A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models

A. Ronald Gallant
and Halbert White

Basil Blackwell

First published 1988

Basil Blackwell Ltd.
108 Cowley Road, Oxford, OX4 1JF, UK

Basil Blackwell Inc.
432 Park Avenue South, Suite 1503
New York, NY 10016, USA

British Library Cataloguing in Publication Data

Gallant, A Ronald
A unified theory of estimation and inference
for nonlinear dynamic models.
1. Econometric models 2. Nonlinear theories
I. Title II. White, Halbert
330'.028 HB141

ISBN 0-631-15765-4

Library of Congress Cataloging-in-Publication Data

t, A. Ronald, 1942—
A unified theory of estimation and inference for nonlinear dynamic models.

Includes index.
1. Econometric models. 2. Linear
I. White, Halbert. II. Title.
HB141.G33 1988 330'.028 87-25602
ISBN 0-631-15765-4

Typeset in 11 on 13pt Monotype Times Maths
by Advanced Filmsetters (Glasgow) Ltd
Printed in Great Britain by TJ Press, Padstow

Contents

Chung-Ming Kuan of UCSD, Whitney Newey of Princeton, Peter Phillips of Yale, and B. M. Pötscher of the Institut für Ökonometrie, Vienna. Particular thanks are due Angelo Melino of the University of Toronto, who provided detailed comments on the manuscript, and to Donald W. K. Andrews of Yale, whose insightful comments and suggestions immeasurably improved the technical standard of this work. We are also grateful for the excellent comments and helpful suggestions of three anonymous reviewers. Although the authors would like to blame these people and many others besides for any errors and deficiencies in the current work, this would not be noble. Instead, each author blames the other.

Finally, the authors owe a very substantial debt to Leigh Henry who expertly, meticulously and cheerfully prepared the final version of the manuscript for publication, and to Annetta Whiteman who flawlessly typed an earlier version of the manuscript.

A. Ronald Gallant
Raleigh, NC

Halbert White
La Jolla, CA

1 Introduction

It has been over forty years since economists and econometricians first began coming to grips with the fact that economic processes are dynamic and simultaneous in nature. The pioneering work of Mann and Wald (1943), Koopmans (1945), and Marschak (1947) was fundamental in establishing approaches to modeling and estimating linear simultaneous systems of dynamic equations. This work is still the primary foundation for much of the research done in economics and econometrics today.

Subsequently, it was realized that although linear models were conceptually convenient and analytically tractable, they did not always provide an adequate framework for modeling economic behavior. Nonlinear models could provide greater flexibility, and techniques for estimating the parameters of nonlinear models with exogenous explanatory variables and independent identically distributed (i.i.d.) errors, developed in the important work of Jennrich (1969) and Malinvaud (1970), were extended to nonlinear dynamic models by Hannan (1971) for single equations and by Robinson (1972) for systems of equations.

Next, it was realized that economically plausible nonlinear simultaneous systems of equations did not necessarily have a convenient (or even analytically derivable) reduced form. This led investigators to study methods of estimation and inference for implicit, nonlinear simultaneous systems of equations. Pioneers in this work were Amemiya (1977) and Gallant (1977) and Gallant and Jorgenson (1979). A broad unification was achieved by Burguete, Gallant, and Souza (1982). This work does not allow for dynamics. More recently, treatments of implicit nonlinear simultaneous systems allowing for fairly rich dynamics have been given by Hansen (1982) and Gourieroux, Monfort, and Trognon (1985).

In the work just cited, it is typically assumed either that the explanatory variables are fixed (i.e. nonstochastic) as in an experiment, and that the innovations to the system are independent and identically distributed; or that the explanatory variables and errors are stationary and ergodic. Such assumptions are by no means necessary, as suggested by the results of Robinson (1978) for linear and multilinear processes with heteroskedastic martingale difference innovations. Nor do such assumptions necessarily provide a realistic description of economic time series, as pointed out long ago by Koopmans (1937) and as argued more recently by Hendry and Richard (1983) and White and Domowitz (1984). A more satisfactory description of economic time series is that they are dependent and heterogeneous.

Thus, a general theory applicable to the study of economic phenomena should be able to treat nonlinear dynamic simultaneous systems of implicit equations with errors and explanatory variables which are dependent and heterogeneously distributed. In addition, it is desirable to have a theory which permits treatment of the wide variety of estimators now applied in modern econometrics: maximum likelihood, instrumental variables, method of moments, and *m*-estimation techniques, for example. In his general and very elegant monograph, Bierens (1981) builds on numerous contributions to the mathematical statistics literature to produce a theory of econometric estimation which very nearly satisfies this description. Bierens assumes that the errors of the structural equations are i.i.d., but his treatment otherwise contains all the features just mentioned.

