

INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

- 1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.**
- 2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.**
- 3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.**
- 4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.**
- 5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.**

University Microfilms International

300 North Zeeb Road

Ann Arbor, Michigan 48106 USA

St. John's Road, Tyler's Green

High Wycombe, Bucks, England HP10 8HR

78-4797

RADECKI, Lawrence John, 1950-
SIMULTANEOUS EQUATION MODELS WITH LAGGED
ENDOGENOUS VARIABLES AND AUTOREGRESSIVE
DISTURBANCES.

The University of Michigan, Ph.D., 1977
Economics, general

University Microfilms International, Ann Arbor, Michigan 48106

SIMULTANEOUS EQUATION MODELS WITH LAGGED ENDOGENOUS VARIABLES
AND AUTOREGRESSIVE DISTURBANCES

by
Lawrence John Radecki

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in The University of Michigan
1977

Doctoral Committee:

Professor Jan Kmenta, Chairman
Professor Robert S. Holbrook
Professor E. Philip Howrey
Associate Professor Edward D. Rothman

ACKNOWLEDGEMENTS

I wish to thank the members of my committee and especially my chairman for their assistance and patience in the writing of this dissertation.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	ii
LIST OF TABLES	iv
LIST OF APPENDICES	vi
INTRODUCTION	1
CHAPTER	
I AUTOREGRESSIVE DISTURBANCES	5
II ESTIMATION METHODS	22
III TEIGEN'S MODEL	48
IV PERRY'S MODEL	66
V SAMPLING EXPERIMENTS	82
VI CONCLUSION	131
APPENDICES	137
BIBLIOGRAPHY	142

LIST OF TABLES

Table	Page
1 Estimates of Structural Coefficients	56
2 Demand Elasticities	60
3 Supply Elasticities	60
4 Single Equation and Structural Estimates of Steady-State Elasticities	62
5 Predicted Values of the Endogenous Variables	64
6 Evaluation of the Predictions	64
7 Estimates of Structural Coefficients	73
8 Phillips Curves	77
9 Simulated Values for $w(t)$	79
10 Evaluation of the Simulations	80
11 Mean Estimates of the Coefficients, Strong Simple Autoregression $\rho=.8$	110
12 Frequency of Estimates Around True Parameter Values, Strong Simple Autoregression	112
13 Histogram of β_1 Estimates (Method D, Sample Size 20, True Value ¹ of $\beta_1=2$)	95
14 Histogram of β_1 Estimates (Method D, Sample Size 100, True Value ¹ of $\beta_1=2$)	96
15 Frequency of Incorrect Confidence Interval (95% Level of Confidence), Strong Simple Autoregression	97
16 Frequency of Failure to Reject Null Hypothesis, Strong Simple Autoregression	115
17 Mean Estimates of the Coefficients, Moderate Simple Autoregression $\rho=.4$	116

Table		Page
18	Frequency of Estimates in Intervals Around True Parameter Values, Moderate Simple Autoregression	118
19	Frequency of Incorrect Confidence Interval (95% Level of Confidence), Moderate Simple Autoregression	101
20	Frequency of Failure to Reject Null Hypothesis, Moderate Simple Autoregression	121
21	Mean Estimates of the Coefficients, Vector Autoregression	122
22	Direction of Mean Bias	104
23	Frequency of Estimates in Intervals Around True Parameter Values, Vector Autoregression	124
24	Frequency of Incorrect Confidence Interval (95% Level of Confidence), Vector Autoregression	105
25	Frequency of Failure to Reject Null Hypothesis, Vector Autoregression	128

LIST OF APPENDICES

	Page
Appendix A	137
Appendix B	140

INTRODUCTION

The purpose of this dissertation is to investigate the estimation methods designed for simultaneous equation systems with lagged endogenous variables and autoregressive disturbances. This study is a worthwhile pursuit because in their work researchers often encounter an econometric model with these features. Economic models frequently involve a set of relationships that simultaneously explain the behavior of a number of variables. For example, supply and demand models of a commodity market and macro models of the economy easily fit into the framework of a simultaneous equation system. Furthermore, these models are often specified to contain lagged endogenous variables because the system's structure is dynamic. A dynamic structure is one in which a substantial amount of time may elapse between a change in the value of an explanatory variable and its effect on the value of the dependent variable. This delayed reaction is made explicit by the inclusion of lagged variables in the model. Rationalizations for the delayed reaction include adaptive expectations, partial stock adjustments, and institutional arrangements. Finally, in the case of models estimated from time-series data, the standard assumption of nonautocorrelated disturbances is frequently violated. This means that the effects on the model that are summarized by the disturbances carry over from one period into following periods. The assumption of autoregressive disturbances typically replaces the basic assumption of nonautocorrelated

disturbances. These three features of many econometric models - simultaneity, lagged endogenous variables, and autoregressive disturbances - pose a common, although interesting estimation problem on which this dissertation will focus.

The investigation of this compound econometric problem will begin by discussing the specification of autoregressive disturbances in the context of a system of equations. If autocorrelated disturbances are believed to be present, then the approach usually taken is to specify a first-order simple autoregressive scheme for each equation in the system. This is the autoregressive scheme introduced for single-equation models in all econometric textbooks and stressed in the classroom. A new approach, beginning to appear more frequently in journal articles, is to specify a first-order vector autoregressive scheme. The econometric justifications and logical appeal of both autoregressive schemes will be considered. A survey will then be presented of the estimation methods developed to handle properly the econometric problem of simultaneity, lagged endogenous variables, and autoregressive disturbances. These methods may give more satisfactory results than methods ignoring autoregression because in this context estimates generated by a method developed under the assumption of nonautocorrelated disturbances lack all desirable properties. A few tests of hypotheses concerning the specification of the regression disturbances are also included. Some of the estimators and tests included in the survey will be used in the work that follows.

Next, the following two previously estimated models will be reestimated:

1. Ronald L. Teigen, "Demand and Supply Functions for Money: Another Look at Theory and Measurement," Econometrica 44 (1976): 377-85. In this paper the author presents a small model of the monetary sector, including equations for the demand for money, the supply of money, and a reduced form equation for income.
2. George L. Perry, Unemployment, Money Wage Rates, and Inflation (Cambridge, Massachusetts: The M.I.T. Press, 1966). Chapter 5 contains a four equation model of the wage-price subsystem of the economy that was designed to study the process of inflation; the model consists of equations explaining profit rates, manufacturing output prices, and wages.

These models were originally estimated under the hypothesis of non-autocorrelated disturbances even though it was likely that this assumption was invalid. The reestimation of these models will be undertaken to determine what differences, if any, in the empirical results will occur if the models are estimated under the assumption of first-order autoregressive disturbances. If autoregressive disturbances are detected, then the new estimates will be used to retest the hypotheses to which each model was directed. This reestimation work will serve to confirm or dispute the original conclusions of these studies.

Following the reestimation of the models, sampling experiments will be conducted in order to learn more about the relative performance of various estimation methods when the disturbances of a

simultaneous equation model having lagged endogenous variables follow a first-order autoregressive scheme. It is expected that these experiments will show:

1. the sensitivity of the coefficient estimates to the specification of the autoregressive scheme. A high sensitivity would indicate that researchers should be extremely careful when they consider the nonautocorrelated, simple autoregressive, and vector autoregressive specifications for their models' disturbances.
2. the gain in estimation accuracy that results from assuming more complicated autoregressive schemes for regression disturbances. Information of this sort would assist researchers in allocating their limited resources between the estimation of their models and the other aspects of their work.
3. the unreliability of hypothesis tests based on the coefficient estimates, their estimated standard errors, and the t distribution. This work should indicate if the likelihood of making correct inferences from a particular estimate corresponds to the stated level of significance.

The ability of some tests to detect the presence of autoregression will be investigated as a supplement to the experiments dealing with the estimation methods. It is hoped that this work represents a comprehensive study on the estimation of simultaneous equation models with lagged endogenous variables and autoregressive disturbances.

CHAPTER I

AUTOREGRESSIVE DISTURBANCES

The linear regression model is usually formulated as

$$y(t) = \beta_0 + \beta_1 x_1(t) + \beta_2 x_2(t) + \dots + \beta_K x_K(t) + u(t)$$

where $u(t)$ is the random disturbance of the regression equation and $E[u(t)] = 0$. The regression disturbance makes stochastic the economic relationship between the dependent variable and its explanatory variables. The following reasons justify the presence of the disturbance term and the stochastic nature of the model:¹

1. The data for the model's variables contain measurement errors due to imperfect methods and inevitable mistakes in the collecting and processing of economic data. In applied econometric work, researchers usually discount this rationale for the disturbance term by assuming that the measurement errors contribute so little to the total regression disturbance that it can be safely disregarded.
2. Human behavior is inherently erratic; an individual may react differently to identical situations. Even though the explanatory variables (prices, income, family size) may take on the same set of values during two or more observations in the sample for a household's demand equation, the dependent variable (purchases of a product) takes on different values because the rational consumer may act on a whim.

3. The model omits relevant explanatory factors whose effects are lumped into the regression disturbance. Relevant explanatory factors are omitted from regression models because:
 - a. individually some factors have so little influence on the dependent variable that their effect cannot be measured accurately; nevertheless, these same factors can collectively have a significant influence on the dependent variable;
 - b. certain psychological and qualitative factors (for example, tastes and expectations) can neither be measured statistically nor incorporated satisfactorily into the model through binary variables; and
 - c. random factors (such as acts of God) have unique and unpredictable effects on the model.

The disturbance term in a regression equation, therefore, can be viewed as representing the pooled effect of lumping together measurement errors, the randomness of human behavior, and omitted factors. Among the basic assumptions that researchers frequently make concerning the regression disturbances is the absence of autocorrelation. That is, the regression disturbance's value for a given observation has no linear effect on its value for any other observation in the sample; thus, $u(s)$ and $u(t)$ are uncorrelated, or equivalently, $E[u(s) \cdot u(t)] = 0$, $(s \neq t)$. When cross-sectional data comprise the sample, a strong argument can often be made to defend the assumption of nonautocorrelated disturbances. If the observations are taken over time, however, it is more reasonable to believe that, for a given observation in the sample, the effect on the model

represented by the regression disturbance carries over into subsequent periods, so that $u(s)$ and $u(t)$ are correlated; $E[u(s) \cdot u(t)] \neq 0$, ($s \neq t$). The following reasons account for this correlation:²

1. for two or more successive periods, the same error may be made in the measurement of an economic variable;
2. since most economic time-series data are highly autocorrelated, the collective effect of the omitted, individually insignificant variables will be autocorrelated;
3. random factors and factors impossible to quantify exert influence that spans more than one observation in the sample; and,
4. the model is misspecified. (If this is the source of the autocorrelation, then the researcher should correct the specification of the model rather than change the assumptions concerning the disturbance term.)

When the researcher incorrectly assumes nonautocorrelated disturbances, the mistake may have serious and undesirable effects on estimation, hypothesis testing, and forecasting with the model. If the autocorrelation scheme for the regression disturbances is correctly specified, however, estimation methods that account for the autocorrelated nature of the disturbances can then be used. These methods restore at least the desirable asymptotic properties for estimation and forecasting and permit asymptotically valid hypothesis testing.³

The most popular specification for an autocorrelated disturbance is the first-order simple autoregressive scheme:

$$u(t) = \rho u(t-1) + \varepsilon(t)$$

where ρ is the autoregression coefficient, $|\rho| < 1$, and $\varepsilon(t)$ is a disturbance that fulfills all the basic assumptions of the linear regression model and is assumed to be independent of $u(t-1)$. This autoregressive scheme interprets the current regression disturbance, $u(t)$, as the combined result of the preceding regression disturbance, $u(t-1)$, and a new random effect, $\varepsilon(t)$. Furthermore, it implies that the disturbance for a particular observation has a diminishing effect on future disturbances as the number of periods between the two observations increases. In contrast, $|\rho| > 1$ is implausible because that means that the effect of a particular disturbance would "intensify" as the number of periods separating the disturbances increases and that the variance of the disturbance term $u(t)$ would be infinite.

If a system of G simultaneous equations (or a group of G seemingly unrelated regression equations) is to be estimated, then each equation in the system can follow its own first-order simple autoregressive scheme:

$$u_g(t) = \rho_g u_g(t-1) + \varepsilon_g(t)$$

for $g=1, \dots, G$. The first-order simple autoregressive scheme for the entire system can be represented in matrix form as

$$\begin{bmatrix} u_1(t) \\ u_2(t) \\ \cdot \\ \cdot \\ \cdot \\ u_G(t) \end{bmatrix} = \begin{bmatrix} \rho_1 & 0 & \dots & 0 \\ 0 & \rho_2 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & \rho_G \end{bmatrix} \begin{bmatrix} u_1(t-1) \\ u_2(t-1) \\ \cdot \\ \cdot \\ \cdot \\ u_G(t-1) \end{bmatrix} + \begin{bmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_G(t) \end{bmatrix}$$

or more compactly as

$$u(t) = Ru(t-1) + \varepsilon(t)$$

where R is the diagonal matrix of autoregression parameters. Furthermore, while working in the context of a system of simultaneous regression equations, the researcher should consider the possibility of contemporaneous correlation across equations of the nonautocorrelated disturbance, $\varepsilon_g(t)$, in each equation's autoregressive scheme. This type of correlation among the regression disturbances is the motivation for grouping together seemingly unrelated regression equations. For example, the consumer demands for products are likely to share determining factors besides those that the researcher chose to include in the regression models. Hence, the regression disturbances of the demand equations are likely to be contemporaneously correlated. The researcher can then take this information into account, and thereby improve the estimation of the regression equations. A similar correlation is likely to be present among the disturbances of a simultaneous equation system whose structural equations are obviously related. For example, consider a small macro model of the United States' economy that includes structural equations for consumption, investment, and the money market, and a GNP identity. It seems reasonable to believe that common factors outside the model simultaneously influence households' demand for goods and services and their demand for money or simultaneously influence households' consumption spending and firms' and households' investment spending. Hence the disturbances of the system's structural equations are likely to be contemporaneously correlated. This type

of correlation among the disturbances is an ancillary feature of simultaneous equation models. Researchers sometimes assume that it is absent or simply choose to disregard it, especially if there are doubts concerning the model's specification.

If contemporaneous correlations exist among the nonautocorrelated disturbances, then $E[\epsilon_g(t) \cdot \epsilon_h(t)] \neq 0$ for $g \neq h$. Expressed in matrix form, this means that

$$E \begin{bmatrix} \epsilon_1(t) \\ \epsilon_2(t) \\ \vdots \\ \epsilon_G(t) \end{bmatrix} [\epsilon_1(t) \ \epsilon_2(t) \ \dots \ \epsilon_G(t)] = E[\epsilon(t)\epsilon(t)'] = \Phi$$

where Φ is a symmetric, positive definite matrix. Only if these contemporaneous correlations are absent is $E[\epsilon_g(t) \cdot \epsilon_h(t)] = 0$, for $g \neq h$, and Φ a diagonal matrix. Unless there is some reason to believe that the nonautocorrelated disturbances are contemporaneously uncorrelated across equations, then, the matrix should not be limited to being diagonal.

Vector Autoregression

Some econometricians have suggested that the first-order simple autoregressive scheme should be altered for simultaneous equation systems and groups of seemingly unrelated regression equations.⁴ They feel that the autoregressive scheme should be generalized by lifting the diagonality restriction on the R matrix. This means that the current value of the regression disturbance for a particular equation depends not only on its own past values but also on the past

values of the regression disturbances in other equations. This autoregressive scheme for the regression disturbances, called first-order vector autoregression, can be represented as

$$u_g(t) = \rho_{g1}u_1(t-1) + \rho_{g2}u_2(t-1) + \dots + \rho_{gg}u_g(t-1) + \dots \\ + \rho_{gG}u_G(t-1) + \varepsilon_g(t)$$

where the ρ_{gh} 's are autoregression parameters; or, in vector form

$$\begin{bmatrix} u_1(t) \\ u_2(t) \\ \vdots \\ u_G(t) \end{bmatrix} = \begin{bmatrix} \rho_{11} & \rho_{12} & \dots & \rho_{1G} \\ \rho_{21} & \rho_{22} & \dots & \rho_{2G} \\ \vdots & \vdots & & \vdots \\ \rho_{G1} & \rho_{G2} & \dots & \rho_{GG} \end{bmatrix} \begin{bmatrix} u_1(t-1) \\ u_2(t-1) \\ \vdots \\ u_G(t-1) \end{bmatrix} + \begin{bmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \\ \vdots \\ \varepsilon_G(t) \end{bmatrix}$$

Note that by suitable transformations $u_g(t)$ can be reduced from a first-order equation depending on G regression disturbances - $u_1(t-1)$, $u_2(t-1)$, $\dots$, $u_G(t-1)$ - to a G th order equation depending only on its own past values.⁵

In the two equation case,

$$u_1(t) = \rho_{11}u_1(t-1) + \rho_{12}u_2(t-1) + \varepsilon_1(t)$$

$$u_2(t) = \rho_{21}u_1(t-1) + \rho_{22}u_2(t-1) + \varepsilon_2(t)$$

can be transformed to

$$u_1(t) = (\rho_{11} + \rho_{22})u_1(t-1) + (\rho_{12}\rho_{21} + \rho_{11}\rho_{22})u_1(t-2) + \delta_1(t)$$

$$u_2(t) = (\rho_{11} + \rho_{22})u_2(t-1) + (\rho_{12}\rho_{21} + \rho_{11}\rho_{22})u_2(t-2) + \delta_2(t)$$

where

$$\delta_1(t) = \rho_{21}\varepsilon_2(t-1) - \rho_{22}\varepsilon_1(t-1) + \varepsilon_1(t)$$

$$\delta_2(t) = \rho_{12}\varepsilon_1(t-1) + \rho_{11}\varepsilon_2(t-1) + \varepsilon_2(t)$$

Furthermore, the vector autoregressive scheme requires restrictions on the autoregression parameters to guarantee that the effect of a particular disturbance diminishes as the number of periods separating the disturbances increases. For the two equation case, the necessary and sufficient conditions are:

$$1 + (\rho_{11} + \rho_{22}) + (\rho_{12}\rho_{21} + \rho_{11}\rho_{22}) > 0$$

$$1 - (\rho_{12}\rho_{21} + \rho_{11}\rho_{22}) > 0$$

$$1 - (\rho_{11} + \rho_{22}) + (\rho_{12}\rho_{21} + \rho_{11}\rho_{22}) > 0$$

(They are derived from the stability conditions of a second-order difference equation; that is, they are the necessary and sufficient conditions for both roots of the auxiliary equation to be less than one in absolute value.) Apparently Amemiya overlooked these requirements and in a published article considered cases in which the R matrix violated these conditions.⁶

The appeal of the vector autoregressive scheme seems to come from the opportunity to relax the diagonality restriction on R, the matrix of autoregression parameters. In econometrics, the researcher ought to impose restrictions on parameter values only when prior information can be incorporated into the estimation procedure.⁷ In general, the unrestricted version of any model should be estimated unless prior information indicates otherwise. In the case of autoregressive disturbances, some econometricians have pointed out that the simple autoregressive scheme is a special case of the more general vector autoregressive scheme.⁸ Extending their argument, the simple autoregressive scheme with its accompanying diagonality

restriction on R should be specified only if it is believed that the regression disturbance for each regression equation depends on its own lagged values, but in no way on the lagged values of regression disturbances of other equations. If there is no justification for this restriction, however, the more general vector autoregressive scheme, which places no restriction on R , must then be used if estimates with desirable asymptotic properties are to be obtained. Consequently, researchers who specify the simple autoregressive scheme when they suspect that their simultaneous equation model violates the basic assumption of nonautocorrelated disturbances may be unintentionally placing an unwarranted restriction upon their models. Moreover, allowing R to be a general matrix is analogous to allowing Φ to be a general matrix. Neither the R nor the Φ matrices should be limited to being diagonal unless the restrictions are justified.

