c:c} I_{10} & V_{1961} \\ I_{10} & V_{1962} \\ \vdots & \vdots \\ I_{10} & V_{1978} \end{array}\right] \begin{bmatrix} A \\ B \end{bmatrix} + \begin{bmatrix} e_{1961} \\ e_{1962} \\ \vdots \\ e_{1978} \end{bmatrix} \qquad \text{(II.49e)}$$

(180X1) (180X10) (180X5) (15X1) (180X1)

$$y_t = \begin{bmatrix} y_{1,t} \\ y_{2,t} \\ \\ y_{10,t} \end{bmatrix} \text{(10X1)} \qquad V_t = \begin{bmatrix} x_{1,t} & z_{1,t} & u_{1,t} & y_{1,t-1} & w_{1,t} \\ x_{2,t} & z_{2,t} & u_{2,t} & y_{2,t-1} & w_{2,t} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{10,t} & z_{10,t} & u_{10,t} & y_{10,t-1} & w_{10,t} \end{bmatrix}$$

(10 X 5)

and

$$A = \begin{bmatrix} a_1 \\ a_2 \\ \\ a_{10} \end{bmatrix} \qquad B = \begin{bmatrix} b \\ c \\ d \\ e \\ f \end{bmatrix} \qquad e_t = \begin{bmatrix} e_{1,t} \\ e_{2,t} \\ \vdots \\ e_{10,t} \end{bmatrix}$$

We may conveniently condense the above by

$$Y = V\,C + E$$

where

$$C = \begin{bmatrix} A \\ B \end{bmatrix} \qquad V = \left[\begin{array}{c:c} I_{10} & V_{1961} \\ I_{10} & V_{1962} \\ \vdots & \vdots \\ I_{10} & V_{1978} \end{array}\right]$$

Applying generalized least squares to this form (with $COV(e_t) = W$ positive definite) we may construct the matrix of observed errors

$$\hat{e} = \underset{(10\ X\ 18)}{\left[\hat{e}_{1961}\ \hat{e}_{1962} \text{----} \hat{e}_{1962}\right]}$$

and then compute

$$\hat{W} = \underset{(10X10)}{\hat{e}\,\hat{e}'} / 18$$

which is the sample interregional covariance of the errors whose (u,v) entry is

$$\hat{W}_{u,v} = \sum_{t=1961}^{1978} \hat{e}_{ut}\,\hat{e}_{vt}\,' / 18$$

Finally we constitute the full matrix of variance-covariances for the errors

$$\underset{(180X180)}{\hat{\Sigma}} = \begin{bmatrix} \hat{W} & 0 \\ 0\ \hat{W} & \vdots \\ 0\ \text{---} & \hat{W} \end{bmatrix} \qquad \text{(II.49f)}$$

Since the above matrix is invertible, we may calculate

$$\hat{\Sigma}^{-1} = \begin{bmatrix} \hat{W}^{-1} & & 0 \\ 0 & \hat{W}^{-1} & \vdots \\ \vdots & & \\ 0 & \text{----} & \hat{W}^{-1} \end{bmatrix} \qquad \text{(II.49g)}$$

Having thus summarised all the known and assumed information on the covariance of errors by (II.49 f and g), Freedman and Peters estimated the variance-covariance matrix of $\hat{b}$ by the generalized least squares formula

$$\hat{cov}(\hat{b}) = \underset{(15X15)}{\left[V'\ \hat{\Sigma}^{-1} V \right]^{-1}}$$

the results of these computations were held by Freedman and Peters to be unreliable. We summarise their argument here. Statisticians often use an iterative procedure to estimate the variance matrix Σ . First, they calculate $\hat{\Sigma}$ as above; and calling it $\hat{\Sigma}_{(0)}$ they can obtain an initial estimate of b. by calculating

$$\underset{(pX1)}{\hat{b}_{(1)}} = \underset{(pXp)}{\left[V'\ \hat{\Sigma}_{(0)}^{-1} V \right]^{-1}} V'\ \hat{\Sigma}_{(0)}^{-1} Y$$

and using this, compute

$$\hat{e}_{(1)} = Y - V \hat{b}_{(1)}$$

which is used to compute

$$\hat{W}_{(1)} = \hat{e}_{(1)} \hat{e}'_{(1)} / n$$

Incorporating this in a new estimate $\hat{b}_{(2)}$ of b, they continue with this process for a fixed number k of steps or until the value of

$$\hat{b}_{(k)} = \left[V' \hat{\Sigma}^{-1}_{(k-1)} V \right]^{-1} V' \hat{\Sigma}^{-1}_{(k-1)} Y$$

appears to stabilize, allowing a relatively constant value of

$$\hat{cov}(\hat{b}) = \left[V' \hat{\Sigma}^{-1}_{(k)} V \right]^{-1}$$

But this procedure depends of course on $\hat{\Sigma}^{-1}_{(0)}$, the original estimate, being a "good" estimate of Σ. In the case under study, Freedman and Peters seriously questioned its validity, because of the relatively small sample size (n = 18), compared to the large number of parameters (15) to be estimated. In regression problems, if one has a sample size near to or equal the number of parameters m to estimate, results are deemed to be unreliable. To see this, one may consider a geometrical explanation. In the case of m = n, one may draw only one plane in R^m which exactly fits the data; for a different data set, a wholly different plane may result. (this is also approximately true for m near n.)

Hence a different procedure had to be found. To assess the true variability of the coefficients, Freedman and Peters applied the bootstrap in the following way. Using the errors

$$\hat{e}_{st} = y_{st} - (\hat{a}_s + \hat{b} x_{st} + \hat{c} z_{st} + \hat{d} u_{st} + \hat{e} y_{s,t-1} + \hat{f} w_{st})$$

they constructed

$$\hat{e}'_t = \left[\hat{e}_{1,t}, \hat{e}_{2,t}, \cdots, \hat{e}_{10,t} \right]$$

To bootstrap, they used a version of Algorithm "a", in the three (3) steps which follow:

1) they assigned $\hat{F} = 1/18$ as empirical distribution for the 18 10-vector set of errors $\hat{e}_t$

2) Generation of a set of 18 (10X1) vectors $\hat{e}_t^*$ by a process of simple random sampling with replacement from the distribution $\hat{F}$; then, computation of the pseudo-data $\hat{y}^*$ recursively, using the coefficient estimates and the resampled errors, i.e.

$$y_t^* = \left[I_{10} \vdots U_t \vdots y_{t-1}^* \right] \begin{bmatrix} \hat{A} \\ \hat{B} \\ \hat{e} \end{bmatrix} + \hat{e}_t^*$$

where

$$U_t = \begin{bmatrix} x_{1,t} & z_{1,t} & u_{1,t} & w_{1,t} \\ x_{2,t} & z_{2,t} & u_{2,t} & w_{2,t} \\ | & | & | & | \\ x_{10,t} & z_{10,t} & u_{10,t} & w_{10,t} \end{bmatrix} \qquad y_{t-1}^* = \begin{bmatrix} y_{1,t-1}^* \\ y_{2,t-1}^* \\ | \\ y_{10,t-1}^* \end{bmatrix}$$

(10 X 4) (10 X 1)

(t=1961,...,1978)

$$\hat{A} = \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \\ \hat{a}_{10} \end{bmatrix} \qquad \hat{B} = \begin{bmatrix} \hat{b} \\ \hat{c} \\ \hat{d} \\ \hat{e} \\ \hat{f} \end{bmatrix}$$

where

$$y_{1960}^* = y_{1960}$$

(Note that we have juggled the notation used in (II.49e), to illustrate the recursiveness of the equation)

Note also that the assumptions of the RDFOR model have been respected, in generating the pseudo-data: the assumed interregional correlation structure has not been disturbed; also, the errors are

selected independently from one another, preserving the assumption of independence in time.

3) From the pseudo-data, a vector of coefficients

$$\hat{C}^* = \begin{bmatrix} \hat{A}^* \\ \hat{B}^* \\ \hat{e}^* \end{bmatrix}$$

was calculated, by

$$\hat{C}^* = \left[V^{*\prime} \hat{\Sigma}^{*-1} V^* \right]^{-1} V^{*\prime} \hat{\Sigma}^{*-1} Y^*$$

4) the steps (1) - (3) were repeated 100 times (B = 100)

Results of the bootstrapping, as well as the results based on the standard asymptotic calculations, are presented here in TABLE 2.3 reproduced from Freedman and Peters (1984)

Nominal SE (col. 1) values present the GLS estimates for the standard errors of the coefficients. They are the root of the diagonal entries of $\hat{cov}(\hat{C})$. Calculations for columns (2),(3) and (4) were done using the following

$$SD = \left\{ \text{diagonal entries of } \left[100^{-1} \sum_{b=1}^{100} \left[\hat{C}_b^* - \hat{C}_\cdot^* \right] \left[\hat{C}_b^* - \hat{C}_\cdot^* \right]' \right] \right\}^{0.5}$$

where

$$\hat{C}_\cdot^* = 100^{-1} \sum_{b=1}^{100} \hat{C}_b^*$$

they also used

$$\text{RMS Nominal SE} = 100^{-0.5} \left[\sum_{b=1}^{100} \text{Nominal SE}_b^{\,2} \right]^{0.5} \qquad (II.50)$$

Table 2.3

Asymptotic and bootstrap results for the RDFOR Model

(From Freedman and Peters (1984))

	GLS results	Bootstrap results		
	(1)	(2)	(3)	(4)
	Nominal SE	SD	RMS Nom. SE	RMS Boot. SE
Coefficient				
a_1	.31	.54	.19	.43
a_2	.31	.55	.19	.43
a_3	.31	.55	.19	.43
a_4	.30	.53	.18	.41
a_5	.32	.55	.19	.44
a_6	.30	.53	.18	.41
a_7	.32	.55	.19	.44
a_8	.32	.55	.19	.44
a_9	.29	.51	.18	.40
a_{10}	.31	.54	.19	.42
b	.013	.025	.0084	.020
c	.031	.052	.019	.043
d	.019	.028	.011	.022
e	.025	.042	.017	.034
f	.021	.039	.014	.029

The table clearly shows, comparing the results in column (1) to those in column (2), that, as the authors claim, the standard errors computed using the traditionnal asymptotics undervalue the variability of the coefficient estimate $\hat{C}$. The authors have used the statistic "RMS Nominal SE" to standardize for the fact that the bootstrapped results are the result of resampling, giving the equivalent of a much larger sample than the original; as such, it appears more relevant to compare the results of columns (2) and (3): these illustrate even more clearly the undervaluing of the variability of the traditionnal asymptotic estimates. Note also the consistently lower values of column (3), with respect to column (1): we shall use this result later. (see Chapter III)

The authors also attempted to verify the validity of SD, the bootstrap estimator of the variability of the random vector $\hat{C}$; they did so by resampling 100 times with replacement from each of the 100 vectors of residuals $\hat{e}_t^*$ which had been themselves generated in the original bootstrap; they thus generated sets of vectors $\hat{e}_t^{**}$ and, correspondingly, set of pseudovalues of y through

$$y_t^{**} = \left[I_{10} \vdots U_t \vdots y_{t-1}^{**} \right] \begin{bmatrix} \hat{A}^* \\ \hat{B}^* \\ \hat{e}^* \end{bmatrix} + \hat{e}_t^{**}$$

Note that each generation of pseudo-values here uses the bootstrap estimates of coefficients from the base set. This allowed calculation of

$$\begin{bmatrix} \hat{A}^{**} \\ \hat{B}^{**} \\ \hat{e}^{**} \end{bmatrix}$$

and of a measure of variability defined as

$$\text{RMS Boot SE} = \left[B^{*-1} \sum_{b=1}^{B^*} (SD_b^*)^2 \right]^{0.5}$$

where

$$(SD_b^*)^2 = B^{*-1} \left[\sum_{b=1}^{B^*} \left[\hat{C}_b^{**} - \hat{C}_\cdot^{*} \right] \left[\hat{C}_b^{**} - \hat{C}_\cdot^{*} \right]' \right]$$

where

$$\hat{C}_\cdot^{*} = B^{*-1} \sum_{b=1}^{B^*} \hat{C}_b^{**}$$

The results of these calculations are in column (4) of TABLE 2.3. Note that the values in col. (4) , although all lower than those in col (2), are much higher than those of col.(3) and still consistently and considerably higher than those of col.(1).