As flexible as nonlinear dynamic systems of equations may be, there is no guarantee that the equations specified by the researcher provide a correct description of economic reality. They may be misspecified in any number of ways for any number of reasons. Thus, it is important to have available a theory which explicitly recognizes the potential for misspecification and which provides a satisfying theory of estimation and inference despite the presence of misspecification. The problem of misspecification in the context of maximum likelihood was considered by Silvey (1959), and was given a thorough and elegant treatment in the i.i.d. case by Huber (1967). A Bayesian treatment is also available in the work of Berk (1966; 1970). The problem of estimation of systems of misspecified dynamic equations by any of a variety of "prediction error" methods (including maximum likelihood) is treated in the series of papers by Ljung (1976; 1978), Caines and Ljung (1976), and Ljung and Caines (1978) and is given a comprehensive treatment in Ljung's (1987) recent book on system identification.

In this work, we use results of Domowitz and White (1982) and an estimation framework proposed by Bates and White (1985) to provide a theory of estimation and inference which applies to a fairly broad class of estimators (including maximum likelihood, method of moments, and *m*-estimators) for possibly misspecified models. These models may be nonlinear dynamic systems of implicit simultaneous equations (or may also be censored, truncated, or limited dependent variables models) with explanatory variables, dependent variables, and/or errors which may be dependent and heterogeneously distributed. Our results thus extend those of Ljung and his collaborators by allowing for implicit equations and estimation techniques not necessarily based on prediction error methods. We also provide somewhat weaker dependence conditions. The work of Bates and White (1985) only considers the issue of consistency. Here we treat asymptotic distribution, covariance matrix estimation, and issues of inference as well. Furthermore, we relax the restrictive assumption made by Bates and White (1985) that the functions defining the estimator depend on only a finite number of recent lagged values of the dependent and explanatory variables of the model. The results of Domowitz and White (1982), White and Domowitz (1984), and Bates and White (1985) contain a number of other restrictive assumptions and occasional errors that are eliminated here.

Although explicitly formulated for time series, our results also apply to experimental, cross-section, and panel data. The results given here therefore represent an extension of the unification achieved by Burguete, Gallant, and Souza (1982) to the case of dynamic models. Despite the apparent generality of our results, they still contain a number of somewhat restrictive assumptions which prevent their applicability in particular cases of interest. These assumptions and their implications will be discussed as we progress.

It should be evident from the foregoing discussion that our results rely heavily on the work of many who have gone before. In many cases, the work of those already cited is similarly dependent. Thus, our results should clearly be viewed as an elaboration of the fundamental, classical, and ingenious work of Doob (1934; 1953), Wald (1949), and Le Cam (1953). We also wish to acknowledge our heavy reliance on the work of Billingsley (1968) and McLeish (1975a; 1975b), whose general and elegant results allow us to adopt the relatively general stochastic framework used throughout.

The plan of this monograph is as follows. In chapter 2 we discuss the underlying data generation process and establish the existence of the estimators of interest, members of a particular class of extremum estimators. In chapter 3 we establish the consistency of these estimators under general conditions. Chapter 4 contains useful results of a technical nature concerning the property of near epoch dependence, which plays a crucial role in our analysis. Chapter 5 considers the asymptotic normality of our class of estimators under general conditions, and chapter 6 presents consistent estimators for the asymptotic covariance matrix. In chapter 7 we present a unified theory of inference under null and locally alternative hypotheses. Chapter 8 contains concluding remarks and a discussion of directions for further research. For convenience, all of the various assumptions made at different points are collected together and presented in an appendix to the book.

REFERENCES

Amemiya, T. 1977: The maximum likelihood and the nonlinear three-stage least squares estimator in the general nonlinear simultaneous equation model, *Econometrica* 45, 955–68.

Bates, C. and H. White 1985: A unified theory of consistent estimation for parametric models, *Econometric Theory* 1, 151–78.

Berk, R. H. 1966: Limiting behavior of posterior distributions when the model is incorrect, *Annals of Mathematical Statistics* 37, 51–8.

Berk, R. H. 1970: Consistency *a posteriori*, *Annals of Mathematical Statistics* 41, 894–906.

Bierens, H. 1981: *Robust Methods and Asymptotic Theory in Econometrics.* New York: Springer-Verlag.

Billingsley, P. 1968: *Convergence of Probability Measures.* New York: John Wiley and Sons.

Burguete, J., A. R. Gallant, and G. Souza 1982: On the unification of the asymptotic theory of nonlinear econometric models, *Econometric Reviews* 1, 151–212.

Caines, P. E. and L. Ljung 1976: Asymptotic normality and accuracy of prediction error estimates, University of Toronto, Department of Electrical Engineering research report 7602.

Domowitz, I. and H. White 1982: Misspecified models with dependent observations, *Journal of Econometrics* 20, 35–58.

Doob, J. L. 1934: Probability and statistics, *Transactions of the American Mathematical Society* 36, 759–75.

Doob, J. L. 1953: *Stochastic Processes.* New York: John Wiley and Sons.

Gallant, A. R. 1977: Three-stage least-squares estimation for a system of simultaneous, nonlinear, implicit equations, *Journal of Econometrics* 5, 71–88.