One might disagree with the above interpretation of the simple and vector autoregressive schemes. The simple autoregressive scheme is based upon the plausible (although less than totally satisfactory) argument that was presented earlier. The vector autoregressive scheme has a shakier foundation. Guilkey endorses the vector autoregressive specification because it permits the model to account for all the relationships across equations among the regression disturbances and their lagged values.⁹ Guilkey expects these relationships to exist among the regression disturbances because the regression equations are themselves interdependent, which provides the motive for grouping the equations into a system. In a two equation

model, however, one should not necessarily specify the vector autoregressive scheme:

$$u_1(t) = \rho_{11}u_1(t-1) + \rho_{12}u_2(t-1) + \varepsilon_1(t)$$

$$u_2(t) = \rho_{21}u_1(t-1) + \rho_{22}u_2(t-1) + \varepsilon_2(t)$$

because - in regard to the first equation - $u_1(t)$ and $u_2(t-1)$ are correlated. The correlation between $u_1(t)$ and $u_2(t-1)$ can be attributed to a simple autoregressive scheme for each equation in which the disturbances $\varepsilon_1(t)$ and $\varepsilon_2(t)$ are contemporaneously correlated - a specification for the regression disturbances that has some support. In this specification (R diagonal, $\rho_{12} = \rho_{21} = 0$; ϕ general), $u_1(t)$ and $u_2(t-1)$ are correlated because

1. $u_1(t-1)$ and $u_2(t-1)$ are correlated due to the contemporaneous correlation between $\varepsilon_1(t-1)$ and $\varepsilon_2(t-1)$; and
2. $u_1(t)$ depends on its lagged value, $u_1(t-1)$.

To justify the vector autoregressive scheme shown above (R general; ϕ diagonal or general), the researcher needs to know that $u_1(t)$ depends upon $u_2(t-1)$. Therefore, in contrast to the earlier argument, a researcher with limited or incomplete information concerning the regression disturbances may have insufficient grounds for the vector autoregressive specification.

Furthermore, the vector autoregressive scheme has in itself no plausible economic interpretation. For a G equation system, the vector autoregressive scheme says that $u_1(t)$, the amount the dependent variable of the first equation deviates from its systematic part during period t , depends on $u_1(t-1)$, $\varepsilon_1(t)$, and $u_2(t-1)$, ... , $u_G(t-1)$, the amounts the dependent variables of the other structural equations

deviated from their systematic parts during period $t-1$. In the context of a small macro model, this means that the amount consumption departs from its systematic part during period t depends in part on the amount investment departed from its systematic part one period earlier. But there is no econometric principle that corresponds to the behavior of the model's regression disturbances suggested by the vector autoregressive scheme.

While specifying the vector autoregressive scheme, one might argue that the timing is different for the effects the omitted factors have on the model's equations. In a small macro model, certain factors outside the model may influence both consumption and investment, but their effects on consumption and investment begin at different times. These omitted factors may have an immediate effect on the investment equation and a proportional effect, delayed by one period, on the consumption equation. If so, then the following autoregressive scheme is called for in the two equation case:

$$u_1(t) = \rho_1 u_1(t-1) + \delta_{12} \varepsilon_2(t-1) + \varepsilon_1(t)$$

$$u_2(t) = \rho_2 u_2(t-1) + \delta_{21} \varepsilon_1(t-1) + \varepsilon_2(t)$$

This autoregressive scheme says that $u_1(t)$, the current disturbance of the first equation, depends upon

1. $u_1(t-1)$, the lingering effect the omitted factors began to have on the first equation one or more periods ago;
2. $\varepsilon_2(t-1)$, the effect that the omitted factors had immediately on the second equation, and that they have with a one period delay on the first equation; and
3. $\varepsilon_1(t)$, the immediate effect the omitted factors have on the first equation.

This autoregressive scheme differs slightly from the vector autoregressive scheme: $\varepsilon_g(t-1)$ replaces $u_g(t-1)$ in both equations. This autoregressive scheme nevertheless possesses what the vector autoregressive scheme lacks: a plausible econometric interpretation.

Through successive substitutions, $u_1(t)$ from the above autoregressive scheme can be shown to equal a linear combination of initial conditions - $u_1(0)$, $\varepsilon_2(0)$ - and sample values - $\varepsilon_1(1)$, $\varepsilon_1(2)$, $\dots$, $\varepsilon_1(t)$; $\varepsilon_2(1)$, $\varepsilon_2(2)$, $\dots$, $\varepsilon_2(t-1)$. Similarly $u_2(t)$ from the vector autoregressive scheme can be shown to equal a linear combination of initial conditions - $u_1(0)$, $u_2(0)$ - and the same sample values - $\varepsilon_1(1)$, $\varepsilon_1(2)$, $\dots$, $\varepsilon_1(t)$; $\varepsilon_2(1)$, $\varepsilon_2(2)$, $\dots$, $\varepsilon_2(t-1)$. If $\varepsilon_2(0) = u_2(0)$, then either autoregressive scheme can generate the same set of sample values for $u_1(t)$, $t=1, \dots, T$. In this light, the vector autoregressive scheme may be considered to be an approximation of the true autoregressive scheme from which it receives the econometric justification for its use.

Still, the vector autoregressive scheme requires substantially more information concerning the nature of the model's regression disturbances than does the simple autoregressive scheme. The vector autoregressive schemes requires that the correlation between $u_g(t)$ and $u_h(t-1)$ be at least partly due to the dependence of $u_g(t)$ on $\varepsilon_h(t-1)$, and not due solely to the correlation between $\varepsilon_g(t)$ and $\varepsilon_h(t)$. That is, it requires that some omitted factors disturb certain equations immediately and disturb other equations with a delay, while

there may be still other omitted factors that disturb all equations simultaneously. In fact, it seems as if the vector autoregressive scheme requires so much information concerning the nature of the regression disturbances that it is highly vulnerable to Griliches' criticism of complex autoregressive schemes.¹⁰ Griliches fears that complicated estimation methods designed in conjunction with a complex autoregressive scheme specified for the regression disturbances may only be covering up a misspecified model. He feels that econometricians may earn greater payoffs by working to identify the sources of the autoregression and to remove them from their models than by devising sophisticated estimation methods that "take care" of the problem. Furthermore, Kmenta has expressed in conversation his dislike for the way a first-order vector autoregressive scheme for a system with G equations implies a G th-order simple autoregressive scheme for each regression equation. Since a G equation system can be converted into a system with more or fewer equations by adding or deleting equations (for example, through substitution), the order of each equation's simple autoregressive scheme is rather arbitrary. The model builder and not the forces underlying the model determines the order of the autoregressive scheme. Kmenta further argues that if a G equation system has been specified to have a G th-order autoregressive scheme, then a first-order vector autoregressive scheme is a compact alternative for the model's autoregressive specification. On the other hand, if there are no grounds for assuming a G th order simple autoregressive scheme for each of the G equations, then the vector autoregressive scheme may be inappropriate.

In summary, the strongest argument for specifying the vector autoregressive scheme is its greater generality than the simple autoregressive scheme. While specifying the vector autoregressive scheme, a researcher must also keep in mind its stability conditions and the fact that it may be only an approximation of the true autoregressive scheme. The arguments against the vector autoregressive scheme concern its heavy demands on the researcher's prior knowledge. Aside from these considerations, it must also be noted that the structural disturbances may not follow the vector autoregressive scheme in the model's original form, and do so only when the model is transformed for estimation. A precedent for this possibility was set in the context of a single equation model whose dependent variable is explained by the "expected" or "permanent" level of the independent variable. The original form of the model is the following:

$$y(t) = \alpha_0 + \alpha_1 x^*(t) + \varepsilon(t)$$

where $x^*(t) = (1-\lambda)x(t) + \lambda x^*(t-1)$, $0 \leq \lambda < 1$,

and the model's disturbance term, $\varepsilon(t)$, fulfills all the basic assumptions. By a Koyck transformation, the unobservable variable $x^*(t)$ is eliminated and the model is put into the form that is actually estimated:

$$y(t) = \beta_0 + \beta_1 x(t) + \beta_2 y(t-1) + u(t)$$

where $u(t) = \varepsilon(t) - \lambda \varepsilon(t-1)$. Consequently, in the form that is estimated, the equation's disturbance follows a first-order moving average process, although originally the disturbance term fulfills all the basic assumptions. In a similar fashion, consider the following two equation model in its original form:

$$y_1(t) = \alpha_0 + \alpha_1 y_2^*(t) + \alpha_2 x_1(t) + u_1(t)$$

$$y_1(t) = \beta_0 + \beta_1 y_2^*(t) + \beta_2 x_2(t) + u_2(t)$$

where y_2^* is the permanent level of y_2 defined by

$$y_2^*(t) = (1-\lambda)y_2(t) + \lambda y_2^*(t-1), \quad 0 \leq \lambda < 1,$$

and let the model's disturbances follow the first-order simple autoregressive scheme; that is, $u_g(t) = \rho_g u_g(t-1) + \varepsilon_g(t)$ for $g=1,2$.

Since $y_2^*(t)$ is unobservable, the researcher might use a judicious substitution and estimate the model in the following form:

$$y_1(t) = \gamma_0 + \gamma_1 y_2(t) + \gamma_2 x_1(t) + \gamma_3 x_2(t-1) + v_1(t)$$

$$y_1(t) = \delta_0 + \delta_1 y_2(t) + \delta_2 x_2(t) + \delta_3 x_1(t-1) + v_2(t)$$

where

$$v_1(t) = \rho_1 u_1(t-1) - (\beta_1/\beta_2)u_2(t-1) + \varepsilon_1(t)$$

$$v_2(t) = \rho_2 u_2(t-1) - (\beta_2/\beta_1)u_1(t-1) + \varepsilon_2(t)$$

Subsequently, the specification of the transformed equation's structural disturbances, $v_1(t)$ and $v_2(t)$, can be rearranged by the elimination of $u_1(t-1)$ and $u_2(t-1)$ to resemble the first-order vector autoregressive scheme, where

$$v_1(t) = \rho_{11}v_1(t-1) + \rho_{12}v_2(t-1) + w_1(t)$$

$$v_2(t) = \rho_{21}v_1(t-1) + \rho_{22}v_2(t-1) + w_2(t)$$

for which $w_1(t)$ and $w_2(t)$ are complicated terms, based on current and lagged values of ε_1 and ε_2 ; and the ρ_{gh} coefficients depend on the model's α_k , β_k , λ , and ρ_g parameters. Consequently in the form that is actually estimated, the equation's disturbances follow a pattern that resembles the first-order vector autoregressive scheme, while in the original form, they do not. What this shows is that the "discovery" of vector autoregression by a researcher in the estimation

of an econometric model may be the product of transforming the model for the purpose of estimation. This leaves open the question whether vector autoregression is present in a model's original form.

Consequences of Autoregressive Disturbances

Autoregressive disturbances in the presence of lagged endogenous variables cause single or simultaneous equation model estimators to lose all the desirable properties for estimation, hypothesis testing, and prediction which they might possess when all assumptions are fulfilled.¹¹ Maximum likelihood estimation methods break down because the likelihood function based on nonautocorrelated disturbances no longer applies. Least squares estimation methods break down because autoregressive disturbances result in a contemporaneous correlation between the lagged endogenous variables acting as explanatory variables and the structural disturbances. In the simultaneous equation model $By(t) + \Gamma x(t) = u(t)$ where $u(t) = Ru(t-1) + \varepsilon(t)$, $y(t)$ is correlated with $u(t)$ and $y(t-1)$ is correlated with $u(t-1)$; hence, the lagged endogenous variables contained in $x(t)$ are contemporaneously correlated with $u(t)$. Therefore, econometricians have had to develop tests for the presence of autoregressive disturbances and other methods to obtain estimates of the structural coefficients possessing desirable properties in the context of autoregressive disturbances and lagged endogenous variables. A survey of these estimators and some tests follows.

CHAPTER I

FOOTNOTES

¹A. Koutsoyiannis, Theory of Econometrics (London: Macmillan Co., 1973), pp. 50-2.

²Ibid., pp. 188-9.

³These methods will be described in Chapter 2.

⁴The earliest delineation of the vector autoregressive scheme that I came across appeared in J.D. Sargan, "The Maximum Likelihood Estimation of Economic Relationships," Econometrica 29 (1961): 414-26. Since then, Hendry, Dhrymes, Fair, and others have used it.

⁵Giancarlo Gandolfo, Mathematical Methods and Models in Economics Dynamics (Amsterdam: North-Holland Publishing Company, 1971), Chapter 8.

⁶Takeshi Amemiya, "Specification Analysis in the Estimation of Parameters of a Simultaneous Equation Model with Autoregressive Errors," Econometrica 34 (1966): 302.

⁷Jan Kmenta, Elements of Econometrics (New York: Macmillan Co., 1971), pp. 430-451; of course, computational difficulties also dictate the imposition of certain restrictions. For example, the specification of linearity in the variables is rarely based on prior information.

⁸Sargan, "Maximum Likelihood Estimation," p. 414; David K. Guilkey, "Estimation of Systems of Regression Equations with First-Order Vector Autoregressive Errors" (Ph.D. dissertation, University of North Carolina, 1973), pp. 4-8.

⁹Guilkey, "Systems of Regression Equations," p. 8.

¹⁰Zvi Griliches, "A Note on Serial Correlation Bias in Estimates of Distributed Lags," Econometrica 27 (1961): 72.

¹¹The properties of estimators referred to here are those discussed in Kmenta, Elements of Econometrics, pp. 155-70.

CHAPTER II

ESTIMATION METHODS

Sargan was the first to deal with the compound estimation problem of simultaneity, first-order vector autoregressive disturbances, and lagged endogenous variables.¹ He merged the econometric methods designed for estimation of simultaneous equation models having non-autocorrelated disturbances with the methods designed for estimation of single-equation models with first-order autoregressive disturbances. He specified the model with G equations in its structural form as

$$By(t) + \Gamma x(t) = u(t)$$

where $x(t)$ includes lagged endogenous variables in addition to exogenous variables, and $u(t)$, the model's disturbances, are first-order vector autoregressive. Using the Orcutt transformation modified for vector autoregression, the model can be rewritten as

$$[By(t) + \Gamma x(t)] - R[By(t-1) + \Gamma x(t-1)] = \varepsilon(t)$$

where R is the matrix of autoregression parameters and $\varepsilon(t)$ represents the transformed model's nonautocorrelated disturbances, which may be contemporaneously correlated across equations. Sargan then derived the full-information maximum likelihood estimator accounting for first-order vector autoregressive disturbances analogous to the full-information maximum likelihood (FIML) estimator developed under the assumption of nonautocorrelated disturbances.² Note that "full-information" estimation methods will refer to estimators that take

into account the contemporaneous correlation across equations of the nonautocorrelated disturbances; "limited-information" estimation methods will refer to estimators that disregard this source of information. In the autoregressive case, the logarithmic likelihood function for the $T-1$ observations ($t = 2, 3, \dots, T$) of $y(t)$ conditional on the values of $x(t)$ and $y(1)$ is:

$$L = -(1/2)G(T-1) - (1/2)(T-1)\log|\phi| + (T-1)\log|B| \\ - (1/2)\sum_{t=2}^T \{ [By(t) + \Gamma x(t)] - R[By(t-1) + \Gamma x(t-1)] \}' \phi^{-1} \{ [By(t) + \Gamma x(t)] \\ - R[By(t-1) + \Gamma x(t-1)] \}$$

where $\phi = E[\varepsilon(t) \cdot \varepsilon(t)']$

The maximum likelihood estimates of B , Γ , R and ϕ are then obtained by maximizing the likelihood function with respect to these parameters. These estimates are consistent, asymptotically efficient, and asymptotically normal.³ Hendry was the first to report estimates produced from this estimation method. He took Klein's Model I and found the estimates generated by Sargan's method to be quite different from those estimates obtained under the assumption that the matrix R is diagonal or equals 0.⁴

After deriving the full-information maximum likelihood estimation method for the most general case (no zero restrictions on the elements of the matrices R and ϕ), Sargan developed a single-equation maximum likelihood estimation method for which simple autoregressive disturbances are assumed and the contemporaneous correlation across equations of the nonautocorrelated disturbances is neglected (which means that R and ϕ are considered to be diagonal matrices). This

estimator corresponds to the limited-information maximum likelihood (LIML) estimator developed in the absence of autocorrelation.⁵ If R is a diagonal matrix, then this method is consistent and asymptotically efficient among all limited-information estimators; however, if the restriction on R is improper, then no desirable asymptotic properties can be claimed because the likelihood function is improper.

Least Squares Estimation With Simple Autoregressive Disturbances

Amemiya extended Sargan's work with maximum likelihood estimators by deriving the limited-information least squares estimator of models with simple autoregressive disturbances.⁶ This estimator of simultaneous equation systems with simple autoregressive disturbances is a direct extension of the two-stage least squares (2SLS) estimator developed in the context of nonautocorrelated disturbances.⁷ For a description of this method, consider the first equation in the system:

$$y_1(t) = Y_1(t)\beta_1 + X_1(t)\gamma_1 + u_1(t)$$

where $u_1(t)$ follows the first-order simple autoregressive scheme; that is, $u_1(t) = \rho_1 u_1(t-1) + \varepsilon_1(t)$. After the application of the Orcutt transformation, the first equation becomes:

$$\begin{aligned} [y_1(t) - \rho_1 y_1(t-1)] &= [Y_1(t) - \rho_1 Y_1(t-1)]\beta_1 \\ &\quad + [X_1(t) - \rho_1 X_1(t-1)]\gamma_1 + \varepsilon_1(t) \end{aligned}$$

Then $Y_1(t)$ is replaced by its fitted values, $\hat{Y}_1(t)$, which are obtained by regressing each endogenous variable in $Y_1(t)$ on all the variables in the augmented reduced form - $Y(t-1)$, $X(t)$, $X(t-1)$.