Hence, in this application, if one accepts the premises and application made of the bootstrap, the latter method has shown itself a more valid tool in evaluating the true variability of the coefficient estimates. We note finally that Freedman and Peters provide some theoretical justification why the traditional asymptotics tend to underestimate the variability of coefficient estimates.

We end this section and this chapter, by describing some Bootstrap asymptotic results that have been derived by Freedman (1984) for some simultaneous equation models. From now on, we shall be using the latter's notation, showing equivalences, when needed, with the results given in Chapter I for simultaneous equations.

The results taken from Freedman (1984) that we will be describing are applicable to cross-sectional data, where we draw a simple random sample

from a population. Omitting, to simplify notation, all subscripting for sampling order (usually denoted by t=1, 2,..., n), we can consider the following model, for the 1th equation in an q equation system

$$\underset{(1X1)}{y_1} = \underset{(1Xp)}{U_1}\,\underset{(pX1)}{A_1} + \underset{(1X1)}{e_1} \qquad \text{(II.51a)}$$

where

$$U_1 = \left[\, Y_1 \,;\, X_1 \,\right]$$

where Y_1 is the (1 X m_1) random vector of endogenous variables included in equation "1" whereas X_1 is the (1 X k_1) random vector of included exogenous variables ($m_1 + k_1 = p$) . Define also an (r X 1) vector of instruments V , ($r \geq p$), which can be written as

$$V = \begin{bmatrix} X_1' \\ X_1^{*\prime} \end{bmatrix} \qquad \text{(II.51b)}$$

where X_1^* is the subvector of X of all exogenous variables not in the 1th equation; the condition ($r \geq p$) assures the identifiability of the equations: the Order Condition of Identifiability as stated by Goldberger (1962) states this condition as being $k_1^* \geq m_1$, where k_1^* is the number of exogenous variables excluded from the 1th equation, and m_1, the number of included RHS endogenous variables. This can be written as

$$r = k_1^* + k_1 \geq k_1 + m_1 = p$$

in the current context.

Assume also, for the above model, the convergence, in Mallows metric terms, of the empirical distribution of the vector (y_1, U_1, V_1) and of its theoretical distribution, to the same distribution. (We shall hereafter omit the italic subscript "1"; but the derivations will always refer to the 1th equation of the system.) Finally, assume $E(V_1 e_1) = 0$

Let

$$Q = E(V\ y) \qquad R = E(V\ U) \qquad S = E(V\ V') \qquad \text{(II.51c)}$$
$$(r\text{X}1) \qquad (r\text{X}p) \qquad (r\text{X}r)$$

where Q, R, and S are constant. These conditions are essential for the uniqueness of the coefficient vector. Multiplying (II.51a) by V and taking expectations, we will obtain a unique value for A, if r(R) = p, from

$$Q = R\ A$$

(Note that the above model assumes stronger conditions than those described in our Simultaneous Equation model in Chapter I: there, it was assumed that the sample was part of a time series, with non-zero correlations between adjacent members of the series; here we are assuming that total independence exists between members of the sample. A fortiori, we will therefore obtain consistent estimates).

We now take n observations in the random vectors U, V, y and e. Let

$$\left[y_t,\ U_t,\ V_t,\ e_t \right]$$

be independent and distributed as (y, U, V, e). Let

$$E\left[V_t\ e_t \right] = 0$$

$$y_t = U_t\ A + e_t \qquad \text{(II.52)}$$

The 2SLS estimate of A is achieved as follows. Denote

$$Q_n = n^{-1} \sum_{t=1}^{n} V_t\ y_t \qquad S_n = n^{-1} \sum_{t=1}^{n} V_t\ V_t'$$

$$R_n = n^{-1} \sum_{t=1}^{n} V_t\ U_t \qquad \Delta_n = n^{-1} \sum_{t=1}^{n} V_t\ e_t$$

Note that we may also write the above expressions as

$$R_n = n^{-1} \underset{(rXn)}{\left[V_1 \vdots V_2 \vdots \ldots \vdots V_n \right]} \underset{(nXp)}{\begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{bmatrix}} = \underset{(rXp)}{V\,U/\,n}$$

$$Q_n = n^{-1} \underset{(rXn)}{\left[V_1 \vdots V_2 \vdots \ldots \vdots V_n \right]} \underset{(nX1)}{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}} = \underset{(r\ 1)}{V\,y/\,n}$$

$$S_n = n^{-1} \underset{(rXn)}{\left[V_1 \vdots V_2 \vdots \ldots \vdots V_n \right]} \underset{(nXr)}{\begin{bmatrix} V_1' \\ V_2' \\ \vdots \\ V_n' \end{bmatrix}} = \underset{(rXr)}{V\,V'/\,n}$$

$$\Delta_n = n^{-1} \underset{(rXn)}{\left[V_1 \vdots V_2 \vdots \ldots \vdots V_n \right]} \underset{(nXp)}{\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}} = \underset{(rX1)}{V\,e/\,n}$$

Pre-multiplying (II.52) by V_t and summing over (t=1,2,...,n), we obtain, using the above notation

$$\underset{(rX1)}{Q_n} = \underset{(rXp)(pX1)}{R_n\ A} + \underset{rX1)}{\Delta_n} \qquad (II.53)$$

A unique solution exists for this system, which can be found by least squares. In fact, A can be estimated using generalized least squares (GLS) where the components of Δ_n will have a covariance structure Σ allowing non-zero correlation among them, but all component having identical variance

$$\mathrm{Var}(\Delta_n/V) = E\left[\left[n^{-1} \sum_{t=1}^{n} V_t\, e_t \right] \left[n^{-1} \sum_{t=1}^{n} e_t\, V_t' \right] / V \right]$$

$$= n^{-2}\, E\left[V\, ee'V/V \right] = n^{-2}\, \sigma^2\, VV' = n^{-1}\sigma^2\, S_n$$

where

$$E(ee') = \sigma^2 I_n$$

we are conditionning on the matrix of instruments, treating them here as constants. This is one possibility, since the instruments, including the lagged endogenous variables, are independent from the current operation of the system (Theil, 1971), and may be considered either as constants, or as uncorrelated random variables (a different set of results could be derived, if one were to consider them to be the latter type).

And so, applying GLS to (II.53), we obtain

$$\hat{A}_n = \left[R_n' S_n^{-1} R_n\right]^{-1} R_n' S_n^{-1} Q_n \qquad \text{(II.54)}$$

provided S_n^{-1} exists.

This is the conventionnal 2SLS estimator. Note that we may rewrite the above in the form submitted in Chapter I, with which we proved the consistency of this estimator. Translating Freedman's notation into ours, we have, for the first equation,

$$U_t = \left[Y_{1,t} \vdots X_{1,t} \right]$$

$$V_t = \begin{bmatrix} X_{1,t}' \\ X^*_{1,t} \end{bmatrix} = X_t'$$

$$y = y_{1,t}$$

the estimator given in (II.54) becomes

$$\hat{A}_n = \left[U'V'/n \; (V'V/n)^{-1} \; VU/n \right]^{-1} U'V'/n \; (V'V/n)^{-1} \; Vy/n$$

$$= n\, n^{-1} \left\{ \begin{bmatrix} Y_1' \\ X_1' \end{bmatrix} X \, (X'X)^{-1} X' \left[Y_1 \vdots X_1 \right] \right\}^{-1} \begin{bmatrix} Y_1' \\ X_1' \end{bmatrix} X \, (X'X)^{-1} X' y_1 \qquad \text{(II.55a)}$$

Now, since

$$V\,U = \sum_{t=1}^{n} V_t\,U_t = \sum_{t=1}^{n} \begin{bmatrix} X_{1,t}' \\ X_{1,t}^{*\prime} \end{bmatrix} \begin{bmatrix} Y_{1,t} : X_{1,t} \end{bmatrix}$$

$$= \begin{bmatrix} X_1' Y_1 & X_1' X_1 \\ X_1^{*\prime} Y_1 & X_1^{*\prime} X_1 \end{bmatrix} = X' \begin{bmatrix} Y_1 : X_1 \end{bmatrix}$$

then (II.55a) can be written as

$$\hat{A}_n = \begin{bmatrix} Y_1'X\,(X'X)^{-1}X'Y_1 & Y_1'X\,(X'X)^{-1}X'X_1 \\ X_1'X\,(X'X)^{-1}X'Y_1 & X_1'X\,(X'X)^{-1}X'X_1 \end{bmatrix}^{-1} \begin{bmatrix} Y_1'X\,(X'X)^{-1}X'y_1 \\ X_1'X\,(X'X)^{-1}X'y_1 \end{bmatrix}$$

$$= \begin{bmatrix} Y_1'X\,(X'X)^{-1}X'Y_1 & Y_1'X(X'X)\;X'X \begin{bmatrix} I_{K_1} \\ 0 \\ (k-k_1)Xk_1 \end{bmatrix} \\ \begin{bmatrix} I_{K_1} \\ 0 \end{bmatrix}' X'X\,(X'X)^{-1}X'Y_1 & \begin{bmatrix} I_{K_1} \\ 0 \end{bmatrix}' X'X\,(X'X)^{-1}X'Y_1 \end{bmatrix}$$

$$X \begin{bmatrix} Y_1'X\,(X'X)^{-1}X'y_1 \\ \begin{bmatrix} I_{K_1} \\ 0 \end{bmatrix}' X'X\,(X'X)^{-1}X'y_1 \end{bmatrix}$$

$$= \begin{bmatrix} Y_1'X\,(X'X)^{-1}X'Y_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} Y_1'X\,(X'X)^{-1}X'y_1 \\ X_1'y_1 \end{bmatrix}$$

We note that this is equivalent to premultiplying equation (II.51a) by the matrix constructed from the vectors

$$V_{1,t} = \begin{bmatrix} Y_{1,t}'X\,(X'X)^{-1}X' \\ X_{1,t}' \end{bmatrix} = \begin{bmatrix} \hat{Y}_{1,t}' \\ X_{1,t}' \end{bmatrix} \quad (t=1,\ 2,\ldots,n) \qquad \text{(II.55b)}$$

Of course, we must rewrite equation (II.51a) in the following form before the premultiplication

$$y_{1,t} = Y_{1,t} A_{11} + X_{1,t} A_{21} + e_{1,t}$$

Finally, the estimator for the asymptotic covariance of the estimator $\hat{A}_n$ that we obtained in (I.28) was

$$\hat{VAR}(\hat{A}_n) = s^2 \begin{bmatrix} Y_1'X\,(X'X)^{-1}X'Y_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix}^{-1}$$

which, translated into Freedman's notation becomes

$$s^2 n^{-1} \left[R_n' \, S_n^{-1} \, R_n \right]^{-1} \qquad \text{(II.56)}$$

The traditionnal asymptotics for the distribution of

$$n^{0.5} \left[\hat{A}_n - A \right]$$

can now be given. With the vector of coefficients in the form (II.54), the formula in (II.32b) becomes

$$n^{0.5} \left[\hat{A}_n - A \right] \sim N_r \left[0, \sigma^2 (R'S^{-1}R)^{-1} \right]$$

with

$$R_n \xrightarrow{p} R \quad \text{and} \quad S_n \xrightarrow{p} S$$

that is the matrices R_n and S_n converge weakly to respective constants R and S.

