14097
NATIONAL LIBRARY
OTTAWA

BIBLIOTHÈQUE NATIONALE
OTTAWA

NAME OF AUTHOR.....Rajendra Nath Pandya.....
TITLE OF THESIS....."A Class of Bootstrap Estimators for.....
.....Identification of Linear Discrete-Time..
.....Models"
UNIVERSITY.....Carleton University.....
DEGREE FOR WHICH THESIS WAS PRESENTED.....Ph.D.....
YEAR THIS DEGREE GRANTED.....1972.....

Permission is hereby granted to THE NATIONAL LIBRARY
OF CANADA to microfilm this thesis and to lend or sell copies
of the film.

The author reserves other publication rights, and
neither the thesis nor extensive extracts from it may be
printed or otherwise reproduced without the author's
written permission.

(Signed).....*R. Pandya*.....

PERMANENT ADDRESS:

Deptt. of Communication Engrg
Govt. Engineering College,
.....JABALPUR.....(INDIA)

DATED...*June 7*.....1972

A CLASS OF BOOTSTRAP ESTIMATORS FOR IDENTIFICATION
OF LINEAR DISCRETE-TIME MODELS

by

Rajendra Nath Pandya, M.A. Sc.

A thesis submitted to the Faculty of
Graduate Studies in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy

Faculty of Engineering

Carleton University

Ottawa, Ontario

1 May 1972

ii

The undersigned recommend to the Faculty
of Graduate Studies acceptance of the thesis

*A Class of Bootstrap Estimators for
Identification of Linear Discrete-
Time Models*

submitted by Rajendra Nath Pandya, M.A.Sc.,
in partial fulfilment of the requirements for
the degree of Doctor of Philosophy

.....*Bernard Pagurek*.....
Thesis Supervisor

.....*[Signature]*.....
Dean, Faculty of Engineering

Carleton University

June 6, 1972

ACKNOWLEDGEMENTS

The author wishes to express his sincere thanks to Dr. B. Pagurek of the Division of Systems Engineering for his supervision, constant encouragement and many long and fruitful discussions during the course of this work.

The financial assistance from Carleton University, the National Research Council of Canada and the Government of Ontario is also gratefully acknowledged.

The author especially wishes to express his appreciation for the patient encouragement provided by his wife, Margaret.

ABSTRACT

In a recent survey paper Astrom³ classifies the on-line or recursive estimators for system identification as a 'fiddlers paradise'. His comment is perhaps motivated by the absence of a unified approach to such estimators.

Notwithstanding the above comment, recursive estimators offer the distinct advantage of simplicity in programming and computational efficiency with regard to computer time and storage. Further, if system identification is an integral part of an adaptive control strategy and/or the system parameters have a tendency to drift, there is no other option but to use on-line algorithms for parameter estimation.

Besides presenting a comprehensive study of existing recursive estimators, a class of recursive estimators, viz., Bootstrap Estimators is proposed in this thesis. In general, the bootstrap estimators represent recursive approximations to the Two Stage Least Square (2SLS) estimator (or its more general forms the k-class, the h-class and double k-class estimators) that is so frequently used in the identification of econometric models.

For a single output, discrete-time, linear dynamic process subject to process and output disturbances, the necessary arguments for deriving the Levy's proper canonical model and the equivalent difference equation model are presented.

7

It is shown that consistent estimates for the process parameters in the above model can be obtained by using either the concept of 'transformation of the data' or 'transformation of the parameters'. The former leads to the Generalized Least Square (GLS) estimator and the latter to the Two Stage Least Square (2SLS) estimator. A simplified Two Stage Least Square (S2SLS) estimator is proposed which is computationally more efficient than, and asymptotically equivalent to the 2SLS estimator, provided certain constraints on the input to the process are satisfied.

The concept behind the Instrumental Variable (IV) estimators for estimation of process parameters is discussed in detail from an algebraic as well as geometric viewpoint. It is further shown that a variety of consistent estimators, including the 2SLS and the Yule-Walker estimators, fall within the class of IV estimators. A recursive approximation to the 2SLS (bootstrap) estimator is proposed which is very simple to program and is computationally efficient.

By using the concept of 'structural' and 'incidental' parameters, the definition of 'Bootstrap Estimators' is formalized. Though it is difficult to provide a rigorous proof of stability for such algorithms, it is shown that the necessary (though not sufficient) conditions for stability can be achieved by incorporating some simple stabilizing filters in the algorithm. It is shown that in general, bootstrap estimators represent recursive approximations to the 2SLS estimator or its more general forms the k-class, h-class and the double k-class estimators.

7

Recursive estimators (within the class of bootstrap estimators) for the estimation of noise parameters when process parameters are known, and for the estimation of process as well as noise parameters simultaneously in a single experiment are proposed.

The class of Bootstrap estimators proposed in this thesis are used to estimate the model parameters for a Glucose Tolerance Test, conducted at the Faculty of Medicine, University of Ottawa. The results are compared with those obtained by using MLE and other existing on line estimators.

Brief comments on such related problems as identification in the presence of closed-loop control, tracking of time-varying parameters, identification of multivariable systems and a priori estimation of process order, are also included.

LIST OF SYMBOLS

The symbols that have been used throughout the text are listed here. The symbols whose use is restricted to a specific section or part thereof are defined in the appropriate section. The numbers in the parenthesis refer to the equation in the text where the symbol occurs for the first time.

$(\bar{\cdot})$	Sample mean of $(\cdot)$
$(\cdot')$	Matrix transpose
$(\hat{\cdot}), (\tilde{\cdot})$	Estimate of $(\cdot)$
k	Discrete time index
n	System order
N	Dimension of the measurement vector
p	Dimension of the vector of process parameters
s	Dimension of the vector of reduced-form parameters
$u(k)$	Deterministic input to the process (2.1)
$x(k)$	Noise-free output (scalar) of the process (2.25)
$y(k)$	Noise corrupted system output (2.1)
F, g, a, h, d	System and output matrices for Levy's Canonical form (2.13)
$\underline{x}(k)$	Vector of state variables in Levy's canonical form (2.13)
$e(k)$	Sequence of innovations in Levy's canonical form (2.13)
λ^2	Variance of $e(k)$ (2.14)
a_i, b_i, c_i	Parameters in the difference equation model (2.15)

$w(k)$	Additive errors $[y(k) - x(k)]$ in the difference equation model (2.26)
$\underline{y}$	Observation vector (3.4)
M	Design matrix for the structural model (3.4)
$\underline{\beta}$	Vector of process parameters
$\underline{v}$	Disturbance vector in the structural model (3.4)
M_1	Partition of M containing inputs $u(k)$ (3.5)
M_2	Partition of M containing outputs $y(k)$ (3.5)
$\underline{\beta}_1$	Parameter vector associated with M_1 (3.5)
$\underline{\beta}_2$	Parameter vector associated with M_2 (3.5)
$v(k)$	Sequence of disturbances associated with the structural model (3.13)
$\underline{m}(k)$	Design vector at time k (3.12)
$\underline{u}_i$	Column vectors of M_1
$\underline{y}_i$	Column vectors of M_2
Ω	Variance-covariance matrix of $\underline{v}$
U	Design matrix in the reduced form model (3.33)
$\underline{\gamma}$	Vector of reduced form parameters (3.33)
$\underline{w}$	Disturbance vector in the reduced form model (3.33)
Z	General Instrumental Variable matrix (4.4)
$z(k)$	General sequence of instrumental variables (4.13)
$\underline{z}(k)$	Vector of instruments corresponding to the design vector $\underline{m}(k)$ (4.14)
X	Matrix corresponding to M in which $y(k)$ are replaced by $x(k)$ (5.11)
$\underline{\delta}$	Vector of noise parameters in the difference equation model (6.26)
$\underline{\theta}$	Vector of process and noise parameters (6.41)

LIST OF ABBREVIATIONS

BIBO	Bounded Input Bounded Output
OLS	Ordinary Least Square
GLS	Generalized Least Square
ILS	Indirect Least Square
2SLS	Two Stage Least Square
S2SLS	Simplified Two Stage Least Square
3SLS	Three Stage Least Square
IV	Instrumental Variable
MLE	Maximum Likelihood Estimator
MA	Moving Average (process)
AR	Auto Regressive (process)
MARMA	Mixed Auto Regressive Moving Average (process)
Var-Covar	Variance-Covariance (matrix)
G. Var	Generalized Variance
N.I.D.	Normally and Independently Distributed
N.D.	Normally Distributed
PRBS	Pseudo Random Binary Sequence
i.i.d.	Independently and Identically Distributed
p.d.f.	Probability Density Function

TABLE OF CONTENTS

	PAGE
CHAPTER 1 - INTRODUCTION	1
1.1 Introduction to the Problem	1
1.2 Contents and Organization of the Thesis	3
CHAPTER 2 - MODELLING OF LINEAR DISCRETE-TIME SYSTEMS	8
2.1 Basic State Variable Model	8
2.2 A Canonic Form with a Minimum Parameter Set	13
2.3 Levy's Proper Canonical Representation	16
2.4 Astrom's Difference Equation Model	18
2.5 Some Special Cases	22
CHAPTER 3 - SOME IDENTIFICATION TECHNIQUES FOR STRUCTURAL MODELS	26
3.1 Preliminary Remarks	26
3.2 Least Squares Estimates of Process Parameters	28
3.3 Transformation of Data - Generalized Least Squares Estimator	33
3.4 Transformation of Parameters - Reduced Form Model	36
3.5 Results	48
3.6 Concluding Remarks	53
CHAPTER 4 - INSTRUMENTAL VARIABLE ESTIMATORS	62
4.1 Introduction to the Instrumental Variable Technique	62
4.2 Choice of Instrumental Variables	67
4.3 Some Practical IV Estimators	76
4.4 Identification of Force-free Models	88
4.5 Results	96
4.6 Concluding Remarks	99
CHAPTER 5 - IDENTIFICATION OF LINEAR DYNAMIC MODELS WITH ADDITIVE OUTPUT NOISE - BOOTSTRAP ESTIMATORS	127
5.1 Introduction	127
5.2 The Model and the 'Idealized' Estimators for its Parameters	130
5.3 Practical Approximations to $\hat{\beta}^1, \hat{\beta}^2, \hat{\beta}^3$ - Bootstrap Estimators	135
5.4 Relationship of Bootstrap Estimators to General k-class, h-class and double k-class Estimators	153

TABLE OF CONTENTS (continued)

5.5 Results	160
5.6 Concluding Remarks	163
CHAPTER 6 - ESTIMATION OF PARAMETERS OF THE ENTIRE MODEL	179
6.1 Preliminary Remarks	179
6.2 Estimation of Noise Parameters, Given Consistent Estimates of the Process Parameters	180
6.3 Simultaneous Estimation of Process and Noise Parameters in a Single Experiment - Bootstrap Estimators	198
6.4 Identification of Glucose Tolerance Test Model - Application to Experimental Data	216
6.5 Concluding Remarks	223
CHAPTER 7 - SOME RELATED PROBLEMS AND FINAL CONCLUSIONS	228
7.1 Some Related Problems	228
7.2 Final Conclusions	233
REFERENCES	236
APPENDIX A	240
A.1 Preliminaries	240
A.2 Properties of Estimators for a Given Fixed N	242
A.3 Properties of Estimators for $N \rightarrow \infty$	245
A.4 Some Relevant Theorems	250
APPENDIX B	256
B.1 Properties of Ω , the Noise Covariance Matrix	256
B.2 Asymptotic Distribution of the IV Estimator	259
APPENDIX C	264
C.1 Generation of N.I.D. Random Variables	264
C.2 Evaluation of Subroutine WNOISE	266

LIST OF ILLUSTRATIONS

FIGURE		PAGE
2.1	Basic State Space Model	12
2.2	Block Diagram Representation of the Difference Equation Model	21
3.1	Convergence of $(\text{Bias})^2$ for OLS, ILS, S2SLS & 2SLS Estimators - Model (A) Second Order	58
3.2	Convergence of MSE for OLS, ILS, S2SLS & 2SLS Estimators - Model (A) Second Order	59
3.3	Convergence of $(\text{Bias})^2$ for OLS, ILS, and S2SLS Estimators - Model (B) Fourth Order	60
3.4	Convergence of MSE for OLS, ILS & S2SLS Estimators - Model (B) Fourth Order	61
4.1(a)	Geometry of OLS Estimator $E(M'y) \equiv 0$	65
4.1(b)	Geometry of the IV & OLS Estimator $E(M'y) > 0$	65
4.2	Recursive IV Estimator in which the IV Sequence is Generated by an Auxiliary Model	83
4.3	Detailed Structure for the Recursive Approximation to 2SLS Using an Auxiliary Model	86
4.4	$(\text{Bias})^2$ Convergence of IV Estimators Based on Lagged Outputs as Instruments - Model (A) Second Order	106
4.5	MSE Convergence of IV Estimators Based on Lagged Outputs as Instruments - Model (A) Second Order	107
4.6	$(\text{Bias})^2$ Convergence of On-line IV Estimators that Use an Auxiliary Model to Generate the Instruments Recursively - Model (A) Second Order	110
4.7	MSE Convergence of On-line IV Estimators that Use an Auxiliary Model to Generate the Instruments Recursively - Model (A) Second Order	111
4.8	$(\text{Bias})^2$ Convergence of On-line IV Estimators that Use an Auxiliary Model to Generate the Instruments Recursively - Model (B) Fourth Order	112

LIST OF ILLUSTRATIONS (continued)

FIGURE		PAGE
4.9	MSE Convergence of On-line Estimators that Use an Auxiliary Model to Generate the Instruments Recursively - Model (B) Fourth Order	113
4.10	(Bias) ² Convergence for IV Estimators that Use (a) Lagged Inputs; (b) Lagged Outputs and (c) Recursive Estimates of Noise-free Outputs as Instruments Model (A) Second Order	116
4.11	MSE Convergence for IV Estimators that Use (a) Lagged Inputs; (b) Lagged Outputs and (c) Recursive Estimates of Noise-free Outputs as Instruments - Model (A) Second Order	117
4.12	(Bias) ² Convergence for IV Estimators that Use (a) Lagged Inputs; (b) Lagged Outputs and (c) Recursive Estimates of Noise-free Outputs as Instruments - Model (B) Fourth Order	118
4.13	MSE Convergence for IV Estimators that Use (a) Lagged Inputs; (b) Lagged Outputs and (c) Recursive Estimates of Noise-free Outputs as Instruments - Model (B) Fourth Order	119
4.14	MSE Convergence of (a) $\hat{\beta}_{(IV)}$ -Pandya (b) S2SLS and (c) $\hat{\beta}_{(IV)}^*$ that Uses True Process Outputs - Model (A) Second Order	122
4.15	MSE Convergence of (a) $\hat{\beta}_{(IV)}$ -Pandya (b) S2SLS and (c) an IV Estimator that Uses the True Process Output as Instruments - Model (B) Fourth Order	123
4.16	Convergence of (Bias) ² for OLS, Yule-Walker(1) and Yule-Walker(2) Estimates - Model (C) Fourth Order Force-free	125
4.17	Convergence of MSE for OLS, Yule-Walker(1) and Yule-Walker(2) Estimators - Model (C) Fourth Order Force-free	126
5.1	Process with Output Additive Disturbance	128
5.2	Basic Structure of a Bootstrap Estimator	138

LIST OF ILLUSTRATIONS (continued)

FIGURE		PAGE
5.3	Some Typical Gain Sequences for Filter (A)	147
5.4	Progressively Refined Bound on $\hat{\lambda}^2(k)$ for Adaptive Filter (B)	151
5.5	Detailed Structure of a Bootstrap Estimator with a Suitable Filter	154
5.6	Effect of Adaptive Filter on $\hat{\lambda}^2(k)$ - Bootstrap Estimator $\hat{\beta}^1$ Applied to Second Order Model	173
5.7	Effect of Adaptive Filter on $\hat{\lambda}^2(k)$ - Bootstrap Estimator $\hat{\beta}^2$ Applied to Second Order Model	174
5.8	Effect of Adaptive Filter on $\hat{\lambda}^2(k)$ - Bootstrap Estimator $\hat{\beta}^3$ Applied To Second Order Model	175
5.9	Effect of Adaptive Filter on $\hat{\lambda}^2(k)$ - Bootstrap Estimator $\hat{\beta}^1$ Applied to Fourth Order Model	176
5.10	Effect of Adaptive Filter on $\hat{\lambda}^2(k)$ - Bootstrap Estimator $\hat{\beta}^2$ Applied to Fourth Order Model	177
5.11	Effect of Adaptive Filter on $\hat{\lambda}^2(k)$ - Bootstrap Estimator $\hat{\beta}^3$ Applied to Fourth Order Model	178
6.1	Bootstrap Estimator for Estimation of Noise Parameters $\hat{\delta}$	192
6.2	Flow Diagram for Implementing 2SLS-(1), 2SLS-(2) and the Bootstrap Estimators for Noise-Parameters	195
6.3	Convergence of 2SLS-(1), 2SLS-(2) and the Bootstrap Estimator for Noise Parameters - II Order Model - Astrom	197
6.4	Adaptive Recursive Least Squares Algorithm - Panuska	209
6.5	Block Diagram Representation of Recursive Estimators for the System Parameters Based on Instrumental Variables	212
6.6	Implementation of Recursive Estimator for System Parameters Based on Instrumental Variables	213

LIST OF ILLUSTRATIONS (continued)

FIGURE		PAGE
6.7	Convergence of Recursive Algorithm-Panuska and Recursive Algorithm-Pandya Second Order Model (A)	218
6.8	Convergence of Recursive Algorithm-Panuska and Recursive Algorithm-Pandya. Second Order Model (B) ...	219
6.9	Noise-free Output - Bootstrap Estimator	224
6.10	Noise-free Output - MLE	225
6.11	Noise-free Output - LSE	226
C.1	Average Probability Distribution of Output Samples from WNOISE	272

LIST OF TABLES

TABLE		PAGE
3.1	Comparison of Estimates Obtained from OLS, ILS, S2SLS and 2SLS Estimators - Model (A) Second Order ..	55
3.2	Comparison of Estimates Obtained from OLS, ILS and S2SLS Estimators - Model (B) Fourth Order	56
3.3	Comparison of Estimates Obtained from 2SLS and S2SLS Estimators - Model (A), Second Order	57
4.1	Comparison of IV Estimators Based on Lagged Outputs as Instruments	103
4.2	Comparison of IV Estimators Based on Lagged Outputs as Instruments - Model (B) Fourth Order	104
4.3	(Bias) ² and MSE Convergence of IV Estimators Based on Lagged Outputs as Instruments - Model (B) Fourth Order	105
4.4	Comparison of On-line IV Estimators that Use an Auxiliary Model to Generate the Instruments Recursively - Model (A) Second Order	108
4.5	Comparison of On-Line IV Estimators that Use an Auxiliary Model to Generate the Instruments Recursively - Model (B) Fourth Order	109
4.6	Comparison of IV Estimators that Use (a) Lagged Inputs; (b) Lagged Outputs and (c) Recursive Estimates of Noise-free Outputs as Instruments - Model (A) Second Order	114
4.7	Comparison of IV Estimators that Use (a) Lagged Inputs; (b) Lagged Outputs and (c) Recursive Estimates of the Noise-free Outputs and Instruments - Model (B) Fourth Model	115
4.8	Comparison of (a) $\hat{\beta}_{(IV)}$ -Pandya, (b) S2SLS and (c) $\hat{\beta}_{(IV)}^*$ that Uses True Process Output as Instruments - Model (A) Second Order	120
4.9	Comparison of (a) $\hat{\beta}_{(IV)}$ -Pandya, (b) S2SLS and (c) $\hat{\beta}_{(IV)}^*$ that Uses True Process Output as Instruments - Model (B) Fourth Order	121

LIST OF TABLES (continued)

TABLE		PAGE
4.9	Comparison of (a) $\hat{\beta}_{(IV)}$ -Pandya, (b) S2SLS and (c) $\hat{\beta}_{(IV)}^*$ that Uses True Process Output as Instruments - Model (B) Fourth Order	121
4.10	Comparison of OLS with Estimates for Use in Force-free Models - Model (C) Fourth Order (force-free)	124
5.1	Comparison of Idealized Estimators $\tilde{\beta}^1$, $\tilde{\beta}^2$ and $\tilde{\beta}^3$ - Second Order Model	165
5.2	Comparison of Bootstrap Estimators $\hat{\beta}^1$, $\hat{\beta}^2$ and $\hat{\beta}^3$ - Second Order Model and No Filter	166
5.3	Comparison of Bootstrap Estimators $\hat{\beta}^1$, $\hat{\beta}^2$ and $\hat{\beta}^3$ - Second Order Model and Non-Adaptive Filter	167
5.4	Comparison of Bootstrap Estimators $\hat{\beta}^1$, $\hat{\beta}^2$ and $\hat{\beta}^3$ - Second Order Model and Adaptive Filter	168
5.5	Comparison of Idealized Estimator $\tilde{\beta}^1$, $\tilde{\beta}^2$ and $\tilde{\beta}^3$ - Fourth Order Model	169
5.6	Comparison of Bootstrap Estimators $\hat{\beta}^1$, $\hat{\beta}^2$ and $\hat{\beta}^3$ - Fourth Order Model and No Filter	170
5.7	Comparison of Bootstrap Estimators $\hat{\beta}^1$, $\hat{\beta}^2$, and $\hat{\beta}^3$ - Fourth Order Model with Non-adaptive Filter	171
5.8	Comparison of Bootstrap Estimators $\hat{\beta}^1$, $\hat{\beta}^2$, and $\hat{\beta}^3$ - Fourth Order Model and Adaptive Filter	172
6.1	Estimates of Noise Parameters Obtained from (i) 2SLS-(1), (ii) 2SLS-(2) and (iii) Bootstrap Estimators	196
6.2	Comparison of Estimates Obtained from Recursive Algorithm- Panuska and Recursive Algorithm-Pandya - Second Order Model-(A)	216
6.3	Comparison of Estimates Obtained from Recursive Algorithm-Panuska and Recursive Algorithm-Pandya - Second Order Model-(B)	217
6.4	Parameter Estimates for the Glucose Tolerance Test	222
C.1	Sample Mean and Variance for Output of WNOISE	268
C.2	Average Serial Correlations for Output of WNOISE	269
C.3	Average Amplitude Distribution (Percent) for Output of WNOISE	270
C.4	Coefficients of Skewness and Kurtosis for Output of WNOISE	273

CHAPTER 1

INTRODUCTION

1.1 Introduction to the Problem

Any design procedure for a control strategy presupposes a known 'mathematical model' for the process to be controlled. Hence 'process modelling' and 'process identification' must precede such a design. The aim of process modelling is to provide a suitable mathematical structure to the given process and the purpose of process identification is to estimate the numerical coefficients associated with such a structure.

For dynamic processes, this structure is provided by a set of differential or difference equations. Together with a set of suitable coefficients, these equations constitute the 'mathematical model' for the process and this model is supposed to represent the process behavior in as accurate and as simple a manner as possible.

In the classical approach, linear control system design generally evolves around a Transfer Function model of the process. For a non-linear process, a hybrid model is usually employed, in which the principal nonlinearity is represented either graphically or analytically in the time domain and the rest of the process components by their transfer functions. In either case, the composite model is based on measurements carried out on smaller viable units of the total process.

Given an accurate mathematical model, recent advances in control systems theory and computer technology often provide the necessary theore-

tical and computational frame-work for controlling large and/or intricate processes with any specified degree of accuracy. However, the accuracy of control cannot exceed the accuracy of the process model, so that the increased level of sophistication in the control system design techniques demands a matching improvement in the techniques of process modelling and identification.

The intricate nature of the process implies that the mathematical model can no more be based on measurements made on smaller component units of the total process. Further, if greater accuracy in the process model is to be achieved, one cannot disregard the interaction between the physical process and the stochastic environment in which it usually operates. As an example, a physiological process is much too complicated to be considered as a simple inter-connection of small component units with measurable input-output relationships. Moreover, any measurements made on the process are not exactly reproducible due to the random effects that are present within and without the process.

It is therefore apparent that one or more of the following factors may have to be taken into account while modelling a given process:

1. The mathematical model may have to be based on the input-output measurements on the total process.
2. The input-output measurements may be contaminated with noise from the measuring transducers.
3. The output being measured (in the absence of transducer noise) may itself be random due to the disturbance genera-

ted within the process or those introduced at the input. This implies that a suitable filter may have to be used for estimating the true process output, if accuracy of control is to be maintained.

4. Since the design of a suitable filter (e.g., a Weiner-Kalman filter) for state estimation assumes a knowledge of the noise covariance structure, the mathematical model should not only include the process dynamics but also the dynamics of the noise processes involved and the way they interact with the physical process.

Thus, the successive steps involved in the design of a control strategy for an unknown physical process embedded in a stochastic environment are:

1. System modelling
2. System identification
3. State estimation
- and 4. Controller design.

1.2 Contents and Organization of the Thesis

The major effort in this thesis is directed towards the problem of System Identification in the presence of disturbances. Attention is further restricted to the identification of stable, linear, single output, discrete-time models with stationary parameters in the presence of process and output measurement noise. The techniques for parameter estimation for this class of models are rooted in the extensive literature on

'econometrics', 'time series analysis' and 'the theory of the general linear hypothesis'. The initial part of the thesis is devoted to the interpretation and adaptation of the relevant models and methods in the above mentioned disciplines for use in the frame-work of systems theory. One of the recursive estimators proposed in this section is then used to develop and formalize the concept of 'Bootstrap Estimators' which in general represent recursive approximations to Two Stage Least Square (2SLS) technique in econometrics.

A complete survey of existing literature on System Identification has not been included in this thesis - mainly because some excellent survey papers^{1,2,3,4} are available. The papers are not only relatively recent but the last two papers contain extensive bibliographies on the subject.

As far as possible, the notation used in the thesis conforms to the vector-matrix notation currently employed in econometrics and Kalman filtering theory. Results have been placed at the end of relevant chapters. For experiments conducted on simulated data, 20 different noise sequences were used. Necessary statistical tests on the noise sequences were conducted so as to ensure that the assumptions associated with the model under study were satisfied.*

Including this introductory chapter, the thesis contains 7 chapters. Chapter 2 is an exposition on the problems associated with the modelling of an unknown process operating in a stochastic environment. Using a set of qualitative constraints, it is shown how the Levy's Proper

* See Appendix C for details.

Canonical Representation and the equivalent Difference Equation model may be derived from the basic State-space model for the class of systems under study. Author's contribution in this chapter does not extend beyond a systematic arrangement of material from scattered sources so that the underlying arguments behind the popularity of these models are exposed.

Chapter 3 starts by demonstrating the shortcomings of the Ordinary Least Squares (OLS) technique as applied to the parametric models developed in Chapter 2. It is shown that consistent estimates of the process parameters can be obtained by using the concept of either 'transformation of data' or 'transformation of the parameters'. The former leads to the Generalized Least Square (GLS) estimator whereas the latter leads to such estimators as Indirect Least Squares (ILS) and Two Stage Least Squares (2SLS) estimators. A Simplified Two Stage Least Squares (S2SLS) estimator is proposed which is equivalent to 2SLS provided a Pseudo Random Binary Sequence (PRBS) is used as the input signal. The proposed estimator leads to considerable saving in computer time.

Chapter 4 is devoted to a discussion of Instrumental Variable (IV) technique for estimation of process parameters. The philosophy behind the IV estimator is discussed both from an algebraic as well as geometric viewpoint. It is shown that ILS, S2SLS and 2SLS estimators belong to the general class of IV estimators. A short survey of existing IV estimators is given and it is shown that the IV estimator proposed by

Mayne³² is identical to an estimator which utilizes suitably lagged values of the output as instruments and which is computationally much simpler. A recursive approximation to 2SLS (bootstrap) estimator which is computationally efficient and whose convergence properties compare well with the existing recursive IV estimators is proposed⁵. The problem of parameter estimation in force-free model is also discussed and it is established that the Yule-Walker estimator proposed by Mehra¹⁴ and Gersch¹⁵ is equivalent to the IV estimator that utilizes lagged outputs as instruments⁶.

In Chapter 5, by using a simpler model (process disturbed by output noise only) and the concept of 'structural' and 'incidental' parameters the definition of a general 'Bootstrap Estimator' is formalized. It is shown that the necessary (though not sufficient) conditions for the stability of such bootstrap algorithms can be satisfied by incorporating a simple stabilizing filter in the algorithm. The underlying relationship of the resulting estimators to the general k-class, h-class and double k-class estimators is demonstrated.

In Chapter 6 the techniques for estimating the noise parameters when process parameters (or their estimates) are known are discussed. A set of 2SLS estimators and the corresponding bootstrap estimator are formulated for this purpose. The problem of estimating the process as well as noise parameters simultaneously and in a single experiment is also discussed in this chapter. The adaptive recursive least squares algorithm proposed by Panuska⁴⁵ is compared with an algorithm proposed by the author, which consists of two consecutive bootstrap estimators.

The class of Bootstrap estimators proposed in the thesis are tested on experimental data obtained from a Glucose Tolerance Test⁷.

Chapter 7 contains brief comments on such related topics as identification in the presence of closed-loop control, tracking of time-varying parameters, identification of multivariable system and a priori estimation of system order. This chapter also contains the final conclusions.

The three appendices to the thesis contain relevant definitions and theorems that occur repeatedly in the text, and the test results for the Random Noise Generator used for the simulation experiments presented in the thesis.

CHAPTER 2

MODELLING OF LINEAR DISCRETE-TIME SYSTEMS

2.1 Basic State Variable Model

As pointed out in Section 1.1, modelling of the total system* (physical process embedded in its stochastic environment) requires that a 'Stochastic Model' be proposed that embodies the dynamics of the process, the dynamics of the stochastic environment and the way they interact with each other. Further, it is desirable that the proposed model should meet the following obvious constraints:

1. The model should be as simple as possible;
2. The model should represent the physical and stochastic processes involved as accurately as possible; and
3. If possible, the proposed model should facilitate the subsequent operations; viz. identification, state estimation and controller design.

The first two constraints are conflicting in nature, and lead to engineering tradeoffs that must be based on the available knowledge

* Henceforth the term 'system' will be used to include both the physical process and the associated noise structure whereas the term 'process' will be restricted to the physical process only.

of, and the intuitive feel for the system. Modelling, therefore, starts off by trying to answer the following questions about the process and its stochastic environment under the first two constraints:

1. *Is the process stationary?*

If the process is not time-invariant, the class of estimators that can be used for parameter identification is immediately restricted to the ones that can operate in real time, and the problem of identification, state estimation and control cannot be separated from each other. The special problems that are associated with the identification of non-stationary processes are discussed in Chapter 7.

2. *Is the process stable?*

Stability of the process per se is not essential as far as the identification of the process from a theoretical view point is concerned. However, for the identification techniques discussed in this thesis to be effective, it is necessary that the process being identified be stable.

3. *Is the process linear?*

Linearity here refers to the linearity of the dynamic relationships as opposed to the linearity in the parameters of a regression model³. Needless to say, identification of non-linear systems poses some special problems.

4. *Is the process controllable and observable?*

Since system identification procedures are to be based on the input-output measurements, for the purpose of this thesis, it is assumed that the process is controllable and observable. There is no experimental procedure (in terms of the input-output measurements) which will provide an adequate answer to this question. If it is known a priori that this assumption does not hold then any and all identification procedures based on input-output measurements will identify only that part of the process which is controllable from the input and observable at the output.

5. *What is the order of the process?*

Accuracy of identification, particularly for parametric or regression models, depends heavily on a correct answer to this question. Though, a priori estimation of the process order is a difficult problem, goodness of fit tests, using the residuals obtained from the fitted model, are usually employed to confirm whether the order of the fitted model is adequate.

6. *How does the stochastic environment interact with the process?*

Random disturbances may enter the process through the input-output transducers and/or operate directly on the process states through unknown dynamics.

7. *What are the statistical properties of the disturbances?*

Though it would be unrealistic to assume a complete knowledge of these statistical properties, some familiarity with the

source and general characteristics of the noise is essential. Specifically, the analysis is considerably simplified if the following assumptions are true:

- i) disturbances entering the process at various points are statistically independent of each other and also of the process input signal.
- ii) disturbances are normally distributed with zero mean.
- iii) disturbances are either white or have a rational spectrum.

The answers to the questions posed above are usually based on one or more of the following:

- 1. Preliminary experimentation.
- 2. Intuition or engineering judgement.
- 3. Outright a priori assumptions.

Assuming that the process is to be controlled by a digital computer (controller) and/or the input-output data are available at regularly spaced time intervals, we start with the following discrete-time, linear state-space model for the system (Fig. 2.1):

$$\begin{aligned}\underline{x}^o(k+1) &= \underline{F}^o \underline{x}(k) + \underline{g}^o u(k) + \underline{f}^o \eta(k) \\ y(k) &= \underline{h}^o' \underline{x}^o(k) + d u(k) + \xi(k)\end{aligned}\tag{2.1}$$

It is assumed that preliminary investigations assign the following characteristics to the process and the disturbances:

Fig. 2.1. Basic State Space Model

1. The input-output relationship of the physical process can be adequately described by a linear dynamic model of known order n . Further the model is time-invariant BIBO stable, and controllable and observable.
2. The process states are disturbed by additive noise $\eta(k)$ through $\underline{f}^\circ$ and the output is disturbed by additive measurement noise $\xi(k)$.
3. The noise processes $\eta(k)$ and $\xi(k)$ are sample sequences from zero mean, white* stationary Gaussian processes with finite variance that are statistically independent of each other and of the process input $u(k)$.

For the model represented by Eqn. (2.1), the identification problem may be formulated as follows:

"Given $\underline{h}^\circ$ and a record of the input-output data, estimate the $(n^2 + 2n + 1)$ elements of $\underline{F}^\circ$, $\underline{g}^\circ$, $\underline{f}^\circ$ and d ."

2.2 A Canonic Form with a Minimum Parameter Set

In the derivation of the model represented by Eqn. (2.1), the only constraints imposed were those pertaining to the simplicity and

* If the noise processes are not white but have an auto regressive structure (correlated noise with stable rational spectrum) the proposed model is still valid if the state vector is suitably augmented.

accuracy of the model. We are now ready to apply the third constraint discussed in Section 2.1. One way the identification procedures can be facilitated is by minimizing the number of parameters to be estimated, i.e., by using a system model that expresses the input-output relationships using a minimal set of parameters. Since the input-output relationships of a given linear system are invariant under a linear transformation of the states, we look for a suitable non-singular similarity transformation which will transform the model of Eqn. (2.1) to an equivalent system having a minimum number of parameters.

For a single input-single output linear system, a number of canonic representations are available, which specify the I/O relationship in terms of a minimum parameter set^{*}. One such realization is obtained by using the observability matrix:

$$T = \begin{bmatrix} h^0 \\ h^0 F^0 \\ \vdots \\ h^0 F^{0n-1} \end{bmatrix} \quad (2.2)$$

as the transformation^{8,9}. The transformation in this case is unique, because the observability matrix is $n \times n$ and non-singular.

Thus, applying the transformation (2.2) to the state-space model (2.1), we get the following equivalent system:

^{*} A detailed discussion of such canonic forms is beyond the scope of this thesis. See Desour, C.A. "Notes for a Second Course on Linear Systems", Van Nostrand, 1970, pp 99-104.

$$\begin{aligned}\underline{x}(k+1) &= F \underline{x}(k) + \underline{g} u(k) + \underline{f} \eta(k) \\ y(k) &= \underline{h}' \underline{x}(k) + d u(k) + \xi(k)\end{aligned}\quad (2.3)$$

where

$$\underline{x}(k) = T \underline{x}^{\circ}(k) \rightarrow \underline{x}^{\circ}(k) = T^{-1} \underline{x}(k) \quad (2.4)$$

$$F = T F^{\circ} T^{-1} = \begin{bmatrix} \underline{0} & I_{(n-1)} \\ -a_n & -a_{n-1} \dots -a_1 \end{bmatrix} \quad (2.5)$$

$$\underline{h}' = \underline{h}^{\circ'} T^{-1} = [1 \ 0 \ 0 \ \dots \ 0] \quad (2.6)$$

$$\underline{g} = T \underline{g}^{\circ} \quad (2.7)$$

$$\underline{f} = T \underline{f}^{\circ} \quad (2.8)$$

In the equivalent model (2.3), the vector $\underline{x}(k)$ is the vector of state variables and all elements of the system matrix 'F' except the last row are pre-determined, the latter being given by

$$[-a_n \ -a_{n-1} \ \dots \ -a_2 \ -a_1] = \underline{h}^{\circ'} (F^{\circ})^2 T^{-1} \quad (2.9)$$

Thus the total number of parameters to be estimated in the phase variable canonical model (Eqn. (2.3)) is $(3n + 1)$ as opposed to $(n^2 + 2n + 1)$ in the original model (Eqn. (2.1)).

2.3 Levy's Proper Canonical Representation

Assuming that the $(3n + 1)$ parameters in the preceeding canonical model have been consistently estimated, the next step towards the design of a suitable controller is to estimate the n vector of state variables $\underline{x}(k)$. Design of a Kalman filter for this purpose requires a knowledge of the second moments of $\eta(k)$ and $\xi(k)$. Since these moments are not known a priori, we may wish to estimate them during or after the identification procedure. However, the effects of the two independent noise sequences $\eta(k)$ and $\xi(k)$ on the measured output $y(k)$ cannot be easily separated, which means that their second moments cannot be estimated separately.

Levy's Proper Canonical model avoids this difficulty by specifying the input-output relations of the system in terms of a single noise sequence - namely the 'innovations sequence'. Thus, if the canonical form (2.3) is considered as the message model, the corresponding optimum linear filter (Kalman filter) is given by:

$$\begin{aligned}\hat{\underline{x}}(k+1/k+1) &= \hat{\underline{x}}(k+1/k) + \underline{K}(k)[y(k+1) - \underline{h}' \hat{\underline{x}}(k+1/k) - d u(k+1)] \\ \hat{\underline{x}}(k+1/k) &= \underline{F} \hat{\underline{x}}(k/k) + \underline{g} u(k)\end{aligned}\tag{2.10}$$

where standard notation from Kalman filtering theory has been used and $\underline{K}(k)$ is the n vector of Kalman filter gains.

The 'innovations process' $e(k)^{10,11}$ is defined as:

$$e(k) \triangleq [y(k) - \underline{h}' \hat{\underline{x}}(k/k-1) - d u(k)] \quad (2.11)$$

where $e(k)$ is a scalar sequence composed of samples from a zero mean, stationary Gaussian process, provided $\eta(k)$ and $\xi(k)$ have statistical properties that were assigned to them in Section 2.1.

Writing the steady state Kalman filter equations, (2.10), in terms of the innovations $e(k)$ leads to the Levy's Proper Canonical Representation:

$$\begin{aligned} \hat{\underline{x}}(k+1/k) &= F \hat{\underline{x}}(k/k-1) + \underline{g} u(k) + F \underline{K} e(k) \\ y(k) &= \underline{h}' \hat{\underline{x}}(k/k-1) + d u(k) + e(k) \end{aligned} \quad (2.12)$$

where $\underline{K}$ is the constant vector of steady state Kalman gains. By appropriately redefining the variables in Eqn. (2.11), the model may be simply written as:

$$\begin{aligned} \underline{x}(k+1) &= F \underline{x}(k) + \underline{g} u(k) + \underline{\alpha} e(k) \\ y(k) &= \underline{h}' \underline{x}(k) + d u(k) + e(k) \end{aligned} \quad (2.13)$$

$$\text{with} \quad e(k) \sim \text{N.I.D.}(0, \lambda^2) \quad (2.14)$$

and $\underline{\alpha} = F \underline{K}$ = constant vector.

In the Levy's proper canonical model (2.13), the n vector $\underline{\alpha}$ contains the steady state Kalman gains which in turn contain the necessary information on the second moments of $\eta(k)$ and $\xi(k)$ and the elements of $\underline{f}$ (Eq. (2.3)). Further, once the parameters of this model are known, the estimates $\hat{\underline{x}}(k+1/k)$ can be directly obtained from the model and the second moment of $e(k)$ can be consistently estimated by

$$\frac{1}{k} \sum_{i=1}^k [y(i) - \underline{h}' \hat{\underline{x}}(i/i-1) - d u(i)]^2$$

2.4 Astrom's Difference Equation Model

Since the only measurable quantities in Eqn. (2.13) are the output $y(k)$ and the input $u(k)$, it is appropriate to eliminate the state variables $\underline{x}(k)$ from the model. Recalling that $\underline{x}(k)$ is the n vector of phase variables and $\underline{h}' = [1 \ 0 \ 0 \ \dots \ 0]$, model (2.13) can be transformed into the following n^{th} order difference equation:

$$\begin{aligned} y(k) + a_1 y(k-1) + \dots + a_n y(k-n) &= b_0 u(k) + b_1 u(k-1) \dots \\ &\dots + b_n u(k-n) + e(k) + c_1 e(k-1) \dots + c_n e(k-n) \end{aligned} \quad (2.15)$$

where

$$[-a_n - a_{n-1} \dots - a_2 - a_1] = \text{Last row in } F \quad (2.16)$$

$$b_0 = d \quad (2.17)$$

$$b_i = g_i^* + a_i d, \quad i = 1, 2, \dots, n \quad (2.18)$$

where g_i^* is the i^{th} element of

$$\begin{bmatrix} g_1^* \\ g_2^* \\ \vdots \\ g_n^* \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ a_1 & 1 & 0 & 0 & \dots & 0 & 0 \\ a_2 & a_1 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n-1} & a_{n-2} & \dots & a_1 & 1 & \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_n \end{bmatrix} \quad (2.19)$$

$$c_i = \alpha_i^* + a_i, \quad i = 1, 2, \dots, n \quad (2.20)$$

where α_i^* is the i^{th} element of

$$\begin{bmatrix} \alpha_1^* \\ \alpha_2^* \\ \vdots \\ \alpha_n^* \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ a_1 & 1 & 0 & 0 & \dots & 0 & 0 \\ a_2 & a_1 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n-1} & a_{n-2} & \dots & a_1 & 1 & \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \quad (2.21)$$

It is worth noting that even though the parameters represented by the elements of F , $\underline{g}$, $\underline{\alpha}$, and d in the state space model (2.13) may be completely independent of each other, the parameters b_i ; $i = 1, 2, \dots, n$ (also c_i ; $i = 1, 2, \dots, n$) in the corresponding difference equation model (2.15) not only depend on each other but also on a_i ; $i = 1, 2, \dots, n-1$.

This interdependence may be avoided by using the dual of the 'phase variable' canonic form (instead of using the canonic form (2.3))^{*}.

