TSCSREG:
A SAS PROCEDURE FOR THE ANALYSIS OF TIME SERIES CROSS-SECTION DATA

by

Douglas J. Drummond
A. Ronald Gallant

Mimeograph Series No. 1107
Raleigh, N. C. February 1977

Table of Contents

THE TSCSREG PROCEDURE

TSCSREG, an acronym for Time Series Cross-Section REGression, is designed to analyze a class of linear econometric models which commonly arise when combining time series and cross-sectional data. Such models may be viewed as a two-way design with covariates:

$$y_{ij} = \sum_{k=1}^{p} x_{ijk}\beta_k + u_{ij} \qquad \begin{array}{l} i = 1, 2, \ldots, CS \\ j = 1, 2, \ldots, TS \end{array}$$

As is always the case, the performance of any estimation procedure for the model regression parameters β_k, $k = 1, 2, \ldots, p$ will depend upon the statistical characteristics of the error components, u_{ij}, in the model. For this reason, the package TSCSREG was written to allow a user to study the estimates of the regression parameters in the above model under three of the more common error structures in the recent literature. Namely,

1. A variance components model.
2. A first order autoregressive model with contemporaneous correlation.

and

3. A mixed variance-component moving average error process.

In doing this, the TSCSREG procedure makes no assumptions concerning the rank of the matrix of explanatory variables X. Moreover, only range-preserving estimators are used for all required error structure parameters.

OUTPUT

Use of PROC TSCSREG causes all information and references for the procedure options to be printed out. These enable one to locate the original

research report and thus obtain a complete description of the three types of error structures mentioned in the preceding section. In addition, for each model entered, the TSCSREG procedure gives the model variables, underlying error structure parameter estimates, and regression parameter estimates and analysis. Included under the umbrella of "analysis" is the name of the SAS variable with which it is associated, a t-statistic for testing whether the corresponding beta is zero, the significance probability of the t-statistic, and the standard error of the b-value including the degrees of freedom upon which it is based. Further, wherever possible, we have adhered to the notation of the original reference.

The only output option available with PROC TSCSREG is to have the variance-covariance and/or the correlation matrix of the resulting regression parameter estimates for each model and assumed error structure printed out.

The Procedure TSCSREG Statement

The PROCEDURE TSCSREG statement is of the form

PROC TSCSREG TS = __ CS = __ < DATA = data-set name>
<FULLER> <PARKS> <DASILVA M = __>;

Parameters for PROC TSCSREG Statement

Data - Input data set. See paragraph on 'STRUCTURE OF INPUT DATA SET' in later part of documentation.

TS - The number of observations in each time series.

CS - The number of observations in each cross-section.

FULLER - Analysis of model assuming a Type I error structure.

PARKS - Analysis of model assuming a Type II error structure.

DASILVA - Analysis of model assuming a Type III error structure

M - The length of the assumed moving average process in method DASILVA ($M \leq TS-2$)

PROCEDURE INFORMATION STATEMENTS

PARMCARDS Statement

There is only one type of Parameter card, namely, MODEL.

PARAMETER CARD SPECIFICATION

The MODEL statement is of the following form:

MODEL Numeric_variable = numeric_variable <additional numeric variables>

/<NOINT / NOMEAN> <VAR> <CORR> #

Options and Parameters:

NOMEAN - When this option is absent, the procedure will include an intercept
NOINT term in the model. When NOMEAN or NOINT is specified, no intercept will be included.

VAR - This option causes the variance-covariance matrix of the model parameter estimates to be printed.

CORR - This option causes the correlation matrix of the model parameter estimates to be printed.

TREATMENT OF MISSING VALUES

If the value of any one of the SAS variables used for the model presently being analyzed is missing from an observation, analysis on this and all subsequent models will terminate with an accompanying error message.

STRUCTURE OF INPUT DATA SET

PROC TSCSREG assumes that the input data set consists of N = CS * TS observations which have previously been sorted according to

```
PROC SORT <OUT = data_set_name><DATA = data_set_name>
BY  CS_ID  TS_ID;
```

No check is made on the above structure except that of whether the actual number of observations on the input data set agrees with the product of CS and TS. If it does not, analysis on this model and all subsequent ones will terminate with an accompanying error message.

STANDARD FIX-UPS TAKEN

For procedure PARKS , the first-order autocorrelation coefficient must be estimated for each cross-section. Let RHO be the CS-VECTOR of true parameters and R the corresponding vector of estimates. Then to ensure that only range-preserving estimates are used in PROC TSCSREG the following form for R is taken:

$$
\begin{aligned}
R(I) &= R(I) && \text{if } |R(I)| < 1 \\
&= \max(.95, RMAX) && \text{if } R(I) \geq 1 \\
&= \min(-.95, RMIN) && \text{if } R(I) \leq -1
\end{aligned}
$$

where

$$
\begin{aligned}
RMAX &= 0 \text{ , if } R(I) < 0 \text{ or } R(I) \geq 1 \text{ for all } I \\
&= \max_J \{R(J): \ 0 \leq R(J) < 1\} \text{ , otherwise.}
\end{aligned}
$$

and similarly

$$RMIN = 0 \text{ , if } R(I) > 0 \text{ or } R(I) \leq -1 \text{ for all } I$$

$$= \min_J R(J): \; -1 < R(J) \leq 0 \quad \text{otherwise.}$$

whenever this fix-up is taken, an error message is printed.

JOB CONTROL STATEMENT REQUIREMENTS AT TUCC

PROC TSCSREG is written to run under SAS(1976) and is available in the SAS Supplementary Library. To run a job at the Triangle University Computing Center requires the following job control language cards

```
//job_name  JOB   xxx.yyy.zzz,programmer_name
//STEP      EXEC SAS
//SAS.FT16F001 DD DSN=&&UT2,UNIT=SYSDA,SPACE=(TRK,(10,10)),
//    DCB-(RECFM=VBS,LRECL=2602,BLKSIZE=2606)
//SAS.SYSIN DD *
          :
          SAS STATEMENTS AND DATA CARDS
          :
/*
```

EXAMPLE

For completeness we begin by creating a data set WORK.DS1 from cards. PROC SORT is then used to rearrange the data into the required time series and cross-section format. Finally, PROC TSCSREG is applied to analyze the data under all three error structure models. All relevant output is included in this documentation including a listing of the sorted data set via PROC PRINT.

```
DATA WORK.DS1; INPUT Y 4-8  X_1 21-30 X_2 31-40 X_3 41-50 STATE $ 61-
     YEAR 71-74 ; CARDS;

PROC SORT OUT=WORK.DS1 ; BY STATE YEAR ;

PROC TSCSREG DATA=WORK.DS1 TS=15 CS=8 FULLER PARKS DASILVA M=5;
PARMCARDS;
MODEL Y = X_1 X_2 X_3 / VAR CORR#
;

PROC PRINT  DATA=WORK.DS1  ;
```

```
- - - - - - - - - - -  OPTIONS SPECIFIED - - - - - - - - - - -


OPTION FULLER SPECIFIED

REFERENCE:
W. A. FULLER & G. E. BATTESE, ESTIMATION OF LINEAR MODELS
WITH CROSSES-ERROR STRUCTURE, JOURNAL OF ECONOMETRICS, V. 2,
NO. 1, MAY 1974, P. 67-78


OPTION PARKS SPECIFIED

REFERENCE:
J. KMENTA, ELEMENTS OF ECONOMETRICS, NEW YORK: MACMILLAN CO.
1971, P. 512-514.


OPTION DA SILVA SPECIFIED WITH M =      5

REFERENCE:
J. G. C. DA SILVA, THE ANALYSIS OF CROSS-SECTIONAL TIME
SERIES DATA, INSTITUTE OF STATISTICS MIMEOGRAPH SERIES, NO.
1011, RALEIGH, N. C.: NORTH CAROLINA STATE UNIVERSITY, 1975,
(PH.D. DISSERTATION).


NUMBER OF OBSERVATIONS IN A CROSS-SECTION =          8

NUMBER OF OBSERVATIONS IN A TIME SERIES =           15


NOTE:
IF THE DATA SET WAS NOT SORTED ACCORDING TO THE INSTRUCTIONS
IN THE DOCUMENTATION OF PROC TSCSREG THE RESULTS WILL BE
ERRONEOUS.  E.G. STATEMENTS SIMILAR TO :
     PROC SORT DATA=DATA_SET_NAME OUT=DATA_SET_NAME;
     BY CROSS_ID TIME_ID;
SHOULD PRECEDE PROC TSCSREG.
```

****

MODEL

****

DEPENDENT VARIABLE

Y

INDEPENDENT VARIABLES

$INT

X_1

X_2

X_3

****
****

FULLER AND BATTESE METHOD ESTIMATES

SOURCE	B VALUES	T FOR H:B=0	PROB>|T|	STD ERR B
$INT	1.65092	3.4741	0.0007	0.47520
X_1	0.287052	C.85364	0.3951	0.33627
X_2	-0.898527	-1.1456	0.2543	0.78434
X_3	0.580770	2.2946	0.0236	0.25310

DEGREES OF FREEDOM FOR T-STATISTICS = 116

VARIANCE COMPONENT ESTIMATES

VARIANCE COMPONENT FOR CROSS-SECTIONS = 0.34670784
(SIGMA_SQUARED_SUB_V)

VARIANCE COMPONENT FOR TIME SERIES = 0.076498770
(SIGMA_SQUARED_SUB_E)

VARIANCE COMPONENT FOR ERROR = 0.53583838
(SIGMA_SQUARED_SUB_EPSILON)

TRANSFORMED REGRESSION M.S.E. = 0.53143436
(M.S.E. DEGREES OF FREEDOM = 116)

****
*
*
* VARIANCE-COVARIANCE MATRIX FOR
* FULLER AND BATTESE PARAMETER ESTIMATES
*
*
****

	$INT	X_1	X_2	X_3
$INT	0.22582	-0.052143	-0.28883	-0.028777
X_1	-0.052143	0.11308	-0.0020331	-0.010987
X_2	-0.28883	-0.0020331	0.61518	0.011855
X_3	-0.028777	-0.010987	0.011855	0.064062

****
*
*
* CORRELATION MATRIX FOR
* FULLER AND BATTESE PARAMETER ESTIMATES
*
*
****

	$INT	X_1	X_2	X_3
$INT	1.0000	-0.32631	-0.77493	-0.23926
X_1	-0.32631	1.0000	-0.0077086	-0.12909
X_2	-0.77493	-0.0077086	1.0000	0.059716
X_3	-0.23926	-0.12909	0.059716	1.0000

NOTE:
THE FULLER AND BATTESE METHOD USED 0.35 SECONDS
AND REQUIRED AN INCREMENT OF 1640 BYTES OF STORAGE.

