

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

U·M·I

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

PREVIEW

Order Number 9314427

Evaluation of order determination procedures in ARMA models

Parkhurst, Anne Mullins, Ph.D.

The University of Nebraska - Lincoln, 1992

Copyright ©1993 by Parkhurst, Anne Mullins. All rights reserved.

U·M·I

300 N. Zeeb Rd.
Ann Arbor, MI 48106

PREVIEW

EVALUATION OF ORDER DETERMINATION PROCEDURES IN ARMA MODELS

by

Anne M. Parkhurst

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Interdepartmental Area of Engineering

Under the Supervision of Professor John L. Ballard

Lincoln, Nebraska

December, 1992

DISSERTATION TITLE

Evaluation of Order Determination Procedures in ARMA Models

BY

Anne M. Parkhurst

SUPERVISORY COMMITTEE:

APPROVED

DATE

John L. Ballard
Signature

11/20/92

John L. Ballard
Typed Name

Morris H. Schneider
Signature

11/20/92

Morris H. Schneider
Typed Name

K. P. Rajurkar
Signature

11/20/92

K. P. Rajurkar
Typed Name

James A. DeShazer
Signature

11/20/92

James A. DeShazer
Typed Name

Signature

Typed Name

Signature

Typed Name



EVALUATION OF ORDER DETERMINATION PROCEDURES IN ARMA MODELS

Anne M. Parkhurst, Ph.D.

University of Nebraska, 1992

Adviser: John L. Ballard

Autoregression-moving average (ARMA) models provide insight into many biological systems. One of the most difficult decisions in ARMA modeling is identifying the order of the model. Many procedures have been proposed. The purpose of this study was to compare Pandit and Wu's multi-objective test criterion with Akaike's information criterion and Schwartz's Bayesian criterion, SBC.

Several versions of the criteria were compared for 10 processes which were simulations of ARMA models ranging in order from (1,1) to (8,7). Batches of 100 realizations were generated for each model. The performance of the criteria varied depending on how closely a process complied with the assumptions, length of series and alpha levels chosen. The error variance and method of parameter estimation had no effect on the relative merits of the criteria.

The proposed strategy uses Pandit and Wu's $(2p, 2p-1)$ portfolio and SBC (or F-ratio, if sample size ≤ 125) to identify a neighborhood for the model on the first pass. The portfolio is modified to include all models between adjacent primary models which define the neighborhood. In the second

pass, a model is selected using SBC. Three classical time series were assessed by this strategy. The results compare favorably to those in the literature.

The proposed strategy was utilized to identify a sensible heat loss model for non-laying hens. The ARMA(2,0) model which was selected for all four hens, explained 70 to 89% of the total variation. The proposed strategy was also used to identify a tympanic temperature model for steers housed in a constant thermoneutral environment. The ARMA(3,0) model was adequate for both animals in the study, although the model accounted for only 15 to 27% of the total variation.

This dissertation is divided into four sections. The first contains a literature review of time series relative to dynamic systems. The second provides a comparison of order determination procedures and introduces a new strategy. The third describes the utilization of proposed strategy in modeling biological processes. The last section discusses various aspects in the construction of simulation software.

ACKNOWLEDGEMENTS

The road to completing this study was long and winding. I could not have made it without the help and support of my professors, friends and colleagues. As I began the program, the first person to encourage me and expedite the process was Prof. Morris H. Schneider - without him there would have been no beginning.

I would like to express my sincere appreciation to my advisor, Prof. John L. Ballard whose support, counsel, and encouragement kept me going through many trials and tribulations. Without his patience there would have been no end.

Of all those who helped me through the middle, I'd especially like to thank the members of the reading committee, Kamalakar P. Rajurkar and James A. DeShazer. To Prof. Rajurkar I express gratitude for introducing me to data dependent systems, and providing valuable constructive criticism and suggestions.

To Prof. DeShazer, I am deeply indebted for the tutorial in environmental systems, for sharing his creativity and friendship, and for introducing me to other biological engineers, especially G. Leroy Hahn.

I wish to thank Dr. Hahn for sharing his enthusiasm, for his receptiveness to new ideas, for the introduction to chaos and for the many thoughtful and extremely stimulating scientific discussions.