The difference equation model may be written more compactly by defining the following polynomials in the unit delay operator q^{-1} :

$$\begin{aligned} A(q^{-1}) &= 1 + a_1 q^{-1} + a_2 q^{-2} + \dots + a_n q^{-n} \\ B(q^{-1}) &= b_0 + b_1 q^{-1} + b_2 q^{-2} + \dots + b_n q^{-n} \\ C(q^{-1}) &= 1 + c_1 q^{-1} + c_2 q^{-2} + \dots + c_n q^{-n}. \end{aligned} \quad (2.22)$$

so that Eqn. (2.15) may be written as:

$$A(q^{-1})y(k) = B(q^{-1})u(k) + C(q^{-1})e(k) \quad (2.23)$$

or

$$y(k) = \left[\frac{B(q^{-1})}{A(q^{-1})} \right] u(k) + \left[\frac{C(q^{-1})}{A(q^{-1})} \right] e(k) \quad (2.24)$$

where $[B(q^{-1})/A(q^{-1})]$ is the 'Pulse Transfer Function' of the 'noise-free process' and if

$$x(k) = \left[\frac{B(q^{-1})}{A(q^{-1})} \right] u(k), \quad (2.25)$$

is defined as the output of the physical process in the absence of the stochastic effects and

^{*} For the necessary transformation to obtain this particular canonic form see Chen, C.T. "Introduction to Linear System Theory", Holt, Rinehart and Winston, New York, 1970, pp 225-267.

$$w(k) = \left[\frac{C(q^{-1})}{A(q^{-1})} \right] e(k), \quad (2.26)$$

as the Mixed Auto-regressive Moving Average Process (MARMA) obtained by filtering $e(k)$ through $[C(q^{-1})/A(q^{-1})]$, Eqn. (2.23) may be represented by the block diagram in Fig. 2.2.*

Fig. 2.2. Block Diagram Representation of the Difference Equation Model

* This is a specific representation in which all the noise is referred to the output. Other representations, that are less restrictive, are possible.

Statistical model (2.23) is attributed to Astrom¹² and has received wide acceptance in the relevant literature. An alternative model due to Box and Jenkins¹³, where $w(k)$ is generated by filtering $e(k)$ through a system represented by the pulse transfer function $[F(q^{-1})/G(q^{-1})]$, is also used by some authors.

2.5 Some Special Cases

Levy's proper canonical model (Eqn. (2.13)) and the corresponding difference equation model (Eqn. (2.15)) is sufficiently general in that it can be reduced to a number of special cases that are frequently encountered in practice, by assigning appropriate values to the parameters. Some of these special cases are presented here.

1. $a_i = c_i; i = 1, 2, \dots, n$

Corresponding models for this case simplify to:

$$\begin{aligned}\underline{x}(k+1) &= F \underline{x}(k) + g u(k) \\ y(k) &= \underline{h}' \underline{x}(k) + e(k)\end{aligned}\tag{2.27}$$

or

$$A(q^{-1})y(k) = B(q^{-1})u(k) + A(q^{-1})e(k)\tag{2.28}$$

or

$$\begin{aligned}\underline{x}(k) &= \left[\frac{B(q^{-1})}{A(q^{-1})} \right] u(k) \\ y(k) &= \underline{x}(k) + e(k)\end{aligned}\tag{2.29}$$

with $e(k) \sim \text{N.I.D.}(0, \lambda^2)$.

This simple model seems to fit a large number of situations in systems theory and econometrics. In systems theory this model adequately describes a linear dynamic process whose output is corrupted by additive white noise. In econometrics, the model implies the presence of distributed lags together with measurement error in the regresand and some of the regressors.

2. $b_i = 0; i = 0, 1, 2, \dots, n$

The models corresponding to this situation simplify to

$$\begin{aligned}\underline{x}(k+1) &= F \underline{x}(k) + \underline{\alpha} e(k) \\ y(k) &= \underline{h}' \underline{x}(k) + e(k)\end{aligned}\tag{2.30}$$

or

$$A(q^{-1})y(k) = C(q^{-1})e(k)\tag{2.31}$$

This case implies the absence of any exogenous variables (deterministic inputs) in the system. The problem of parameter and state estimation in such force-free systems is frequently encountered in orbit determination of satellites¹⁴. In 'time series analysis' the model represents a 'mixed auto-regressive moving average'¹⁵ (MARMA) process. This case warrants special attention because in the absence of known inputs, identification techniques based on the 'reduced form model' are not applicable.

3. $c_i = 0; i = 1, 2, \dots, n$

$$\begin{aligned}\underline{x}(k+1) &= F \underline{x}(k) + \underline{g} u(k) + \underline{\alpha} e(k) \\ y(k) &= \underline{h}' \underline{x}(k) + e(k)\end{aligned}\quad (2.32)$$

with

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = - \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ a_1 & 1 & 0 & \dots & 0 \\ \vdots & & & & \\ a_{n-1} & a_{n-2} & \dots & a_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \quad (2.33)$$

or

$$A(q^{-1})y(k) = B(q^{-1})u(k) + e(k) \quad (2.34)$$

This model is generally referred as the 'Mann and Wald model'¹⁶, who have shown that consistent estimates of its parameters can be obtained by applying the Ordinary Least Squares (OLS) techniques. This is also the model used by Lee⁸. This model is encountered more frequently in econometrics than in systems theory. It is evident from Eqn. (2.33) that the model assumes a very special structure for $\underline{\alpha}$ and hence for $\underline{f}^o$ and $\eta(k)$ in Fig. 2.1, and it is difficult to justify its use in the systems context.

4. $b_i = 0; i = 0, 1, 2, \dots, n, c_i = 0; i = 1, 2, \dots, n$

$$A(q^{-1})y(k) = e(k) \quad (2.35)$$

This model represents a pure auto-regressive (AR) process of finite order and its parameters can be consistently estimated by ordinary least squares techniques.

$$5. \quad \underline{b_i = 0; i = 0, 1, 2, \dots, n, a_i = 0; i = 1, 2, \dots, n}$$

$$y(k) = c(q^{-1})e(k) \quad (2.36)$$

The model describes a pure moving average (MA) process. Unlike the auto-regressive process, its parameters cannot be estimated easily. It is further evident that under certain stability conditions a moving average process of finite order can be converted to an auto-regressive process of infinite order and vice-versa.

If the preliminary investigations indicate that the system under study can adequately be represented by one of these special cases, then it may be possible to use a simplified identification procedure. Though the ultimate aim is to estimate the parameters of the Levy's proper canonical model, some attention has been paid, in this thesis, to these special cases, as they do represent a fairly large class of practical situations.

CHAPTER 3

SOME IDENTIFICATION TECHNIQUES FOR STRUCTURAL MODELS

3.1 Preliminary Remarks

Once a suitable model for the process and its stochastic environment has been postulated, the next step is to estimate the parameters associated with this model. Thus in general, the problem we pose is the following:

"Given the n^{th} order model:

$$\begin{aligned}\underline{x}(k+1) &= \underline{F} \underline{x}(k) + \underline{g} u(k) + \underline{\alpha} e(k) \\ y(k) &= \underline{h}' \underline{x}(k) + d u(k) + e(k)\end{aligned}\tag{3.1}$$

or

$$A(q^{-1})y(k) = B(q^{-1})u(k) + C(q^{-1})e(k)\tag{3.2}$$

with

$$e(k) \sim \text{N.I.D. } (0, \lambda^2)\tag{3.3}$$

and a set of input-output measurements, obtain consistent estimates of the $(3n + 1)$ parameters $\underline{F}$, $\underline{g}$, $\underline{\alpha}$, and d in model (3.1) or

$[b_0, b_1, \dots, b_n, -a_1, -a_2, \dots, -a_n, c_1, \dots, c_n]$ in model (3.2)."

The assumption on the normality of $e(k)$ is not critical, except when 'interval estimation' is to be carried out, in which case a knowledge of λ^2 or its estimate becomes necessary. More often than

not, the assumption is quite valid because the stochastic processes $\eta(k)$ and $\xi(k)$ (Eqn. (2.1)) represent the average effect of a number of independent noise sources and the normality of $\eta(k)$ and $\xi(k)$ and hence that of $e(k)$ is assured by the Central Limit Theorem. Further, if $e(k)$ are normally distributed, the analysis is considerably simplified, e.g., the asymptotic distributions for various consistent estimators can be derived without going into the complicated theory of Laws of Large Numbers and the Central Limit Theorems.

As already mentioned in Section 1.1, the process states can be estimated and hence an effective control strategy can be designed if accurate estimates of both the process parameters, $[b_0, b_1, \dots, b_n, -a_1, -a_2, \dots, -a_n]$ and the noise parameters, $[c_1, c_2, \dots, c_n]$ are available. The process and noise parameters may either be estimated simultaneously or one may first choose to obtain consistent estimates of the process parameters which may be used subsequently to estimate the noise parameters.

In this and the next two chapters, attention is restricted to the estimation of process parameters. The problem of estimating the noise parameters, given consistent estimates of the process parameters and the problem of estimating both these sets of parameters in a single experiment is discussed in Chapter 6.

3.2 Least Squares Estimates of Process Parameters

Assuming $(N+n)$ pairs of input-output measurements to be available, the set of N difference equations obtained from Eqn. (3.2), each with $p = 2n + 1$ unknown process parameters may be written in the following vector-matrix notation:

$$\underline{y} = M\underline{\beta} + \underline{v} \quad (3.4)$$

or

$$\underline{y} = M_1\underline{\beta}_1 + M_2\underline{\beta}_2 + \underline{v}$$

or

$$\underline{y} = [M_1 | M_2] \begin{bmatrix} \underline{\beta}_1 \\ \underline{\beta}_2 \end{bmatrix} + \underline{v} \quad (3.5)$$

where

$$\begin{matrix} \underline{y} \\ (N \times 1) \end{matrix} = [y(n+1) y(n+2) \dots y(n+N)]' \quad (3.6)$$

$$\begin{matrix} M \\ (N \times p) \end{matrix} = [M_1 | M_2] = \begin{bmatrix} u(n+1) & u(n) & \dots & u(1) & | & y(n) & \dots & y(1) \\ u(n+2) & u(n+1) & \dots & u(2) & | & y(n+1) & \dots & y(2) \\ \vdots & \vdots & & \vdots & | & \vdots & & \vdots \\ u(n+N) & u(n+N-1) & \dots & u(N) & | & y(n+N-1) & \dots & y(N) \end{bmatrix} \quad (3.7)$$

$$\underset{(px1)}{\underline{\beta}} = \begin{bmatrix} \underline{\beta}_1 \\ -\underline{\beta}_2 \end{bmatrix} = [b_0 b_1, \dots, b_n | -a_1 - a_2, \dots, -a_n]' \quad (3.8)$$

$$\underset{(Nx1)}{\underline{v}} = \begin{bmatrix} v(n+1) \\ v(n+2) \\ \vdots \\ v(n+N) \end{bmatrix} = \begin{bmatrix} e(n+1) + c_1 e(n) + \dots + c_n e(1) \\ e(n+2) + c_1 e(n+1) + \dots + c_n e(2) \\ \vdots \\ e(n+N) + c_1 e(n+N-1) + \dots + c_n e(N) \end{bmatrix} \quad (3.9)$$

Remarks

1. In general, $N > p$, i.e., the number of measurements exceeds the number of parameters to be estimated.
2. For a single equation or one of the rows in Eqn. (3.4)

$$y(k) = b_0 u(k) + \dots + b_n u(k-n) - a_1 y(k-1) - \dots - a_n y(k-n) + v(k)$$

the independent variables $u(k), u(k-1), \dots, u(k-n), y(k-1), y(k-2), \dots, y(k-n)$ are referred as the 'regressors' and the dependent variable $y(k)$ on the left hand side is referred as the 'regressand'.

3. In the vector matrix model (3.4), the following terminology is often used:

$\underline{y}$ = Observation vector

M = Design matrix

$\underline{\beta}$ = Parameter vector

$\underline{v}$ = Disturbance vector.

4. The partitioning of the parameter vector $\underline{\beta}$ and the design matrix M indicated by Eqn. (3.5), (3.7) and (3.8) is meant to emphasize the fact that the elements of M_1 consist of input (or 'exogenous') variables only, which are generally not correlated with the elements of the disturbance vector $\underline{v}$, and similarly, the elements of M_2 consist of output (or 'exogenous') variables that may (or may not) be correlated with the elements of $\underline{v}$. We shall find repeated use for this partitioned form of the model in the subsequent pages.
5. The following partitioning of the design matrix M into its columns and rows is also some times used:

$$\underset{(N \times p)}{M} = [\underline{u}_0 \ \underline{u}_1 \ \dots \ \underline{u}_n \ y_1 \ y_2 \ \dots \ y_n] \quad (3.10)$$

$$M = \begin{bmatrix} \underline{m}'(n+1) \\ \underline{m}'(n+2) \\ \vdots \\ \underline{m}'(k) \\ \vdots \\ \underline{m}'(n+N) \end{bmatrix} \quad (3.11)$$

where

$$\underline{m}(k) = [u(k) \ u(k-1) \ \dots \ u(k-n) \ y(k-1) \ \dots \ y(k-n)]' \quad (3.12)$$

so that

$$y(k) = \underline{m}'(k) \underline{\beta} + v(k) \quad (3.13)$$

Needless to say, model (3.4) and its variations presented above, will occur again and again in the following pages.

If the matrix M has full rank ($= p$), which, in practice can be assured by proper input signal design, model (3.4) leads to the OLS estimator:

$$\hat{\underline{\beta}}_{LS} = (M'M)^{-1} M' \underline{y} \quad (3.14)$$

Model (3.4), which is the basis of the OLS estimator $\hat{\underline{\beta}}_{LS}$ and which will dictate its statistical properties, has the following characteristics:

1. It will be assumed that

$$\text{p.lim}_{N \rightarrow \infty} \left[\frac{1}{N} M'M \right] \text{ exists and is non-singular} \quad (3.15)$$

2. Since each element of the disturbance vector $\underline{v}$ ($v(i); i = n+1, n+2, \dots, n+N$) is a moving average process characterized by Eqn. (3.9), the additive noise terms $v(k), v(k+j)$ ($j \leq n$) are correlated, i.e.,

$$E(\underline{v}) = \underline{0}; \quad E(\underline{v} \underline{v}') = \Omega \neq \lambda_v^2 I \quad (3.16)$$

3. Regressors $y(k-1), y(k-2), \dots, y(k-n)$ are lagged values of the regressand $y(k)$.
4. The effective disturbance $v(k)$ at any given sampling instant k is correlated with the regressors $y(k-1), y(k-2), \dots, y(k-n)$, i.e.,

$$E[\underline{m}(k) \underline{v}(k)] \neq \underline{0}$$

therefore

$$\text{p.lim}_{N \rightarrow \infty} \left[\frac{1}{N} \underline{M}' \underline{v} \right] \neq \underline{0} \quad (3.17)$$

Though property 4 above is a consequence of properties 2 and 3, its effect by far is the most drastic in that it renders the OLS estimates of the process parameters in model (3.4) inconsistent. This assertion can easily be proved as follows:

$$\text{p.lim}_{N \rightarrow \infty} [\hat{\underline{\beta}}_{LS}] = \underline{\beta} + \text{p.lim}_{N \rightarrow \infty} \left[\left(\frac{1}{N} \underline{M}' \underline{M} \right)^{-1} \left(\frac{1}{N} \underline{M}' \underline{v} \right) \right] \quad (3.18)$$

or

$$\text{p.lim}_{N \rightarrow \infty} [\hat{\underline{\beta}}_{LS}] = \underline{\beta} + \text{p.lim}_{N \rightarrow \infty} \left[\frac{1}{N} \underline{M}' \underline{M} \right]^{-1} \text{p.lim}_{N \rightarrow \infty} \left[\frac{1}{N} \underline{M}' \underline{v} \right] \quad (3.19)$$

where Eqn. (3.19) is obtained from Eqn. (3.18) by invoking Slutsky's theorem^{A,17,18*}. In Eqn. (3.19), the first limit on the right hand side exists by assumption (3.15) and the second limit does not vanish (Eqn. (3.17)), which confirms that

$$\text{p.lim}_{N \rightarrow \infty} [\hat{\underline{\beta}}_{LS}] \neq \underline{\beta} \quad (3.20)$$

* Lettered references pertain to corresponding Appendices at the end of this thesis.

The following points, however, should be noted:

1. The inconsistency in $\underline{\beta}_{LS}$ will persist, even if the process is disturbed by additive white noise at the output, i.e., when $c_i = a_i$ and the simpler model represented by Eqn. (2.29) holds.
2. The OLS estimator in model (3.4) will be consistent (though biased for finite sample size) iff. $c_i = 0$; $i = 1, 2, \dots, n$, so that $E(\underline{v} \underline{v}') = \lambda_v^2 I$, i.e., $\underline{\beta}_{LS}$ is a consistent estimator of the process parameters in model (3.4) if the system can either be represented by the Mann and Wald model (Eqn. (2.34)) or by a pure Auto-regressive process (Eqn. (2.35)), where $c_i = 0$, $b_i = 0$.

3.3 Transformation of Data - Generalized Least Squares Estimator

Thus the application of OLS to model (3.2) or (3.4) leads to inconsistent estimates of the process parameters. However we may recall from Section 3.2 that $\underline{\beta}$ can be consistently estimated by an OLS estimator provided the elements of the disturbance vector $\underline{v}$ are independently and identically distributed, i.e., $E(\underline{v} \underline{v}') = \lambda_v^2 I$. We may therefore look for a linear transformation that would transform model (3.4) into a model in which the above condition on the disturbance vector is satisfied. If OLS is now applied to the transformed model, the resulting estimator for $\underline{\beta}$ would be consistent.

The Generalized Least Squares (GLS) estimator, also known as the Atken estimator utilizes this concept, with the required transformation provided by the $N \times N$ non-singular matrix R' which satisfies:

$$R R' = \Omega^{-1} \quad (3.21)$$

where

$$\Omega \triangleq E(\underline{v} \underline{v}') = \text{Cov}(\underline{v}) \quad (3.22)$$

so that

$$E(R' \underline{v} \underline{v}' R) = \text{Cov}(R' \underline{v}) = R' \Omega R \equiv I \quad (3.23)$$

Existence of R' that satisfies (3.23) is assured by the fact that Ω is a variance-covariance matrix for a moving average process and hence positive definite^{B,19}.

The transformed model obtained by premultiplying both sides of (3.4) by R' is

$$R' \underline{y} = R' M \underline{\beta} + R' \underline{v} \quad (3.24)$$

or

$$\underline{y}^* = M^* \underline{\beta} + \underline{v}^* \quad (3.25)$$

with

$$E(\underline{v}^*) = 0, \quad E(\underline{v}^* \underline{v}^{*'}) = I \quad (3.26)$$

OLS now applied to model (3.25) provides the GLS estimator

$$\hat{\underline{\beta}}_{GLS} = (M^*{}' M^*)^{-1} M^*{}' \underline{y}^*$$

or

$$\hat{\underline{\beta}}_{GLS} = (M' R R' M)^{-1} M' R R' \underline{y}$$

or

$$\hat{\underline{\beta}}_{GLS} = (M' \Omega^{-1} M)^{-1} M' \Omega^{-1} \underline{y} \quad (3.27)$$

An alternative way of deriving the GLS estimator from the model (3.4) is to minimize the quadratic form

$$Q^* = (\underline{y} - M\underline{\beta})' \Omega^{-1} (\underline{y} - M\underline{\beta}) \quad (3.28)$$

with respect to $\underline{\beta}$, as opposed to the quadratic form

$$Q = (\underline{y} - M\underline{\beta})' (\underline{y} - M\underline{\beta}) \quad (3.29)$$

Since the model (3.25) - (3.26) now conforms to the Mann and Wald model, the GLS estimator is consistent, though still biased for finite sample size²⁰. The latter deficiency results from the presence of distributed lags in the model. The asymptotic variance-covariance matrix for the estimator is given by*

$$\text{Cov} (\hat{\underline{\beta}}_{GLS}) = (M' \Omega^{-1} M)^{-1} \quad (3.30)$$

* For definition of asymptotic covariance matrix see Appendix A and B.

and asymptotically the estimator is Best Linear Consistent and hence asymptotically efficient²¹.

For a known error covariance matrix Ω and for large sample size, the GLS estimates are also the Maximum Likelihood (MLE) estimates, which have the desirable properties of consistency and asymptotic efficiency. However, the assumption concerning the a priori knowledge of Ω places a severe restriction on its practical applications. A variety of approximations to the GLS estimator can be found in the engineering^{7,22,23,24} as well as econometric^{21,23} literature, where a suitable estimate of Ω is used in Eqn. 3.23 or equivalently, the output $y(k)$ is passed through a suitable linear filter so that the filtered output would fit (at least approximately) the model 3.25-26. Use of a consistent estimate of Ω instead of its true value, of course, introduces certain loss in asymptotic efficiency²¹, except in the absence of distributed lags.

3.4 Transformation of Parameters - Reduced Form Model

It was noted in Section 3.2 that the OLS inconsistency in model (3.4) results from the combined effects of the presence of lagged endogenous variables and serially correlated errors. The method outlined in the preceding section removes the latter deficiency through a suitable transformation of the data. An alternative approach would be to remove the lagged endogenous variables from the model - though the effective disturbances may still be serially correlated. This can be achieved by reparameterizing the model, i.e., by using a suitable transformation (usually

nonlinear) on the parameters, so that the independent variable is expressed solely in terms of exogenous variables and the additive (though serially correlated) disturbances. Such a model is generally referred as the 'Reduced Form Model' in econometrics and as the 'Impulse Response Model' in Systems engineering. Parameters of such a model can be consistently estimated by OLS.

Thus writing the model represented by Eqn. (3.2) as the sum

$$y(k) = x(k) + w(k) \quad (3.31)$$

where

$$x(k) = \frac{B(q^{-1})}{A(q^{-1})} u(k) , \quad (3.32)$$

$$w(k) = \frac{C(q^{-1})}{A(q^{-1})} e(k) , \quad (3.33)$$

$$e(k) \sim \text{N.I.D.}(0, \lambda^2), \quad (3.34)$$

and recognizing that $[B(q^{-1})/A(q^{-1})]$ is the 'Pulse Transfer Function' of the process, the output $y(k)$ may be written in terms of the input sequence alone.

Theoretically, the number of parameters in the reduced form model should be infinity. In practice, however, the sequence can be terminated after a finite number of terms, because the process is assumed to be stable. The number of terms retained in the model will depend on the relative stability of the process, which in turn is determined by the characteristic roots of $A(q^{-1})$. Thus we may write

$$y(k) = \gamma_0 u(k) + \gamma_1 u(k-1) + \dots + \gamma_n u(k-n) + \dots + \gamma_{s-1} u(k-s-1) + w(k) \quad (3.35)$$

Let $N+s$ be the number of input-output data, then the N equations obtained from (3.35), each with s unknown parameters can be written as:

$$\underline{y} = U\underline{\gamma} + \underline{w} \quad (3.36)$$

with

$$E(\underline{w}) = \underline{0}, \quad E(\underline{w} \underline{w}') = \psi \neq \lambda^2 \underline{w} I \quad (3.37)$$

$$E(U' \underline{w}) \equiv \underline{0} \quad (3.38)$$

where

$$U = N \times s$$

and

$$\underline{\gamma} = s \times 1$$

Assuming a full rank for 'U', the OLS estimator for $\underline{\gamma}$ is given by

$$\hat{\underline{\gamma}} = (U'U)^{-1} U' \underline{y} \quad (3.39)$$

The unbiasedness (for finite sample size) and consistency of $\hat{\underline{\gamma}}$ is assured by Eqn. (3.38). On the other hand, the fact that $w(k)$, in general, are serially correlated prevents $\hat{\underline{\gamma}}$ from being Best Linear Unbiased Estimates (BLUE) of $\underline{\gamma}$. $\hat{\underline{\gamma}}$ will be BLUE iff.

$$C(q^{-1}) \equiv A(q^{-1}) \rightarrow c_i \equiv a_i \rightarrow w(k) \equiv e(k) \quad (3.40)$$

i.e., iff. the process is disturbed by additive white noise at the output.

Though extensive work has been done on the identification of reduced form or impulse response models, the model as such suffers from the following disadvantages:

1. As already mentioned, depending on the relative stability of the process, the dimension of U can become very large, thus entailing a substantial amount of computation.
2. Most system design techniques are based on the finite dimensional structural models - such as Eqn. (3.1) or (3.2).

However, a knowledge of the reduced form parameters (or its estimates) is still an asset, because this knowledge may be used in formulating consistent estimators for the structural parameters β_i . In general

$$\beta_i = f_i(\gamma_j); \quad i = 1, 2, \dots, p; \quad j = 1, 2, \dots, s$$

where f_i are non-linear functions, and it is difficult to solve this set of nonlinear algebraic equations directly. The difficulty may however be avoided by using a somewhat different approach. Thus:

$$\underline{y} = M \underline{\beta} + \underline{v}$$

also

$$\underline{y} = U \underline{\gamma} + \underline{w}$$

or

$$\underline{y} = U \hat{\underline{\gamma}} + \hat{\underline{w}} \quad (3.41)$$

where

$$\hat{\underline{w}} = \underline{y} - U \hat{\underline{\gamma}} \quad (3.42)$$

represents the vector of reduced form residuals. It therefore follows that

$$U \hat{\underline{y}} = M \underline{\beta} + (\underline{v} - \hat{\underline{w}}) \quad (3.43)$$

or

$$U'U \hat{\underline{y}} = U'M \underline{\beta} + U'\underline{v} - U'\hat{\underline{w}}$$

where

$$U'\hat{\underline{w}} = U'[\underline{y} - U'(U'U)^{-1} U'\underline{y}] = \underline{0},$$

so that

$$U'U[(U'U)^{-1} U'\underline{y}] = U'M \underline{\beta} + U'\underline{v}$$

or

$$U'\underline{y} = U'M \underline{\beta} + U'\underline{v} \quad (3.44)$$

Since the elements of U and $\underline{v}$ are not correlated with each other

$$\text{p.lim}_{N \rightarrow \infty} \left[\frac{1}{N} U'\underline{v} \right] = \underline{0} \quad (3.45)$$

and we may attempt to estimate the structural parameters $\underline{\beta}$ from the relation:

$$U'\underline{y} = U'M \hat{\underline{\beta}} \quad (3.46)$$

which may be considered as the set of "normal equations".

The matrix $(U'M)$ in the above relation has the dimension $(s \times p)$ and we have already assumed that both U and M have full rank. Depending on the number of columns in the U matrix or the number of significant impulse response coefficients, we may distinguish the following three cases:

Case 1: $s < p$

This case implies that the number of reduced form parameters is less than the number of parameters β_i ; $i = 1, 2, \dots, p$ in the structural model. Thus the number of 'normal equations' (3.46) is less than the number of unknown parameters $\underline{\beta}$ and an unique and exact solution for $\hat{\underline{\beta}}$ cannot be found. The model is referred as an 'unidentified' or 'under-identified' model. If the model has been correctly specified (process order has not been over-guessed) this contingency ($s < p$) will occur with probability zero. Best approximate solutions for $\underline{\beta}$ in an under-identified model can be obtained by using a suitable 'generalized inverse' or a 'pseudo-inverse'^{26,27}.

Case 2: $s = p$

This case implies a 'just-identified' model and such a situation is possible though not probable. It is likely to occur if all the roots of the polynomial $A(q^{-1})$ are very close to the origin of the unit circle. For the 'just-identified' case, the matrix $(U'M)$ in Eqn. (3.46) is $(p \times p)$ and under the assumption that

$$p.\lim_{N \rightarrow \infty} \left[\frac{1}{N} U'M \right] \text{ exists as a non singular matrix} \quad (3.47)$$

$\hat{\underline{\beta}}$ may be consistently estimated as:

$$\hat{\underline{\beta}} = (U'M)^{-1} U'y \quad (3.48)$$

For a just-identified model estimator (3.48) is known as the Indirect Least Squares (ILS) estimator.

Case 3: $s > p$

This case represents the 'over-identified' model and is most commonly encountered in practice. Since by assumption, all the columns of U are linearly independent, different estimators for $\underline{\beta}$ can be obtained from Eqn. (3.46), depending on which p columns of U are retained for estimating $\underline{\beta}$. In other words, if we insist that the estimator retain the simple format of the ILS estimator (Eqn. (3.48)), there is no unique solution. In fact one can obtain $s!/p!(s-p)!$ different estimators, and further since each of these estimators utilizes only a part of the information, they will not be efficient - though they will retain their consistency.

However, if we multiply through by $(U'M)'$ in (3.46), we get the estimator:

$$\hat{\underline{\beta}} = (M'UU'M)^{-1} M'UU'y \quad (3.49)$$

This is a new estimator proposed by the author and as we shall see, for the types of test input signals that are often used for system identification, and for large samples, estimator (3.49) not only provides the same results as the well known Two Stage Least Squares Estimator (to be discussed next) but leads to a considerable reduction in computation time and complexity. For this reason, this estimator will be referred to as the Simplified Two Stage Least Squares (S2SLS) estimator.

An alternative approach to deal with an over-identified model is provided by the Two Stage Least Squares (2SLS) technique, in which the first stage consists of replacing the lagged endogenous variables

$y(k-1), y(k-2), \dots, y(k-n)$ in the right hand side of model (3.4) by their estimates given by:

$$\hat{y}(k) = \hat{\gamma}_0 u(k) + \hat{\gamma}_1 u(k-1) + \dots + \hat{\gamma}_{s-1} u(k-s-1) \quad (3.50)$$

where $\hat{\gamma}_i$ are the OLS estimates of the reduced form parameters. In the second stage, OLS is applied to the resulting model, leading to a consistent estimator for the structural parameters β_i .

The following steps are, therefore, necessary for obtaining the 2SLS estimates of β :

1. For the reduced form model:

$$\underline{y} = \underline{U} \underline{\gamma} + \underline{w}.$$

estimate the reduced form parameters $\underline{\gamma}$ using the OLS estimator

$$\hat{\underline{\gamma}} = (\underline{U}'\underline{U})^{-1} \underline{U}'\underline{y}$$

2. Obtain the 'fitted' value of $\underline{y}$ as:

$$\hat{\underline{y}} = \underline{U} \hat{\underline{\gamma}}$$

3. In the structural model:

$$\underline{y} = \underline{M} \underline{\beta} + \underline{v}$$

replace the lagged endogenous variables $y(k); k = 1, 2, \dots, N+n-1$ on the r.h.s. by their fitted values from 2. above, which gives the model:

$$\underline{y} = \hat{\underline{M}} \underline{\beta} + \underline{v} + \underline{\varepsilon} \quad (3.51)$$

4. Use OLS on model (3.51) to give the 2SLS estimator:

$$\hat{\beta}_{2SLS} = (\hat{M}'\hat{M})^{-1} \hat{M}'\underline{y} \quad (3.52)$$

where $\underline{\varepsilon}$ in model (3.51) represents the error vector introduced by using $\hat{M}$ instead of M .

The OLS estimator for $\underline{y}$ in the first stage and the one for $\underline{\beta}$ in the second stage can be combined to give a single expression for the 2SLS estimator. Thus considering the following partitioned form of model (3.4):

$$\underline{y} = M_1\underline{\beta}_1 + M_2\underline{\beta}_2 + \underline{v} \quad (3.53)$$

where M_1 contains only the input variables and M_2 only the lagged output variables, replacement of the elements in M_2 by their estimates is equivalent to the model:

$$\underline{y} = M_1\underline{\beta}_1 + U(U'U)^{-1} U'M_2 + \underline{v} + \hat{W}_n\underline{\beta}_2 \quad (3.54)$$

where

$$\begin{aligned} (U'U)^{-1}U'M_2 &= (U'U)^{-1}U' [\underline{y}_1 | \underline{y}_2 | \dots | \underline{y}_n] \\ &= [\hat{\underline{y}}_1 | \hat{\underline{y}}_2 | \dots | \hat{\underline{y}}_n] \end{aligned} \quad (3.55)$$

and

$$\hat{W}_n = \begin{bmatrix} \hat{w}_1 & \hat{w}_2 & \dots & \hat{w}_n \end{bmatrix} \quad (3.56)$$

where

$$\hat{w}_j = \underline{y}_j - U \hat{\underline{y}}_j \quad (3.57)$$

is the vector of residuals corresponding to $\underline{y}_j$ and $\hat{\underline{y}}_j$.

However, the matrix U may be partitioned as:

$$U = [M_1 : U_2] \rightarrow M_1 = U \begin{bmatrix} I_{n+1} \\ 0 \end{bmatrix} \quad (3.58)$$

so that

$$U(U'U)^{-1} U'M_1 = M_1 \quad (3.59)$$

and we may simply write (3.54) as

$$y = U(U'U)^{-1} U'M \beta + v + \hat{W}_p \beta \quad (3.60)$$

where

$$\hat{W}_p = [0 : \hat{W}_n] = (M - \hat{M}) \quad (3.61)$$

OLS now applied to model (3.60) gives the 2SLS estimator

$$\hat{\beta}_{2SLS} = [M'U(U'U)^{-1}U'M]^{-1} M'U(U'U)^{-1}U y \quad (3.62)$$

where the error term $\hat{W}_p \beta$ (Eqn. (3.60)) does not enter into the normal equations for $\hat{\beta}_{2SLS}$, because

$$\begin{aligned} M'U(U'U)^{-1}U'\hat{W}_p \beta &= M'U(U'U)^{-1}U'(M - \hat{M}) \beta \\ &= M'U(U'U)^{-1}U' [M - U(U'U)^{-1}U'M] \beta \\ &= M'U(U'U)^{-1} [U'M - (U'U)(U'U)^{-1}U'M] \beta \\ &= 0. \end{aligned}$$

The 2SLS estimator may be tested for consistency by putting $y = M\beta + v$ in Eqn. (3.62) and taking probability limits on both sides so that we get:

$$\begin{aligned} \text{p.lim}_{N \rightarrow \infty} (\hat{\beta}_{2\text{SLS}} - \beta) &= \text{p.lim}_{N \rightarrow \infty} \left\{ \left[\left(\frac{1}{N} M'U \right) \left(\frac{1}{N} U'U \right)^{-1} \left(\frac{1}{N} U'M \right) \right]^{-1} \times \right. \\ &\quad \left. \times \left(\frac{1}{N} M'U \right) \left(\frac{1}{N} U'U \right)^{-1} \left(\frac{1}{N} U'v \right) \right\} \end{aligned}$$

we may now apply Slutsky's theorem and evaluate the limiting values of the individual terms. Since the input variables $u(k)$ and the effective disturbances $v(k)$ are not correlated $\text{p.lim}_{N \rightarrow \infty} (1/N U'v) = 0$ and since the probability limits of all other terms exist by assumption, the estimator is consistent.

The 2SLS may also be considered as equivalent to the OLS applied to the model

$$\begin{aligned} y(k) &= b_0 u(k) + b_1 u(k-1) + \dots + b_n u(k-n) - \\ &\quad - a_1 \hat{x}(k-1) - \dots - a_n \hat{x}(k-n) + \hat{w}(k) \end{aligned} \quad (3.63)$$

where $\hat{x}(k)$ are the estimates of the noise-free process output obtained by fitting the reduced form model and $w(k)$ are the corresponding residuals. Since the regressors $\hat{x}(k-1), \hat{x}(k-2), \dots, \hat{x}(k-n)$ in the model (3.63) are minimally correlated with the effective disturbance $\hat{w}(k)$, the 2SLS estimator is consistent. In general, $\hat{w}(k)$ are serially correlated and therefore the 2SLS estimator will not always be efficient. It will be consistent as well as asymptotically efficient iff.

$$c_i \equiv a_i; \quad i = 1, 2, \dots, n \rightarrow w(k) \sim \text{N.I.D.}(0, \lambda^2) \quad (3.64)$$

i.e., if the process is disturbed only by additive white noise at the output²⁰.

To obtain improved asymptotic efficiency when the above condition is not satisfied, Theil and Zellner²⁸ have proposed a Three Stage Least Squares (3SLS) estimator in which the residuals from the 2SLS procedure are used to estimate the Variance-Covariance matrix of the errors. This estimated Variance-Covariance matrix is used in the third stage, to formulate an approximate GLS (Aitken) estimator to obtain increased asymptotic efficiency.

The various estimators discussed in this section, which originate from the concept of a reduced form model (estimators (3.48), (3.49) and (3.62)) have the following characteristics:

1. The 2SLS estimator (3.62) represents the general structure of such estimators because by a suitable choice of the matrix U , it may be reduced to either the ILS estimator (3.48) or the S2SLS estimator (3.49). The single expression (3.62), for the 2SLS estimator was derived by combining the two consecutive stages involved in this procedure. If the entire data record is to be processed in a batch, this single expression can be used to implement the 2SLS estimator directly, i.e., without explicitly estimating the reduced form parameters $\underline{\gamma}$. In practice, however, it is found expedient to estimate the reduced form parameters explicitly. In the latter case, it becomes possible to use the recursive forms of the individual estimators for $\underline{\gamma}$ and $\underline{\beta}$, thus avoiding the two matrix inversions indicated in Eqn. (3.62).

2. If the model is just-identified, so that $(U'M)$ is a $p \times p$ matrix that satisfies condition (3.45), then the estimators (3.49) and (3.62) automatically reduce to the single ILS estimator (3.47).

3. If the input sequence $u(k)$; $k = 1, 2, \dots$ is such that it satisfies:

$$\text{p.lim}_{N \rightarrow \infty} \left[\frac{1}{N} U'U \right] = \lambda_u^2 I, \quad (3.65)$$

then estimator (3.62) reduces to the S2SLS estimator (3.49). In other words, if the input sequence is obtained either as samples from a white noise process or as a suitably designed PRBS, asymptotically, the reduced form parameters may be estimated from:

$$\hat{\underline{Y}} = \frac{1}{\lambda_u^2 N} U' \underline{Y} \quad (3.66)^*$$

and the resulting 2SLS estimator is given by (3.49).

4. If the model (3.2) does not contain any exogenous variables ($b_i = 0$; $i = 0, 1, \dots, n$) i.e., if the process is force-free, the identification techniques discussed in this section cannot be used because the reduced form parameters cannot be estimated.

3.5 Results

Simulation results presented in this section attempt to bring out the comparative performance of the following four estimators:

$$\text{OLS} \quad \hat{\underline{\beta}} = (M'M)^{-1} M' \underline{Y}$$

$$\text{ILS} \quad \hat{\underline{\beta}} = (U_p'M)^{-1} U_p' \underline{Y}; \quad U_p = N \times p$$

$$\text{S2SLS} \quad \hat{\underline{\beta}} = (M'UU'M)^{-1} M'UU' \underline{Y}; \quad U = N \times s \quad s > p$$

$$\text{2SLS} \quad \hat{\underline{\beta}} = [M'U(U'U)^{-1}U'M]^{-1} M'U(U'U)^{-1}U' \underline{Y}; \quad U = N \times s, \quad s > p.$$

* This result is used extensively in correlation experiments.

To avoid any matrix inversion routines, recursive forms of these estimators were used. Recursive form for an estimator that has the general structure

$$\hat{\underline{\beta}} = (\underline{Z}'\underline{M})^{-1} \underline{Z}' \underline{y}$$

has been presented in the next chapter (Eqns. (3.14) - (4.18) and the OLS and ILS estimators can easily be converted to this form. The last two estimators (S2SLS and 2SLS) involve two stage procedures. The first stage consists of estimating the reduced form parameters $\underline{y}$ and in the second stage the $\underline{M}'$ matrix in the OLS estimator is replaced by a corresponding matrix which contains the fitted outputs (from the first stage) of the reduced form model. The difference between S2SLS and the 2SLS estimator results from the way the reduced form parameters are estimated. Thus in the S2SLS technique, the reduced form parameters are estimated as the simple average

$$\hat{\underline{y}} = \frac{1}{\lambda_u^2} (\underline{U}' \underline{y}) \quad (3.67)$$

whereas in the 2SLS technique, these parameters are estimated by

$$\hat{\underline{y}} = (\underline{U}'\underline{U})^{-1} \underline{U}' \underline{y} \quad (3.68)$$

If the input sequence is so designed that $\text{p.lim}_{N \rightarrow \infty} (1/N \underline{U}'\underline{U})$ approximates an identity matrix, for large sample size, the parameter estimates obtained from (3.67) and (3.68) will be the same.

Both (3.67) and (3.68) can be reduced to a recursive form and it is needless to say that (3.67) provides substantial saving in computing time and storage and is much simpler to program.

Once the reduced form parameters have been estimated, we can replace the matrix 'M' in the OLS estimator by a matrix that contains the fitted outputs instead of the measured outputs, and the structural parameters β can be estimated recursively.

The simulation experiments were conducted on the XDS-SIGMA 7 digital machine at the Carleton University Computing Centre and the relevant details are given below:

Number of Trials = 20

Maximum Number of Input-Output Data Pairs = 1,000

Input Signals: PRBS (1,1,127) and PRBS (1,1,255) where PRBS (a,t,T) implies a Pseudo Random Binary Sequence of amplitude 'a', clock time 't' and period 'T' — the last two being measured in terms of the sampling rate.

Initial Conditions: At each trial, the recursive estimator under test was primed by a recursive OLS estimator, operating on the first 25 data pairs, the latter being initiated by $\hat{\beta}_{LS}(0) = \underline{0}$ and $C(0) = 10XI$.

Models: The estimators were tested on the following two models:

Model (A) Second Order

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+2) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -0.7 & 1.5 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + \begin{bmatrix} 2.5 \\ 4.05 \end{bmatrix} u(k) + \begin{bmatrix} 0.5 \\ 0.25 \end{bmatrix} e(k)$$

$$y(k) = x_1(k) + u(k) + e(k)$$

$$e(k) \sim \text{N.I.D. } (0,1)$$

$$u(k) \sim \text{PRBS } (1,1,127)$$

The state vector in the above Levy's Canonical form consists of the phase variables so that the corresponding difference equation model is given by

$$y(k) = u(k) + u(k-1) + u(k-2) + 1.5 y(k-1) - 0.7 y(k-2) + e(k) - e(k-1) + 0.2 e(k-2)$$

$$e(k) \sim \text{N.I.D. } (0,1)$$

$$u(k) \sim \text{PRBS } (1,1,127)$$

This model has been used by Mayne³⁴ and the poles of the Pulse Transfer Function are located at

$$p_1 = 0.75 + j 0.37$$

$$p_1^* = 0.75 - j 0.37$$

Model (B) Fourth Order

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ x_3(k+1) \\ x_4(k+1) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -0.656 & 0.784 & -0.18 & 1.0 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \\ x_4(k) \end{bmatrix} + \begin{bmatrix} 1.00 \\ 1.00 \\ 0.82 \\ 1.424 \end{bmatrix} u(k) + \begin{bmatrix} 0.2957 \\ 0.6239 \\ 1.251 \\ 0.7854 \end{bmatrix} e(k)$$

$$y(k) = x_1(k) + e(k)$$

$$e(k) \sim \text{N.I.D. } (0,1)$$

$$u(k) \sim \text{PRBS } (1,1,255)$$

The corresponding difference equation model is given by

$$y(k) = u(k-1) + y(k-1) - 0.18 y(k-2) + 0.784 y(k-3) - 0.656 y(k-4) + \\ + e(k) - 0.704 e(k-1) + 0.508 e(k-2) - 0.104 e(k-3) + 0.071 e(k-4)$$

$$e(k) \sim \text{N.I.D. } (0,1)$$

$$u(k) \sim \text{PRBS } (1,1,255)$$

and its poles are located at

$$p_1 = 0.9 + j 0.1$$

$$p_1^* = 0.9 - j 0.1$$

$$p_2 = -0.4 + j 0.8$$

$$p_2^* = -0.4 - j 0.8.$$

This model has been used by Lee⁸ and Mehra¹¹ and the P.T.F. $1/A(q^{-1})$ represents a hypothetical missile with the short period and the first bending moment included.