The above procedure assumes homoscedasticity. As was done in (II.32c), and for the same reasons, we need to center the errors, which will assure weak convergence of $\hat{A}_n$ to a constant; hence whether we have homoscedasticity or not, convergence will occur.

The centering is done as follows. Let the errors be

$$\hat{e}_t = y_t - U_t \hat{A}_n$$

next, calculate

$$\tilde{e}_t = \hat{e}_t - \hat{V}_t' \hat{b}_n' \qquad \text{(II.58)}$$

where

$$\hat{b}_n = S_n^{-1} n^{-1} \sum_{t=1}^{n} V_t \hat{e}_t$$

(Note that

$$n^{-1} \sum_{t=1}^{n} V_t \tilde{e}_t = n^{-1} \sum_{t=1}^{n} V_t \hat{e}_t - n^{-1} \sum_{t=1}^{n} V_t V_t' S_n^{-1} \left[\sum_{t=1}^{n} V_t \hat{e}_t \right]$$

$$= \hat{\Delta}_n - \hat{\Delta}_n = 0 \qquad \text{(II.59)}$$

and so we are now fully respectful of the assumption of orthogonality of the instruments with the errors)

We may now apply the bootstrap.

1) Let u be the empirical distribution of

$$\left[U_t, V_t, \tilde{e}_t \right] \quad (t=1,2,\ldots,n)$$

$$u : 1/n$$

Thus u assigns measure 1/n to each vector above.

2) Resampling proceeds as follows. Given the data $\left[y_t, U_t, V_t \right]$ $(t=1,2,\ldots,n)$, let $\left[U_s, V_s, e_s \right]$ $(s=1,2,\ldots,n)$ be independent with distribution u.

Using these resampled vectors, generate the pseudo data

$$y_s^* = U_s^* \hat{A}_n + e_s^*$$

Note that this procedure preserves any relationship there may be between

instruments and disturbances. Finally, calculating

$$Q_n^* = n^{-1} \sum_{s=1}^{n} V_s^* y_s^* \qquad S_n^* = n^{-1} \sum_{s=1}^{n} V_s^* (V_s^*)'$$

$$R_n^* = n^{-1} \sum_{s=1}^{n} V_s^* U_s^* \qquad \Delta_n^* = n^{-1} \sum_{s=1}^{n} V_s^* \hat{e}_s^*$$

we may obtain

$$\hat{A}_n^* = \left[R_n^{*\prime} S_n^{*-1} R_n^*\right]^{-1} R_n^{*\prime} S_n^{*-1} Q_n^*$$

$$= \hat{A}_n + \left[R_n^{*\prime} S_n^{*-1} R_n^*\right]^{-1} R_n^{*\prime} S_n^{*-1} \Delta_n^*$$

A number of conclusions may be reached on the asymptotics of this procedure. The bootstrap principle assumes that the error structure of the starred estimates imitates that in the original estimates. The following theorem formalizes this, from Freedman (1984). We omit the proof.

THEOREM Along almost all sample sequences, as $n \longrightarrow \infty$,

a) $R_n^* \xrightarrow{p} R$, $S_n^* \xrightarrow{p} S$ and $Q_n \xrightarrow{p} Q$, conditionnal on the sample data;

b) the conditional law of (1) $(n^{0.5} \Delta_n^*)$ and the unconditional law of (2) $(n^{0.5} \Delta_n)$ converge to the same limit . Further, the Mallows metric distance d_r^2 between (1) and (2) tends to 0.

c) the conditional law of $n^{0.5}\left[\hat{A}_n^* - \hat{A}_n\right]$ and the conditional law of $n^{0.5}\left[\hat{A}_n - A\right]$ have the same limit, also.

The above model and bootstrapping procedure are for cross-sectional data. We shall see, in the next chapter, how the dynamic nature of the

model we used was taken into consideration in the specific bootstrap process adopted. For now, let us simply state that the asymptotic results for the dynamic model can be shown to be the same, in terms of the statements in parts (a), (b) and (c) of the above theorem, as for the model using cross-sectional data (Freedman (1984)).

CHAPTER III

In this chapter, we shall describe the application of the bootstrap to a dynamic simultaneous system of equations, Klein's Model I. The last section of Chapter II described Freedman's (1984) asymptotic results for bootstrapping the 2SLS estimator in linear simultaneous equation systems which use cross-sectional data; but he also describes the asymptotics for dynamic models, in the same paper. In this Chapter, we shall first describe Freedman's asymptotic results for the bootstrapping of the 2SLS estimator in dynamic simultaneous linear equation models; then, after a description of Klein's Model I, we shall apply a method extracted from Freedman's results to bootstrap the 2SLS estimates calculated from Klein's data.

Before describing Freedman's bootstrapping of the 2SLS estimator for dynamic models, we will first lay out the assumptions for dynamic models. Using Freedman's notation, we may consider observable random vectors X_t and Y_t, obtained for each time period t (t=1,2,...,n); the components of these vectors are considered to interrelate in an m-equation system which can be written as

$$\underset{(1Xm)}{Y_t} = \underset{(1Xm)(mXm)}{Y_t\ A} + \underset{(1Xm)\ (mXm)}{Y_{t-1}\ B} + \underset{(1Xk)(kXm)}{X_t\ C} + \underset{(1Xm)}{e_t} \quad \text{(III.1)}$$

Where A, B and C are coefficient matrices where certain coefficient terms are constrained to 0, and where

$$X_t = \left[\, 1,\ X_{1t},\ X_{2t},\ \ldots\ ,X_{kt} \,\right]$$

Also, assume that the vectors $\left[\, X_t \,:\, e_t \,\right]$ are i.i.d ; Note that this assumption is stronger than the analogous assumption given in the general linear simultaneous model where the e_t are assumed simply uncorrelated

(and not necessasily independent) and where the X_t are part of a stationnary multivariate stochastic process (SMSP) with rapidly diminishing dependence.

Equally, assume that the empirical probabilities of the sample $\{X_t, e_t\}$ and the theoretical probabilities from which the $\{X_t, e_t\}$ are drawn converge in a Mallows metric sense.

We further assume that the X_t are orthogonal to all the e_t, in the sense $E\left[X_t' e_t \right] = 0$ and that $E(e_t) = 0$.

Finally, we must admit some assumptions on the generation of the vectors Y_t. Assume the Y_t are generated by an SMSP. We may compute the Y_t in a recursive way. Using (III.1), we may write

$$
\begin{aligned}
Y_t &= (Y_{t-1}B + X_t C + e_t)(I - A)^{-1} \\
&= \left[(Y_{t-2}B + X_{t-1} C + e_{t-1})(I - A)^{-1}B \right. \\
&\qquad \left. + X_t C + e_t\right](I - A)^{-1} \\
&= Y_{t-s+1}\left[B (I - A)^{-1}\right]^{s+1} + X_{t-s}\left[C (I - A)^{-1}\right]\left[B (I - A)^{-1}\right]^{s} \\
&\quad + \ldots + X_t \left[C (I - A)^{-1}\right] + e_{t-s} (I - A)^{-1}\left[B (I - A)^{-1}\right]^{s} + \ldots \\
&\quad + e_{t-s} (I - A)^{-1} \\
&= Y_{t-s+1}\left[B (I - A)^{-1}\right]^{s+1} + \sum_{s=0}^{\infty}\left[(X_{t-s}C + e_{t-s})(I - A)^{-1}\left[B (I - A)^{-1}\right]^{s}\right]
\end{aligned}
\qquad \text{(III.2)}
$$

The last equation is identical to equation (I.6) in Chapter I, with

$$C_1^{s+1} = \left[B\,(I - A)^{-1}\right]^{s+1}$$

$$C_2\,C_1^{\tau} = C\,(I - A)^{-1}\left[B\,(I - A)^{-1}\right]^{s}$$

$$V_t = e_t\,(I - A)^{-1}$$

Hence, by the reasoning given after equation (I.6), the present system will be stable if

$$\lim_{\tau\to\infty}\left[B\,(I - A)^{-1}\right]^{\tau} = 0 \qquad \text{(III.3)}$$

which requires that all the eigenvalues of $\left[B\,(I - A)^{-1}\right]^{\tau}$ be less than 1 in absolute value (assuming that there are no multiple eigenvalues). If this condition is respected, then (III.2) becomes

$$Y_t = \sum_{s=0}^{\infty}\left[(X_{t-s}C + \varepsilon_{t-s})(I - A)^{-1}\left[B\,(I - A)^{-1}\right]^{s}\right] \qquad \text{(III.4)}$$

The above description applies to the entire system of m equations. If we now focus on, say, the first equation, we may write, using Freedman's (1984) notation, with slight modifications:

$$\underset{(1X1)}{y_t} = \underset{(1x1)}{U_t\alpha} + \delta_t \qquad \text{(III.5)}$$

Basically, this is equation (III.1) where all LHS endogenous variables have been dropped, except the one assigned to the first equation, i.e y_t ; and where

$$U_t = \underset{(1Xm_1)}{\left[\,Y_{1,t}\right.} \;;\; \underset{(1Xm_1')}{Y_{1,t-1}} \;;\; \underset{(1Xk_1)}{\left.X_{1,t}\,\right]}$$

here m_1 , m_1' and k_1 are the numbers of endogenous, lagged endogenous and exogenous variables included in the first equation, respectively. The coefficient vector α is therefore the first columns of A, B, and C in

(III.1) stacked one on top of one another, but from which we have omitted the 0's .