(Note that some variables in $Y(t-1)$ will be redundant if they have already been included in $X(t)$.) Then β_1 , γ_1 , and ρ_1 (and their standard errors) in the equation

$$[y_1(t) - \rho_1 y_1(t-1)] = [\hat{Y}_1(t) - \rho_1 \hat{Y}_1(t-1)]\beta_1 \\ + [X_1(t) - \rho_1 X_1(t-1)]\gamma_1 + \varepsilon_1^*(t)$$

(where ε_1^* is the composite disturbance)

are estimated by a least squares method adapted for simple autoregressive disturbances. With no lagged endogenous variables among the predetermined variables, a two step method that obtains $\hat{\rho}_1$ separately from $(\hat{\beta}_1, \hat{\gamma}_1)$, such as Durbin's procedure, is consistent and asymptotically efficient among limited-information estimators as long as $\hat{\rho}_1$ is a consistent estimate of ρ_1 . The distribution of $(\hat{\beta}_1, \hat{\gamma}_1)$ is independent of the distribution of $\hat{\rho}_1$ in the absence of lagged endogenous variables, and therefore, only a consistent estimate of ρ_1 is required to establish the desirable asymptotic properties of $(\hat{\beta}_1, \hat{\gamma}_1)$. This means that a consistent estimate of ρ_1 is a substitute for knowledge of ρ_1 . With lagged endogenous variables among the predetermined variables, a two step method that obtains $\hat{\rho}_1$ separately from $(\hat{\beta}_1, \hat{\gamma}_1)$ achieves only consistency if $\hat{\rho}_1$ is a consistent estimate of ρ_1 . A consistent estimate of ρ_1 in this situation is inferior to knowledge of ρ_1 . To achieve asymptotic efficiency, the estimates of (β_1, γ_1) and ρ_1 must be found jointly because the distributions of $(\hat{\beta}_1, \hat{\gamma}_1)$ and $\hat{\rho}_1$ are not independent.⁸ Three equivalent methods for jointly estimating (β_1, γ_1) and ρ_1 are nonlinear least squares, the Hildreth-Lu search procedure, and the Cochrane-Orcutt iterative procedure.⁹ If the covariance of $(\hat{\beta}_1, \hat{\gamma}_1)$

and $\hat{\rho}_1$ is asymptotically negligible (which Fair assumes), then the approximate variance-covariance matrix of $(\hat{\beta}_1, \hat{\gamma}_1)$ is estimated by:¹⁰

$$\begin{aligned} s_{\varepsilon}^2[1](\hat{Z}_1^* \hat{Z}_1^*)^{-1} &= s_{\varepsilon}^2[1]\{[\hat{Z}_1 - \hat{\rho}_1 Z(-1)]' [\hat{Z}_1 - \hat{\rho}_1 Z(-1)]\}^{-1} \\ &= s_{\varepsilon}^2[1]\{(\hat{Y}_1^* \ X_1^*)' (\hat{Y}_1^* \ X_1^*)\}^{-1} \\ &= s_{\varepsilon}^2[1]\{[\hat{Y}_1 - \hat{\rho}_1 Y_1(-1) \ X_1 - \hat{\rho}_1 X_1(-1)]' \\ &\quad [\hat{Y}_1 - \hat{\rho}_1 Y_1(-1) \ X_1 - \hat{\rho}_1 X_1(-1)]\}^{-1} \end{aligned}$$

where $Z_1 = (Y_1 \ X_1)$

$s_{\varepsilon}^2[1] = [1/(T-N_1^*-1)]e_1'e_1$, the estimated variance of $\varepsilon_1(t)$

$$e_1 = y_1^* - \hat{Y}_1^* \hat{\beta}_1 - X_1^* \hat{\gamma}_1$$

N_1^* = the number of explanatory variables appearing in the first equation. The approximate variance of $\hat{\rho}_1$ is estimated by $(1-\hat{\rho}_1)^2/T-1$.¹¹

Although Amemiya never pointed this out, it is easy to continue on to the full-information least squares estimator designed for simple first-order autoregressive disturbances. This method accounts for the contemporaneous correlation among the ε 's, the non-autocorrelated disturbances. It parallels the procedures in the three-stage least squares (3SLS) estimator developed under the assumption of nonautocorrelated disturbances.¹² After estimating each equation by Amemiya's limited-information least squares method described above, we estimate $\sigma_{\varepsilon}[g,h]$, the variance and covariance of the ε 's, by the usual method:

$$s_{\varepsilon}[g,h] = [1/(T-N_g^*-1)]e_g'e_h$$

where $N_g^* > N_h^*$ and g and $h = 1, 2, \dots, G$. Then β_g and γ_g are estimated for all G equations at once by

$$\hat{\alpha} = (\hat{Z}^* \hat{\phi}^{-1} \hat{Z}^*)^{-1} (\hat{Z}^* \hat{\phi}^{-1} y^*)$$

where $\hat{\phi}$ replaces $\phi_* = E(\varepsilon^* \varepsilon^{*\prime})$, whose elements are unknown, and is composed of the submatrices $s_{\varepsilon}[g, h] \cdot I_{T-1}$. The approximate asymptotic variance-covariance matrix of $\hat{\alpha}$ is estimated by $(\hat{Z}^* \hat{\phi}^{-1} \hat{Z}^*)^{-1}$.

Besides developing the autoregressive analogue to the 2SLS estimator, Amemiya compared the performance of his method (limited-information, least squares, simple autoregression) with the 2SLS estimator when the model has vector autoregressive disturbances. That is, he compared these two estimators when both estimators mis-specify the model's autoregressive scheme: the 2SLS method specifies $R=0$, Amemiya's method R diagonal, the true autoregressive scheme is R general. This work was motivated by Sargan's suggestion that 2SLS estimation of a single equation in a model that actually had vector autoregressive disturbances would be improved by incorrectly assuming simple autoregressive disturbances. In order to make this comparison through computations of the estimators' variance-covariance matrices instead of sampling experiments, Amemiya specified that all predetermined variables in the model were exogenous and then slightly modified his estimation method. Amemiya's modified estimator is consistent whether the R matrix is general or diagonal; however, it is never asymptotically efficient among limited-information estimators. From his calculations, Amemiya concludes that in almost every case estimation is indeed improved by assuming simple autoregressive disturbances rather than nonautocorrelated disturbances even though the disturbances were actually vector autoregressive.

While Amemiya's conclusion will not be disputed (in fact, the sampling experiments reported in chapter V somewhat confirm his conclusion), it should be pointed out again that some of the cases he considered involved vector autoregressive processes whose values were implausible. (This oversight was noted in chapter I.)

Fair was dissatisfied with the prospects of implementing Amemiya's limited-information least squares estimator (the analogue to the 2SLS estimator).¹³ The successful application of Amemiya's estimator requires that the number of observations in the sample be greater than the number of variables in the augmented reduced form; $T > G + 2K + 1$ minus the number of redundant variables. The available sample will often be undersized, or too small to meet this condition, if the model contains more than a few equations. This means that the fitted values to be used in the second stage cannot be obtained. Consequently, Fair modified Amemiya's method so that estimation would require fewer observations. Fair's method differs from Amemiya's by the number of variables used in forming $\hat{Y}_1(t)$, the fitted values of $Y_1(t)$. Instead of using all predetermined variables in the augmented reduced form - $Y(t-1)$, $X(t)$, $X(t-1)$ - only a subset is used. Each endogenous variable in $Y_1(t)$ is regressed only on those predetermined variables which appear in the first equation after the transformation - $y_1(t-1)$, $Y_1(t-1)$, $X_1(t)$, $X_1(t-1)$. Thus, equations which were impossible to estimate by Amemiya's method due to an insufficient number of observations ($T \leq G + 2K + 1$ minus the number of redundant variables) will become feasible to estimate by

Fair's method if $T > G_g^{\Delta} + 2K_g^* + 1$ minus the number of redundant variables. Although more widely applicable than Amemiya's method, this method is not asymptotically efficient among all limited-information estimation methods. Fair's method employs a subset of the variables in the augmented reduced form to obtain the fitted values $\hat{Y}_1(t)$, while Amemiya's method employs all the variables in the augmented reduced form. Therefore, in Fair's method $\hat{Y}_1(t)$ is not as highly correlated with the variable that it replaces, $Y_1(t)$, as it is in Amemiya's method. This means that Fair's method must have a higher asymptotic variance than Amemiya's method, which is asymptotically efficient.¹⁴ Consequently, the practice of using only a subset of the variables in the augmented reduced form to obtain the fitted values, $\hat{Y}_1(t)$, causes the loss of asymptotic efficiency for Fair's method. The loss of asymptotic efficiency is trivial, however, since Fair's method is needed only if the sample is undersized. The finite sample performance of Fair's method is all that is relevant. The approximate asymptotic variance-covariance matrix of $(\hat{\beta}_1, \hat{\gamma}_1)$ and the approximate asymptotic variance of $\hat{\rho}_1$ are estimated by the same formulae as those used in Amemiya's method (page 26). Fair used his method to estimate a small (eight equation) macro model of the United States' economy.¹⁵ He reports that the predictive performance of this small macro model, estimated with his method which accounts for simple first-order autoregressive disturbances, equals the predictive performance of far more elaborate macro models that were estimated under the assumption of nonautocorrelated disturbances.

Fair's method has been criticized because an endogenous variable which appears as an explanatory variable in two or more equations will have different sets of fitted values for each equation.¹⁶ This is because $y_g(t-1)$, $Y_g(t-1)$, $X_g(t)$, and $X_g(t-1)$ will be different in each case. It was never explained though why this feature of Fair's method poses a difficulty for those who use it.

Dhrymes, Berner and Cummins developed a variation to the methods of Amemiya and Fair.¹⁷ In its second stage, this estimator, like the previous estimators, uses the least squares method to estimate a single equation of the model after:

1. predicted values from the first stage are substituted for the actual values of the endogenous variables appearing as explanatory variables; and
2. a consistent estimate of the autoregression parameter, ρ_g , is obtained to transform the equation.

It is the procedure through which fitted values are obtained in the first stage which distinguishes this method from the previous limited-information least squares methods. The first stage involves these steps:

1. Obtain consistent estimates of each equation in the model through instrumental variable estimation without transforming the equation or estimating ρ_g . Instruments must be used for both current and lagged endogenous variables appearing as explanatory variables because the disturbances are autoregressive.
2. Use these estimates of the structural form to derive a consistent estimate of the model's reduced form (Not the augmented

reduced form) by using the identity $\Pi = -B^{-1}\Gamma$. (The previous methods estimated either a portion or the entire augmented reduced form directly.)

3. Obtain predicted values for the endogenous variables of the model by dynamic simulation, given initial conditions. That is, predicted values are obtained recursively by using generated values of the endogenous variables in the place of the actual values of the lagged endogenous variables and in conjunction with the derived reduced form estimates. (The previous methods used the actual values of the lagged endogenous variables. Note that this method, like Fair's, permits estimation with samples too small for Amemiya's method; that is, with sample size $T \leq G+2K+1$ minus the number of redundant variables in the augmented reduced form. This method, however, circumvents the criticism of Fair's method mentioned above.)

This estimation method is significantly more complex than those of Amemiya and Fair, yet it does not achieve asymptotic efficiency among limited-information estimators. An unusual feature of this method is that iterating the procedure with a new estimate of ρ_g will not alter its asymptotic properties.

Least Squares Estimation With Vector Autoregressive Disturbances

Subsequently, Dhrymes reconsidered the problem of full-information estimation of a model with vector autoregressive disturbances. He extended Sargan's work with maximum likelihood estimation methods by deriving the full-information least squares estimator under the

assumption of vector autoregressive disturbances.¹⁸ This estimator is an adaptation of the three-stage least squares (3SLS) estimator developed in the context of nonautocorrelated disturbances. Dhrymes' estimation procedure is as follows:

1. Estimate each structural equation consistently through instrumental variables by
 - a. obtaining fitted values for the endogenous variables by regressing each endogenous variable on just the exogenous variables in the model;
 - b. replacing the actual values of all endogenous variables acting as explanatory variables - lagged and unlagged - with these fitted values; and
 - c. estimating each equation by ordinary least squares.

2. From these consistent estimates of the structural model:

- a. obtain $\hat{u}(t)$, a consistent predictor of $u(t)$ for all t ;
- b. estimate each of the G rows in the matrix R by applying ordinary least squares to

$$\begin{aligned} \hat{u}_g(t) = & \rho_{g1} \hat{u}_1(t-1) + \rho_{g2} \hat{u}_2(t-1) + \dots + \rho_{gg} \hat{u}_g(t-1) + \dots \\ & + \rho_{gG} \hat{u}_G(t-1) + w_g(t) \end{aligned}$$

for $g = 2, \dots, G$; where $w_g(t)$ is a composite disturbance.

- c. estimate the elements of Φ , the variance-covariance matrix of the nonautocorrelated disturbances, by

$$s_e[g,h] = (1/T-1) \sum_{t=2}^T e_g(t) e_h(t)$$

for $g,h = 1, \dots, G$

$$\text{where } e_g(t) = \hat{u}_g(t) - \hat{\rho}_{g1} \hat{u}_1(t-1) - \hat{\rho}_{g2} \hat{u}_2(t-1) \dots - \hat{\rho}_{gG} \hat{u}_G(t-1)$$

3. Use $\tilde{R}$ to transform the system $y(t) = Z(t)\alpha + u(t)$, where $u(t)$ is first-order vector autoregressive, to

$$[y(t) - \tilde{R}y(t-1)] = [Z(t) - \tilde{R}Z(t-1)]\alpha + [u(t) - \tilde{R}u(t-1)]$$
 or, $y^*(t) = Z^*(t)\alpha + \tilde{\varepsilon}(t)$
 where the * superscript indicates variables transformed by $\tilde{R}$ and $\tilde{\varepsilon}(t)$ is the composite disturbance.
4. Estimate the augmented reduced form - which includes $Y(t-1)$, $X(t)$, and $X(t-1)$ - to obtain new fitted values for the endogenous variables. In the next stage, these fitted values will replace the actual values of only the current endogenous variables appearing as explanatory variables in the structural equations. This is in contrast with the earlier stage in which the actual values of both current and lagged endogenous variables were replaced
5. Estimate α consistently by

$$\hat{\alpha} = [\hat{Z}^{*'}(\hat{\Phi}^{-1} \otimes I_{T-1})\hat{Z}^*]^{-1}[\hat{Z}^{*'}(\hat{\Phi}^{-1} \otimes I_{T-1})y^*]$$
6. For asymptotically efficient estimates, iterate the procedure with a new estimate of R until the estimates of α and R converge. (Iteration is necessary due to the presence of lagged endogenous variables; convergence, however, is not guaranteed.)
7. The asymptotic variance-covariance matrix of the final estimates of α and R has been worked out by Dhrymes and Erlat.¹⁹

By setting Φ equal to a diagonal matrix, Dhrymes' limited-information least squares estimator for a model with vector autoregressive disturbances is obtained. This method is asymptotically

efficient among all limited-information estimators. An illustration of the inner workings of this limited-information estimator will elucidate Dhrymes' econometric techniques. Suppose the model to be estimated contains two equations; then it can be put into the form:

$$\begin{bmatrix} y_1(1) \\ y_1(2) \\ . \\ . \\ . \\ y_1(T) \\ \hline y_2(1) \\ y_2(2) \\ . \\ . \\ . \\ y_2(T) \end{bmatrix} = \begin{bmatrix} 1 & W_1(1) & 0 & 0 \\ 1 & W_1(2) & 0 & 0 \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ 1 & W_1(T) & 0 & 0 \\ \hline 0 & 0 & 1 & W_2(1) \\ 0 & 0 & 1 & W_2(2) \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ 0 & 0 & 1 & W_2(T) \end{bmatrix} \begin{bmatrix} \alpha_{10} \\ \alpha_{11} \\ \hline \alpha_{20} \\ \alpha_{21} \end{bmatrix} + \begin{bmatrix} u_1(1) \\ u_1(2) \\ . \\ . \\ . \\ u_1(T) \\ \hline u_2(1) \\ u_2(2) \\ . \\ . \\ . \\ u_2(T) \end{bmatrix}$$

where $W_1(t)$ and $W_2(t)$ are row vectors containing each equation's explanatory variables. After the transformation, the model takes the form

$$\begin{bmatrix} y_1(2) - \rho_{11}y_1(1) - \rho_{12}y_2(1) \\ y_1(3) - \rho_{11}y_1(2) - \rho_{12}y_2(2) \\ . \\ . \\ . \\ y_1(T) - \rho_{11}y_1(T-1) - \rho_{12}y_2(T-1) \\ \hline y_2(2) - \rho_{21}y_1(1) - \rho_{22}y_2(1) \\ y_2(3) - \rho_{21}y_1(2) - \rho_{22}y_2(2) \\ . \\ . \\ . \\ y_2(T) - \rho_{21}y_1(T-1) - \rho_{22}y_2(T-1) \end{bmatrix} =$$

$$\begin{bmatrix}
 1-\rho_{11} & W_1(2)-\rho_{11}W_1(1) & -\rho_{12} & -\rho_{12}W_2(1) \\
 1-\rho_{11} & W_1(3)-\rho_{11}W_1(2) & -\rho_{12} & -\rho_{12}W_2(2) \\
 . & . & . & . \\
 . & . & . & . \\
 . & . & . & . \\
 1-\rho_{11} & W_1(T)-\rho_{11}W_1(T-1) & -\rho_{12} & -\rho_{12}W_2(T-1) \\
 \hline
 -\rho_{21} & -\rho_{21}W_1(1) & 1-\rho_{22} & W_2(2)-\rho_{22}W_2(1) \\
 -\rho_{21} & -\rho_{21}W_1(2) & 1-\rho_{22} & W_2(3)-\rho_{22}W_2(2) \\
 . & . & . & . \\
 . & . & . & . \\
 . & . & . & . \\
 -\rho_{21} & -\rho_{21}W_1(T-1) & 1-\rho_{22} & W_2(T)-\rho_{22}W_2(T-1)
 \end{bmatrix}
 \begin{bmatrix}
 \alpha_{10} \\
 \alpha_{11} \\
 \alpha_{20} \\
 \alpha_{21}
 \end{bmatrix}
 +
 \begin{bmatrix}
 \varepsilon_1(2) \\
 \varepsilon_1(3) \\
 . \\
 . \\
 . \\
 \varepsilon_1(T) \\
 \varepsilon_2(2) \\
 \varepsilon_2(3) \\
 . \\
 . \\
 . \\
 \varepsilon_2(T)
 \end{bmatrix}$$

Once the matrix R has been consistently estimated and the actual values of the endogenous variables acting as explanatory variables have been replaced by their fitted values, apply ordinary least squares to this form of the model for consistent estimates. Asymptotic efficiency (among limited-information methods) requires

1. a correction for heteroskedasticity if $\text{Var}(\varepsilon_1) \neq \text{Var}(\varepsilon_2)$, and
2. iteration of the procedure with a new estimate of R until convergence occurs (convergence not guaranteed).

If R is set equal to a diagonal matrix, Dhrymes' method simplifies to the limited and full-information variants of Amemiya's least squares estimator. When either nonautocorrelation or simple autoregression is specified for the disturbances of a system of equations, note that limited-information estimators can be characterized as single-equation estimation methods, which estimate each equation in the system separately; and full-information estimators can be

characterized as system methods of estimation, which compress the system into one equation and estimate all G equations of the model together.²⁰ When vector autoregressive disturbances are specified, however, this distinction no longer holds because both the limited and full-information estimators are system methods of estimation. Consequently all estimators designed for models with vector autoregressive disturbances, not just their full-information versions, are sensitive to specification errors, a difficulty common to all system methods of estimation. Since a system method estimates all the model's equations together, the effects of a specification error in one equation are spread throughout the model, and cannot be isolated in the misspecified equation.

Dhrymes and Erlat note that iterating the procedure with new estimates of R and ϕ will provide different, asymptotically efficient estimates.²¹ These estimates will have different numerical values, but the same asymptotic properties as those estimates obtained by iterating only with a new estimate of R . This is because the estimator's asymptotic distribution only requires a consistent estimate of ϕ for asymptotic efficiency: iteration with a new estimate of ϕ does not matter.