Thus the dimension of $\underline{\beta}$ in both these models is 5×1 and for the S2SLS and the 2SLS estimators the dimension of $\underline{y}$, the reduced form parameter vector was taken as 25×1 , i.e., the dimension of the U matrix in these estimators is $1,000 \times 25$.

Tables 3.1, 3.2: Results in these tables pertain to 1,000 data pairs, $\bar{\hat{\beta}}_i$ implies the average value of the parameter over 20 trials and RMSE indicates the corresponding sample Root Mean Squared Error.

Figures 3.1 to 3.4: To provide an idea of the overall convergence properties of the estimators, the following statistics have been plotted as a function of the number of input-output data pairs:

$$1. \text{ Normalized Total (Bias)}^2 = \frac{||\underline{\beta} - \hat{\underline{\beta}}(k)||^2}{||\underline{\beta} - \hat{\underline{\beta}}_{LS}(25)||^2}$$

$$2. \text{ Normalized Total MSE} = \frac{||\underline{\beta} - \hat{\underline{\beta}}(k)||^2}{||\underline{\beta} - \hat{\underline{\beta}}_{LS}(25)||^2}$$

The Normalizing Factor is the value of the corresponding statistic for the Initial Estimates which at each trial are obtained by using OLS on the first 25 Data Pairs.

The models, initial conditions, etc. detailed above have also been used for simulation studies at the end of the next chapter.

3.6 Concluding Remarks

In this chapter, some standard techniques used for parameter estimation in econometrics have been presented. Though the 2SLS estimator is very popular in econometrics, in its classical form, it has seldom been applied for the identification of engineering systems. Its principal drawback lies in the fact that it is computationally very inefficient (Table 3.1).

The author has proposed the Simplified Two Stage Least Squares (S2SLS) estimator which gives the same results as the actual 2SLS estimator when the input sequence is so designed that the matrix $p.\lim_{N \rightarrow \infty} [1/N U'U]$ approximates an identity matrix (Table 3.3). This estimator is computationally far more efficient as compared to the 2SLS estimator (Table 3.1) and it is well suited for identification of engineering systems, where the choice of the input signal is generally open.

From the results presented in the preceeding section, inconsistency of the OLS estimator and the lower asymptotic efficiency of the ILS estimator (compared to S2SLS and 2SLS estimators) is evident. The ILS estimator, however, is consistent (Figs. 3.1, 3.3).

MODEL (A) SECOND ORDER

OF DATA PAIRS = 1,000

OF TRIALS = 20

ESTIMATOR	$\hat{\beta}_1$	RMSE	$\hat{\beta}_2$	RMSE	$\hat{\beta}_3$	RMSE	$\hat{\beta}_4$	RMSE	$\hat{\beta}_5$	RMSE	EST. TIME 1000 POINTS
True Values	1.000	-	1.000	-	1.000	-	1.500	-	-0.700	-	-
Initial Estimates	0.798	0.270	1.329	0.430	1.143	0.360	1.415	0.100	-0.623	0.100	-
OLS	1.022	0.033	1.128	0.135	1.152	0.165	1.389	0.110	-0.599	0.100	9 secs.
ILS	1.023	0.034	1.004	0.035	0.958	0.100	1.512	0.040	-0.715	0.045	9 secs.
S2SLS	1.023	0.035	1.017	0.033	0.975	0.070	1.499	0.010	-0.698	0.007	12 secs.
2SLS	1.023	0.035	1.017	0.033	0.975	0.070	1.499	0.010	-0.698	0.007	90 secs.

TABLE 3.1. Comparison of Estimates Obtained from OLS, ILS, S2SLS and 2SLS Estimators - Model (A) Second Order

MODEL (B) FOURTH ORDER
 # OF DATA PAIRS = 1:000
 # OF TRIALS = 20

ESTIMATOR	$\hat{\beta}_1$	RMSE	$\hat{\beta}_2$	RMSE	$\hat{\beta}_3$	RMSE	$\hat{\beta}_4$	RMSE	$\hat{\beta}_5$	RMSE	EST. TIME 1000 POINTS
True Values	1.000	-	1.000	-	-0.180	-	0.784	-	-0.656	-	-
Initial Estimates	1.011	0.195	0.719	0.325	0.117	0.340	0.645	0.190	-0.489	0.240	-
OLS	0.985	0.035	0.787	0.215	0.077	0.260	0.674	0.115	-0.591	0.068	9 secs.
ILS	0.987	0.038	1.016	0.050	-0.197	0.060	0.787	0.062	-0.651	0.085	9 secs.
S2SIS	0.988	0.035	0.994	0.028	-0.173	0.040	0.788	0.031	-0.662	0.024	12 secs.

TABLE 3.2. Comparison of Estimates Obtained from OLS, ILS and S2SIS Estimators -
 Model (B) Fourth Order

MODEL (A) SECOND ORDER, # OF DATA PAIRS = 1,000

TRUE VALUES	-1.0000		1.0000		1.0000		1.5000		-0.7000	
TRIAL NUMBER	2SLS	S2SLS	2SLS	S2SLS	2SLS	S2SLS	2SLS	S2SLS	2SLS	S2SLS
1	1.0330	1.0330	1.0013	1.0014	0.9849	0.9851	1.4990	1.4989	-0.6998	-0.6997
2	1.0330	1.0331	0.9856	0.9856	1.0111	1.0112	1.4995	1.4996	-0.7000	-0.7001
3	0.9962	0.9961	1.0628	1.0628	0.9211	0.9210	1.5077	1.5077	-0.6063	-0.7064
4	1.0293	1.0293	1.0326	1.0322	0.8951	0.8945	1.5083	1.5087	-0.7046	-0.7050
5	1.0030	1.0030	1.0395	1.0395	0.9319	0.9319	1.5042	1.5043	-0.7029	-0.7029
6	0.9720	0.9720	1.0603	1.0608	0.9648	0.9655	1.4940	1.4935	-0.6928	-0.6925
7	1.0290	1.0289	1.0048	1.0052	1.0582	1.0587	1.4884	1.4880	-0.6882	-0.6877
8	1.0491	1.0491	1.0243	1.0245	0.9508	0.9510	1.4885	1.4883	-0.6870	-0.6868
9	1.0430	1.0430	1.0299	1.0302	0.8588	0.8594	1.5148	1.5144	-0.7109	-0.7106
10	1.0496	1.0496	0.9522	0.9519	1.0549	1.0545	1.4943	1.4946	-0.6954	-0.6957

TABLE 3.3. Comparison of Estimates Obtained from 2SLS and S2SLS Estimators - Model (A), Second Order

Fig. 3.1. Convergence of $(\text{Bias})^2$ for OLS, ILS, S2SLS & 2SLS Estimators
- Model (A) Second Order

MODEL (A) SECOND ORDER NORMALIZING FACTOR = 0.412

Fig. 3.2. Convergence of MSE for OLS, ILS, S2SLS & 2SLS Estimators
- Model (A) Second Order

Fig. 3.3. Convergence of (Bias)² for OLS, ILS & S2SLS Estimators
- Model (B) Fourth Order

Fig. 3.4. Convergence of MSE for OLS, ILS & S2SLS Estimators
- Model (B) Fourth Order

CHAPTER 4

INSTRUMENTAL VARIABLE ESTIMATORS

4.1 Introduction to the Instrumental Variable Technique

The Instrumental Variable (IV) technique which provides consistent estimates of the structural parameters in model (3.2) was originally proposed by Riersol²⁹ and a definitive analysis of its properties for estimation of economic relationships was carried out by Saragan³⁰. In the engineering context, it was first applied by Joseph, Lewis and Tou³¹ (implicitly) and since has been discussed by Lacoss³², Wong and Polak³³, Mayne³⁴, Rowe³⁵, Young³⁶, Pandya^{5,6}, etc.

The technique is conceptually simple and straightforward in that it aims at removing the source of inconsistency in OLS estimators in a direct manner. The technique is also general and powerful, in that the consistent estimators of $\underline{\beta}$ discussed so far, can be interpreted as IV estimators and consistent estimators can be formulated for the parameters of a force-free model using the IV technique⁶.

The classical IV estimator for the parameter vector in the model:

$$\underline{y} = M\underline{\beta} + \underline{v} \quad (4.1)$$

$$e(\underline{v}) = \underline{0}; \quad E(\underline{v}\underline{v}') = \Omega \neq \lambda^2 \underline{v}'\underline{I} \quad (4.2)$$

$$\text{p.lim}_{N \rightarrow \infty} \left[\frac{1}{N} M' \underline{v} \right] \neq \underline{0} \quad (4.3)$$

is obtained by using a $p \times N$ transformation Z' which satisfies the conditions:

$$\text{p.lim}_{N \rightarrow \infty} \left[\frac{1}{N} Z' \underline{v} \right] = \underline{0} , \quad (4.4)$$

and

$$\text{p.lim}_{N \rightarrow \infty} \left[\frac{1}{N} Z' M \right] \text{ exists as a non-singular matrix,} \quad (4.5)$$

leading to the transformed model:

$$Z' \underline{y} = Z' M \underline{\beta} + Z' \underline{v}, \quad (4.6)$$

and in view of (4.4) and (4.5), to the IV estimator:

$$\hat{\underline{\beta}}_{IV} = (Z' M)^{-1} Z' \underline{y}, \quad (4.7)$$

where the $N \times p$ matrix Z is called the Instrumental Variable (IV) matrix and its elements the Instrumental Variables or simply Instruments.

Conditions (4.4) and (4.5) imposed on the IV matrix imply that the elements in the corresponding rows of Z and $\underline{v}$ be minimally correlated and the elements in the corresponding rows of Z and M be maximally correlated. Since

$$\text{p.lim}_{N \rightarrow \infty} \left[\hat{\underline{\beta}}_{IV} \right] = \underline{\beta} + \text{p.lim}_{N \rightarrow \infty} \left[\frac{1}{N} Z' M \right]^{-1} \text{p.lim}_{N \rightarrow \infty} \left[\frac{1}{N} Z' \underline{v} \right] \quad (4.8)$$

conditions (4.4) and (4.5) ensure the consistency of $\hat{\underline{\beta}}_{IV}$.

The residuals obtained from the IV estimator are given by

$$\hat{\underline{v}} = (\underline{y} - \hat{\underline{y}}) = (\underline{y} - M\hat{\underline{\beta}}_{IV}) \quad (4.9)$$

and since

$$Z'(\underline{y} - M\hat{\underline{\beta}}_{IV}) = Z'(\underline{y} - M(Z'M)^{-1}Z'\underline{y}) = \underline{0} \quad (4.10)$$

it follows that the error vector $\hat{\underline{v}}$ is orthogonal to the linear subspace generated by the columns of the IV matrix Z . In geometric terms, therefore, the IV estimates are obtained by taking a projection from the observation vector $\underline{y}$ on to L_M and orthogonal to L_Z where, L_M and L_Z are, respectively, the linear subspaces generated by the column vectors of the design matrix M and the IV matrix Z . In comparison, the OLS estimates are obtained by projecting $\underline{y}$ onto L_M and orthogonal to L_M . Figure 4.1 illustrates the geometry of the IV estimator for a single parameter ($p = 1$) model under the following conditions³⁷:

- (a) $\underline{\beta} = 1 \times 1$, $M = N \times 1$, $Z = N \times 1$, $\underline{y} = N \times 1$.
- (b) $E(\underline{u}) = \underline{0}$.
- (c) $E(M'\underline{u}) > 0$.
- (d) $E(Z'\underline{v}) = 0$.

Assumption (d) above is somewhat stronger than (4.4) in that it requires that Z be orthogonal to $\underline{v}$ 'on the average', rather than asymptotically.

For the limiting case $E(M'\underline{v}) = 0$, the ellipsoid of concentration (Fig. 4.1) for $\underline{v}$ is a sphere centered at $M\underline{\beta}$ and $\hat{\underline{\beta}}_{LS}$ become consistent.

Considering the partitioned form of model (4.1)

$$\underline{y} = M_1\underline{\beta}_1 + M_2\underline{\beta}_2 + \underline{v} \quad (4.11)$$

Fig. 4.1(a). Geometry of OLS Estimator $E(M'\underline{y}) \equiv 0$

Fig. 4.1(b). Geometry of the IV & OLS Estimator $E(M'\underline{y}) > 0$

in which M_1 contains only input variables and M_2 only output variables, we have

$$M'v = \begin{bmatrix} M_1'v \\ M_2'v \end{bmatrix}$$

and

$$\lim_{N \rightarrow \infty} \left[\frac{1}{N} M'v \right] = \frac{\lim_{N \rightarrow \infty} \left(\frac{1}{N} M_1'v \right)}{\lim_{N \rightarrow \infty} \left(\frac{1}{N} M_2'v \right)} = \begin{bmatrix} 0 \\ \lim_{N \rightarrow \infty} \left(\frac{1}{N} M_2'v \right) \end{bmatrix} \quad (4.12)$$

where the vector above the partition is null because M_1 contains only the input variables $u(k)$ which are uncorrelated with (in fact statistically independent of) the disturbances. This indicates that the first $n+1$ columns of the design matrix M , v.i.z., the submatrix M_1 can be retained as the first $n+1$ columns of the IV matrix Z . The IV matrix Z will therefore have the following structure:

$$\begin{matrix} Z \\ N \times p \end{matrix} = \begin{bmatrix} u(n+1) & u(n) & \dots & u(1) & | & z(n) & \dots & z(1) \\ u(n+2) & u(n+1) & \dots & u(2) & | & z(n+1) & \dots & z(2) \\ \vdots & \vdots & \dots & \vdots & | & \vdots & \dots & \vdots \\ u(n+N) & u(n+N-1) & \dots & u(N) & | & z(n+N-1) & \dots & z(N) \end{bmatrix} \quad (4.13)$$

and it has been established^{33,36} that the IV estimator (4.7) has the following recursive form:

$$\hat{\underline{\beta}}_{IV}(k+1) = \hat{\underline{\beta}}_{IV}(k) + \frac{C(k) \underline{z}(k+1)(y(k+1) - \underline{m}'(k+1) \hat{\underline{\beta}}_{IV}(k))}{1 + \underline{m}'(k+1) C(k) \underline{z}(k+1)} \quad (4.14)$$

$$C(k+1) = C(k) - \frac{C(k) \underline{z}(k+1) \underline{m}'(k+1) C(k)}{1 + \underline{m}'(k+1) C(k) \underline{z}(k+1)} \quad (4.15)$$

where

$$\underset{(p \times p)}{C(k)} \triangleq (Z'M)^{-1} \text{ at sample \#k} \quad (4.16)$$

$$\underset{(p \times 1)}{\underline{m}(k+1)} = [u(k+1) \ u(k) \ \dots \ u(k-n+1) \mid y(k) \ \dots \ y(k-n+1)]' \quad (4.17)$$

$$\underset{(p \times 1)}{\underline{z}(k+1)} = [u(k+1) \ u(k) \ \dots \ u(k-n+1) \mid z(k) \ \dots \ z(k-n+1)]' \quad (4.18)$$

The recursive form of the estimator has the obvious advantage in that it avoids any matrix inversions and permits updating of the parameters in real time. In the present context the latter advantage is of much greater significance because it not only implies that the problem of estimating slowly time-varying parameters can easily be handled but, if necessary, the instrumental variables may also be generated in real time. Thus, if the recursive form of the IV estimator is used, it is not essential that all the elements of the IV matrix Z be known a priori.

4.2 Choice of Instrumental Variables

One of the basic problems associated with the IV estimators for structural parameters is the choice of a suitable set of instrumental variables. Since most system models encountered in practice are over-

identified, there is more than one set of instrumental variables that satisfy conditions (4.4) and (4.5). Obviously we would wish to choose the set which provides estimates with the lowest asymptotic variance.

Under the following assumptions:

1. Disturbance vector $\underline{v}$ in model (4.1) is identically and normally distributed with mean $\underline{0}$ and variance-covariance matrix Ω ,
2. Probability limit of $[1/N \underline{Z}'\underline{M}]$ exists as a non-stochastic and non-singular matrix, and
3. The elements of the IV matrix $\underline{Z}$ are either deterministic or they are statistically independent of the elements of $\underline{v}$,

the asymptotic distribution of the IV estimator for $\underline{\beta}$ is derived in Appendix B and is given by:

$$\lim_{N \rightarrow \infty} \sqrt{N} [\hat{\underline{\beta}}_{IV} - \underline{\beta}] \sim \text{N.D.} (\underline{0}, \psi_{\hat{\underline{\beta}}}) \quad (4.19)$$

where

$$\psi_{\hat{\underline{\beta}}} = \text{p.lim}_{N \rightarrow \infty} \left[\left(\frac{1}{N} \underline{Z}'\underline{M} \right)^{-1} \left(\frac{1}{N} \underline{Z}'\underline{\Omega}\underline{Z} \right) \left(\frac{1}{N} \underline{M}'\underline{Z} \right)^{-1} \right] \quad (4.20)$$

so that for sufficiently large samples or asymptotically we have

$$E[\hat{\underline{\beta}}_{IV}] = \underline{\beta} \quad (4.21)$$

$$\text{covar} [\hat{\underline{\beta}}_{IV}] = (\underline{Z}'\underline{M})^{-1} (\underline{Z}'\underline{\Omega}\underline{Z}) (\underline{M}'\underline{Z})^{-1} \quad (4.22)$$

A measure for the dispersion of the estimates around its true value is provided by the determinant of the asymptotic variance-covariance matrix³⁸, which is generally referred as the 'Generalized Variance' of $\hat{\beta}$, i.e., asymptotically

$$G.Var [\hat{\beta}_{IV}] = |(Z'M)^{-1}(Z'\Omega Z)(M'Z)^{-1}| \quad (4.23)$$

and we wish to choose the Z that will minimize this measure of dispersion.

If the noise covariance matrix Ω were known the obvious choice for Z would be $\Omega^{-1}M$ which leads to the GLS estimator

$$\hat{\beta}_{GLS} = (M'\Omega^{-1}M)^{-1} M'\Omega^{-1} y$$

with the asymptotic covariance matrix

$$Covar [\hat{\beta}_{GLS}] = (M'\Omega^{-1}M)^{-1}$$

Since the noise covariance matrix Ω is unknown, Wong¹⁸ has suggested that a consistent estimate of Ω may be used in its place and the resulting estimator is optimal in the sense that it provides the lowest Generalized Variance of all the available IV estimators. However, it is evident from Eqn. (2.12) that the noise covariance matrix Ω is dependent on the elements of the unknown state-transition matrix F or on the parameters $a_1, a_2, \dots, a_n$ in model (3.2), that we are actually trying to estimate. Thus at this stage, it is unrealistic to assume that even

a consistent estimate of Ω is known. Under this contingency, it is difficult, if not impossible, to find an IV matrix that will lead to the lowest possible generalized variance amongst all the available IV estimators, operating under the assumption that Ω is unknown.

The situation is somewhat analogous to the case of a regression model that satisfies all the conditions of the Gauss Markov theorem except that the disturbances are serially correlated with a completely unknown covariance structure. The only option available in this case is to pretend that all the conditions of the Gauss Markov theorem are satisfied and use the OLS estimator. This is equivalent to minimizing the simple sum of squared errors (Eqn. (3.29)) as opposed to the sum of squared errors appropriately weighted by the noise covariance terms (Eqn. (3.28)) that leads to the GLS estimator. The resulting OLS estimates are unbiased and consistent, though not efficient. The actual loss in efficiency is determined by the degree of serial correlation of the disturbances.

In the present context also, we have a similar option, viz. we can 'pretend' that the disturbances are independently and identically distributed - though they are correlated with some of the regressors so that straight OLS is inappropriate, i.e.,

$$E[\underline{v} \underline{v}'] = \lambda^2 \underline{I}, \quad E[\underline{m}(k) v(k)] \neq 0 \quad (4.24)$$

and instead of minimizing the asymptotic generalized variance (Eqn. (4.23)) we are content to minimize the determinant:

$$\tilde{G} \text{ Var } [\hat{\beta}_{IV}] = \lambda_v^2 |(Z'M)^{-1}(Z'Z)(M'Z)^{-1}| \quad (4.25)$$

which is obtained by substituting (4.24) in (4.23).

This procedure, of course, does not lead to an IV estimator that is asymptotically efficient under general conditions. However, as we shall presently see, it does provide asymptotically efficient estimates under certain special conditions on the model and in general will provide consistent estimates that are reasonably efficient in most practical situations. Assuming that the elements of Z and M are measured as deviations from their sample means, (4.25) may be written in terms of the (sample) coefficient of vector correlation³⁸ between Z and M ($\rho_{Z,M}$) as

$$\tilde{G} \text{ Var } [\hat{\beta}_{IV}] = \frac{1}{|M'M| \rho_{(Z,M)}} \quad (4.26)$$

where

$$\rho_{(Z,M)} \triangleq \frac{|(Z'M)(Z'Z)^{-1}(M'Z)|}{|M'M|} \quad (4.27)$$

Since $|M'M|$ is invariant under any choice of Z, we must choose the IV matrix Z such that it maximizes $\rho_{(Z,M)}$ and at the same time satisfies the conditions (4.4) and (4.5). In the light of the last two conditions, it is evident that the instrumental variables should have an influence on the measured output $y(k)$ and at the same time they should be minimally correlated with the disturbance vector. This, of course, leaves us with the exogenous (or input) variables in the model or some suitable function of these input variables.

In the reduced form (or impulse response) model:

$$\underline{y} = U \underline{\gamma} + \underline{w}$$

the necessary information on the noise-free part of $\underline{y}$ is contained in the $N \times s$ matrix U , whose elements are also statistically independent of the disturbances in the model. Dhrymes³⁸ has shown that the coefficient of vector correlation between U and M provides an upper bound for

$\rho(Z, M)$ i.e.,

$$\rho(U, M) \geq \rho(Z, M) \quad \forall Z \in S \quad (4.28)$$

where S is the set of all potential IV matrices that satisfy (4.4) and (4.5).

If U is used as an IV matrix we get the over-identified system of normal equations:

$$\begin{matrix} U' \underline{y} \\ (s \times 1) \end{matrix} = \begin{matrix} U' M \hat{\underline{\beta}} \\ (s \times 1) \end{matrix} \quad (4.29)$$

and a solution for $\hat{\underline{\beta}}$ is given by

$$\hat{\underline{\beta}} = (M' U U' M)^{-1} M' U U' \underline{y} \quad (4.30)$$

which is identical to the S2SLS estimator (3.49) that was discussed in Section 3.4.

Alternatively, we may look for a suitable linear transformation of U that satisfies the following conditions:

$$(a) \quad Z^* = UA \quad \text{is a } N \times p \text{ matrix}$$

$$\text{and} \quad (b) \quad \rho(Z^*, M) = \rho(U, M)$$

The required transformation is given by

$$A = (U'U)^{-1}U'M$$

because

$$Z^* = U(U'M)^{-1}U'M \quad \text{is an } N \times p \text{ matrix}$$

and

$$\begin{aligned} \rho(Z^*, M) &= \frac{|(Z^{*'}M)(Z^{*'}Z)^{-1}(M'Z^*)|}{|M'M|} \\ &= \frac{|(M'U(U'U)^{-1}U'M)(M'U(U'U)^{-1}U'M)^{-1}(M'U(U'U)^{-1}U'M)|}{|M'M|} \\ &= \frac{|(M'U)(U'U)^{-1}(U'M)|}{|M'M|} \\ &= \rho(U, M) \end{aligned}$$

The IV estimator corresponding to the choice

$$Z = Z^* = U(U'U)^{-1}U'M. \quad (4.31)$$

is given by

$$\hat{\beta}_{IV} = (M'U(U'U)^{-1}U'M)^{-1}M'U(U'U)^{-1}U'y \quad (4.32)$$

which is identical to the 2SLS estimator (3.62). As already pointed out in Section 3.4, if

$$\lim_{N \rightarrow \infty} \left[\frac{1}{N} U'U \right] = \lambda^2_u I$$

i.e., if the input sequence is composed of samples from a zero mean white noise process or a suitable approximation to it (say a properly designed PRBS), estimator (4.32) reduces to the estimator (4.30).

At a first glance, estimator (4.32) appears to be even more complicated than (4.30). However, if we recall from Section 3.4, that

$$(i) \quad (U'U)^{-1} U'Y = \hat{Y}$$

$$(ii) \quad M = [M_1 \mid M_2]$$

$$(iii) \quad M_2 = [y_1 \mid y_2 \mid \dots \mid y_n]$$

so that

$$(U'U)^{-1} U'M_2 = [\hat{Y}^1 \mid \hat{Y}^2 \mid \dots \mid \hat{Y}^n]$$

and

$$(iv) \quad (U'U)^{-1} U'M_1 = M_1$$

it can be seen that

$$\begin{aligned} Z^* &= U(U'U)^{-1} U'M \\ &= [U(U'U)^{-1} U'M_1 \mid U(U'U)^{-1} U'M_2] \\ &= [M_1 \mid U\hat{Y}^1 \mid U\hat{Y}^2 \mid \dots \mid U\hat{Y}^n] \\ &= [M_1 \mid \hat{X}_2] \end{aligned}$$

or

$$Z^* = \left[\begin{array}{cccc|cc} u(n+1) & u(n) & \dots & u(1) & \hat{x}(n) & \dots & \hat{x}(1) \\ u(n+2) & u(n+1) & \dots & u(2) & \hat{x}(n+1) & \dots & \hat{x}(2) \\ \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ u(n+N) & u(n+N-1) & \dots & u(N) & \hat{x}(n+N-1) & \dots & \hat{x}(N) \end{array} \right] \quad (4.33)$$

i.e., the submatrix $\hat{X}_2$ is obtained from M_2 by replacing the lagged endogenous variables $y(k)$; $k = 1, 2, \dots, n+N-1$ by the estimates of their noise-free components which are obtained by fitting the reduced form (or the impulse response) model. In other words, $\hat{x}(k)$; $k = 1, 2, \dots, n+N-1$ are the estimates of the 'noise-free process output' $x(k)$, (Fig. 2.2). The submatrix M_1 which contains input variables only, however, remains unchanged.

From the discussion of the 2SLS estimator in Section 3.4 it follows that the resulting estimates of the structural parameters β are consistent and asymptotically efficient if the process is disturbed only by additive white noise at the output. In the absence of this condition the 2SLS estimator will not be asymptotically efficient - though it will continue to be consistent. The actual loss in asymptotic efficiency will be determined by the variance-covariance matrix of $\underline{w}$, the effective disturbance vector in the reduced form model. Since the elements of $\underline{w}$ are generated by filtering $e(k)$ through the linear system characterized by the Pulse Transfer Function $C(q^{-1})/A(q^{-1})$ (Eqns. (3.33) - (3.34), the degree of serial correlation between $w(k)$ and $w(k-j)$; $j = 0, 1, 2, \dots$ is de-

terminated by the location of the poles and zeros of $C(q^{-1})/A(q^{-1})$ within the origin centered unit circle. If the zeros of $C(q^{-1})/A(q^{-1})$ are in close proximity of its poles, i.e., if the major source of noise is the additive output noise, we can expect the 2SLS estimator to provide consistent estimates of the process parameters that have reasonably good asymptotic efficiency. If the additive output noise is not the major source of disturbance in the model and further, if the poles of $C(q^{-1})/A(q^{-1})$ (or the characteristic roots of $A(q^{-1}) = 0$) are close to the unit circle, the 2SLS estimator could become very inefficient.

4.3 Some Practical IV Estimators

In this section some of the existing IV estimators are reviewed. A recursive approximation to the 2SLS estimator is proposed which is not only computationally efficient but whose statistical properties (as judged by simulation experiments) compare favorably with the existing recursive IV estimators.

The difference equation model under discussion is

$$y(k) = b_0 u(k) + b_1 u(k-1) + \dots + b_n u(k-n) - a_1 y(k-1) - a_2 y(k-2) - \dots - a_n y(k-n) + e(k) + c_1 e(k-1) + \dots + c_n e(k-n)$$

or

$$y(k) = \underline{m}'(k) \underline{\beta} + v(k); \quad b = 1, 2, \dots, n+N \quad (4.34)$$

where

$$\underline{m}(k) = [u(k)u(k-1) \dots u(k-n) \mid y(k-1) \dots y(k-n)]' \quad (4.35)$$

is the design vector at sampling instant k .

The disturbance term at sampling instant k

$$v(k) = e(k) + c_1 e(k-1) + \dots + c_n e(k-n); \quad e(k) \sim \text{N.I.D.}(0, \lambda^2)$$

is a moving average process of order n , so that

$$\begin{aligned} \text{(i)} \quad E[v(k)] &= 0; \quad E[v(k)v(k+j)] \neq 0, \quad j \leq n \\ &= 0, \quad j > n \end{aligned} \quad (4.36)$$

$$\text{(ii)} \quad E[v(k)u(k+j)] = 0, \quad \forall j \quad (4.37)$$

$$\begin{aligned} \text{(iii)} \quad E[v(k)y(k-j)] &\neq 0 \quad j \leq n \\ &= 0 \quad j > n \end{aligned} \quad (4.38)$$

$$\text{and} \quad \text{(iv)} \quad E[v(k)x(k+j)] = 0 \quad \forall j \quad (4.39)$$

where $x(k)$ is the noise-free process output.

From the discussion in the preceeding two sections it is evident that if we wish to use an IV estimator for $\underline{\beta}$, we need at each sampling instant k , a $p \times 1$ IV vector $\underline{z}(k)$, whose elements are minimally correlated with the disturbance $v(k)$ and maximally correlated with the design vector $\underline{m}(k)$. When these constraints on the IV vector $\underline{z}(k)$ are viewed in the light of Eqns. (4.36) - (4.39) and the fact that instrumental variables are needed for only those elements of the design vector $\underline{m}(k)$ which are correlated with $v(k)$, we have the following candidates for use as instrumental variables:

1. *Suitably lagged values of the input $u(k)$ as instruments*

For this choice, the $p \times 1$ IV vector corresponding to $\underline{m}(k)$ has the form

$$\underline{z}(k) = [u(k)u(k-1) \dots u(k-n) \mid u(k-n-1) \dots u(k-2n)]' \\ k = 2n+1, 2n+2, \dots \quad (4.40)$$

and if the resulting $N \times p$ IV matrix is represented by Z_u the estimator is given by

$$\hat{\underline{\beta}}_{IV_u} = (Z_u' M)^{-1} Z_u' \underline{y} \quad (4.41)$$

As already pointed out, estimator (4.41) is equivalent to the ILS estimator (3.48). The estimator, in general, is inefficient unless the model is just-identified, i.e., all the information on the noise-free part of $y(k)$ is contained in the $p = 2n+1$ past inputs, which constitute the IV vector (4.40).

This estimator has been used by Joseph, Lewis and Tou³¹ and also by Wong and Polak³³.

2. *Suitably lagged values of the output $y(k)$ as instruments.*

For this choice

$$\underline{z}(k) = [u(k)u(k-1) \dots u(k-n) \mid y(k-n-1) \dots y(k-2n)]' \\ k = 2n+1, 2n+2, \dots \quad (4.42)$$

and if the resulting IV matrix is denoted by Z_y , the estimator is given by

$$\hat{\beta}_{IV_y} = (Z_y' M)^{-1} Z_y' y \quad (4.43)$$

This particular choice for the instrumental variables has been suggested by Goldberger¹⁷, and Pandya⁶ has shown that in a force-free model ($b_i = 0$; $i = 0, 1, 2, \dots, n$) the resulting estimator is equivalent to the modified Yule-Walker estimator used by Mehra¹⁴ and Gersch¹⁵. This equivalence is discussed in Section 4.4.

Mayne³⁴ has attempted to improve on the IV estimator (4.43) by considering a two stage procedure, similar to the 2SLS technique discussed in Section 3.4. Considering the partitioned form of the model (Eqn. (4.11)), 2SLS estimator is obtained by first regressing the columns of the submatrix M_2 on the $p \times s$ matrix U which contains only the input variables $u(k)$. In the second stage, the submatrix M_2 is replaced by $\hat{M}_2$, where the elements of $\hat{M}_2$ are obtained from the least squares fit in the first stage. Application of OLS to this modified model leads to 2SLS. The same result can also be obtained by using $[M_1 | \hat{M}_2]$ as the IV matrix in the second stage. In Mayne's method the columns of the submatrix M_2 are not regressed on the matrix U but on a matrix which contains some input variables and some lagged output variables that are minimally correlated with the disturbance. One such matrix, of course, is the IV matrix Z_y derived from Eqn. (4.42) and used in the IV estimator (4.43).

If we replace the matrix U by Z_y in the 2SLS estimator, the resulting estimator is given by:

$$\hat{\underline{\beta}} = [M'Z_y(Z_y'Z_y)^{-1}Z_y'M]^{-1} M'Z_y(Z_y'Z_y)^{-1} Z_y' y \quad (4.44)$$

and if $\text{p.lim}_{N \rightarrow \infty} 1/N [Z_y'M]$ and $\text{p.lim}_{N \rightarrow \infty} 1/N [Z_y'Z_y]$ exist as non-singular matrices, the above estimator reduces to the simple IV estimator (4.43). Thus the extra computational load implicit in the estimator (4.44) provides little improvement over the simple IV estimator (4.43). This fact is borne out by the simulation results presented at the end of this chapter.

3. Estimates of the noise-free process output $x(k)$ as instruments

The IV vector in this case has the form

$$\underline{z}(k) = [u(k)u(k-1) \dots u(k-n) ; \hat{x}(k-1) \dots \hat{x}(k-n)]' \quad (4.45)$$

$k = n+1, n+2, \dots$

and the estimator is given by

$$\hat{\underline{\beta}}_{IV_x} = (Z_x'M)^{-1} Z_x' y \quad (4.46)$$

where Z_x is the IV matrix corresponding to the IV vector (4.45).

It has already been shown that when $\hat{x}(k)$ are obtained by utilizing the OLS estimates of the reduced form parameters (impulse response coefficients), the resultant estimator (4.46) is equivalent to the 2SLS estimator (4.32). Estimator (4.32) involves two separate matrix inversions, viz. $(U'U)^{-1}$ and $(M'U(U'U)^{-1}U'M)^{-1}$. Depending on the relative stability of the process, dimension of $(U'U)$ may become excessively large so that its inversion would entail a substantial computational effort.

These computational difficulties can, however, be avoided if the recursive form of the IV estimator (Eqns. (4.14) - (4.18)) is used. In this case the required instrumental variables $\hat{x}(k)$, $k = 1, 2, \dots, n+N-1$ can also be generated on-line as the output of an auxiliary model which is linear, has the same order as the process model and is subjected to the same input sequence $u(k)$, $k = 1, 2, \dots, n+N$. The questions that still remain to be answered pertain to (a) the initial conditions and (b) the parameters to be used in the auxiliary model.

Since the auxiliary model is linear and if it is assumed to be stable, the effects of the initial conditions become negligible after a finite time, so that a knowledge of exact initial conditions is not critical.

To answer the second question, let us assume that we use any arbitrary set of parameters that ensure the stability of the auxiliary model. Since the output of a linear, stable dynamic system is a function of (a) the parameters, (b) the initial conditions and (c) the inputs, or for a given set of parameters and initial conditions it is a function of the past and present inputs, the output of the auxiliary model will satisfy the conditions (4.4) and (4.5) required of any instrumental variables. Thus, if it was only a question of obtaining another sequence of instrumental variables, any set of parameters that ensure the stability of the auxiliary model would be adequate. For example, Laccos³² has used $\hat{\beta}_{LS}$, the OLS estimates of β as the parameters of the auxiliary model.

However, what we are striving for, is a recursive approximation to the 2SLS estimator, in which the reduced form outputs $\hat{x}(k)$ are un-

biased and consistent estimates of the true 'noise-free' process outputs $x(k)$. We would therefore achieve this objective if the output of the auxiliary model also satisfies the above properties of unbiasedness and consistency. Though it is difficult to ensure the unbiasedness (for small sample size) of $\hat{x}(k)$, it is possible to obtain consistent estimates of the noise-free process output by the simple recursive algorithm shown in Fig. 4.2.

The recursive relations governing the algorithm of Fig. 4.2 are given by

$$\hat{\beta}(k+1) = \hat{\beta}(k) + \frac{C(k) \underline{z}(k+1)(y(k+1) - \underline{m}'(k+1)\hat{\beta}(k))}{1 + \underline{m}'(k+1)C(k) \underline{z}(k+1)} \quad (4.47)$$

$$C(k+1) = C(k) - \frac{C(k) \underline{z}(k+1) \underline{m}'(k+1) C(k)}{1 + \underline{m}'(k+1)C(k) \underline{z}(k+1)} \quad (4.48)$$

$$\underline{m}(k+1) = [u(k+1)u(k) \dots u(k-n+1) \vdots y(k) \dots y(k-n+1)]' \quad (4.49)$$

$$\underline{z}(k+1) = [u(k+1)u(k) \dots u(k-n+1) \vdots \hat{x}(k) \dots \hat{x}(k-n+1)]' \quad (4.50)$$

$$\hat{x}(k) = \underline{z}'(k) \hat{\beta}(k) \quad (4.51)$$

Thus the output of the auxiliary model is generated by the time-varying difference equation (4.51). In terms of the input sequence $u(k)$, the output of the auxiliary model may be represented by

Fig. 4.2. Recursive IV Estimator in Which the IV Sequence is Generated by an Auxiliary Model

$$\hat{x}(k) = f(\hat{\beta}(k), u(1), u(2), \dots, u(k), \hat{x}(0))$$

and if $\hat{\beta}(k)$ are such that they ensure the stability of the auxiliary model for all k , $\hat{x}(k)$; $k = 1, 2, \dots$ are suitable candidates for use as instrumental variables, i.e.; under the above assumption

$$\text{p.lim}_{N \rightarrow \infty} [\hat{\beta}(k)] = \underline{\beta}$$

and since $\hat{x}(k)$ is a continuous function of $\hat{\beta}(k)$ we may invoke Slutsky's theorem to assert that

$$\text{p.lim}_{N \rightarrow \infty} [\hat{x}(k)] = x(k)$$

The results indicate that the slight correlation between the instruments $\hat{x}(k)$ and the disturbance $v(k)$, due to the presence of $\hat{\beta}(k)$ in Eqn. (4.51), has negligible effect. In this context one may note that the condition, $\text{p.lim. } (1/N \underline{Z}'\underline{v}) = 0$, does not imply that the elements of $\underline{Z}$ and $\underline{v}$ are completely uncorrelated. Consistency of the 2SLS estimator, in which the estimates of the reduced form parameters (and hence the instruments) as well as the structural parameters are obtained from the same set of measurements, further confirms this fact.

Wong and Polak³³ and Young³⁶ have used similar algorithms where the instrumental variable sequence is estimated by

$$\hat{x}(k) = \underline{z}'(k) \hat{\beta}(k-r), \quad r > n \quad (4.52)$$

with the argument that $\hat{\beta}(k-r)$, and hence the resulting estimates of $x(k)$ are uncorrelated with the noise terms $e(k), e(k-1), \dots, e(k-n)$, i.e., uncorrelated with the last n measurements $y(k), y(k-1), \dots, y(k-n)$. Use of (4.51) however, greatly simplifies the algorithm from the computational viewpoint and at the same time the net effect on the final estimates of $x(k)$ is minimal because with the increase in the number of data pairs, the dependence of $\hat{\beta}(k)$ on the last n measurements becomes marginal. A further advantage of (4.51) lies in the fact that it involves no 'back tracking' which is undesirable for estimation of time-varying parameters.

A detailed layout for the identification algorithm that utilizes Eqn. (4.51) to estimate $x(k)$ is shown in Fig. 4.3. Unlike the 2SLS estimator, the estimates of $x(k)$ furnished by Eqn. (4.51) or (4.52) have no explicit dependence on the measured output $y(k)$, and hence the convergence of such an algorithm cannot be guaranteed under completely arbitrary initial conditions. For this reason Wong and Polak³³ include a Jury stability test in their algorithm and Young³⁶ suggests the use of a low pass filter to remove any sudden fluctuations in $\hat{\beta}(k)$. The former method of stabilization complicates the algorithm, specially for higher order systems and the latter tends to reduce the rate of convergence. Extensive simulation studies carried out by the author, however, indicate that the simple algorithm of Fig. 4.3 with the initial conditions as specified earlier, is well behaved with a very high degree of probability and no special stabilizing efforts seem to be warranted.

In the present context, it would be pertinent to mention an algorithm proposed by Rowe³⁵ which also utilizes an auxiliary model for

Fig. 4.3. Detailed Structure for the Recursive Approximation to 2SIS Using an Auxiliary Model

generating the required instrumental variables. However, in this recursive algorithm, the auxiliary model is re-initiated at every sampling instant k , by the delayed design vector $\underline{m}(k-n+1)$, i.e., the instruments $\hat{x}(k), \hat{x}(k-1), \dots, \hat{x}(k-n+1)$ required to obtain $\hat{\beta}(k+1)$ from $\hat{\beta}(k)$ are generated as follows:

$$\hat{x}(k-n+1) = \underline{m}'(k-n+1) \hat{\beta}(k)$$

or

$$\begin{aligned} \hat{x}(k-n+1) &= [u(k-n+1) \dots u(k-2n+1) y(k-n) y(k-n-1) \dots y(k-2n+1)] \hat{\beta}(k) \\ \hat{x}(k-n+2) &= [u(k-n+2) \dots u(k-2n+2) \hat{x}(k-n+1) y(k-n) \dots y(k-2n+2)] \hat{\beta}(k) \\ \hat{x}(k-n+3) &= [u(k-n+3) \dots u(k-2n+3) \hat{x}(k-n+2) \hat{x}(k-n+1) \dots y(k-2n+3)] \hat{\beta}(k) \\ &\vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots \\ \hat{x}(k) &= [u(k) \dots u(k-n) \hat{x}(k-1) \hat{x}(k-2) \dots y(k-n)] \hat{\beta}(k) \end{aligned}$$

As opposed to Eqn. (4.51) which is initialized only once in the beginning and which runs concurrently with the recursive estimator for $\hat{\beta}$, i.e., at each sampling instant we have only one output $\hat{x}(k)$ corresponding to the parameter estimates $\hat{\beta}(k)$, Rowe's algorithm is initialized* at each sampling instant and it is required to generate the complete

* Note that Rowe's algorithm also uses the latest estimates $\hat{\beta}(k)$ (and not $\hat{\beta}(k-r)$) to generate the instruments.

set of instruments $\hat{x}(k), \hat{x}(k-1), \dots, \hat{x}(k-n+1)$ corresponding to the parameter estimate $\hat{\beta}(k)$. This continuous re-initialization of the auxiliary model however has the advantage that the sequence of instruments is always tied to the measured outputs $y(k)$ and the likelihood of the IV sequence becoming unbounded is reduced.

Assuming for a moment, that the true initial conditions and the true process parameters were known, then Eqn.(4.51) will generate the true output of the noise-free process whereas Rowe's auxiliary model will generate unbiased (though not consistent) estimates of the noise-free process output.