Further, we may define an ((m+k) X1) ($m + k \geq m_1 + m_1' + k_1$)vector of instruments

$$V_t = \begin{bmatrix} Y_{t-1}' \\ X_t' \end{bmatrix}$$

As already demonstrated in Chapter II, we may use, equivalently,

$$V_t = \begin{bmatrix} \hat{Y}_{1,t}' \\ Y_{1,t-1}' \\ X_{1,t} \end{bmatrix}$$

where the subscript designates only those variables included in the first equation. Here, since we are dealing with the dynamic model, and since the lagged endogenous variables can be shown to be independent of the current operation of the system, we have included the latter in the vector of instruments. See equation (II.55b); (here, the notation $\left[X_{1,t}\right]$ of that equation is seen to be equivalent to $\left[Y_{1,t-1} \vdots X_{1,t}\right]$

Of course, as for the full system, we should have $E\{V_t \delta_t\} = 0$ (the brackets {} refer to the entire set of variables in t, (t=1,2,..,n)). If the V_t are considered constants, then we shall assume $E\{\delta_t\} = 0$. Here too, $\left[y_t, U_t, V_t, \delta_t \right]$ is stationnary and ergodic.

Assume, as in the cross-sectional model, that

$$Q = \underset{(rX1)}{E\{V_t y_t\}}$$

101

$$R = \underset{(rXp)}{E\{V_t U_t\}}$$

$$S = \underset{(r \times r)}{E\{V_t V_t'\}}$$

where Q, R, and S are matrices of constants with respect to t. There is no dependence on t, due to the assumed stationarity. Premultiplying (III.5) by V_t and taking expectations, we will have a unique vector α and an identifiable system $Q = R\alpha$, provided $r(R) = p$ (Otherwise α will not be unique). The condition of uniqueness of α also necessitates that $(r \geq p)$. Assume finally that the determinant of S is non-zero.

We then apply data for periods $(t=1,\ldots,n)$ into our model. As for the cross-sectioned data, we can estimate $\hat{\alpha}$ by instrumental variables regression. Let

$$Q_n = n^{-1} \sum_{t=1}^{n} V_t y_t \qquad R_n = n^{-1} \sum_{t=1}^{n} V_t U_t$$

$$S_n = n^{-1} \sum_{t=1}^{n} V_t V_t' \qquad \Delta_n = n^{-1} \sum_{t=1}^{n} V_t \delta_t$$

It can be shown, by the ergodic theorem, that

$$Q_n \xrightarrow{p} Q \qquad R_n \xrightarrow{p} R \qquad S_n \xrightarrow{p} S$$

As for the cross-sectional model the standard 2SLS estimator to be bootstrapped is

$$\hat{\alpha}_n = \left[R_n' S_n^{-1} R_n\right]^{-1} R_n' S_n^{-1} Q_n$$

with estimator for asymptotic covariance given by

$$\hat{COV}(\hat{\alpha}_n) = s^2 n^{-1} \left[R_n' S_n^{-1} R_n\right]^{-1}$$

Of course, again as for the cross-sectional model, we can obtain the observed errors from

$$\hat{\delta}_t = y_t - U_t \hat{\alpha}_n$$

and we may obtain $V_t \perp \delta_t$ by centering the errors using

$$\tilde{\delta}_t = \hat{\delta}_t - \hat{b}_n' V_t$$

where

$$\hat{b}_n = S_n^{-1} n^{-1} \sum_{t=1}^{n} V_t \hat{\delta}_t$$

(see equation II.58 and II.59)

All this can be computed for any equation of the system. Having thus estimated α for each equation, we may construct, by padding with the 0 coefficient constraints, the matrices

$$\underset{(mXm)}{\hat{A}_n} \qquad \underset{(mXm)}{\hat{B}_n} \qquad \underset{(kXm)}{\hat{C}_n}$$

Also, we build $\tilde{e}_t$ (1 X m) from the errors of the various equations put together

$$\tilde{e}_t = \left[\tilde{\delta}_{1t}, \tilde{\delta}_{2t}, \ldots, \tilde{\delta}_{mt} \right]$$

Now, for the bootstrap, as suggested by Freedman; we do the following

1) let F : 1/n be the empirical distribution of $(X_t, \tilde{e}_t)$ $(t=1,2,\ldots,n)$.

2) Let us choose the random vectors (X_s^*, e_s^*) from F independently

$$(X_s^*, e_s^*) \sim_{iid} F$$
$$(s=1,2,\ldots,n)$$

s = 0 some arbitrary point.

3) assume that the Y_t^* converge to finite values and that

$$Y_t^* = \sum_{s=0}^{\infty} \left[(X_{t-s}^* \hat{C}_n + e_{t-s}^*)(I - \hat{A}_n)^{-1} \left[\hat{B}_n (I - \hat{A}_n)^{-1} \right] \right]^s \quad \text{(III.6)}$$

Then, the pseudo-data thus generated will satisfy the model

$$Y_t^* = Y_t^* \hat{A}_n + Y_{t-1}^* \hat{B}_n + X_t^* \hat{C}_n + e_t^*, \quad X_t^* \perp e_t^*$$

Having thus selected our bootstrap procedure to respect the basic assumptions of the model, one can proceed to the calculation of bootstrap coefficient and $\hat{COV}_*$ estimates in the same way as in the original estimates. That is, for example, for the first equation, dropping the components of Y_t^*, Y_{t-1}^*, and X_t^* not in the equation, we estimate $\hat{\alpha}_n$ in the model

$$\underset{(1X1)}{y_t^*} = U_t^* \hat{\alpha}_n + \underset{(1x1)}{\delta_t^*} \quad \text{(III.7)}$$

with

$$\hat{\alpha}_n^* = \left[R_n^{*\prime} S_n^{*-1} R_n^* \right]^{-1} R_n^{*\prime} S_n^{*-1} Q_n^* \quad \text{(III.8)}$$

with covariance matrix

$$\hat{COV}_*(\hat{\alpha}_n^*) = (s^2)^* n^{-1} \left[R_n^{*\prime} S_n^{*-1} R_n^* \right]^{-1} \quad \text{(III.9)}$$

where

$$Q_n^* = n^{-1} \sum_{t=1}^{n} V_t^* y_t^* \qquad R_n^* = n^{-1} \sum_{t=1}^{n} V_t^* U_t^*$$

$$S_n^* = n^{-1} \sum_{t=1}^{n} V_t^* V_t^{*\prime}, \quad (s^2)^* = n^{-1} \hat{\delta}_t^{*\prime} \hat{\delta}_t^* \quad \text{(III.10)}$$

Note that $\hat{\alpha}_n^*$ can also be expressed as

$$\hat{\alpha}_n^* = \left[R_n^{*\prime} S_n^{*-1} R_n^* \right]^{-1} R_n^{*\prime} S_n^{*-1} \left[R_n^* \hat{\alpha}_n + \Delta_n^* \right]$$

$$= \hat{\alpha}_n + \left[R_n^{*\prime} S_n^{*-1} R_n^{*} \right]^{-1} R_n^{*\prime} S_n^{*-1} \Delta_n^{*}$$

Finally, we state, without proof, the bootstrap principle, as given by Freedman

THEOREM

Along almost all sample sequences, as $n \longrightarrow \infty$, and conditionnally on the data:

a) $Q_n^{*} \xrightarrow{P} Q$ and $R_n^{*} \xrightarrow{P} R$ and $S_n^{*} \xrightarrow{P} S$,

(in conditionnal probability)

b) the conditionnal law of $n^{0.5}\Delta_n^{*}$ has the same limit, as the unconditionnal law of $n^{0.5}\Delta_n$.

We note in passing that the unconditionnal law of $n^{0.5}\Delta_n$ can be expressed as

$$n^{0.5}\left[R_n' S_n^{-1} R_n \right]^{-1} R_n' S_n^{-1} \Delta_n \sim N_r\left[0, \sigma^2 (R' S^{-1} R)^{-1} \right]$$

(see equation (II.56), and since $n^{0.5}\left[\hat{\alpha}_n^{*} - \hat{\alpha}_n^{*}\right]$ is equal to the LHS of above.

Having thus described Freedman's (1984) results, which show the convergence of the bootstrap, as applied by him, to standard results, we may apply the bootstrap to a specific model. We therefore chose a very well-known and straightforward model, one, to quote Theil, "...which is a favorite drilling ground of theoretical econometricians," that is,

Klein's Model I.

Let us first describe this model; the description is from Theil (1971). The equations of the model are, for time t (t=1,...,n)

$$(I)\quad C_t = b_0 + b_1 P_t + b_2 P_{t-1} + b_3 (W_t + W_t') + e_t$$

$$(II)\quad I_t = b_0' + b_1' P_t + b_2' P_{t-1} + b_3' K_{t-1} + e_t'$$

$$(III)\quad W_t = b_0'' + b_1'' X_t + b_2'' X_{t-1} + b_3'' (t - 1931) + e_t''$$

$$(IV)\quad X_t = C_t + I_t + G_t$$

$$(V)\quad P_t = X_t - W_t - T_t$$

$$(VI)\quad K_t = K_{t-1} + I_t$$

Where the endogenous variables were

C_t : aggregate consumption in year t.

P_t : total profits in year t.

W_t : wage bill paid by industry on year t.

K_t : capital stock in year t.

I_t : net investment in year t.

X_t : total production of private industry in year t.

and where the predetermined variables were

1 : (nX1) unit vector.

W_t' : government wage bill in year t.

T_t : taxes in year t.

G_t : government non-wage expenditures

t: year (t = 1920,...,1941)

Also P_{t-1} , K_{t-1} , and X_{t-1} were included, being lagged endogenous.

Equations (I), (II) and (III) are statistical, inexact; equations (IV), (V) and (VI) are definitional. Note that the simultaneity of equations (I), (II) and (III) is made explicit when one considers the definitional equations.

Using the data provided by Theil (1971) we computed the coefficient estimates and their covariance matrices, for equations (I), (II) and (III) . The calculations were done on VAX 8550 (Concordia University), utilising the SAS statistical package, and are shown in Table 3.1.

Table 3.1 reports the estimates for the coefficients, as well as, in parentheses and below each, their respective standard errors. The latter are computed as the square roots of the diagonal elements of $\hat{COV}(\hat{\alpha}_n)$ their variance-covariance matrix.

The above results almost exactly match results given by Goldberger (1962) and Theil (1971). Only two (2) asymptotic standard error estimates differ from these two sources, and there only by a margin of less than 1/10 % .