PROGRAMMER, FULLER AND BATTESE METHOD:

DOUGLAS J. DRUMMOND
INSTITUTE OF STATISTICS
NORTH CAROLINA STATE UNIVERSITY
RALEIGH, NORTH CARCLINA 27607

****
*
*
* PARKS METHOD ESTIMATES
*
*
****

SOURCE	B VALUES	T FOR H:B=0	PROB>|T|	STD ERR B
SINT	1.82506	7.6144	0.0000	0.23969
X_1	0.0617632	0.27765	0.7818	0.22245
X_2	-1.21082	-4.3419	0.0000	0.27887
X_3	0.676353	4.2902	0.0000	0.15765

DEGREES OF FREEDOM FOR T-STATISTICS = 116

ESTIMATED VALUES OF RHO(I)
(FIRST ORDER AUTOREGRESSIVE PARAMETER ESTIMATES)

I =	1	0.33044
I =	2	0.70512
I =	3	0.38124
I =	4	0.18514
I =	5	0.29976
I =	6	0.29833
I =	7	0.73599
I =	8	0.34145

ESTIMATED PHI MATRIX

	COL 1	COL 2	COL 3	COL 4	COL 5	COL 6	COL 7	COL
1	0.38928	0.22902	-0.089207	0.13964	-0.021634	0.21287	0.11522	-0.0907
2	0.22902	0.88173	0.0044749	-0.21682	-0.30819	0.25355	-0.44083	-0.324
3	-0.089207	0.0044749	0.30685	0.088768	0.032994	0.048281	-0.10146	-0.0286
4	0.13964	-0.21682	0.088768	0.64587	0.28847	0.24491	0.34085	0.169
5	-0.021634	-0.30819	0.032994	0.28847	0.76448	0.0029048	0.48981	0.409
6	0.21287	0.25355	0.048281	0.24491	0.0029048	0.43311	0.065165	-0.0811
7	0.11522	-0.44083	-0.10146	0.34085	0.48981	0.065165	0.77489	0.333
8	-0.090791	-0.32429	-0.028638	0.16911	0.40931	-0.081160	0.33335	0.634

TRANSFORMED REGRESSION M.S.E. = 1.0071936
(M.S.E. DEGREES OF FREEDOM = 116)

VARIANCE-COVARIANCE MATRIX FOR PARKS PARAMETER ESTIMATES

	$INT	X_1	X_2	X_3
$INT	0.057449	-0.024841	-0.047571	-0.012853
X_1	-0.024841	0.049485	0.00056828	-0.0012826
X_2	-0.047571	0.00056828	0.077769	0.0060672
X_3	-0.012853	-0.0012826	0.0060672	0.024854

CORRELATION MATRIX FOR PARKS PARAMETER ESTIMATES

	$INT	X_1	X_2	X_3
$INT	1.0000	-0.46590	-0.71170	-0.34016
X_1	-0.46590	1.0000	0.0091606	-0.036572
X_2	-0.71170	0.0091606	1.0000	0.13800
X_3	-0.34016	-0.036572	0.13800	1.0000

NOTE:
PARKS METHOD USED 0.35 SECONDS AND REQUIRED AN
INCREMENT OF 5752 BYTES OF CORE STORAGE.

PROGRAMMER, PARKS METHOD:

DOUGLAS J. DRUMMOND
INSTITUTE OF STATISTICS
NORTH CAROLINA STATE UNIVERSITY
RALEIGH, NORTH CAROLINA 27607

DA SILVA METHOD ESTIMATES

SOURCE	B VALUES	T FOR H:B=0	PROB>|T|	STD ERR B
$INT	2.48618	6.2664	0.0000	0.39675
X_1	-1.71005	-15.058	0.0	0.11357
X_2	-0.833112	-1.1493	0.2528	0.72489
X_3	0.934267	21.830	0.0	0.042798

DEGREES OF FREEDOM FOR T-STATISTICS = 116

VARIANCE COMPONENT ESTIMATES

VARIANCE COMPONENT FOR CROSS-SECTIONS = 0.25413869
(SIGMA_SQUARED_SUB_A)

VARIANCE COMPONENT FOR TIME SERIES = 0.017052898
(SIGMA_SQUARED_SUB_B)

AUTOCOVARIANCES

LAG	0	0.59280
LAG	1	0.18490
LAG	2	0.10431
LAG	3	-0.012287
LAG	4	-0.13475
LAG	5	0.093570

TRANSFORMED REGRESSION M.S.E. = 8.1190853
(M.S.E. DEGREES OF FREEDOM = 116)

VARIANCE-COVARIANCE MATRIX FOR DA SILVA PARAMETER ESTIMATES

	$INT	X_1	X_2	X_3
$INT	0.15741	-0.0059601	-0.24306	-0.00035290
X_1	-0.0059601	0.012897	-0.00022752	-0.0012295
X_2	-0.24306	-0.00022752	0.52546	0.00033894
X_3	-0.00035290	-0.0012295	0.00033894	0.0018316

CORRELATION MATRIX FOR DA SILVA PARAMETER ESTIMATES

	$INT	X_1	X_2	X_3
$INT	1.0000	-0.13228	-0.84515	-0.020784
X_1	-0.13228	1.0000	-0.0027637	-0.25296
X_2	-0.84515	-0.0027637	1.0000	0.010925
X_3	-0.020784	-0.25296	0.010925	1.0000

NOTE:
DA SILVA METHOD USED 1.83 SECONDS AND REQUIRED AN INCREMENT OF 10752 BYTES OF CORE STORAGE.

PROGRAMMER, DA SILVA METHOD:

DR. A. RONALD GALLANT
INSTITUTE OF STATISTICS
NORTH CAROLINA STATE UNIVERSITY
RALEIGH, NORTH CAROLINA 27607

OBS	Y	X_1	X_2	X_3	STATE	YEAR
1	3.259	0.2	0.4	0.7	NH	1956
2	1.401	0.4	0.4	0.4	NH	1957
3	0.975	0.3	0.4	0.9	NH	1958
4	0.913	0.2	0.4	0.5	NH	1959
5	0.486	0.2	0.4	0.2	NH	1960
6	1.094	0.8	0.4	0.4	NH	1961
7	2.261	0.1	0.4	0.6	NH	1962
8	2.128	0.6	0.4	0.3	NH	1963
9	1.916	0.9	0.4	0.4	NH	1964
10	3.421	0.9	0.4	0.3	NH	1965
11	1.046	0.5	0.4	0.3	NH	1966
12	2.225	0.9	0.4	0.3	NH	1967
13	2.715	0.5	0.4	0.6	NH	1968
14	2.079	0.4	0.4	0.4	NH	1969
15	2.671	0.8	0.4	0.9	NH	1970
16	1.862	0.2	0.8	0.0	NY	1956
17	3.023	0.4	0.8	0.7	NY	1957
18	1.877	0.3	0.8	0.8	NY	1958
19	1.567	0.2	0.8	0.0	NY	1959
20	0.312	0.2	0.8	0.0	NY	1960
21	2.225	0.8	0.8	0.8	NY	1961
22	2.422	0.1	0.8	0.6	NY	1962
23	2.483	0.6	0.8	0.4	NY	1963
24	1.229	0.9	0.8	0.5	NY	1964
25	1.497	0.9	0.8	0.0	NY	1965
26	0.072	0.5	0.8	0.9	NY	1966
27	1.262	0.9	0.8	0.1	NY	1967
28	0.996	0.5	0.8	0.4	NY	1968
29	1.320	0.4	0.8	0.0	NY	1969
30	2.994	0.8	0.8	0.4	NY	1970
31	1.117	0.2	0.3	0.7	CONN	1956
32	1.621	0.4	0.3	0.5	CONN	1957
33	1.471	0.3	0.3	0.2	CONN	1958
34	0.734	0.2	0.3	0.5	CONN	1959
35	1.028	0.2	0.3	0.0	CONN	1960
36	0.281	0.8	0.3	0.0	CONN	1961
37	1.394	0.1	0.3	0.1	CONN	1962
38	2.052	0.6	0.3	0.3	CONN	1963
39	2.717	0.9	0.3	0.7	CONN	1964
40	2.800	0.9	0.3	0.6	CONN	1965
41	2.119	0.5	0.3	0.8	CONN	1966
42	3.067	0.9	0.3	0.1	CONN	1967
43	1.944	0.5	0.3	0.6	CONN	1968
44	2.008	0.4	0.3	0.5	CONN	1969
45	2.180	0.8	0.3	0.5	CONN	1970
46	2.335	0.2	0.2	0.9	TEXAS	1956
47	2.196	0.4	0.2	0.6	TEXAS	1957
48	2.131	0.3	0.2	0.2	TEXAS	1958
49	2.981	0.2	0.2	0.6	TEXAS	1959
50	0.457	0.2	0.2	0.2	TEXAS	1960
51	1.948	0.8	0.2	0.7	TEXAS	1961
52	2.883	0.1	0.2	0.9	TEXAS	1962
53	2.930	0.6	0.2	0.5	TEXAS	1963
54	2.866	0.9	0.2	0.3	TEXAS	1964
55	3.449	0.9	0.2	0.5	TEXAS	1965
56	1.862	0.5	0.2	0.5	TEXAS	1966
57	3.891	0.9	0.2	0.8	TEXAS	1967
58	3.618	0.5	0.2	0.7	TEXAS	1968
59	3.959	0.4	0.2	0.4	TEXAS	1969
60	4.268	0.8	0.2	0.9	TEXAS	1970

OBS	Y	X_1	X_2	X_3	STATE	YEAR
61	3.473	0.2	0.3	0.1	UTAH	1956
62	2.921	0.4	0.3	0.7	UTAH	1957
63	2.248	0.3	0.3	0.2	UTAH	1958
64	2.091	0.2	0.3	0.0	UTAH	1959
65	0.707	0.2	0.3	0.0	UTAH	1960
66	2.503	0.8	0.3	0.9	UTAH	1961
67	2.802	0.1	0.3	0.7	UTAH	1962
68	2.657	0.6	0.3	0.1	UTAH	1963
69	1.895	0.9	0.3	0.5	UTAH	1961
70	1.033	0.9	0.3	0.0	UTAH	1965
71	0.562	0.5	0.3	0.9	UTAH	1966
72	2.058	0.9	0.3	0.6	UTAH	1967
73	1.775	0.5	0.3	0.5	UTAH	1968
74	2.221	0.4	0.3	0.6	UTAH	1969
75	2.748	0.8	0.3	0.9	UTAH	1970
76	0.831	0.2	0.7	0.0	OHIO	1956
77	1.011	0.4	0.7	0.8	OHIO	1957
78	0.390	0.3	0.7	0.2	OHIO	1958
79	0.052	0.2	0.7	0.3	OHIO	1959
80	1.202	0.2	0.7	0.3	OHIO	1960
81	0.819	0.8	0.7	0.8	OHIO	1961
82	1.018	0.1	0.7	0.0	OHIO	1962
83	1.533	0.6	0.7	0.7	OHIO	1963
84	0.570	0.9	0.7	0.7	OHIO	1964
85	1.627	0.9	0.7	0.8	OHIO	1965
86	0.111	0.5	0.7	0.0	OHIO	1966
87	2.190	0.9	0.7	0.4	OHIO	1967
88	2.316	0.5	0.7	0.7	OHIO	1968
89	1.916	0.4	0.7	0.5	OHIO	1969
90	1.040	0.8	0.7	0.1	OHIO	1970
91	0.919	0.2	0.9	0.1	MAINE	1956
92	0.547	0.4	0.9	0.1	MAINE	1957
93	0.442	0.3	0.9	0.3	MAINE	1958
94	0.656	0.2	0.9	0.6	MAINE	1959
95	0.622	0.2	0.9	0.7	MAINE	1960
96	0.896	0.8	0.9	0.5	MAINE	1961
97	1.534	0.1	0.9	0.4	MAINE	1962
98	1.805	0.6	0.9	0.3	MAINE	1963
99	0.992	0.9	0.9	0.7	MAINE	1964
100	0.560	0.9	0.9	0.1	MAINE	1965
101	0.937	0.5	0.9	0.5	MAINE	1966
102	0.033	0.9	0.9	0.5	MAINE	1967
103	0.957	0.5	0.9	0.0	MAINE	1968
104	0.367	0.4	0.9	0.3	MAINE	1969
105	1.542	0.8	0.9	0.7	MAINE	1970
106	0.516	0.2	0.1	0.4	IOWA	1956
107	0.766	0.4	0.1	0.5	IOWA	1957
108	1.135	0.3	0.1	0.0	IOWA	1958
109	1.198	0.2	0.1	0.1	IOWA	1959
110	3.249	0.2	0.1	0.5	IOWA	1960
111	0.717	0.8	0.1	0.7	IOWA	1961
112	0.301	0.1	0.1	0.6	IOWA	1962
113	1.133	0.6	0.1	0.9	IOWA	1963
114	0.372	0.9	0.1	0.7	IOWA	1964
115	0.541	0.9	0.1	0.6	IOWA	1965
116	0.703	0.5	0.1	0.4	IOWA	1966
117	1.401	0.9	0.1	0.8	IOWA	1967
118	0.428	0.5	0.1	0.6	IOWA	1968
119	0.952	0.4	0.1	0.9	IOWA	1969
120	0.823	0.8	0.1	0.1	IOWA	1970

Chapter II

Description of the Methods

1. Fuller and Battese Method

Various Components Model

Fuller and Battese [1974] have studied the model in which the random errors U_{ij} have the decomposition

$$U_{ij} = V_i + e_j + \epsilon_{ij} \qquad \begin{matrix} i = 1, 2, \ldots, N \\ j = 1, 2, \ldots, T \end{matrix}$$

with the errors V_i, e_j and ϵ_{ij} being independently distributed with zero means and variances $\sigma_v^2 \geq 0$, $\sigma_e^2 \geq 0$ and $\sigma_\epsilon^2 > 0$ respectively. As such, the authors are studying the common two way-random effects model with covariates.