To my loving parents - James and Josephine Mullins
and

my darling children

Lars Jimmy James Paul Mullins Parkhurst

Vera Elizabeth Josephine Anne Mullins Parkhurst

The past and the future,
May they forever be entwined

PREVIEW

TABLE OF CONTENTS

I. BACKGROUND*1. Stochastic Processes*

1.a. AUTOCORRELATION AND AUTOCOVARIANCE . . .	3
1.b. STATIONARITY	4
1.c. INVERTIBILITY	9

2. Standard Models

2.a. AUTOREGRESSIVE MODELS	11
i. Zero-Order Autoregressive Models, $AR(0)$	11
ii. First-Order Autoregressive Model, $AR(1)$	12
(1) Stationarity by Substitution	13
(2) Inverse Autoregressive Operator. .	15
(3) Stationarity from Green's Function .	16
iii. Second-order Autoregressive Model, $AR(2)$	18
(1) Physical Interpretation	22
iv. Review of Complex Numbers and Trigonometry	26
v. General-order Autoregressive Process, $AR(p)$	32
2.b. MOVING AVERAGE MODELS	34

	ii
2.c. MIXED MODELS	40
i. The ARMA(2,1) process	41
ii. Green's function for the Mixed Model	42
iii. Relationship between Autocorrelation and Green's function	43

3. Methods Of Parameter Estimation

3.a. LEAST SQUARES - CONDITIONAL	48
3.b. MAXIMUM LIKELIHOOD	53
3.c. UNCONDITIONAL LEAST SQUARES	58

4. Diagnostic Techniques based on Residual Analysis

4.a. VISUAL INSPECTION	59
i. Residuals Over Time	59
ii. Residuals versus Predicted Values	60
iii. Histogram of Residuals	61
iv. Normal Probability Plots	61
4.b. TESTS OF HYPOTHESES	64
i. Correlation Between Residual and Normal Scores	64
ii. Runs Test	64
4.c. CORRELOGRAM OR AUTOCORRELATION FUNCTION	65
4.d. PORTMANTEAU LACK OF FIT TEST	68

5. Order Determination Procedures

5.a. SEQUENTIAL HYPOTHESIS TESTING	70
i. Trickle-down Selection	71
ii. Gush-up Selection - Pandit and Wu's Procedure	73
5.b. PATTERN RECOGNITION	78
i. Autocorrelation Function	78
ii. Partial Autocorrelation Function	81
iii. Inverse Autocorrelation Function	83
iv. S- or R- Arrays	86
v. Generalized Partial Autocorrelation	88
vi. Residual Variance Plots	91
5.c. PREDICTIVE POWER	92
i. Final Prediction Error	92
ii. Parzen's CAT	94
5.d. INFORMATION CRITERIA	96
i. Akaike's AIC	97
ii. Relationships among AIC, FPE and Residual Variance Plots	99
iii. Akaike's Bayesian Extension of AIC	101
iv. Schwarz's Bayesian Criterion, SBC	102
v. Hannan-Quinn Information Criterion, HQC	104

6. References

6.a. LITERATURE CITED	105
---------------------------------	-----

II. COMPARISON OF ORDER DETERMINATION CRITERIA

1. Introduction

1.a. PURPOSE	111
1.b. SOME CURRENT PROBLEMS	114
1.c. SOME CLASSICAL PROBLEMS	116
1.d. OBJECTIVE	118

2. Materials and Methods

2.a. APPROACH	120
2.b. ORDER DETERMINATION TECHNIQUES	121
i. Pattern Recognition	121
ii. Information Criteria	122
iii. Sequential Hypothesis Tests	123
2.c. STRATEGIES FOR CREATING A PORTFOLIO OF TENTATIVE MODELS	124
2.d. SIMULATION METHODOLOGY	127
2.e. DESIGN OF SIMULATION EXPERIMENTS	132
i. Sample Size, Variance, Estimation	132
ii. Theoretical Models	133
2.f. ASSUMPTIONS RESTATED	142

3. *Results and Discussion*

3.a. SIMULATIONS OF ORDER CRITERIA	144
i. Comparisons over Correct Model	145
ii. Criteria over Portfolio	148
iii. Estimation Methods	187
iv. Magnitude of Variation	191
v. Sample Size over Criteria	194
vi. Significance Levels	205
3.b. MODIFICATION OF CRITERIA	208
3.c. ASSESSMENT OF CLASSICAL EXAMPLES BY PROPOSED STRATEGY	218
i. Chemical Concentrations	219
ii. Wolfer's Sunspots	225
iii. Canadian Lynx Trappings	229