Note that the Klein system may be written in the form of (III.1); incorporating the above estimates, as well as the values of observed residuals calculated from each equation, we may write for (t=1,2,...,n)

$$\underset{(21X6)}{Y} = \underset{(21X6)}{Y}\underset{(6X6)}{\hat{A}_n} + \underset{(21X6)}{\bar{Y}}\underset{(6X6)}{\hat{B}_n} + \underset{(21X5)}{X}\underset{(5X6)}{\hat{C}_n} + \underset{(1X6)}{\hat{e}}$$

where $\bar{Y}$ expresses the (21 X 6) matrix of lagged endogenous vectors:

Table 3.1

Two-stage Least-squares Estimates (Std. error)

$$C_t = \underset{(1.321)}{16.555} + \underset{(0.117)}{0.017}\, P_t + \underset{(0.107)}{0.216}\, P_{t-1} + \underset{(0.040)}{0.810}\, (W_t + W_t')$$

$$I_t = \underset{(7.523)}{20.278} + \underset{(0.173)}{0.150}\, P_t + \underset{(0.162)}{0.616}\, P_{t-1} - \underset{(0.036)}{0.158}\, K_{t-1}$$

$$W_t = \underset{(1.147)}{1.500} + \underset{(0.036)}{0.439}\, X_t + \underset{(0.039)}{0.147}\, X_{t-1} + \underset{(0.029)}{0.130}\, (t-1931)$$

and where

$$\hat{A}_n = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ .017 & 0 & 0 & .150 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & .439 & 0 & 0 & 0 \\ .810 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\hat{B} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ .216 & 0 & 0 & .616 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & .147 & 0 & 0 & 0 \\ 0 & 0 & 0 & -.158 & 0 & 0 \end{bmatrix}$$

$$\hat{C}_n = \begin{bmatrix} 16.555 & 0 & 1.50 & 20.278 & 0 & 0 \\ 0 & 0 & 0.130 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and finally where

$$\hat{e} = \left[\hat{\delta}_I \mid \hat{\delta}_{II} \mid \ldots \mid \hat{\delta}_{VI} \right] \qquad \text{(III.11)}$$

(21X6)

are the observed errors of the inexact equations (I), (II) and III respectively. The last three (3) columns of $\hat{e}$ are simply the values of the endogenous variables that do not have their own inexact relations: P_t, K_t and X_t. (Note that their respective columns in $\hat{A}_n$, $\hat{B}_n$, and $\hat{C}_n$ are all columns of 0's).

To check the stability of the Klein system, hence verifying the condition in (III.3), we calculated the eigenvalues of $\hat{B}_n(I - \hat{A}_n)^{-1}$. They were $\lambda_1 = .334$; $\lambda_2, \lambda_3 = .838 \left[\cos(.435) \pm i \sin(.435)\right]$.

Hence, their absolute value (or moduli) were all less than 1 in absolute value. We were then ready to bootstrap.

As described in our report of the Freedman (1984) results previously, we first calculated

$$\hat{\delta}_t = y_t - U_t \hat{\alpha}_n$$

for each of equations (I), (II) and (III); we then centered the errors by calculating

$$\tilde{\delta}_t = y_t - \hat{b}_n' V_t \qquad \text{(III.12)}$$

where

$$V_t = \begin{bmatrix} Y_{t-1}' \\ X_t' \end{bmatrix}$$

((m+k) X 1)

$$\hat{b}_n = S_n^{-1} n^{-1} \sum_{t=1}^{n} V_t \hat{\delta}_t$$

for each equation and all t.

The calculation of the pseudovalues according to Freedman assumes the availability of a very large sample, which for all intents and purposes could be considered infinite (equation (III.6)): However, as we were dealing with a limited sample, we utilized another form, equivalent to the above, which used the recursive form

$$Y_t^* = (Y_{t-1}^* \hat{B}_n + X_t^* \hat{C}_n + \tilde{e}_t^*)(I - \hat{A}_n)^{-1} \qquad \text{(III.14)}$$

where $\hat{A}_n$, $\hat{B}_n$ and $\hat{C}_n$ have been given previously.

The Y_t^* could then be generated recursively by their predecessors Y_{t-1}^* using the resampled vectors $(X_t^*, \tilde{e}_t^*)$. The latter were generated in the following manner.

(a) First, we concatenated the (21X1) vectors of centered errors $\tilde{\delta}_t$ from equation (III.12), adding to them the raw values of the endogenous variables that did not possess an inexact relation in the system

$$\tilde{e} = \left[\tilde{\delta}_I \mid \tilde{\delta}_{II} \mid \ldots \mid \tilde{\delta}_{VII} \right]$$

(21X6)

Next, we concatenated X, the full data (21 X 5) matrix of exogenous vectors X_t, with $\tilde{e}$, to form

$$\underset{(21X12)}{\left[X \vdots \tilde{e} \right]}$$

b) Resampling: Given Klein's data, we let $\mu : 1/n$ be the empirical distribution of $(X_t, \tilde{e}_t)$ in the above matrix. We generated the $(X_t^*, \tilde{e}_t^*)$ with replacement using μ, in the following manner:

(i) first, we constructed a (21 X 1) vector u of values by selecting a sample (n = 21) from U(0,1) through the subroutine RANUNI on SAS. The following operation was then performed on the components of the vector u to give them discrete values between 1 and 21:

$$u_t\text{˘} = \left[u_t \cdot 20 \right] + 1 \quad (t=1,2,\ldots,n)$$

where u_t is the tth component of u and $\left[\ \right]$ represents truncation to the next lowest integer. The values u_t˘ are then put together to form the new vector u˘ .

(ii) Using the newly-formed (21 X 1) vector u˘ , the rows in the matrix $\left[X \vdots \tilde{e} \right]$ corresponding to the values of the u_t˘ were selected one-by-one, starting with the row number corresponding to value u_1˘ through to u_{21}˘ to make up the matrix of resampled vectors

$$\underset{(21X12)}{\left[X^* \vdots \tilde{e}^* \right]}$$

3) Generation of Y_t^* : the pseudo-values Y_t^* were then generated using (III.14)

$$Y_t^* = (Y_{t-1}^* \hat{B}_n + X_t^* \hat{C}_n + \tilde{e}_t^*)(I - \hat{A}_n)^{-1}$$

to start this recursive process, we used $Y_0^* = Y_0$.

4) Calculation of $\hat{\alpha}^*$: once all the pseudo-value vectors Y_t^*, X_t^* and $\hat{e}_t^*$ had been generated, The procedure followed that of (III.7-.10) to calculate

$$\hat{\alpha}_n^* = \left[R_n^{*\prime} S_n^{*-1} R_n^* \right]^{-1} R_n^{*\prime} S_n^{*-1} Q_n^*$$

and

$$\hat{COV}_*(\hat{\alpha}_n^*) = (s^2)^* n^{-1} \left[R_n^{*\prime} S_n^{*-1} R_n^* \right]^{-1}$$

5) The calculation of $\hat{\alpha}_n^*$ was repeated 900 times, and the vector

$$MSE_{1*}(\hat{\alpha}_{nb}^*) = \text{diagonal of } \hat{COV}_{1*}(\hat{\alpha}_{nb}^*) \qquad \text{(III.15)}$$

was calculated, where

$$\hat{COV}_{1*}(\hat{\alpha}_{nb}^*) = B^{-1} \sum_{b=1}^{B} (\hat{\alpha}_{nb}^* - \hat{\alpha}_{\cdot}^*)(\hat{\alpha}_{nb}^* - \hat{\alpha}_{\cdot}^*)' \qquad (B = 900)$$

where the $\hat{\alpha}_{nb}^*$ were the coefficient vector estimates calculated, and where

$$\hat{\alpha}_{\cdot}^* = \sum_{b=1}^{B} \hat{\alpha}_{nb}^*$$

for each of equations (I), (II) and (III) of the system. In addition, following a procedure proposed by Freedman and Peters (1984) as described in Chapter II (see equation II.50), we calculated

$$\text{MS of Nominal SE} = B^{-1} \sum_{b=1}^{B} \text{Nominal SE}_b^2 \qquad \text{(III.16)}$$

where Nominal SE_b^2 is the vector of diagonal entries of $\hat{COV}_*(\hat{\alpha}_n^*)$. The rationale behind the use of such an estimator lies in the need to compare our results with the original results of Klein , and Goldberger (1962)

and Theil (1971), where the traditionnal estimates of asymptotic variability were used.

The results of our calculations are shown in Table 3.2; averages of bootstrap regression coefficients are also shown, to give an idea of the part bias plays in the MSE values given. In this connection, we also calculated

$$MSE_{2^*}(\hat{\alpha}^*_{nb}) = \text{diagonal of } \hat{COV}_{2^*}(\hat{\alpha}^*_{nb}) \qquad \text{(III.17)}$$

where

$$\hat{COV}_{2^*}(\hat{\alpha}^*_{nb}) = B^{-1} \sum_{b=1}^{B} (\hat{\alpha}^*_{nb} - \hat{\alpha}_n)(\hat{\alpha}^*_{nb} - \hat{\alpha}_n)' \quad (B = 900) \qquad \text{(III.17)}$$

where $\hat{\alpha}_n$ is the vector of original Klein values. Hence $MSE_{2^*}(\hat{\alpha}^*_{nb})$ is an estimate of variability which includes the relative bias between the original Klein estimates and the bootstrap average $\hat{\alpha}^*_{\cdot}$.

We may offer several empirical comments on the results of TABLE 3.2. Note first that for seven (7) of the twelve coefficient values, the $MSE_{1^*}(\hat{\alpha}^*_{nb})$, as well as the MS of Nominal SE values, are lower than those of the original asymptotic estimates; note also that the variability of the coefficients for the intercepts goes down much more dramatically than that of other coefficients: no explanation can be offerred as yet for this; note in addition that, for those values of $MSE_{1^*}(\hat{\alpha}^*_{nb})$ that are higher than the original estimates, there is a seemingly sizeable difference between the original coefficient value and the bootstrap average (the one exception to this is MSE_{1^*} for b_1'). This relative bias

Table 3.2

Bootstrapping of Klein Model I Using Freedman's (1984) Method

(B = 900)

	Coeff. est.		Variability measures			
Coeff	Original Klein	Bootstrap average $(B^{-1}\sum_{b=1}^{B}\hat{\alpha}^{*}_{nb})$	Original var b_i	MSE_{1^*} (III.15)	MS of Nom. SE	MSE_{2^*} (III.17)
b_0	16.555	16.550	1.745	.0036	.0043	.0036
b_1	.017	.749	.0137	.0239	.0298	.5603
b_2	.216	.216	.0114	.0004	.0004	.0004
b_3	.810	.332	.0016	.0420	.0553	.2709
b_0'	20.278	20.272	56.596	.0150	.0290	.0143
b_1'	.150	.670	.0300	.0283	.0900	.2982
b_2'	.616	.611	.0260	.0079	.0287	.0079
b_3'	-.158	-.158	.0013	.00004	.0002	.00004
b_0''	1.500	1.477	1.316	.047	.049	.048
b_1''	.439	.423	.0013	.0385	.044	.0388
b_2''	.147	.147	.0015	.0006	.0006	.0005
b_3''	.130	.0007	.0008	.0015	.0015	.0182

is reflected in the corresponding values of $MSE_{2*}(\hat{\alpha}^{*}_{nb})$ which are much higher than either the $MSE_{1*}(\hat{\alpha}^{*}_{nb})$ values, or even the original estimates of asymptotic variance. Finally, we note that the MS of Nominal SE values, although higher than the $MSE_{1*}(\hat{\alpha}^{*}_{nb})$ values, follow the same downward trend as the latter, which is an additionnal check on the bootstrap's value as a tool to reduce the variability of coefficient estimates.

The Freedman bootstrapping method was not the only method explored in this thesis. Freedman recommended

$$V_t = \begin{bmatrix} Y'_{t-1} \\ \\ X'_t \end{bmatrix}$$

as vector of instruments for both the calculation of the 2SLS coefficients and the centering of errors. In the Klein model, there is a possibility of seven (7) lagged variables, and five (5) exogenous variables.