In such a model, the covariance matrix for the vector of random errors $\underline{U}$ can be expressed as

$$E(\underline{U}\,\underline{U}') = V = \sigma_\epsilon^2 I_{NT} + \sigma_V^2 A + \sigma_e^2 B$$

where A is the Kronecker product matrix of I_N and J_T; B is the Kronecker product matrix of J_N and I_T; I_{NT}, I_N and I_T are identity matrices of order NT, N and T, respectively; J_N and J_T are $(N \times N)$ and $(T \times T)$ matrices, respectively, having all elements equal to one.

Given such an error structure, the variance components in V are estimated by the "fitting of constants" method [e.g., see Searle (1971b)] with the proviso that any negative variance component is set to zero for parameter estimation purposes. Estimated generalized least-squares are then performed via

$$\hat{\underline{\beta}} = (X'\hat{V}^{-1}X)^{-}X'\hat{V}^{-1}\underline{Y}$$

with the transformed model mean squared-error adjusted so as to estimate σ_ϵ^2. It should be pointed out that the authors obtain substantial computational

advantages in their procedure by presenting the fitting-of-constants estimators for the variance components in terms of deviations from appropriate means instead of creating the dummy variables for use in regressions.

Finally, Fuller and Battese give sufficient conditions in order for $\hat{\underline{\beta}}$ to be unbiased and asymptotically normally distributed.

2. Parks Method

Autoregressive Model

Parks [1967] considered the first-order autoregressive model in which the random errors U_{ij} $\begin{pmatrix} i = 1, 2, \ldots, N \\ j = 1, 2, \ldots, T \end{pmatrix}$ have the structure

$$E(U_{ij}^2) = \sigma_{ii} \qquad \text{(heteroscedasticity)}$$

$$E(U_{ij}U_{kj}) = \sigma_{ik} \qquad \text{(contemporaneously correlated)}$$

$$U_{ij} = \rho_i U_{i,j-1} + \epsilon_{ij} \qquad \text{(autoregression)}$$

where

$$E(\epsilon_{ij}) = 0$$

$$E(U_{i,j-1}\epsilon_{kj}) = 0$$

$$E(\epsilon_{ij}\epsilon_{kj}) = \phi_{ik}$$

$$E(\epsilon_{ij}\epsilon_{k\ell}) = 0 \qquad (j \neq \ell)$$

$$E(U_{io}) = 0$$

$$E(U_{io}^2) = \frac{\phi_{ii}}{1-\rho_i^2}$$

$$E(U_{io}U_{jo}) = \frac{\phi_{ij}}{1-\rho_i\rho_j}$$

As such, the model assumed is first-order autoregressive with contemporaneous correlation between cross-sections.

In such a model, the covariance matrix for the vector of random errors $\underline{U}$ can be expressed as

$$E(\underline{U}\,\underline{U}') = V = \begin{bmatrix} \sigma_{11}P_1 & \sigma_{12}P_{12} & \cdots & \sigma_{1N}P_{1N} \\ \sigma_{21}P_{21} & \sigma_{22}P_{22} & \cdots & \sigma_{2N}P_{2N} \\ \vdots & & & \\ \sigma_{N1}P_{N1} & \sigma_{N2}P_{N2} & \cdots & \sigma_{NN}P_{NN} \end{bmatrix}$$

where

$$P_{ij} = \begin{bmatrix} 1 & \rho_j & \rho_j^2 & \cdots & \rho_j^{T-1} \\ \rho_i & 1 & \rho_j & \cdots & \rho_j^{T-2} \\ \rho_i^2 & \rho_i & 1 & \cdots & \rho_j^{T-3} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_i^{T-1} & \rho_i^{T-2} & \rho_i^{T-3} & \cdots & 1 \end{bmatrix}$$

Given such an error structure, V is estimated by a 2-stage procedure leaving $\underline{\beta}$ to be estimated by the usual estimated generalized least-squares.

The first step in estimating V involves the use of ordinary least-squares to estimate $\underline{\beta}$ and obtain the fitted residuals

$$\hat{\underline{U}} = \underline{Y} - X\hat{\underline{\beta}}_{OLS}$$

A consistent estimator of the first-order autoregressive parameter can then be obtained in the usual manner via

$$\hat{\rho}_i = \frac{\sum_{j=2}^{T} \hat{U}_{ij}\hat{U}_{i,j-1}}{\sum_{j=2}^{T} \hat{U}^2_{i,t-1}} \qquad i = 1, 2, \ldots, N$$

Lastly, the autoregressive characteristic of the data can be removed (asymptotically) by the usual transformation of taking weighted differences.

That is, for $i = 1, 2, \ldots, N$

$$Y_{i1}\sqrt{1-\hat{\rho}_i^2} = \sum_{k=1}^{p}\sqrt{1-\hat{\rho}_i^2}\, X_{i1k}\beta_k + U_i\sqrt{1-\hat{\rho}_i^2}$$

$$Y_{ij} - \hat{\rho}_i Y_{i,j-1} = \sum_{k=1}^{p}(X_{ijk} - \hat{\rho}_i X_{i,j-1,k})\,\beta_k + U_{ij} - \hat{\rho}_i U_{i,j-1}$$

$$j = 2, 3, \ldots, T$$

which we write as

$$Y^*_{ij} = \sum_{k=1}^{P} X^*_{ijk}\,\beta_k + U^*_{ij} \qquad \begin{matrix} i = 1, 2, \ldots, N \\ j = 1, 2, \ldots, T \end{matrix}$$

Notice that the transformed model has not lost any observations like Kmenta (1971, pg. 512-14) does for simplicity.

The second step in estimating the covariance matrix V consists of applying ordinary least-squares to the above transformed model to obtain

$$\hat{\underline{U}}^* = \underline{Y}^* - X^*\hat{\underline{\beta}}^*_{OLS}$$

from which

$$S_{ij} = \frac{\hat{\phi}_{ij}}{1-\hat{\rho}_i\hat{\rho}_j}$$

where

$$\hat{\phi}_{ij} = \frac{1}{T-p}\sum_{k=1}^{T}\hat{U}^*_{ik}\,\hat{U}^*_{jk}$$

provides a consistent estimator of σ_{ij}. Estimated generalized least squares then proceeds in the usual fashion via

$$\hat{\underline{\beta}} = (X'\hat{V}^{-1}X)^{-}X'\hat{V}^{-1}Y$$

where $\hat{V}$ is the derived consistent estimator of V. For computational purposes, it should be pointed out that $\hat{\underline{\beta}}$ is obtained directly from the transformed model via

$$\hat{\underline{\beta}} = (X^{*'}\hat{\phi}^{-1}X^{*})^{-}X^{*'}\hat{\phi}^{-1}\underline{Y}^{*}$$

which has definite advantages. Further, the above procedure is obviously equivalent to Zellner's (1962) 2-stage methodology applied to the transformed model.

Parks is able to show that his estimator is consistent and asymptotically normally distributed with

$$\text{Var}(\hat{\underline{\beta}}) \doteq (X'\hat{V}^{-1}X)^{-1}$$

3. Da Silva Method

Mixed variance-component moving average model

Suppose there is available a sample of observations at T time points on each of A cross-sectional units. It is assumed that the observed value of the dependent variable at the $t^{\underline{th}}$ time point on the $i^{\underline{th}}$ cross-sectional unit can be expressed as:

$$y_{it} = \underline{x}'_{it}\underline{\beta} + a_i + b_t + e_{it} ,$$

$i = 1, \ldots, A$; $t = 1, \ldots, T$, where $\underline{x}'_{it} = (x_{it1} \cdots x_{itp})$ is a vector of explanatory variables for the $t^{\underline{th}}$ time point and $i^{\underline{th}}$ cross-sectional unit, $\underline{\beta} = (\beta_1 \cdots \beta_p)'$ is the vector of parameters, a_i is a time invariant cross-sectional unit effect, b_t is a cross-sectional unit invariant time effect,

and e_{it} is a residual effect unaccounted for by the explanatory variables and the specific time and cross-sectional unit effects.

If we arrange the observations first by individual units, then by time periods within individuals, we may write the equations in matrix notation as

$$\underline{y} = X\underline{\beta} + \underline{u} ,$$

where

$$\underline{u} = (\underline{a} \otimes \underline{1}_T) + (\underline{1}_A \otimes \underline{b}) + \underline{e} ,$$

$y = (y_{11} \cdots y_{1T}\ y_{21} \cdots y_{AT})'$, $X = (\underline{x}_{11} \cdots \underline{x}_{1T}\ \underline{x}_{21} \cdots \underline{x}_{AT})'$, $\underline{a} = a_1 \ldots a_A)'$,

$\underline{b} = (b_1 \ldots b_T)'$, $\underline{e} = (e_{11} \ldots e_{1T} e_{21} \ldots e_{AT})'$,

$\underline{1}_A$ is an $A \times 1$ vector with all elements equal to 1 , and $\otimes$ denotes Kronecker product.

It is assumed that:

Assumption 1.

$\{\underline{x}_{it}\}_{i=1,t=1}^{A,T}$ is a sequence of nonstochastic known $p \times 1$ vectors in R^p whose elements are uniformly bounded in R^p . The rank of X is p.

Assumption 2.

$\underline{\beta}$ is a $p \times 1$ constant vector of unknown parameters.

Assumption 3.

$\underline{a}$ is a vector of uncorrelated random variables such that $E(a_i) = 0$ and $Var(a_i) = \sigma_a^2$ (unknown) , $\sigma_a^2 > 0$, $i = 1, \ldots, A$.

Assumption 4.

$\underline{b}$ is a vector of uncorrelated random variables such that $E(b_t) = 0$ and $Var(b_t) = \sigma_b^2$ (unknown) , $\sigma_b^2 > 0$, $t = 1, \ldots, T$.

Assumption 5.

$\underline{e}_i = (e_{i1} \cdots e_{iT})'$ is a sample of a realization of a finite moving average time series of order M ($< T-1$) for each i ; hence

$e_{it} = \alpha_0 \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \cdots + \alpha_M \varepsilon_{t-M}$, $t = 1, \ldots, T$ $(i=1, \ldots A)$,

where α_0 , $\alpha_1, \cdots, \alpha_M$ are unknown constants such that $\alpha_0 \neq 0$ and $\alpha_M \neq 0$, and $\{\varepsilon_j\}_{j=-\infty}^{\infty}$ is a white noise process, i.e., a sequence of uncorrelated random variables with $E(\varepsilon_t) = 0$ and $E(\varepsilon_t^2) = \sigma^2$ (unknown) , $\sigma_\varepsilon^2 > 0$.

Assumption 6.

The sets of random variables $\{a_i\}_{i=1}^{A}$, $\{b_t\}_{t=1}^{T}$ and $\{e_{it}\}_{t=1}^{T}$ $(i = 1, \ldots, A)$ are mutually uncorrelated.

Assumption 7.

The random terms have normal distributions: $a_i \sim N(0,\sigma_a^2)$, $b_t \sim N(0,\sigma_b^2)$ and $\varepsilon_{t-k} \sim N(0,\sigma_\varepsilon^2)$, $i = 1, \ldots, A$; $t = 1, \ldots, T$; $k = 1, \ldots, M$.