Dhrymes' full and limited-information estimators for models with vector autoregressive disturbances suffer from the same problem as Amemiya's full and limited-information estimators with simple autoregressive disturbances; if the sample is undersized, then estimation cannot be carried out. The large number of variables in the

augmented reduced form restricts the use of this method to models with only a few equations; otherwise, fitted values cannot be obtained for the second stage. This limitation once again stimulated Fair to modify an existing estimation method so that fewer observations would be required for successful estimation of a model.²²

Fair uncovered the necessary modifications to Dhrymes' method by extending Brundy and Jorgenson's efficient instrumental variables estimation method to models with autoregressive disturbances.²³

Brundy and Jorgenson developed their method for models with non-autocorrelated disturbances because even relatively large samples will often be insufficient to allow the use of 2SLS or 3SLS estimation. If there are fewer observations in the sample than predetermined variables in the model, which happens frequently among even moderate sized models, the estimation again breaks down because fitted values cannot be obtained for the second stage. Brundy and Jorgenson's breakthrough consists of obtaining fitted values for the second stage while bypassing a direct estimate of the reduced form in the first stage. They suggest estimating each structural equation consistently and then making use of the identity $\Pi = -B^{-1}\Gamma$ to derive a consistent estimate of the reduced form from which the fitted values are obtained. Fair subsequently extends the Brundy and Jorgenson technique to models with autoregressive disturbances, for which the problem of sample size is more prevalent. The necessary adjustment to Dhrymes' method takes place in step 4 (page 33); one does not estimate the augmented reduced form to obtain new fitted values; instead, one can use the consistent estimate of R from step

2, and the identities $\Pi_1 = B^{-1}RB$, $\Pi_2 = -B^{-1}\Gamma$, $\Pi_3 = B^{-1}R\Gamma$ to derive a consistent estimate of the augmented reduced form;

$$\begin{aligned} y(t) &= \Pi_1 y(t-1) + \Pi_2 x(t) + \Pi_3 x(t-1) + w(t) \\ &= B^{-1}Ry(t-1) - B^{-1}\Gamma x(t) + B^{-1}R\Gamma x(t-1) + B^{-1}\varepsilon(t) \end{aligned}$$

From these derived reduced form estimates, we can obtain the fitted values that replace the actual values of the endogenous variables acting as explanatory variables. Fair claims consistency and asymptotic efficiency (provided that the procedure is iterated with a new estimate of R until convergence is reached) for his method. Asymptotic efficiency is an empty claim since this method need only be used with undersized samples; only the finite sample performance of the estimator is meaningful. If the limited-information version of this method is used with the assumption that R is diagonal, then this method is an alternative to Fair's earlier method.

Both Dhrymes' and Fair's methods require iteration with a new estimate of R because α and R must be estimated jointly in the presence of lagged endogenous variables in order to achieve asymptotic efficiency. Iterative procedures are a problem, however, because they are computationally expensive and there is no guarantee that the estimates will converge. To circumvent this difficulty, Hatanaka developed an asymptotically efficient two step estimator that does not require iteration.²⁴ It is easiest to understand Hatanaka's innovation in the context of single-equation models, for which, in fact, he developed it.²⁵ Consider the following regression equation which contains a lagged dependent variable:

$$y(t) = \beta_0 + \beta_1 x(t) + \beta_2 y(t-1) + u(t)$$

where $u(t)$ is a first-order autoregressive disturbance. As stated earlier, a two step procedure that uses a consistent estimate of ρ obtained separately from $\hat{\beta}$ subsequently to transform the equation leads to consistent estimation of a regression equation containing a lagged dependent variable. Joint estimation of ρ and the β 's is necessary to achieve asymptotic efficiency. In order to avoid iteration or searching over all possible values of ρ while still achieving asymptotic efficiency, Hatanaka suggests using this procedure: take $\tilde{u}(t)$, the consistent predictor of $u(t)$, from which the consistent estimate $\tilde{\rho}$ came, and use $\tilde{u}(t-1)$ as an additional explanatory variable in the transformed equation. That is, after obtaining an initial consistent estimate of ρ from the consistent predictor $\tilde{u}(t)$, apply ordinary least squares to the equation

$$\begin{aligned} [y(t) - \tilde{\rho}y(t-1)] &= \beta_0(1-\tilde{\rho}) + \beta_1[x(t) - \tilde{\rho}x(t-1)] \\ &\quad + \beta_2[y(t-1) - \tilde{\rho}y(t-2)] \\ &\quad + \rho^* \tilde{u}(t-1) + \tilde{\varepsilon}(t) \end{aligned}$$

Then the estimates $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\rho} = (\tilde{\rho} + \hat{\rho}^*)$ are consistent and asymptotically efficient estimates of the equation's parameters without iteration since the β 's and ρ are estimated jointly. The estimate of $\hat{\rho}^*$ can be considered to be a correction of ρ 's original estimate $\tilde{\rho}$. Although the residual adjustment estimator avoids iteration and the possibility that the estimates will not converge, this is done with the cost that the final estimate of ρ , $\hat{\rho} = (\tilde{\rho} + \hat{\rho}^*)$, may be implausible ($|\hat{\rho}| > 1$).

Extending this technique to simultaneous equation systems requires a relatively minor adjustment to Dhrymes' and Fair's methods. Step 5 of Dhrymes' procedure (page 33) calls for the application of the generalized least squares (GLS) method with ϕ_* replaced by its consistent estimate $\hat{\phi}$ to the equation

$$y^*(t) = \hat{Z}^*(t)\alpha + \varepsilon^*(t)$$

where the variables have been transformed by $\tilde{R}$, the initial estimate of the matrix R . With the incorporation of Hatanaka's technique, step 5 now involves the application of the generalized least squares method to the equation

$$y^*(t) = \hat{Z}^*(t)\alpha + [I_g \otimes \tilde{u}'(t-1)]R^* + \varepsilon^*(t)$$

where $\tilde{R}$ transforms the variables. Hatanaka and also Dhrymes and Taylor claim that $\hat{\alpha}$ and $\hat{R} = (\tilde{R} + \hat{R}_m^*)$ - where $\hat{R}_m^*$, the correction to R 's initial estimate, $\tilde{R}$, is a $(G \times G)$ matrix arranged in accordance with R and obtained from $\hat{R}^*$, a $(G^2 \times 1)$ column vector - are consistent and asymptotically efficient estimates of α and R without iteration since α and R are estimated jointly.²⁶ Hatanaka's method will be referred to as the full-information residual adjusted estimator designed for models with vector autoregressive disturbances.

Assuming or ignoring the fact that the nonautocorrelated disturbances are contemporaneously uncorrelated across equations results in the limited-information counterpart of Hatanaka's estimator. Furthermore, if the disturbances follow the simple rather than the vector autoregressive scheme, then Hatanaka's method reduces to full and limited-information residual adjusted estimators that are residual adjusted variants of Amemiya's least squares estimators. Hatanaka's

residual adjustment innovation can be used in conjunction with many least squares estimators designed for models with autoregressive disturbances. (The survey of estimation methods just completed was in general presented in chronological order. Appendix A contains a chart that stresses the relationships that the estimators in the survey have to one another and to some other familiar estimators.)

To carry out the estimation work and sampling experiments in the following chapters, Amemiya's limited-information least squares estimator designed for models with simple autoregressive disturbances (which will be referred to as 2SLS-AUTO) and Hatanaka's limited-information residual adjusted estimator designed for models with vector autoregressive disturbances (which will be referred to as 2SLS(RA)-VECTOR) will be used. Amemiya's and Hatanaka's methods were chosen from among the estimation methods described above because:

1. these methods are asymptotically efficient among all limited-information estimators in the context of the autoregressive scheme for which they were designed; and
2. these methods require the least computational effort.

The limited-information rather than the full-information versions of the estimators will be used because there is reasonable concern over the specification of the models that are to be reestimated.

Tests for Autoregressive Disturbances

To determine if one of the preceding methods, developed under the assumption of autoregressive disturbances, is required for asymptotically efficient estimates of the model, tests of the disturbances' specification are needed. The three specifications for the disturbances under consideration are nonautoregression, simple autoregression, and vector autoregression. Three tests are thus called for:

1. A test of the hypothesis of nonautoregression, $H_0: \rho_g = 0$, against the alternative of simple autoregression, $H_A: \rho_g \neq 0$; (for all $g=1, \dots, G$)
2. A test of the hypothesis of nonautoregression, $H_0: R = 0$, against the alternative of vector autoregression, $H_A: R \neq 0$; and
3. A test of the hypothesis of simple autoregression, H_0 : the matrix R is diagonal, against the alternative of vector autoregression, H_A : the matrix R is general.

The test for the presence of simple autoregressive disturbances will be carried out by testing whether the estimated value of the autoregression parameter, ρ_g , is significantly different from zero. If $\rho_g/s_\epsilon[g] < 2$, then the estimate will be considered significant.²⁷ Note that the Durbin-Watson test is inappropriate here due to the presence of lagged endogenous variables.²⁸

Several ways have been devised to perform the second test - the test of the hypothesis of nonautoregression against the alternative of vector autoregression. Hendry²⁹ and Goldfeld and Quandt³⁰ suggest the use of the likelihood ratio test. Guilkey proposes a rather complicated test which he describes as a generalization of

Durbin's procedure for the single equation model.³¹ Goldfeld and Quandt also point out that a very simple test would be to reject the null hypothesis of nonautoregression if the estimate of any individual element in R is significantly different from zero.³² This test can be improved upon in a straightforward manner: instead of using the OLS estimation method separately on the G equations of the form

$$\tilde{u}_g(t) = \rho_{g1}\tilde{u}_1(t-1) + \dots + \rho_{gG}\tilde{u}_G(t-1) + \varepsilon_g^*(t)$$

to obtain $\tilde{R}$, we can compress these G equations into one large equation

$$\begin{bmatrix} u_1(t) \\ u_2(t) \\ \vdots \\ u_G(t) \end{bmatrix} = \begin{bmatrix} u_1(t-1) \dots u_G(t-1) & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & u_1(t-1) \dots u_G(t-1) & 0 & \dots & 0 \\ \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 & u_1(t-1) \dots u_G(t-1) \end{bmatrix} \begin{bmatrix} \rho_{11} \\ \vdots \\ \rho_{1G} \\ \vdots \\ \rho_{G1} \\ \vdots \\ \rho_{GG} \end{bmatrix}$$

$$+ \begin{bmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \\ \vdots \\ \varepsilon_G(t) \end{bmatrix}$$

and through the generalized least squares (GLS) method estimate the entire R matrix at once. (The GLS method is needed to account for heteroskedasticity and contemporaneous correlation among the nonauto-correlated disturbances.) Then the F -test for the overall significance of the regression can perform the test of the hypothesis of nonautoregression, $R = 0$, against the alternative of vector autoregression, $R \neq 0$.³³ This is an improvement of Goldfeld and Quandt's test because it is possible for none of the estimated elements of the R matrix to be significantly different from zero, while the F -test rejects the null hypothesis of no relationship between the dependent and explanatory variables. When it is necessary, this test will be used in the later chapters.

Finally, several ways have also been devised to test for vector autoregression when simple autoregression is the null hypothesis. Again the use of the likelihood ratio test is one possibility; a test developed by Hendry is another.³⁴ A test related to the F -test for the significance of the regression can be employed, however, when the model's autoregressive scheme is compressed into one equation. The maintained hypothesis of simple autoregression implies that all off diagonal elements of R , ρ_{gh} for $g \neq h$, equal zero; the alternative hypothesis of vector autoregression implies that this is not true. Therefore, an F -test for the relevance of additional explanatory variables in a regression equation can be used to carry out the test of simple versus vector autoregression.³⁵ If the F -test rejects the maintained hypothesis, $H_0: \rho_{gh} = 0$ for $g \neq h$, then the presence of vector autoregressive disturbances is indicated. This test will be used as needed in later chapters.

CHAPTER II

FOOTNOTES

¹J.D. Sargan, "The Maximum Likelihood Estimation of Economic Relationships," Econometrica 29 (1961): 414-26.

²Jan Kmenta, Elements of Econometrics (New York: Macmillan Company, 1971), p. 578.

³Sargan did not prove these claims. Instead, see P.J. Dhrymes and John B. Taylor, "On an Efficient Estimator for Dynamic Simultaneous Equation Models with Autoregressive Errors," International Economic Review 17 (1976): 362-76.

⁴David F. Hendry, "Maximum Likelihood Estimation of Systems of Simultaneous Regression Equations with Errors Generated by a Vector Autoregressive Process," International Economic Review 12 (1971): 257-72; "Correction" 15 (1974): 260.

⁵Kmenta, Elements, p. 567.

⁶Takeshi Amemiya, "Specification Analysis in the Estimation of Parameters of a Simultaneous Equation Model with Autoregressive Residuals," Econometrica 34 (1966): 283-306.

⁷Kmenta, Elements, p. 559.

⁸G.S. Maddala, "Generalized Least Squares with an Estimated Covariance Matrix," Econometrica 39 (1971): 23-33.

⁹Kmenta, Elements, p. 285 and 446-7; the convergence criteria and the equivalence of the new procedures is discussed in W. Oberhofer and Jan Kmenta, "A General Procedure for Obtaining Maximum Likelihood Estimates in Generalized Regression Models," Econometrica 42 (1974): 579-90. It does not seem that convergence is guaranteed on account of the presence of a lagged exogenous variable. In practice, according to Fair, this method always seems to converge.

¹⁰The * superscript indicates variables transformed by the estimated value of ρ .

¹¹Ray C. Fair, "The Estimation of Simultaneous Equation Models with Lagged Endogenous Variables and First-Order Serially Correlated Errors," Econometrica 38 (1970): 514.

- ¹²Kmenta, Elements, p. 573.
- ¹³Fair, "Simultaneous Equation Models," pp. 507-16.
- ¹⁴Kmenta, Elements, p. 311.
- ¹⁵Ray C. Fair, A Short-Run Forecasting Model for the United States Economy (Lexington, Mass.: Heath Lexington Books, 1971).
- ¹⁶P.J. Dhrymes, R. Berner, and D. Cummins, "A Comparison of Some Limited Information Estimators for Dynamic Simultaneous Equations Models with Autocorrelated Errors," Econometrica 42 (1974): 313.
- ¹⁷This method was first described in P.J. Dhrymes, R. Berner, and D. Cummins, "Limited Information Estimation of Simultaneous Equations Models with Lagged Endogenous Variables and Autocorrelated Errors" (Discussion paper No. 183, University of Pennsylvania, 1970), and later in Dhrymes, Berner, and Cummins, "A Comparison of Some Limited Information Estimators," pp. 311-32.
- ¹⁸P.J. Dhrymes, "Full Information Estimation of Dynamic Simultaneous Equations Models with Autoregressive Errors" (Discussion paper No. 203, University of Pennsylvania, 1971).
- ¹⁹P.J. Dhrymes and H. Erlat, "Asymptotic Properties of Full Information Estimators in Dynamic Autoregressive Equation Models," Journal of Econometrics 2 (1974): 255.
- ²⁰Kmenta, Elements, Sections 13-3 and 13-4.
- ²¹Dhrymes and Erlat, "Asymptotic Properties," pp. 247-59.
- ²²Ray C. Fair, "Efficient Estimation of Simultaneous Equations with Autoregressive Errors by Instrumental Variables," Review of Economics and Statistics 54 (1972): 444-9.
- ²³J.M. Brundy and D.W. Jorgenson, "Efficient Estimation of Simultaneous Equations by Instrumental Variables," Review of Economics and Statistics 53 (1971): 207-24.
- ²⁴Michio Hatanaka, "Several Efficient Two-Step Estimators for the Dynamic Simultaneous Equations Model with Autoregressive Disturbances," Journal of Econometrics 4 (1976): 189-204.
- ²⁵Michio Hatanaka, "An Efficient Two-Step Estimator for the Dynamic Adjustment Model with Vector Autoregressive Errors," Journal of Econometrics 2 (1974): 199-220.
- ²⁶Dhrymes and Taylor, "On an Efficient Two-Step Estimator," pp. 362-76.

²⁷This test may be asymptotically equivalent to the test outlined by Durbin appearing in J. Durbin, "Testing for Serial Correlation in Least-Squares Regression When Some of the Regressors are Lagged Dependent Variables," Econometrica 38 (1970): 410-21.

²⁸M. Nerlove and K.F. Wallis, "Use of the Durbin-Watson Statistic in Inappropriate Situations," Econometrica 34 (1966): 235-8.

²⁹Hendry, "Maximum Likelihood Estimation."

³⁰S.M. Goldfeld and R.E. Quandt, Nonlinear Methods in Econometrics (Amsterdam: North-Holland Publishing Company, 1972), p. 188.

³¹David K. Guilkey, "A Test for the Presence of First-Order Vector Autoregressive Errors When Lagged Endogenous Variables are Present," Econometrica 43 (1975): 711-17.

³²Goldfeld and Quandt, Nonlinear Methods.

³³Kmenta, Elements, p. 366.

³⁴Hendry, "Maximum Likelihood Estimation."

³⁵Kmenta, Elements, p. 370.

CHAPTER III

TEIGEN'S MODEL

Teigen formulated, estimated, and tested a small macro model of the monetary sector of the United States' economy. He produced his original results with end-of-the-quarter data,¹ but in response to Gibson's remarks,² revised his study by using quarterly-average data.³ The model includes a reduced form equation for income and equations for the demand for and supply of money, and incorporates a role for the commercial banking system in determining the money stock. When interest rates in the money markets make it advantageous to do so, member banks should acquire reserves, make loans, and expand the money stock. Member banks can obtain the additional reserves by exercising their discounting privileges with the Federal Reserve System, by entering the Federal Funds market, or by selling securities in their portfolios. Therefore, interest rate are embodied in both the money demand and supply equations, and the exogeneity of the money stock becomes a dubious assumption. Consequently, Teigen assumed that the Federal Reserve System controls only the reserves it supplies and incorporated the money stock into his model as an endogenous variable. The amount of reserves supplied by the Federal Reserve System, not the money stock, is the exogenous variable. This feature of Teigen's model contrasts with nearly all previous and many later studies on the demand-for-money function that treat the money stock

as an exogenous variable. Teigen's approach removed the potential simultaneous equations bias that permeates estimates of the demand-for-money function in single-equation models. Teigen's estimates can therefore be used to improve the estimated interest and income elasticities for the demand for and supply of money, if the specification of a simultaneous equations model is necessary.

Structure of Teigen's Model

The supply-of-money function reflects the behavior of commercial banks in the following fashion. When the profitability of loans rises, other relevant factors kept equal, banks will adjust their portfolios toward loans. By expanding loans and demand deposits, the banks cause the actual money stock to rise relative to the potential money stock. The potential money stock is determined by the amount of reserves supplied by the Federal Reserve System and the amount of member bank borrowing, along with banking rules and behavioral relationships. When the profitability of loans falls, banks will adjust their portfolios away from loans and cause the actual money stock to fall relative to the potential money stock. The supply-of-money function takes the form

$$(M/M^*) = X(\text{profitability of loans})$$

where

M is the observed money stock;

M* is the potential money stock based on reserves supplied by the Federal Reserve System; it is not the potential money stock from all reserves since it excludes reserves obtained through member bank borrowing; and

M/M^* is the extent to which the potential money stock (based on Federal Reserve System supplied reserves) is realized. Teigen then approximated the profitability of loans for commercial banks as the difference between the return from lending and the cost of lending. The interest rate on four to six month commercial paper measures the return to commercial banks from lending. The cost of lending is the opportunity cost of using excess reserves for commercial loans instead of loaning excess reserves to other banks; this cost could be measured by the Federal Funds rate. Alternatively, the cost of lending is the cost of acquiring additional reserves, which can be measured by the Fed's discount rate or the Federal Funds rate. In his study, Teigen chose to use the discount rate.