Another possibility is using

$$V_t = \begin{bmatrix} \hat{Y}'_{1,t} \\ \\ Y'_{1,t-1} \\ \\ X_{1,t} \end{bmatrix}$$

as instruments for the calculation of 2SLS coefficients, and

$$V_t = \begin{bmatrix} \hat{Y}_t^{\vee\prime} \\ Y_{t-1}^{\vee\prime} \\ X_t^{\vee\prime} \end{bmatrix}$$

for the centering of errors.

The latter set regroups all the components of Y_t which are estimated by $\hat{Y}_t$ for equations (I) (II) and (III), estimates which are used in the second stage of the 2SLS estimation method, and which have been written as $\hat{Y}_t^{\vee}$. It also gathers all the lagged endogenous and exogenous variables actually contained in equations (I), (II) and (III): this is written as $Y_{t-1}^{\vee}$ and $X_t^{\vee}$. Note that this combination of instruments is more restricted than the original Freedman instruments; making double use of the estimates $\hat{Y}$ in the 2SLS calculations, it is closer to the Klein calculations.

Also, since the predetermined variables were not at all expected to be linked to the errors in the model, we decided to resample the errors only, instead of resampling the vectors $(X_t, \tilde{e}_t)$ as described earlier.

Results for bootstrapping done with the above combination of instruments , and with resampling of the errors only, are given in TABLE 3.3. The same comments apply as for TABLE 3.2, except that in this case, the relative bias shows up for only three (3) coefficients, with here again, the associated larger variability. (except, MSE_{1*} for b_1)

Having thus bootstrapped the Klein equations by two methods and obtained 900 estimates of the coefficient vectors, we then

Table 3.3

Bootstrapping of Klein Model I Using Klein Instruments and Resampling the errors Only

(B = 900)

	Coeff. est.		Variability measures			
Coeff	Original Klein	Bootstrap average $(B^{-1}\sum_{b=1}^{B}\hat{\alpha}^{*}_{nb})$	Original var b_i	MSE_{1^*} (III.15)	MS of Nom. SE	MSE_{2^*} (III.17)
b_0	16.555	16.550	1.745	.0057	.0077	.0057
b_1	.017	.717	.0137	.0246	.0346	.5140
b_2	.216	.217	.0114	.0005	.0006	.0005
b_3	.810	.329	.0016	.0333	.0521	.2649
b_0'	20.278	20.275	56.596	.0220	.0207	.0220
b_1'	.150	.727	.0300	.0269	.0330	.3594
b_2'	.616	.615	.0260	.0067	.0067	.0067
b_3'	-.158	-.158	.0013	.0004	.0005	.0004
b_0''	1.500	1.477	1.316	.014	.016	.014
b_1''	.439	.440	.0013	.0063	.0091	.0063
b_2''	.147	.146	.0015	.0001	.0015	.0001
b_3''	.130	.130	.0008	.0004	.0004	.0004

plotted histograms of the data and tested its normality . Histograms of $\left[\hat{\alpha}^{*}_{nb} - \hat{\alpha}^{*}_{.}\right]$ were drawn up for the results of both methods of bootstrapping described. The histograms can all be seen to be roughly symmetric and bell-shaped. (see Fig. 3.1) ; see appendix B for the full set of histograms).

A univariate normal analysis was performed on the components of $\left[\hat{\alpha}^{*}_{nb} - \hat{\alpha}^{*}_{.}\right]$ for equation (I). First, the ranked values of the latter were plotted against

$$z_b = \Psi\,[(r_b - .375)/(B + .25)] \qquad (B = 900)$$

where r_b is the rank of the bth value, and Ψ is the inverse cumulative normal function; following the test for normality proposed by Filliben (1975), the Pearson correlation coefficients were then calculated between the v_b and the z_b , using

$$r = \sum_{b=1}^{B} (v_b - v_.)(z_b) \Big/ \left[\sum_{b=1}^{B} (v_b - v_.)^2 \sum_{b=1}^{B} (z_b)^2 \right]^{0.5}$$

where

$$v_. = \sum_{b=1}^{B} v_b$$

An example of the obtained plots, which were all of approximately the same shape, shows them to be of a very elongated 'S'-form, suggesting a continuous smooth and symmetric distribution not unlike the normal distribution, but with unusually long tails (see FIG 3.2). This applies to both Freedman-method results and to the results from the other bootstrapping method described . The correlation coefficient (r) values, on the other hand, range from 0.97735 to 0.99016, values all leading to

Fig 3.1

Bar Chart of bootstrapped values $\left[\hat{\alpha}^{*}_{nb_{13}} - \hat{\alpha}^{*}_{\cdot 13} \right]$

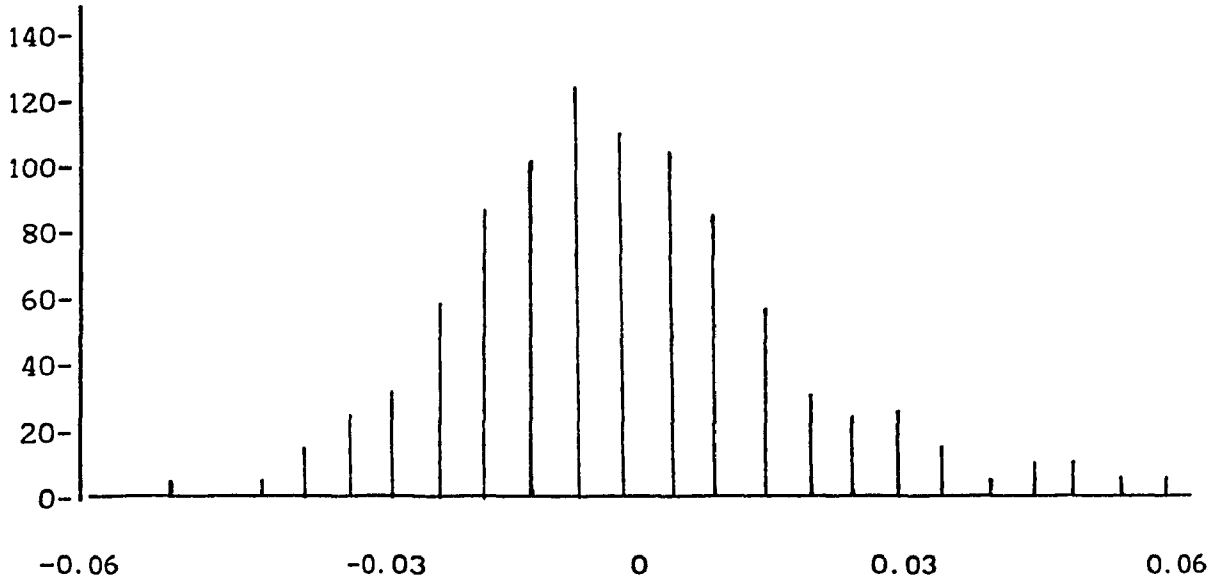
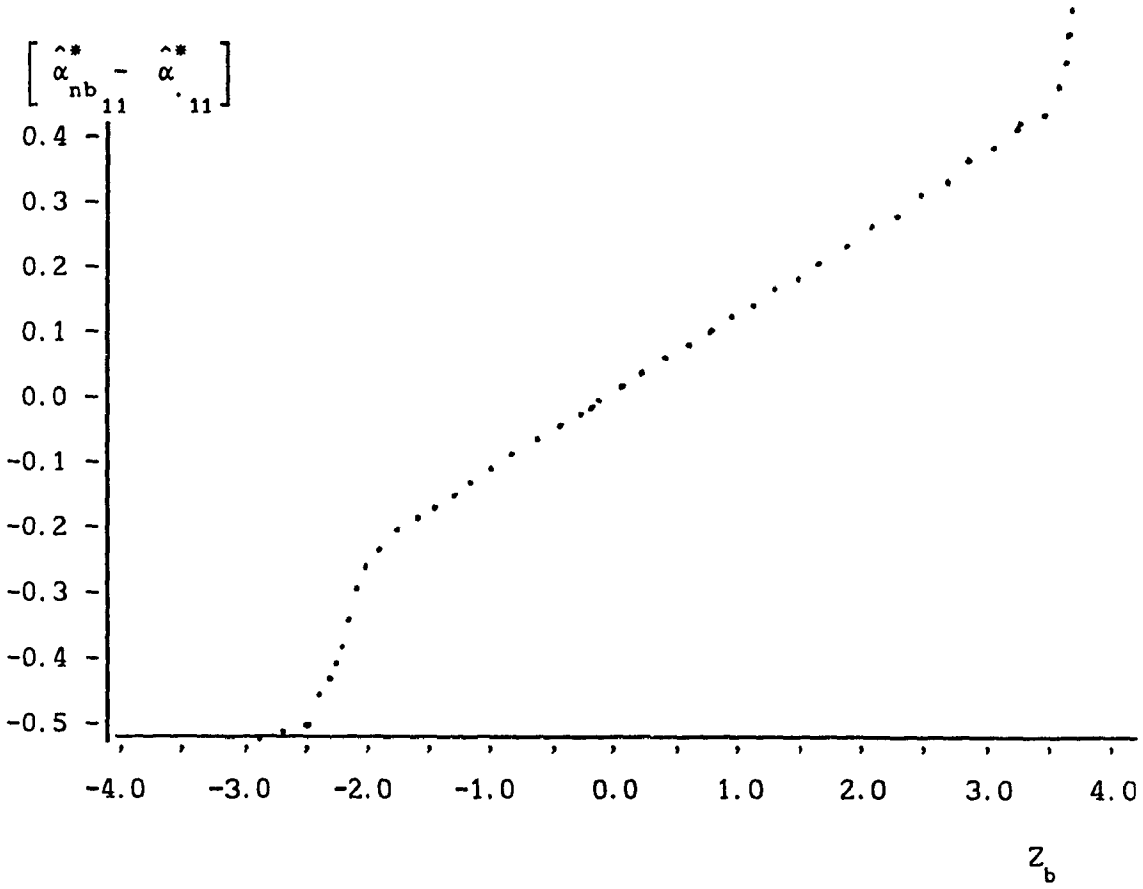


Fig 3.2

Q-Q plot of $\left[\hat{\alpha}^{*}_{nb_{11}} - \hat{\alpha}^{*}_{\cdot_{11}} \right]$ versus normal percentiles



rejection of the hypothesis of non-normality of the data $\left[H_0: \rho=0\right]$, but at a very low value of significance ($P < .005$), near the limit of the region of rejection of the null hypothesis; we note that intrapolation was necessary to arrive at the above conclusions, and that the ranges in Filliben's Table of r values did not allow precise extrapolation for the bootstrap sample size used in this thesis. One may therefore assign the normal distribution to the univariate marginal distributions of the individual components of $\left[\hat{\alpha}^*_{nb} - \hat{\alpha}^*_{\cdot}\right]$; but we must add that this assignation is of a borderline nature and in need of confirmation, under the test and information given to us by Filliben. One may even cast doubt on the applicability of Filliben's test for such large sample sizes as used in this thesis.