If Assumptions 1-6 are satisfied, then

$$E(\underline{y}) = X\underline{\beta}$$

and

$$\mathrm{Var}(\underline{y}) = \sigma_a^2(I_A \otimes J_T) + \sigma_b^2(J_A \otimes I_T) + (I_A \otimes \Gamma_T) ,$$

where I_A is the identity matrix of size A , J_A is an $A \times A$ matrix with all elements equal to 1 and Γ_T is a $T \times T$ matrix with $ts^{\underline{th}}$ element

$$\mathrm{Cov}(e_{it},e_{is}) = \begin{cases} \gamma(|t-s|) & \text{if } |t-s| \le M \\ 0 & \text{if } |t-s| > M \end{cases} .$$

The covariance matrix, which we denote by V, can be written in the form

$$V = \sigma_a^2(I_A \otimes J_T) + \sigma_b^2(J_A \otimes I_T) + \sum_{k=0}^{M} \gamma(k)(I_A \otimes \Gamma_T^{(k)}) \quad \text{where}$$

$\Gamma_T^{(0)} = I_T$ and, for $k = 1, \ldots, M$, $\Gamma_T^{(k)}$ is a band matrix whose k^{th} off-diagonal elements are 1 and all other elements are 0.

Thus, the covariance matrix of the vector of observations $\underline{y}$ has the form

$$V(\underline{\nu}) = \sum_{k=1}^{M+3} \nu_k V_k ,$$

where $\underline{\nu} = (\nu_1, \ldots, \nu_{M+3})'$,

$$\nu_1 = \sigma_a^2$$

$$\nu_2 = \sigma_b^2$$

$$\nu_k = \gamma(k-3) , \quad k = 3, \ldots, M+3$$

$$V_1 = I_A \otimes J_T$$

$$V_2 = J_A \otimes I_T$$

$$V_k = I_A \otimes \Gamma_T^{(k-3)} , \quad k = 3, \ldots, M+3 .$$

The estimator of $\underline{\beta}$ is a two step GLS type estimator - GLS with the unknown covariance matrix replaced by a suitable estimator of V. It is obtained by substituting estimators for the scalar multiples ν_k, $k = 1, 2, \ldots, M + 3$

Seely (1969) presents a general theory of unbiased estimation when the choice of estimators is restricted to finite dimensional vector spaces, with special emphasis to quadratic estimation of functions of the form

$$\sum_{i=1}^{m} \delta_i \nu_i ,$$

where ν_i $(i = 1, \ldots, m)$ are parameters associated with a linear model $E(\underline{y}) = X\underline{\beta}$ with covariance matrix

$$\sum_{i=1}^{m} \nu_i V_i$$

and V_i $(i=1, \ldots, m)$ are real symmetric matrices. The method is also discussed in Seely (1970a, 1970b) and Seely and Zyskind (1971). Seely and Soong (1971) consider the MINQUE principle using an approach along the lines of Seely (1969).

The method employed here is the one developed by Seely. The concepts and results necessary for our purposes are presented next.

Let $\underline{y}$ be an $n \times 1$ random vector with expectation $X\underline{\beta}$ and covariance matrix

$$\sum_{i=1}^{m} \nu_i V_i ,$$

where X is a known $n \times p$ matrix of rank p, each V_i is a known $n \times n$ real symmetric matrix and $\underline{\beta} = (\beta_1 \ldots \beta_p)'$ and $\underline{\nu} = (\nu_1 \ldots \nu_m)'$ are vectors of unknown parameters.

Let $\mathcal{C}$ denote the set of the $n \times n$ real symmetric matrices and let $\bar{\mathcal{C}}$ denote the space of quadratic estimators with matrices from $\mathcal{C}$, i.e., $\bar{\mathcal{C}} = \{\underline{y}'A\underline{y}: A \in \mathcal{C}\}$. We shall restrict considerations to quadratic estimators $\underline{y}'A\underline{y}$, $A \in \mathcal{C}$, that satisfy the condition $AX = 0$. So, the

space of estimators we consider is the subspace of $\bar{\mathcal{C}}$, $\bar{n} = \{\underline{y}'A\underline{y}: A \in \mathcal{C}, AX = 0\}$.

Definition

A parametric function

$$\sum_{i=1}^{m} \delta_i \nu_i$$

is said to be estimable if there is an $A \in \mathcal{C}$, $AX = 0$, such that $\underline{y}'A\underline{y}$ is an unbiased estimator of the parametric function.

One reason for limiting considerations to the subspace $\bar{n}$ of quadratic estimators is the desirable property that a quadratic estimator $\underline{y}'A\underline{y}$ be invariant under the class of transformations of the form $\underline{y} + X\underline{\alpha}$ for arbitrary $\underline{\alpha}$. This seems to be a reasonable requirement of estimators for linear functions of the parameters in the covariance matrix. Also, the class of multivariate normal distributions $\{N_n(X\underline{\beta}, \sum_{i=1}^{m} \nu_i V_i): \beta \in R^p, \underline{\nu} \in R^m\}$ remains invariant under this class of transformations. Another criterion which leads to considering only the subspace $\bar{n}$ is to require that the variance of a quadratic form $\underline{y}'A\underline{y}$ be independent of the parameter $\underline{\beta}$. Hsu (1938) used this criterion and showed that requiring $\mathrm{Var}(\underline{y}'A\underline{y})$ to be independent of $\underline{\beta}$ implied that $AX = 0$.

Theorem

The parameters $\nu_1, \ldots, \nu_m$ are all estimable if and only if the matrix B with ij^{th} element $\mathrm{tr}(NV_iNV_j)$,

$$N = I - X(X'X)^{-1}X',$$

is nonsingular.

In checking for the estimability of the variance and covariance components, the following result is of interest.

Corollary

The parameters $\nu_1, \ldots, \nu_m$ are all estimable if and only if the matrices $NV_1N, \ldots, NV_mN$ are linearly independent.

Theorem

If the parameters $\nu_1, \ldots, \nu_m$ are all estimable, then the unique unbiased estimator $\hat{\underline{\nu}}$ of $\underline{\nu}$ is the solution to the set of equations

$$B\hat{\underline{\nu}} = \underline{c} ,$$

where B is the matrix defined in the previous Theorem and $\underline{c}$ is the $m \times 1$ vector with $i^{\underline{th}}$ element $\underline{y}'NV_iN\underline{y}$.

It can be seen that Seely's estimator is, in fact, an unweighted MINQUE estimator, i.e., a MINQUE estimator with the prior estimate of the covariance matrix replaced by the identity matrix I , as representing complete ignorance of the true variance and covariance components.

The reason for choosing Seely's estimator is its computational advantage, since the derivation of the MINQUE estimator would require the inversion of the $AT \times AT$ matrix representing prior information about the covariance matrix. Besides, as usually no reasonable guess on variance and covariance components is available, one may do better by not using any prior information in the estimation procedure.

The matrix V has a complicated structure, because its characteristic vectors, as well as characteristic roots, are functions of the parameters σ_a^2 , σ_b^2 and $\gamma(h)$, $h = 0, 1, \ldots, M$. As a consequence, the asymptotic theory with the unknown covariance matrix V is intractable. What we shall

do is to follow Hannan (1963) and substitute an approximating matrix $\overset{*}{V}$ whose characteristic roots depend on the unknown parameters and whose characteristic vectors do not. This matrix is defined by

$$\overset{*}{V} = \sigma_a^2(I_A \otimes J_T) + \sigma_b^2(J_A \otimes I_T) + (I_A \otimes \overset{*}{\Gamma}_T) \quad ,$$

where $\overset{*}{\Gamma}_T$ is the circular symmetric matrix of dimension T, i.e., $\overset{*}{\Gamma}_T = Q_T \Lambda_T Q_T'$, where Λ_T is the diagonal matrix with the $t^{\underline{th}}$ diagonal element

$$d_t = \begin{cases} \gamma(0) + 2\sum_{h=1}^{M} \gamma(h)\cos(\pi th/T) \\ \qquad t = 2, 4, \ldots, T-1 \text{ (T odd) or } T \text{ (T even)} \\ \gamma(0) + 2\sum_{h=1}^{M} \gamma(h)\cos[\pi(t-1)h/T] \\ \qquad t = 1, 3, \ldots, T \text{ (T odd) or } T-1 \text{ (T even)} \end{cases}$$

and Q_T' is the $T \times T$ orthonormal matrix with $ts^{\underline{th}}$ element

$$q_{ts} = \begin{cases} \sqrt{1/T}\,, & t = 1 \\ \sqrt{2/T}\cos[\pi t(s-1)/T]\,, & t = 2, 4, \ldots, T-2 \text{ (T even)} \\ & \text{or } T-1 \text{ (T odd)} \\ \sqrt{2/T}\sin[\pi(t-1)(s-1)/T]\,, & t = 3, 5, \ldots, T-1 \\ & (T \text{ even}) \text{ or } T \text{ (T odd)} \\ \sqrt{1/T}\,(-1)^{s+1}\,, & t = T \text{ if } T \text{ is even}\,, \end{cases}$$

$s = 1, \ldots, T$.

The fundamental result obtained by Hannan (1963) is a consequence of particularly relevant properties of the covariance structure of stationary time series. The properties manifest themselves in the following theorem. Proofs of the theorem are given by Grenander and Szego (1958), Amemiya and Fuller (1967) and Fuller (1971b).

Theorem

Let Γ_T be the covariance matrix of a realization of T observations from a stationary time series with absolutely summable covariance function. Then, for every $\epsilon > 0$ there exists a T_ϵ such that for $T > T_\epsilon$

$$Q_T'\Gamma_T Q_T - \Lambda_T = E_T ,$$

where every element of the matrix E_T has absolute value less than ϵ ; the matrices Q_T and Λ_T are as defined above.

This theorem establishes that, asymptotically, the matrix Q_T will diagonalize all covariance matrices associated with stationary time series with absolutely summable covariance function. This fact was the motivation for the result established by Hannan (1963) which we now state in the following theorem:

Theorem

Consider the regression equations $\underline{Y} = \Phi\underline{\beta} + \underline{v}$, where $\underline{Y}$ is a vector of T components, Φ is a $T \times p$ matrix, $\underline{\beta}$ is a $p \times 1$ vector of parameters, and $\underline{v} = (v_1 \ldots v_T)'$ is independent of Φ . Suppose v_t , $t = 1, \ldots T$, is generated by a stationary process with absolutely summable covariance function and that Φ is susceptible to a "generalized harmonic analysis". Let $\Delta_T = \text{diag}(\Phi'\Phi)$.. Then, as $T \to \infty$, $\Delta_T(\underline{b} - \underline{\beta})$ has the same limiting covariance

matrix as $\Delta_T(\tilde{\underline{\beta}} - \underline{\beta})$ where $\tilde{\underline{\beta}}$ is the GLS estimator, $\tilde{\underline{\beta}} = (\Phi'\Gamma_T^{-1}\Phi)^{-1}\Phi'\Gamma_T^{-1}\underline{Y}$, and $\underline{b}$ is the estimator with the same expression as $\tilde{\underline{\beta}}$ except that Γ_T is replaced by $\overset{*}{\Gamma}_T$ defined above.

The assumptions on Φ amount to the following, where $\Phi = (\underline{\phi}_1 \cdots \underline{\phi}_p)$, $\underline{\phi}_i = (\phi_{1i} \cdots \phi_{Ti})'$,

(i) $$\lim_{T\to\infty} \frac{\sum_{t=1}^{T+h} \phi_{ti}^2}{\sum_{t=1}^{T} \phi_{ti}^2} = 1 \quad \text{for all } i \text{ and finite } h ;$$

(ii) $$\lim_{T\to\infty} \frac{\sum_{t=1}^{T-h} \phi_{ti}\phi_{t+h,j}}{\sum_{t=1}^{T} \phi_{ti}^2 \sum_{t=1}^{T} \phi_{tj}^2} \quad \text{exists and is finite for}$$

all i, j, h .

The implication of this result is that in regression problems with stationary errors the transformation of the observations by the matrix Q_T yields an approximately uncorrelated set of data. In this way the problem is reduced to one in ordinary weighted regression in which the weights can be estimated from the data. The weights to be estimated are, in fact, proportional to the spectral density of the process at frequencies $\pi t/T$ corresponding to the columns of Q_T. This fact suggests the terminology "conversion to the frequency domain" used to describe the procedure by several authors.

