BEYOND REGRESSION:
NEW TOOLS FOR PREDICTION AND ANALYSIS
IN THE BEHAVIORAL SCIENCES

A thesis presented

by

Paul John Werbos

to

The Committee on Applied Mathematics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

statistics

Harvard University

Cambridge, Massachusetts

August, 1974

Copyright reserved by the author

PREFACE

The initial impetus for this thesis came from a suggestion by Prof. Karl Deutsch, my thesis supervisor, that I look more closely at the prediction of national assimilation and political mobilization, by use of the Deutsch-Solow model; his comments have been of major help to me with all the empirical work on nationalism, and in revising the original structures of Chapters (V) and (VI). The earlier work reported in section (ii) of Chapter (VI) was carried out under his supervision and support, through a research project funded by the Cambridge Project. I have been surprised more than once by the sensitivity and receptiveness of his intuition, in suggesting areas of research which looked unworkable at first but which led in the end to useful and surprising innovations. The opinions expressed in Chapter (V), however, remain my own responsibility, particularly in their more bull-headed aspects.

Professor Mosteller, in the Harvard Statistics Department, has helped by introducing me to current work on robust estimation, by suggesting the use of simulation tests, and by monitoring the general content of the first four chapters. Professors Anderson and Bossert, of the Committee on Applied Mathematics, and

Professor Dempster, of the Harvard Statistics Department, have helped me to find a more orderly way of presenting the mathematical ideas here, help which was sorely needed in the summer of 1973. To the Cambridge Project, and its sponsors at ARPA, must be given thanks, not only for all the computer time used in this research and in its documentation, but also for the opportunity to translate some of these ideas from theory into operational systems without the extensive delays more common in such research; without their support, this research could never have been brought to a state final enough to allow its continuation.

The personal and financial conditions which led to the completion of this thesis were rather complex, and debts are owed to more individuals than should properly be cited here. Certainly I have a strong debt to my parents in this respect; a debt to Richard Ney, whose ideas on the stock market financed a large part of my activity in this period; a moral debt to Prof. Nazli Choucri, of M.I.T., who, in a brief discussion in the spring of 1973, encouraged me to continue this work; a debt to a colleague, Gopal Krishna, for helping me see more clearly that the psychological hurdles I faced at first were not totally unique; a debt to Dr. Karreman, of the Bockus Research Institute, who, in one of those

1960's programs to encourage high school students, connected with the Moore School of Electronics at the University of Pennsylvania, got me started asking questions about the phenomenon of intelligence, questions which led me to the dynamic feedback concept, which was applied only later to formal statistics.

TABLE OF CONTENTS

CHAPTER OR SECTION	PAGE
PREFACE	(ii)
TABLE OF CONTENTS	(v)
LIST OF FIGURES	(viii)
LIST OF TABLES	(ix)
SYNOPSIS	(xii)
(I) GENERAL INTRODUCTION AND SUMMARY	I-1
(II) DYNAMIC FEEDBACK, STATISTICAL ESTIMATION AND SYSTEMS OPTIMIZATION: THE GENERAL TECHNIQUES	II-1
(i) INTRODUCTION	II-1
(ii) ORDINARY REGRESSION	II-4
(iii) NONLINEAR REGRESSION AND DYNAMIC FEEDBACK	II-15
(iv) MODELS WITH MEMORY	II-28
(v) NOISE, AND THE CONCEPT OF THE TRUTH OF A MODEL, IN STATISTICS	II-36
(vi) ORDINARY REGRESSION AND THE MAXIMUM LIKELIHOOD APPROACH	II-43
(vii) THE NEED FOR SOPHISTICATED NOISE MODELS	II-48
(viii) HOW TO ESTIMATE EXPLICIT SOPHISTICATED NOISE MODELS	II-60
(ix) PATTERN ANALYSIS	II-65
(x) OPTIMIZATION	II-72
(xi) THE METHOD OF "RELAXATION" WITH MEASUREMENT-NOISE-ONLY MODELS	II-78