Teigen based the demand-for-money function on the square root inventory rule for the transactions demand for money. Baumol and Tobin developed this idea over twenty years ago. According to this rule, the desired average amount of money held for transactions depends on the level of income, interest rates, institutional factors, and the transactions costs involved in shifting money in and out of liquid, interest earning assets until it is needed. Furthermore, Teigen assumed that a partial adjustment or habit persistence mechanism determines the actual demand for money. The demand-for-money equation then takes the form

$$M = M[\text{income, interest rates, } M(-1)]$$

where M is the actual (not the desired) level of money demanded, and transactions costs and institutional factors are assumed to be constant.

Finally, the reduced form equation for income is derived from the simple Keynesian model that Teigen developed.⁴ This equation takes the form

$$Y = Y[E, NW, Y(-1)]$$

where

E is exogenous spending - investment, exports, and government expenditures; and

NW is net worth.

Econometric Issues

Teigen estimated the following form of his model:

1. Demand for money

$$M = \alpha_0 + \alpha_1(r \cdot Y) + \alpha_2 Y + \alpha_3 M(-1) + \alpha_4 S_1 + \alpha_5 S_2 + \alpha_6 S_3 + u_1$$

2. Supply of money

$$(M/M^*) = \beta_0 + \beta_1(r - r_D) + \beta_2 S_4 + u_2$$

3. Income

$$Y = \gamma_0 + \gamma_1 E + \gamma_2 NW + \gamma_3 Y(-1) + \gamma_4 [S_1 \cdot Y(-1)] \\ + \gamma_5 [S_2 \cdot Y(-1)] + \gamma_6 [S_3 \cdot Y(-1)] + u_3$$

where S_1 , S_2 , and S_3 are seasonal binary variables, and S_4 is a structural shift binary variable which takes on the value one for the period from 1953-I to 1958-IV.

The endogenous variables are the money stock, the interest rate, and income. Therefore, the model does take into account the simultaneous determination of the interest rate and the money stock. The model is linear in its parameters, but nonlinear in its endogenous variables since it contains the endogenous functions $r \cdot Y$ and M/M^* . This characteristic of the model must be carefully considered during estimation.

Teigen estimated his model in the following manner:

1. Ordinary least squares method (OLS) was used for the income equation.
2. The fitted values for the endogenous variable Y and the endogenous functions $r \cdot Y$ and $r - r_D$ were obtained by regressing Y , $r \cdot Y$, and $r - r_D$ on all predetermined variables in the model; this corresponds to the first step of two-stage least squares (2SLS) estimation.
3. The money demand and supply equations were estimated by the OLS method after the actual values of Y , $r \cdot Y$, and $r - r_D$ had been replaced by their fitted values; this corresponds to the second stage of 2SLS estimation.

Concerning Teigen's econometric methods, two major objections can be registered. First, the model's nonlinearity in endogenous variables is ignored. For example, the reduced form equation for the interest rate (r) is a nonlinear function of the predetermined variables in the monetary sector - $\hat{Y}$, $M(-1)$, M^* , r_D , S_1 , S_2 , S_3 , S_4 :

$$r = \frac{-\alpha_0 - \alpha_1 \hat{Y} - \alpha_3 M(-1) - \alpha_4 S_1 - \alpha_5 S_2 - \alpha_6 S_3 + \beta_0 M^* - \beta_1 (r_D \cdot M^*) + \beta_2 (S_4 \cdot M^*)}{(\alpha_2 \hat{Y} - \beta_1 M^*)} + \frac{u_1 + (u_2 \cdot M^*)}{(\alpha_2 \hat{Y} - \beta_1 M^*)}$$

Teigen never worked with this equation, his model's reduced form equation for the interest rate, in order to obtain fitted values for r . Instead he estimated a linear equation which includes the monetary sector's predetermined variables:

$$r = \delta_0 + \delta_1 \hat{Y} + \delta_2 M(-1) + \delta_3 M^* + \dots + (\text{disturbance})$$

The disturbance term in this linear equation represents the combined product of the model's structural disturbances and the misspecification of the nonlinear reduced form equation. While using fitted values obtained from estimates of the linear equation in the second stage of 2SLS estimation - in which the OLS method is applied to the equation $y_g = \hat{Y}_g \beta_g + X_g \gamma_g + u_g^*$ where u_g^* is the composite disturbance - Teigen cannot correctly claim that the equation's disturbance u_g^* is uncorrelated with each of the explanatory variables contained in $\hat{Y}_g$ and X_g . Since the composite disturbance term u_g^* contains the effect of the misspecification of the reduced form equation, it will be correlated with the equation's predetermined variables, X_g . Consequently, his estimates lack all desirable properties and his tests of hypotheses are invalid.

Second, Teigen assumed that the model's disturbances are nonautocorrelated. This may have been a defensible assumption for his original study. In that study he used data for which all stock variables are measured as of each quarterly call date; that is, measured in points of time roughly 90 days apart. This implies that the effect of factors lumped into the regression disturbances must last from 91 to 180 days if these factors are to influence two successive observations and cause correlation between two successive disturbances. It is possible that the effects of these small, random factors in the model's disturbance term are sufficiently temporary for Teigen's assumption to be valid. In the later version of his study, however, he used data that are quarterly averages of daily figures. This implies that the effects of the small, random

factors must last from only 1 to 90 days to influence two successive observations and cause autocorrelated disturbances. Thus, Teigen's assumption of nonautocorrelated disturbances for the revised study is likely to be invalid.⁵

Our reestimation of Teigen's model with quarterly-average data will account for the model's nonlinearities and will incorporate the assumption of autoregressive disturbances. Reestimation will involve the application of Kelejian's nonlinear 2SLS method,⁶ modified for simple autoregressive disturbances. Kelejian's method corresponds to the usual 2SLS estimation procedure with only the reduced form adapted for the model's nonlinear structure. The usual 2SLS method requires a consistent estimate of the reduced form equations in order to obtain fitted values for the endogenous variables. These fitted values replace the actual values of the endogenous variables among the explanatory variables to purge their correlation with the structural disturbances. This correlation is the cause of inconsistency in the OLS estimates. When a model is nonlinear, however, expressions for the endogenous variables and endogenous functions in terms of just the predetermined variables and predetermined functions can become highly complex. Therefore, fitted values can be obtained from the reduced form only with great difficulty - and often not at all. In place of estimating the reduced form itself, Kelejian proposes to estimate a polynomial of the model's predetermined variables that approximates the reduced form and from which fitted values for the endogenous variables and functions can be obtained. These fitted values will be uncorrelated with the

structural disturbances and the second stage estimates will be consistent as the sample size tends to infinity, provided that the degree of the polynomial is allowed to increase with sample size. That is, let the degree of the polynomial increase at the rate $(T)^k$ where $0 < k < 1$, so that as the sample size (T) tends to infinity so will the degree of the polynomial. The value of k , however, must be sufficiently small in order that the number of parameters to be estimated from the polynomial grows at a slower rate than the increase in sample size.

New Estimates

Several new estimates of Teigen's model are reported in table 1. The first set of estimates is Teigen's. The second set of estimates was made with the OLS method, adapted for simple autoregressive disturbances (OLS-AUTO). Since this method ignores simultaneity, it is unsatisfactory, but its estimates are included for comparison. The third set of estimates is given by the nonlinear 2SLS method, adapted for simple autoregressive disturbances (2SLS-AUTO). In the application of this method, fitted values for the endogenous functions $r-r_D$ and $r \cdot Y$ were obtained by estimating a truncated second degree polynomial of the predetermined variables in the transformed version of the monetary sector. Truncating the polynomial approximating the reduced form is necessary because this procedure results in a greater number of terms than observations in the sample. Fortunately, the findings of Goldfeld and Quandt suggest that a quadratic polynomial is adequate for Teigen's model. From sampling experiments with two nonlinear models, Goldfeld and

TABLE 1
ESTIMATES OF STRUCTURAL COEFFICIENTS

Demand for Money Equation

	Constant	r·Y	Y	M(-1)	S ₁	S ₂	S ₃	ρ_{11}	ρ_{12}	ρ_{13}	R ²	DW
Teigen (1974)	12.8 (3.7)	-.0023 (.0006)	.037 (.009)	.83 (.05)	-1.65 (.52)	-3.76 (.44)	-1.49 (.35)	...	...	...	.995	NR ^a
OLS-AUTO	20.9 (5.0)	-.0022 (.0007)	.057 (.011)	.70 (.05)	-.59 (.55)	-3.07 (.44)	-1.11 (.12)	.55 (.12)	...	...	.997	1.95
Nonlinear 2SLS-AUTO	16.7 (3.8)	-.0027 (.0006)	.054 (.009)	.75 (.05)	-.89 (.45)	-3.31 (.38)	-1.32 (.25)	.37 (.14)	...	...	.997	1.78
Nonlinear 2SLS (RA)-VECTOR	20.4 (4.8)	-.0024 (.0006)	.061 (.011)	.69 (.07)	-.41 (.56)	-2.98 (.47)	-1.12 (.28)	.45 (.14)	-2.06 (3.32)	-.053 (.035)	.996	...

Supply of Money Equation

	Constant	r-r _D	S ₄	ρ_{21}	ρ_{22}	ρ_{23}	R ²	DW
Teigen (1974)	.85 (.006)	.094 (.015)	.016 (.008)	...	...	...	.578	NR
OLS-AUTO	.86 (.011)	.058 (.020)	.011 (.014)	...	.51 (.13)	...	.578	1.98
Nonlinear 2SLS-AUTO	.86 (.011)	.059 (.019)	.009 (.014)	...	.52 (.13)	...	.584	2.07
Nonlinear 2SLS (RA)-VECTOR	.86 (.011)	.067 (.019)	.011 (.014)	.004 (.006)	.50 (.13)	.0018 (.0013)	.429	...

TABLE 1 - Continued

Income Equation

	Constant	E	NW	Y(-1)	S ₁ Y(-1)	S ₂ Y(-1)	S ₃ Y(-1)	ρ_{31}	ρ_{32}	ρ_{33}	R ²	DW
Teigen (1974)	37.3 (3.2)	.89 (.08)	.096 (.010)	.26 (.06)	-.100 (.006)	-.055 (.004)	-.066 (.004)	...	...	...	.998	1.41
OLS-AUTO	39.8 (4.1)	.92 (.08)	.102 (.012)	.21 (.07)	-.097 (.006)	-.056 (.004)	-.067 (.003)	...	...	.32 (.14)	.998	2.13
Nonlinear 2SLS (RA)-VECTOR	39.2 (4.2)	.90 (.08)	.105 (.012)	.22 (.07)	-.097 (.006)	-.058 (.004)	-.067 (.003)	1.29 (.69)	30.82 (16.47)	.26 (.17)	.998	...

^aNR means not reported

Quandt show that Kelejian's nonlinear 2SLS estimation utilizing a quadratic polynomial "uniformly outperforms" standard 2SLS utilizing a first-degree polynomial. Furthermore, they see "no reason to expect continued improved performance" as the degree of the polynomial increases over two.⁷ If the assumption of simple autoregressive disturbances is valid, then this estimation method would possess the desirable asymptotic properties and would generate the preferred estimates for Teigen's model. (The inference that the preferred estimates come from a method with the desirable asymptotic properties will be challenged later in the sampling experiments.) Finally, the fourth set of estimates was obtained from Hatanaka's limited-information, residual adjusted estimation method for models with vector autoregressive disturbances (2SLS(RA)-VECTOR) after it was altered along the lines of nonlinear 2SLS so that the model's nonlinearities were recognized. If the assumption of vector autoregressive disturbances is valid, then this method would possess the desirable asymptotic properties and the preferred estimates of Teigen's model would be obtained from this estimation method.

The estimates of the autoregression parameter and its standard error from both the OLS-AUTO method and the nonlinear 2SLS-AUTO methods cause us to reject the null hypothesis of nonautoregressive disturbances for all three equations. In each instance, the estimate of the autoregression parameter is at least twice as great as its estimated standard error. The nonlinear 2SLS(RA)-VECTOR method's estimates of the R matrix, however, do not cause us to reject the null hypothesis of simple autoregressive disturbances in favor of

the alternative hypothesis of vector autoregressive disturbances, according to the third test (the test for additional relevant explanatory variables) described in chapter II. The calculated value of the test statistic is 2.27, which is less than the critical value of 2.44 (at the .05 level of significance). Consequently, the estimates of Teigen's model obtained from the nonlinear 2SLS-AUTO method will be used in the empirical work that follows.

Among these four sets of estimates of the money demand and supply equations, there exists some variation in the estimates of the explanatory variables' coefficients. The sharpest contrasts between Teigen's estimates and estimates based on the assumption of simple autoregressive disturbances are found in the estimates of the income (Y) coefficient in the demand equation and the profitability of loans ($r - r_D$) coefficient in the supply equation. Also, the coefficient of the structural variable S_4 becomes insignificant when estimated under the assumption of simple autoregressive disturbances. In comparison to Teigen's findings, the estimates from the nonlinear 2SLS-AUTO method indicate that the demand for money is more sensitive to income and the interest rate and adjusts more slowly to the desired level; and the supply of money is less sensitive to profitability. Furthermore, a comparison of the estimates obtained for the autoregression parameters shows substantial differences; for example, the estimates of ρ_{11} range from .37 to .55.

Empirical Results

From the coefficient estimates obtained from the nonlinear 2SLS-AUTO method, new estimates of demand and supply elasticities can be calculated and compared to Teigen's.

TABLE 2
DEMAND ELASTICITIES

The Elasticity of the Demand for Money with Respect to (Computed at Means)	<u>Teigen</u>		<u>Radecki</u>	
	<u>Short Run</u>	<u>Steady State</u>	<u>Short Run</u>	<u>Steady State</u>
r	-.014	-.081	-.016	-.065
Y	.089	.517	.132	.527

Relative to the results obtained from the nonlinear 2SLS-AUTO method, Teigen overestimated demand's steady-state interest elasticity by 25 per cent and underestimated its short-run interest elasticity by 12 per cent, its short-run income elasticity by 48 per cent, and its steady-state income elasticity by 2 per cent.

TABLE 3
SUPPLY ELASTICITIES

The Elasticity of the Supply of Money with Respect to (Computed at Means)	<u>Study</u>	
	<u>Teigen</u>	<u>Radecki</u>
r	.232	.146
r _D	-.202	-.127

Making the same comparison, Teigen overestimated the supply elasticities by 59 per cent. From these comparisons, it is obvious

that in several instances Teigen's estimates of the elasticities and the new estimates contrast sharply. These differences are attributable to the alternative treatment of the model's nonlinearities and the autoregressive disturbances. The new estimates of the interest rate elasticity of the demand and supply of money imply that monetary policy is more effective and fiscal policy is less effective than indicated by Teigen's results.

Besides recomputing the estimates of the elasticities, the data also allow a new comparison of single-equation and structural estimates of the demand and supply elasticities. In his original study, Teigen reported that single-equation, steady-state estimates of

1. demand's elasticity with respect to the interest rate are biased downwards in absolute value by 7 per cent;
2. demand's elasticity with respect to income are unbiased; and
3. supply's elasticities were biased downwards by 41 per cent.

Using the estimates from the OLS-AUTO method, the results presented in table 4 were obtained. In contrast to Teigen's findings, the demand elasticities calculated from the single-equation method seem to contain substantial simultaneous equations bias, while the supply elasticities are relatively free of such bias.

An implicit assumption contained throughout this chapter is that the application of the preferred estimation method (nonlinear 2SLS-AUTO) will improve the estimates of the model's parameters and the elasticities. Theoretically, the preferred method possesses the

TABLE 4

SINGLE EQUATION AND STRUCTURAL ESTIMATES OF
STEADY-STATE ELASTICITIES

<u>The Elasticity of the Demand for Money with Respect to (Computed at Means)</u>	<u>Single- Equation</u>	<u>Structural</u>	<u>Simultaneous Equations Bias</u>
r	-.044	-.065	downwards, 48%
Y	.484	.527	downwards, 9%
<u>The Elasticity of the Supply of Money with Respect to (Computed at Means)</u>			
r	.143	.146	downwards, 2%
r _D	-.125	-.127	downwards, 2%

desirable asymptotic properties; however, a finite sample of 48 observations is used to conduct the reestimation work, whose results are reported above. It is possible that in small samples estimation methods lacking the desirable asymptotic properties could outperform methods that do possess the desirable asymptotic properties. It is unclear whether the nonlinear 2SLS-AUTO method generates superior estimates of Teigen's model. The results from sampling experiments designed to compare the small sample performances of the estimators will be reported in chapter V and hopefully will remove these ambiguities (to the extent to which sampling experiments are capable of doing so).

As an alternative to the sampling experiments, the ability to forecast can be used to compare the estimates of Teigen's model. Presumably, more accurate predictions of the endogenous variables will reflect better estimates of the model's parameters and the demand and supply elasticities. Table 5 presents the predictions obtained from Teigen's estimates and the nonlinear 2SLS-AUTO estimates for the endogenous variables - the interest rate, the money stock, and income - for the four quarters beyond the sample period. These predicted values were obtained dynamically, which means that predicted rather than actual values were used for the lagged endogenous variables. The forecasting performances of the two sets of estimates are compared in table 6. On this basis, neither set of estimates can be considered superior. A judgment must be postponed until the results from the sampling experiments are reported.

TABLE 5

PREDICTED VALUES OF THE ENDOGENOUS VARIABLES

	<u>Quarters</u>			
	<u>I</u>	<u>II</u>	<u>III</u>	<u>IV</u>
Actual r	3.30	3.60	4.19	4.76
Teigen's r	3.28	3.99	4.99	4.67
Nonlinear 2SLS-AUTO r	3.60	4.21	4.50	4.80
Actual M	143.03	142.36	143.00	144.83
Teigen's M	142.28	140.36	140.46	142.77
Nonlinear 2SLS-AUTO M	140.10	139.83	139.73	142.10
Actual Y	452.4	484.6	477.2	519.6
Teigen's Y	452.9	485.6	481.2	518.9
Nonlinear 2SLS-AUTO Y	450.8	483.9	478.9	517.1

TABLE 6

EVALUATION OF THE PREDICTIONS

	<u>Sum of Errors</u>	<u>Sum of Squared Errors</u>
Teigen's r	1.30	.79
Nonlinear 2SLS-AUTO r	1.26	.56
Teigen's M	7.4	15.4
Nonlinear 2SLS-AUTO M	9.5	25.5
Teigen's Y	6.2	17.7
Nonlinear 2SLS-AUTO Y	7.5	14.6

CHAPTER III

FOOTNOTES

¹Ronald L. Teigen, "Demand and Supply Functions for Money in the United States: Some Structural Estimates," Econometrica 32 (1964): 476-509.

²William E. Gibson, "Demand and Supply Functions for Money in the United States: Theory and Measurement," Econometrica 40 (1972): 361-70.

³William E. Gibson, "Demand and Supply Functions for Money: A Comment," Econometrica 44 (1976): 387-9; Ronald L. Teigen, "Demand and Supply Functions for Money: Another Look at Theory and Measurement," Econometrica 44 (1976): 377-85.

⁴Teigen, "Demand and Supply Functions for Money in the United States," p. 507.

⁵Gibson also assumed nonautoregressive disturbances for his model while working with quarterly-average data.

⁶Harry H. Kelejian, "Two-Stage Least Squares and Econometric Systems Linear in Parameters but Nonlinear in the Endogenous Variables," Journal of the American Statistical Association 66 (1971): 573-4; also described in S. Goldfeld and R.E. Quandt, Nonlinear Methods in Econometrics (Amsterdam: North-Holland Publishing Company, 1972), pp. 232-8.

⁷Goldfeld and Quandt, Nonlinear Methods, Chapter 8.

⁸Of course, this may not be true if the model is misspecified or its coefficients undergo a structural shift.