The spectral density of a stationary process $\{v_t\}_{t=\infty}^{\infty}$ is the Fourier series

$$f(w) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h)\cos(wh) ,$$

where $\gamma(h) = \mathrm{Cov}(v_t, v_{t+h})$. In our case, $\gamma(h) = 0$ for $|h| > M$; so, the spectral density becomes

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-M}^{M} \gamma(h)\cos(\omega h) ,$$

and can be estimated by

$$\hat{f}(\omega) = \frac{1}{2\pi} \sum_{h=-M}^{M} \hat{\gamma}(h)\cos(\omega h) ,$$

where $\hat{\gamma}(h)$, $h = -M, \ldots, -1, 0, 1, \ldots, M$, are appropriate estimators of the autocovariances (see Anderson (1971), Fuller (1971b), or any of the standard references in time series analysis for details).

Hannan's method has been applied to a variety of regression problems by Hamon and Hannan (1963). Duncan and Jones (1966) present a practical computing technique for the application of the method. Engle (1974) presents and extended version of Hannan's method of regression analysis in the frequency domain. Engle argues that frequency domain regression analyses have the standard small sample properties, that they do not require smoothing of the periodogram for the estimation of the spectrum and are computationally easy to use. The application of the technique has been extended to distributed lag analysis by Hannan (1965, 1967), Amemiya and Fuller (1967), Fishman (1969), Dhrymes (1971) and Sims (1971).

We next establish the spectral decomposition of $\overset{*}{V}$.

Theorem

The matrix $\overset{*}{V}$, defined previously, has spectral decomposition

$$\overset{*}{V} = (A\sigma_b^2 + T\sigma_a^2 + d_1)(o_1 \otimes q_1)(o_1 \otimes q_1)'$$

$$+ \sum_{t=2}^{T} (A\sigma_b^2 + d_t)(o_1 \otimes q_t)(o_1 \otimes q_t)'$$

$$+ \sum_{i=2}^{A} (T\sigma_a^2 + d_1)(o_i \otimes q_1)(o_i \otimes q_1)'$$

$$+ \sum_{t=2}^{T} \sum_{i=2}^{A} d_t(o_i \otimes q_t)(o_i \otimes q_t)' ,$$

where q_t , $t = 1, \ldots, T$, is the $t^{\underline{th}}$ column of the matrix Q_T defined previously; $o_1 = \frac{1}{\sqrt{A}} 1_A$; o_i , $i = 2, \ldots, A$, is any set of A-1 orthonormal contrasts of dimension A , and d_t , $t = 1, \ldots, T$, as above.

Let $\hat{\sigma}_a^2$, $\hat{\sigma}_b^2$ and $\hat{\gamma}(h)$, $h = 0, \ldots, M$, denote Seely's estimators of the corresponding unknown parameter values. Let $\hat{d}_t$ be defined by

$$\hat{d}_t = \begin{cases} \hat{\gamma}(0) + 2\sum_{h=1}^{M} \hat{\gamma}(h)\cos(\pi th/T) \\ \qquad t = 2, 4, \ldots, T-1 \text{ (T odd) or T (T even)} \\ \hat{\gamma}(0) + 2\sum_{h=1}^{M} \hat{\gamma}(h)\cos[\pi(t-1)h/T] \\ \qquad t = 1, 3, \ldots, T \text{ (T odd) or T-1 (T even)} \end{cases}$$

These estimators are not guaranteed to be nonnegative definite. In order to get a positive definite estimator for the covariance matrix, we define

$$\tilde{\sigma}_a^2 = \begin{cases} \hat{\sigma}_a^2 , & \text{if } \hat{\sigma}_a^2 > 0 \\ 0 , & \text{otherwise} \end{cases}$$

$$\tilde{\sigma}_b^2 = \begin{cases} \hat{\sigma}_b^2 , & \text{if } \hat{\sigma}_b^2 > 0 \\ 0 , & \text{otherwise} \end{cases}$$

$$\tilde{d}_t = \begin{cases} \hat{d}_t , & \text{if } \hat{d}_t > c \\ c , & \text{otherwise} \end{cases}$$

where c is an arbitrary positive real number. In case $\hat{d}_t > 0$ for at least one t, we may set

$$c = \min_{t \in T} \{\hat{d}_t : \hat{d}_t > 0\} .$$

We use as estimator of the covariance matrix V in the GLS phase of the estimation procedure the matrix

$$\tilde{\overset{*}{V}} = \tilde{\sigma}_a^2(I_A \otimes J_T) + \tilde{\sigma}_b^2(J_A \otimes I_T) + (I_A \otimes Q_T \tilde{\Lambda}_T Q_T') ,$$

where $\tilde{\Lambda}_T = \text{diag}(\tilde{d}_1 \ldots \tilde{d}_T)$ and the other matrices on the right-hand side are as defined above.

The proposed estimator of the vector of regression parameters $\underline{\beta}$ is

$$\tilde{\underline{\beta}} = (X' \tilde{\overset{*}{V}}^{-1} X)^{-1} X' \tilde{\overset{*}{V}}^{-1} \underline{y} .$$

Chapter III

Monte Carlo Simulation

The matrix $\tilde{\overset{*}{V}}$ is chosen over the matrix V with the parameter estimators replacing unknown parameters, $\tilde{V}$, because:

(i) $\tilde{\overset{*}{V}}$ has a known spectral decomposition, making its use computationally feasible for large T;

(ii) the substitution of $\tilde{\overset{*}{V}}$ for $\tilde{V}$ allows the use of known results to prove the unbiasedness and symptotic normality of the regression coefficient estimators.

The use of $\tilde{\overset{*}{V}}$ instead of $\hat{\overset{*}{V}}$, i.e., $\overset{*}{V}$ with Seely's estimators in place of the unknown parameters, guarantees a positive definite estimator for the covariance matrix.

Simulation Model

The primary purpose of the simulation study was to provide a final check that the computer routines were reasonably correct. At the same time, however, we recognized that the results might be useful to individuals wanting

(1) To study the efficiencies of a model compared to ordinary least-squares.

and

(2) To compare the robustness of the various models under competing error structures.

For these reasons, a large amount of auxillary information is included at the end of this section.

For simplicity we took a four parameter model (i.e. p=4) with 10 cross-sections (i.e. N=10) and each time series of length 15 (i.e. T=15). The fixed part of the model was then specified by X given in Table 3.0.

and

$$\underline{\beta}' = (1, 1, 1, 1)$$

Notice that X has one factor constant only over time, one only over cross-sections, one constant over both times series and cross-sections (i.e. the mean) and a factor which varies over time series and cross-sections. It was felt that such a structure occurs frequently in combining time series and cross-sectional data and would show to good advantage the lack of any rank assumptions imposed in our models.

Next, we have the random part of the model. For this we chose

$$\underline{U} \sim N(\underline{0}, V)$$

Table 3.0. The Design Matrix

Row	X-1	X-2	X-3	X-4	Row	X-1	X-2	X-3	X-4
1	1.0	0.2	0.3	0.1	76	1.0	0.2	0.4	0.7
2	1.0	0.4	0.3	0.7	77	1.0	0.4	0.4	0.4
3	1.0	0.3	0.3	0.2	78	1.0	0.3	0.4	0.9
4	1.0	0.2	0.3	0.0	79	1.0	0.2	0.4	0.5
5	1.0	0.2	0.3	0.0	80	1.0	0.2	0.4	0.2
6	1.0	0.8	0.3	0.9	81	1.0	0.8	0.4	0.4
7	1.0	0.1	0.3	0.7	82	1.0	0.1	0.4	0.6
8	1.0	0.6	0.3	0.1	83	1.0	0.6	0.4	0.3
9	1.0	0.9	0.3	0.5	84	1.0	0.9	0.4	0.4
10	1.0	0.9	0.3	0.0	85	1.0	0.9	0.4	0.3
11	1.0	0.5	0.3	0.9	86	1.0	0.5	0.4	0 3
12	1.0	0.9	0.3	0.6	87	1.0	0.9	0.4	0.3
13	1.0	0.5	0.3	0.5	88	1.0	0.5	0.4	0.6
14	1.0	0.4	0.3	0.6	89	1.0	0.4	0.4	0.4
15	1.0	0.8	0.3	0.9	90	1.0	0.8	0.4	0.9
16	1.0	0.2	0.3	0.2	91	1.0	0.2	0.7	0.0
17	1.0	0.4	0.3	0.4	92	1.0	0.4	0.7	0.8
18	1.0	0.3	0.3	0.2	93	1.0	0.3	0.7	0.2
19	1.0	0.2	0.3	0.1	94	1.0	0.2	0.7	0.3
20	1.0	0.2	0.3	0.7	95	1.0	0.2	0.7	0.3
21	1.0	0.8	0.3	0.0	96	1.0	0.8	0.7	0.8
22	1.0	0.1	0.3	0.3	97	1.0	0.1	0.7	0.0
23	1.0	0.6	0.3	0.7	98	1.0	0.6	0.7	0.7
24	1.0	0.9	0.3	0.6	99	1.0	0.9	0.7	0.7
25	1.0	0.9	0.3	0.6	100	1.0	0.9	0.7	0.8
26	1.0	0.5	0.3	0.8	101	1.0	0.5	0.7	0.0
27	1.0	0.9	0.3	0.4	102	1.0	0.9	0.7	0.4
28	1.0	0.5	0.3	0.6	103	1.0	0 5	0.7	0.7
29	1.0	0.4	0.3	0.0	104	1.0	0.4	0.7	0 5
30	1.0	0.8	0.3	0.5	105	1.0	0.8	0.7	0.1
31	1.0	0.2	0.2	0.9	106	1.0	0.2	0.4	0.5
32	1.0	0.4	0.2	0.6	107	1.0	0.4	0.4	0.3
33	1.0	0.3	0.2	0.2	108	1.0	0.3	0.4	0.8
34	1.0	0.2	0.2	0.6	109	1.0	0.2	0.4	0 3
35	1.0	0.2	0.2	0.2	110	1.0	0.2	0.4	0.8
36	1.0	0.8	0.2	0.7	111	1.0	0.8	0.4	0.3
37	1.0	0.1	0.2	0.9	112	1.0	0.1	0.4	0.0
38	1.0	0.6	0.2	0.5	113	1.0	0.6	0.4	0.5
39	1.0	0.9	0.2	0.3	114	1.0	0.9	0.4	0.0
40	1.0	0 9	0.2	0 5	115	1.0	0.9	0.4	0.9
41	1.0	0.5	0.2	0.5	116	1.0	0.5	0.4	0.1
42	1.0	0.9	0.2	0.8	117	1.0	0.9	0.4	0.3
43	1.0	0.5	0.2	0.7	118	1.0	0.5	0.4	0.7
44	1.0	0.4	0.2	0.4	119	1.0	0.4	0.4	0.0
45	1.0	0.8	0.2	0.9	120	1.0	0.8	0.4	0.9
46	1.0	0.2	0.3	0.7	121	1.0	0.2	0.9	0.1
47	1.0	0.4	0.3	0.5	122	1.0	0.4	0.9	0.1
48	1.0	0.3	0.3	0.2	123	1.0	0.3	0.9	0.3
49	1.0	0.2	0.3	0.5	124	1.0	0.2	0.9	0.6
50	1.0	0.2	0.3	0.0	125	1.0	0.2	0.9	0.7
51	1.0	0.8	0.3	0.0	126	1.0	0.8	0.9	0.5
52	1.0	0.1	0.3	0.1	127	1.0	0.1	0.9	0.4
53	1.0	0.6	0.3	0.3	128	1.0	0.6	0.9	0.3
54	1.0	0.9	0.3	0.7	129	1.0	0.9	0.9	0.7
55	1.0	0.9	0.3	0.6	130	1.0	0.9	0.9	0.1
56	1.0	0 5	0.3	0.8	131	1.0	0.5	0.9	0.5
57	1.0	0.9	0.3	0.1	132	1.0	0.9	0.9	0.5
58	1.0	0.5	0.3	0.6	133	1.0	0.5	0.9	0.0
59	1.0	0.4	0.3	0.5	134	1.0	0.4	0.9	0.3
60	1.0	0.8	0.3	0.5	135	1.0	0.8	0.9	0.7
61	1.0	0.2	0.8	0.0	136	1.0	0.2	0.1	0.4
62	1.0	0.4	0.8	0.7	137	1.0	0.4	0.1	0.5
63	1.0	0.3	0.8	0.8	138	1.0	0.3	0.1	0.0
64	1.0	0.2	0.8	0.0	139	1.0	0.2	0.1	0.1
65	1.0	0.2	0.8	0.0	140	1.0	0.2	0.1	0.5
66	1.0	0.8	0.8	0.8	141	1.0	0.8	0.1	0.7
67	1.0	0.1	0.8	0.6	142	1.0	0.1	0.1	0.6
68	1.0	0.6	0.8	0.4	143	1.0	0.6	0.1	0.9
69	1.0	0.9	0.8	0.5	144	1.0	0.9	0.1	0.7
70	1.0	0.9	0.8	0.0	145	1.0	0.9	0.1	0.6
71	1.0	0.5	0.8	0.9	146	1.0	0.5	0.1	0.4
72	1.0	0.9	0.8	0.1	147	1.0	0.9	0.1	0.8
73	1.0	0.5	0.8	0.4	148	1.0	0.5	0.1	0.6
74	1.0	0.4	0.8	0.0	149	1.0	0.4	0.1	0.9
75	1.0	0.8	0.8	0.4	150	1.0	0.8	0.1	0.1

where V was specified by the following components

(i) Fuller-Battese Model

Here we took

$$\sigma_V^2 = \sigma_e^2 = \sigma_\varepsilon^2 = .5$$

NOTE: On searching the literature we could find no evidence that any one component is dominant and hence this equal allocation:

(ii) Parks Method

Here we specified ϕ and $\underline{\rho}$ as follows

POPULATION PHI MATRIX

$$\phi = \begin{bmatrix}
0.29400 & 0.09800 & 0.13859 & 0.13859 & 0.16974 & 0.16974 & 0.19600 & 0.19600 & 0.21913 & 0.21913 \\
0.09800 & 0.29400 & 0.13859 & 0.13859 & 0.16974 & 0.16974 & 0.19600 & 0.19600 & 0.21913 & 0.21913 \\
0.13859 & 0.13859 & 0.58800 & 0.19600 & 0.24005 & 0.24005 & 0.27719 & 0.27719 & 0.30990 & 0.30990 \\
0.13859 & 0.13859 & 0.19600 & 0.58800 & 0.24005 & 0.24005 & 0.27719 & 0.27719 & 0.30990 & 0.30990 \\
0.16974 & 0.16974 & 0.24005 & 0.24005 & 0.88200 & 0.29400 & 0.33948 & 0.33948 & 0.37955 & 0.37955 \\
0.16974 & 0.16974 & 0.24005 & 0.24005 & 0.29400 & 0.88200 & 0.33948 & 0.33948 & 0.37955 & 0.37955 \\
0.19600 & 0.19600 & 0.27719 & 0.27719 & 0.33948 & 0.33948 & 1.17600 & 0.39200 & 0.43827 & 0.43827 \\
0.19600 & 0.19600 & 0.27719 & 0.27719 & 0.33948 & 0.33948 & 0.39200 & 1.17600 & 0.43827 & 0.43827 \\
0.21913 & 0.21913 & 0.30990 & 0.30990 & 0.37955 & 0.37955 & 0.43827 & 0.43827 & 1.47000 & 0.49000 \\
0.21913 & 0.21913 & 0.30990 & 0.30990 & 0.37955 & 0.37955 & 0.43827 & 0.43827 & 0.49000 & 1.47000
\end{bmatrix}$$

POPULATION AUTOREGRESSIVE PARAMETERS

$$\underline{\rho} = \begin{bmatrix}
0.60000 \\
0.80000 \\
0.60000 \\
0.80000 \\
0.60000 \\
0.80000 \\
0.60000 \\
0.80000 \\
0.60000 \\
0.80000
\end{bmatrix}$$

so that observations on the same time period had the following covariance structure over cross-sections

POPULATION ERROR STRUCTURE

$$\begin{bmatrix}
0.45937 & 0.18846 & 0.21655 & 0.26652 & 0.26522 & 0.32642 & 0.30625 & 0.37692 & 0.34240 & 0.42141 \\
0.18846 & 0.81667 & 0.26652 & 0.38498 & 0.32642 & 0.47150 & 0.37692 & 0.54444 & 0.42141 & 0.60871 \\
0.21655 & 0.26652 & 0.91875 & 0.37692 & 0.37508 & 0.46163 & 0.43310 & 0.53305 & 0.48422 & 0.59597 \\
0.26652 & 0.38498 & 0.37692 & 1.63333 & 0.46163 & 0.66681 & 0.53305 & 0.76996 & 0.59597 & 0.86084 \\
0.26522 & 0.32642 & 0.37508 & 0.46163 & 1.37812 & 0.56538 & 0.53044 & 0.65285 & 0.59305 & 0.72991 \\
0.32642 & 0.47150 & 0.46163 & 0.66681 & 0.56538 & 2.45000 & 0.65285 & 0.94301 & 0.72991 & 1.05431 \\
0.30625 & 0.37692 & 0.43310 & 0.53305 & 0.53044 & 0.65285 & 1.83750 & 0.75385 & 0.68480 & 0.84283 \\
0.37692 & 0.54444 & 0.53305 & 0.76996 & 0.65285 & 0.94301 & 0.75385 & 3.26667 & 0.84283 & 1.21741 \\
0.34240 & 0.42141 & 0.48422 & 0.59597 & 0.59305 & 0.72991 & 0.68480 & 0.84283 & 2.29687 & 0.94231 \\
0.42141 & 0.60871 & 0.59597 & 0.86084 & 0.72991 & 1.05431 & 0.84283 & 1.21741 & 0.94231 & 4.08333
\end{bmatrix}$$

This selection resulted in

$$\underset{i = 1,2,\ldots,N}{\text{med}} \quad \text{Var}(U_{it}) \doteq 1.5$$

That is, variance between the models was standardized in this manner. This allows for the heteroscedastic nature of the cross-sections, varying correlations, and different autoregressive parameters in the cross-sections.

In retrospect, perhaps a preferred way of introducing these characteristics would have been to equate

$$E(U_{it}U_{jt} \mid \text{Fuller}) = E(U_{it}U_{jt} \mid \text{Parks})$$

$$\text{i.e. } \phi_{ij} = \begin{cases} (1-\rho_i^2)\,(\sigma_v^2 + \sigma_e^2 + \sigma_\varepsilon^2) & i \neq j \\ (1-\rho_i\rho_j)\,\sigma_v^2 & i \neq j \end{cases}$$

Hence, by varying ρ the cross-sections could have been made heteroscedastic, with different autoregressive parameters, and varying correlations while at the same time matching variances and covariances - a much cleaner approach. Even

here, however, there is no matching on covariances on different time periods even for the same cross-section.

(iii) Da Silva Method

Here we took

$$\sigma_A^2 = .5$$

$$\sigma_B^2 = .5$$

$$\alpha_i = \frac{\rho^i}{Q} \qquad i = 0, 1, 2, \ldots, 7$$

such that

$$\sum_{i=0}^{7} \alpha_i^2 = 1$$

and

$$\rho = .7$$

$$M = 7$$

This produced the following autocovariance function:

$$\gamma = \begin{bmatrix} .50000 \\ .34879 \\ .24241 \\ .16721 \\ .11351 \\ .07440 \\ .04485 \\ .02107 \end{bmatrix}$$

NOTE: $\gamma(0) = \sigma_\epsilon^2 = .5$

Notice that every observation then has a variance of 1.5 (as in (i)).

In order to make the desired comparison of the possible error structures we ran a simulation on each of the three error structures and analyzed all 4 candidate alternatives.

Analysis Assuming Error Structure \ Generating Error Structure	Variance Component	Autoregressive	Moving Average
Fuller-Battese	1	5	9
Parks	2	6	10
Da Silva	3	7	11
Ordinary Least Squares	4	8	12

Details of Simulation Runs

To generate the normal errors we used the IMSL routines GGUSN and GGNOF to obtain N(0,1) independent errors which were appropriately transformed to provide the required error structures - program listings are available upon request. It was decided to do 500 runs on each generating error structure (which requires approximately 2 hours total connect time for all three generating structures). To be consistent with the aim of the study, we collected data on the following questions

(i) Are the estimators unbiased?

(ii) When the assumed distribution is correct, do efficiencies result with respect to reduced variance in the parameter estimates?

and

(iii) Are the estimates approximately normally distributed so that valid inferences can be made?

In addition, error structure parameters were estimated and compared with their input value as a further check on the accuracy of the program and the simulation itself.

In presenting the results of the simulation, we decided to exclude from analysis any cases where standard fix-ups (or no estimation at all!) takes place. A count is given, however, of how often this occurs.

Table 3.1 summarizes the pertinent point estimate features of the models. For convenience, the rows of each model cell represent the four parameters in each model. Next, from an inferential point of view, it is instructive to compare the marginal empirical distribution of the standardized estimates ($t^* = \frac{\hat{\beta} - 1}{s.e.(\hat{\beta})}$) . Tables 3.2 - 3.4 do this for the usual three error generating structures (t_i^* corresponds to the t-statistic for testing β_i).

Finally, the estimates of the error structure parameters based on 450 iterations are given in Table 3.5. In addition, ϕ in the Parks model had average estimated component in 450 iterations given by

$$\hat{\phi} = \begin{bmatrix} .29340 & .08715 & .11339 & .10179 & .13517 & .13324 & .16401 & .14049 & .17747 & .14581 \\ & .29036 & .12155 & .12367 & .13374 & .13468 & .17307 & .16536 & .17094 & .17815 \\ & & .56925 & .17472 & .19098 & .19483 & .23566 & .23092 & .25221 & .23564 \\ & & & .56619 & .19250 & .20471 & .23853 & .22972 & .25954 & .24582 \\ & & & & .77420 & .21895 & .26352 & .24590 & .28535 & .29032 \\ & & & & & .80642 & .26726 & .28204 & .30862 & .32905 \\ & \text{symmetric} & & & & & 1.07532 & .31621 & .33445 & .34521 \\ & & & & & & & 1.08544 & .35912 & .34790 \\ & & & & & & & & 1.29233 & .40955 \\ & & & & & & & & & 1.33411 \end{bmatrix}$$

Table 3.1. Summary statistics (estimates) for simulation study.