4.4 Identification of Force-free Models

The model considered in the preceeding sections is characterized by the presence of exogenous or input variables which are measurable without error. Most of the estimators that have been discussed, utilize this fact. Such estimators as ILS, 2SLS, 3SLS or the recursive approximation to the 2SLS presented in the last section become meaningless if the model is force-free, i.e., if $b_i = 0$; $i = 0, 1, 2, \dots, n$ in Eqns. (4.34). Force-free models occur quite frequently in orbit determination of satellites and in Time Series Analysis as Auto-regressive (AR) or Mixed Auto-regressive Moving Average (MARMA) processes.

Thus the model to be identified can be represented as

$$y(k) = -a_1 y(k-1) - a_2 y(k) \dots - a_n y(k-n) + e(k) + c_1 e(k-1) + \dots + c_n e(k-n) \quad (4.53)$$

or

$$y(k) = -a_1 y(k-1) - a_2 y(k-2) \dots - a_n y(k-n) + v(k) \\ k = 1, 2, \dots, n+N \quad (4.54)$$

or

$$\underline{y} = M_2 \underline{\beta}_2 + \underline{v} \quad (4.55)$$

Equation (4.53) represents a MARMA (n,n) process. However, if $c_i = 0$; $i = 1, 2, \dots, n$, Eqn. (4.53) reduces to a pure AR process and in Time Series Analysis, parameters of such a process are generally estimated from the Yule-Walker equations³⁹. For the AR process

$$y(k) = -a_1 y(k-1) - a_2 y(k-2) \dots - a_n y(k-n) + e(k) \quad (4.56) \\ k = 1, 2, \dots, n+N$$

or

$$\underline{y} = M_2 \underline{\beta}_2 + \underline{e}$$

where $y(k)$ are assumed to be zero mean and stationary random variables, the Yule-Walker equations are obtained by multiplying both sides of (4.56) by $y(k-j)$ and taking expectations, i.e.,

$$E[y(k)y(k-j)] = -a_j E[y(k-1)y(k-j)] - \dots - a_n E[y(k-n)y(k-j)] + \\ + E[e(k)y(k-j)] \quad (4.57)$$

Since the last term in (4.57) is zero for $j \neq 0$ we may write it as

$$\sigma_j = -a_1 \sigma_{j-1} - a_2 \sigma_{j-2} - \dots - a_n \sigma_{j-n} \quad (4.58)$$

where

$$\sigma_i \triangleq E[y(k)y(k-i)] \quad (4.59)$$

If j is now allowed to take values from 1 to n and since $\sigma_j = \sigma_{-j}$ we have

$$\begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \vdots \\ \sigma_n \end{bmatrix} = \begin{bmatrix} \sigma_0 & \sigma_1 & \dots & \sigma_{n-1} \\ \sigma_1 & \sigma_0 & \dots & \sigma_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n-1} & \sigma_{n-2} & \dots & \sigma_0 \end{bmatrix} \begin{bmatrix} -a_1 \\ -a_2 \\ \vdots \\ -a_n \end{bmatrix} \quad (4.60)$$

or

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_n \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \dots & \rho_{n-1} \\ \rho_1 & 1 & \dots & \rho_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \dots & 1 & \end{bmatrix} \begin{bmatrix} -a_1 \\ -a_2 \\ \vdots \\ -a_n \end{bmatrix} \quad (4.61)$$

where

$$\rho_i = \rho_{-i} \triangleq \frac{\sigma_i}{\sigma_0}, \quad \rho_0 = 1. \quad (4.62)$$

The set of equations (4.61) are usually referred as the Yule-Walker equations.

In practice, the correlation terms $\rho_1, \rho_2, \dots, \rho_n$ may not be known, in which case the Yule-Walker estimates of $-a_1, -a_2, \dots, -a_n$ are obtained as

$$\begin{bmatrix} -\hat{a}_1 \\ -\hat{a}_2 \\ \vdots \\ -\hat{a}_n \end{bmatrix} = \begin{bmatrix} 1 & \hat{\rho}_1 & \dots & \hat{\rho}_{n-1} \\ \hat{\rho}_1 & 1 & \dots & \hat{\rho}_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\rho}_{n-1} & \hat{\rho}_{n-2} & \dots & 1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \vdots \\ \hat{\rho}_n \end{bmatrix} \quad (4.63)$$

where the autocorrelation $\hat{\rho}_j$ for shift j is consistently estimated by the sample correlation

$$\hat{\rho}_j = \frac{\hat{\sigma}_j}{\hat{\sigma}_0} = \frac{\frac{1}{N} \sum_N y(k)y(k-j)}{\frac{1}{N} \sum_N y(k)y(k)} \quad (4.64)$$

i.e.,

$$\begin{bmatrix} -\hat{a}_1 \\ -\hat{a}_2 \\ \vdots \\ -\hat{a}_n \end{bmatrix} = \begin{bmatrix} \frac{1}{N\hat{\sigma}_0} \sum_N [y(k-1)]^2 & \frac{1}{N\hat{\sigma}_0} \sum_N y(k-1)y(k-2) & \dots & \frac{1}{N\hat{\sigma}_0} \sum_N y(k-1)y(k-n) \\ \vdots & \frac{1}{N\hat{\sigma}_0} \sum_N [y(k-2)]^2 & \dots & \frac{1}{N\hat{\sigma}_0} \sum_N y(k-2)y(k-n) \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \frac{1}{N\hat{\sigma}_0} \sum_N [y(k-n)]^2 \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{N\hat{\sigma}_0} \sum_N y(k)y(k-1) \\ \frac{1}{N\hat{\sigma}_0} \sum_N y(k)y(k-2) \\ \vdots \\ \frac{1}{N\hat{\sigma}_0} \sum_N y(k)y(k-n) \end{bmatrix}$$

or

$$\begin{bmatrix} -\hat{a}_1 \\ -\hat{a}_2 \\ \vdots \\ -\hat{a}_n \end{bmatrix} = \begin{bmatrix} \sum_N [y(k-1)]^2 & \sum_N y(k-1)y(k-2) & \dots & \sum_N y(k-1)y(k-n) \\ \vdots & \sum_N [y(k-2)]^2 & \dots & \sum_N y(k-2)y(k-n) \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \sum_N [y(k-n)]^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_N y(k)y(k-1) \\ \sum_N y(k)y(k-2) \\ \vdots \\ \sum_N y(k)y(k-n) \end{bmatrix}$$

or

$$\hat{\beta}_2 = [M_2' M_2]^{-1} M_2' y \quad (4.65)$$

In other words, the Yule-Walker estimates of β_2 in model (4.55) are nothing but the OLS estimates whose consistency is assured by the Mann and Wald theorem¹⁶.

For the MARMA(n,n) process represented by Eqn. (4.53) Yule-Walker equations (4.61) cannot be derived because the derivation of Eqn. (4.61) would require that $E[v(k)y(k-j)]$ be zero for $j = 1, 2, \dots, n$, which, of course, is not true in this case. Mehra¹⁴ and Gersch¹⁵ use a modified form of Yule-Walker equations which are obtained by multiplying Eqn. (4.54) by $y(k-j)$ and taking the necessary expectations, where j now varies from $n+1$ to $2n$ so that the condition $E[v(k)y(k-j)]$ is satisfied. The modified Yule-Walker equations that result are given by

$$\begin{bmatrix} \sigma_{n+1} \\ \sigma_{n+2} \\ \vdots \\ \sigma_{2n} \end{bmatrix} = \begin{bmatrix} \sigma_n & \sigma_{n-1} & \dots & \sigma_1 \\ \sigma_{n+1} & \sigma_n & \dots & \sigma_2 \\ \vdots & \vdots & & \vdots \\ \sigma_{2n-1} & \sigma_{2n-2} & \dots & \sigma_n \end{bmatrix} \begin{bmatrix} -a_1 \\ -a_2 \\ \vdots \\ -a_n \end{bmatrix} \quad (4.66)$$

or

$$\begin{bmatrix} \rho_{n+1} \\ \rho_{n+2} \\ \vdots \\ \rho_{2n} \end{bmatrix} = \begin{bmatrix} \rho_n & \rho_{n-1} & \dots & \rho_1 \\ \rho_{n+j} & \rho_n & \dots & \rho_2 \\ \vdots & \vdots & & \vdots \\ \rho_{2n-1} & \rho_{2n-2} & \dots & \rho_n \end{bmatrix} \begin{bmatrix} -a_1 \\ -a_2 \\ \vdots \\ -a_n \end{bmatrix} \quad (4.67)$$

where σ_j and ρ_j are defined respectively by Eqns. (4.59) and (4.62).

The $n \times n$ matrix on the right hand side of Eqn. (4.66) or (4.67) is not symmetric and does not have the usual structure of a covariance matrix. However, it is a Hankel matrix and Mehra¹⁴ has shown that it is non-singular. Again the Yule-Walker estimates of β_2 are obtained by replacing the true correlations (that are unknown) in Eqn. (4.67) by the sample correlations, i.e.,

$$\begin{bmatrix} -\hat{a}_1 \\ -\hat{a}_2 \\ \vdots \\ -\hat{a}_n \end{bmatrix} = \begin{bmatrix} \frac{1}{N\hat{\sigma}_0} \sum_N y(k-1)y(k-n-1) & \dots & \frac{1}{N\hat{\sigma}_0} \sum_N y(k-n)y(k-n-1) \\ \frac{1}{N\hat{\sigma}_0} \sum_N y(k-1)y(k-n-1) & \dots & \frac{1}{N\hat{\sigma}_0} \sum_N y(k-n)y(k-n-2) \\ \vdots & & \vdots \\ \frac{1}{N\hat{\sigma}_0} \sum_N y(k-1)y(k-2n) & \dots & \frac{1}{N\hat{\sigma}_0} \sum_N y(k-n)y(k-2n) \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{N\hat{\sigma}_0} \sum_N y(k)y(k-n-1) \\ \frac{1}{N\hat{\sigma}_0} \sum_N y(k)y(k-n-2) \\ \vdots \\ \frac{1}{N\hat{\sigma}_0} \sum_N y(k)y(k-2n) \end{bmatrix} \quad (4.68)$$

which is identical to the estimator

$$\hat{\underline{\beta}}_2 = [\underline{Z}'_y M_2]^{-1} \underline{Z}'_y \underline{y}^* \quad (4.69)$$

where $\underline{Z}_y$ is the IV matrix constructed from the IV vector

$$\underline{z}(k) = [y(k-n-1)y(k-n-2) \dots y(k-2n)]', \quad k = 2n, \dots, N+2n \quad (4.70)$$

Thus the estimators proposed by Gersch and Mehra based on the modified Yule-Walker equations (4.67) are identical to an IV estimator that utilizes suitably lagged values of the regressors as instruments⁶. Under this equivalence, the so-called Yule-Walker estimator (Eqn. (4.68)) can be put into the recursive formulation of Eqns. (4.14) to (4.18) in which $\underline{z}(k)$ is defined by Eqn. (4.70).

The Yule-Walker equations (4.66) were obtained by multiplying Eqn. (4.54) by $y(k-j)$, $j = n+1, n+2, \dots, 2n$ and taking expectations. Mehra¹⁴ has suggested an alternate estimator in which the Yule-Walker equations are obtained by the same procedure as above but with the difference that j is now allowed to vary from $n+1$ to $m > 2n$. Thus, instead of having a set of n normal equations we have $(m-n) > n$ equations:

* This estimator could also have been obtained by putting $M_1 \equiv \underline{0}$; $\hat{\underline{\beta}}_1 \equiv \underline{0}$ in Eqn. (4.42).

$$\begin{bmatrix} \rho_{n+1} \\ \rho_{n+2} \\ \vdots \\ \rho_{2n} \\ \vdots \\ \rho_m \end{bmatrix}_{(m-n) \times 1} = \begin{bmatrix} \rho_n & \rho_{n-1} & \dots & \rho_1 \\ \rho_{n+2} & \rho_n & \dots & \rho_2 \\ \vdots & \vdots & & \vdots \\ \rho_{2n-1} & \rho_{2n-2} & & \rho_n \\ \vdots & \vdots & & \vdots \\ \rho_{m-1} & \rho_{m-2} & & \rho_{m-n} \end{bmatrix}_{(m-n) \times n} \begin{bmatrix} -a_1 \\ -a_2 \\ \vdots \\ -a_n \end{bmatrix}_{n \times 1} \quad (4.71)$$

For estimation of $\underline{\beta}_2$ the true correlations have to be replaced by the sample correlations defined by Eqn. (4.64) and under these conditions it is straight forward to show⁶ that the resulting set of normal equations:

$$\begin{bmatrix} \hat{\rho}_{n+1} \\ \hat{\rho}_{n+2} \\ \vdots \\ \hat{\rho}_m \end{bmatrix} = \begin{bmatrix} \hat{\rho}_n & \hat{\rho}_{n-1} & \dots & \hat{\rho}_1 \\ \hat{\rho}_{n+1} & \hat{\rho}_n & \dots & \hat{\rho}_2 \\ \vdots & \vdots & & \vdots \\ \hat{\rho}_{m-1} & \hat{\rho}_{m-2} & \dots & \hat{\rho}_{m-n} \end{bmatrix} \begin{bmatrix} -\hat{a}_1 \\ -\hat{a}_2 \\ \vdots \\ -\hat{a}_n \end{bmatrix} \quad (4.72)$$

is equivalent to

$$\underline{Z}'_{ya} \underline{y} = \underline{Z}'_{ya} M_2 \hat{\underline{\beta}}_2 \quad (4.73)$$

where the $N \times (m-n)$ IV matrix $\underline{Z}_{ya}$ corresponds to the augmented IV vector:

$$\underline{z}(k) = [y(k-n-1)y(k-n-2) \dots y(k-2n) \dots y(k-m)]' \quad (4.74)$$

A solution for $\hat{\beta}_2$ can now be obtained as

$$\hat{\beta}_2 = [M_2' Z_{ya} Z_{ya}' M_2]^{-1} M_2' Z_{ya} Z_{ya}' \underline{y} \quad (4.75)$$

Estimator (4.75) is similar in structure to the S2SLS estimator (3.47) and the associated computational problems have already been discussed in Chapter 3.

4.5 Results

Simulation studies presented in this section pertain to the IV estimators that fall in the three basic categories discussed in Section 4.3, and the estimators that are suitable for use in a force-free model which were discussed in Section 4.4. Thus the various IV estimators presented here have been grouped as follows:

1. Suitably lagged inputs as Instruments

$$\hat{\beta} \text{ (IV)} = (Z_u' M)^{-1} Z_u' \underline{y} \rightarrow \text{ILS}$$

2. Suitably lagged outputs as Instruments

$$\hat{\beta} \text{ (IV)} - \text{Mayne's Method}$$

$$\hat{\beta} \text{ (IV)} = (Z_y' M)^{-1} Z_y' \underline{y} \rightarrow \text{lagged outputs as IV}$$

3. Estimates of the noise-free process output as Instruments

a. *Instruments obtained by fitting the reduced form model*

$$\hat{\underline{\beta}} = (\mathbf{M}'\mathbf{U}\mathbf{U}'\mathbf{M})^{-1} \mathbf{M}'\mathbf{U}\mathbf{U}' \underline{\mathbf{y}} \rightarrow \text{S2SLS}$$

$$\hat{\underline{\beta}} = (\mathbf{M}'\mathbf{U}(\mathbf{U}'\mathbf{U})^{-1} \mathbf{M}'\mathbf{U}(\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}') \underline{\mathbf{y}} \rightarrow \text{2SLS.}$$

b. *Instruments obtained from an auxiliary dynamic model*

IV Estimator - Pandya

IV Estimator - Wong & Polak (without Jury stability test)

IV Estimator - Rowe

4. Estimators suitable for use in force-free models

$$\hat{\underline{\beta}}_2 = (\mathbf{Z}_y' \mathbf{M}_2)^{-1} \mathbf{Z}_y' \underline{\mathbf{y}} \rightarrow \text{Yule-Walker (1)}$$

$$\hat{\underline{\beta}}_2 = (\mathbf{M}_2' \mathbf{Z}_{ya} \mathbf{Z}_{ya}' \mathbf{M}_2)^{-1} \mathbf{M}_2' \mathbf{Z}_{ya} \mathbf{Z}_{ya}' \underline{\mathbf{y}} \rightarrow \text{Yule-Walker (2)}$$

Estimators under category 1 and 3.a viz., ILS, S2SLS and 2SLS estimators have already been discussed in Chapter 3 and pertinent simulation results were presented in Section 3.5. Since all the above estimators, except the ones in category 4, were tested on the same two models and under exactly the same conditions that were specified in Section 3.5, results pertaining to the ILS and S2SLS estimators have been brought forward from Section 3.5 for cross comparison. Similarly for the sake of comparison, estimates were also obtained from an IV estimator that utilized the true noise-free process output as Instruments.

In the results that follow, estimators in each category were compared with each other and where pertinent cross-comparisons have been included.

As already mentioned, the estimators in categories 1, 2, and 3, were tested on the same two models used in Section 3.5 and these are rewritten here for easy reference.

MODEL (A) Second Order

$$y(k) = u(k) + u(k-1) + u(k-2) + 1.5 y(k-1) - 0.7 y(k-2) + e(k) - e(k-1) \\ + 0.2 e(k-2)$$

$$e(k) \sim \text{N.I.D. } (0,1)$$

$$u(k) \sim \text{PRBS } (1,1,127)$$

MODEL (B) Fourth Order

$$y(k) = u(k-1) + y(k-1) - 0.18 y(k-2) + 0.764 y(k-3) - 0.656 y(k-2) \\ + e(k) - 0.704 e(k-1) + 0.508 e(k-2) - 0.104 e(k-3) \\ + 0.071 e(k-4)$$

$$e(k) \sim \text{N.I.D. } (0,1)$$

$$u(k) \sim \text{PRBS } (1,1,255)$$

The estimators suitable for force-free models (Category 4) were tested on the following Fourth Order model from Mehra¹⁴:

MODEL (C) Fourth Order

MODEL (C) Fourth Order

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+2) \\ x_3(k+1) \\ x_4(k+1) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -0.656 & 0.784 & -0.18 & 1.0 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \\ x_4(k) \end{bmatrix} + \begin{bmatrix} 0.2757 \\ 0.6239 \\ 1.251 \\ 0.7854 \end{bmatrix} e(k)$$

$$y(k) = x_1(k) + e(k)$$

$$e(k) \sim \text{N.I.D. } (0, 2.557)$$

which is equivalent to the following difference equation model:

$$\begin{aligned} y(k) = & y(k-1) - 0.18 y(k-2) + 0.784 y(k-3) - 0.656 y(k-4) \\ & + e(k) - 0.704 e(k-1) + 0.508 e(k-2) - 0.104 e(k-3) + 0.071 e(k-4) \end{aligned}$$

$$e(k) \sim \text{N.I.D. } (0, 2.557)$$

For the Yule-Walker estimator with an augmented IV matrix (Yule-Walker (a)), a Z_{ya} matrix of dimension 1,000x8 was used. All other conditions for the simulation experiments were the same as in Section 3.5.

4.6 Concluding Remarks

A comprehensive study of the Instrumental Variable technique for parameter estimation in linear models has been presented in this chapter. It has been shown that the IV estimator has a structure that is general enough to encompass such well known estimators as the 2SLS and

the Yule-Walker estimators. The power of the IV technique is further enhanced by the fact that it has an exact recursive formulation so that, if necessary, the required Instruments can also be generated recursively. A recursive approximation to the 2SLS estimator is proposed which uses an auxiliary dynamic model to generate the required Instruments -- it is simple in concept, is easy to program and provides results that are comparable, if not better, than those provided by similar, though more complex, estimators found in the existing literature.

For the models used in the simulation experiments and subject to the conditions under which they were tested, we may draw the following inference from the results presented in Section 4.5:

Tables 4.1-2-3, Figs. 4.4-5

The simple IV estimator that uses suitably lagged outputs as Instruments gives better results than Mayne's IV estimator. In fact Mayne's method does not seem to work on the fourth order model (MODEL B) whose poles are closer to (though inside) the origin centred unit circle with one pair of poles to the left of the origin, indicating a larger settling time and a higher natural frequency for the process.

Tables 4.4-5, Figs. 4.6-7-8-9

The on-line approximation to 2SLS estimator proposed by the author gives results that compare favorably with the on-line estimators due to Wong and Polak and Rowe that also use an auxiliary dynamic model to generate the Instruments.

Tables 4.6-7, Figs. 4.10-11-12-13

The on-line estimator proposed by the author is compared with IV estimators that use suitably lagged inputs and lagged outputs as Instruments and the results indicate that the latter two estimators provide estimates with larger deviations from the true values.

Tables 4.8-9, Figs. 4.14-15

The on-line estimator is compared with the S2SLS estimator and and IV estimator that uses the true noise-free process outputs as Instruments and asymptotically the three estimators give essentially the same results.

Tables 4.10, Figs. 4.16-17

The two estimators in Category 4 that are suitable for estimating the parameters of a force-free model are compared and it may be noted that Yule-Walker (2) estimator does not provide any marked improvement over the simple Yule-Walker (1) estimator.

From the results presented in the last section it would appear that little is gained in using more complicated estimators within each of the categories for which comparisons have been made. The cross-comparisons presented indicate that the on-line approximation to the 2SLS estimator (IV-Pandya) is not only simple but provides estimates with low dispersion around the true values of the parameters.

As already pointed out in Section 4.3, one disadvantage that is common to all the on-line estimators (Category 3b) is that their

convergence cannot be guaranteed under completely arbitrary initial conditions. As opposed to the 2SLS estimator, its recursive approximations provide estimates of the noise-free output that has no explicit dependence on the measured output $y(k)$. In the next chapter, an attempt has been made to remove this deficiency by proposing the use of some simple filters.

MODEL (A) SECOND ORDER
 # OF DATA PAIRS = 1.000
 # OF TRIALS = 20

ESTIMATORS	$\hat{\beta}_1$	RMSE	$\hat{\beta}_2$	RMSE	$\hat{\beta}_3$	RMSE	$\hat{\beta}_4$	RMSE	$\hat{\beta}_5$	RMSE	EST. TIME 1000 POINTS
True Values	1.000	-	1.000	-	1.000	-	1.500	-	-0.700	-	
Initial Estimates	0.798	0.270	1.329	0.430	1.143	0.360	1.415	0.100	-0.623	0.100	
OLS	1.022	0.033	1.128	0.135	1.152	0.165	1.389	0.110	-0.599	0.100	9 secs.
IV-Mayne's	1.023	0.035	1.006	0.033	0.957	0.085	1.510	0.024	-0.709	0.022	18 secs.
IV-Lagged Outputs	1.023	0.035	1.013	0.034	0.696	0.065	1.502	0.006	-0.702	0.004	9 secs.

TABLE 4.1. Comparison of IV Estimators Based on Lagged Outputs as Instruments

MODEL (B) FOURTH ORDER
 # OF DATA PAIRS = 1.000
 # OF TRIALS = 20

ESTIMATOR	$\hat{\beta}_1$		$\hat{\beta}_2$		$\hat{\beta}_3$		$\hat{\beta}_4$		$\hat{\beta}_5$		EST. TIME	
	RMSE		RMSE		RMSE		RMSE		RMSE		RMSE	1000 POINTS
True Values	1.000	-	1.000	-	-0.180	-	0.784	-	-0.656	-	-	-
Initial Estimates	1.011	0.195	0.719	0.325	0.117	0.340	0.645	0.190	-0.489	0.240		
OLS	0.985	0.035	0.787	0.215	0.077	0.260	0.674	0.115	-0.591	0.068	9 secs.	
IV-Mayne's	0.991	0.034	1.041	0.160	-0.233	0.275	0.813	0.255	-0.672	0.110	18 secs.	
IV-Lagged Outputs	0.988	0.036	0.994	0.041	-0.173	0.042	0.797	0.067	-0.660	0.023	9 secs.	

TABLE 4.2. Comparison of IV Estimators Based on Lagged Outputs as Instruments -
 Model (B) Fourth Order

MODEL (B) FOURTH ORDER

# OF DATA PAIRS k	NORMALIZED TOTAL (BIAS) ² (N.F. = 0.214)		NORMALIZED TOTAL MSE (N.F. = 0.3485)	
	LAGGED OUTPUTS AS IV	MAYNE'S METHOD	LAGGED OUTPUTS AS IV	MAYNE'S METHOD
100	0.0834	1.3212	0.4105	14.8693
200	0.0425	0.1447	0.1786	1.2667
300	0.0061	0.5229	0.0946	2.5049
400	0.0100	0.0335	0.0746	0.9219
500	0.0106	0.0105	0.0559	1.4520
600	0.0046	0.1037	0.0380	1.1842
700	0.0029	0.0448	0.0303	0.9159
800	0.0030	0.0365	0.0235	0.5377
900	0.0025	0.0647	0.0222	0.9895
1000	0.0028	0.0265	0.0207	0.5174

TABLE 4.3. (Bias)² and MSE Convergence of IV Estimators Based on Lagged Outputs as Instruments - Model (B) Fourth Order

Fig. 4.4. (Bias)² Convergence of IV Estimators based on Lagged Outputs as Instruments - Model (A) Second Order

Fig. 4.5. MSE Convergence of IV Estimators Based on Lagged Outputs as Instruments - Model (A) Second Order

MODEL (A) SECOND ORDER
 # OF DATA PAIRS = 1.000
 # OF TRIALS = 20

ESTIMATORS	$\hat{\beta}_1$	RMSE	$\hat{\beta}_2$	RMSE	$\hat{\beta}_3$	RMSE	$\hat{\beta}_4$	RMSE	$\hat{\beta}_5$	RMSE
True Values	1.000	-	1.000	-	1.000	-	1.500	-	-0.700	-
Initial Estimates	0.798	0.270	1.329	0.430	1.143	0.360	1.415	0.100	-0.623	0.100
OLS	1.022	0.033	1.128	0.135	2.152	0.165	1.389	0.110	-0.599	0.100
IV Wong & Polak	1.023	0.035	1.014	0.032	0.969	0.072	1.502	0.01	-0.702	0.0083
IV Rowe	1.023	0.035	1.016	0.032	0.972	0.071	1.500	0.0093	-0.699	0.008
IV Pandya	1.023	0.035	1.016	0.032	0.973	0.071	1.500	0.0093	-0.699	0.009

TABLE 4.4. Comparison of On-line IV Estimators that use an Auxiliary Model to Generate the Instruments Recursively - Model (A) Second Order

MODEL (B) FOURTH ORDER

OF DATA PAIRS = 1.000

OF TRIALS = 20

ESTIMATOR	$\hat{\beta}_1$	RMSE	$\hat{\beta}_2$	RMSE	$\hat{\beta}_3$	RMSE	$\hat{\beta}_4$	RMSE	$\hat{\beta}_5$	RMSE
True Values	1.000	-	1.000	-	-0.18	-	0.784	-	-0.656	-
Initial Estimates	1.011	0.195	0.719	0.325	0.117	0.340	0.645	0.190	-0.489	0.240
OLS	0.985	0.035	0.787	0.215	0.077	0.260	0.674	0.115	-0.591	0.068
IV Wong & Polak	0.988	0.040	1.016	0.052	-0.215	0.105	0.812	0.080	-0.668	0.030
IV Rowe	0.987	0.036	0.991	0.024	-0.169	0.035	0.784	0.029	-0.659	0.015
IV Pandya	0.988	0.036	0.995	0.028	-0.174	0.038	0.786	0.029	-0.660	0.023

TABLE 4.5. Comparison of On-line IV Estimators that use an Auxiliary Model to Generate the Instruments Recursively - Model (B) Fourth Order

Fig. 4.6. (Bias)² Convergence of On-line IV Estimators that Use an Auxiliary Model to Generate the Instruments Recursively - Model (A) Second Order

Fig. 4.7. MSE Convergence of On-line IV Estimators that Use an Auxiliary Model to Generate the Instruments Recursively - Model (A) Second Order

Fig. 4.8. (Bias)² Convergence of On-line IV Estimators that Use an Auxiliary Model to generate the Instruments Recursively - Model (B) Fourth Order

Fig. 4.9. MSE Convergence of On-line Estimators that Use an Auxiliary Model to Generate the Instruments Recursively - Model (B) Fourth Order

MODEL (A) SECOND ORDER

OF DATA PAIRS = 1.000

OF TRIALS = 20

ESTIMATORS	$\hat{\beta}_1$	RMSE	$\hat{\beta}_2$	RMSE	$\hat{\beta}_3$	RMSE	$\hat{\beta}_4$	RMSE	$\hat{\beta}_5$	RMSE	EST. TIME 1000 SAMPLES
True Values	1.000	-	1.000	-	1.000	-	1.500	-	-0.700	-	-
Initial Estimates	0.789	0.270	1.319	0.430	1.143	0.340	1.415	0.100	-0.623	0.100	-
OLS	1.022	0.033	1.128	0.135	1.152	0.165	1.389	0.110	-0.599	0.100	9 secs.
IV Lagged Inputs	1.023	0.034	1.004	0.035	0.958	0.100	1.512	0.040	-0.715	0.045	9 secs.
IV Lagged Outputs	1.023	0.035	1.013	0.034	0.969	0.065	1.502	0.006	-0.702	0.004	9 secs.
IV Pandya	1.023	0.035	1.016	0.032	0.973	0.071	1.500	0.009	-0.699	0.008	9 secs.

TABLE 4.6: Comparison of IV Estimators that use (a) Lagged Inputs, (b) Lagged Outputs and (c) Recursive Estimates of Noise-free Outputs as Instruments - Model (A) Second Order

MODEL (B) FOURTH ORDER
 # OF DATA PAIRS = 1.000
 # OF TRIALS = 20

ESTIMATORS	$\hat{\beta}_1$		$\hat{\beta}_2$		$\hat{\beta}_3$		$\hat{\beta}_4$		$\hat{\beta}_5$		EST. TIME	
	RMSE	$\hat{\beta}_1$	RMSE	$\hat{\beta}_2$	RMSE	$\hat{\beta}_3$	RMSE	$\hat{\beta}_4$	RMSE	$\hat{\beta}_5$	RMSE	1000 SAMPLES
True Values	1.000	-	1.000	-	-0.180	-	0.784	-	-0.656	-	-	-
Initial Estimates	1.011	0.195	0.719	0.325	0.117	0.340	0.645	0.190	-0.489	0.240	-	-
OLS	0.985	0.035	0.787	0.215	0.077	0.260	0.674	0.115	-0.591	0.068	9 secs.	9 secs.
IV Lagged Inputs	0.987	0.038	1.016	0.050	-0.197	0.060	0.787	0.062	-0.651	0.085	9 secs.	9 secs.
IV Lagged Outputs	0.988	0.036	0.994	0.041	-0.173	0.042	0.797	0.067	-0.671	0.080	9 secs.	9 secs.
IV Pandya	0.988	0.036	0.995	0.028	-0.174	0.038	0.786	0.028	-0.660	0.023	9 secs.	9 secs.

TABLE 4.7: Comparison of IV Estimators that use (a) Lagged Inputs, (b) Lagged Outputs and (c) Recursive Estimates of the Noise-free Outputs and Instruments - Model (B) Fourth Order

Fig. 4.10. (Bias)² Convergence for IV Estimators that Use
 (a) Lagged Inputs (b) Lagged Outputs and
 (c) Recursive Estimates of Noise-free Outputs as
 Instruments - Model (A) Second Order

Fig. 4.11. MSE Convergence for IV Estimators that Use (a) Lagged Inputs (b) Lagged Outputs and (c) Recursive Estimates of Noise-free Outputs as Instruments - Model (A) Second Order

Fig. 4.12. (Bias)² Convergence for IV Estimators that Use (a) Lagged Inputs (b) Lagged Outputs and (c) Recursive Estimates of Noise-free Outputs as Instruments - Model (B) Fourth Order

Fig. 4.13. MSE Convergence for IV Estimators that Use (a) Lagged Inputs (b) Lagged Outputs and (c) Recursive Estimates of Noise-free Outputs as Instruments - Model (B) Fourth Order

MODEL (A) SECOND ORDER

OF DATA PAIRS = 1.000

OF TRIALS = 20

ESTIMATOR	$\hat{\beta}_1$	RMSE	$\hat{\beta}_2$	RMSE	$\hat{\beta}_3$	RMSE	$\hat{\beta}_4$	RMSE	$\hat{\beta}_5$	RMSE	EST. TIME 1000 POINTS
True Values	1.000	-	1.000	-	1.000	-	1.500	-	-0.700	-	-
Initial Estimates	0.798	0.270	1.329	0.430	1.143	0.360	1.415	0.100	-0.623	0.100	-
$\hat{\beta}_{(IV)}$ Pandya	1.023	0.035	1.016	0.032	0.973	0.071	1.500	0.009	-0.699	0.008	9 secs.
S2SLS	1.023	0.035	1.017	0.033	0.975	0.070	1.499	0.010	-0.698	0.007	12 secs.
$\hat{\beta}^*$ (IV)	1.023	0.035	1.014	0.031	0.969	0.074	1.502	0.010	-0.701	0.008	9 secs.

TABLE 4.8. Comparison of (a) $\hat{\beta}_{(IV)}$ - Pandya, (b) S2SLS and (c) $\hat{\beta}^*_{(IV)}$ - that uses True Process Output as Instruments - Model (A) Second Order

MODEL (B) FOURTH ORDER

OF DATA PAIRS = 1.000

OF TRIALS = 20

ESTIMATOR	$\hat{\beta}_1$	RMSE	$\hat{\beta}_2$	RMSE	$\hat{\beta}_3$	RMSE	$\hat{\beta}_4$	RMSE	$\hat{\beta}_5$	RMSE	EST. TIME 1000 POINTS
True Values	1.000	-	1.000	-	-0.180	-	0.784	-	-0.656	-	-
Initial Estimates	1.011	0.195	0.719	0.325	0.117	0.340	0.645	0.190	-0.489	0.240	-
$\hat{\beta}_{(IV)}$ Pandya	0.988	0.036	0.995	0.028	-0.174	0.038	0.786	0.028	-0.660	0.023	9 secs.
S2SLS	0.988	0.035	0.994	0.028	-0.173	0.040	0.788	0.031	-0.662	0.024	12 secs.
$\hat{\beta}_{(IV)}^*$	0.988	0.036	1.004	0.028	-0.187	0.040	0.796	0.032	-0.665	0.023	9 secs.

TABLE 4.9. Comparison of (a) $\hat{\beta}_{(IV)}$ - Pandya, (b) S2SLS and (c) $\hat{\beta}_{(IV)}^*$ - that uses True Process Output as Instruments - Model (B) Fourth Order

Fig. 4.14. MSE Convergence of (a) $\hat{\beta}_{(IV)}$ -Pandya (b) S2SLS and (c) $\hat{\beta}_{(IV)}^*$ THAT USES TRUE PROCESS OUTPUTS - MODEL (A) Second Order

Fig. 4.15. MSE Convergence of (a) $\hat{\beta}_{(IV)}$ - Pandya (b) S2SLS and (c) an IV Estimator that Uses the True Process Output as Instruments - Model (B) Fourth Order

MODEL (C) FOURTH ORDER
(Force-free)

OF DATA PAIRS = 1.000

OF TRIALS = 20

ESTIMATOR	$\hat{\beta}_1$	RMSE	$\hat{\beta}_2$	RMSE	$\hat{\beta}_3$	RMSE	$\hat{\beta}_4$	RMSE	EST. TIME 1000 POINTS
True Values	1.000	-	-0.180	-	0.784	-	-0.656	-	-
Initial Estimates	0.522	0.568	0.188	0.410	0.543	0.325	-0.350	0.406	-
OLS	0.605	0.400	0.206	0.390	0.647	0.135	-0.509	0.150	8 secs.
YULE-WALKER(1)	0.985	0.050	-0.167	0.051	0.792	0.057	-0.664	0.032	8 secs.
YULE-WALKER(2)	0.997	0.042	-0.170	0.065	0.783	0.054	-0.663	0.045	10 secs.

TABLE 4.10. Comparison of OLS with Estimates for use in Force-free Models -
Model (C) Fourth Order (force-free)

Fig. 4.16. Convergence of $(\text{Bias})^2$ for OLS, Yule-Walker (1) and Yule-Walker (2) Estimates - Model (C) Fourth Order Force-Free

Fig. 4.17. Convergence of MSE for OLS, Yule-Walker (1) and Yule-Walker (2) Estimators - Model (C) Fourth Order Force-free

CHAPTER 5

IDENTIFICATION OF LINEAR DYNAMIC MODELS WITH ADDITIVE
OUTPUT NOISE - BOOTSTRAP ESTIMATORS *5.1 Introduction

In the preceeding two chapters, a number of techniques for process parameter estimation in a linear dynamic model were considered. The results presented in Chapter 4 indicate that the recursive approximation to the 2SLS estimator provides a simple and effective technique for on-line estimation of process parameters. In this chapter, we shall restrict our attention to the class of on-line estimators that utilize an auxiliary dynamic model to generate (recursively) the estimates of the noise-free output. Further, we shall consider the simple, though not infrequent, case of a process, modelled by a stable, linear difference equation of known order, whose output is subject to additive measurement noise which is independently and identically distributed (Fig. 5.1). As pointed out in Section 3.2, even for this simple model' OLS estimates of the process parameters are inconsistent.

Parameter estimation problem for this simple model is approached from a somewhat different viewpoint. To start with, it is assumed that the true process output $x(k)$; $k = 1, 2, \dots$ is measured without error. This hypothetical situation provides us with a set of

* The term 'Bootstrap Estimator' is due to Rowe³⁵.

idealized estimators for the process parameters, which are consistent. Practical approximations to these idealized estimators are then obtained by replacing the true process outputs $x(k)$ by their recursive estimates $\hat{x}(k)$.

Fig. 5.1. Process with Output Additive Disturbance

The set of on-line estimators that result from this procedure require that the true process output $x(k)$ be estimated explicitly in real time. Needless to say, one of the estimators in this set of practical estimators, is the IV estimator that utilizes estimates of $x(k)$ as Instruments. An on-line algorithm for this estimator was presented in Chapter 4. The recursive estimator for $x(k)$ used in this algorithm (Fig. 4.3, Eqns. (4.47-51)) was a time varying difference equation (Eqn. 4.51) with $\hat{\beta}(k)$, the latest parameter estimates, as the coefficients. Convergence of such an algorithm results from the mutual and cumulative effects of the parameter estimates $\hat{\beta}(k)$ converging to $\underline{\beta}$ and the estimates $\hat{x}(k)$ of the process output converging to $x(k)$. The stability of such an algorithm, however, is by no means guaranteed. If either of these

estimates start to diverge, we shall have a run-away condition and the final values of the parameter estimates may be completely erroneous.

The class of recursive estimators under discussion, may be considered as a Kalman filter, with the parameter vector $\underline{\beta}(k)$ as the state vector to be estimated in the following message model:

$$\hat{\underline{\beta}}(k+1) = \phi \hat{\underline{\beta}}(k)$$

$$y(k) = [u(k) \dots u(k-n) x(k-1) \dots x(k-n)] \underline{\beta}(k) + e(k)$$

where

$$\phi \equiv I \quad \text{if } \underline{\beta} \text{ is a constant vector.}$$

A necessary (though not sufficient) condition for a Kalman filter to be stable is that the elements of the matrices in the message model be uniformly bounded^{39,40}. Thus if $x(k-1), x(k-2), \dots, x(k-n)$ in the output matrix are replaced by their estimates, then for the resulting algorithm to be stable it is necessary that these estimates be uniformly bounded at all times.

The recursive estimator for the process output that was used in the last chapter (Eqn. (4.51)) is an open loop estimator in the sense that its output $\hat{x}(k)$ does not explicitly depend on the measured output $y(k)$. Assuming that the known input sequence $u(k); k = 1, 2, \dots$ is bounded, the output $\hat{x}(k)$ of such an open loop auxiliary model will be bounded if its BIBO (Bounded Input Bounded Output) stability is assured.

This approach has been utilized by Wong and Polak³³ in their recursive IV algorithm where they include a Jury stability test to ensure BIBO stability of the auxiliary model. In essence, this approach places bounds on certain elements of $\hat{\beta}(k)$ to achieve the necessary bounds on $\hat{x}(k)$. For low order models it is relatively simple to incorporate the Jury's stability test in the algorithm. However, with increase in the model order, this procedure tends to become more complicated.

An alternative method for ensuring the stability of the so-called 'Bootstrap' algorithm is proposed. The method consists in filtering the outputs of the auxiliary model in a suitable fashion. The technique is not only simple, but it can be made adaptive in the sense that the filter is activated only if the algorithm tends to diverge — otherwise the filter remains dormant. The presence of such a filter makes the bootstrap algorithms more flexible and it is possible to demonstrate their underlying relationship to the k-class, h-class and double k-class estimators³⁸ that are sometimes used in econometrics for improving the small sample behavior of 2SLS estimators.

5.2 The Model and the 'Idealized' Estimators for its Parameters

The simple model under consideration is shown in Fig. 5.1 and the true process outputs $x(k)$; $k = 1, 2, \dots$ are generated by the following n^{th} order linear difference equation:

$$x(k) = b_0 u(k) + b_1 u(k-1) + \dots + b_n u(k-n) - a_1 x(k-1) - \dots - a_n x(k-n) \quad (5.1)$$

or

$$x(k) = \frac{B(q^{-1})}{A(q^{-1})} u(k) \quad (5.2)$$

The measured outputs $y(k)$; $k = 1, 2, \dots$ are given by

$$y(k) = x(k) + e(k) \quad (5.3)$$

where

$$e(k) \sim \text{N.I.D. } (0, \lambda^2) \quad (5.4)$$

so that

$$y(k) = b_0 u(k) + b_1 u(k-1) + \dots + b_n u(k-n) - a_1 x(k-1) - \dots - a_n x(k-n) + e(k) \quad (5.5)$$

or

$$y(k) = b_0 u(k) + b_1 u(k-1) + \dots + b_n u(k-n) - a_1 y(k-1) - \dots - a_n y(k-n) + v(k) \quad (5.6)$$

where

$$v(k) = e(k) + a_1 e(k-1) + \dots + a_n e(k-n) \quad (5.7)$$

Alternatively, the model may be specified in the state variable form as

$$\begin{aligned} \underline{x}(k+1) &= F \underline{x}(k) + \underline{g} u(k) \\ y(k) &= \underline{h}' \underline{x}(k) + d u(k) + e(k) \end{aligned} \quad (5.8)$$

where $\underline{x}(k)$ is the vector of state variables, $\underline{h}' = [1, 0, \dots, 0]$ and the matrices F , $\underline{g}$ and d are related to the parameters a_i and b_i through Eqs. (2.15-18)).

For $N+n$ pairs of input-output measurements, the set of N difference equations obtained from Eqn. (5.6), each with $p = 2n+1$ unknown process parameters may be written as

$$\underline{y} = M\underline{\beta} + \underline{v} \quad (5.9)$$

where the above model differs from model (3.4), used in the preceeding two chapters in that the coefficients of the moving average process $v(k)$ in Eqn. (5.7) are a_i instead of c_i ; $i = 1, 2, \dots, n$. The observation vector $\underline{y}$, the parameter vector $\underline{\beta}$ and the design matrix M are as defined in Eqns. (3.6), (3.8), and (3.7) respectively.