4. *Conclusions*

4.a. SUGGESTIONS FOR FURTHER RESEARCH	233
4.b. SUMMARY	234

5. *References*

5.a. LITERATURE CITED	238
---------------------------------	-----

6. APPENDIX	243
6.a. SEEDS FOR ARMA SIMULATIONS	243
6.b. TABLES OF SIMULATION RESULTS	244
<i>i. ARMA(1,1)</i>	<i>244</i>
<i>ii. ARMA(2,0)</i>	<i>276</i>
<i>iii. ARMA(2,1)</i>	<i>298</i>
<i>iv. ARMA(3,1)</i>	<i>320</i>
<i>v. ARMA(3,2)</i>	<i>342</i>
<i>vi. ARMA(4,0)</i>	<i>364</i>
<i>vii. ARMA(4,3)</i>	<i>386</i>
<i>viii. ARMA(6,0)</i>	<i>408</i>
<i>ix. ARMA(7,1)</i>	<i>430</i>
<i>x. ARMA(8,7)</i>	<i>452</i>
6.c. SAS PROGRAMS FOR ANALYSIS OF SIMULATIONS	474
<i>i. Criteria vs Correct Model</i>	<i>474</i>
<i>ii. Criteria vs Portfolio</i>	<i>475</i>
<i>iii. Criteria vs Size</i>	<i>477</i>
<i>iv. Distribution of Parameter Estimates . .</i>	<i>478</i>
<i>v. Box Plots</i>	<i>480</i>
<i>vi. Response Surfaces</i>	<i>482</i>
<i>vii. Generate and Plot ARIMA Process</i>	<i>484</i>
<i>viii. Sample Program for calling Macros . . .</i>	<i>488</i>
6.d. DATA AND PROGRAMS FOR APPLICATIONS . . .	489
<i>i. Chemical Concentrations</i>	<i>489</i>
<i>ii. Wolfer's Sunspots</i>	<i>491</i>
<i>iii. Canadian Lynx Trappings</i>	<i>494</i>

III. IMPLEMENTATION - Modeling Biological Processes of

Farm Animals

1. Introduction

1.a. BIORHYTHMS, THERMOREGULATION AND STATISTICAL MODELS	497
---	-----

2. Dynamical Systems Approach

2.a. MODELING TECHNIQUE	499
2.b. PROPOSED STRATEGY	500
2.c. EXAMPLE I: EGG PRODUCTION IN HENS	503
i. Sensible Heat Loss	505
ii. Sample: O% Productivity	506
2.d. EXAMPLE II: HEAT STRESS IN STEERS	508
i. Tympanic Temperature	508
ii. Sample: Constant Thermoneutral Regime	509

3. Results and Discussion

3.a. MODEL SELECTION for HENS	512
3.b. MODEL SELECTION for STEERS	520

4. Conclusions

4.a. SUMMARY	529
------------------------	-----

5. References

5.a. LITERATURE CITED	530
--	------------

6. Appendix

6.a. SAS PROGRAMS FOR MODEL SELECTION	534
6.b. SAS PROGRAMS FOR GRAPHICS	536
6.c. DATA FOR HENS	542
6.d. DATA FOR STEERS	546

PREVIEW

IV. SIMULATION - DEVELOPMENT OF PROGRAMS

1. Introduction

1.a. PURPOSE OF SIMULATION	547
--------------------------------------	-----

2. Materials and Methods

2.a. METHODOLOGY	547
2.b. SPECIFICATIONS	549

3. Programs

3.a. BATCH SIMULATION OF TIME SERIES	551
3.b. SHELL FOR SIMULATION OF THEORETICAL MODEL	556
3.c. REALIZATIONS FROM AN ARMA MODEL	558
3.d. PORTFOLIO FOR FIRST PASS	560
3.e. COLLECTING STATISTICS FOR EACH REALIZATION	563
3.f. ORDER DETERMINATION CRITERIA	566
3.g. TABULATE SIMULATIONS	571
3.h. AIDS FOR DEBUGGING ARIMABAT SAS PROGRAM .	573

4. Conclusions

4.a. SUMMARY	574
-------------------------------	------------

5. References

5.a. LITERATURE CITED	575
--	------------

6. Appendix

6.a. PROGRAM: COEFFICIENTS GIVEN ROOTS	577
6.b. PROGRAM: CHARACTERISTICS OF ROOTS GIVEN COEFFICIENTS	578
6.c. PROCEDURE to DOWNLOAD PERCENT SELECTED TABLES	583
6.d. PROCEDURE to EDIT PERCENT SELECTION TABLES	584
6.e. PROGRAM: BATCH SIMULATION	586

I. BACKGROUND

This section contains both a literature review and a review of the concepts in time series analysis that are relevant to the study of dynamic systems. Numerous journal articles are referenced as well as classical textbooks. The textbooks range from the elementary (Chatfield, 1989) to the intermediate (Box and Jenkins, 1976; Pandit and Wu, 1983) to the more advanced (Priestly, 1981).