Model	Mean	Variance	MSE	$(\text{bias})^2$	Scale	No. iterations
1	1.01682	0.37470	0.37423	0.00028	0.50008	500
	0.92155	0.44725	0.45251	0.00616		
	1.05468	0.84886	0.85016	0.00299		
	1.00079	0.04627	0.04618	0.00001		
2	0.98768	0.45895	0.45818	0.00015	1.01102	499
	0.93717	0.58257	0.58535	0.00395		
	1.07050	1.00520	1.00815	0.00497		
	0.99723	0.08387	0.08371	0.00001		
3	1.01804	0.40382	0.40332	0.00033	321.039	493
	0.91422	0.46396	0.47037	0.00736		
	1.04990	0.84448	0.84525	0.00249		
	1.00527	0.17406	0.17373	0.00003		
4	1.02103	0.38338	0.38305	0.00044	1.34320	500
	0.92334	0.44910	0.45408	0.00588		
	1.05303	0.85166	0.85276	0.00281		
	0.99089	0.08286	0.08278	0.00008		
5	0.96188	0.80842	0.80824	0.00145	0.92288	494
	1.04793	0.97488	0.97520	0.00230		
	1.00393	0.84230	0.84061	0.00002		
	1.00248	0.07655	0.07640	0.00001		
6	0.99752	0.53179	0.53067	0.00001	1.00494	472
	1.00263	0.23268	0.23219	0.00001		
	0.97294	0.78236	0.78143	0.00073		
	1.01257	0.04447	0.04453	0.00016		
7	0.90412	0.70474	0.71086	0.00919	115.816	229
	1.03174	0.44053	0.43962	0.00101		
	1.04364	0.89614	0.89414	0.00190		
	1.02735	0.11507	0.11532	0.00075		
8	0.96779	0.83661	0.83598	0.00104	1.61893	500
	1.04394	0.96044	0.96045	0.00193		
	0.99837	0.85431	0.85261	<0.00001		
	0.99782	0.08730	0.08713	<0.00001		
9	1.02165	0.44958	0.44915	0.00047	0.38313	500
	1.00299	0.44328	0.44241	0.00001		
	0.95321	1.02495	1.02509	0.00219		
	0.99732	0.03224	0.03218	0.00001		
10	1.02404	0.47698	0.47660	0.00058	1.00988	493
	0.99684	0.61937	0.61813	0.00001		
	0.95165	1.13919	1.13922	0.00234		
	1.00677	0.04207	0.04204	0.00005		
11	1.03465	0.44572	0.44598	0.00120	208.456	476
	0.99184	0.43748	0.43663	0.00001		
	0.94387	0.99305	0.99411	0.00315		
	0.98988	0.06372	0.06369	0.00010		
12	1.02155	0.44773	0.44730	0.00046	1.32475	500
	1.00295	0.44601	0.44513	0.00001		
	0.95325	1.01999	1.02014	0.00216		

Tabular Values		C	-2.57580	-1.96000	-1.64480	-0.67450	0.00000	0.67450	1.64480	1.96000	2.5758
		$P(z \leq C)$	0.00500	0.02500	0.05000	0.25000	0.50000	0.75000	0.95000	0.97500	0.9950
	Fuller	$P(t_1^* \leq C)$	0.01000	0.03200	0.04800	0.24000	0.48200	0.72200	0.92800	0.97000	0.9940
		$P(t_2^* \leq C)$	0.00600	0.04400	0.07000	0.29000	0.51800	0.78800	0.97400	0.98800	0.9940
		$P(t_3^* \leq C)$	0.02400	0.05000	0.06800	0.24800	0.48600	0.69600	0.92000	0.95200	0.9760
		$P(t_4^* \leq C)$	0.00600	0.01400	0.04400	0.26800	0.47400	0.74600	0.96400	0.98800	0.9960
	Parks	$P(t_1^* \leq C)$	0.22645	0.27054	0.30461	0.39479	0.50501	0.57715	0.69339	0.72545	0.7835
		$P(t_2^* \leq C)$	0.17635	0.24248	0.29058	0.42084	0.52705	0.63327	0.77756	0.81363	0.8677
		$P(t_3^* \leq C)$	0.28657	0.32665	0.35271	0.43086	0.48497	0.52305	0.58918	0.61323	0.6553
irical		$P(t_4^* \leq C)$	0.16834	0.22846	0.27455	0.37074	0.48297	0.59920	0.75150	0.79760	0.8777
tributions	DaSilva	$P(t_1^* \leq C)$	0.01326	0.04463	0.07302	0.27586	0.48479	0.70385	0.90457	0.94320	0.9858
		$P(t_2^* \leq C)$	0.02434	0.05882	0.09128	0.32860	0.53144	0.75862	0.95335	0.97566	0.9878
		$P(t_3^* \leq C)$	0.02637	0.05071	0.07708	0.26775	0.48884	0.68154	0.90467	0.94118	0.9736
		$P(t_4^* \leq C)$	0.33266	0.33671	0.36308	0.43205	0.48682	0.56187	0.62272	0.63489	0.6470
	OLS	$P(t_1^* \leq C)$	0.10600	0.18400	0.21200	0.35800	0.48800	0.60000	0.75000	0.80200	0.8820
		$P(t_2^* \leq C)$	0.12200	0.20000	0.23000	0.40200	0.51600	0.66800	0.84200	0.88000	0.9320
		$P(t_3^* \leq C)$	0.13600	0.20000	0.24000	0.36200	0.48400	0.60600	0.70400	0.76600	0.8420
		$P(t_4^* \leq C)$	0.00000	0.01600	0.02400	0.21800	0.52200	0.78800	0.97400	0.98800	0.9980

Table 3.2 Comparison of nominal asymptotic critical points with standardized empirical distributions generated assuming Type I errors.

Tabular Values		C	-2.57580	-1.96000	-1.64480	-0.67450	0.00000	0.67450	1.64480	1.96000	2.57580
		$P(z \leq C)$	0.00500	0.02500	0.05000	0.25000	0.50000	0.75000	0.95000	0.97500	0.99500
rical	Fuller	$P(t_1^* \leq C)$	0.05466	0.12753	0.17409	0.34413	0.49190	0.66194	0.87854	0.91498	0.95961
		$P(t_2^* \leq C)$	0.04049	0.08705	0.13765	0.32794	0.48785	0.65182	0.82186	0.87652	0.96154
		$P(t_3^* \leq C)$	0.01012	0.04453	0.08300	0.26518	0.49190	0.74494	0.93117	0.94332	0.98785
		$P(t_4^* \leq C)$	0.00202	0.02429	0.04656	0.23684	0.48988	0.75101	0.96964	0.98785	1.00000
	Parks	$P(t_1^* \leq C)$	0.23093	0.28814	0.32415	0.43644	0.50212	0.56568	0.68008	0.71822	0.78178
		$P(t_2^* \leq C)$	0.15678	0.22034	0.24576	0.38559	0.51271	0.59110	0.72669	0.78178	0.84534
		$P(t_3^* \leq C)$	0.19068	0.25212	0.29237	0.42373	0.50636	0.61017	0.72034	0.75000	0.80932
		$P(t_4^* \leq C)$	0.10593	0.16949	0.19915	0.35805	0.47881	0.62076	0.75636	0.79449	0.84958
ributions	DaSilva	$P(t_1^* \leq C)$	0.05677	0.13100	0.19651	0.36681	0.54148	0.69432	0.88210	0.92140	0.96507
		$P(t_2^* \leq C)$	0.00873	0.04804	0.07860	0.30131	0.47598	0.67686	0.88646	0.92140	0.97380
		$P(t_3^* \leq C)$	0.00873	0.05240	0.06987	0.26201	0.48035	0.71179	0.92576	0.95633	0.99127
		$P(t_4^* \leq C)$	0.33188	0.36681	0.37555	0.42358	0.47598	0.50218	0.56332	0.58515	0.59389
	OLS	$P(t_1^* \leq C)$	0.21400	0.26800	0.29000	0.40200	0.49000	0.58800	0.72200	0.77200	0.84400
		$P(t_2^* \leq C)$	0.15000	0.20400	0.25200	0.40000	0.48600	0.59600	0.71600	0.75800	0.81600
		$P(t_3^* \leq C)$	0.12800	0.19400	0.22200	0.38600	0.49400	0.62200	0.79000	0.82400	0.88000
		$P(t_4^* \leq C)$	0.00000	0.00600	0.02800	0.20600	0.50800	0.78400	0.98800	0.99600	1.00000

Table 3.3 Comparison of nominal asymptotic critical points with standardized empirical distributions generated assuming Type II errors.

Tabular Values		C	-2.57580	-1.96000	-1.64480	-0.67450	0.00000	0.67450	1.64480	1.96000	2.5758
		$P(z \leq C)$	0.00500	0.02500	0.05900	0.25000	0.50000	0.75000	0.95000	0.97500	0.9950
irical tributions	Fuller	$P(t_1^* \leq C)$	0.01400	0.05000	0.06400	0.23600	0.48600	0.73600	0.93000	0.96400	0.9880
		$P(t_2^* \leq C)$	0.01200	0.03000	0.05800	0.23600	0.50600	0.73200	0.94000	0.97200	0.9940
		$P(t_3^* \leq C)$	0.01400	0.04200	0.07600	0.29200	0.51600	0.75600	0.93600	0.94800	0.9760
		$P(t_4^* \leq C)$	0.00200	0.01400	0.03800	0.25200	0.52400	0.74600	0.95400	0.98200	0.9960
	Parks	$P(t_1^* \leq C)$	0.19878	0.25152	0.28398	0.39959	0.49493	0.57809	0.68154	0.71602	0.7707
		$P(t_2^* \leq C)$	0.17241	0.22718	0.25963	0.38945	0.51521	0.59838	0.72617	0.76065	0.8357
		$P(t_3^* \leq C)$	0.35294	0.38337	0.39148	0.46653	0.51116	0.55984	0.61460	0.64706	0.6734
		$P(t_4^* \leq C)$	0.13793	0.20892	0.24544	0.38945	0.51521	0.60446	0.75862	0.79513	0.8336
	DaSilva	$P(t_1^* \leq C)$	0.02101	0.04832	0.06513	0.23950	0.47479	0.70168	0.90956	0.94748	0.9873
		$P(t_2^* \leq C)$	0.02521	0.04412	0.07563	0.27521	0.51681	0.72059	0.91176	0.95378	0.9852
		$P(t_3^* \leq C)$	0.01050	0.05672	0.08613	0.30882	0.52521	0.75000	0.93487	0.95378	0.9768
		$P(t_4^* \leq C)$	0.46639	0.47689	0.48319	0.50840	0.52311	0.54832	0.56723	0.57773	0.5882
	OLS	$P(t_1^* \leq C)$	0.12000	0.19400	0.21400	0.37200	0.49000	0.59600	0.75200	0.80400	0.860C
		$P(t_2^* \leq C)$	0.11000	0.16400	0.20000	0.37600	0.50800	0.60800	0.77800	0.82400	0.900C
		$P(t_3^* \leq C)$	0.19200	0.27000	0.30600	0.43200	0.52000	0.59400	0.74600	0.79000	0.848C
		$P(t_4^* \leq C)$	0.00000	0.00400	0.01600	0.19200	0.50800	0.82400	0.98000	0.99400	1.000C

Table 3.4 Comparison of nominal asymptotic critical points with standardized empirical distributions generated assuming Type III errors.

Table 3.5. Error parameter estimates obtained from 450 iterations on correct model.

Error type	Error parameter	Average error parameter estimate	Variance of estimated error parameters
I	σ_v^2	0.47696	0.06976
	σ_e^2	0.50943	0.04807
	σ_ϵ^2	0.50140	0.00386
II	ρ_1	0.49864	0.06914
	ρ_2	0.60593	0.05828
	ρ_3	0.49816	0.00532
	ρ_4	0.62619	0.04450
	ρ_5	0.46667	0.04888
	ρ_6	0.59901	0.09918
	ρ_7	0.46353	0.04780
	ρ_8	0.64978	0.04231
	ρ_9	0.43538	0.05744
	ρ_{10}	0.63381	0.04435
III	σ_A^2	0.39235	0.07473
	σ_B^2	0.39396	0.04006
	$\gamma(0)$	0.50772	0.02849
	$\gamma(1)$	0.30347	0.04362
	$\gamma(2)$	0.20074	0.04593
	$\gamma(3)$	0.12511	0.04620
	$\gamma(4)$	0.06423	0.05559
	$\gamma(5)$	0.04818	0.05505
	$\gamma(6)$	0.00656	0.05678
	$\gamma(7)$	0.01314	0.05500

Interpretation of Simulation Results

As in any simulation study, nothing has been "proved" by the undertaking at hand. We do feel, however, that the following observations are worth making.

1. "Correctness" of computer routines

From all indications, there is no reason to doubt the accuracy of the computer implementation of the models. In every case, the estimates seem unbiased and the scales in the model approximately correct. The only exception to the latter statement is the Da Silva method whose true scale is 1 in model 11 but is estimated on the average of 476 iterations to be 208.456! This we attribute to the use of $\tilde{V}^*$ in lieu of $\tilde{V}$ as described in § 2.3. It is pointed out that with the exception of several minor bugs, no major programming errors were uncovered by the Monte Carlo study.

2. Empirical estimation properties of models

From Table 3.1, it appears that all estimators produce (as they should) unbiased estimators. As far as efficiencies go, only the Parks method (model 6) does substantially better in terms of mean-square-error than the obvious ordinary least-squares competitor. Both Fuller (model 1) and Da Silva (model 11) produce only minor improvements over the OLS procedure (models 4 and 12, respectively). In all fairness, however, all models perform reasonably well on the competing error structures especially in light of the fact that parameter values of magnitude 1 are being picked out of noise having variance 1.5 in the study. This requirement we feel is not at all unreasonable for some types of economic data.