CHAPTER IV

PERRY'S MODEL

Perry investigated the process of inflation by designing, estimating, and analyzing a small macro model of the United States economy's wage-price-profit subsystem.¹ The model's focus is on the quarterly behavior of wages; however, Perry elaborated from a single-equation model because he recognized that wage rates interact with prices and profit rates. The model's structure consists of three equations and an identity that explain changes in profit rates, manufacturing output prices, and aggregate wages.² Changes in aggregate wages can be partly explained by the average profit rate and its first difference; in general, labor can be expected to obtain a greater than average increase in wages during periods of abnormally high or rising profits. Suppose, as Perry did, that prices in manufacturing industries are determined by a markup process over costs. Since wages are a component of a firm's costs, a change in wage rates can be expected to affect a firm's price for its product, and also its profits. The behavior of wage rates is thus simultaneously determined with the behavior of prices and profits. Consequently, wages, prices, and profits appear as endogenous variables in a structural model. The incorporation of the aggregate wage function, or Phillips "curve," in a simultaneous equation model distinguishes Perry's work from many other studies that

ignore the feedback effect of wages on costs and profits. This approach removes the potential simultaneous equation bias that could spoil estimates of aggregate wage behavior in single-equation models. Estimates of Perry's model can be used to reexamine the Phillips "curve" and, since the model is dynamic, to analyze the inflationary process through time.

Structure of the Model

The wage equation explains the percentage change in the average wage rate over the past year. The change in the wage rate is a function of the unemployment rate, the change in the cost of living, the rate of profits, and expectations for the rate of profits. The unemployment rate reflects the willingness of employers to grant wage increases and indicates the probable success of gaining wage increases. Periods of low unemployment are associated with large wage increases because

1. employers will be anxious to avoid strikes, since periods of low unemployment correspond to periods of strong consumer demand;
2. employers will be more able to pass on higher labor costs when consumer demand is strong; and
3. certain employers in the tight labor market, characterized by excess demand, may be trying to bid labor away from other employers.

The change in the cost of living affects the size of the wage increase that results because a rise in the cost of living contributes to the strength of labor's demands and to the pressure on the union from its members to obtain a wage increase. The current and expected

average profit rate indicates the ability of employers to pay wage increases now and in the future; also, the current profit rate measures the cost to the firm of withstanding a strike. Therefore, a high current or expected profit rate lowers employers' resistance to a wage increase and allows employees to obtain large wage increases. The change in the profit rate is used as a proxy for the expected profit rate since increases in profits cause unions to expect higher future profits. The wage equation that is estimated is the following:

$$w(t) = \alpha_0 + \alpha_1[U(t)]^{-1} + \alpha_2c(t-1) + \alpha_3R(t-1) + \alpha_4[\Delta R(t)] + u_1(t)$$

where

$w(t)$ is the percentage change in the rate of straight-time average hourly earnings over the past year (four quarters);

$U(t)$ is the average percentage unemployment rate during the past year;

$c(t)$ is the percentage change in the consumer price index over the past year;

$R(t)$ is the average percentage rate of return on stockholders' equity over the past year; and

$\Delta R(t) = R(t) - R(t-1)$, the first difference of $R(t)$.

The price equation explains the percentage change in the index of output prices for the manufacturing sector over the past year. It is based on a modified version of the markup theory of oligopoly pricing. Perry maintained that output prices are

basically determined by a pure markup over unit costs, which are a function of the prices of variable inputs and the extent to which fixed inputs are utilized. Labor's wages and the prices of raw materials are considered to be the main items in variable costs; the change in the index of capacity utilization indicates movements along a firm's short run cost curves. The markup of prices over unit costs, however, is significantly influenced by the strength of consumer demand, which is represented by an index of capacity utilization. Presumably, greater increases in output prices will occur during periods of pressure on capacity due to an increase in the markup over costs. Perry estimated the output price equation in the following form:

$$p_m(t) = \beta_0 + \beta_1 w(t) + \beta_2 p_r(t) + \beta_3 [\Delta(Q/K)(t)] \\ + \beta_4 (Q/K)(t) + u_2(t)$$

where

$p_m(t)$ is the annual percentage change in the index of manufacturing output prices;

$p_r(t)$ is the annual percentage change in the index of raw material prices;

$(Q/K)(t)$ is the average over the past year of the index of capacity utilization in manufacturing; and

$\Delta(Q/K)(t) = (Q/K)(t) - (Q/K)(t-1)$ is the first difference in (Q/K) .

The third equation explains the change in the average profit rate in terms of changes in unit cost, capacity utilization, and the price of output. As in the previous equation, changes in

labor's wages and the prices of raw materials reflect changes in a firm's variable costs. Changes in the price of manufactured output cause changes in a firm's revenue. Changes in the level of capacity utilization capture the effect on profits of a) changes in unit costs from the degree to which fixed costs are utilized, and b) changes in the level of output and sales. Perry estimated the profit equation in the following form:

$$\begin{aligned}\Delta R(t) = & \gamma_0 + \gamma_1 w(t-1) + \gamma_2 p_r(t-1) + \gamma_3 p_m(t) \\ & + \gamma_4 [\Delta(Q/K)](t) + u_3(t)\end{aligned}$$

where all variables are previously defined. The identity concerning the profit rate, $\Delta R(t) = R(t) - R(t-1)$, completes the model.

Econometric Issues

Perry estimated his simultaneous equation model of the wage-price-profit subsystem with the OLS method. He justified the use of the OLS estimation method on the grounds that:

1. the OLS estimates, by his conjecture, contain little simultaneous equation bias.³ This notion is based on a comparison of OLS and 2SLS estimates of the model after all lagged endogenous variables are made current.
2. the 2SLS method could not remove from OLS estimates the source of a bias more serious than simultaneity.⁴ Perry believed that the bias in the OLS estimates is for the most part the result of having autoregressive disturbances in a model with lagged endogenous variables. Consequently, applying the 2SLS estimation method would fail to remedy the problem.

Perry's approach to the estimation of his model must be considered inadequate for several econometric reasons. First, there is no basis for the judgment that the simultaneous equation bias is so insignificant that it is unnecessary to use a simultaneous equation estimator - such as the 2SLS method - since

1. this conclusion was based on estimates of essentially a different model than the one under consideration.
2. the cost of obtaining 2SLS estimates for this three equation model is minimal. To conclude that the additional cost of producing 2SLS estimates that remove simultaneous equation bias outweighs the benefits from doing so suggests that the work associated with the model may be of little value or that the problem under study is inconsequential.
3. a small numerical change in a coefficient estimate could potentially alter the stability of the model's oscillations or imply substantial changes for public policy.

Besides, the claim that the simultaneous equation bias is insignificant weakens the credibility of the specification of a simultaneous equation model to explain aggregate wage behavior. A second fault with Perry's approach to estimation relates to the problem of autoregressive disturbances. Since Perry felt that autoregressive disturbances were a much more significant source of bias than was simultaneity, he could have used this knowledge by employing a single-equation estimation method that corrects for autoregressive disturbances. Third, if autoregressive disturbances are present in a structural model, then what is called for is an econometric method

designed for the compound estimation problem. If Perry had reason to prefer the OLS method to more complicated estimators - perhaps because he believed that the OLS method would provide estimates with the least mean square error (MSE) in the context of a structural model with autoregressive disturbances, lagged endogenous variables, possible errors of measurement, doubts concerning the third equation's specification, and a finite sample of only 51 observations - then he failed to make that clear.

New Estimates

Table 7 presents three sets of estimates for Perry's model. The first set of estimates is Perry's, obtained from the OLS method. These estimates suffer from the problems mentioned above. Estimates from the 2SLS method and from the OLS method adapted for simple autoregressive disturbances (OLS-AUTO) were also calculated; these estimates will not be reported here because the estimation methods ignore a violation of a basic assumption. It should be mentioned, however, that if Perry employed the OLS-AUTO method (the method that should have been used if one believed that the model's disturbances are simple autoregressive and the effect of simultaneity is negligible) he would have obtained highly significant estimates of each equation's autoregression coefficient ($\hat{\rho}_1 = .43$, $\hat{\rho}_2 = .72$, and $\hat{\rho}_3 = .62$) that suggest a high degree of autoregression in the model's disturbances. The 2SLS method adapted for simple autoregressive disturbances (2SLS-AUTO) provides the second set of estimates.⁵ If the assumption of simple autoregressive disturbances is valid, then this estimation method would possess the desirable asymptotic

TABLE 7
ESTIMATES OF STRUCTURAL COEFFICIENTS

Wage Equation	Constant	c(-1)	(U) ⁻¹	R(-1)	ΔR	ρ_{11}	ρ_{12}	ρ_{13}	DW	R ²
Perry	-4.31 (NR)	.37 (.05)	14.71 (2.2)	.42 (.11)	.79 (.17)	...	...	...	NR ^a	.87
2SLS-AUTO	-5.01 (1.21)	.38 (.08)	15.25 (3.0)	.48 (.10)	.91 (.23)	.43 (.13)	...	...	1.92	.87
2SLS(RA)-VECTOR	-5.94 (1.77)	.34 (.12)	14.03 (3.9)	.59 (.16)	.88 (.31)	.44 (.14)	.091 (.13)	.073 (.40)	...	.83
Price Equation	Constant	w	P _F	Δ(Q/K)	(Q/K)	ρ_{21}	ρ_{22}	ρ_{23}	DW	R ²
Perry	-15.03 (NR)	.47 (.12)	3.44 (.23)	-.61 (.12)	.15 (.05)	...	...	...	NR	.92
2SLS-AUTO	-10.60 (9.4)	.55 (.27)	2.91 (.29)	-.47 (.13)	.10 ^b (.10)	...	.74 (.10)	...	1.58	.96
2SLS(RA)-VECTOR	-11.96 (5.6)	.38 (.19)	3.16 (.26)	-.58 (.12)	.12 (.06)	-.015 (.17)	.37 (.14)	-1.19 (.45)	...	.90

TABLE 7 - Continued

Profit Equation	Constant	w(-1)	P _r (-1)	P _m	$\Delta(Q/K)$	ρ_{31}	ρ_{32}	ρ_{33}	DW	R ²
Perry	.068 (NR)	-.069 (.032)	-.36 (.11)	.10 (.03)	.25 (.03)	...	...	...	NR	.73
2SLS-AUTO	.122 (.18)	-.106 (.036)	-.41 (.09)	.13 (.03)	.24 (.03)	...	...	.58 (.12)	1.71	.89
2SLS(RA)-VECTOR	.297 (.21)	-.163 (.040)	-.41 (.10)	.18 (.03)	.21 (.03)	.14 (.06)	-.04 (.04)	.44 (.15)	...	.66

^aNR means not reported

^binsignificant structural coefficients at the .95 level of significance

properties and would generate the preferred estimates of Perry's model. In fact, the estimates of the autoregression coefficient and its standard error allow us to reject the null hypothesis of nonautoregressive disturbances for all three equations. In each case, the ratio of the coefficient estimate to its standard error is greater than two. The last set of estimates is given by the limited-information version of Hatanaka's residual adjusted method designed for models with vector autoregressive disturbances (2SLS(RA)-VECTOR). If the assumption of vector autoregressive disturbances is correct, then this estimation method possesses the desirable asymptotic properties and produces the preferred estimates of Perry's model. The estimates obtained for the matrix R, the matrix of autoregression coefficients, and the standard errors do lead us to reject the null hypothesis of simple autoregressive disturbances in favor of the alternative of vector autoregressive specification. The estimates of two off-diagonal elements of the R matrix, ρ_{31} and ρ_{23} , are significant; that is, their estimates are more than twice their standard errors. Consequently, the 2SLS(RA)-VECTOR estimates of Perry's model will be used in the empirical work that will follow. The estimates obtained from the 2SLS-AUTO method will also be used. These estimates are included because

1. the available evidence indicates that they may be as good or better than the 2SLS(RA)-VECTOR estimates obtained from a finite sample; and
2. Perry was uncertain about the third equation's specification, and the estimates from a method designed for vector autoregressive

disturbances might be especially sensitive to specification errors; it happens that the significant elements of the R matrix are associated with the third equation.

Empirical Results

From the coefficient estimates acquired from the 2SLS-AUTO and 2SLS(RA)-VECTOR methods, new estimates can be calculated for the wage change-unemployment-inflation relationship. In order to estimate the tradeoff between the rate of unemployment and the rates of wage and price inflation, more commonly called the Phillips curve, first assume that the rate of change of prices is equal to the rate of change in wages less the rate of increase in labor's productivity; that is, $c(t) = w(t) - p(t)$, where $p(t)$ is the rate of labor productivity increase. For the work that follows, the rate of productivity growth will be set equal to 3 per cent, roughly the average rate in the United States during the period covered by Perry. Furthermore, assume that the relative shares of income going to labor and capital are constant over time; this means that the variable $\Delta R(t) = 0$. Finally, the long-run steady state relationship will be investigated; lags in the explanatory variables will thus be ignored. With different profit and unemployment rates, the results presented in table 8 are found. The three Phillips curves based on the OLS, 2SLS-AUTO and 2SLS(RA)-VECTOR estimates look quite similar. The differences that can be detected are the following:

1. The Phillips curve based on the 2SLS-AUTO estimates implies the highest rates of inflation for a low level of unemployment;

TABLE 8
PHILLIPS CURVES

Percentage Rate of Price Change ^a				
<u>Profit Rate</u>	<u>Unemployment Rate</u>	<u>Perry</u>	<u>2SLS-AUTO</u>	<u>RA-VECTOR</u>
10%	3%	2.88	2.99	2.43
	4%	.94	.95	.66
	5%	-.23	-.28	-.40
	6%	-1.00	-1.09	-1.10
	7%	-1.56	-1.68	-1.61
11%	3%	3.55	3.76	3.32
	4%	1.61	1.72	1.55
	5%	.44	.49	.49
	6%	-.33	-.32	-.21
	7%	-.89	-.91	-.72
12%	3%	4.22	4.53	4.21
	4%	2.28	2.49	2.44
	5%	1.11	1.26	1.38
	6%	.34	.45	.68
	7%	-.22	-.14	.17

^a(Rate of wage change) = 3 + (Rate of price change)

it is also the most steeply sloped of the three curves.

2. The Phillips curve based on 2SLS(RA)-VECTOR estimates is the least sloped of the three curves, but is the most sensitive to changes in the profit rate; it also implies the lowest rates of inflation at a low level of employment; and
3. The Phillips curve calculated from Perry's estimates is the least sensitive of the three curves to the profit rate.

As stated above, these distinctions are visible, but not necessarily significant. There appears to be no basis to favor one set of estimates over another; the three Phillips curves look equally plausible.

The various sets of estimates can also be used to determine the dynamic properties of the model. Assuming, as before, that

- a) the distribution of income to labor and capital is constant, and
- b) prices change at the rate at which wages change less the increase in labor's productivity, the model can be reduced to a third order difference equation in the wage variable. When the wage-price-profit subsystem is tested for stability, each set of estimates reveals that the model is stable, which means that a shock to the system will dampen out.

According to the tests for the specification of the model's disturbances, the 2SLS(RA)-VECTOR method should be the preferred estimator, in the sense that this method possesses the desirable asymptotic properties. As mentioned in regard to Teigen's model, because a finite sample is all that is available, an estimation method inferior in the sense of lacking the desirable asymptotic

properties (methods such as OLS and 2SLS-AUTO) could outperform the preferred method (2SLS(RA)-VECTOR when applied to Perry's model. As an alternative to the sampling experiments, the various sets of estimates can be compared on their ability to simulate the actual values of the endogenous variables. Shown below in table 9 are simulated values for the rate of wage change obtained from OLS, 2SLS-AUTO, and 2SLS(RA)-VECTOR estimates for the last four quarters of the sample period. (A similar set of calculations was made in conjunction with Teigen's model, but was termed forecasting because it was carried out beyond the sample period.)

TABLE 9
SIMULATED VALUES FOR $w(t)$

	<u>Quarters</u>			
	<u>1959-IV</u>	<u>1960-I</u>	<u>1960-II</u>	<u>1960-III</u>
Actual $w(t)$	3.33	3.24	2.78	3.74
Perry's $w(t)$	3.46	3.67	3.55	3.66
2SLS-AUTO $w(t)$	3.24	3.45	3.26	3.25
2SLS(RA)-VECTOR $w(t)$	3.59	3.77	3.48	3.45

These simulated values were obtained dynamically using simulated rather than actual values for lagged endogenous variables. The accuracy of the three sets of simulated values is compared in table 10. On this basis, the 2SLS-AUTO estimates appear to be superior. This result is somewhat anticipated. The 2SLS-AUTO method should achieve greater simulation accuracy compared to the OLS method because there appears to be a fairly high degree of

autoregression in the model's disturbances; on the other hand, it was suspected that the specification errors might sufficiently damage the 2SLS(RA)-VECTOR estimates to cause this method's estimates to fall below the 2SLS-AUTO estimates in the rankings based on the ability to simulate the actual values of the endogenous variables.

TABLE 10
EVALUATION OF THE SIMULATIONS

	<u>Sum of Errors</u>	<u>Sum of Squared Errors</u>
Perry's $w(t)$	1.41	.801
2SLS-AUTO $w(t)$	1.27	.522
2SLS(RA)-VECTOR $w(t)$	1.76	.895

CHAPTER IV

FOOTNOTES

¹George L. Perry, Unemployment, Money Wage Rates, and Inflation (Cambridge, Mass.: The M.I.T. Press, 1966); a single-equation model is estimated in Chapter 2, the simultaneous equation model is estimated in Chapter 5.

²Perry also specified an additional equation; however, this equation contains no variable appearing in the other three equations. Hence, this equation need not be considered part of the model.

³Perry, Unemployment, pp. 96-7.

⁴Perry, Unemployment, pp. 98-9.

⁵The data provided by Perry in the appendix of his book contain an error: the decimal point should be moved one place to the left for the data on the variables p_r and p_f .

CHAPTER V

SAMPLING EXPERIMENTS

This chapter contains the results and analysis of a set of sampling experiments conducted with several simultaneous equation estimation methods and a two equation structural model in which autoregressive disturbances and lagged endogenous variables are present. The estimation methods selected for the experiments include those used to carry out the reestimation work reported in the earlier chapters - Amemiya's limited-information least squares estimator designed for models with simple autoregressive disturbances (2SLS-AUTO), and Hatanaka's limited-information residual adjusted estimator designed for models with vector autoregressive disturbances (2SLS(RA)-VECTOR). In addition, Fair's modification of Amemiya's method for undersized samples was selected for the experiments. The sampling experiments' first purpose is to compare the small sample performance of three estimators - ordinary least squares (OLS), OLS adapted for simple autoregressive disturbances (OLS-AUTO), and two-stage least squares (2SLS) - that fail to deal completely with the compound estimation problem of simultaneity, lagged endogenous variables, and autoregressive disturbances with the performance of those estimators, mentioned above, that do so, but which computationally are more burdensome. This investigation is called for because the desirable asymptotic properties of an estimator dealing completely with a

compound estimation problem may give no indication of its finite sample properties. In small samples it is possible for simpler estimation methods lacking the desirable asymptotic properties to outperform more complicated methods possessing these properties. The second purpose of the sampling experiments is to determine whether the t distribution is a reasonable approximation to the estimators' sampling distribution so that interval estimates and hypothesis tests based on the t distribution are reliable. In practice researchers working with small samples typically refer to the t distribution to calculate confidence intervals and to carry out tests of hypotheses even though this may be an inappropriate procedure. To fulfill these purposes, sampling experiments are used rather than a mathematical description of the estimators' sampling distributions. This is done because the derivation of the sampling distributions of simultaneous equation estimators in small samples is extremely complicated and exceedingly difficult. Basmann and others have succeeded in deriving the mathematical descriptions only for a small number of special cases.¹

Previous Studies

Other studies, whose goals have been to investigate the small sample performance of estimators of equation systems for which autoregressive disturbances were specified, have preceded the sampling experiments conducted as a part of this dissertation. Fair's work consisted of a comparison of the within sample forecasting accuracy of these six estimation methods: OLS, 2SLS, FIML (the maximum likelihood counterpart to 3SLS), OLS-AUTO, Fair's modified version of

Amemiya's 2SLS-AUTO, and FIML-AUTO (the maximum likelihood counterpart to Amemiya's 3SLS-AUTO).² Each estimation method was applied to Fair's eight equation macro forecasting model of the United States' economy and used to generate dynamic within sample predictions of GNP. According to the root mean square errors of the predictions, estimation methods dealing completely with the compound estimation problem (Fair's method and FIML-AUTO) provide more accurate predictions than the other estimation methods. There also appear to be greater gains in forecasting accuracy from using an estimation method that accounts only for autoregressive disturbances than from one that accounts only for simultaneity. Although the ability to forecast is a relevant criterion to assess various estimators, the generality of Fair's results is suspect since they are based on a single sample and a model whose true specification is unknown.