In order to derive the set of idealized estimators which constitute the basis for the bootstrap estimators discussed in the subsequent sections, we assume for the moment that the true process output $x(k)$; $k = 1, 2, \dots, N+n$ is also measurable. Equation (5.5) then provides us with another (though hypothetical) model:

$$\underline{y} = X\underline{\beta} + \underline{e} \quad (5.10)$$

which has the design matrix

$$\begin{matrix} X \\ (N \times p) \end{matrix} = \begin{bmatrix} u(n+1) & u(n) & \dots & u(1) & x(n) & \dots & x(1) \\ u(n+2) & u(n+1) & \dots & u(2) & x(n+1) & \dots & x(2) \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ u(n+N) & u(n+N-1) & \dots & u(N) & x(n+N-1) & \dots & x(N) \end{bmatrix} \quad (5.11)$$

and the disturbance vector:

$$\begin{matrix} \underline{e} \\ (\underline{N} \times 1) \end{matrix} = \begin{bmatrix} e(n+1) \\ e(n+2) \\ \vdots \\ e(n+N) \end{bmatrix} \quad (5.12)$$

Assuming a full rank for the matrices M and X, the two models (5.9) and (5.10) lead to the following set of estimators for $\underline{\beta}$:

$$\underline{\tilde{\beta}}^0 = (\underline{M}'\underline{M})^{-1} \underline{M}'\underline{y} \quad (5.13)$$

$$\underline{\tilde{\beta}}^1 = (\underline{X}'\underline{M})^{-1} \underline{X}'\underline{y} \quad (5.14)$$

$$\underline{\tilde{\beta}}^2 = (\underline{X}'\underline{X})^{-1} \underline{X}'\underline{y} \quad (5.15)$$

$$\underline{\tilde{\beta}}^3 = (\underline{M}'\underline{X})^{-1} \underline{M}'\underline{y} \quad (5.16)$$

Estimators $\underline{\tilde{\beta}}^0$ and $\underline{\tilde{\beta}}^1$ evolve from model (5.9) - $\underline{\tilde{\beta}}^0$ being the OLS estimator and $\underline{\tilde{\beta}}^1$ the ideal IV estimator for this model as it utilizes the true process outputs $x(k)$ as instruments³³. Similarly, $\underline{\tilde{\beta}}^2$ and $\underline{\tilde{\beta}}^3$ represent the Maximum Likelihood (ML) estimator and an IV estimator for model (5.10).

To test for the consistency of these estimators, we may combine model (5.9) with estimators (5.13) and (5.14), and model (5.10) with estimators (5.15) and (5.16) to give

$$\tilde{\underline{\beta}}^0 = \underline{\beta} + \left(\frac{1}{N} M'M \right)^{-1} \left(\frac{1}{N} M'v \right) = \underline{\beta} + (S^0)^{-1}(\underline{R}^0) \quad (5.17)$$

$$\tilde{\underline{\beta}}^1 = \underline{\beta} + \left(\frac{1}{N} X'M \right)^{-1} \left(\frac{1}{N} X'v \right) = \underline{\beta} + (S^1)^{-1}(\underline{R}^1) \quad (5.18)$$

$$\tilde{\underline{\beta}}^2 = \underline{\beta} + \left(\frac{1}{N} X'X \right)^{-1} \left(\frac{1}{N} X'e \right) = \underline{\beta} + (S^2)^{-1}(\underline{R}^2) \quad (5.19)$$

$$\tilde{\underline{\beta}}^3 = \underline{\beta} + \left(\frac{1}{N} M'X \right) \left(\frac{1}{N} M'e \right) = \underline{\beta} + (S^3)^{-1}(\underline{R}^3) \quad (5.20)$$

Estimator $\tilde{\underline{\beta}}^i$ ($i = 0,1,2,3$) will be consistent if the following two conditions are satisfied:

$$\text{p.lim}_{N \rightarrow \infty} (R^i) \equiv \underline{0} \quad (5.21)$$

$$\text{p.lim}_{N \rightarrow \infty} (S^i) \text{ exists as a non-singular matrix} \quad (5.22)$$

The matrices M and X are assumed to have full rank, and it will further be assumed that condition (5.22) is satisfied by all four estimators. However, as we have already established in Chapter 3, the OLS estimator $\tilde{\underline{\beta}}^0$ fails to satisfy condition (5.21) and is therefore inconsistent. Condition (5.21) is satisfied by estimators $\tilde{\underline{\beta}}^1$ and $\tilde{\underline{\beta}}^2$ because the matrix X (Eqn. (5.11)) is composed of the inputs $u(k)$; $k = 1,2,\dots,N+n$ and the true process outputs $x(k)$; $k = 1,2,\dots,N+n-1$, neither of which are correlated with the disturbances. For estimator $\tilde{\underline{\beta}}^3$, $\underline{R}^3$ has the form

$$\underline{R}^3 = \left(\frac{1}{N} M' \underline{e} \right) = \begin{bmatrix} \frac{1}{N} \sum_{j=n}^{N+n} u(j)e(j) \\ \vdots \\ \frac{1}{N} \sum_{j=n}^{N+n} u(j-n+1)e(j) \\ \vdots \\ \frac{1}{N} \sum_{j=n}^{N+n} y(j-1)e(j) \\ \vdots \\ \frac{1}{N} \sum_{j=n}^{N+n} y(j-n+1)e(j) \end{bmatrix} \quad (5.23)$$

and since the measured output $y(k)$ is not correlated with the future values of the disturbances $e(k+1)$, $e(k+2)$ etc., condition (5.21) is also satisfied by this estimator.*

Thus, for this hypothetical situation, where the true process outputs $x(k)$ are assumed to be known, the three estimators $\tilde{\beta}^1$, $\tilde{\beta}^2$ and $\tilde{\beta}^3$ provide consistent estimates of the process parameters. Unfortunately $x(k)$ are not measurable and these idealized estimators cannot be implemented. The only estimator that can be implemented is the OLS estimator $\tilde{\beta}^0$, which happens to be inconsistent.

5.3 Practical Approximations to $\tilde{\beta}^1$, $\tilde{\beta}^2$, $\tilde{\beta}^3$ - Bootstrap Estimators

System model (5.6) is analogous to a model that is frequently encountered in econometrics and is characterized by the fact that the

* If the additive output noise in Eqn. (5.3) is not independently distributed (e.g., model (3.2)) then $\tilde{\beta}^3$ will not be consistent.

regressand and some of the regressors are measured with error. Using the terminology introduced by Neyman and Scot⁴¹, the model contains a fixed set of structural parameters, viz., the vector of process parameters $\underline{\beta}$ and a set of incidental parameters^{*}, in this case, the true process outputs $x(1), x(2), \dots, x(N+n-1)$, whose size increases as the number of measurements is increased. Presence of these incidental parameters renders the model unidentifiable with respect to both sets of parameters - unless some additional information on the incidental parameters is available.

In the present case, this additional information is provided by the fact that the incidental parameters are generated by the linear difference equation (5.1) or equivalently by the state-space model (5.8). Consistent estimates of the structural parameters $\underline{\beta}$ can therefore be obtained if we can estimate the incidental parameters - either implicitly or explicitly by utilizing this additional information.

One way in which the problem can be approached is to minimize the quadratic form

$$Q = (\underline{y} - X\underline{\beta})' (\underline{y} - X\underline{\beta})$$

with respect to $\underline{\beta}$, and subject to the constraint specified by Eqn. (5.1) or equivalently by state equations (5.8). This procedure leads to a Two

* Sometimes these are also referred as nuisance parameters.

Point Boundary Value Problem (TPBVP) whose closed form solution is not readily available¹¹. The TPBVP can, however, be solved by employing numerical minimization techniques^{11,12}. Alternatively, it may be converted to a nonlinear initial value problem and again solved by an iterative procedure (quasi-linearization). Estimation of the incidental parameters $x(k)$; $k = 1, 2, \dots$ is implicit in these techniques.

The approach that we shall consider is based on an attempt to estimate the structural parameters $\underline{\beta}$ as well as the incidental parameters $x(k)$ explicitly in real time. The estimates of the structural parameters $\underline{\beta}$ are obtained recursively from either of the three estimators $\tilde{\underline{\beta}}^1$, $\tilde{\underline{\beta}}^2$ or $\tilde{\underline{\beta}}^3$ with $x(k)$ replaced by their best estimates - which in turn are obtained by utilizing Eqn. (5.1) or (5.8), with the unknown parameters replaced by their most recent estimates. A bootstrapping operation is therefore involved between the two recursive estimators. The basic structure of a bootstrap estimator is shown in Fig. 5.2, which consists of two recursive estimators - one for the structural parameters and the other for the incidental parameters with continuous exchange of information between the two.

Recursive estimators for the process parameters

If we define

Fig. 5.2. Basic Structure of a Bootstrap Estimator

$$\begin{matrix} \mathbf{Z}_x \\ (N \times p) \end{matrix} = \begin{vmatrix} u(n+1) & u(n) & \dots & u(1) & \hat{x}(n) & \dots & \hat{x}(1) \\ u(n+2) & u(n+1) & \dots & u(2) & \hat{x}(n+1) & \dots & \hat{x}(2) \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ u(n+N) & u(n+N-1) & \dots & u(N) & \hat{x}(n+N-1) & \dots & \hat{x}(N) \end{vmatrix} \quad (5.24)$$

as the matrix obtained by replacing the true incidental parameters $x(1), x(2), \dots, x(N+n-1)$ in the matrix X (Eqn. (5.11)) by their recursive estimates $\hat{x}(1), \hat{x}(2), \dots, \hat{x}(N+n-1)$, we get the following bootstrap estimators corresponding to the idealized estimators $\tilde{\beta}^1$, $\tilde{\beta}^2$, and $\tilde{\beta}^3$ respectively:

$$\hat{\beta}^1 = (\mathbf{Z}'_x \mathbf{M})^{-1} \mathbf{Z}'_x \mathbf{y} \quad (5.25)$$

$$\hat{\beta}^2 = (\mathbf{Z}'_x \mathbf{Z}_x)^{-1} \mathbf{Z}'_x \mathbf{y} \quad (5.26)$$

$$\hat{\beta}^3 = (\mathbf{M}' \mathbf{Z}_x)^{-1} \mathbf{M}' \mathbf{y} \quad (5.27)$$

Estimator $\hat{\beta}^1$ is the IV estimator that utilizes the estimates of the noise-free process output as instruments and the recursive form for this type of estimator has already been presented in Chapter 4 (Eqns. (4.47) - (4.51)). Recursive form for $\hat{\beta}^3$ is obtained from that of $\hat{\beta}^1$ by interchanging the vectors $\underline{m}(k+1)$ and $\underline{z}(k+1)$ everywhere. Estimator $\hat{\beta}^2$ has the following recursive form:

$$\hat{\beta}(k+1) = \hat{\beta}(k) + \frac{C(k) \underline{z}(k+1) (y(k+1) - \underline{z}'(k+1) \hat{\beta}(k))}{1 + \underline{z}'(k+1) C(k) \underline{z}(k+1)} \quad (5.28)$$

$$C(k+1) = C(k) - \frac{C(k) \underline{z}(k+1) \underline{z}'(k+1) C(k)}{1 + \underline{z}'(k+1) C(k) \underline{z}(k+1)} \quad (5.29)$$

where the superscript has been dropped for convenience and where

$$C(k) = (Z_x' Z_x)^{-1} \text{ at sampling instant } k \quad (5.30)$$

$$\underline{z}(k+1) = [u(k+1)u(k) \dots u(k-n+1)\hat{x}(k) \dots \hat{x}(k-n+1)]' \quad (5.31)$$

Thus, Estimator #1 in Fig. 5.2 will consist of the recursive forms of either $\hat{\underline{\beta}}^1$, $\hat{\underline{\beta}}^2$ or $\hat{\underline{\beta}}^3$ with $\hat{x}(k)$ furnished by Estimator #2.

Recursive Estimators for the Process Output

The true process outputs $x(k)$; $k = 1, 2, \dots$ evolve according to the linear difference equation (5.1) or equivalently according to the linear state space model (5.8), in which the state vector $\underline{x}(k)$ is the n vector of state variables and $\underline{h}' = [1 \ 0 \ 0 \ \dots \ 0]$. If the transition matrix F , the input matrix $\underline{g}$ and the parameter d in the latter model are replaced by their most recent estimates furnished by Estimator #1 (Fig. 5.2), we may consider the alternate model:

$$\begin{aligned} \underline{x}(k+1) &= \hat{F}(k, k-1) \underline{x}(k) + \hat{\underline{g}}(k) u(k) + \underline{\varepsilon}(k) \\ y(k) &= \underline{h}' \underline{x}(k) + \hat{d}(k) u(k) + v(k) + e(k) \end{aligned} \quad (5.32)$$

where the fictitious system noise $\underline{\varepsilon}(k)$ and the output noise $v(k)$ have been introduced to account for the modelling errors - the errors caused by using uncertain parameters in the model.

For the message model (5.32), if the assumptions associated with a discrete Kalman filter are satisfied and if the noise moments:

$$E(\underline{\varepsilon}(k) \underline{\varepsilon}'(k)) = Q(k) \quad (5.33)$$

$$E(v(k) v(k)) = \sigma^2(k) \quad (5.34)$$

$$E(e(k) e(k)) = \lambda^2 \quad (5.35)$$

are known, the minimum variance estimates of the state vector $\underline{x}(k)$ can be obtained from the filter:

$$\hat{\underline{x}}(k/k-1) = \hat{F}(k, k-1) \hat{\underline{x}}(k-1/k-1) + \hat{g}(k-1) u(k-1) \quad (5.36)$$

$$\hat{\underline{x}}(k/k) = \hat{\underline{x}}(k/k-1) + \underline{K}(k) [y(k) - \underline{h}' \hat{\underline{x}}(k/k-1) - \hat{d}(k)u(k)] \quad (5.37)$$

$$\underline{K}(k) = P(k/k-1) \underline{h} [\underline{h}' P(k/k-1) \underline{h} + \sigma^2(k) + \lambda^2]^{-1} \quad (5.38)$$

$$P(k/k-1) = \hat{F}(k, k-1) P(k-1/k-1) \hat{F}'(k, k-1) + Q(k) \quad (5.39)$$

$$P(k/k) = P(k/k-1) - \underline{K}(k) \underline{h}' P(k/k-1) \quad (5.40)$$

$$\hat{\underline{x}}(0/0) = E[\underline{X}(0)] \quad (5.41)$$

$$P(0/0) = \text{Cov} [\underline{X}(0) - \hat{\underline{x}}(0/0)] \quad (5.42)$$

The filtered estimates of the scalar process output*

$$x(k) = \underline{h}' \underline{x}(k) + d u(k) \quad (5.43)$$

are given by

$$\hat{x}(k/k) = \underline{h}' \hat{\underline{x}}(k/k) + \hat{d}(k) u(k) \quad (5.44)$$

or

$$\hat{x}(k/k) = \hat{x}(k/k-1) + K_1(k) [y(k) - \hat{x}(k/k-1)] \quad (5.45)$$

where $K_1(k)$ = first element of $\underline{K}(k)$ and K_1 satisfies the relation

$$0 \leq K_1(k) \leq 1$$

and

$$\hat{\underline{x}}(k/k-1) = \underline{h}' \hat{\underline{x}}(k/k-1) + \hat{d}(k) u(k) \quad (5.46)$$

Use of fictitious system and output noise to cover modelling errors was originally suggested by Schmidt⁴¹ who proposed the so-called Schmidt-Kalman filter, for models with persistent parameter uncertainties. Though this approach is quite general and hence attractive, its practical utility is limited because:

- (a) It is unrealistic to assume that the elements of $Q(k)$ and $\sigma^2(k)$ are either known a priori or they can be accurately estimated during the filtering process.

* Note that $\underline{x}(k)$ is the n vector of phase variables whereas $x(k)$ (without the bar) is the noise-free process output defined by (5.43) and it is a scalar.

- and (b) It is difficult to justify the fact that the standard assumptions associated with a discrete Kalman filter will hold in this situation. More likely than not, $\underline{\varepsilon}(k)$ and $v(k)$ and/or $\underline{\varepsilon}(k)$ and $\hat{x}(0)$ will be correlated. This not only complicates the filtering equations but also requires a knowledge of these cross-correlations.

An important effect of the presence of $Q(k)$ is to increase the error covariance matrix (Eqn. (5.39)), thereby preventing the filter gain to become zero, thus keeping the filter open for incoming data. Assuming that the parameter d is known (so that $v(k) \equiv 0$), we can achieve this effect by using an artificially designed matrix $Q(k)$ of the form:

$$Q(k) = d_1 I; \quad d_1 = \text{positive constant} \quad (5.47)$$

if the parameter uncertainties are persistent. On the other hand, if we expect the parameter uncertainties to decrease progressively - which happens to be our main concern in the present case, we may set

$$Q(k) = f(k) \times I \quad (5.48)$$

where $f(k)$ is a scalar function which satisfies the conditions:

- (i) $0 \leq f(k) \leq d_2 \quad \forall k, \quad d_2 = \text{positive constant}$
- (ii) $f(k+1) \leq f(k) \quad \forall k$
- (iii) $\lim_{N \rightarrow \infty} f(k) = 0$

A more attractive choice is provided by Jazwinsky⁴², through his adaptive filtering technique, whereby, the elements of $Q(k)$ (taken as a diagonal matrix for simplicity) are made to depend on the difference between the expected and the actual behaviour of the residuals - the actual behavior being judged by observing some suitable statistic of the residuals.

As we may recall, for the bootstrap algorithm, at the sampling instant $(k+1)$, we only need the single estimate of the scalar process output $x(k)$ (and not the complete state vector). In which case, the structure of the required filter is given by

$$\hat{x}(k/k) = \hat{x}(k/k-1) + \alpha(k) [y(k) - \underline{h}' \hat{x}(k/k-1) - d(k)u(k)]$$

$$\hat{x}(k/k) = \hat{x}(k/k-1) + \alpha(k) [y(k) - \hat{x}(k/k-1)]$$

or

$$\hat{x}(k/k) = [1 - \alpha(k)] \hat{x}(k/k-1) + \alpha(k) y(k) \quad 0 \leq \alpha \leq 1 \quad (5.49)$$

where $\alpha(k)$ is a general gain sequence - not necessarily an optimum one, and where the predictions $\hat{x}(k/k-1)$ are given by Eqns. (5.36) and (5.46), or equivalently by the linear, time-varying difference equation:

$$\begin{aligned} \hat{x}(k/k-1) = & \hat{b}_0(k)u(k) + \hat{b}_1(k)u(k-1) + \dots + \hat{b}_n(k)u(k-n) \\ & - \hat{a}_1(k) \hat{x}(k-1/k-1) - \dots - \hat{a}_n(k) \hat{x}(k-n/k-n) \end{aligned} \quad (5.50)$$

The choice of the gain sequence $\alpha(k)$ is dictated by the following considerations:

1. The sequence of residuals or more specifically the innovations sequence defined by

$$\hat{e}(k) = y(k) - \hat{x}(k/k-1) \quad (5.51)$$

should remain bounded - preferably within some pre-specified bounds.

2. The filter is most likely to diverge in the early stages of the estimation procedure, when the parameter uncertainties are the greatest. Thus, if we can control $\hat{x}(k)$ from diverging in these initial stages, in all probability, the bootstrap algorithm will be stable.

In view of the above considerations, we may choose the scalar gain sequence $\alpha(k)$ in one of the following two ways, which correspond respectively to the two choices of $Q(k)$ discussed earlier.

(A) $\alpha(k) = \text{some suitable function of } k$

One of the possible functions that may characterize this class of filters is given by

$$\alpha(k) = \left[1 - \left(\frac{k-1}{k} \right)^\zeta \right] , \quad \zeta \geq 0 \quad (5.52)$$

in which the limiting values of the parameter ζ define the range of variation in $\alpha(k)$. Thus

$$\zeta = 0 \rightarrow \alpha(k) \equiv 0 \quad \forall k$$

and the filter output is independent of the data, and

$$\zeta = \infty \rightarrow \alpha(k) \equiv 1 \quad \forall k$$

and the filter output is identically equal to $y(k)$ and the bootstrap estimators $\hat{\beta}^1$, $\hat{\beta}^2$ and $\hat{\beta}^3$ are all reduced to OLS estimators. Depending on how much weight is to be assigned to the observations $y(k)$, we may choose a suitable value for ζ within its range. Obviously, if ζ is too small, there is a risk that $\hat{x}(k)$ may start to diverge and the bootstrap algorithms may become unstable. On the other hand, if ζ is too large, the algorithm may converge very slowly. Some typical gain sequences $\alpha(k)$ that may be used in practice are shown in Fig. 5.3. The case where

$$\zeta = \infty \rightarrow \alpha(k) = 1 \quad \text{for } k \leq k_m$$

$$\zeta = 0 \rightarrow \alpha(k) = 0 \quad \text{for } k > k_m$$

implies that the estimator for the process parameters $\underline{\beta}$ (Fig. 5.2) is primed by an OLS estimator, operating on an initial data stream of length k_m , subsequent to which unfiltered estimates of $x(k)$ are used. Bootstrap estimators, for linear models with stationary as well as slowly time-varying parameters, and for this choice of $\alpha(k)$ have been discussed by Pandya⁵.

Though the filters in this category are very easy to implement, they have the following drawbacks:

Fig. 5.3. Some Typical Gain Sequences for Filter (A)

- (a) The choice of ζ or the nature of the filter gain $\alpha(k)$ has to be decided either by experimentation or set in some arbitrary manner. More often than not, one is tempted to decide in favor of a slowly converging algorithm rather than risk instability.
- (b) The filter for $x(k)$ operates independently of the estimator for $\underline{\beta}$. In other words, even if the bootstrap algorithm tends to be unstable, the filter for $x(k)$ is unable to take any corrective action, because the value of ζ (or the nature of $\alpha(k)$) once set, remains unchanged.

(B) $\alpha(k)$ = *suitable function of the residuals*

This method of setting the filter gains $\alpha(k)$ is a simple adaptation of Jazwinsky's technique. It not only takes out some of the guess work associated with the simple filter discussed above but because of its adaptive nature its performance is superior. As mentioned earlier, the technique is based on the philosophy that if the filter is converging, then the actual behavior of the residuals, as judged on the basis of some suitable statistic, should be consistent with their expected behavior.

With reference to the model under discussion, if the bootstrap estimator is converging ($\hat{\underline{\beta}} \xrightarrow{P} \underline{\beta}$), then the sequence of residuals defined by Eqn. (5.51) should converge in probability to a normal random variable with zero mean and a variance that is close to λ^2 , the variance of the additive output noise. This means that if we consider the statistic

Text complete; leaf 149 omitted in numbering.

$$\hat{\lambda}^2(k) = \frac{1}{k} \sum_{i=1}^k [y(i) - \hat{x}(i/i-1)]^2 = \frac{1}{k} \sum_{i=1}^k [\hat{e}(i)]^2 \quad (5.53)$$

or

$$\hat{\lambda}^2(k) = \left(\frac{k-1}{k} \right) \hat{\lambda}^2(k-1) + \frac{1}{k} [\hat{e}(k)]^2 \quad (5.54)$$

then it must satisfy the property

$$\text{p.lim}_{k \rightarrow \infty} [\hat{\lambda}^2(k)] = \lambda^2 \quad (5.55)$$

Thus, $\hat{\lambda}^2(k)$ is a consistent estimator of λ^2 , and if we were to observe $\hat{\lambda}^2(k)$ at statistically significant intervals, the following inequality should hold with a high degree of probability:

$$\hat{\lambda}^2(r) \geq \hat{\lambda}^2(2r) \geq \dots \geq \lambda^2 \quad (5.56)$$

For a converging algorithm therefore, $\hat{\lambda}^2(k)$ are bounded both from above and below, with probability one.

In practice it is not difficult to obtain a very conservative estimate of λ^2 .^{*} If not, we can always start with a large positive constant as an a priori estimate of λ^2 and when the algorithm starts to converge, we can refine it periodically by using the inequality (5.56), as shown in Fig. 5.4.

^{*} e.g., from a preliminary step response test.

Fig. 5.4. Progressively Refined Bound on $\hat{\lambda}^2(k)$ for Adaptive Filter (B)

If the algorithm is converging, we expect that the values of $\hat{\lambda}^2(k)$, calculated recursively by Eqn. (5.54) should fall below the bound $\Lambda^2(k)$ (Fig. 5.4). If the algorithm starts to diverge, $\hat{\lambda}^2(k)$ will tend to violate the bound and a corrective action is called for. This suggests that we may set the filter gains $\alpha(k)$ adaptively, as follows:

$$\begin{aligned} \alpha(k) &= 0 \quad \text{if } \hat{\lambda}^2(k) \leq \Lambda^2(k) \\ \alpha(k) &= \left[1 - \frac{\Lambda^2(k)}{\hat{\lambda}^2(k)} \right] \quad \text{if } \hat{\lambda}^2(k) > \Lambda^2(k) \end{aligned} \quad (5.57)$$

At each sampling instant, the value of $\Lambda^2(k)$ acts as a threshold and we may notice that

$$\begin{aligned} \Lambda^2(k) &\equiv 0 \rightarrow \alpha(k) \equiv 1 \quad \forall k \rightarrow \text{OLS} \\ \Lambda^2(k) &\equiv \infty \rightarrow \alpha(k) \equiv 0 \quad \forall k \rightarrow \text{no filtering} \end{aligned}$$

In Eqn. (5.54), recursive estimates of $\hat{\lambda}^2(k)$ are obtained by assigning equal weights to all the residuals. To make the filter for $x(k)$ respond faster to any diverging tendencies, it is advisable to assign greater weights to more recent residuals. This can be achieved by using a recursive estimator of the form:

$$\hat{\lambda}^2(k) = \left(\frac{k-1}{k} \right)^\eta \hat{\lambda}^2(k-1) + \left[1 - \left(\frac{k-1}{k} \right)^\eta \right] \hat{e}^2(k); \quad \eta > 1 \quad (5.58)$$

where $\eta = 2$ is generally adequate.

In view of the above discussion, the recursive algorithm for one of the bootstrap estimators - say

$$\hat{\underline{\beta}}^1 = (\underline{Z}'_X \underline{M})^{-1} \underline{Z}'_X \underline{y}$$

may be written as:

$$\hat{\underline{\beta}}(k+1) = \underline{\beta}(k) + \frac{C(k) \underline{z}(k+1) [y(k+1) - \underline{m}'(k+1) \hat{\underline{\beta}}(k)]}{1 + \underline{m}'(k+1) C(k) \underline{z}(k+1)} \quad (5.59)$$

$$C(k) = C(k) - \frac{C(k) \underline{z}(k+1) \underline{m}'(k+1) C(k)}{1 + \underline{m}'(k+1) C(k) \underline{z}(k+1)} \quad (5.60)$$

$$\underline{m}(k+1) = [u(k+1)u(k) \dots u(k-n+1) y(k) \dots y(k-n+1)]' \quad (5.61)$$

$$\underline{z}(k+1) = [u(k+1)u(k) \dots u(k-n+1) \hat{x}(k/k) \dots \hat{x}(k-n+1/k-n+1)]' \quad (5.62)$$

$$\hat{x}(k/k) = \hat{x}(k/k-1) + \alpha(k) [y(k) - \hat{x}(k/k-1)]; \quad 0 \leq \alpha(k) \leq 1 \quad (5.63)$$

$$\hat{x}(k/k-1) = \underline{z}'(k) \hat{\underline{\beta}}(k) \quad (5.64)$$

which may be implemented in practice, as shown in Fig. 5.5.

5.4 Relationship of Bootstrap Estimators to General k-class, h-class and double k-class Estimators

If we put $\alpha(k) = 0 \forall k$, in Eqn. (5.63), the on-line algorithm for $\hat{\underline{\beta}}^1$ reduces to the recursive approximation to 2SLS (Eqns. (4.47-51),

Fig. 4.3), that was proposed in Chapter 4. However, as long as $\lim_{k \rightarrow \infty} \alpha(k) = 0$, asymptotically the two estimators are equivalent. On the other hand, if $\alpha(k) = 1 \quad \forall k$, the on-line estimator $\hat{\underline{\beta}}^1$ (also $\hat{\underline{\beta}}^2$ and $\hat{\underline{\beta}}^3$) reduces to a recursive least squares estimator. Thus, by incorporating a filter (Eqn. (5.63)) in the algorithm for $\hat{\underline{\beta}}^1$, we have made it more flexible in the sense that the performance of the estimator now depends on the nature of the filter gain sequence $\alpha(k)$.

In the present case, the decision to use such a filter was motivated by stability considerations for the on-line algorithms. However, as we shall presently see, the explicit use of a filter, relates these algorithms to the so-called general k-class, h-class and double k-class estimators in econometrics^{38,43}. The general k-class estimator may be formulated as follows:

Consider the partitioned form of the model $\underline{y} = M\underline{\beta} + \underline{v}$ given by:

$$\underline{y} = M_1 \underline{\beta}_1 + M_2 \hat{\underline{\beta}}_2 + \underline{v} \quad (5.65)$$

or

$$\underline{y} = [M_1 \mid M_2] \begin{bmatrix} \underline{\beta}_1 \\ \underline{\hat{\beta}}_2 \end{bmatrix} + \underline{v} \quad (5.66)$$

where M_1 contains only the input variables that are not correlated with the elements of $\underline{v}$, and M_2 contains the output variables that are correlated with the elements of $\underline{v}$.

As we may recall from Section 4.2, the 2SLS estimator can be obtained by using the following Instrumental Variable (IV) matrix in model (5.66);

$$Z = [M_1 : \hat{X}_2] \quad (5.67)$$

where $\hat{X}_2$ is the matrix obtained by replacing the elements of M_2 , by their noise free estimates - the latter being obtained by fitting the reduced form (or impulse response) model using OLS. The 2SLS estimator is therefore given by

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}_{2SLS} = \begin{bmatrix} M_1' M_1 & M_1' M_2 \\ X_2' M_1 & \hat{X}_2' M_2 \end{bmatrix}^{-1} \begin{bmatrix} M_1' \\ \hat{X}_2' \end{bmatrix} Y \quad (5.68)$$

or

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}_{2SLS} = \begin{bmatrix} M_1' M_1 & M_1' M_2 \\ (M_2 - \hat{W})' M_1 & (M_2 - \hat{W})' M_2 \end{bmatrix}^{-1} \begin{bmatrix} M_1' \\ (M_2 - \hat{W})' \end{bmatrix} Y \quad (5.69)$$

where $\hat{W}$ is the matrix of residuals resulting from the reduced form fit.

Since the design matrix U , in the reduced form model has the form:

$$U = [M_1 : U_2] \quad (5.70)$$

i.e., M_1 is a submatrix of U with equal number of rows, and since

$$\begin{aligned}
 U'\hat{W} &= U'[M_2 - \hat{X}_2] \\
 &= U'[M_2 - U(U'U)^{-1} U'M_2] \\
 &= 0
 \end{aligned}$$

we have

$$\begin{bmatrix} M_1' \\ \hline U_2' \end{bmatrix} \hat{W} = \begin{bmatrix} 0 \\ \hline 0 \end{bmatrix} \rightarrow M_1' \hat{W} = \hat{W}' M_1 = 0 \quad (5.71)$$

Further

$$\hat{W}' M_2 = \hat{W}' [\hat{X}_2 + \hat{W}] = \hat{W}' \hat{X}_2 + \hat{W}' \hat{W}$$

but

$$\begin{aligned}
 \hat{W}' \hat{X}_2 &= [M_2 - U(U'U)^{-1} U'M_2]' [U(U'U)^{-1} U'M_2] \\
 &= M_2' U(U'U)^{-1} U'M_2 - M_2' U(U'U)^{-1} U' U(U'U)^{-1} U'M_2 \\
 &= 0
 \end{aligned}$$

so that

$$\hat{W}' M_2 = \hat{W}' \hat{W} \quad (5.72)$$

and in view of Eqns. (5.71) and (5.72), the 2SLS estimator (5.69) may be written as:

$$\begin{bmatrix} \hat{\beta}_1 \\ \hline \hat{\beta}_2 \end{bmatrix}_{2SLS} = \begin{bmatrix} M_1' M_1 & M_1' M_2 \\ \hline M_2' M_1 & M_2' M_2 - \hat{W}' \hat{W} \end{bmatrix}^{-1} \begin{bmatrix} M_1' \\ \hline (M_2 - \hat{W})' \end{bmatrix} y \quad (5.73)$$

However, the 2SLS estimator above may be considered as a special case of the more general estimator:

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}_{k\text{-class}} = \begin{bmatrix} M_1' M_1 & M_1' M_2 \\ M_2' M_1 & M_2' M_2 - k \hat{W}' \hat{W} \end{bmatrix}^{-1} \begin{bmatrix} M_1' \\ (M_2 - k \hat{W})' \end{bmatrix} y \quad (5.74)$$

with $k = 1$.*

Estimator (5.74) is known as the general k-class estimator.

The following properties of this estimator are immediately apparent:

1. $k = 1 \rightarrow$ 2SLS estimator
2. $k = 0 \rightarrow$ OLS estimator
3. If k is a function of time t , then the estimator is consistent iff.

$$\lim_{t \rightarrow \infty} k(t) = 1$$

and the asymptotic behavior of the resulting estimator is the same as that of 2SLS estimator.

Though the foregoing derivation of the general k-class estimator was based on a generalization of 2SLS, an alternate way to arrive at the same result is to use the following IV matrix in Model (5.66):

$$\begin{aligned} Z^* &= [M_1 : (M_2 - k \hat{W})] \\ &= [M_1 : (\hat{X}_2 + (1-k) \hat{W})] \\ &= [M_1 : \hat{X}_2^*] \end{aligned} \quad (5.77)$$

* Symbol k here does not signify the time index. For the rest of this section only, we shall use t as the discrete time index.

and if we consider an element of the matrix $\hat{X}_2^*$, we have

$$\hat{x}^*(t) = \hat{x}(t) + [1 - k(t)]\hat{w}(t)$$

or

$$\hat{x}^*(t) = \hat{x}(t) + [1 - k(t)][y(t) - \hat{x}(t)] \quad (5.78)$$

where k is assumed to be a function of time. Equation (5.78) may be written as

$$\hat{x}^*(t) = \hat{x}(t) + \alpha(t) [y(t) - \hat{x}(t)] \quad (5.79)$$

where

$$\alpha(t) \triangleq [1 - k(t)]$$

The relationship between the general k -class estimator and the recursive algorithm for $\hat{\beta}^1$ (Fig. 5.5) becomes apparent when we compare Eqn. (5.79) with Eqn. (5.63). In fact, Eqns. (5.59-64) may be considered as a recursive approximation to the general k -class estimator - just as Eqns. (4.47-51) in Chapter 4, represent a recursive approximation to the 2SLS estimator.

It is not difficult to show that a similar relationship exists between the recursive algorithms for the bootstrap estimators $\hat{\beta}^2$ and $\hat{\beta}^3$, and what are respectively known as h -class and double k -class estimators given by:

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}_{h\text{-class}} = \begin{bmatrix} M_1' M_1 & M_1' M_2 \\ M_2' M_1 & M_2' M_2 - (2h - h^2) \hat{W}' \hat{W} \end{bmatrix}^{-1} \begin{bmatrix} M_1' \\ (M_2 - h \hat{W})' \end{bmatrix} \underline{y} \quad (5.80)$$

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}_{\text{double } k\text{-class}} = \begin{bmatrix} M_1' M_1 & M_1' M_2 \\ M_2' M_1 & M_2' M_2 - k_1 \hat{W}' \hat{W} \end{bmatrix}^{-1} \begin{bmatrix} M_1' \\ (M_2 - k_2 \hat{W})' \end{bmatrix} \underline{y} \quad (5.81)$$

where $h = k_1 = 1 - \alpha(t)$ and $k_2 = 0$.

5.5 Results

The Bootstrap estimators $\hat{\beta}^1$, $\hat{\beta}^2$ and $\hat{\beta}^3$ were tested on artificially generated data from the following two models:

Second Order Model (Astrom¹²)

$$x(k) = 1.5 x(k-1) - 0.7 x(k-2) + u(k-1) + 0.5 u(k-1)$$

$$y(k) = x(k) + e(k)$$

$$e(k) \sim \text{N.I.D. } (0, \lambda^2); \quad \lambda^2 = 1.$$

$$u(k) \sim \text{PRBS } (1, 1, 127) + \text{PRBS } \left(\frac{1}{2}, 127, 889\right)$$

Fourth Order Model (Lee⁸)

$$x(k) = x(k-1) - 0.18 x(k-2) + 0.784 x(k-3) - 0.656 x(k-4)$$

$$y(k) = x(k) + e(k)$$

$$e(k) \sim \text{N.I.D. } (0, \lambda^2); \quad \lambda^2 = 1.$$

$$u(k) \sim \text{PRBS } (1, 1, 255) + \text{PRBS } \left(\frac{1}{2}, 255, 765\right)$$

where PRBS(a,t,T) represents a Pseudo Random Binary Sequence of amplitude a, clock time t and period T - the last two being measured in terms of the sampling rate.

The input signals represented by PRBS (1/2,127,889) in the first model and PRBS(1/2,255,765) in the second model have been included to account for the fact that, in practice, the process to be identified may already be in operation. The above signals are much slower and represent process inputs in the absence of any special test signals that may be applied for identifying the process parameters.

Recursive estimates of the process output $x(k)$, required for each of the bootstrap estimators, were obtained by using the equations (5.63), (5.64). For the filter represented by Eqn. (5.63), the following three filter gain sequences were used:

1. No Filter

$$\alpha(k) \equiv 0 \quad \forall k$$

2. Non-adaptive Filter

$$\alpha(k) = \left[1 - \left(\frac{k-1}{k} \right)^\zeta \right]; \quad \zeta = 2$$

3. Adaptive Filter

$$\alpha(k) = 0 \quad \text{if} \quad \hat{\lambda}^2(k) \leq \Lambda^2(k)$$

$$\alpha(k) = \left[1 - \frac{\Lambda^2(k)}{\hat{\lambda}^2(k)} \right] \quad \text{if} \quad \lambda^2(k) > \Lambda^2(k)$$

$$\Lambda^2(k) = 3 \lambda^2 = \text{constant}$$

where $\hat{\lambda}^2(k)$ is obtained recursively from Eqn. (5.58).

For comparison, the estimates obtained from the OLS estimator and the 'Idealized estimators' $\underline{\hat{\beta}}^1$, $\underline{\hat{\beta}}^2$, and $\underline{\hat{\beta}}^3$ (Eqns. (5.13-16)) have also been included.

The following initial conditions were used in the recursive algorithms (Fig. 5.5):

$$\begin{aligned} \underline{\hat{\beta}}(0) &= \underline{0} \\ C(0) &= 10 \times I \\ z(0) &= \underline{m}(0) \end{aligned}$$

Results presented in Table 5.1 and 5.5 clearly indicate that the idealized estimators $\underline{\hat{\beta}}^1$, $\underline{\hat{\beta}}^2$, and $\underline{\hat{\beta}}^3$ are consistent, whereas the OLS estimator for this model is inconsistent.

For the above initial conditions, bootstrap estimators $\underline{\hat{\beta}}^1$ and $\underline{\hat{\beta}}^2$ converge even without the use of a filter, but estimator $\underline{\hat{\beta}}^3$ does require a filter to achieve convergence. (Tables 5.2-4 and 5.6-8.) For

the fourth order model, the latter converges only in the presence of the adaptive filter. We may therefore conclude that estimators $\hat{\beta}^1$ and $\hat{\beta}^2$ exhibit better convergence properties than $\hat{\beta}^3$. Further, when the algorithms converge, the presence of the filter does not appreciably affect the large sample properties of the estimators as reflected in the sample RMS error of the estimates calculated from 20 trials.

The fact that the presence of the adaptive filter keeps the estimates of the process outputs bounded can be seen from Figs. (5.6-11). In these graphs the sample mean squared error of $\hat{\lambda}^2(k)$, the recursive estimates of the noise variance, are plotted as a function of k for the different bootstrap estimators operating with and without the adaptive filter.

5.6 Concluding Remarks

In this chapter an attempt has been made to formalize the concept of 'Bootstrap Estimators'. In essence such estimators are characterized by the fact that the 'structural parameters' as well as the 'incidental parameters' are estimated recursively in such a way that the latest estimates of the 'structural parameters' are used in estimating the 'incidental parameters' and vice versa. Obviously, such a procedure is possible only if the two estimators are recursive.

To alleviate the stability problems associated with such estimators, an adaptive filter for the incidental parameters is proposed. The results on simulated data indicate that such a filter definitely has a stabilizing effect on the bootstrap algorithms.

In Section 5.4 it is shown that use of a filter for the incidental parameters, relates the resulting estimators to the general k-class, h-class and double k-class estimators encountered in econometrics.