3. Empirical distributional properties of models

Tables 3.2 - 3.4 provide strong empirical evidence that the inferential powers of ordinary least-squares are badly distorted under our error structures. In fact, the only method showing general robustness to the underlying error structure was Fuller-Battese. For this reason alone we could recommend its general use. Another favorable feature of this method is the great economy in core storage realized in its implementation in SAS over its competitors. It is interesting to note that although the Parks method (model 6) seemed preferable to its competitors (i.e. models 5, 7, 8) from a m.s.e. point of view, its inferential credentials are deplorable. Finally, with the exception of parameter β_4, Da Silva exhibits at best "tolerable" robustness in the study.

Perhaps the above results will move some researchers to further study their candidate estimators under competing error structures and ensure that they retain some acceptable robustness in addition to desirable estimation properties.

4. Estimation of underlying error distribution parameters

Most applied statisticians are interested in the estimation of the error distribution parameters only insofar as they affect the statistical properties of the resulting parameter estimates. For completeness in the study, however, we find it interesting that the Da Silva method had so much trouble obtaining range-preserving estimates for its error parameters under model 7. In addition, notice in Table 3.5 that the error parameters are consistently underestimated for both the Type II and III distributions. This undoubtedly is the cause for the apparent flatness in the empirical distribution of the Parks method throughout the study.

IV. Bibliography

Amemiya, T. (1971). The estimation of the variances in a variance-components model. International Economic Review, 12:1-13.

Amemiya, T. and W. A. Fuller. (1967). A comparative study of alternative estimators in a distributed lag model. Econometrica, 35:509-529.

Anderson, T. W. (1969). Statistical inference for covariance matrices with linear structure. In P. R. Krishnaiah (Editor), Multivariate Analysis-II, 55-66. Academic Press, Inc., New York, New York.

Anderson, T. W. (1970). Estimation of covariance matrices which are linear combinations of whose inverses are linear combinations of given matrices. In R. C. Bose, I. M. Chakravarti, P. C. Mahalanobis, C. R. Rao, and K. J. C. Smith (Editors), Essays in Probability and Statistics, 1-24. The University of North Carolina Press, Chapel Hill, N. C.

Anderson, T. W. (1971). The Statistical Analysis of Time Series. John Wiley and Sons, Inc., New York, New York.

Anderson, T. W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. The Annals of Statistics, 1:135-141.

Balestra, P. and M. Nerlove. (1966). Pooling cross section and time series data in the estimation of a dynamic model: The demand for natural gas. Econometrica, 34:585-612.

Bellman, R. (1960). Introduction to Matrix Analysis. McGraw-Hill Book Company, New York, New York.

Blair, R. D. and J. Kraft. (1974). Estimation of elasticity of substitution in American manufacturing industry from pooled cross section and time series observations. The Review of Economics and Statistics, 56:343-347.

Brehm, C. T. and T. R. Saving. (1964). The demand for general assistance payments. The American Economic Review, 54:1002-1018.

Cochrane, D. and G. H. Orcutt. (1949). Application of least squares regression to relationships containing autocorrelated error terms. Journal of the American Statistical Association, 44:32-61.

Cramer, H. (1946). Mathematical Methods of Statistics. Princeton University Press, Princeton, New Jersey.

DeBord, D. V. (1974). Demand for and cost of salt marsh mosquito abatement. Ph.D. Dissertation. Department of Economics, North Carolina State University, Raleigh, N. C. University Microfilms, Ann Arbor, Michigan.

Dhrymes, P. J. (1971). Distributed Lags: Problems of Estimation and Formulation. Holden-Day, Inc., San Francisco, California.

Duncan, D. B. and R. H. Jones. (1966). Multiple regression with stationary errors. Journal of the American Statistical Association, 61:917-928.

Engle, R. F. (1974). Band spectrum analysis. International Economic Review, 15:1-11.

Fishman, G. S. (1969). Spectral Methods in Econometrics. Harvard University Press, Cambridge, Massachusetts.

Fuller, W. A. (1976). Introduction to Statistical Time Series. John Wiley and Sons, Inc., New York City, New York.

Fuller, W. A. and G. E. Battese. (1974). Estimation of linear models with crossed-error structure. Journal of Econometrics, 2:67-78.

Graybill, F. A. (1969). Introduction to Matrices With Applications in Statistics. Wadsworth Publishing Company, Inc., Belmont, California.

Graybill, F. A. and R. A. Hultquist. (1961). Theorems concerning Eisenhart's Model II. The Annals of Mathematical Statistics, 32:261-269.

Grenander, U. and G. Szego. (1958). Toeplitz Forms and Their Applications. University of California Press, Berkeley, California.

Hamon, B. V. and E. J. Hannan. (1963). Estimating relations between time series. Journal of Geophysical Research, 68:6033-6041.

Hannan, E. J. (1960). Time Series Analysis. Methuen and Company, Ltd., London, England.

Hannan, E. J. (1963). Regression For Time Series. In M. Rosenblatt (Editor), Proceedings of the Symposium on Time Series Analysis, 17-37. John Wiley and Sons, Inc., New York City, New York.

Hannan, E. J. (1965). The estimation of relationships involving distributed lags. Econometrica, 33:206-224.

Hannan, E. J. (1967). The estimation of lagged regression relation. Biometrika, 54:409-418.

Henderson, C. R. (1953). Estimation of variance and covariance components. Biometrics, 9:226-252.

Henderson, C. R. (1971). Comment on "The Use of Error Components Models in Combining Cross Section with Time Series Data". Econometrica, 39:397-401.

Hildreth, C. (1949). Preliminary considerations regarding time series and/or cross section studies. Cowles Commission Discussion Paper, Statistics No. 333. University of Chicago, Chicago, Illinois (Unpublished).

Hildreth, C. (1950). Combining cross section data and time series data. Cowles Commission Discussion Paper, No. 347. University of Chicago,. Chicago, Illinois (Unpublished).

Hoch, I. (1962). Estimation of production function parameters combining time series and cross section data. Econometrica, 30:34-53.

Hsu, P. L. (1938). On the best unbiased quadratic estimate of the variance. In J. Neyman and E. S. Pearson (Editors), Statistical Research Memoirs, 2:91-104. University of London, London, England.

Hussain, A. (1969). A mixed model for regressions. Biometrika, 56:327-336.

Johnson, P. R. (1960). Land substitutes and changes in corn yields. Journal of Farm Economics, 42:294-306.

Johnson, P. R. (1964). Some aspects of estimating statistical cost functions. Journal of Farm Economics, 46:179-187.

Kakwani, N. C. (1967). The unbiasedness of Zellner's seemingly unrelated regression equations estimators. Journal of the American Statistical Association, 62:141-142.

Kmenta, J. (1971). Elements of Econometrics. The Macmillan Company, New York City, New York.

Kmenta, J. and R. F. Gilbert. (1968). Small sample properties of alternative estimators of seemingly unrelated regressions. Journal of the American Statistical Association, 63:1180-1200.

Kmenta, J. and R. F. Gilbert. (1970). Estimation of seemingly unrelated regressions with autoregressive disturbances. Journal of the American Statistical Association, 65:186-197.

Kuh, E. (1959). The validity of cross-sectionally estimated behavior equations in time series applications. Econometrica, 27:197-214.

Maddala, G. S. (1971). The use of variance components models in pooling cross section and time series data. Econometrica, 39:341-358.

Mann, H. B. and A. Wald. (1943). On stochastic limit and order relationships. Annals of Mathematical Statistics, 14:217-226.

Marschak, J. (1939). On combining market and budget data in demand studies: a suggestion. Econometrica, 7:332-335.

Marschak, J. (1943). Money illusion and demand analysis. The Review of Economic Statistics, 25:40-48.

Mitra, S. K. and C. R. Rao. (1968). Some results in estimation and tests of linear hypothesis under the Gauss-Markoff model. Sankhya, Series A, 30:281-290.

Mundlak, Y. (1961). Empirical production function free of management bias. Journal of Farm Economics, 43:44-56.

Mundlak, Y. (1963). Estimation of production and behavioral functions from a combination of cross-section and time-series data. In C. F. Christ et al. (Editors), Measurement In Economics, 138-166. Stanford University Press, Stanford, California.

Nerlove, M. (1967). Experimental evidence on the estimation of dynamic economic relations from a time series of cross sections. Economic Studies Quarterly, 18:42-74.

Nerlove, M. (1971a). Further evidence on the estimation of dynamic economic relations from a time series of cross sections. Econometrica, 39:359-382.

Nerlove, M. (1971b). A note on error components models. Econometrica, 39:383-396.

Parks, R. W. (1967). Efficient estimation of a system of regression equations when distrubances are both serially and contemporaneously correlated. Journal of the American Statistical Association, 62:500-509.

Rao, C. R. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. In L. M. Le Cam and J. Neyman (Editors), Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1:355-372. University of California Press, Berkeley, California.

Rao, C. R. (1970). Estimation of heteroscedastic variances in linear models. Journal of The American Statistical Association, 65:161-172.

Rao, C. R. (1971a). Estimation of variance and covariance components - MINQUE theory. Journal of Multivariate Analysis, 1:257-275.

Rao, C. R. (1971b). Minimum variance quadratic unbiased estimation of variance components. Journal of Multivariate Analysis, 1:445-456.

Rao, C. R. (1972). Estimation of variance and covariance components in linear models. Journal of the American Statistical Association, 67:112-115.

Rao, C. R. (1973). Linear Statistical Inference and Its Applications. Second Edition. John Wiley and Sons, Inc., New York City, New York.

Schultz, H. (1938). The Theory and Measurement of Demand. The University of Chicago Press, Chicago, Illinois.

Searle, S. R. (1971a). Linear Models. John Wiley and Sons, Inc. New York City, New York.

Searle, S. R. (1971b). Topics in variance component estimation. Biometrics, 27:1-76.

Seely, J. (1969). Estimation in finite-dimensional vector spaces with application to the mixed linear model. Ph.D. Dissertation. Department of Statistics, Iowa State University, Ames, Iowa. University Microfilms, Ann Arbor, Michigan.

Seely, J. (1970a). Linear spaces and unbiased estimation. The Annals of Mathematical Statistics, 41:1725-1734.

Seely, J. (1970b). Linear spaces and unbiased estimation - application to the mixed linear model. The Annals of Mathematical Statistics, 41:1735-1748.

Seely, J. and S. Soong. (1971). A note on MINQUE's and quadratic estimability. Mimeographed report. Oregon State University, Corvallis, Oregon.

Seely, J. And G. Zyskind. (1971). Linear spaces and minimum variance unbiased estimation. The Annals of Mathematical Statistics, 42:691-703.

Sims, G. A. (1971). Are there exogeneous variables in short-run production relations? Center of Economic Research, University of Minnesota, Minneapolis, Minnesota. Discussion Paper No. 10.

Stone, R. (1954). The Measurement of Consumer's Expenditure and Behavior in the United Kingdom, 1920-1938, Vol. 1. University Press, Cambridge, England.

Swamy, P. A. V. B. and S. S. Arora. (1972). The exact finite sample properties of the estimators of coefficients in the error components regression models. Econometrica, 40:261-275.

Theil, H. (1971). Principles of Econometrics. John Wiley and Sons, Inc., New York City, New York.

Tobin, J. (1950). A statistical demand function for food in the U.S.A. Journal of the Royal Statistical Society, Series A, 113:113-141.

Wallace, T. D. and A. Hussain. (1969). The use of error components models in combining cross section with time series data. Econometrica, 37:55-72.

Wold, H. and L. Jureen. (1953). Demand Analysis - A study in Econometrics. John Wiley and Sons, Inc., New York City, New York.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. Journal of the Ameriacn Statistical Association, 57:348-368.

Zellner, A. (1963). Estimators of seemingly unrelated regressions: Some exact finite sample results. Journal of the American Statistical Association, 58:977-992.

Zellner, A. and D. S. Huang. (1962). Further properties of efficient estimators for seemingly unrelated regression equations. International Economic Review, 3:300-313.

Chapter I

Procedure Documentation