Hurd investigated the small sample properties of estimation methods for simultaneous equations with autoregressive disturbances through sampling experiments on the following model:

$$y(t) = \alpha z(t) + u(t)$$

$$z(t) = \beta x(t) + v(t)$$

where $u(t)$ follows a simple first-order autoregressive scheme - $u(t) = \rho u(t-1) + \varepsilon(t)$ - and $\varepsilon(t)$ and $v(t)$ are contemporaneously correlated.³ This correlation between $\varepsilon(t)$ and $v(t)$ causes the variable $z(t)$ to be correlated with the disturbance term of the equation in which it appears as an explanatory variable and makes OLS estimation of the first equation inconsistent. To carry out the experiments,

Hurd used the OLS, OLS-AUTO, 2SLS, and 2SLS-AUTO estimation methods; he compared their small sample properties according to the values of their mean square errors (MSE). His results show for samples composed of 10 to 30 observations that:

1. except for very few combinations of parameter values, the simultaneous equation estimation methods (2SLS and 2SLS-AUTO) have a larger mean square error than the single-equation methods (OLS and OLS-AUTO); and
2. when the degree of simultaneity is substantial ($\text{Cov}(\varepsilon, v) \geq .4$), the OLS-AUTO method has a smaller mean square error than the OLS method only if the value of ρ is large ($\rho \geq .7$).

Hurd cautiously concludes that the OLS method may be the best estimator in samples composed of 10 to 30 observations since it is usually profitable to ignore simultaneity and often profitable to ignore autoregression also. This means that the desirable asymptotic properties of the 2SLS-AUTO method do not carry over to finite samples composed of 10 to 30 observations. This conclusion, however, is based on a model with little resemblance to most models used by researchers. Hurd's model is recursive and its first equation contains no predetermined variables; that is, the first equation includes among its explanatory variables a single endogenous variable unaccompanied by either exogenous variables (not even a constant term) or lagged endogenous variables.

Goldfeld and Quandt elaborated Hurd's sampling experiments by specifying this more realistic structural model:

$$y_1(t) = \beta_1 y_2(t) + \beta_2 x_1(t) + \beta_3 x_2(t) + u_1(t)$$

$$y_2(t) = \gamma_1 y_1(t) + \gamma_2 x_3(t) + \gamma_3 x_4(t) + u_2(t)$$

for which the disturbances $u_1(t)$ and $u_2(t)$ follow the first-order simple autoregressive scheme.⁴ According to the MSE criterion, Goldfeld and Quandt compared the sampling distributions of these five estimation methods: OLS, OLS-AUTO, 2SLS, FIML, and FIML-AUTO. Their results show that for small samples ($T \leq 90$) there are usually significant gains in accuracy from using an estimation method that corrects for both simultaneity and autoregressive disturbances. This means that the desirable asymptotic properties do carry over to small samples, which is a finding that conflicts with Hurd's. When the degree of autoregression is high ($\rho_g = .9$) relative to the degree of simultaneity ($\beta_1 = \gamma_1 = .5$), however, there is little loss in accuracy from correcting only for autoregressive disturbances. While Goldfeld and Quandt's study is indeed a significant contribution, it is still somewhat unsatisfactory since the structural model used contains only exogenous variables among the predetermined variables. Typically researchers work with models containing lagged endogenous variables, which are a complicating factor in the estimation process.

None of the studies cited above have incorporated the vector autoregressive specification. Guilkey, however, specified both the simple and vector autoregressive schemes for a set of sampling experiments conducted with a group of two seemingly unrelated regression equations.⁵ His model was estimated by the OLS method, Zellner's procedure for efficient estimates of groups of seemingly

unrelated regression equations, Parks' adaptation of Zellner's procedure to the case of first-order simple autoregressive disturbances, and Guilkey's adaptation of Zellner's procedure to the case of first-order vector autoregressive disturbances. His results show that even in samples composed of as few as twenty observations the asymptotic properties are "a good guide to the finite sample properties." An exception to this conclusion is the case for which the diagonal elements of the R matrix were specified to be greater in value than the off diagonal elements ($\rho_{11}=\rho_{22}=.6$, $\rho_{12}=\rho_{21}=.2$; which implies a stable process.) Here the estimation method designed under the assumption of simple autoregressive disturbances (Parks' procedure) outperformed the method designed under the assumption of vector autoregressive disturbances (Guilkey's procedure) at sample size 20. Although Guilkey's results are interesting and valuable, they may not hold in the context of simultaneous equation models with lagged endogenous variables.

Format of Sampling Experiments

The survey of the previous work indicates a strong need for further research in this area. Neither lagged endogenous variables nor vector autoregressive disturbances have been specified for the simultaneous equation models used in the context of these earlier experiments. The experiments conducted as a part of this dissertation represent an effort to investigate models with these characteristics. Various estimation methods are applied to samples of data generated from a two equation structural model. The relative performances of the selected estimation methods are then compared on the basis of their performance in repeated trials.

The two equation model has the following structural form:

$$y_1(t) = \beta_0 + \beta_1 y_2(t) + \beta_2 y_1(t-1) + \beta_3 x_1(t) + u_1(t)$$

$$y_2(t) = \gamma_0 + \gamma_1 y_2(t) + \gamma_2 y_2(t-1) + \gamma_3 x_2(t) + u_2(t)$$

with first-order autoregressive disturbances such that

$$u_1(t) = \rho_{11} u_1(t-1) + \rho_{12} u_2(t-1) + \varepsilon_1(t)$$

$$u_2(t) = \rho_{21} u_1(t-1) + \rho_{22} u_2(t-1) + \varepsilon_2(t)$$

Both structural equations are overidentified according to the usual order conditions for identification.⁶ Note that the first equation contains as an explanatory variable the lagged value of the endogenous variable on the left hand side; whereas the second equation contains as an explanatory variable the lagged value of the endogenous variable appearing on the right hand side. The structural model is simple, and yet elaborate enough to resemble a number of small, but common economic models. For instance, consider the demand and supply model for a product:

$$\text{(demand)} \quad Q^*(t) = \alpha_0 + \alpha_1 P(t) + \alpha_2 Y(t) + u_1^*(t)$$

$$\text{(supply)} \quad Q(t) = Q^D(t) + Q^F(t)$$

$$Q^D(t) = \beta_{10} + \beta_{11} P(t-1) + u_{21}(t)$$

$$Q^F(t) = \beta_{20} + \beta_{21} P(t) + \beta_{22} I(t) + u_{22}(t)$$

where

Q = quantity

Q^* = the desired level of demand (based on a habit persistence
or partial adjustment mechanism)

P = price

Y = income

Q^D = domestic supply

Q^F = foreign supply

I = level of inventories held by foreign suppliers

(The variables Q , Q^D , Q^F , and P are endogenous; Y and I are exogenous.) If the sample does not provide separate information on the domestic and foreign supplies, then the model's equations to be estimated will have the form:

$$\text{(demand)} \quad Q(t) = \gamma_0 + \gamma_1 P(t) + \gamma_2 Q(t-1) + \gamma_3 Y(t) + u_1(t)$$

$$\text{(supply)} \quad Q(t) = \delta_0 + \delta_1 P(t) + \delta_2 P(t-1) + \delta_3 I(t) + u_2(t)$$

These equations are precisely the same as those used in the sampling experiments. Certainly other similar small models might be found.

The model's structural form implies the following for its reduced form:

$$\begin{aligned} y_1(t) &= [1/(\beta_1 - \gamma_1)] [\beta_1 \gamma_0 - \beta_0 \gamma_1 + \beta_1 \gamma_2 y_2(t-1) - \beta_2 \gamma_1 y_1(t-1) \\ &\quad + \beta_1 \gamma_3 x_2(t) - \beta_3 \gamma_1 x_1(t)] \\ &\quad + [1/(\beta_1 - \gamma_1)] [\beta_1 u_2(t) - \gamma_1 u_1(t)] \\ y_2(t) &= [1/(\beta_1 - \gamma_1)] [\gamma_0 - \beta_0 + \gamma_2 y_2(t-1) - \beta_2 y_1(t-1) \\ &\quad + \gamma_3 x_2(t) - \beta_3 x_1(t)] \\ &\quad + [1/(\beta_1 - \gamma_1)] [u_2(t) - u_1(t)] \end{aligned}$$

The reduced form shows that β_1 cannot equal γ_1 . For simplicity, let the model's structural coefficients take on the following values:

$$y_1(t) = -25 + 2y_2(t) + .5y_1(t-1) + x_1(t) + u_1(t)$$

$$y_1(t) = -15 + y_2(t) + .5y_2(t-1) + x_2(t) + u_2(t)$$

Three sampling experiments were conducted; each consisted of 100 replications with samples comprised of 20, 50, and 100 observations.⁷ The three experiments correspond to these three specifications for the structural disturbances (each implies a stable

process):

Experiment I strong simple autoregression $\rho_{11}=\rho_{22}=.8$; $\rho_{12}=\rho_{21}=0$

Experiment II moderate simple autoregression $\rho_{11}=\rho_{22}=.4$; $\rho_{12}=\rho_{21}=0$

Experiment III vector autoregression $\rho_{11}=\rho_{22}=.5$; $\rho_{12}=\rho_{21}=.3$

The sample values⁸ for the nonautocorrelated disturbances, $\varepsilon_1(t)$ and $\varepsilon_2(t)$, were drawn from a normal distribution with mean 0 and variance 2. This value for the variance gave high R^2 values ($>.9$) as are usually found when working with macroeconomic time-series data. The sample values⁹ of the exogeneous variables, $x_1(t)$ and $x_2(t)$, were drawn from a normal distribution with mean 10 and variance 1. The sample values of the endogenous variables were then calculated from the values of the exogeneous variables and the disturbance terms.

(In the discussion of the results, the six estimation methods used in the experiments will be referred to as:

Method A - ordinary least squares (OLS) estimator

Method B - OLS estimation adapted for simple autoregression (OLS-AUTO)

Method C - two-stage least squares (2SLS) estimator

Method D - Amemiya's 2SLS-AUTO estimator

Method E - Fair's modification of Amemiya's method for undersized samples

Method F - Hatanaka's 2SLS(RA)-VECTOR estimator.)

The results of the sampling experiments are presented in tables 11-25. (For convenience, tables 11-12, 16-18, 20-21, 23, and 25 are grouped together at the end of the chapter.) The criterion for

judging the estimators' performance will be the relative degree of concentration of the estimators' distributions around the true parameter value. In the past, researchers conducting sampling experiments have typically calculated the distribution's mean, variance, and MSE to describe its basic characteristics and thereby measure its concentration around the true parameter value. In some cases, however, this may be an inappropriate practice. Basmann investigated the distribution function of the 2SLS estimator in small samples for a few specific cases and showed that the 2SLS estimator, at times, has an infinite variance.¹⁰ When an estimator has no finite variance, the estimates of the variance or MSE calculated from the sampling experiments are difficult to interpret and may be meaningless. To avoid the pitfall of nonexistent second moments, we will compare the accuracy of the estimators by computing the number of times each method provides estimates within selected intervals around the true parameter value. This method of assessing the performance of alternative estimation methods was used by Zellner to analyze a set of sampling experiments with sampling theory and Bayesian estimators of simultaneous equation models.¹¹ Since the model used in this dissertation closely resembles that model in the case for which Basmann found the 2SLS estimator's variance to be infinite, Zellner's alternative, which circumvents the use of moments, is recommended here.

Experiment I: Strong Simple Autoregression

The first experiment is characterized by a rather high value, $\rho=.8$, of the autoregression coefficients. This somewhat extreme

case was chosen so that the effects of simple autoregressive disturbances would be pronounced. Table 11 contains the mean estimates of the parameter values from methods A through E.¹² Calculation of the mean estimate appears to be a valid practice since Basmann's work suggests that the estimators' distributions should have a finite small sample mean for the model being used here.¹³ These results show that for all sample sizes method D (2SLS-AUTO) regularly produces estimates of β and γ parameters with smaller mean biases than methods A, B, or C; and that method A (OLS) produces the largest biases. (Method E will be discussed separately.) For sample sizes 50 and 100, the mean biases of method D's estimates are small and generally becoming smaller as sample size increases. This suggests that method D's asymptotic unbiasedness becomes evident in these experiments at sample size 50. Methods B and D provide very similar mean estimates of the ρ coefficients. For sample size 20, their mean estimates of ρ are substantially biased toward zero; for sample size 100, these biases are greatly reduced, but not entirely removed.

Table 12 contains the frequency distributions of the estimates in intervals surrounding the true values of the parameters. For sample size 100, the estimates of the β and γ coefficients obtained from method D are distributed around the true value of the parameter with a concentration as dense or denser than that of estimates from the other methods in the intervals ± 0.05 , ± 0.10 , ± 0.15 , ± 0.20 . Method B is second best, uniformly superior to methods A and C. The small dispersion of method A's estimates combined with a relatively

large bias results in no estimates of β_0 or β_1 falling in the interval $\pm .05$ around the true parameter value. For sample size 50, in most instances the estimates produced by method D are still the most highly concentrated around the true value of the parameters; while the differences between methods B and D are diminished or have vanished. For sample size 20, the advantage of method D over method B disappears; no distinction can be made with any degree of certainty between the performance of these two methods. As a supplement to the information conveyed by the estimators' frequency distributions, a check was made to determine the frequency of insignificant slope estimates, based on the t-ratio, that were produced by each method. This analysis showed that there were 11 cases for which the null hypothesis of the slope coefficient equalling zero could not be rejected at the 5 per cent level of significance by methods A and C; this serious error was never made by methods B or D. Thus, the relatively poor performance of methods A and C is verified. The result that method D is in most cases best when sample size reaches 50 roughly indicates the lower limit for the sample size at which the property of asymptotic unbiasedness becomes evident in this experiment. The information contained together in tables 11 and 12 shows that the distributions of method D's estimates are collapsing around the true parameter value as sample size increases, in accordance with the property of consistency. Regarding the estimates of the autoregression coefficients, there is virtually no difference between the frequency distributions of method B's and D's estimates of ρ_1 and ρ_2 in the intervals surrounding .8 at each sample size.

To illustrate an estimator's sampling distribution, histograms are presented in tables 13 and 14 for the estimates of β_1 obtained in this experiment from method D and samples of size 20 and 100. The histograms illustrate that the estimator's sampling distribution is collapsing around the true value of the parameter ($\beta_1=2$) as sample size increases. Furthermore, the sampling distribution for the estimator with sample size 20 seems to be somewhat skewed to the right and has a taller peak (leptokurtic) than in the case of the normal distribution. (Basmann's work showed that the distribution of the 2SLS estimator in finite samples is leptokurtic.) Also, the single observation in the interval furthest to the right appears to be an example of an outlying value. The phenomenon of "outliers" has frequently been encountered in sampling experiments investigating the small sample properties of 2SLS-type estimators, and may be responsible for the small sample distribution of the 2SLS estimator having an infinite variance in some cases. The sampling distribution for the estimator with sample size 100 is more symmetrical, but it has a shorter peak (platykurtic) than in the case of the normal distribution. The fact that this sampling experiment consists of only 100 replications may account for this finding. No outliers were evident.

Next, 50 interval estimates were computed for each of the model's six slope coefficients based on the t distribution and the point estimates of method D from samples of size 20 and 100, and the the point estimate of method B from samples of size 20. This work was done to check the frequency for which the interval estimates,

TABLE 13

HISTOGRAM OF β_1 ESTIMATES (METHOD D, SAMPLE SIZE 20,
TRUE VALUE OF $\beta_1 = 2$)

<u>Interval Midpoint</u>	<u>Count</u>	<u>(Each x = 1)</u>
1.7870	2	xx
1.8388	8	xxxxxxxx
1.8906	15	xxxxxxxxxxxxxxxx
1.9424	17	xxxxxxxxxxxxxxxx
1.9942	19	xxxxxxxxxxxxxxxx
2.0460	23	xxxxxxxxxxxxxxxx
2.0978	7	xxxxxxx
2.1496	5	xxxxx
2.2014	3	xxx
2.2532	0	
2.3050	1	x
TOTAL	100	(interval width = .0518)

Descriptive Statistics

Minimum	1.787
Maximum	2.305
Mean	1.990
Mean Bias	-.010
Standard Deviation	.097
Mean Square Error	9.51×10^{-3}
Skewness	.353
Kurtosis	.224

TABLE 14

HISTOGRAM OF β_1 ESTIMATES (METHOD D, SAMPLE SIZE 100,
TRUE VALUE OF $\beta_1 = 2$)

<u>Interval Midpoint</u>	<u>Count</u>	<u>(Each x = 1)</u>
1.9000	2	xx
1.9186	4	xxxx
1.9372	4	xxxx
1.9558	12	xxxxxxxxxxxx
1.9744	14	xxxxxxxxxxxx
1.9930	20	xxxxxxxxxxxxxxxxxxxx
2.0116	18	xxxxxxxxxxxxxxxxxxxx
2.0302	11	xxxxxxxxxx
2.0488	7	xxxxxxx
2.0674	6	xxxxxx
2.0860	2	xx
TOTAL	100	(interval width = .0186)

Descriptive Statistics

Minimum	1.900
Maximum	2.086
Mean	1.997
Mean Bias	-.003
Standard Deviation	.040
Mean Square Error	1.61×10^{-3}
Skewness	-.153
Kurtosis	-.214

based on the t distribution at the 5 per cent level of probability actually covered the true parameter value. The results are recorded in table 15. The standard practice of using the t distribution for confidence intervals and hypothesis testing appears sound since the frequency of failures is close to the expected number of failures (15), given the level of confidence.