SECOND ORDER MODEL
OF DATA PAIRS = 1,000
OF TRIALS = 20

ESTIMATOR	$\tilde{\beta}_1$	RMSE	$\tilde{\beta}_2$	RMSE	$\tilde{\beta}_3$	RMSE	$\tilde{\beta}_4$	RMSE	$\tilde{\chi}^2$	RMSE
True Values	1.000	-	0.500	-	1.500	-	-0.700	-	1.000	-
Starting Values	0.0	1.000	0.0	0.500	0.0	1.500	0.0	0.700	3.0	2.000
$\tilde{\beta}^0$ (OLS)	0.989	0.036	0.915	0.420	1.036	0.455	-0.277	0.425	1.543	0.556
$\tilde{\beta}^1$	0.994	0.041	0.460	0.096	1.511	0.026	-0.708	0.019	0.988	0.042
$\tilde{\beta}^2$	1.011	0.030	0.483	0.044	1.503	0.019	-0.701	0.017	0.983	0.037
$\tilde{\beta}^3$	1.011	0.030	0.483	0.045	1.502	0.023	-0.700	0.022	0.985	0.036

TABLE 5.1. Comparison of Idealized Estimators $\tilde{\beta}^1, \tilde{\beta}^2$ and $\tilde{\beta}^3$ -
Second Order Model

SECOND ORDER MODEL

OF DATA PAIRS = 1,000

OF TRIALS = 20

NO FILTER

ESTIMATOR	$\hat{\beta}_1$	RMSE	$\hat{\beta}_2$	RMSE	$\hat{\beta}_3$	RMSE	$\hat{\beta}_4$	RMSE	$\bar{\chi}^2$	RMSE
True Values	1.000	-	0.500	-	1.500	-	-0.700	-	1.000	-
Starting Values	0.0	1.000	0.0	0.500	0.0	1.500	0.0	0.700	3.0	2.000
$\hat{\beta}$ (OLS)	0.989	0.036	0.915	0.420	1.036	0.455	-0.277	0.425	1.543	0.556
$\hat{\beta}_1^1$	0.994	0.041	0.474	0.088	1.496	0.022	-0.694	0.017	1.073	0.096
$\hat{\beta}_2^2$	1.011	0.046	0.472	0.105	1.506	0.030	-0.705	0.026	1.110	0.185
$\hat{\beta}_3^3$	0.819	1.872	1.075	1.845	1.170	0.780	-0.520	0.440	<u>DIVERGES</u>	

TABLE 5.2. Comparison of Bootstrap Estimators $\hat{\beta}_1^1$, $\hat{\beta}_2^2$ and $\hat{\beta}_3^3$ -
Second Order Model and No Filter

SECOND ORDER MODEL

OF DATA PAIRS = 1,000

OF TRIALS = 20

FILTER (A) - NON-ADAPTIVE

ESTIMATOR	$\hat{\beta}_1$	RMSE	$\hat{\beta}_2$	RMSE	$\hat{\beta}_3$	RMSE	$\hat{\beta}_4$	RMSE	$\bar{\chi}^2$	RMSE
True Values	1.000	-	0.500	-	1.500	-	-0.700	-	1.000	-
Starting Values	0.0	1.000	0.0	0.500	0.0	1.500	0.0	0.700	3.0	2.000
$\hat{\beta}$ (OLS)	0.989	0.036	0.915	0.420	1.036	0.455	-0.277	0.425	1.543	0.556
$\hat{\beta}_1^1$	0.993	0.040	0.477	0.087	1.492	0.024	0.689	0.020	1.089	0.115
$\hat{\beta}_2^2$	1.013	0.047	0.467	0.106	1.507	0.031	-0.706	0.026	1.096	0.180
$\hat{\beta}_3^3$	1.015	0.050	0.465	0.118	1.510	0.036	-0.709	0.031	1.113	0.185

TABLE 5.3. Comparison of Bootstrap Estimators $\hat{\beta}_1^1$, $\hat{\beta}_2^2$ and $\hat{\beta}_3^3$ -
Second Order Model and Non-adaptive Filter

SECOND ORDER MODEL

OF DATA PAIRS = 1,000

OF TRIALS = 20

FILTER (B) - ADAPTIVE

ESTIMATOR	$\hat{\beta}_1$	RMSE	$\hat{\beta}_2$	RMSE	$\hat{\beta}_3$	RMSE	$\hat{\beta}_4$	RMSE	$\hat{\chi}^2$	RMSE
True Values	1.000	-	0.500	-	1.500	-	-0.700	-	1.000	-
Starting Values	0.0	1.000	0.0	0.500	0.0	1.500	0.0	0.700	3.0	2.000
$\hat{\beta}$ (OLS)	0.989	0.036	0.915	0.420	1.036	0.455	-0.277	0.425	1.543	0.556
$\hat{\beta}^1$	0.994	0.040	0.475	0.089	1.494	0.024	-0.692	0.019	1.084	0.110
$\hat{\beta}^2$	1.010	0.050	0.472	0.110	1.506	0.030	-0.705	0.027	1.108	0.185
$\hat{\beta}^3$	1.018	0.055	0.458	0.127	1.512	0.038	-0.711	0.032	1.141	0.237

TABLE 5.4. Comparison of Bootstrap Estimators $\hat{\beta}^1$, $\hat{\beta}^2$ and $\hat{\beta}^3$ -
Second Order Model and Adaptive Filter

FOURTH ORDER MODEL

OF DATA PAIRS = 1,000

OF TRIALS = 20

ESTIMATOR	$\tilde{\beta}_1$	RMSE	$\tilde{\beta}_2$	RMSE	$\tilde{\beta}_3$	RMSE	$\tilde{\beta}_4$	RMSE	$\tilde{\beta}_5$	RMSE	$\tilde{\lambda}^2$	RMSE
True Values	1.000	-	1.000	-	-0.180	-	0.784	-	-0.656	-	1.000	-
Starting Values	0.0	1.000	0.0	1.000	0.0	0.180	0.0	0.784	0.0	0.656	3.0	2.000
$\tilde{\beta}^0$ (OLS)	1.017	0.040	0.680	0.340	0.175	0.355	0.373	0.410	-0.284	0.360	1.373	0.385
$\tilde{\beta}^1$	0.994	0.048	1.004	0.018	-0.191	0.034	0.797	0.037	-0.661	0.021	0.983	0.035
$\tilde{\beta}^2$	0.993	0.023	1.012	0.030	-0.191	0.033	0.786	0.027	-0.656	0.016	0.982	0.034
$\tilde{\beta}^3$	0.993	0.023	1.012	0.043	-0.195	0.055	0.795	0.044	-0.662	0.028	0.993	0.034

TABLE 5.5. Comparison of Idealized Estimator $\tilde{\beta}^1$, $\tilde{\beta}^2$ and $\tilde{\beta}^3$ - Fourth Order Model

FOURTH ORDER MODEL

OF DATA PAIRS = 1,000

OF TRIALS = 20

NO FILTER

ESTIMATOR	$\hat{\beta}_1$	RMSE	$\hat{\beta}_2$	RMSE	$\hat{\beta}_3$	RMSE	$\hat{\beta}_4$	RMSE	$\hat{\beta}_5$	RMSE	$\hat{\chi}^2$	RMSE
True Values	1.000	-	1.000	-	-0.180	-	0.784	-	-0.656	-	1.000	-
Starting Values	0.0	1.000	0.0	1.000	0.0	0.18	0.0	0.784	0.0	0.656	3.0	2.000
$\hat{\beta}$ (OLS)	1.017	0.040	0.680	0.340	0.175	0.355	0.373	0.410	-0.284	0.360	1.373	0.385
$\hat{\beta}_1$	0.995	0.048	0.991	0.023	-0.172	0.035	0.770	0.039	-0.641	0.027	1.255	0.340
$\hat{\beta}_2$	0.994	0.0173	1.009	0.023	-0.200	0.036	0.789	0.040	-0.649	0.031	1.340	0.485
$\hat{\beta}_3$	2.333	4.000	0.787	0.495	-0.150	0.115	0.672	0.31	-0.489	0.430	<u>DIVERGES</u>	

TABLE 5.6. Comparison of Bootstrap Estimators $\hat{\beta}^1$, $\hat{\beta}^2$ and $\hat{\beta}^3$ - Fourth Order Model and No Filter

FOURTH ORDER MODEL

OF DATA PAIRS = 1,000

OF TRIALS = 20

FILTER (A) - NON-ADAPTIVE

ESTIMATOR	$\hat{\beta}_1$	RMSE	$\hat{\beta}_2$	RMSE	$\hat{\beta}_3$	RMSE	$\hat{\beta}_4$	RMSE	$\hat{\beta}_5$	RMSE	$\hat{\lambda}^2$	RMSE
True Values	1.000	-	1.000	-	-0.180	-	0.784	-	-0.656	-	1.000	-
Starting Values	0.0	1.000	0.0	1.000	0.0	0.180	0.0	0.784	0.0	0.656	3.0	2.000
$\hat{\beta}$ (OLS)	1.017	0.040	0.680	0.340	0.175	0.355	0.373	0.410	-0.284	0.360	1.373	0.385
$\hat{\beta}_1^1$	0.995	0.048	0.987	0.025	-0.168	0.038	0.765	0.041	-0.636	0.030	1.280	0.365
$\hat{\beta}_2^2$	0.994	0.019	1.009	0.023	-0.200	0.036	0.787	0.038	-0.649	0.027	1.168	0.205
$\hat{\beta}_3^3$	1.038	0.15	0.989	0.05	-0.194	0.055	0.785	0.084	-0.634	0.115	<u>DIVERGES</u>	

TABLE 5.7. Comparison of Bootstrap Estimators $\hat{\beta}_1^1$, $\hat{\beta}_2^2$, and $\hat{\beta}_3^3$ - Fourth Order Model with Non-adaptive Filter

FOURTH ORDER MODEL

OF DATA PAIRS = 1,000

OF TRIALS = 20

FILTER (B) - ADAPTIVE

ESTIMATOR	$\hat{\beta}_1$	RMSE	$\hat{\beta}_2$	RMSE	$\hat{\beta}_3$	RMSE	$\hat{\beta}_4$	RMSE	$\hat{\beta}_5$	RMSE	$\hat{\lambda}^2$	RMSE
True Values	1.000		1.000		-0.180		0.784		-0.656		1.000	
Starting Values	0.0	1.000	0.0	1.000	0.0	0.180	0.0	0.784	0.0	0.656	3.0	2.000
$\hat{\beta}$ (OLS)	1.017	0.040	0.680	0.340	0.175	0.355	0.373	0.410	-0.284	0.360	1.373	0.385
$\hat{\beta}^1$	0.995	0.047	0.987	0.026	-0.168	0.039	0.764	0.043	-0.636	0.031	1.320	0.420
$\hat{\beta}^2$	0.992	0.017	1.011	0.024	-0.201	0.038	0.791	0.040	-0.653	0.027	1.170	0.185
$\hat{\beta}^3$	0.996	0.019	1.005	0.029	-0.196	0.050	0.802	0.049	-0.663	0.031	1.224	0.25

TABLE 5.8. Comparison of Bootstrap Estimators $\hat{\beta}^1$, $\hat{\beta}^2$, and $\hat{\beta}^3$ - Fourth Order Model and Adaptive Filter

Fig. 5.6. Effect of Adaptive Filter on $\hat{\lambda}^2(k)$ - Bootstrap Estimator $\hat{\beta}_x^1$ Applied to Second Order Model

Fig. 5.7. Effect of Adaptive Filter on $\hat{\lambda}^2(k)$ - Bootstrap Estimator $\hat{\beta}_x^2$ Applied to Second Model

Fig. 5.8. Effect of Adaptive Filter on $\hat{\lambda}^2(k)$ - Bootstrap Estimator $\hat{\beta}^3$ Applied to Second Order Model

Fig. 5.9. Effect of Adaptive Filter on $\hat{\lambda}^2(k)$ - Bootstrap Estimator $\hat{\beta}_x^1$ Applied to Fourth Order Model

Fig. 5.10. Effect of Adaptive Filter on $\hat{\lambda}^2(k)$ - Bootstrap Estimator $\hat{\beta}^2$ applied to Fourth Order Model

Fig. 5.11. Effect of Adaptive Filter on $\hat{\lambda}^2(k)$ - Bootstrap Estimator $\hat{\beta}^3$ Applied to Fourth Order Model

CHAPTER 6

ESTIMATION OF PARAMETERS FOR THE ENTIRE MODEL

6.1 Preliminary Remarks

As discussed in Section 3.1, system identification generally entails the estimation of the process parameters $(b_0, b_1, \dots, b_n, -a_1, -a_2, \dots, -a_n)$ as well as the noise parameters $(c_1, c_2, \dots, c_n)$ in the difference equation model:

$$\begin{aligned} y(k) = & b_0 u(k) + b_1 u(k-1) + \dots + b_n u(k-n) \\ & - a_1 y(k-1) - a_2 y(k-2) - \dots - a_n y(k-n) \\ & + c_1 e(k-1) + c_2 e(k-2) + \dots + c_n e(k-n) + e(k) \end{aligned} \quad (6.1)$$

$$e(k) \sim \text{N.I.D. } (0, \lambda^2)$$

or the estimation of the unknown elements of the matrices F , $\underline{g}$, d and $\underline{\alpha}$ in the Levy's canonical model:

$$\begin{aligned} \underline{x}(k+1) &= F \underline{x}(k) + \underline{g} u(k) + \underline{\alpha} e(k) \\ y(k) &= \underline{h}' \underline{x}(k) + d u(k) + e(k) \\ e(k) &\sim \text{N.I.D. } (0, \lambda^2) \end{aligned} \quad (6.2)$$

where $\underline{x}(k)$ is the n vector of phase variables and $\underline{h}' = [1 \ 0 \ 0 \ \dots \ 0]$.

An estimate of λ^2 is also required, if interval estimation (error analysis) on the estimated parameters is to be carried out and/or the identified model is to be used for the design of a suitable control strategy.

The process and noise parameters may either be estimated simultaneously or one may first choose to obtain consistent estimates of the process parameters which may be subsequently used to estimate the noise parameters. So far, we have confined our attention to the estimation of process parameters only. In Chapters 3 and 4, consistent estimators for process parameters in model (6.1) were presented. In Chapter 5, we considered a simple, though not infrequent special case of model (6.1), where $c_i \equiv a_i$; $i = 1, 2, \dots, n$ (or $\underline{\alpha} \equiv \underline{0}$ in model (6.2)), which represents a linear dynamic system subject to additive white noise at the output.

In this chapter, the following problems will be considered:

1. estimation of noise parameters, given consistent estimates of the process parameters,
2. simultaneous estimation of process and noise parameters in a single experiment.

6.2 Estimation of Noise Parameters, Given Consistent Estimates of the Process Parameters

In model (6.1), if the process parameters ($b_0, b_1, \dots, b_n, -a_1, -a_2, \dots, -a_n$) are known, and since $u(k), y(k)$; $k = 1, 2, \dots, N$ constitute the measurements, the problem of estimating the noise parameters ($c_1, c_2, \dots, c_n$) is equivalent to estimating the parameters of the following n^{th} order moving average (MA) process:

$$v(k) = c_1 e(k-1) + c_2 e(k-2) + \dots + c_n e(k-n) + e(k) \quad (6.3)$$

$$e(k) \sim \text{N.I.D. } (0, \lambda^2)$$

in which the variables on the left hand side

$$v(k) = y(k) + a_1 y(k-1) + \dots + a_n y(k-n) - b_0 u(k) - \dots - b_n u(k-n) \quad (6.4)$$

are the only known quantities. For simplicity we shall assume that the initial conditions on the MA process (6.3) are identically zero, i.e.,

$$e(j) = 0 \quad \forall \quad j \leq 0 \quad (6.5)$$

In obtaining the moving average process (6.3), it was assumed that the true process parameters are known, so that exact values of $v(k)$ can be calculated from Eqn. (6.4). In practice, however, we may be constrained to work with estimates of the process parameters provided by a suitable consistent estimator. In the latter case $v(k)$; $k = 1, 2, \dots, N$ in Eqn. (6.3) are replaced by their consistent* estimates $\hat{v}(k)$. A consistent estimator for the noise parameters $(c_1, c_2, \dots, c_n)$ that is a continuous function of $v(k)$; $k = 1, 2, \dots, N$, will continue to be consistent when $v(k)$ are replaced by their consistent estimates $\hat{v}(k)$. This leads to the conclusions:

* The term consistent estimates is used here in the sense that the estimates are generated by a consistent estimator.

1. Availability of consistent estimates for the process parameters is a pre-requisite, if consistent estimates of the noise parameters are to be obtained from the model.

$$\hat{v}(k) = c_1 e(k-1) + c_2 e(k-2) + \dots + c_n e(k-n) + e(k)$$

in which $\hat{v}(k)$ are calculated from

$$\hat{v}(k) = y(k) + \hat{a}_1 y(k-1) + \dots + \hat{a}_n y(k-n) - \hat{b}_0 u(k) - \dots - \hat{b}_n u(k-n)$$

2. There is no loss in generality, if the exact model (6.3) is used for formulating the required estimators for the noise parameters. This will keep the notation from becoming unnecessarily complicated.

The estimators for the noise parameters presented in this section are based on the important assumption that the roots of the algebraic equation:

$$q^n + c_1 q^{n-1} + c_2 q^{n-2} + \dots + c_n q^0 = 0 \quad (6.6)$$

are located inside the origin centred unit circle. In other words, the linear system represented by the z-Transfer Function:

$$\frac{1}{C(z^{-1})} = \frac{1}{1 + c_1 z^{-1} + c_2 z^{-2} + \dots + c_n z^{-n}} \quad (6.7)$$

is stable.

A. Method of Spectrum Factorization

The moving average process (6.3), with initial conditions (6.5) may be written as

$$\underline{v} = \Delta \underline{e} \quad (6.8)$$

where

$$\underline{v} = [v(1) \ v(2) \ \dots \ v(N)]' \quad (6.9)$$

$$\underline{e} = [e(1) \ e(2) \ \dots \ e(N)]' \quad (6.10)$$

and

$$\Delta = \begin{bmatrix} 1 & & & & & & & & & \\ c_1 & 1 & & & & & & & & \\ c_2 & c_1 & 1 & & & & & & & \\ \vdots & \vdots & \vdots & \ddots & \vdots & & & & & \\ c_n & c_{n-1} & \dots & c_1 & 1 & & & & & \\ \bigcirc & \text{---} & \text{---} & \text{---} & c_n & c_{n-1} & \dots & c_1 & 1 & \end{bmatrix} \quad (6.11)$$

It is shown in Appendix B that the Var - Covar matrix for $\underline{v}$ is given by:

$$E[\underline{v} \underline{v}'] = \Omega = \lambda^2 \Delta \Delta' \quad (6.12)$$

where Δ is the unit lower triangular matrix defined by (6.11).

A consistent estimate $\hat{\Omega}$ of Ω can be obtained by estimating its individual elements by their sample correlations* as

$$\frac{1}{N-n} \sum_{i=1}^N v(i)v(i+j) \xrightarrow{P} E[v(k)v(k+j)] \quad (6.13)$$

where it is not necessary to estimate all $N \times N$ elements of Ω - firstly because Ω is symmetric and secondly because $v(k)$ being a moving average process of order n implies that:

$$\frac{1}{N-n} \sum_{i=1}^N v(i)v(i+j) \xrightarrow{P} E[v(k)v(k+j)] = 0 \quad \forall j > n \quad (6.14)$$

Once all the unknown elements of Ω have been estimated, we may write

$$\hat{\Omega} = \hat{\lambda}^2 \hat{\Delta} \hat{\Delta}' \quad (6.15)$$

and if the LUD decomposition represented by (6.15) can be found and if such a decomposition is unique, we would have the required estimates of the noise parameters, $(c_1, c_2, \dots, c_n)$ and λ^2 .

Alternatively, this problem may be considered as that of designing an optimum (realizable) whitening filter for the stochastic process $v(k)$. Since the auto-covariance structure of the input ($\hat{\Omega} = \hat{\lambda}^2 \hat{\Delta} \hat{\Delta}'$) and the cross-covariance structure between the input and output ($\hat{\lambda}^2$) is

* It may be recalled that $v(k)$ are zero mean.

known (within the constant factor $\hat{\lambda}^2$), such a filter may be designed either in the frequency domain or in the time domain. In the frequency domain, the procedure would involve the factorization of the discrete process spectral density matrix generated by $\hat{\Omega}$.

Rowe³⁵ has successfully applied the spectrum factorization technique for the estimation of noise parameters and has outlined in detail, the conditions for uniqueness and stability of the resulting whitening filter.

B. The Two Stage Least Squares (2SLS) Method

The basic difficulty in estimating the parameters of the moving average process (6.3), arises from the fact that the so-called independent variables $e(k-1), e(k-2), \dots, e(k-n)$ on the right hand side are not known. If consistent estimates of these variables can be obtained, then model (6.3) may be replaced by

$$v(k) = c_1 \hat{e}(k-1) + c_2 \hat{e}(k-2) + \dots + c_n \hat{e}(k-n) + \hat{e}(k) + v(k) \quad (6.16)$$

where $v(k)$ takes care of the errors introduced by using $\hat{e}(k), \hat{e}(k-1), \dots, \hat{e}(k-n)$ instead of their true values. OLS may now be applied to model (6.16), to obtain the required estimates of the noise parameters. Thus the first stage in the 2SLS procedure would consist of obtaining the estimates of $e(k)$; $k = 1, 2, \dots, N$, and in the second stage the noise parameters are estimated by applying OLS to the modified model (6.16). The procedure is outlined below:

Stage 1: Estimate $e(k)$; $k = 1, 2, \dots, N$

(a) Convert the n^{th} order moving average process

$$v(k) = C(q^{-1}) e(k) \quad (6.17)$$

$$C(q^{-1}) = 1 + c_1 q^{-1} + c_2 q^{-2} + \dots + c_n q^{-n}$$

into the infinite order auto-regressive (AR) process

$$e(k) = \left[\frac{1}{C(q^{-1})} \right] v(k) \quad (6.18)$$

or

$$e(k) = \Gamma(q^{-1}) v(k) \quad (6.19)$$

where

$$\Gamma(q^{-1}) = 1 + \gamma_1 q^{-1} + \gamma_2 q^{-2} + \dots + \gamma_s q^{-s} + \dots$$

or

$$e(k) = v(k) + \gamma_1 v(k-1) + \dots + \gamma_s v(k-s) + \dots \quad (6.20)$$

(b) Under the stability assumption for $C(q^{-1})$ (Eqns. (6.6), (6.7)) truncate the above AR process after a suitable number of terms (say at $v(k-s)$; $s > n$) to give

$$v(k) = -\gamma_1 v(k-1) - \gamma_2 v(k-2) - \dots - \gamma_s v(k-s) + e(k) \quad (6.21)$$

$$k = 1, 2, \dots, N$$

or

$$\underline{v} = -V \underline{\gamma} + \underline{e} \quad (6.22)$$

(c) Estimate $\underline{\gamma} = [\gamma_1 \ \gamma_2 \ \dots \ \gamma_s]$ by the OLS estimator

$$\underline{\gamma} = - (V'V)^{-1} V' \underline{v} \quad (6.23)$$

(d) Estimate the vector $\underline{e}$ by the residuals

$$\hat{\underline{e}} = \underline{v} + V \hat{\underline{\gamma}} \quad (6.24)$$

We now have consistent estimates of $e(k)$, provided the dimension of $\underline{\gamma}$ is large enough to include all the pertinent impulse response coefficients of model (6.18). We are therefore ready to go to the second stage.

Stage 2: Estimate c_i ; $i = 1, 2, \dots, n$

(a) Write model (6.3) in terms of the estimates of $e(k)$ obtained in stage 1, i.e.,

$$v(k) = c_1 \hat{e}(k-1) + c_2 \hat{e}(k-2) + \dots + c_n \hat{e}(k-n) + \hat{e}(k) + \underline{v}(k) \quad (6.25)$$

or

$$\underline{v} = \hat{E} \underline{\delta} + (\hat{\underline{e}} + \underline{v}) \quad (6.26)$$

where

$$\underline{\delta} = [c_1 c_2, \dots, c_n]' \quad (6.27)$$

and $\hat{E}$ is the design matrix resulting from (6.25).

(b) Estimate the vector of noise parameters $\underline{\delta}$ by the OLS estimator

$$\underline{\hat{\delta}} = (\hat{E}'\hat{E})^{-1} \hat{E}' \underline{y} \quad (6.28)$$

Alternatively, since the vector $\underline{\hat{e}}$ is known, it can be brought to the left hand side of Eqn. (6.26) leading to the estimator

$$\underline{\hat{\delta}}^* = (\hat{E}'\hat{E})^{-1} \hat{E}'(\underline{y} - \underline{\hat{e}}) \quad (6.29)$$

which would generally provide better results than estimator (6.28). However, as we shall presently see, if $e(k)$ are to be estimated directly from model (6.3) (and not from the reduced form model (6.21)), we will be restricted to use estimator (6.28).

Results from simulation experiments presented later in this section indicate that the 2SLS technique is very effective. However, the method tends to be computationally inefficient, particularly if the roots of the algebraic Eqn. (6.6) are too close to (though inside) the origin centred unit circle, thus making the dimension of $\underline{y}$ excessively large.

C. Recursive Approximation to 2SLS - Bootstrap Estimator

Considerable saving in computation time and storage can be achieved, if it were possible to estimate both the noise parameters $\underline{\delta}$ as well as the unknown disturbance terms $e(k)$; $k = 1, 2, \dots, N$, directly

from model (6.3). The problem is analogous to the one considered in Chapters 4 and 5, where the process parameters $\underline{\beta}$ as well as the noise-free process outputs $x(k)$ were estimated on-line from the structural model - without using the reduced form model for estimating $x(k)$.

To achieve this objective in the present case, assume for the moment, that estimates of $e(k)$; $k = 1, 2, \dots, N$ in model (6.3) are known, so that we may consider the model:

$$v(k) = c_1 \hat{e}(k-1) + c_2 \hat{e}(k-2) + \dots + c_n \hat{e}(k-n) + \hat{e}(k) + v(k)$$

$$k = 1, 2, \dots, N$$

or

$$\underline{v} = \hat{E} \underline{\delta} + (\underline{\hat{e}} + \underline{v}) \quad (6.30)$$

which leads to the OLS estimator:

$$\underline{\hat{\delta}} = (\hat{E}' \hat{E})^{-1} \hat{E}' \underline{v} \quad (6.31)$$

If however, we use the recursive form of the estimator (6.31), then the requirement that all the estimates of $e(k)$; $k = 1, 2, \dots, N$ be known a priori can be dropped - provided these estimates are made available (one at a time) at each sampling instant. Supposing that we are at the sampling instant k (where k could be the time at which the recursive estimator is started, i.e., $k = 1$) and the measurement $v(k)$ has just been received. The value of $\hat{\delta}(k)$ can be estimated from the recursive relation

$$\hat{\underline{\delta}}(k) = \hat{\underline{\delta}}(k-1) + \frac{S(k-1) \hat{\underline{\epsilon}}(k) [v(k) - \hat{\underline{\epsilon}}'(k) \hat{\underline{\delta}}(k-1)]}{1 + \hat{\underline{\epsilon}}'(k) S(k-1) \hat{\underline{\epsilon}}(k)} \quad (6.32)$$

where

$$\hat{\underline{\epsilon}}(k) = [\hat{e}(k-1) \hat{e}(k-2) \dots \hat{e}(k-n)]' \quad (6.33)$$

and

$$S(k-1) = (\hat{E}'\hat{E})^{-1} \text{ at time } (k-1) \quad (6.34)$$

provided $\hat{\underline{\epsilon}}(k)$, $S(k-1)$ and $\hat{\underline{\delta}}(k-1)$ are available as initial conditions.

To estimate $\hat{\underline{\delta}}(k+1)$ from $\hat{\underline{\delta}}(k)$, at the next measurement, we require $S(k)$, and $\hat{\underline{\epsilon}}(k+1)$. The former is given by

$$S(k) = S(k-1) - \frac{S(k-1) \hat{\underline{\epsilon}}(k) \hat{\underline{\epsilon}}'(k) S(k-1)}{1 + \hat{\underline{\epsilon}}'(k) S(k-1) \hat{\underline{\epsilon}}(k)} \quad (6.35)$$

The vector $\hat{\underline{\epsilon}}(k+1)$ consists of

$$\hat{\underline{\epsilon}}(k+1) = [\hat{e}(k) ; \hat{e}(k-1) \dots \hat{e}(k-n+1)]' \quad (6.36)$$

in which all the elements except the first are known. The unknown term $\hat{e}(k)$ may be obtained from the structural model as

$$\hat{e}(k) = v(k) - \hat{\underline{\epsilon}}'(k) \hat{\underline{\delta}}(k) \quad (6.37)$$

or

$$\hat{e}(k) = v(k) - \hat{c}_1(k)\hat{e}(k-1) - \hat{c}_2(k)\hat{e}(k-2) - \dots - \hat{c}_n(k)\hat{e}(k-n) \quad (6.38)$$

Obviously, the on-line algorithm represented by the Eqns. (6.32), (6.35) and (6.37) constitutes a bootstrap estimator, in which we again have two recursive estimators - one for the noise parameters $\underline{\delta}$ (structural parameters) and the other for the unknown error terms $e(k)$ (the incidental parameters for this case), with continuous updating of one estimator with the help of the other, and vice-a-versa.

When these two recursive estimators are considered separately, the stability and consistency of either one of them may be argued, provided the other one is assumed to satisfy these properties. However, the stability of the overall bootstrap estimator is not obvious, because the behavior of one of the estimators affects the operation of the other.

The variance of $e(k)$ can be estimated as

$$\hat{\lambda}^2(N) = \frac{1}{N} \sum_{i=1}^N [\hat{e}(i)]^2 \quad (6.39)$$

which has the recursive form:

$$\hat{\lambda}^2(k) = \left(\frac{k-1}{k} \right) \hat{\lambda}^2(k-1) + \frac{1}{k} [\hat{e}(k)]^2 \quad (6.40)$$

The recursive algorithm represented by Eqns. (6.32), (6.35) and (6.37) may be implemented in practice, as shown in Fig.6.1. Use of $\hat{v}(k)$ instead of $v(k)$ in the figure indicates that they are obtained by using consistent estimates of the process parameters $\underline{\beta}$ (in Eqn. (6.4)) furnished by a separate estimator, e.g., one of the IV estimators discussed in Chapter 4.

Fig. 6.1. Bootstrap Estimator for Estimation of Noise Parameters $\underline{\delta}$

Results

The 2SLS estimators represented by Eqns. (6.29) and (6.28) (in which the estimates of $e(k)$; $k = 1, 2, \dots, N$ are provided by Eqns. (6.23-24)) and the recursive approximation to the 2SLS represented by Eqns. (6.32), (6.35) and (6.37) (Fig. 6.1) were tested on simulated data. For the purpose of presenting the results, these three estimators are respectively designated as:

$$\text{2SLS - (1)} \rightarrow \hat{\underline{\delta}}^1 = (\hat{E}'\hat{E})^{-1} \hat{E}'(\hat{\underline{v}} - \hat{\underline{e}})$$

$$\text{2SLS - (2)} \rightarrow \hat{\underline{\delta}}^2 = (\hat{E}'\hat{E})^{-1} \hat{E}' \hat{\underline{v}}$$

$$\text{Bootstrap} \rightarrow \hat{\underline{\delta}}^3 = \text{Recursive Approximation to 2SLS}$$

To this author's knowledge, the above three estimators have not been used before, for the estimation of noise parameters in model (6.1). A recursive estimator based on 'adaptive recursive least square estimator' (discussed in the next section) has been suggested by Young⁴⁴. However, no results have been presented to indicate its convergence properties.

The data, for testing the above three estimators, were generated from the following second order model due to Astrom¹²:

$$y(k) = u(k-1) + 0.5 u(k-2) + 1.5 y(k-1) - 0.7 y(k-2) - e(k-1) + 0.2 e(k-2) + e(k)$$

$$e(k) \sim \text{N.I.D. } (0, \lambda^2); \quad \lambda^2 = 1.$$

$$u(k) \sim \text{PRBS}(1,1,127) + \text{PRBS}\left(\frac{1}{2}, 127, 889\right)$$

The basic flow diagram used for implementing these three estimators is given in Fig. 6.2. As before, 1,000 sample points were used and each estimator was tested for 20 different noise sequences.

As shown in Fig. 6.2, in the first experiment, consistent estimates of the process parameters $\underline{\beta}$ were obtained by using the recursive IV algorithm ($\hat{\underline{\beta}}_{\text{IV}}$ - Pandya) presented in Chapter 4. These estimates were used to generate $\hat{v}(k)$; $k = 1, 2, \dots, N$ using Eqn. (6.4). For 2SLS-(1) and 2SLS-(2), the estimates for $\underline{\gamma}$ and $\underline{\delta}$ could have been obtained by batch processing $\hat{v}(k)$ and $\hat{e}(k)$; $k = 1, 2, \dots, N$ respectively. However, these estimates were obtained by using the corresponding recursive forms for estimators $\hat{\underline{\gamma}}$ and $\hat{\underline{\delta}}^1$ and $\hat{\underline{\delta}}^2$ - firstly to avoid the matrix inversions and secondly to get some idea about the rate of convergence of estimators $\hat{\underline{\delta}}^1$ and $\hat{\underline{\delta}}^2$. For each estimator, λ^2 was estimated by the recursive relation (6.40).

Results for $N = 1,000$ that are averaged over 20 trials are given in Table 6.1, which also includes the sample RMS error (over 20 trials) for the estimates, and the approximate time for processing 1,000 pieces of data (i.e., $\hat{v}(k)$; $k = 1, 2, \dots, 1,000$).

To provide some indication of the convergence rate of the three estimators, the Total RMS error normalized to the initial ($k = 0$) Total RMS error is plotted against the number of data points in Fig. 6.3.

Fig. 6.2. Flow Diagram for Implementing 2SLS-(1), 2SLS-(2) and the Bootstrap Estimators for Noise-Parameters

OF DATA PAIRS = 1,000; # OF TRIALS = 20

PROCESS PARAMETERS: STARTING VALUES: $\hat{\beta}(0) = \underline{0}$

ESTIMATOR	$\hat{\beta}_1$	RMSE	$\hat{\beta}_2$	RMSE	$\hat{\beta}_3$	RMSE	$\hat{\beta}_4$	RMSE
True Values	1.000	-	0.500	-	1.500	-	-0.700	-
$\hat{\beta}_{IV}$ - Pandya	1.007	0.035	0.473	0.059	1.501	0.017	-0.699	0.013

NOISE PARAMETERS: STARTING VALUES: $\hat{\delta}(0) = \underline{0}$

ESTIMATOR	$\hat{\delta}_1$	RMSE	$\hat{\delta}_2$	RMSE	$\hat{\lambda}^2$	RMSE	TIME 1,000 SAMPLES
True Values	-1.000	-	0.200	-	1.000	-	-
2SLS-(1)	-0.998	0.036	0.197	0.048	1.005	0.035	~ 36 secs
2SLS-(2)	-1.000	0.037	0.196	0.049	0.993	0.035	~ 36 secs
Bootstrap	-0.982	0.046	0.166	0.072	1.003	0.034	~ 2.0 secs

TABLE 6.1. Estimates of Noise Parameters Obtained from (i) 2SIS-(1), (ii) 2SIS-(2) and (iii) Bootstrap Estimators

Fig. 6.3 Convergence of 2SLS-(1), 2SLS-(2) and the Bootstrap Estimator for Noise Parameters - II Order Model - Astrom

It is evident from the results (Table 6.1, Fig. 6.3) that the 2SLS-(1) provides, by far, the best estimates (among the three estimators tested) of the noise parameters $\underline{\delta}$. However, compared to the Bootstrap estimator (whose rate of convergence is rather slow), the 2SLS estimators are computationally very inefficient.

6.3 Simultaneous Estimation of Process and Noise Parameters in a Single Experiment - Bootstrap Estimators

In the preceding section, some estimators for the noise parameters $\underline{\delta} = [c_1 \ c_2, \dots, c_n]'$ in model (6.1) were considered. These estimators require that consistent estimates of the process parameters $\underline{\beta} = [b_0 \ b_1 \ \dots \ b_n \ -a_1 \ -a_2 \ \dots \ -a_n]'$, should be available. This procedure for identifying the parameters of the entire model would generally require more than one experiment or pass.

An experiment (or pass) here refers to 'an operation either on the complete set of actual measurements or on a sequence of random variables that are functionally dependent on the measurements'. For example, the entire set of measured data may have to be used to estimate the process parameters. These estimates are then used to generate a second sequence of random variables $\hat{v}(k)$ (Eqn. (6.4)) that are functionally dependent on the original measurements. The noise parameters are then estimated by a suitable operation on this second set of random variables, and would therefore, constitute a separate experiment. Even if the process parameters are estimated recursively, so that $\hat{v}(k)$ can

also be obtained recursively, the situation remains unchanged as long as $\hat{v}(k)$; $k = 1, 2, \dots, N$ are stored for future use.*

The term 'simultaneous' estimation of process and noise parameters means that 'at the end of each iteration or recursion (as the case may be), the estimator should provide an estimate of the parameter vector for the entire model'. If the model parameters are being estimated on-line, this definition does not exclude the use of two separate estimators (one for $\underline{\beta}$ and the other for $\underline{\delta}$), provided an estimate of the process as well as noise parameters is obtained before the next pair of input-output measurements is brought in.

In view of the above discussion, the mode of operation of a given estimator for the system parameters $\underline{\theta}' = [\underline{\beta}' ; \underline{\delta}']$ may be classified under one of the following four categories:

Category 1. Neither simultaneous nor one pass

The estimators that have been considered so far, wherein the process and noise parameters are estimated separately and in separate experiments belong to this set.

Category 2. Simultaneous but not one pass

An example of an estimator in this category would be the Maximum Likelihood method of Astrom and Bohlin¹². Though the iterative mini-

* If the method of Spectrum Factorization is to be used for estimating the noise parameters, then sample correlations of $\hat{v}(k)$ would be stored.

mization routine used in this technique provides an estimate of the entire parameter vector $\underline{\theta}$ at the end of each iteration, the number of passes (or iterations) on the measured data is, in general, greater than one.

Category 3. One pass but not simultaneous

If the record of measured data is sufficiently long, Young⁴⁴ suggests that a part of the data may be used to estimate the process parameters by using a recursive IV algorithm. These estimates of $\underline{\beta}$, together with the remaining part of the data are then used to estimate the noise parameters $\underline{\delta}$ using a recursive estimator. Though the available measurements are processed only once, estimates of the noise parameters are not available while the first part of the data are being processed. Unfortunately, Young does not provide any results and it is not clear how the available data should be divided to give the best possible estimates.

Category 4. Simultaneous and one pass

If the process is to be controlled on-line, then it becomes necessary that the system parameters $\underline{\theta}$ should be estimated by a technique that belongs to this category. Of the existing procedures, the Recursive GLS method of Hastings-James and Sage²⁴ and the Adaptive Recursive Least Squares algorithm of Panuska⁴⁵, provide on-line estimates of both the process as well as noise parameters.

For the difference equation model (6.1) under discussion, the Recursive GLS technique proposed by Hastings-James and Sage would provide on-line estimates of the process parameters $\underline{\beta}$ and the parameters $\underline{\gamma}$ of the

auto-regressive process defined by (6.21) - and not the parameters $\underline{\delta}$ of the MA process (6.3). The technique, however, is useful for the estimation of process parameters $\underline{\beta}$ and demonstrates the use of approximate GLS techniques for estimation of $\underline{\beta}$.

The Adaptive Recursive Least Squares algorithm proposed by Panuska (also by Young^{4,6}) is based on the fact that in the difference equation model:

$$\begin{aligned} y(k) = & b_0 u(k) + b_1 u(k-1) + \dots + b_n u(k-n) \\ & - a_1 y(k-1) - a_2 y(k-2) - \dots - a_n y(k-n) \\ & + c_1 e(k-1) + c_2 e(k-2) + \dots + c_n e(k-n) + e(k) \end{aligned}$$

or

$$y(k) = \underline{u}'(k) \underline{\theta} + e(k); \quad k = 1, 2, \dots, N \quad (6.41)$$

$$e(k) \sim \text{N.I.D. } (0, \lambda^2)$$

if the error terms $e(k-1), e(k-2), \dots, e(k-n)$ were known, it would meet the requirements of the Mann and Wald model^A (with two inputs), so that consistent estimates of the system parameters $\underline{\beta}$ can be obtained by using an OLS estimator. Since the error terms $e(k); k = 1, 2, \dots, N$ are not known a priori, Panuska suggests the use of so-called 'computed errors' (i.e., estimates of $e(k)$) that can be generated on-line, if the recursive form of the OLS estimator for model (6.41) is used.

Before considering Panuska's algorithm which provides on-line estimates of the system parameters $\underline{\theta}$, a hypothetical case will be considered. Besides introducing the notation, this hypothetical case will also provide some insight into the statistical and computational properties of the on-line algorithm.

The Hypothetical Case

If it is assumed that the error terms $e(k-1), e(k-2), \dots, e(k-n)$ are known, then the hypothetical model represented by Eqn. (6.41) may be written in the vector-matrix notation as*

$$\underline{y} = [M \mid E] \underline{\theta} + \underline{e}; \quad \underline{e} \sim \text{N.I.D.}(\underline{0}, \lambda^2 I) \quad (6.42)$$

where

$$\underline{y} = [y(1) \ y(2) \ \dots \ y(N)]' \quad (6.43)$$

$$\underline{e} = [e(1) \ e(2) \ \dots \ e(N)]' \quad (6.44)$$

$$\begin{array}{c} \underline{\theta} \\ (3n+1) \times 1 \end{array} = \begin{bmatrix} \underline{\beta} \\ \underline{\delta} \end{bmatrix} = \begin{array}{c} [b_0 b_1 \ \dots \ b_n \ -a_1 - a_2 - \dots - a_n \mid c_1 c_2 \ \dots \ c_n]' \\ \begin{array}{cc} & 1 \times (2n+1) \end{array} & \begin{array}{cc} & 1 \times n \end{array} \end{array} \quad (6.45)$$

and $[M \mid E]$ is the design matrix corresponding to the design vector:

* Note that at each sampling instant k , the error term $e(k)$ is assumed unknown and it acts as the stochastic disturbance in the model.

$$\underline{u}(k) = [u(k)u(k-1)\dots u(k-n)y(k-1)\dots y(k-n) : e(k-1)\dots e(k-n)]' \quad (6.46)$$

OLS now applied to model (6.42) leads to the idealized estimator

$$\underline{\tilde{\theta}} = \begin{bmatrix} \underline{\tilde{\beta}} \\ \underline{\tilde{\delta}} \end{bmatrix} = \begin{bmatrix} M'M & M'E \\ E'M & E'E \end{bmatrix}^{-1} \begin{bmatrix} M'y \\ E'y \end{bmatrix} \quad (6.47)$$

or

$$\underline{\tilde{\theta}} = \underline{\theta} + \begin{bmatrix} \frac{1}{N} M'M & \frac{1}{N} M'E \\ \frac{1}{N} E'M & \frac{1}{N} E'E \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{N} M'e \\ \frac{1}{N} E'e \end{bmatrix} \quad (6.48)$$

where it is assumed that the design matrix $[M : E]$ has full rank and its elements are uniformly bounded.*

Under the assumption that the probability limit of the inverse matrix in (6.48) exists and since

$$\text{p.lim}_{N \rightarrow \infty} \begin{bmatrix} \frac{1}{N} M'e \\ \frac{1}{N} E'e \end{bmatrix} = \underline{0} \quad (6.49)$$

the estimator is consistent.**

* Since $e(k) \sim N.I.D.(0, \lambda^2)$, estimator (6.47) is also the Maximum Likelihood estimator.

** The fact that $(1/N M'e \xrightarrow{p} 0)$ was proved in Chapter 5 (Section 5.2) and $(1/N E'e \xrightarrow{p} 0)$ because $e(k)$ are i.i.d.

In practice however, the error terms in model (6.41) are not known and one may decide to use some suitable estimates to implement the estimator (6.47). Such an estimator will be consistent iff

$$\hat{e}(k) \xrightarrow{P} e(k)$$

i.e., iff the estimates of $e(k)$ used in (6.47) are consistent.

Further insight into the operation of the idealized estimator (6.47) may be obtained by considering the normal equations for the estimator, given by

$$\begin{bmatrix} M'y \\ E'y \end{bmatrix} = \begin{bmatrix} M'M & M'E \\ E'M & E'E \end{bmatrix} \begin{bmatrix} \tilde{\beta} \\ \tilde{\delta} \end{bmatrix} \quad (6.50)$$

which lead to the separate (but not independent) estimators for the process parameters $\underline{\beta}$ and the noise parameters $\underline{\delta}$:

$$\tilde{\underline{\beta}} = (M'M)^{-1} M' (\underline{y} - E'\tilde{\underline{\delta}}) = (M'M)^{-1} M' \underline{y}^* \quad (6.51)$$

$$\tilde{\underline{\delta}} = (E'E)^{-1} E' (\underline{y} - M'\tilde{\underline{\beta}}) = (E'E)^{-1} E' \tilde{\underline{y}} \quad (6.52)$$

where the observation vector $\underline{y}^*$, in (6.51) corresponds to the modified (or filtered) observations

$$y^*(k) = y(k) - \tilde{c}_1 e(k-1) - \tilde{c}_2 e(k-2) - \dots - \tilde{c}_n e(k-n) \quad (6.53)$$

and the observation vector $\tilde{v}$ in (6.52) corresponds to the MA process:

$$\tilde{v}(k) = y(k) + \tilde{a}_1 y(k-1) + \dots + \tilde{a}_n y(k-n) - \tilde{b}_0 u(k) - \tilde{b}_1 u(k-1) - \dots - \tilde{b}_n u(k-n) \quad (6.54)$$

It is apparent from Eqn. (6.53) that consistent estimates of $\underline{\beta}$ in (6.47) result from the fact that the measurements $y(k)$ are purged of the error terms that are correlated with the regressors $y(k-1), y(k-2), \dots, y(k-n)$ in model (6.41).