Gallant, A. R. and D. W. Jorgenson 1979: Statistical inference for a system of simultaneous, nonlinear, implicit equations in the context of instrumental variable estimation, *Journal of Econometrics* 11, 275–302.

Gourieroux, C., A. Monfort, and A. Trognon 1985: A general approach to serial correlation, *Econometric Theory* 1, 315–40.

Hannan, E. J. 1971: Non-linear time series regression, *Journal of Applied Probability* 8, 767–80.

Hansen, L. P. 1982: Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029–54.

Hendry, D. and J.-F. Richard 1983: The econometric analysis of economic time series, *International Statistical Review* 51, 111–48.

Huber, P. 1967: The behavior of maximum likelihood estimates under nonstandard conditions, in *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability*, vol. 1, Berkeley: University of California Press, 221–33.

Jennrich, R. 1969: Asymptotic properties of nonlinear least squares estimators, *Annals of Mathematical Statistics* 40, 633–43.

Koopmans, T. 1937: *Linear Regression Analysis of Economic Time Series.* Haarlem: De Erven F. Bohn N.V.

Koopmans, T. 1945: Statistical estimation of simultaneous economic relationships, *Journal of the American Statistical Association* 40, 448–66.

Le Cam, L. 1953: On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates, *University of California Publications in Statistics* 2, 23–53.

Ljung, L. 1976: On consistency and identifiability, in *Mathematical Programming Study* no. 5, New York: North Holland, 169–90.

Ljung, L. 1978: Convergence analysis of parametric identification methods, *IEEE Transactions on Automatic Control* AC-23, 770–83.

Ljung, L. 1987: *System Identification: theory for the user.* Englewood Cliffs: Prentice-Hall.

Ljung, L. and P. E. Caines, 1978: Asymptotic normality of prediction error estimators for approximate system models, Linköping University, Department of Electrical Engineering report.

Malinvaud, E. 1970: *Statistical Methods of Econometrics.* Amsterdam: North Holland.

Mann, H. B. and A. Wald 1943: On the statistical treatment of linear stochastic difference equations, *Econometrica* 11, 173–220.

Marschak, J. 1947: Econometric structure, path policy, and prediction, *American Economic Review* 37, 81–4.

McLeish, D. L. 1975a: A maximal inequality and dependent strong laws, *Annals of Probability* 3, 826–36.

McLeish, D. L. 1975b: Invariance principles for dependent variables, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandete Gebiete* 32, 165–78.

Robinson, P. M. 1972: Non-linear regression for multiple time-series, *Journal of Applied Probability* 9, 758–58.

Robinson, P. M. 1978: On consistency in time series analysis, *Annals of Statistics* 6, 215–23.

Silvey, S. D. 1959: The Lagrangian multiplier test, *Annals of Mathematical Statistics* 30, 389–407.

Wald, A. 1949: Note on the consistency of the maximum likelihood estimate, *Annals of Mathematical Statistics* 20, 595–603.

White, H. and I. Domowitz 1984: Nonlinear regression with dependent observations, *Econometrica* 52, 143–62.

---

2 The Data Generation Process and Optimization Estimators

We suppose that we are interested in analyzing a body of data generated according to the following assumption.

Assumption DG (data generation)

Let (Ω, F, P) be a complete probability space. The observed data are generated as a realization

$$x_t = X_t(\omega) = W_t(\ldots, V_{t-1}(\omega), V_t(\omega), V_{t+1}(\omega), \ldots) \quad \omega \in \Omega$$

of a stochastic process $X_t : \Omega \to \mathbb{R}^{w_t}$, $w_t \in \mathbb{N} \equiv \{1, 2, \ldots\}$, where $V_t : \Omega \to \mathbb{R}^v$, $v \in \mathbb{N}$, and $W_t : \times_{t=-\infty}^{\infty} \mathbb{R}^v \to \mathbb{R}^{w_t}$ are such that X_t is measurable-$F/B(\mathbb{R}^{w_t})$, $t = 0, \pm 1, \pm 2, \ldots$. □

In what follows, any reference to Ω, F, or P will be understood as pertaining to the underlying complete probability space of this definition. The notation $B(\cdot)$ denotes the Borel σ-field generated by the open sets of the indicated set.

The data we analyze are viewed as arising from some transformation W_t of an underlying process V_t. Some or all of the elements of V_t may be unobserved; typically, V_t will consist of unobserved shocks to an economic data generating process. It may (but need not) also include nonendogenous explanatory variables and/or instrumental variable candidates. Observed elements of V_t can also be elements of X_t, so that the corresponding element of W_t is simply an appropriate projection mapping. Note that the dimension of the function W_t may itself depend on t. By allowing this dependence, it is possible to treat situations in which, as t grows, X_t contains a growing number of lagged (or future) values of some underlying process (such as V_t). For simplicity, our examples below will not exploit this possibility; we shall choose $w_t = w$