In view of the nature of these results, an additionnal test for normality was performed on the data, an adjusted Kolmogorov D test (available on the UNIVARIATE package on SAS), which compares the cumulative frequencies of the sample distribution of the data to the frequencies of a normal distribution with mean and variance equal to the sample's mean and variance (Stephens (1974)). The test computes

$$D = \max \left[D^+, D^- \right]$$

where

$$D^+ = \max_{1 \le i \le n} \left[b/B - z_b \right] \qquad D^- = \max_{1 \le i \le n} \left[z_b - (b-1)/B \right]$$

where

$$z_b = \Psi (w_b) \qquad w_b = \left[\hat{\alpha}^*_{nb} - \hat{\alpha}^*_{n\,.}\right] \left[SD_{\hat{\alpha}_{nb}} \right]^{-1}$$

where

$$SD_{\hat{\alpha}_{nb}} = \left[\text{components of diagonal of } \hat{COV}_{1^*}(\hat{\alpha}^*_{nb})\right]^{0.5}$$

(If D excedes the tabled value, one rejects the hypothesis of normality; Stephens (1974), Table 1.3)

The above test has been shown by Stephens (1974), in certain Monte Carlo test circumstances, to have almost the same discriminatory power as the Shapiro-Wilks W statistic. On the other hand, Filliben has shown the test based on the W statistic to be comparable in power to his r test (Filliben, 1975); hence, the r test may be taken to be comparable in power to the D test.

The results with the D statistic categorically reject the hypotheses of normality for all four (4) components of the coefficient vector. This 100% rejection would seem to indicate that a good portion of the coefficient components may not, finally, be normally distributed. The D statistic was also calculated for the coefficient vectors of equations II and III; here, it rejected the normal hypothesis for 6 out of 8 coefficient vector components.

The results of the r test, as well as the latter results, force us to conclude that many (if not all) univariate marginals of the joint distribution of the coefficient vectors are not in the normal family.

Continuing our normal analysis at a higher dimensional level, we then proceeded to evaluate the pairwise normality of the components of the

bootstrap coefficients. We first calculated the quantities

$$d_b^2 = \underset{(1Xp)}{(\Delta\hat{\alpha}^*)}\ \underset{(pXp)}{S^{-1}}\ \underset{(pX1)}{(\Delta\hat{\alpha}^*)'} \qquad \text{(III.18)}$$

where $(\Delta\hat{\alpha}^*)$ designates a size p subvector (p = 2,4) given by

$$(\Delta\hat{\alpha}^*) = (\hat{\alpha}^*_{nb} - \hat{\alpha}^*_{\cdot})$$

and where

$$S^{-1} = \hat{COV}^{-1}_{i*}\ (\hat{\alpha}^*_{nb})$$

which corresponds to the matrix of sample variances and covariances of the components of the subvector $(\Delta\hat{\alpha}^*)$.

The quantities d_b^2 were then ranked, and plotted, for p = 2, against

$$\chi^2_2 = -2 \ln \left[1 - (r_b - 0.5)/B \right] \qquad (B = 900)$$

where r_b is the rank of the d_b^2 values; this is the (100 $(r_b$ - 0.5)/B) percentile of the chi-square distribution with 2 degrees of freedom, obtained from

$$(r_b - 0.5)/B = \int_0^{\chi^2_2} 0.5\ e^{-0.5\ x} dx$$

The pairwise normality was than evaluated by a visual assessment of the linearity of the plot of the d_b^2 versus the χ^2_2.

Also, for p = 4, the results of the SAS function

$$\chi^2_4 = \text{gaminv}(r_b\ ,\ 4)$$

where the r_b are the ranks of the d_b^2 with p = 4, and where "gaminv" designates the inverse gamma, were plotted against the values d_b^2 themselves, as a test for the N_4 joint multivariate normality of the coefficient vectors. Again, closeness to normality (to $N_4(\mu,\Sigma)$, here) was verified by the linearity of the plots.

None of the plots thus obtained, either for p = 2 or for p = 4, were linear in shape; they tended, in fact to be smoothly curvilinear in form (see Fig. 3.3 for a typical example; see also appendix C) displaying a much more rapid rate of increase for the d_b^2 than for the χ_p^2. Note that one might expect this form, given the S-shape of the univariate plots seen earlier: one may rewrite the individual d_b^2 values as

$$d_b^2 = \sum_{i=1}^{p} (\Delta\hat{\alpha}_i^*)^2 \text{var}\ \hat{\alpha}_i^* + \sum_{i<j}^{p} (\Delta\hat{\alpha}_i^*)(\Delta\hat{\alpha}_j^*)(\text{cov}\ \hat{\alpha}_i^*\ \hat{\alpha}_j^*) \quad \text{(III.19)}$$

from (III.18).

And so the slight S-shaped character of the univariate Q-Q plots, which therefore show a tendency for long tails and greater variability, is made more clearly evident in the χ_p^2 plots. Note that $(\text{cov}\ \hat{\alpha}_i^*\hat{\alpha}_j^*)$ in (III.19) may oftentimes be negative, but if so, one of the $(\Delta\hat{\alpha}_i^*)$ will quite likely also be negative, which means that the second component of the above sum will almost always be positive.

To conclude on the remarks made above, the discussion given earlier, which tended to cast great doubt on the normality of the individual components of the coefficient vectors, appears to have received support from the above p-variate analysis of the subvectors. Certainly, we may categorically state that the joint distribution of the coefficient vectors is not $N_4(\mu, \Sigma)$. The r-test described by Filliben (1974) could certainly be adapted to the plotting of the ranked data against the percentiles of the ranks of other distributions (possibly, symmetric alternatives with tails longer than the normal's), as Filliben

Table 3.3

95 percent Bonferroni Confidence Intervals

for Bootstrapped* coefficient estimates

$[\alpha^*_{\blacksquare\cdot} \pm z_{.00312S}(\text{diag MSE}_{1\bullet})^{0.5}]$

EQUATION I

16.555 ± 0.021

0.717 ± 0.42

0.217 ± 0.06

0.329 ± 0.51

EQUATION II

20.275 ± 0.42

0.727 ± 0.45

0.615 ± 0.21

-0.158 ± 0.06

EQUATION III

1.500 ± 0.33

0.440 ± 0.21

0.146 ± 0.027

0.130 ± 0.06

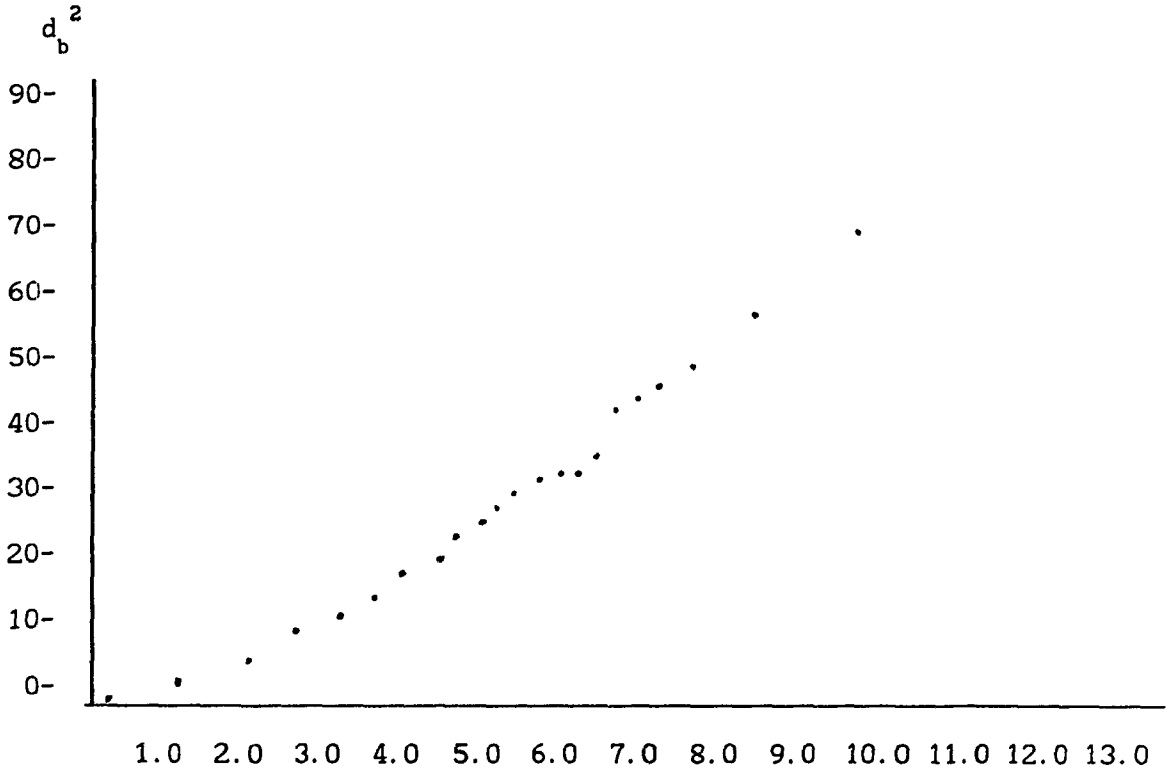
himself suggests. We have not done so at this writing.

Efron (1982, 1984, 1985), and Efron and Tibshirani (1986), have described procedures to calculate confidence intervals for bootstrapped estimators. Assuming the data to be normally distributed, they generated the pseudo-data from $\hat{F}$: $N(\hat{\mu}, \hat{\Sigma})$, where $\hat{\mu}$ is the vector of samplemeans, and $\hat{\Sigma}$ the sample variance-covariance matrix.

But in our case, having adopted $\hat{F}$: 1/n for the resampling in this study, we could not follow the Efron procedure. We therefore attempted to derive confidence intervals using other means. Bonferroni confidence intervals were calculated; results are in Table 3.3. The previous results on the distributions of the coefficient vectors lead us to conclude that these intervals are probably too narrow, since the bootstrapped data appears as a whole to be distributed with tails that are longer than for the normal distribution .

Fig 3.3

Plot of d_b^2 against χ_4^2 percentiles



Appendix A

Probability limits and stationary stochastic processes

This thesis makes frequent use of the concept of the probability limit, specifically in applications to sample statistics from stationary stochastic processes. The following is from from Goldberger (1964).

(A) To begin with, some definitions and basic results for probability limits:

i) Let $\{X_n\} = X_1, X_2, \ldots, X_n$ be a sequence of random variables. Suppose that

$$\lim_{n \to \infty} \text{Prob}\left[|X_n - \mu| \geq \delta\right] = 0$$

for every $\delta > 0$, where μ is a finite constant. Then μ is said to be the probability limit of the sequence $\{X_n\}$, and we write

$$\text{plim } X = \mu$$

This is also called convergence in probability; as $n \to \infty$, the mass or density function becomes entirely concentrated on the point μ. Finally, we also say that X is consistent for μ.

ii) If $\lim_{n \to \infty} E(X_n) = \mu$ and

$$\lim_{n \to \infty} \left[n^{-1} \lim_{n \to \infty} E\left[n^{0.5}(X_n - E(X_n)) \right]^2 \right] = 0,$$

that is, if the expectation approaches a constant as $n \to \infty$, and if the variance also goes to nil (the distribution is said hence to collapse

on a single point), then

$$\text{plim } X = \mu$$

X is said thence to be consistent for μ .

iii) An important property of probability limits is that the probability limit of a continuous function is the function of the probability limit; thus if $\text{plim } X = \mu$ and if $g(X)$ is continuous, then

$$\text{plim } g(X) = g(\mu)$$

This is called Slutzky's theorem.