Before leaving this hypothetical case, consider the effect of using the Instrumental Variable (IV) matrix $[Z \ ; \ E]$ on model (6.42), where the matrix Z satisfies the conditions:

$$\left[\frac{1}{N} \ Z' M \right] \xrightarrow{P} \text{A non-singular matrix} \quad (6.55)$$

$$\left[\frac{1}{N} \ Z' \underline{v} \right] = \left[\frac{1}{N} \ Z' (E\underline{\delta} + \underline{e}) \right] \xrightarrow{P} \underline{0} \quad (6.56)$$

Multiplying both sides of Eqn. (6.42) by $[Z' \ ; \ E']'$, we have

$$Z' \underline{y} = Z' M \underline{\beta} + Z' \underline{v} \quad (6.57)$$

$$E'(\underline{y} - M\underline{\beta}) = E'E\underline{\delta} + E'\underline{e} \quad (6.58)$$

In view of conditions (6.55-56) and the fact that $[1/N \ E'e] \xrightarrow{P} \underline{0}$ asymptotically, the parameters $\underline{\beta}$ and $\underline{\delta}$ may be estimated by:

$$\tilde{\underline{\beta}}_{IV} = (\underline{Z}'\underline{M})^{-1} \underline{Z}'\underline{y} \quad (6.59)$$

$$\tilde{\underline{\delta}}_{IV} = (\underline{E}'\underline{E})^{-1} \underline{E}'(\underline{y} - \underline{M}\hat{\underline{\beta}}_{IV}) \quad (6.60)$$

As opposed to the corresponding set of estimators (6.51-52), in which the convergence of the two estimators is interdependent and cumulative, in the above set of estimators, $\tilde{\underline{\beta}}_{IV}$ is completely independent of $\tilde{\underline{\delta}}_{IV}$. In other words, even if the estimates of the noise parameters $\underline{\delta}$ are greatly in error, they will not adversely affect the quality of $\hat{\underline{\beta}}_{IV}$. The reverse of the above statement, however, is not true.

The fact that the process parameters can be estimated independently of the noise parameters in estimator (6.59-60) also implies that the computation time for this estimator will be less than that for estimator (6.47) (or (6.51-52)). This follows from the fact that the number of multiplications involved in solving a set of p linear algebraic equations is generally proportional to (p^3) . For estimator (6.47), $(3n+1)$ normal equations have to be solved, simultaneously, whereas for estimator (6.59-60), the first $(2n+1)$ equations are independent of the last n equations.

Recursive approximations to the idealized estimators (6.47) and (6.49-60) will now be presented. The algorithms are approximations, because the error terms $e(k)$; $k = 1, 2, \dots, N$, used in the idealized estimators are not known and only their estimates, generated recursively, are available.

A. Adaptive Recursive Least Squares Algorithm - Panuska

This algorithm is a recursive approximation to the idealized estimator (6.47) and is characterized by the recursive relations:

$$\hat{\underline{\theta}}(k+1) = \hat{\underline{\theta}}(k) + \frac{H(k) \hat{\underline{\mu}}(k+1) [y(k+1) - \hat{\underline{\mu}}'(k+1) \hat{\underline{\theta}}(k)]}{1 + \hat{\underline{\mu}}'(k+1) H(k) \hat{\underline{\mu}}(k+1)} \quad (6.61)$$

$$H(k+1) = H(k) - \frac{H(k) \hat{\underline{\mu}}(k+1) \hat{\underline{\mu}}'(k+1) H(k)}{1 + \hat{\underline{\mu}}'(k+1) H(k) \hat{\underline{\mu}}(k+1)} \quad (6.62)$$

$$H(k) = \begin{bmatrix} M'M & M'\hat{\underline{E}} \\ \hat{\underline{E}}'M & \hat{\underline{E}}'\hat{\underline{E}} \end{bmatrix}^{-1} \quad \text{at time } k \quad (6.63)$$

$$\underline{\mu}(k+1) = [u(k+1) \dots u(k-n+1) y(k) \dots y(k-n+1) : \hat{e}(k) \dots \hat{e}(k-n+1)]' \quad (6.64)$$

$$\hat{\underline{\theta}}(k) = \begin{bmatrix} \hat{\underline{\beta}}(k) \\ \hat{\underline{\delta}}(k) \end{bmatrix} = [\hat{b}_0(k) \dots \hat{b}_n(k) - \hat{a}_1(k) - \dots - \hat{a}_n(k) : \hat{c}_1(k) \dots \hat{c}_n(k)]' \quad (6.65)$$

$$\hat{e}(k) = y(k) - \hat{\underline{\mu}}'(k) \hat{\underline{\theta}}(k) \quad (6.66)$$

$$\hat{\lambda}^2(k) = \left[\frac{k-1}{k} \right] \hat{\lambda}^2(k-1) + \frac{1}{k} [\hat{e}(k)]^2 \quad (6.67)$$

The recursive relations (6.61), (6.62) and (6.66) can be implemented in practice, as shown in Fig. 6.4.

It is quite apparent that this algorithm belongs to the class of Bootstrap estimators whose basic characteristics were formalized in Chapter 5. The algorithm consists of two recursive estimators - one for the unknown structural parameters θ and the other for the unknown incidental parameters, which in this case happen to be $e(k)$; $k = 1, 2, \dots, N$. The convergence of the algorithm is dependent on a bootstrapping operation between these two recursive estimators.

Using the theory of Stochastic Approximations, Panuska has proved the mean square convergence of this algorithm. However, his proof depends on the critical assumption that the recursive estimator for $e(k)$ (Eqn. (6.66)) is stable.

B. Recursive Algorithm Based on Instrumental Variables

While discussing the hypothetical case, it was pointed out that some saving in computation time and storage requirements can be achieved, if the $(2n+1)$ process parameters β and the n noise parameters δ are estimated consecutively (in the same recursion). Further, use of a recursive IV estimator for the process parameters β , effectively decouples the estimator for β from the estimator for δ . In many practical situations, both these features would be desirable. Recursive relations for this type of algorithm estimates of noise-free output are used as instruments are given by:

Fig. 6.4 Adaptive Recursive Least Squares Algorithm - Panuska

$$\hat{\underline{\beta}}(k+1) = \hat{\underline{\beta}}(k) + \frac{C(k) \underline{z}(k+1) [y(k+1) - \underline{m}'(k+1) \hat{\underline{\beta}}(k)]}{1 + \underline{m}'(k+1) C(k) \underline{z}(k+1)} \quad (6.68)$$

$$C(k+1) = C(k) - \frac{C(k) \underline{z}(k+1) \underline{m}'(k+1) C(k)}{1 + \underline{m}'(k+1) C(k) \underline{z}(k+1)} \quad (6.69)$$

$$C(k) = (Z'M)^{-1} \text{ at time } k$$

$$\underline{m}(k+1) = [u(k+1) \dots u(k-n+1) \mid y(k) \dots y(k-n+1)]'$$

$$\begin{aligned} \underline{z}(k+1) &= [u(k+1) \dots u(k-n+1) \mid z(k) \dots z(k-n+1)]' \\ &= [u(k+1) \dots u(k-n+1) \mid \hat{x}(k) \dots \hat{x}(k-n+1)]' \end{aligned}$$

$$z(k) = \hat{x}(k) = \underline{z}'(k) \hat{\underline{\beta}}(k) \quad (6.70)$$

$$\hat{v}(k+1) = y(k+1) - \underline{m}'(k+1) \hat{\underline{\beta}}(k+1) \quad (6.71)$$

$$\hat{\underline{\delta}}(k+1) = \hat{\underline{\delta}}(k) + \frac{S(k) \hat{\underline{\epsilon}}(k+1) [\hat{v}(k+1) - \hat{\underline{\epsilon}}'(k+1) \hat{\underline{\delta}}(k)]}{1 + \hat{\underline{\epsilon}}'(k+1) S(k) \hat{\underline{\epsilon}}(k+1)} \quad (6.72)$$

$$S(k+1) = S(k) - \frac{S(k) \hat{\underline{\epsilon}}(k+1) \hat{\underline{\epsilon}}'(k+1) S(k)}{1 + \hat{\underline{\epsilon}}'(k+1) S(k) \hat{\underline{\epsilon}}(k+1)} \quad (6.73)$$

$$S(k) = (\hat{E}'\hat{E})^{-1} \text{ at time } k$$

$$\hat{\underline{\epsilon}}(k+1) = [\hat{\underline{\epsilon}}(k) \hat{\underline{\epsilon}}(k-1) \dots \hat{\underline{\epsilon}}(k-n+1)]'$$

$$\hat{e}(k) = \hat{v}(k) - \hat{e}'(k) \hat{\delta}(k) \quad (6.74)$$

$$\hat{\lambda}^2(k) = \left[\frac{k-1}{k} \right] \hat{\lambda}^2(k-1) + \frac{1}{k} [\hat{e}(k)]^2 \quad (6.75)$$

The block diagram representing this recursive algorithm is shown in Fig. 6.5 and the estimator can be implemented in practice as shown in Fig. 6.6.*

Though this algorithm provides estimates of the process parameters $\underline{\beta}$ that are not affected by the quality of the estimates for the noise parameter $\underline{\delta}$, the reverse, however is not true. Thus if the initial estimates of $v(k)$ are quite poor then so will be the initial estimates of $\underline{\delta}$. In practice, therefore, it may become necessary to give less weight to the initial values of $\hat{v}(k)$. As discussed in Chapter 5, this can be achieved by modifying the error covariance matrix $S(k)$ for estimator $\hat{\delta}$ as:

$$S(k) = S(k) + f(k)xI \quad (4.76)$$

where

$$\lim_{k \rightarrow \infty} f(k) = 0; \quad f(k) \leq f(k-1) \quad \forall k.$$

* In Eqns. (6.68-74) and Figs. 6.6-7, the model is assumed to contain deterministic inputs $u(k)$. If the model is force-free, suitably lagged values of the output $y(k)$ can be used as instruments.

Fig. 6.5. Block Diagram Representation of Recursive Estimators for the System Parameters Based on Instrumental Variables.

Fig. 6.6. Implementation of Recursive Estimator for System Parameters Based on Instrumental Variables

and I is the identity matrix. The types of scalar functions $f(k)$ that may be used for this purpose were discussed in Chapter 5.

Results

The 'adaptive recursive least squares estimator' (Recursive Algorithm - Panuska) represented by Eqns. (6.61), (6.62) and (6.66), Fig. 6.4 and the recursive estimator based on Instrumental Variables (Recursive Algorithm - Pandya) represented by Eqns. (6.68-74), Figs. 6.5-6, were tested on data generated from the following models:

1. Second Order Model - (A)

$$A(q^{-1})y(k) = B(q^{-1})u(k) + C(q^{-1})e(k)$$

$$A(q^{-1}) = (1 - 1.5 q^{-1} + 0.7 q^{-2}) = (q^{-1} - 0.75 + j0.37)(q^{-1} - 0.75 - j0.37)$$

$$B(q^{-1}) = (q^{-1} + 0.5 q^{-2})$$

$$C(q^{-1}) = (1 - q^{-1} + 0.2 q^{-2}) = (q^{-1} - 0.72)(q^{-1} - 0.28)$$

$$e(k) \sim \text{N.I.D. } (0, \lambda^2); \quad \lambda^2 = 1.0$$

$$u(k) \sim \text{PRBS}(1, 1, 127) + \text{PRBS}(\frac{1}{2}, 127, 889)$$

2. Second Order Model - (B)

$$A(q^{-1})y(k) = B(q^{-1})u(k) + C(q^{-1})$$

$$A(q^{-1}) = (1 - 1.8 q^{-1} + 0.82 q^{-2}) = (q^{-1} - 0.9 + j0.1)(q^{-1} - 0.9 - j0.1)$$

$$B(q^{-1}) = (q^{-1} + 0.5 q^{-2})$$

$$C(q^{-1}) = (1 - 1.0 q^{-1} + 0.5 q^{-2}) = (q^{-1} - 0.5 + j0.5)(q^{-1} - 0.5 - j0.5)$$

$$e(k) \sim \text{N.I.D.}(0, \lambda^2); \quad \lambda^2 = 1.0$$

$$u(k) \sim \text{PRBS}(1,1,127) + \text{PRBS}(\frac{1}{2}, 127, 889)$$

The results are presented in Tables 6.2-3 and Figs. 6.7-8. It may be observed that the recursive algorithm proposed by the author and which uses an IV estimator for the process parameters $\underline{\beta}$, requires less time to compute. Under the conditions of the experiment, it provides estimates of process parameters $\hat{\underline{\beta}} = [\hat{\theta}_1 \hat{\theta}_2 \hat{\theta}_3 \hat{\theta}_4]'$ whose RMS error is lower and the RMS error for the noise parameters $\hat{\underline{\delta}} = [\hat{\theta}_5 \hat{\theta}_6]'$ and λ^2 is either the same as or better than Panuska's algorithm.

6.4 Identification of Glucose Tolerance Test Model - Application to Experimental Data

Test results on artificially generated input-output data, that have been presented so far, would indicate that the class of Bootstrap estimators proposed in this thesis provide a computationally efficient method for on-line identification of linear, discrete-time models. However, in a practical experiment some of the assumptions associated with a given identification method may not hold. If the accuracy of the identi-

SECOND ORDER MODEL - (A)

OF DATA PAIRS = 1,000; # OF TRIALS = 20

ESTIMATORS	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$	$\hat{\theta}_5$	$\hat{\theta}_6$	$\hat{\lambda}^2$	TIME 1,000 SAMPLES
True Values	1.000	0.500	1.500	- 0.700	- 1.000	0.200	1.000	
Starting Values	0.00	0.00	0.00	0.00	0.00	0.00	2.00	
$\hat{\theta}$ -Panuska	1.010	0.461	1.508	- 0.705	- 0.972	0.146	1.009	≈ 9 secs
(RMS ERROR)	(0.037)	(0.082)	(0.028)	(0.024)	(0.055)	(0.087)	(0.048)	
$\hat{\theta}$ -Pandya	1.007	0.473	1.501	- 0.699	- 0.963	0.154	1.002	≈ 5 secs
(RMS ERROR)	(0.035)	(0.071)	(0.017)	(0.013)	(0.055)	(0.064)	(0.038)	

TABLE 6.2. Comparison of Estimates Obtained from Recursive Algorithm -
Panuska and Recursive Algorithm - Pandya - Second Order
Model - (A)

SECOND ORDER MODEL - (B)

OF DATA PAIRS = 1,000; # OF TRIALS = 20

ESTIMATORS	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$	$\hat{\theta}_5$	$\hat{\theta}_6$	$\hat{\lambda}^2$	TIME 1,000 SAMPLES
True Values	1.000	0.500	1.800	- 0.820	- 1.000	0.500	1.000	
Starting Values	0.000	0.000	1.000	0.000	0.000	0.000	2.000	
$\hat{\theta}$ -Panuska	1.017	0.459	1.807	- 0.827	- 0.988	0.447	0.999	$\approx$ 9 secs
(RMS ERROR)	(0.036)	(0.071)	(0.013)	(0.013)	(0.040)	(0.085)	(0.047)	
$\hat{\theta}$ -Pandya	1.000	0.473	1.803	- 0.823	- 0.984	0.473	1.004	$\approx$ 5 secs
(RMS ERROR)	(0.034)	(0.066)	(0.009)	(0.009)	(0.033)	(0.047)	(0.038)	

TABLE 6.3. Comparison of Estimates Obtained from Recursive Algorithm - Panuska and Recursive Algorithm - Pandya - Second Order Model - (B)

Fig. 6.7. Convergence of Recursive Algorithm-Panuska and Recursive Algorithm-Pandya
Second Order Model (A)

Fig. 6.8. Convergence of Recursive Algorithm - Panuska and Recursive Algorithm - Pandya.
Second Order Model (B)

fication technique is unduly sensitive to these assumptions, the technique may lead to erroneous results.

In this section, the performance of the proposed Bootstrap estimators is tested on a set of experimental data. The input-output data used for this purpose were obtained from a Glucose Tolerance Test on a dog. The experiment was carried out at the Faculty of Medicine, University of Ottawa and it is a part of a cooperative project between the Division of Systems Engineering at Carleton, the Faculty of Medicine at University of Ottawa and the Ottawa Civic Hospital. The project is currently in progress and is aimed towards the design of an automatic regulator for blood sugar concentration in diabetics.

The Glucose Tolerance Test Model

In essence, the Glucose Tolerance Test on human (or animal) glucose metabolism consists of analyzing the variation in blood sugar level due to an intravenous glucose load. A suitable model for prediction and control may be fitted to the input output data⁷. The system is characterized by the presence of a large transport delay (several minutes) and a constant fasting level (base level). The following second order, linear, discrete-time model was fitted to the experimental input-output data:

$$A(q^{-1}) y(k) = q^{-\tau} B(q^{-1}) u(k) + \alpha + \lambda C(q^{-1}) e(k)$$

$$A(q^{-1}) = 1 + a_1 q^{-1} + a_2 q^{-2}$$

$$B(q^{-1}) = b_1 q^{-1} + b_2 q^{-2}$$

$$C(q^{-1}) = 1 + c_1 q^{-1} + c_2 q^{-2}$$

where Sampling Interval = 10 seconds
 No. of I/O samples = 400
 Transport Delay = 60 samples

The transport delay was extrapolated from a plot of the input/output data. The constant term α included in the model may be considered as a constant fictitious input, such that, in the absence of the external input $u(k)$, it would provide a constant output level = the fasting level. Thus given the value of α , the fasting level (F.L.) may be calculated as:

$$\text{F.L.} = \frac{\alpha}{1 + a_1 + a_2}$$

Results

Parameter estimates obtained by applying the various on-line identification techniques to this set of experimental data are given in Table 6.4 which also includes the Maximum Likelihood estimates, for comparison. The notation Pandya(1) and Pandya(2) implies that in the former case the process and noise parameters were estimated in separate passes whereas in the latter case the parameters of the entire model were estimated in a single pass by using two consecutive Bootstrap estimators as discussed in Section 6.3.

To obtain some idea about the nature of fit provided by using the Bootstrap estimators on this set of experimental data, estimated noise-free output:

GLUCOSE TOLERANCE TEST

NO. OF I/O DATA = 400

METHOD	$\hat{a}_1$	$\hat{a}_2$	$\hat{b}_1$	$\hat{b}_2$	$\hat{\alpha}$	$\hat{F.L.}$	$\hat{c}_1$	$\hat{c}_2$	$\hat{\lambda}^2$
OLS	-1.004	0.044	3.707	-0.467	2.077	52.20	-	-	-
Wong & Polak	-1.346	0.359	3.180	-1.499	0.5392	43.96	-	-	-
Rowe	-1.047	0.072	3.649	-0.912	1.179	47.16	-	-	-
Panuska	-1.177	0.195	3.556	-1.136	0.758	42.11	-0.216	-0.405	2.435
Pandya(1)	-1.076	0.098	3.604	-1.011	1.025	46.3	-0.135	-0.405	2.299
Pandya(2)	-1.076	0.0983	3.604	-1.011	1.025	46.3	-0.192	-0.376	2.32
MLE	-1.076	0.0985	3.791	-0.018	0.999	45.2	-0.153	-0.393	2.296

TABLE 6.4. Parameter Estimates for Glucose Tolerance Test Data

$$\hat{x}(k) = q^{-\tau} \left[\frac{B(q^{-1})}{A(q^{-1})} \right] u(k) + \left[\frac{\hat{\alpha}}{A(q^{-1})} \right]$$

is plotted in Fig. 6.9. A similar plot for the Maximum Likelihood estimates is given in Fig. 6.10.

Thus, the performance of the Bootstrap estimators when used on the Glucose Tolerance Test data has been very encouraging and it is being incorporated as an integral part of an overall stochastic control algorithm for regulation of blood sugar concentration in future tolerance tests.

6.5 Concluding Remarks

In this chapter, the problem of estimating the noise parameters $\underline{\delta}$, given consistent estimates of the process parameters $\underline{\beta}$, and the problem of estimating the process as well as noise parameters simultaneously and in a single experiment were considered.

The 2SLS technique was applied, for the first time, to estimate the noise parameters in model (6.1). It was shown that, based on the 2SLS concept, two estimators, viz., 2SLS-(1) and 2SLS-(2) can be formulated - of which 2SLS-(1) exhibits excellent convergence properties (Fig. 6.6). However, if the available time to compute and/or storage space are at a premium, these two estimators may become unacceptable. In this case, the recursive approximation to 2SLS-(2) (Bootstrap estimator $\hat{\underline{\delta}}^3$) may be used. Though its convergence is slower, it provides considerable saving

Fig. 6.9. Estimated Noise-free Output - Bootstrap Estimator

225

Fig. 6.10. Estimated Noise-free Output - MLE

Fig. 6.11. Estimated Noise-free Output - LSE

in both the time to compute and storage requirements, because the intermediate step of estimating $\underline{Y}$, the parameters of the AR process (6.21), is eliminated.

For estimating the process and the noise parameters simultaneously in the same experiment, the 'adaptive recursive least squares estimator' of Panuska is compared with the recursive algorithm proposed by the author. The latter algorithm has the advantage that, unlike the former algorithm, the estimator for the process parameters operates independently of the estimator for the noise parameters. This property is, in general, desirable because it not only means that estimates of the process parameters will not be adversely affected if the estimates of the noise parameters tend to be poor, but it also leads to some improvement in computational efficiency.

The performance of the Bootstrap estimators was tested on experimental data obtained from a Glucose Tolerance Test. The results confirm that these estimators provide a computationally efficient method for on-line identification of linear, discrete-time models.

CHAPTER 7

SOME RELATED PROBLEMS AND FINAL CONCLUSIONS

7.1 Some Related Problems

Levy's canonical model or the equivalent difference equation model, together with its various special cases outlined in Chapter 2, seem to fit a wide range of practical systems. However, under certain circumstances it may become necessary to modify either the method of identification or the system model or both. Some of the problems that may require special attention, are briefly discussed below:

1. Identification of a Process under Closed-loop Control

Throughout the preceding text it was tacitly assumed that the input sequence $u(k)$ was functionally independent of the system output $y(k)$, i.e., the system to be identified had an open-loop structure. However, in those cases where identification is an integral part of an adaptive control strategy, this condition will not be satisfied. The identification problem becomes particularly acute, if the feedback signal is directly proportional to the system output (e.g., in a linear regulator). Thus, if

$$u(k) = \alpha y(k); \quad k = 1, 2, \dots, \dots \quad (7.1)$$

where α is a constant, it is apparent that the input terms $u(k-1), u(k-2), \dots, u(k-n)$ in the difference equation model:

$$y(k) = b_0 u(k) + b_1 u(k-1) + \dots + b_n u(k-n) - a_1 y(k-1) - \dots - a_n y(k-n) + v(k) \quad (7.2)$$

are linearly related to the corresponding output terms $y(k-1), y(k-2), \dots, y(k-n)$ respectively. This would imply that in the equivalent matrix representation

$$\underline{y} = M\underline{B} + \underline{v} \quad (7.3)$$

the design matrix M ceases to be of full rank and the model is rendered unidentifiable.

This condition, wherein some of the explanatory variables or regressors are heavily correlated with each other is referred as 'multicollinearity' in econometrics and the case represented by Eqn. (7.1) implies 'perfect multicollinearity'. An introductory discussion on the detection of and correction for multicollinearity in econometric models may be found in such standard texts as Goldberger¹⁷, Johnson²⁰, etc.

Fisher⁴⁷ has shown that a process under closed-loop control is identifiable provided the feedback is nonlinear. Bohlin⁴⁸ has briefly discussed the implications of closed-loop control on model identifiability and has shown that such a model is identifiable if there is a known transport delay in the forward path or if a small perturbation signal is added to the input. Pagurek, Riordon and Mahmoud⁷ have successfully implemented an adaptive control strategy for the control of blood-sugar level in a 'Glucose Tolerance Test' on human subjects, where such a transport delay is inherent.

The most general approach to the problem is provided by the Dual Control theory of Fel'dbaum⁴⁹. Though the theory in its original form entails exorbitant computational problems, some ad hoc simplifications may be possible in specific cases.

2. Identification of Time-varying Parameters

If the process to be identified is non-stationary, the problem of identification and control cannot be separated from each other, i.e., the 'separation hypothesis' will no longer hold. In this case, the parameters have to be estimated in real time and in the presence of closed-loop conditions. For the difference equation model considered in this thesis, a non-stationary process also implies that the effective noise $v(k)$ in Eqn. (7.2) will no more be covariance-stationary, i.e., the noise parameters $\underline{\delta} = [c_1 c_2 \dots c_n]'$ will also be time-varying.

The difference equation model with time-varying parameters may be written as:

$$\underline{\theta}(k+1) = \psi \underline{\theta}(k) \quad (7.4)$$

$$y(k) = [u(k) \dots u(k-1)y(k-1) \dots y(k-n)e(k-1) \dots e(k-n)] \underline{\theta}(k) + e(k)$$

or

$$y(k) = \underline{\mu}'(k) \underline{\theta}(k) + e(k) \quad (7.5)$$

$$e(k) \sim \text{N.I.D.}(0, \lambda^2) \quad (7.6)$$

where ψ is the state transition matrix representing the dynamics of the parameter vector $\underline{\theta}(k)$ and it may or may not be constant. If $\psi = I$, then the model reverts back to one with constant parameters.

If ψ and the error terms $e(k-1), e(k-2), \dots, e(k-n)$ are known, then Eqns. (7.4-6) represents a simple message model for a Kalman Filter in which $\underline{\theta}(k)$ is the state vector to be estimated and $\underline{u}(k)$ is the output matrix. The time-varying parameters may therefore be tracked by using the resulting Kalman filter.

Since the error terms $e(k-1), e(k-2), \dots, e(k-n)$ are not known a priori, a sub-optimal filter for $\underline{\theta}(k)$ in which the error terms are estimated on-line may still be designed by using the techniques discussed in Chapter 6. However, if ψ is unknown, even a sub-optimal filter cannot be designed for tracking $\underline{\theta}(k)$, specially if their time variation is either random or too fast.

If the variation of $\underline{\theta}(k)$ with time is relatively slow (slowly time-varying parameters), one may assume that $\psi = I$ and use one of the on-line methods discussed in the preceding chapter to track the parameters. Such a filter would have an inherent tendency to diverge after it has been running for a while, unless effective steps are taken to control the divergence. A number of techniques for controlling error divergence in Kalman filters have been discussed by Jazwinsky⁴⁰.

The general problem of tracking, when the parameters are subject to random fluctuations has not been effectively solved. If the transition matrix ψ (Eqn. (7.4)) could be estimated even approximately, this approximate estimate of ψ combined with the available techniques for controlling error divergence could lead to a suitable tracking filter for a general non-stationary system.

3. Identification of Multivariable Systems

The basic model considered in this thesis was characterized by the presence of a scalar output sequence $y(k)$. Identification techniques discussed in this thesis can easily be extended to include multivariable models. For example, 2SLS technique is frequently used for identification of simultaneous equation models in econometrics and Rowe³⁵ and Mayne³⁴ have successfully used recursive IV estimators for parameter estimation in multivariable models.

The basic problem with multivariable systems is to find a suitable representation for the system. As mentioned in Chapter 2, unlike the scalar output case, a multivariable system does not admit a unique canonical representation. Though the problem of assigning an optimum structure (for identification purposes) to multivariable systems has not yet been resolved, the canonical form proposed by Rowe⁹ has the advantage that it ensures the use of latest measurements for estimation of model parameters.

4. a priori Estimation of Process Order

In view of the fact that the accuracy of identification in parametric models depends heavily on a knowledge of the correct process

order, the importance of an adequate solution to this problem is self evident. Generally the adequacy of the order of the fitted model can be inferred by performing goodness-of-fit tests on the residuals. However, in the absence of any knowledge whatsoever, about the process order, it may become necessary to fit a large number of models each of a different order and then select the one that provides the best fit (as judged from a suitable test statistic). Obviously, such a procedure can become computationally very wasteful. Thus, if a suitable approximation to the true process order can be obtained on an a priori basis, it would lead to a considerable saving in computational effort.

For noise-free processes, Lee⁸ has proposed a test based on the non-singularity of the product moment matrix for $n' > n$, where n is the true process order. However, the test is not applicable in the presence of noise in the system. Woodside⁵⁰ has proposed a set of tests for a priori estimation of process order in the presence of moderate amounts of noise. The tests require large amounts of data and also some knowledge of the noise dynamics. An alternative approach that needs exploring is the design and use of special test input signals for this specific purpose.

7.2 Final Conclusions

A comprehensive study on recursive algorithms for parameter estimation in a class of linear dynamic systems has been presented in this thesis. The author's comments based on analytical and computational results presented in this thesis are summarized below:

1. In any identification experiment, accurate estimation of the process parameters is of primary importance.
2. If the model contains exogenous variables (deterministic inputs), the recursive approximation to 2SLS, viz., $\hat{\beta}_{IV}$ -Pandya is a conceptually simple and computationally efficient estimator for the process parameters and it provides estimates that have low dispersion around their true values. (Chapter 4).
3. If the model does not contain any exogenous variables (force-free model), the recursive IV estimator that utilizes suitably lagged values of the output $y(k)$ as instruments and which is equivalent to the modified Yule-Walker estimator may be used to advantage for the estimation of process parameters (Chapter 4).
4. Though the convergence of the Bootstrap algorithms ($\hat{\beta}_{IV}$ - Pandya being one of them) has not been proved under completely arbitrary initial conditions, in practice, the recursive approximation to the 2SLS proposed by the author exhibits excellent convergence behavior. If necessary, the stability of the algorithm may be further ensured by the use of the simple adaptive filter, for the noise-free process output, proposed by the author (Chapter 5).
5. Having estimated the process parameters, if the noise parameters are also to be estimated, there is again a choice between a straight 2SLS procedure and its recursive approximation (Bootstrap estimator). Though the computational efficiency of the former is poor, its

convergence rate is fast. For the latter the situation is just the reverse. (Chapter 6).

6. If the parameters of the entire model, i.e., both the process as well as noise parameters are to be estimated simultaneously and in a single experiment, the recursive algorithm proposed by the author ensures that the estimates of the process parameters are not adversely affected by the quality of the estimates of the noise parameters. This algorithm also offers some saving in computing time and storage requirements. (Chapter 7).

REFERENCES

1. Eykhoff, P., *Process Parameter and State Estimation*, Automatica, Vol. 4, pp. 205-233, 1968.
2. Cuenod, M. and A.P. Sage, *Comparison of Some Methods used for Process Identification*, Survey paper - IFAC Symposium on Problems of Identification in Automatic Control, Prague, June 1967.
3. Astrom, K.J. and P. Eykhoff, *System Identification - a Survey*, Automatica, Vol. 7, No. 2, pp. 123-162.
4. Nieman, R.E., D.G. Fisher and D.E. Seborg, *A Review of Process Identification and Parameter Estimation Techniques*, International Journal of Control, Vol. 13, No. 2, February 1971.
5. Pandya, R.N., *A Class of Bootstrap Estimators for Linear System Identification*, International Journal of Control, Vol. 15, No. 6, June 1972.
6. Pandya, R.N., *On the Equivalence of Yule-Walker Estimator and an Instrumental Variable Estimator*, Division of Systems Engineering, Carleton University, Ottawa, August 1971.
7. Pagurek, B., J.S. Riordon and S. Mahmoud, *Adaptive Control of the Human Glucose Regulatory System*, Medical and Biological Engineering (to appear November 1972).
8. Lee, R.C.K., *Optimal Identification, Estimation and Control*, M.I.T. Press, Cambridge, Mass., 1964.
9. Rowe, I.H., *A Statistical Model for the Identification of Multivariable Systems*, Proc. IFAC Symposium on Multivariable Control Systems, Dusseldorf, 1968.
10. Kailath, T., *An Innovations Approach to Least Squares Estimation Part I: Linear Filtering in Additive White Noise*, IEEE Transactions on Automatic Control, Vol. AC 13, No.6, December 1968.
11. Mehra, R.K., *Identification of Stochastic Linear Dynamic Systems*, Paper 6-f-1, 1968 IEEE Symposium on Adaptive Processes, University Park, Pa., U.S.A.
12. Astrom, K.J. and T. Bohlin, *Numerical Identification of Linear Dynamical Systems from Normal Operating Records*, in Theory of Self Adaptive Systems, edited by P.H. Hammond, Plenum Press, New York, 1966.

13. Box, G.P. and G.M. Jenkins, *Time Series Analysis - Forecasting and Control*, Holden Day Publishers, San Francisco, 1970.
14. Mehra, R.K., *On-line Identification of Linear Dynamic Systems with Applications to Kalman Filtering*, IEEE Transactions on Automatic Control, Vol. AC-16, No. 2, February 1971.
15. Gersch, W., *Estimation of the Autoregressive Parameters of a Mixed Autoregressive Moving Average Time Series*, IEEE Transactions on Automatic Control, Vol. AC-15, No. 5, October 1970.
16. Mann, H.B. and A. Wald, *On the Statistical Treatment of Linear Stochastic Difference Equations*, Econometrica, Vol. 11, pp. 173-220, 1943.
17. Goldberger, A.S., *Econometric Theory*, John Wiley and Sons Inc., New York, 1964.
18. Wong, K.Y., *Estimation of Parameters of Linear Systems using Instrumental Variable Techniques*, Ph.D. Thesis (1966), University of California, Berkeley.
19. Grenander, U. and M. Rosenblatt, *Statistical Analysis of Stationary Time Series*, John Wiley and Sons, New York, 1957.
20. Johnson, J., *Econometric Methods*, McGraw Hill Book Co., Inc., New York, 1963.
21. Maddala, G.S., *Generalized Least Squares with an Estimated Var-Covar Matrix*, Econometrica, Vol. 39, No. 1, January 1971.
22. Steiglitz, K. and L.E. McBride, *Techniques for the Identification of Linear Systems*, Vol. AC-10, No. 4, pp. 461-44, October 1965.
23. Clark, D.H., *Generalized Least Squares Estimation of a Dynamic Model*, Proc. IFAC Symposium on the Problems of Identification in Automatic Control, Prague, June 1967.
24. Hastings-James, R. and M.W. Sage, *Recursive Generalized Least Squares Procedure for On-line Identification of Process Parameters*, Proc. I.E.E., Vol. 116, No. 12, December 1969.
25. Amemiya, T. and W. Fullter, *A Comparative Study of Alternative Estimators in a Distributed Lag Model*, Econometrica, Vol. 35, July 1967.

26. Soderstrom, T., *Notes on Pseudo-inverses - Applications to Identification*, Report 7003, July 1970, Lund Institute of Technology, Division of Automatic Control.
27. Boullion, T.L. and P.L. Odell, *Generalized Inverse Matrices*, Wiley Interscience, New York, 1971.
28. Zellner, A. and H. Theil, *Three Stage Least Squares: Simultaneous Estimation of Simultaneous Equations*, *Econometrica*, Vol. 30, 1962, pp. 54-78.
29. Riersol, O., *Confluence Analysis by means of Instrumental Sets of Variables*, *Archives for Mathematics, Astronomy and Physics*, Vol. 32, 1945.
30. Saragan, J.D., *The Estimation of Economic Relationships Using Instrumental Variables*, *Econometrica*, Vol. 26, pp. 393-415, 1958.
31. Joseph, P. and J. Lewis and J. Tou, *Plant Identification in the Presence of Disturbances and Application to Digital Adaptive Systems*, *A.I.E.E. Transactions on Applications and Industry*, March 1961.
32. Iacoss, R., *Identification of Systems Using the Instrumental Variable Method*, Ph.D. Thesis (1965), University of California, Berkeley.
33. Wong, K.Y. and E. Polak, *Identification of Linear Discrete Time Systems Using the Instrumental Variable Method*, *IEEE Transactions on Automatic Control*, Vol. AC-12, No. 6, December 1967.
34. Mayne, D.Q., *A Method for Estimating Discrete Time Transfer Functions*, *I.E.E. Conference of Computers in Control*, University of Bristol, Paper No. c-2, April 1967.
35. Rowe, I.H., *A Bootstrap Method for the Statistical Estimation of Model Parameters*, *International Journal of Control*, Vol. 12, No. 4, April 1970.
36. Young, P.C., *An Instrumental Variable Method for Real Time Identification of Noisy Processes*, *Reprints, Fourth IFAC Congress, Warsaw, June 1967*, Paper No. 26.6.
37. Wonnacott, A.J. and T.H. Wonnacott, *Econometrics*, John Wiley and Sons, New York, 1970.
38. Dhrymes, P.J., *Econometrics - Statistical Foundations and Applications*, Harper & Row, New York, 1970.

39. Deyst, J.J. and C.F. Price, *Conditions for Asymptotic Stability of the Discrete Minimum Variance Linear Estimator*, IEEE Transactions on Automatic Control, Vol. AC-13, No.6, pp. 702-5, December 1968.
40. Jazwinsky, A.H., *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970.
41. Schmidt, S.F., *Applications of State Space Methods to Navigational Problems*, in *Advances in Control Systems*, Vol. 3, Academic Press, New York, 1966.
42. Jazwinsky, A.H., *Adaptive Filtering*, Automatica, Vol. 5, pp. 475-85, 1967.
43. Theil, H., *Economic Forecasts and Policy*, Second Edition, North Holland Publishing Company, 1961.
44. Young, P.C., *An Extension to the Instrumental Variable Method for the Identification of a Noisy Dynamic Process*, Department of Engineering, University of Cambridge, Technical Note CN/70/1, May 1970.
45. Panuska, V., *An Adaptive Recursive Least Squares Identification Algorithm*, in *Proceedings of IEEE Symposium on Adaptive Processes (9th) - Decision and Control*, pp. IV 6-1 to IV 6-8.
46. Young, P.C., *Comments on On-line Identification of Linear Dynamic Systems with Applications to Kalman Filtering*, Letters to the editor, IEEE Transactions on Automatic Control, Vol. AC-17, No. 2, April 1972.
47. Fisher, E.E., *Identification of Linear Systems*, Preprints, JAAC (1965), pp. 473-75.
48. Bohlin, T., *On the Problem of Ambiguities in Maximum Likelihood Identification*, 2nd Prague IFAC Symposium on Identification and Process Parameter Estimation, paper no. 3.2.
49. Fel'dbaum, A.A., *Optimal Control Systems*, Academic Press, New York, 1965.
50. Woodside, C.M., *Estimation of the Order of Linear Systems*, Automatica, Vol. 7, pp. 727-733.

A P P E N D I X A

A number of definitions and theorems with relevant comments have been collected in this Appendix so that it may serve as a ready-reference for the various statistical concepts encountered in the main text. No effort has been made to present any proofs. Necessary references, however, have been included.

A.1 Preliminaries

Let $(x_t: t = 1, 2, \dots, N)$ be a set of N random variables. In the present context, it would be appropriate to consider the elements of the set as a sample sequence from a weakly stationary process $x(t; \omega)$ whose indexing parameter t (in this case time) is discrete and is defined on a set of positive integers $t = 1, 2, \dots, N$.

Let $F(x_1, x_2, \dots, x_N; \theta_1, \theta_2, \dots, \theta_p) = F(x_1, x_2, \dots, x_N; \underline{\theta})$ be the joint probability density function (p.d.f.) for the above set of random variables and let their marginal densities be given by $f_1(x_1; \underline{\theta}), f_2(x_2; \underline{\theta}), \dots, f_N(x_N; \underline{\theta})$ where $\underline{\theta}$ is a constant p vector of parameters which is common to all the marginal densities and is therefore carried over to the joint p.d.f.

Definition 1. Random Sample

The set of random variables $(x_1, x_2, \dots, x_N)$ is said to constitute a random sample if its elements are independently and identically distributed (i.i.d.), i.e., if:

$$F(x_1, x_2, \dots, x_N; \underline{\theta}) = f_1(x_1; \underline{\theta}) \cdot f_2(x_2; \underline{\theta}) \cdot \dots \cdot f_N(x_N; \underline{\theta}) \quad (\text{A.1.1})$$

and

$$f_1(x_1; \underline{\theta}) = f_2(x_2; \underline{\theta}) = \dots = f_N(x_N; \underline{\theta}) = f(\cdot; \underline{\theta}) \quad (\text{A.1.2})$$

Definition 2. Regular Density

Let $(x_1, x_2, \dots, x_N)$ be a random sample. It is said to be from a population characterized by a regular density function $f(\cdot; \underline{\theta})$, if the following regularity conditions are satisfied:

- i. The range of the random variables is independent of the parameter $\underline{\theta}$;
- and ii. The p.d.f. $f(\cdot; \underline{\theta})$ possesses derivatives of at least third order with respect to $\underline{\theta}$, and these derivatives are bounded by integrable functions of x_t .

(Dhrymes, P.J., *Econometrics*, Harper & Rowe, 1970, pp. 115)

Definition 3. Estimator and Estimate

For the set of random variables $(x_1, x_2, \dots, x_N)$ (not necessarily a random sample), with the joint p.d.f. $F(x_1, x_2, \dots, x_N; \underline{\theta})$, the vector function

$$\underline{g}(x_1, x_2, \dots, x_N) = \hat{\underline{\theta}}_N \quad (\text{A.1.3})$$

constitutes an estimator for the parameter vector $\underline{\theta}$ and the specific value attained by $\hat{\underline{\theta}}_N$ for a particular realization of the set of random variables $(x_1, x_2, \dots, x_N)$ is called an estimate of $\underline{\theta}$.

Remark 1. The index N in $\hat{\underline{\theta}}_N$ implies that the vector function $\underline{g}(\cdot)$ operates on the entire sample (of size N) so that $\hat{\underline{\theta}}_N$, in general, is a sequence

of vector random variables indexed by the set of positive numbers that are defined by the sample size N . Thus the statistical properties of the estimator not only depend on the form of the function $\underline{g}(\cdot)$, but also on the sample size N .

Remark 2. Some times an estimator is also referred as a 'joint statistic for the elements of $\underline{\theta}$ '.

Definition 4. Linear Estimator

If $\underline{g}(\cdot)$ is a linear function (mapping) of the set of random variables $(x_1, x_2, \dots, x_N)$, then it is said to be a linear estimator of $\underline{\theta}$.

A.2 Properties of Estimators for a Given Fixed N

The following properties of estimators refer to the case where the sample size N is fixed. These are some times called the small sample properties. For this case we shall drop the index N from $\hat{\underline{\theta}}_N$ and it is to be understood that the sample is of size N and the number of parameters to be estimated is $p < N$.

Definition 5. Unbiased Estimator

An estimator $\hat{\underline{\theta}}$ is said to be an unbiased estimator of $\underline{\theta}$ if $E(\hat{\underline{\theta}}) = \underline{\theta}$, where $E(\cdot)$ is the expectation operator.