1. Stochastic Processes

A stochastic process is a statistical phenomenon that evolves according to the laws of probability. The term *stochastic* is derived from the Greek word pertaining to chance or conjecture. Stochastic (or random) processes are useful for modeling time series. The phrase "*time series*" has become a generic term for records of events that fluctuate over time or space. Some examples of time series are: mechanical vibrations, the length of a queue, population growth, and air temperature.

Statistically, a stochastic process is a collection of random variables, ordered in time. Usually time is measured at discrete intervals, $t=0,1,2,\dots$ and only one observation is available for each point in time or space. Nevertheless, the observed time series is regarded as one sample from a collection of time series. The infinite population of a

collection of time series is called the *ensemble*, while a single time series is called a *realization*. Hence, time series analysis may be regarded as a way of evaluating parameters of the probability model which generated the observed time series.

1.a. AUTOCORRELATION AND AUTOCOVARIANCE

Autocorrelation measures the association between pairs of observations. In the conventional situation, the correlation between two variables is

$$\rho = \frac{COV(X, Y)}{\sqrt{var(X)} \sqrt{var(Y)}}$$

The association depends primarily on the covariance. The variances merely serve to standardize the result. The covariance is estimated by

$$COV(X, Y) = \frac{1}{N-1} \sum_{t=1}^N (X_t - \bar{X}) (Y_t - \bar{Y})$$

This differs from the present context where there is only one variable and we want to measure the association between observations separated by k units of time,

$$(X_1, X_{k+1}), (X_2, X_{k+2}), \dots, (X_{N-k}, X_N)$$

Here, there are only $N-k$ pairs instead of N pairs of observations. Furthermore, each random variable, X_t , is assumed to have the same mean, μ , so there is no point in

using separate estimates for the mean. As usual, the best estimate is \bar{X} which is based on all the observations. Thus, the estimated autocovariance at lag k is

$$\gamma_k^* = \frac{1}{N-k} \sum_{t=1}^{N-k} (X_t - \bar{X})(X_{t+k} - \bar{X}) \quad (1.a.1)$$

Although this estimate has a relatively small bias, most computer packages use an alternate estimate (Priestly, 1981, p.323)

$$\hat{\gamma}_k = \frac{1}{N} \sum_{t=1}^{N-k} (X_t - \bar{X})(X_{t+k} - \bar{X}) \quad (1.a.2)$$

The estimate, $\hat{\gamma}_k$, is preferred for a couple of reasons. One reason is the assertion (Parzen, 1961) that $\hat{\gamma}_k$ is usually more precise, i.e., has a smaller mean square error. Thus, when k becomes large relative to N , the increase in precision more than compensates for the larger bias. (When k is small relative to N , $\hat{\gamma}$ is still preferred, since there is little difference in precision between the two estimators.)

However, another reason $\hat{\gamma}$ is used is we are interested in the function γ_k for all values of k from 1 to $N-1$. This function is positive semi-definite, as is $\hat{\gamma}_k$. The function γ_k^* is not necessarily positive semi-definite. Positive semi-definiteness means the roots of the autocorrelation matrix are ≥ 0 . (Strictly positive definite implies the autocorrelations lie inside the interval $[-1,1]$). This has important implications for construction of spectra.

1.b. STATIONARITY

In time series analysis, it is usual to confine our attention to processes which are in statistical equilibrium. Such processes are called *stationary*, i.e., their statistical properties do not change over time. As time approaches infinity, the probability distribution of the X_t approximates a limit called the equilibrium distribution which does NOT depend on the initial conditions. Random processes which do not possess this property are called *nonstationary* or *evolutionary* processes, because the statistical properties evolve over time.

There are several ways to describe a stochastic process. One way is to specify the joint probability distribution. The behavior of a time series variable is determined by a probability distribution. The variables associated with different points in the time series are assumed to be statistically dependent. So for two time points, the joint behavior of X_t and X_s is given by the joint bivariate distribution. For k time points, the joint behavior of X_1, X_2, \dots, X_k is given by the multivariate joint probability distribution. This method of describing stochastic processes is complicated and not usually attempted in practice (Chatfield, 1989, p.28).

Another method is to specify the full infinite sequence of moments. This method is simpler and useful, especially