Next, we state two important applications of Slutzky's theorem to matrices. Since the elements of a product matrix are continuous functions of the elements of the component matrices,

$$\text{plim } (A\,B) = (\text{plim } A)\,(\text{plim } B)$$

Also, since the elements of an inverse matrix are continuous functions of the elements of the original matrix,

$$\text{plim } (A^{-1}) = (\text{plim } A)^{-1}$$

(iv) Stochastic processes:

A Stochastic process is defined as a family of random variables $\{X_t\}$ where $t = \ldots,-2,-1,0,1,2,\ldots$ denotes time, such that for every finite set of choices of t $\{t_1, \ldots, t_n\}$, a joint probability distribution is defined for the random variables $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$.

A Stationary Stochastic Process is one in which the joint probability distributions are invariant under translations along the time

axis, so that the joint probability distribution of any finite set $\left\{ X_{t_1}, X_{t_2}, \ldots, X_{t_n} \right\}$ is the same as the joint pdf of the set $\left\{ X_{t_{1+\tau}}, X_{t_{2+\tau}}, \ldots, X_{t_{n+\tau}} \right\}$ for $\tau = \ldots, -2, -1, 0, 1, 2, \ldots$. Hence the following moments, as well as all others, can be defined for the joint pdf:

$$E\, X_t = \mu \qquad E\left[X_t - \mu\right]^2 = \sigma^2 \qquad E\left[X_t - \mu\right]\left[X_{t+\tau} - \mu\right] = \gamma_\tau$$

(B) We next give some basic applications to sample statistics from stationary stochastic processes:

(i) It can be shown that, under the condition that the covariance between X_t and $X_{t+\tau}$ in a Stationary Stochastic Process declines so rapidly that $\lim_{\tau \to \infty} \gamma^\tau = 0$, then

a) $\lim_{n \to \infty} E(\bar{X}) = \mu$, $\bar{X}$ sample mean; and

b) $$\lim_{n \to \infty} \left[n^{-1} \lim_{n \to \infty} E\left[n^{0.5}(\bar{X} - E(\bar{X})) \right]^2 \right] = 0$$

This can be expressed in shorthand form as

$$\text{plim}\ \bar{X} = \mu$$

Also, it can be shown, under the condition that the dependence between distant values of X wears off so rapidly that $\lim_{\tau \to \infty} \gamma^\tau = 0$ and that other limit moment conditions apply, and even though the sampling may be nonrandom, that

(ii) plim $s^2 = \sigma^2$, where $s^2 = n^{-1} \sum_{t=1}^{n} \left[X_t - \bar{X}\right]^2$;

and that

(iii) plim $c_\tau = \gamma_\tau$, where $c_\tau = (n-1)^{-1} \sum_{t=1}^{n} \left[X_t - \bar{X}\right]\left[X_{t+\tau} - \bar{X}\right]$

(C) Multivariate stochastic processes:

A multivariate stochastic process is a family of random (K X 1) vectors $\{X_t\}$ where

$$X_t = \begin{bmatrix} X_{t1} \\ \vdots \\ X_{tK} \end{bmatrix}$$

such that for every finite set of choices of t ($t = t_1, t_2, \ldots, t_n$), a joint probability distribution is defined for the random vectors X_{t_1}, $X_{t_2}, \ldots, X_{t_n}$; hence each distribution for a given finite set of choices of t is a distribution covering an (n X K) array of variables. If these joint distributions are invariant under translations along the time axis, we have a stationary multivariate stochastic process, for which we may define the following moments, assuming they exist:

$$E\, X_t = \mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_K \end{bmatrix}$$

$$E\left[X_t - \mu\right]\left[X_t - \mu\right]' = \Sigma = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1K} \\ & \vdots & \\ \sigma_{K1} & \cdots & \sigma_{KK} \end{bmatrix}$$

$$E\left[X_t - \mu\right]\left[X_{t+\tau} - \mu\right]' = \Gamma_\tau = \begin{bmatrix} \gamma_{\tau 11} & \cdots & \gamma_{\tau 1K} \\ \vdots & & \\ \gamma_{\tau K1} & \cdots & \gamma_{\tau KK} \end{bmatrix}$$

where $\sigma_{ij} = E\left[X_{ti} - \mu_i\right]\left[X_{tj} - \mu_j\right]$ is the contemporaneous covariance between the random variables X_i and X_j $(i, j = 1, \ldots, K)$. And where $\gamma_{\tau ij} = E\left[X_{ti} - \mu_i\right]\left[X_{t+\tau, j} - \mu_j\right]$ is the lag-τ crossvariance between X_i and X_j.

(D) Consistent estimators of moments of stationary multivariate stochastic processes:

By a sample of size n of the stationary multivariate stochastic process $\{X_t\}$ we mean a joint drawing $\left\{X_1, \ldots, X_n\right\}$ from the joint distribution of n consecutive X_t's. Given such a sample we may compute the sample means $\bar{X}_i = n^{-1}\sum_{t=1}^{n} X_{ti}$ and the sample variances and contemporaneous covariances $s_{ij} = n^{-1}\sum_{t=1}^{n}\left[X_{ti} - \bar{X}_i\right]\left[X_{tj} - \bar{X}_j\right]$ and consider using these as estimates of the corresponding population parameters. Expressing the above in matrix form, we have for the sample observations

$$X = \begin{bmatrix} X_{11} & \cdots & X_{1K} \\ & \vdots & \\ X_{n1} & \cdots & X_{nK} \end{bmatrix}$$

where each row gives one observation on the random vector X_t. The sample mean vector is

$$\bar{X} = \begin{bmatrix} \bar{X}_1 \\ \vdots \\ \bar{X}_K \end{bmatrix} = n^{-1}X'\iota, \qquad \underset{(nX1)}{\iota} = \left[1, 1, \ldots, 1\right]'$$

and the sample contemporaneous covariance matrix is

$$S = \begin{bmatrix} s_{11} & \cdots & s_{1K} \\ & \vdots & \\ s_{K1} & \cdots & s_{KK} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{t=1}^{n} X_{t1}^2 - n\bar{X}_1^2 & \cdots & \sum_{t=1}^{n} X_{t1}X_{tK} - n\bar{X}_1\bar{X}_K \\ & \vdots & \\ \sum_{t=1}^{n} X_{tK}X_{t1} - n\bar{X}_K\bar{X}_1 & \cdots & \sum_{t=1}^{n} X_{tK}^2 - n\bar{X}_K^2 \end{bmatrix}$$

$$= n^{-1}X'X - \bar{X}\bar{X}' = n^{-1}X'MX$$

where $M = I_n - n^{-1}\iota\iota'$. Under the conditions stipulated in the univariate case, i.e. in the presence of sufficiently rapidly decreasing dependence between X_t and $X_{t+\tau}$ as τ increases, even under conditions of nonrandom sampling, we have

$$\text{plim } \bar{X} = \mu$$

and

$$\text{plim } S = \Sigma.$$

We finally note that under the conditions stated above on the decreasing dependence between X_t and $X_{t+\tau}$, the sample second moment $n^{-1}X'X$ also has a probability limit:

$$\text{plim } n^{-1}X'X = \text{plim } (S + \bar{X}\bar{X}')$$

$$= \text{plim } S + (\text{plim } \bar{X})(\text{plim } \bar{X})' = \Sigma + \mu\mu'$$

$$= E(X_tX_t' - \mu\mu') + \mu\mu' = EX_tX_t' = \Sigma_{XX}$$

BIBLIOGRAPHY

1. Bickel, P. and Freedman, D. (1981) "Some Asymptotic Theory for the Bootstrap." Ann. Statist. 9, 1196-1217.

2. Donatos, G. (1984) "Monte Carlo Study of econometric estimators for small samples." Thesis (M.Sc.) Concordia University

3. Droge, B. (1987) "A Note on Estimating the MSEP in Nonlinear Regression." Statistics 18 (1987), 4, 499-520.

4. Dwivedi, T.D. "An Identity between double k-class, k-class, Two-Stage Least Squares and Ordinary Least Squares." Proc. of ISI, 1983.

5. Efron, B. (1979a) "Bootstrap Methods: Another Look at the Jackknife." Ann. Statist. 7, 1-26.

6. Efron, B. (1982a) "The Jackknife, the Bootstrap and Other Resampling Plans." Philadelphia, S.I.A.M., J.W. Arrowsmith Ltd

7. Efron, B. (1981) "Nonparametric Estimation of Standard Errors: the Jackknife, the Bootstrap and Other Methods." Biometrika 68, 3, pp 589-99.

8. Efron, B. and Gong, G. (1983) "A Leisurely Look at the Bootstrap, the Jackknife, and Crossvalidation." The American Statistician, Feb. 1983,

Vol. 37, No. 1.

9. Efron, B. and Stein, C. (1981) "The Jackknife Estimate of Variance." Ann. Statist., Vol. 9, No.3, pp. 586-96.

10. Efron, B. and Tibshirani, R. (1979a) "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy." Statistical Science, Vol.1, no. 1, 54-77.

11. Filliben, J.J., (1975) "The Probability Plot Correlation Coefficient Test for Normality." Technometrics, Vol. 17, No. 1, Feb. 1975.

12. Freedman, D., (1981) "Bootstrapping Regression Models." Ann. Statist., 9, 1218-28.

13. Freedman, D., (1984) "On Bootstrapping Two-Stage Least-Squares Estimates in Stationary Linear Models." Ann. Statist., Vol. 12, No. 3, 827-42.

14. Freedman, D., and Peters, S. (1984a) "Bootstrapping a Regression Equation: Some Empirical Results." J. Amer. Statist. Assoc. 79, 97-106.1 Goldberger, A.S. (1964) Econometrics Theory, Wiley and sons, N.Y.

15 Gu, C. (1987) "What Happens when Bootstrapping the Smoothing Spline." Commun. Statist.- Theory Meth.- 16(11) 3275-84.

16 Johnston, J. (1984) Econometric Methods. Third Ed. McGraw-Hill.

17 Miller, R.G. (1974a) "The Jackknife - a Review." Biometrika, 61, 1-15.

18. Quenneville, B. (1986) "Bootstrap Procedure for Testing Linear Hypotheses Without Normality." Statistics 17 (1986) 4, 533-38.

19. Quenouille M.H. (1949) "Problems in plane sampling" Ann. Math. Stat., 20, 355-75.

20. Singh, K. (1981) "On the Asymptotic Accuracy of Efron's Bootstrap." Ann. Statist. 9, 1187-95.

21. Stephens, M.A. (1974) "EDF Statistics for Goodness of Fit and Some Comparisons." JASA, Sept. 1974, Vol. 69, No. 347, Theory and Methods Section, 730-37.

22. Theil, H. (1971) Principles of Econometrics. John Wiley and Sons.

23. Tukey, J.W., (1958) "Bias and Confidence in not-quite large samples." Ann. Math. Stat., 29, 614.