Definition 6. Cramer - Rao Lower Bound

Let $(x_1, x_2, \dots, x_N)$ be a random sample from a population characterized by the regular p.d.f. $f(\cdot; \underline{\theta})$. If L is the log-likelihood function of the sample defined by

$$\begin{aligned}
L &\triangleq \ln [\text{joint p.d.f. of the sample}] \\
&= \ln \left[\prod_{i=1}^N f(., \underline{\theta}) \right] \\
&= \sum_{i=1}^N \ln [f(., \underline{\theta})] \quad (A.2.1)
\end{aligned}$$

and if $\hat{\underline{\theta}}$ is an unbiased estimator of $\underline{\theta}$, then

$$\text{Cov}(\hat{\underline{\theta}}) \leq - \left[E \left(\frac{\partial^2 L}{\partial \theta_i \partial \theta_j} \right) \right]^{-1} \quad i, j = 1, 2, \dots, p \quad (A.2.2)$$

where a positive semi definite matrix A is said to be at least as large as another positive semi definite matrix B, if (A-B) is positive semi definite.

(Dhrymes, P.J., *Econometrics*, Harper & Row, 1970, pp. 125-27)

Definition 7. Minimum Variance Bound (MVB) Estimator

An estimator $\hat{\underline{\theta}}$ is said to be a minimum variance bound (MVB) estimator of $\underline{\theta}$, if $\hat{\underline{\theta}}$ is unbiased and its covariance matrix attains the Cramer-Rao lower bound, i.e., if

$$E(\hat{\underline{\theta}}) = \underline{\theta} \quad \text{and} \quad \text{Cov}(\hat{\underline{\theta}}) = - \left[E \left(\frac{\partial^2 L}{\partial \theta_i \partial \theta_j} \right) \right]^{-1} \quad (A.2.3)$$

Remark 3. The MVB estimator is sometimes called the most efficient estimator in the sense of Cramer-Rao. However, such an estimator need not always exist, in which case it is expedient to speak of efficiency within a specific class of estimators.

Definition 8. Relative Efficiency

Let S be the class of unbiased estimators of $\underline{\theta}$. An estimator $\hat{\underline{\theta}} \in S$ is said to be efficient relative to another estimator $\tilde{\underline{\theta}} \in S$, if $\text{Cov}(\tilde{\underline{\theta}}) \geq \text{Cov}(\hat{\underline{\theta}})$.

Definition 9. Minimum Variance Estimator

Let S be the class of unbiased estimators of $\underline{\theta}$. If $\hat{\underline{\theta}} \in S$ and is such that for any other estimator $\tilde{\underline{\theta}} \in S$, $\text{Cov}(\tilde{\underline{\theta}}) \geq \text{Cov}(\hat{\underline{\theta}})$, then $\hat{\underline{\theta}}$ is said to be a minimum variance (MV) estimator.

Definition 10. Best Linear Unbiased Estimator

Let S be the class of all linear, unbiased estimators of $\underline{\theta}$. If $\hat{\underline{\theta}} \in S$ and is such that for any other estimator $\tilde{\underline{\theta}} \in S$, $\text{Cov}(\tilde{\underline{\theta}}) \geq \text{Cov}(\hat{\underline{\theta}})$, then $\hat{\underline{\theta}}$ is said to be the Best Linear Unbiased Estimator (BLUE) of $\underline{\theta}$.

Definition 11. Sufficient Estimator

Let $(x_1, x_2, \dots, x_N)$ be a random sample from a population characterized by the p.d.f. $f(\cdot; \underline{\theta})$, and let $\hat{\underline{\theta}}$ be an estimator for the parameter vector $\underline{\theta}$. Further, let $\phi^1, \phi^2, \dots, \phi^{N-1}$ be any other set of statistics for $\underline{\theta}$ (i.e., set of estimators for $\underline{\theta}$).

Estimator $\hat{\underline{\theta}}$ is said to be sufficient relative to $\phi^1, \phi^2, \dots, \phi^{N-1}$, if the joint p.d.f. of $\phi^1, \phi^2, \dots, \phi^{N-1}$ conditioned on $\hat{\underline{\theta}}$ is independent of $\underline{\theta}$.

Remark 4. A sufficient estimator therefore implies that it has exhausted all the available information (from the observed sample) about the parameters. A sufficient estimator is frequently referred as a 'jointly sufficient statistic for the elements of $\underline{\theta}$ '.

A.3 Properties of Estimators for $N \rightarrow \infty$

We may recall (Remark 1) that if the sample size N increases monotonically, the resulting estimator

$$\hat{\theta}_N = g(x_1, x_2, \dots, x_N)$$

constitutes a sequence of vector random variables, indexed by the set of positive integers that are defined by the sample size. Before considering the asymptotic or large sample properties of estimators it would be appropriate to explore, briefly, the convergence behavior of such sequences of random variables.

Let $(\underline{z}_1, \underline{z}_2, \dots, \dots)$ be a sequence of vector random variables, in which each element $(\underline{z}_i)$ is a p vector. Further, let $D_1, D_2, \dots$ be the respective distribution functions of $\underline{z}_1, \underline{z}_2, \dots, \dots$.

Definition 12. Convergence in Distribution

The sequence of vector random variables $\underline{z}_1, \underline{z}_2, \dots$ is said to converge in distribution (or in law) to the random vector $\underline{z}$, if the sequence of distribution functions $D_1, D_2, \dots, \dots$ converges to the distribution function D (of $\underline{z}$) at the continuity points of D .

Definition 13. Convergence in Probability

The sequence of vector random variables $\underline{z}_1, \underline{z}_2, \dots, \dots$ converges in probability (or stochastically) to a random vector $\underline{z}$, if given any two scalars $\epsilon, \delta > 0$, there exists an index $k(\epsilon, \delta)$ such that for all $N > k$

$$P_r [||\underline{z}_N - \underline{z}|| < \epsilon] > 1 - \delta \quad (\text{A.3.1})$$

where the sup-norm of a p vector $\underline{n}$ is a scalar defined by

$$||\underline{n}|| \triangleq \max_{i=1,2,\dots,p} [|\eta_i|] \quad (\text{A.3.2})$$

The vector $\underline{z}$ is called the probability limit of $\underline{z}_N$ and is usually written as

$$\underset{N \rightarrow \infty}{p.\lim} [\underline{z}_N] = \underline{z} \quad (\text{A.3.4})$$

Definition 14. Convergence in Quadratic Mean

If $\underline{z}_1, \underline{z}_2, \dots$ is a sequence of vector random variables for which the first and second absolute moments exist, then it is said to converge in quadratic mean to the random vector $\underline{z}$, if

$$\lim_{N \rightarrow \infty} E[(\underline{z}_N - \underline{z})'(\underline{z}_N - \underline{z})] = 0 \quad (\text{A.3.5})$$

and we say that

$$\lim_{N \rightarrow \infty} [\underline{z}_N] = \underline{z} \quad (\text{A.3.6})$$

Definition 15. Convergence with probability one

The sequence of vector random variables $(\underline{z}_1, \underline{z}_2, \dots)$ is said to converge with probability one (or almost surely) to the random vector $\underline{z}$, if given any two scalars $\epsilon, \delta > 0$, there exists an index $k(\epsilon, \delta)$, such that

$$P_r [||\underline{z}_{K+1} - \underline{z}|| < \epsilon, ||\underline{z}_{K+2} - \underline{z}|| < \epsilon, \dots] > 1 - \delta \quad (\text{A.3.7})$$

Remark 5. Since a constant is a degenerate form of a random variable, the last four definitions include the case where the sequence converges to a constant vector.

Remark 6. The various forms of convergence are related as follows:

Definition 16. Consistent Estimator

An estimator $\hat{\theta}_N$ is said to be a consistent estimator of θ if

$$\text{p.lim}_{N \rightarrow \infty} [\hat{\theta}_N] = \theta \quad (\text{A.3.8})$$

Definition 17. Asymptotic Bias and Covariance

If $\hat{\theta}_N$ is an estimator of θ , then the asymptotic bias and asymptotic covariance matrix of $\hat{\theta}_N$ are respectively defined by

$$\lim_{N \rightarrow \infty} [E(\hat{\theta}_N)] - \theta \quad (\text{A.3.9})$$

$$\lim_{N \rightarrow \infty} [\text{Cov}(\hat{\theta}_N)] \quad (\text{A.3.10})$$

provided $E(\hat{\theta}_N)$ and $\text{Cov}(\hat{\theta}_N)$ exist for all sample sizes N .

Remark 7. For the models considered in this thesis, it is possible to formulate more than one consistent estimator for the parameters. We prefer to choose a consistent estimator because by doing so, we at least ensure that the estimates will converge (stochastically) to the true parameters, as the sample size becomes infinite. If the sample size was truly infinite, any consistent estimator (from the available set) may be used and we are spared from making a choice.

In practice, however, we are constrained to work with a finite (though perhaps a large) number of observations, and the fact that the estimator is consistent, throws little light on such important properties as bias and efficiency of the estimates obtained from this large (though finite) sample. Thus, if we are to be able to choose an optimum estimator (from the class of consistent estimators), it becomes necessary to derive expressions for the asymptotic (or large sample) bias and covariance matrix of the estimators. An alternate way to express this, is to say that we need an estimator that not only converges to the true parameter values but does so at a rate that is faster than any other estimator in this class.

Theoretically, the concepts of efficiency discussed in Section A.2 may be extended to the large sample case by utilizing the definitions of asymptotic bias and covariance matrix proposed earlier (Definition 17). These definitions, however, require that the first and second moments of the proposed estimator exist for all sample sizes N . It is not always possible to meet this requirement, because quite frequently these moments may not exist for small values of N . On the other hand, these moments may exist but may be unknown. In either case, this direct approach may

fail to provide the necessary information on the asymptotic behavior of the estimators.

In practice, this difficulty is resolved by utilizing the concept of 'convergence in distribution'. Thus, it is possible to show that under fairly realistic assumptions, the distribution function of $\sqrt{N}(\hat{\theta}_N - \theta)$ converges asymptotically to multivariate normal and generally the convergence to normality is uniform over compact subsets of the admissible parameter space.

Such estimators as Consistent Asymptotically Normal (CAN); Best Asymptotically Normal (BAN), Consistent Asymptotically Uniformly Normal (CAUN), etc., claim their origin from this concept, and we may define asymptotic (relative) efficiency in terms of the moments of the distribution (instead of the asymptotic moments) as follows:

Definition 18. Asymptotic Efficiency (Relative)

Let $(x_1, x_2, \dots, x_N)$ be a random sample from a population characterized by the p.d.f. $f(., \theta)$. Let $\hat{\theta}_N$ and $\tilde{\theta}_N$ be two consistent estimators of θ such that asymptotically:

$$\sqrt{N}(\hat{\theta}_N - \theta) \sim \text{N.D.}(\underline{0}, \hat{\Sigma}); \quad \sqrt{N}(\tilde{\theta}_N - \theta) \sim \text{N.D.}(\underline{0}, \tilde{\Sigma})$$

Then $\hat{\theta}_N$ is said to be asymptotically efficient relative to $\tilde{\theta}_N$ if $\Sigma = \tilde{\Sigma} - \hat{\Sigma}$ is positive semidefinite.

Remark 8. If the asymptotic distribution of $\sqrt{N}(\hat{\theta}_N - \theta)$ is given by $\sqrt{N}(\hat{\theta}_N - \theta) \sim \text{N.D.}(\underline{0}, \hat{\Sigma})$ then we may infer that asymptotically $\hat{\theta}_N \sim \text{N.D.}(\theta, \frac{1}{N} \hat{\Sigma})$

and most often the asymptotic bias and the asymptotic covariance matrix of $\hat{\theta}_N$ is taken to be the mean and covariance matrix of the above asymptotic distribution.

A.4 Some Relevant Theorems

Definition 19. General Linear Model

Let the observations $y(k)$; $k = 1, 2, \dots, N$ be generated by the following linear relationship:

$$y(k) = \beta_1 x_1(k) + \beta_2 x_2(k) + \dots + \beta_p x_p(k) + \eta(k)$$

$$k = 1, 2, \dots, N$$

or

$$y(k) = \sum_{i=1}^p \beta_i x_i(k) + \eta(k) \quad k = 1, 2, \dots, N \quad (\text{A.4.1})$$

where $y(\cdot)$ is said to be the independent variable or the regressand, $x_i(\cdot)$; $i = 1, 2, \dots, p$ are called the dependent or explanatory variables or regressors, $\eta(\cdot)$ is the disturbance and β_i ; $i = 1, 2, \dots, p$ are the $p(< N)$ unknown parameters to be estimated.

Relation A.4.1, together with the following set of assumptions constitutes the general linear model:

1. There are no linear dependencies among the explanatory variables and these variables are known. (A.4.2)
2. The explanatory variables are uniformly bounded by the finite (but perhaps very large) constant X , i.e.,

$$|x_i(k)| \leq \chi \quad \forall \quad i \text{ and } k \quad (\text{A.4.3})$$

3. The disturbance $\eta(k)$ have zero mean, are serially uncorrelated and have a common finite variance σ^2 , i.e.,

$$E[\eta(k)] = 0, \quad k = 1, 2, \dots, N$$

$$E[\eta(k)\eta(j)] = \begin{cases} 0 & \text{if } k \neq j \\ \sigma^2 & \text{if } k = j \end{cases} \quad (\text{A.4.4})$$

$$k, j = 1, 2, \dots, N.$$

4. The explanatory variables are uncorrelated with the disturbances, i.e.,

$$E[x_i(k)\eta(j)] = 0 \quad \forall \quad i, j, k. \quad (\text{A.4.5})$$

It is more common to specify the general linear model in the following vector matrix notation. Thus for $k = 1, 2, \dots, N$ (A.4.1) may be written as:

$$\begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{bmatrix} = \begin{bmatrix} x_1(1) & x_2(1) & \dots & x_p(1) \\ x_1(2) & x_2(2) & \dots & x_p(2) \\ \vdots & \vdots & \dots & \vdots \\ x_1(N) & x_2(N) & \dots & x_p(N) \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \eta(1) \\ \eta(2) \\ \vdots \\ \eta(N) \end{bmatrix} \quad (\text{A.4.6})$$

or

$$\underline{y} = \underline{X}\underline{\beta} + \underline{\eta} \quad (\text{A.4.6})$$

where $\underline{y}$ = $N \times 1$ Observation vector (known)
 $\underline{X}$ = $N \times p$ Design matrix (known)
 $\underline{\beta}$ = $p \times 1$ Parameter vector (unknown)
 $\underline{\eta}$ = $N \times 1$ Disturbance vector (unknown)

and the associated assumptions may be written as:

$$1. \text{rank}(X) = p \quad (\text{A.4.7})$$

$$2. ||X|| \leq \chi \text{ where } ||X|| \text{ is the sup-norm of } X \quad (\text{A.4.8})$$

$$3. E(\underline{\eta}) = 0 \text{ and } \text{Cov}(\underline{\eta}) = \sigma^2 I, \quad \sigma^2 < \infty \quad (\text{A.4.9})$$

$$4. E(X'\underline{\eta}) = 0 \quad (\text{A.4.10})$$

Theorem 1. (Gauss Markov)

In the general linear model (A.4.6-10), the estimator

$$\hat{\underline{\beta}}_{LS} = (X'X)^{-1} X'y \quad (\text{A.4.11})$$

obtained by minimizing the quadratic form

$$Q = (\underline{y} - X\hat{\underline{\beta}})'(\underline{y} - X\hat{\underline{\beta}}) \quad (\text{A.4.12})$$

is linear, unbiased and efficient (within the class of linear unbiased estimators) estimator of $\underline{\beta}$ and further

$$\hat{\sigma}^2 = \frac{Q_{\min}}{N-p} = \frac{1}{N-p} [(\underline{y} - X\hat{\underline{\beta}}_{LS})'(\underline{y} - X\hat{\underline{\beta}}_{LS})] \quad (\text{A.4.13})$$

is an unbiased estimator of σ^2 .

Corollary

In the general linear model (A.4.6-10), if

$$\underline{\eta} \sim \text{N.D.}(\underline{0}; \sigma^2 I) \text{ then:}$$

- (i) The Least Squares and Maximum Likelihood estimators of $\underline{\beta}$ are identical.

- (ii) $\hat{\underline{\beta}}_{LS} \sim \text{N.D.}(\underline{\beta}, \sigma^2(X'X)^{-1})$.
- (iii) $\hat{\underline{\beta}}_{LS}$ is minimum variance bound.
- (iv) $\hat{\underline{\beta}}_{LS}$ is sufficient.

Theorem 2. (Aitken)

In the general linear model (A.4.6-10), if $\underline{u} \sim \text{N.D.}(\underline{0}, \Omega)$ where Ω is a known positive definite matrix, then the Aitken (or Generalized Least Squares) estimator

$$\hat{\underline{\beta}}_{GLS} = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1} \underline{y} \quad (\text{A.4.14})$$

obtained by minimizing the quadratic form

$$Q_{\Omega} = (\underline{y} - X\hat{\underline{\beta}})' \Omega^{-1} (\underline{y} - X\hat{\underline{\beta}}) \quad (\text{A.4.15})$$

is the best, linear unbiased estimator (BLUE) of $\underline{\beta}$ and

$$\hat{\underline{\beta}}_{GLS} \sim \text{N.D.}(\underline{\beta}, (X'\Omega^{-1}X)^{-1}) \quad (\text{A.4.1.})$$

Corollary 1.

For the model specified in Theorem 2, the Ordinary Least Squares (OLS) estimator of $\underline{\beta}$ is unbiased and consistent but inefficient.

Corollary 2.

In the model specified in Theorem 2, if the covariance matrix Ω is unknown but a consistent estimator $\hat{\Omega}$ for Ω is available, then the approximate GLS estimator (also called the 'feasible Aitken estimator')

$$\hat{\underline{\beta}}_{AGLS} = (\underline{X}'\hat{\underline{\Omega}}^{-1}\underline{X})^{-1}\underline{X}'\hat{\underline{\Omega}}^{-1}\underline{Y} \quad (\text{A.4.17})$$

is the best linear consistent estimator of $\underline{\beta}$, i.e., it is

- i) consistent;
- ii) asymptotically unbiased;
- and iii) asymptotically efficient.

(Proof: Maddala, G.S., "Generalized Least Squares with an Estimated Var-Covar Matrix", *Econometrica*, Vol. 39, No. 1, January 1971.)

Theorem 3 (Mann and Wald)

In the general linear model (A.1.1-5), if the explanatory variables $x_i(k) \equiv y(k-i)$, i.e., if (A.1.1) is a linear difference equation of the form:

$$y(k) = \beta_1 y(k-1) + \beta_2 y(k-2) + \dots + \beta_p y(k-p) + \eta(k) \quad (\text{A.4.18})$$

$$k = 1, 2, \dots, N$$

and if the roots of the algebraic equation

$$w^p - \beta_1 w^{p-1} - \beta_2 w^{p-2} - \dots - \beta_p = 0 \quad (\text{A.4.19})$$

lie within the origin centered unit circle, then the OLS estimator $\hat{\underline{\beta}}_{LS}$ is best linear consistent and the asymptotic distribution of $\hat{\underline{\beta}}_{LS}$ is provided by

$$\sqrt{N} (\hat{\underline{\beta}}_{LS} - \underline{\beta}) \sim \text{N.D.} \left[\underline{0}, \sigma^2 \left(\frac{\underline{X}'\underline{X}}{N} \right)^{-1} \right] \quad (\text{A.4.20})$$

(Proof: Mann, N.B. and A. Wald, "On the Statistical Treatment of Linear Difference Equations", *Econometrica*, Vol. 11, pp. 173-220.)

Theorem 4 (Slutsky)

If the sequence of $p \times 1$ vector random variables $(\underline{z}_1, \underline{z}_2, \dots, \dots)$ converges in probability to the constant vector $\underline{c}$, then

$$\text{p.lim}_{N \rightarrow \infty} [\underline{h}(\underline{z}_N)] = \underline{h}(\underline{c}) \quad (\text{A.4.21})$$

where $\underline{h}(\cdot)$ is a $q \times 1$ vector function that is continuous in $\underline{z}_i$.

(Proof: Cramer, H., "Mathematical Methods of Statistics", Princeton University Press, Princeton, 1946).

Corollary

Consider the sequence of matrices $(A_1, A_2, \dots, \dots)$ and $(B_1, B_2, \dots, \dots)$ whose elements are random variables and which converge in probability to the constant matrices A and B . Since the product of two matrices is a continuous function of their elements, we have

$$\text{p.lim}_{N \rightarrow \infty} [A_N \times B_N] = \text{p.lim}_{N \rightarrow \infty} [A_N] \times \text{p.lim}_{N \rightarrow \infty} [B_N] = A \times B \quad (\text{A.4.22})$$

Similarly it follows that

$$\text{p.lim}_{N \rightarrow \infty} [A_N^{-1}] = \left[\text{p.lim}_{N \rightarrow \infty} (A_N) \right]^{-1} = A^{-1} \quad (\text{A.4.23})$$

A P P E N D I X B

In the first half of this appendix, some properties of the noise covariance matrix Ω , in the Astrom's difference equation model are discussed. In the second half, we use these properties to derive the asymptotic distribution of the IV estimator for the process parameters in the above model and show that the elements of the asymptotic covariance matrix of the IV estimator are inversely proportional to N , the number of observations.

B.1 Properties of Ω , the Noise Covariance Matrix

The model being considered is

$$\begin{aligned} y(k) = & b_0 u(k) + b_1 u(k-1) + \dots + b_n u(k-n) - a_1 y(k-1) - a_2 y(k-2) - \dots \\ & - a_n y(k-n) + e(k) + c_1 e(k-1) + \dots + c_n e(k-n) \\ & k = 1, 2, \dots, N \end{aligned} \quad (B.1.1)$$

with

$$e(k) \sim \text{N.I.D.}(0, \lambda^2) \quad (B.1.2)$$

To simplify the discussion we shall assume that the initial conditions in the above model are identically zero, i.e.,

$$y(j) = u(j) = e(j) = 0 \quad \forall j \leq 0 \quad (B.1.3)$$

For $k = 1, 2, \dots, N$, we therefore have:

$$\begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(n+1) \\ \vdots \\ y(N) \end{bmatrix} = \begin{bmatrix} u(1) & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ u(2) & u(1) & 0 & \dots & 0 & y(1) & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ u(n+1) & u(n) & u(n-1) & \dots & u(1) & y(n) & y(n-1) & \dots & y(1) \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ u(N) & u(N-1) & u(N-2) & \dots & u(N-n) & y(N-1) & y(N-2) & \dots & y(N-1) \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \\ -a_1 \\ \vdots \\ -a_n \end{bmatrix} + \begin{bmatrix} v(1) \\ v(2) \\ \vdots \\ v(n+1) \\ \vdots \\ v(N) \end{bmatrix} \quad (B.1.4)$$

or

$$\underline{y} = M \underline{\beta} + \underline{v} \quad (B.1.5)$$

where

$$\underline{\beta} = [b_0 \ b_1 \ \dots \ b_n \ -a_1 \ -a_2 \ \dots \ -a_n]' \quad (B.1.6)$$

and

$$\underline{v} = \begin{bmatrix} v(1) \\ v(2) \\ \vdots \\ v(n+1) \\ \vdots \\ v(N) \end{bmatrix} = \begin{bmatrix} 1 & & & & & \\ c_1 & 1 & & & & \\ \vdots & \vdots & & & & \\ c_n & c_{n-1} & \dots & c_1 & 1 & \\ \vdots & \vdots & & \vdots & \vdots & \\ \vdots & \vdots & & \vdots & \vdots & \\ c_n & c_{n-1} & \dots & c_1 & 1 & \end{bmatrix} \begin{bmatrix} e(1) \\ e(2) \\ \vdots \\ e(n+1) \\ \vdots \\ e(N) \end{bmatrix} \quad (B.1.8)$$

or

$$\underline{v} = \Delta \underline{e}$$

where Δ is a unit lower triangular matrix whose elements (δ_{ij}) have the following properties:

(a) Elements on each diagonal of Δ are identical.

(b) For $i = 1, 2, \dots, N$

$$\delta_{ii} = 1 \rightarrow |\Delta| = 1 \rightarrow \Delta \text{ is non-singular.}$$

(c) For $i = n+2, n+3, \dots, N$ and $j = 1, 2, \dots, N-n-1$

$$\delta_{ij} = 0$$

i.e., there are no more than $n+1$ non-zero elements in a row.

The covariance matrix of $\underline{y}$ is given by

$$\begin{aligned} \Omega &= E[(\underline{y} - E(\underline{y}))(\underline{y} - E(\underline{y}))'] \\ &= E[(\underline{y} - E(\Delta \underline{e}))(\underline{y} - E(\Delta \underline{e}))'] \end{aligned}$$

and since from (B.1.2) $E(\underline{e}) = 0$, we have

$$\Omega = E(\underline{y} \underline{y}') = E(\Delta \underline{e} \underline{e}' \Delta) = \lambda^2 \Delta \Delta' \quad (\text{B.1.9})$$

Thus Ω admits a LUD decomposition and in the light of the properties (a), (b) and (c) possessed by Δ , we may infer that:

1. Ω is a symmetric and positive definite.
2. Ω is a semi-diagonal matrix of order n so that all its sub-diagonals greater than and including the $(n+1)^{\text{st}}$ have identically zero elements.

3. All the information about Ω is carried by the coefficients $c_1, c_2, \dots, c_n$ (Eqn. (B.1.1)) and if these coefficients are known, we can completely assemble Ω , except for the scaling factor λ^2 .

B.2 Asymptotic Distribution of the IV Estimator

We shall again consider the Astrom's difference equation model (Eqns. (B.1.1) and (B.1.2)) which may be rewritten in the matrix notation as:

$$\underline{y} = M\underline{\beta} + \underline{v} \quad (\text{B.2.1})$$

$$\underline{v} = \Delta \underline{e} \quad (\text{B.2.2})$$

$$\underline{e} \sim \text{N.I.D. } (\underline{0}, \lambda^2 \underline{I}), \quad \lambda^2 < \infty \quad (\text{B.2.3})$$

where the variables in the above model are as defined in the previous section. We shall further assume that the elements of M and Δ are uniformly bounded.

Let the IV estimator for $\underline{\beta}$ be given as

$$\hat{\underline{\beta}}_{\text{IV}} = (\underline{Z}'\underline{M})^{-1} \underline{Z}' \underline{y} \quad (\text{B.2.4})$$

where the elements of the IV matrix $\underline{Z}$ satisfy the following conditions:

1. Elements of Z are either deterministic or if they are stochastic, then they are statistically independent of the elements of $\underline{v}$.*
2. Elements of Z are uniformly bounded (with probability one, if they are stochastic).
3. $\text{p.lim}_{N \rightarrow \infty} (1/N Z'M) = \sum_{ZM}$ which is a non-stochastic and non-singular matrix.

We can write (B.2.4) as

$$\hat{\underline{\beta}}_{IV} = \underline{\beta} + \left(\frac{1}{N} Z'M \right)^{-1} \left(\frac{1}{N} Z' \underline{v} \right) \quad (\text{B.2.5})$$

or

$$\hat{\underline{\beta}}_{IV} = \underline{\beta} + \left(\frac{1}{N} Z'M \right)^{-1} \left(\frac{1}{N} Z' \Delta \underline{e} \right) \quad (\text{B.2.6})$$

The problem is to determine the limiting distribution of the sequence $\sqrt{N} (\hat{\underline{\beta}}_{IV} - \underline{\beta})$. This is precisely what is meant by the asymptotic distribution of the IV estimator.** We are not speaking of the distribution of $\hat{\underline{\beta}}_{IV}$ for $N = \infty$, because the latter, being a consistent estimator, will have a degenerate distribution. To reiterate, we are attempting to find out if $\hat{\underline{\beta}}_{IV}$ (or some simple transformation of $\hat{\underline{\beta}}_{IV}$) converges in distribution to some random vector with a well defined p.d.f. before it converges in probability to a constant vector. From (B.2.6) we have

* This condition is stronger than the usual condition $(1/N Z'v) \xrightarrow{P} 0$. It may, however, be justified on the grounds that usually the elements of the IV matrix Z are obtained through a suitable transformation of the input variables $u(k)$, that are independent of the elements of $\underline{v}$.

** Dhrymes, P.J., *Econometrics*, Harper & Rowe, 1970, pp. 190-92.

$$\sqrt{N} (\hat{\beta}_{IV} - \beta) = \left(\frac{1}{N} Z' M \right)^{-1} \left(\frac{1}{\sqrt{N}} Z' \Delta \underline{e} \right) \quad (\text{B.2.7})$$

Since by assumption, the probability limit of $(1/N Z' M)^{-1}$ exists as a non-stochastic matrix, the problem is reduced to finding the limiting distribution of $(1/\sqrt{N} Z' \Delta \underline{e})$. However,

$$\underline{e} \sim \text{N.I.D. } (0, \lambda^2 I)$$

so that $(1/\sqrt{N} Z' \Delta \underline{e})$ is jointly normal with mean zero and the covariance matrix

$$\Phi_N = \lambda^2 \left[\frac{1}{N} Z' \Delta \Delta' Z \right] = \left[\frac{1}{N} Z' \Omega Z \right] \quad (\text{B.2.8})$$

For $N \rightarrow \infty$ therefore, $(1/\sqrt{N} Z' \Delta \underline{e})$ is jointly normal with zero mean and covariance matrix

$$\lim_{N \rightarrow \infty} (\Phi_N) = \lim_{N \rightarrow \infty} \left[\frac{1}{N} Z' \Omega Z \right] \quad (\text{B.2.9})$$

provided the limit exists. The fact, that for the model under discussion, such a limit exists, can be argued as follows:

It was shown in Section B.1 that Ω is symmetric positive definite and that it is a semi-diagonal matrix of order n i.e., it has non-zero elements on the principal diagonal and up to the n^{th} subdiagonal on either side of the main diagonal. We can therefore represent Ω as the sum of $2n+1$ matrices of the form

$$\Omega = \sum_{\ell=-n}^n \Omega_{\ell} \quad (\text{B.2.10})$$

where Ω_{ℓ} is a matrix that has non-zero elements only on the ℓ^{th} sub-diagonal above the main diagonal ($\ell = 0$).

If the $N \times p$ matrix Z is partitioned into its columns as

$$Z = [\underline{Z}_1 \mid \underline{Z}_2 \mid \dots \mid \underline{Z}_p] \quad (\text{B.2.11})$$

then a typical element $\phi_N(i,j)$ of $\Phi_N = (1/N Z' \Omega Z)$ is a scalar of the form

$$\phi_N(i,j) = \frac{1}{N} \underline{Z}_i' \Omega \underline{Z}_j, \quad i,j = 1,2,\dots,p \quad (\text{B.2.12})$$

or

$$\phi_N(i,j) = \frac{1}{N} \underline{Z}_i' \left[\sum_{\ell=-n}^n \Omega_{\ell} \right] \underline{Z}_j \quad (\text{B.2.13})$$

which is a sum of $2n+1$ terms. Since the elements of Z and Ω are assumed to be uniformly bounded, each term in the above summation is bounded. Hence $\phi_N(i,j)$, $i,j = 1,2,\dots,p$ is a finite sum of finite terms and is therefore bounded. We may then conclude that $\phi_N(i,j)$, $N = 1,2,\dots$ is an infinite sequence with bounded elements, and the existence of its limit is assured by the Bolzano Weirstrass theorem.*

Thus, for $N \rightarrow \infty$, $(1/\sqrt{N} Z' \Delta \underline{e})$ is jointly normal with mean zero and covariance matrix Φ , which is the limiting value of Φ_N . From (B.2.7) it follows that the asymptotic distribution of $\hat{\beta}_{IV}$ is given by

* Kreszig, E., *Advanced Engineering Mathematics*, John Wiley & Sons, 1967, pp. 614-615.

$$\sqrt{N} (\hat{\underline{\beta}}_{IV} - \underline{\beta}) \sim \text{N.D.} \left(\underline{0}, \underline{\sum}_{ZM}^{-1} \Phi \underline{\sum}_{MZ}^{-1} \right) \quad (\text{B.2.14})$$

where $\underline{\sum}_{ZM}$ is the limiting value of $(1/N \underline{Z}'\underline{M})$.

From the above asymptotic distribution we may infer that asymptotically or for large sample sizes

$$E(\hat{\underline{\beta}}_{IV}) = \underline{\beta} \quad (\text{B.2.15})$$

$$\text{Cov} (\hat{\underline{\beta}}_{IV}) = \frac{1}{N} \left[\underline{\sum}_{ZM}^{-1} \Phi \underline{\sum}_{MZ}^{-1} \right] \quad (\text{B.2.16})$$

and that the elements of the asymptotic covariance matrix of the IV estimator are inversely proportional to N , the number of observations.

A P P E N D I X C

Generation of Normally and Independently Distributed Random Variables

Monte Carlo studies on the various estimators presented in the thesis require sequences of random numbers that are Normally and Independently Distributed (N.I.D.) with zero mean and a specified variance. In all the simulation experiments, the estimators are tested for 20 different noise sequences and their average properties (over these 20 trials) have been used as a basis of comparison. The same 20 sequences of random numbers were employed throughout, to generate the artificial data, and the method employed for generating these random numbers on the digital computer and their statistical properties are presented in this appendix.

C.1 Generation of N.I.D. Random Numbers

A common method for generating N.I.D. random variables on a digital computer is to use a suitable transformation on a sequence of Uniformly and Independently Distributed (U.I.D.) variates. A variety of techniques for generating uniformly distributed random numbers on a digital computer are available. The choice of a specific method is generally dictated by the speed of generation and the degree of 'randomness' in the generated sequence. Unfortunately, a single and absolute test that will permit us to pronounce that numbers generated by Method A are more random than those generated by Method B, is not available. An excellent survey that includes all the important references on the subject

is provided by Chambers^{*}, in which he concludes that for many applications the 'multiplicative congruential' method (also called 'power residue' method) is a good choice with minimum pitfalls. Similarly, Allard, Dobell and Hull^{**} point out that with a careful choice of the parameters the multiplicative method should be consistently reliable.

In the multiplicative congruential method, the $(n+1)^{\text{st}}$ random number R_{n+1} , is obtained from the n^{th} random number R_n by the recurrence relation:

$$R_{n+1} = a R_n \pmod{m} \quad (\text{C.1.1})$$

where a and m are positive integers ($a < m$). For a digital computer with word length b (bits), the maximum period length, $\frac{1}{2}(2^b)$, can be achieved by choosing $m = 2^b$ and the seed value or starting value R_0 as an odd integer. The multiplier a , affects the randomness of the generated numbers and the value most often recommended is $a = \sqrt{m} + 3$.

For the XDS SIGMA 7 digital machine that was used for simulation experiments, the word length is 32 bits, so that the following parameter values were used:

$$m = 2^{32}; \quad a = \sqrt{m} + 3 = 65539,$$

* Chambers, R.P., "Random Number Generation on Digital Computers", IEEE Spectrum, Vol. 4, pp. 48-56, 1967.

** Allard, J.L., A.R. Dobell and T.E. Hull, "Mixed Congruential Random Number Generators for Decimal Machines", J.A.C.M., Vol. 10, pp. 131-141, 1963.

and with R_0 an odd integer, the relationship C.1.1 provides a sequence of U.I.D. random numbers in the interval (0,1), with the maximum period length $p = 2^{30} > 10^9$. This uniform random number generator is available as a library function RANDOM(IX) on the XDS SIGMA 7.

The normal random variables were obtained by using the 'direct' method recommended by Muller*. This method starts by generating two U.I.D. random numbers R_1 and R_2 in the interval (0,1). Two N.I.D.(0,1) random variables X_1 and X_2 are then generated by using the relationships:

$$\begin{aligned} X_1 &= (-2 \log_e R_1)^{\frac{1}{2}} \cos(2\pi R_2) \\ X_2 &= (-2 \log_e R_1)^{\frac{1}{2}} \sin(2\pi R_2) \end{aligned} \quad (C.1.2)$$

The U.I.D. random numbers R_1 and R_2 are obtained by using the function routine RANDOM(IX) with two different seed values IX1 and IX2. The Subroutine WNOISE(XNOISE,I1,I2,N,VAR) which utilizes relationships C.1.2 to generate a sequence of N normal deviates XNOISE, with mean zero and variance VAR together with Function RANDOM(IX) are listed at the end of this appendix. The odd integers I1 and I2 in WNOISE represent the seed values for RANDOM(IX). The seed values employed for generating the 20 different sequences of N.I.D. deviates used for simulations is also listed as the calling program for the above subroutines.

C.2 Evaluation of Subroutine WNOISE

As mentioned earlier, the simulation experiments require 20 different sequences of N.I.D. random variables with zero mean and a speci-

* Muller, M.E., "A Comparison of Methods for Generating Normal Deviates on Digital Computers", J.A.C.M., Vol. 6, pp. 376-383, 1959.

fied variance. Twenty different sequences, each containing $N = 20,000$ N.I.D. (0,1) random numbers were generated by using Subroutine WNOISE. For each of these sequences:

$$e_i(k); \quad i = 1, 2, \dots, 20, \quad k = 1, 2, \dots, 20,000$$

the Sample Mean, Sample Variance and Sample Serial Correlations with lags of upto and including 10 were calculated. The Sample Means and Variances for the 20 sequences together with their average values (over the 20 trials) and the corresponding standard deviations are given in Table C.1. The estimated means and variances are very close to the expected values of 0 and 1 respectively.

The Sampled Serial Correlations for lags up to and including 10 (averaged over 20 trials) are tabulated in Table C.2. Corresponding standard deviations are also included. It was noted that for every individual sequence and for every value of lag the absolute value of the sample serial correlation was less than 0.015. Further, no regular pattern was observed. It was therefore considered safe to assume that the sequence of random numbers generated by the subroutine WNOISE were not serially correlated in the range of interest.

To test for normality, the cumulative probability distribution of each sequence of random numbers was measured in the range -5 to +5. The average probability distribution (percent) over the 20 trials, together with their standard deviations are given in Table C.3, in which the corres-

No. of Samples = 20,000

TRIAL	SAMPLE MEAN	SAMPLE VARIANCE
1	0.0105	1.0092
2	0.0057	1.0127
3	0.0040	0.9946
4	0.0023	1.0010
5	-0.0038	1.0017
6	0.0064	1.0045
7	0.0064	1.0076
8	-0.0007	1.0006
9	-0.0075	1.0068
10	0.0050	0.9961
11	0.0107	1.0005
12	-0.0117	1.0026
13	0.0074	0.9987
14	0.0067	0.9862
15	0.0005	0.9990
16	0.0049	0.9890
17	0.0011	0.9975
18	-0.0045	0.9994
19	-0.0132	1.0055
20	-0.0003	1.0004
AVERAGE	0.0015±0.0068	1.0007±0.0065

TABLE C.1. Sample Mean and Variance for
Output of WNOISE

No. of Trials = 20
 No. of Samples in Each Trial = 20,000

SHIFT = j	AVERAGE SAMPLE CORRELATION = $\bar{\sigma}_j$		
0	1.0007	$\pm$	0.0063
1	-0.0013	$\pm$	0.0075
2	0.0009	$\pm$	0.0066
3	0.0024	$\pm$	0.0076
4	0.0037	$\pm$	0.0094
5	0.0003	$\pm$	0.0074
6	-0.0011	$\pm$	0.0046
7	-0.0007	$\pm$	0.0082
8	-0.0025	$\pm$	0.0066
9	-0.0030	$\pm$	0.0067
10	-0.0002	$\pm$	0.0064

TABLE C.2. Average Serial Correlations for
 Output of WNOISE

No. of Trials = 20
 No. of Samples in Each Trial = 20,000

z	AVERAGE $F(z)$ (MEASURED)	$F(z)$ (TRUE)
-3.5	0.0225 ± 0.0113	-
-3.0	0.1407 ± 0.0233	0.13
-2.5	0.6217 ± 0.0568	0.62
-2.0	$2.2907 \pm -/-3-$	2.27
-1.5	6.6237 ± 0.2393	6.68
-1.0	15.8082 ± 0.2393	15.87
-0.5	30.8904 ± 0.3050	30.85
0.0	50.0299 ± 0.2932	50.00
0.5	69.0834 ± 0.3261	69.15
1.0	84.0736 ± 0.2130	84.13
1.5	93.2348 ± 0.1293	93.32
2.0	97.6924 ± 0.0904	97.72
2.5	99.3671 ± 0.0668	99.38
3.0	99.8589 ± 0.0302	99.87
3.5	99.9744 ± 0.0082	-

TABLE C.3. Average Amplitude Probability Distribution (Percent) for Output of WNOISE

ponding true value for a standard normal deviate have been included for comparison. The measured probability distribution is also plotted on a Normal Probability paper (Fig. C.1) which indicates that the random numbers generated by WNOISE closely follow the normal distribution up to and beyond 3 standard deviations.

A further check on the 'skewness' and 'kurtosis' of the distribution was obtained by computing the Coefficients of Skewness and Kurtosis respectively for each of the 20 sequences. These are given in Table C.4. These coefficients were found to be quite close to the true values of 0 and 3 for a standard normal deviate.

Since the sequences of random numbers generated by WNOISE were found to be normally distributed with insignificant serial correlation it was concluded that they were also independently distributed.

Fig. C.1. Average Probability Distribution of Output Samples from WNOISE

No. of Samples = 20,000

TRIAL	COEF. OF SKEWNESS	COEF. OF KURTOSIS
1	0.0080	2.9228
2	0.0242	3.0350
3	0.0025	3.0189
4	0.0019	2.9943
5	0.0036	3.0374
6	0.0353	3.0198
7	0.0071	2.9826
8	0.0102	3.0135
9	-0.0278	2.9753
10	-0.0082	3.0704
11	0.0172	2.9862
12	0.0132	3.0187
13	0.0197	3.0289
14	-0.0074	2.9899
15	0.0010	2.9877
16	-0.0044	3.0183
17	0.0109	3.0445
18	-0.0046	3.0193
19	-0.0248	3.0305
20	0.0210	3.0087
AVERAGE	0.0049 ± 0.0155	3.0101 ± 0.0313

TABLE C.4. Coefficients of Skewness and Kurtosis
for the Output of WNOISE

```

      .
      .
      .
      .
      DU 199 N = 1,20
      I1 = 110945-2*N
      I2 = 999835-2*N
      NA = 20000
      VAR = 1.0
      CALL WNOISE (E,I1,I2,NA,VAR)
      .
      .

```

```

      END

```

```

      SUBROUTINE WNOISE (XNOISE,I1,I2,N,VAR)
      DIMENSION XNOISE(20000)
      DO 2 I = 1,N,2
      R1 = RANDOM(I1)
      R2 = RANDOM(I2)
      X1 = -2.0*A LOG(R1)
      X2 = SQRT(X1)
      X3 = +6.283185307*R2
      J = I+1
      X4 = COS(X3)
      X5 = 1.0-X4*X4
      X6 = SQRT(X5)
      IF(R2-0.5)3,3,4
      X6 = 0.0-X6
      XNOISE(I) = SQRT(VAR)*X2*X4
      XNOISE(J) = SQRT(VAR)*X2*X6
      RETURN
      END

```

```

      FUNCTION RANDOM(IX)
      IY = IX*65539
      IF(IY)5,6,6
      IY = IY+2147483647+1
      YFL = IY
      YFL = YFL*0.4656613E-9
      RANDOM = YFL
      IX = IY
      RETURN
      END

```