CHAPTER OR SECTION PAGE

(xii) THE ORDERED DERIVATIVE AND DYNAMIC FEEDBACK..... II-82

APPENDIX: VARIATIONS ON THE STEEPEST ASCENT METHOD FOR EFFICIENT CONVERGENCE... II-89

FOOTNOTES TO CHAPTER (II)..... II-94

(III) THE MULTIVARIATE ARMA(1,1) MODEL: ITS SIGNIFICANCE AND ITS ESTIMATION..... III-1

(i) INTRODUCTION..... III-1

(ii) THE RECONSTRUCTION OF A WHITE NOISE MODEL FROM A VECTOR ARMA MODEL..... III-9

(iii) THE ESTIMATION OF MULTIVARIATE ARMA PROCESSES..... III-19

(iv) DESCRIPTION OF COMPUTER ROUTINE TO ESTIMATE ARMA PROCESSES..... III-32

APPENDIX: NUMERICAL EXAMPLES OF THE BEHAVIOR OF DIFFERENT CONVERGENCE PROCEDURES..... III-43

FOOTNOTES TO CHAPTER (III)..... III-49

(IV) SIMULATION STUDIES OF TECHNIQUES OF TIME-SERIES ANALYSIS..... IV-1

(i) INTRODUCTION..... IV-1

(ii) DEFINITION OF STUDIES CARRIED OUT..... IV-5

(iii) DESCRIPTION OF RESULTS..... IV-25

CHAPTER OR SECTION	PAGE
(V) GENERAL APPLICATIONS OF THESE IDEAS: PRACTICAL HAZARDS AND NEW POSSIBILITIES.....	V-1
(i) INTRODUCTION AND SUMMARY.....	V-1
(ii) THE LIABILITIES OF MATHEMATICAL METHODS IN PRACTICAL DECISION-MAKING.....	V-3
(iii) PREDICTION: A COMMON GOAL FOR VERBAL AND MATHEMATICAL SOCIAL SCIENCE.....	V-6
(iv) POSSIBILITIES FOR STATISTICS AS AN EMPIRICAL TOOL IN REAL-WORLD PREDICTION	V-23
(v) BEYOND NAIVE EMPIRICISM: ADAPTING OUR IDEAS TO FILL THE GAP LEFT BY STATISTICS	V-38
FOOTNOTES TO CHAPTER (v).....	V-49
(VI) NATIONALISM AND SOCIAL COMMUNICATIONS: A TEST CASE FOR MATHEMATICAL APPROACHES....	VI-1
(i) INTRODUCTION AND SUMMARY.....	VI-1
(ii) INITIAL STUDIES OF THE DEUTSCH-SOLOW MODEL.....	VI-14
(iii) LATER STUDIES OF THE DEUTSCH-SOLOW MODEL.....	VI-49
(iv) NATIONALISM, CONFORMITY AND COMMUNICATIONS TERMS: AN EXTENSION OF THE DEUTSCH MODEL.....	VI-85
Local vs. Regional Language-Dominance as a Dynamic Factor.....	VI-90
Some Operational Dimensions of Nationalism: Narcissism, Stereotyping and Aggression.....	VI-94
Toward More General Models.....	VI-98
(v) ASSIMILATION AND COMMUNICATION: THE CASE OF NORWAY.....	VI-107
FOOTNOTES TO CHAPTER (VI).....	VI-133

LIST OF FIGURES

Figure Number	Title	Page
V-1:	Pathways of Correlations With Noisy Data...	V-18
	Legend to Figures VI-1 Through VI-4.....	VI-3
VI-1:	RMS Average Errors In Long-Term Predictions of Assimilated Populations, In Percentages	VI-4
VI-2:	RMS Average Errors In Long-Term Predictions of Differentiated Populations, In Percentages.....	VI-5
VI-3:	RMS Average Errors In Long-Term Predictions of Mobilized Populations, In Percentages..	VI-6
VI-4:	RMS Average Errors In Long-Term Predictions of Underlying Populations, In Percentages.	VI-7

LIST OF TABLES

Table Number	Title	Page
II-1:	Table of Operations for Equation (2.11)..	II-21
II-2:	Table of Operations for Equations (2.14).....	II-31, -32
II-3:	Hypothetical Example of "Ideal Types"....	II-57
II-4:	Table of Operations for Equations (2.17).....	II-61 to 63
II-5:	Table of Operations for Equations (2.18).....	II-74, -75
III-1:	Example of Convergence Results With Early Version of the ARMA Estimation Routine.	III-44
III-2:	Sample Session With Convergence Results For Final Version of the ARMA Estimation Routine.....	III-45 to 47
IV-1:	Average coefficient estimates, and dispersion errors of estimates, for the six estimation routines and twelve simulated processes defined in section (ii).....	IV-34
IV-2:	Prediction Errors as Defined in Section (ii).....	IV-35 to 40
IV-3:	Estimates of Growth Factor, "c"....	IV-41 to 46
IV-4:	Errors in Prediction & Miscellany..	IV-47 to 70
VI-1:	Sample of Results from DELTA.....	VI-15
VI-2:	Regression Statistics for Mobilized and Underlying Populations.....	VI-16
VI-3:	Regression Statistics for Assimilated and Differentiated Populations.....	VI-17

Table Number	Title	Page
VI-4:	Long-Term Prediction Errors with SERIES in Predicting Mobilization.....	VI-18
VI-5:	Long-Term Predictions Errors with SERIES in Predicting the Underlying Population..	VI-19
VI-6:	Long-Term Prediction Errors with SERIES in Predicting the Assimilated Population.	VI-20
VI-7:	Long-Term Prediction Errors with SERIES in Predicting the Differentiated Population.....	VI-21
VI-8:	RMS Average Errors of Predictions of Mobilized and Underlying Populations, by EXTRAP.....	VI-22
VI-9:	RMS Average Errors of Predictions of Assimilated and Differentiated Populations, by EXTRAP.....	VI-23
VI-10:	Indices of "M"(Mobilization) and of "A" (Assimilation) From the Deutsch-Kravitz Data Used For Runs Reported in Tables VI-1 to VI-9.....	VI-24
VI-11:	Statistics Concerning Regression for Mobilized and Underlying Populations....	VI-50
VI-12:	Statistics Concerning Regression for Assimilated and Differentiated Populations.....	VI-51
VI-13:	ARMA Models for Mobilization Processes..	VI-52
VI-14:	ARMA Models of the Assimilation Process.	VI-53
VI-15:	RMS Averages of Percentage Errors With Long-Term Predictions of Mobilized and Underlying Populations, Based on Regression.....	VI-54
VI-16:	RMS Averages of Percentage Errors With Long-Term Predictions of Mobilized and Underlying Populations, Based on the ARMA Method.....	VI-55