TABLE 15

FREQUENCY OF INCORRECT CONFIDENCE INTERVAL
(95% LEVEL OF CONFIDENCE)
STRONG SIMPLE AUTOREGRESSION

<u>Method</u>	<u>Sample Size</u>	<u>β_1</u>	<u>β_2</u>	<u>β_3</u>	<u>γ_1</u>	<u>γ_2</u>	<u>γ_3</u>	<u>Total (300 Cases)</u>
B	20	3	2	2	1	3	2	13
D	20	3	2	3	1	3	3	15
D	100	5	3	3	2	3	1	17

The estimates of the ρ coefficients and their estimated standard errors from method D were used to test the null hypothesis of nonautoregressive disturbances. This test for the presence of simple first-order autoregressive disturbances, described in chapter II, allows to reject the null hypothesis if $\hat{\rho}$ divided by its estimated standard error is less than 2. This work indicates the ability of the test to detect simple autoregressive disturbances. Also, the ability of the Durbin-Watson test (calculated from method C's estimates) to reject the null hypothesis is shown. This is still another investigation into that probably overworked topic: the performance of the Durbin-Watson statistic in appropriate and inappropriate situations. Table 16 contains these results. This analysis suggests that in the context of strong autoregressive disturbances both tests

perform well for large samples (samples with 50 or 100 observations), but not so well for samples of smaller size (sample size 20); the test based on the estimated value of ρ , however, does much better than the Durbin-Watson test for the smallest sample size.

When the sample contains too few observations for estimation of the complete augmented reduced form, method D cannot be used. In its place Fair's adaptation of 2SLS-AUTO - method E - can be employed, since it requires estimation of only a subset of the augmented reduced form in order to obtain fitted values of the endogenous variables for the second stage. The results contained in table 12 indicate that, for all sample sizes, estimates produced by method D are distributed with a less dense concentration around the true parameter value than those from method B. This finding comes as a bit of a surprise since method E possesses the asymptotic property of consistency, whereas method B does not; this property might be expected to become evident in the results obtained from samples of size 100. This finding suggests that there may be payoffs in estimation accuracy from accounting for both simultaneity and autoregressive disturbances only if the entire augmented reduced form can be estimated. That is, if the sample is undersized, then an ordinary least squares estimator correcting only for simple autoregressive disturbances is preferred to method E. The unreliability of method E is confirmed by a check of the t-ratio of the slope coefficients that reveals 21 cases in which this method incorrectly suggested that a variable was irrelevant at the 5 per cent level of significance.

Experiment II: Moderate Simple Autoregression

A moderate degree of autoregression was specified for the second experiment. Here the value of the autoregression coefficient, ρ , is .4, which is one half its value in the first experiment and closer to that which researchers might encounter in their work. Table 17 contains the mean estimates of the parameter values for methods A through E. With respect to the first equation, for all sample sizes method D regularly provides estimates of the β coefficients with mean biases smaller than methods A, B, or C; method A's estimates contain the largest mean biases. (Method E will be discussed separately.) These findings parallel those in the first experiment. With respect to the second equation, however, no clear pattern emerges. Methods A, C, and D are each able to provide estimates of a γ coefficient such that their estimates have the smallest mean biases for the given sample size. It thus seems that method D's property of asymptotic unbiasedness is evident in its estimates of the first, but not the second equation. Unlike the first equation, the mean biases of method D's estimates of the second equation are neither clearly less than those of other methods; nor are they generally becoming smaller as sample size increases, although they were always small, beginning with sample size 20. Methods B and D provide similar mean estimates of the ρ coefficients; the pattern established in the first experiment is again followed.

Table 18 presents the frequency distributions of the estimates in intervals surrounding the true values of the parameters. For each sample size, no estimator emerges superior according to

the criteria established for these experiments. In most cases (13 out of 20) none of the four methods does best in estimating a particular coefficient for a given sample size. One method would provide estimates most frequently within one interval, but not at another interval; hence, no estimator could be determined to be superior. In some cases (7 out of 20), one method does do best: method D 3 times, method B once, methods A and C once each, with 1 tie between them. Here one estimator does provide estimates most frequently in all four intervals ($\pm .05$, $\pm .10$, $\pm .15$, $\pm .20$). When a moderate degree of simple autoregression is present in the disturbances, method D's property of asymptotic efficiency never becomes as evident as it had in the previous experiment. Furthermore, the desirable property of consistency is apparent in the combined information from tables 17 and 18 only for method D's estimates of the first equation. Even though the distributions of method D's estimates of the γ coefficients are collapsing as sample size increases, consistency is not clearly present because asymptotic unbiasedness, a necessary condition for consistency, has not been established. Regarding the estimates of the autoregression coefficient, the pattern is again the same as that found in the first experiment: virtually no difference in the frequency distributions of method B's and D's estimates of ρ_1 and ρ_2 .

Next, 50 interval estimates for each of the second equation's three slope coefficients were computed; they were based on the t distribution at the 5 per cent level of probability and the point estimates of methods A through D from samples of sizes 20 and 100. These results are contained in table 19. For sample size 20, the use of

the t distribution to provide interval estimates is apparently a reasonable procedure to follow in conjunction with all four estimation methods. For sample size 100, this procedure works well only with methods A and D. For methods B and C, the number of failures of the interval estimates to cover the parameter's true value is substantially greater than the expected number of failures (7.5) indicated by the 95 per cent level of confidence. The poorer performance of methods B and C in the construction of the interval estimates may reflect substantial bias in the point estimates rather than the t distribution being a poor approximation of the estimator's sampling distribution around its mean.

TABLE 19

FREQUENCY OF INCORRECT CONFIDENCE INTERVAL
(95% LEVEL OF CONFIDENCE)
MODERATE SIMPLE AUTOREGRESSION

<u>Method</u>	<u>Sample Size</u>	<u>β_1</u>	<u>β_2</u>	<u>β_3</u>	<u>Total (of 150 Cases)</u>
A	20	1	3	4	8
B	20	2	3	3	8
C	20	2	3	3	8
D	20	3	3	4	10
A	100	3	2	3	8
B	100	8	4	6	18
C	100	8	1	9	18
D	100	2	3	3	8

Tests for the presence of simple autoregressive disturbances were performed with both the estimates of ρ and its standard error and the Durbin-Watson statistic. Table 20 presents these results. The two-sided version of the test based on the estimated value of ρ leads to the rejection of the null hypothesis of nonautoregressive disturbances in about 1 of 3 attempts for sample size 20. This weak

performance had to be expected since the test was given an impossible task: for this sample size, the critical value of the test (.43) was greater than the true value of the parameter (.40). For sample size 50, the test still performed poorly; only for sample size 100 did it perform satisfactorily. At all sample sizes, it outperformed the Durbin-Watson test.

In the case of an undersized sample, Fair's method E can be employed in place of method D. Table 17 indicates that method E provides estimates with satisfactorily small mean biases, but the results in table 18 show that method B outperforms method E in the estimation of the β and γ coefficients at all sample sizes (with one exception: γ_2 , at sample size 100). This finding parallels the finding in the first experiment: method E's large sample property of consistency is not evident in the experiment's findings at sample size 100. The unreliability of method E is confirmed by a check of the slope coefficients that reveal 9 cases in which this method incorrectly suggested that a variable was irrelevant at the 5 per cent level of significance; none of the other estimators ever made this mistake.

Experiment III: Vector Autoregression

In contrast to the simple autoregression specified for the two previous experiments, the third experiment is characterized by vector autoregressive disturbances. The values of $\rho_{11}=\rho_{22}=.5$ and $\rho_{12}=\rho_{21}=.3$ were selected for the elements of the matrix R, the matrix of autoregression coefficients, so that a) a high degree of vector autoregression is present (the sum of a row's coefficients equals .8),

and b) an equation's disturbance term is more strongly correlated with its own past value than with the past value of the other equation's disturbance term (this seems more plausible than the opposite situation). Table 21 contains the mean estimates of the parameter values for methods A through F. For sample sizes 50 and 100, method F usually (11 out of 16 cases) provides estimates with the smallest mean biases of the β and γ slope coefficients; for sample size 20, method F does so only once in 8 cases. Also the mean biases of method F's estimates are small and are generally becoming smaller as sample size increases from 50 to 100. These findings suggest that method F's property of asymptotic unbiasedness becomes evident in this experiment at sample size 50. The mean estimates of the R matrix generated by method F exhibit the peculiar pattern shown in table 22. There is no ready explanation for these results. Looking now at the components of the final estimates of the matrix R (as shown at the bottom of table 21), it can be seen that the mean corrections of the preliminary estimates of the R matrix are on the average actually distorting in some cases. The mean of the preliminary estimates is nearer the parameter's true value than the mean of the final estimates of ρ_{11} , sample sizes 50 and 100; ρ_{21} , sample size 20; and ρ_{22} , sample size 100. Again, no ready explanation can be offered for these findings.

TABLE 22
DIRECTION OF MEAN BIAS

<u>Sample Size</u>	<u>Coefficient</u>			
	<u>ρ_{11}</u>	<u>ρ_{12}</u>	<u>ρ_{21}</u>	<u>ρ_{22}</u>
20	-	-	+	-
50	+	-	-	+
100	+	+	+	+

Table 23 contains the frequency distributions of the estimates in intervals surrounding the true values of the parameters. For sample size 100, method F's estimates of all three of the slope coefficients in the first equation and two of the three in the second equation are distributed around the true parameter values in each of the intervals $\pm .05$, $\pm .10$, $\pm .15$, and $\pm .20$ with a concentration as dense or denser than the other methods' estimates. For sample sizes 20 and 50, it is almost always the case that none of the four methods does best in estimating a particular coefficient. The result that method F is superior in most cases only when sample size reaches 100 roughly indicates the lower limit for the sample size at which the property of asymptotic efficiency becomes evident in the estimates of the slope coefficients. The information contained together in tables 21 and 23 shows that the distributions of method F's estimates are collapsing around the true parameter values as sample size increases, in accordance with the property of consistency. There is no basis to discuss method F's estimates of the R matrix, since none of the alternative methods estimate the off diagonal elements.

Next, 25 interval estimates for each of the model's six slope coefficients were computed; they were based on the t distribution (at the 5 per cent level of probability) and the point estimates. These results are contained in table 24. For sample size 20, the estimates from methods A through D are used along with those from method F because the results reveal that none of the five estimators can be rated best in this experiment. Only methods D and F are considered at sample size 50 since these two methods appear better than the others; and at sample size 100, only method F, the method considered superior, is used. For methods C and D (sample size 20), the frequency of incorrect interval estimates is closest to the expected number of failures (7.5) indicated by the 95 per cent level of confidence. The larger than expected number of incorrect confidence intervals coming from method F at all sample sizes suggests that caution be used in work based on method F's point estimates and the t distribution.

TABLE 24

FREQUENCY OF INCORRECT CONFIDENCE INTERVAL
(95% LEVEL OF CONFIDENCE)
VECTOR AUTOREGRESSION

<u>Method</u>	<u>Sample Size</u>	<u>β_1</u>	<u>β_2</u>	<u>β_3</u>	<u>γ_1</u>	<u>γ_2</u>	<u>γ_3</u>	<u>Total (150 Cases)</u>
A	20	5	2	2	1	0	1	11
B	20	5	1	3	2	1	3	15
C	20	4	0	2	0	0	0	6
D	20	2	2	1	1	0	2	8
F	20	4	2	4	1	2	0	13
D	50	2	2	1	3	1	2	11
F	50	4	2	3	1	2	1	13
F	100	3	2	1	0	1	4	11

Tests for the presence of vector autoregressive disturbances were conducted with the F-test described in chapter II to investigate the ability of this test to detect the presence of vector autoregressive disturbances. The hypotheses of nonautoregressive disturbances and simple autoregressive disturbances were both tested against the alternative of vector autoregressive disturbances. Table 25 contains these results and shows that the test performs poorly at the smallest sample size. In addition, table 25 shows the frequency with which the hypothesis of nonautoregressive disturbances would be rejected in favor of the alternative of simple autoregression according to three relevant tests.

Summary

The results of these sampling experiments can be condensed into the following principal findings:

1. In experiment I, the case of strong simple autoregression, Amemiya's 2SLS-AUTO estimator (method D) emerges as the superior method for samples with 50 and 100 observations; the OLS-AUTO estimator (Method B) matches the performance of the 2SLS-AUTO estimator and together they are rated best at sample size 20. In contrast to these results, no estimation method distinguishes itself at any sample size in experiment II, in which there is a moderate degree of simple autoregression. In experiment III, the case of vector autoregression, Hatanaka's 2SLS(RA)-VECTOR estimator (method F) is rated superior at sample size 100; at smaller sample sizes, no estimator is better than any other. These results might suggest that in this context (simultaneity,

lagged endogenous variables, autoregressive disturbances, and 100 or fewer observations in the sample) an estimator with all the desirable asymptotic properties may outperform estimators without these properties only if a strong degree of simple or vector autoregression is present in the model's disturbances.

2. Throughout the two experiments with simple autoregression, Fair's estimator (method E), the adaptation of the 2SLS-AUTO method to models with undersized samples, is outperformed by the OLS-AUTO estimator, which disregards simultaneity. This seems to indicate that it would be better to use an estimation method that accounts for autoregression, but disregards simultaneity if simple autoregressive disturbances are assumed, the sample is undersized, and Fair's idea has to be implemented to obtain fitted values for the second stage. (This result should also hold when vector autoregression is assumed and the sample is undersized.) Furthermore, Fair's finding - that when estimated by his own method his small macro model generated more accurate forecasts than much more elaborate models, estimated by the OLS or 2SLS methods¹⁴ - deserves reconsideration. Since experiment II shows the OLS and 2SLS estimators to be superior to Fair's estimator, the better forecasting record of Fair's model should be attributed to a more solid theoretical framework, more precise data, or chance, but not to the estimation method.
3. The first two experiments indicate that the use of the t distribution to test hypotheses and construct interval estimates may be a fairly reasonable procedure for researchers to follow when

the model is characterized by simple autoregressive disturbances.¹⁵ According to the third experiment, the procedure is less reliable when the model is characterized by vector autoregressive disturbances.

4. The tests for the disturbances' specification perform poorly with samples of size 20, but do much better and perform satisfactorily as sample size increases to 50 and 100 observations. Fortunately, the unreliability of the tests at the smallest sample size ($T=20$) turns out to be a minor concern. This happens because the performances of methods based on an incorrect assumption for the disturbances are comparable to the performance of the method based on the correct assumption.

There is agreement on some points, but not others between these findings and those discussed in the earlier sampling experiments cited at the beginning of this chapter. Both the experiments reported here and those of Goldfeld and Quandt indicate that it is profitable to correct for autoregression when its degree is high, but not when it is low ($\rho_g = .9$ and $.1$ in their study; $\rho_g = .8$ and $.4$ in ours).¹⁶ Furthermore, Goldfeld and Quandt found very little or no gain from correcting for simultaneity as well as autoregression when the degree of autoregression is high; in contrast, our first experiment shows that the gain is significant. Our results agree with Hurd's to the extent that the desirable asymptotic properties have little relevance for the estimators' performance when the sample contains about 20 observations; however, the OLS method fails in our experiments to duplicate the strong performance shown in Hurd's.¹⁷

With respect to vector autoregression, both Guilkey's experiment and our own ($\rho_{11}=\rho_{22}=.6$; $\rho_{12}=\rho_{21}=.2$ in his study; $\rho_{11}=\rho_{22}=.5$; $\rho_{12}=\rho_{21}=.3$ in ours) reveal that an estimation method designed for vector autoregression is best, provided that a sufficiently large sample is used; in Guilkey's experiments with a group of seemingly unrelated regressions, a sample of 50 observations is needed; in our experiments with a simultaneous equation system including lagged endogenous variables, 100 observations are needed.¹⁸ If the results found by Guilkey and by Goldfeld and Quandt are merged with those reported in this dissertation, then a trend can be detected in the combined findings: as the complexity of the multiple equation model with autoregressive disturbances increases--from a group of seemingly unrelated regressions to a simultaneous equation system to a simultaneous equation system with lagged endogenous variables--a larger sample is needed for the estimation methods possessing all the desirable asymptotic properties to outperform estimation methods lacking these properties.

TABLE 11
 MEAN ESTIMATES OF THE COEFFICIENTS
 STRONG SIMPLE AUTOREGRESSION $\rho = .8$

True Value	<u>Equation 1</u>					<u>Equation 2</u>				
	β_0	β_1	β_2	β_3	ρ_1	γ_0	γ_1	γ_2	γ_3	ρ_2
	-25.00	2.000	.500	1.000	.800	-15.00	1.000	.500	1.000	.800
	<u>Sample Size 20</u>									
A	-22.42	1.893	.483	.860		-14.85	.966	.485	1.027	
B	-24.46	1.959	.493	.959	.615	-14.99	1.018	.504	.978	.642
C	-23.44	1.943	.496	.921		-14.94	.910	.475	1.092	
D	-24.86	1.990	.494	.977	.613	-15.00	1.006	.502	.981	.642
E	-24.87	1.990	.501	.977	.592	-15.07	.977	.498	1.023	.618
	<u>Sample Size 50</u>									
A	-20.91	1.815	.485	.770		-14.94	.943	.475	1.063	
B	-24.22	1.968	.492	.954	.714	-15.03	1.024	.509	.965	.746
C	-22.56	1.897	.497	.852		-15.00	.856	.461	1.173	
D	-24.94	1.997	.501	.986	.719	-15.08	1.004	.505	.992	.743
E	-24.22	1.968	.496	.960	.694	-15.14	.983	.502	1.024	.728

Method

TABLE 12

FREQUENCY OF ESTIMATES IN INTERVALS AROUND TRUE PARAMETER VALUES
STRONG SIMPLE AUTOREGRESSION

		<u>Sample Size 20</u>					<u>Sample Size 50</u>					<u>Sample Size 100</u>				
		<u>Intervals</u>					<u>Intervals</u>					<u>Intervals</u>				
		$\pm .05$	$\pm .10$	$\pm .15$	$\pm .20$		$\pm .05$	$\pm .10$	$\pm .15$	$\pm .20$		$\pm .05$	$\pm .10$	$\pm .15$	$\pm .20$	
Coefficient β_1																
<u>Method</u>		$\pm .05$	$\pm .10$	$\pm .15$	$\pm .20$		$\pm .05$	$\pm .10$	$\pm .15$	$\pm .20$		$\pm .05$	$\pm .10$	$\pm .15$	$\pm .20$	
A		28	44	62	74		11	24	46	68		0	15	38	56	
B		43	68	90	99		50	82	99	100		70	96	100	100	
C		26	56	78	89		25	43	77	86		22	53	74	84	
D		37	73	86	97		55	90	98	100		89	100	100	100	
E		29	58	71	84		35	62	82	94		58	80	99	99	
Coefficient β_2																
A		55	86	95	98		67	94	99	100		76	98	100	100	
B		65	96	100	100		87	100	100	100		97	100	100	100	
C		58	85	96	98		67	93	99	100		76	98	100	100	
D		63	96	100	100		91	100	100	100		99	100	100	100	
E		56	83	97	99		80	98	99	100		98	99	99	100	
Coefficient β_3																
A		18	31	43	56		10	19	26	39		2	10	26	41	
B		22	52	73	87		29	65	90	96		50	86	97	100	
C		16	32	48	66		16	30	53	70		14	33	50	72	
D		20	52	70	85		36	71	91	97		68	94	99	100	
E		19	38	57	75		24	52	70	84		48	84	94	98	

TABLE 12 - Continued

		<u>Sample Size 20</u>					<u>Sample Size 20</u>					<u>Sample Size 100</u>				
		<u>Intervals</u>					<u>Intervals</u>					<u>Intervals</u>				
Coefficient γ_1		$\pm .05$	$\pm .10$	$\pm .15$	$\pm .20$		$\pm .05$	$\pm .10$	$\pm .15$	$\pm .20$		$\pm .05$	$\pm .10$	$\pm .15$	$\pm .20$	
Method																
A	31	67	82	90			36	62	86	92		41	67	86	95	
B	42	78	94	100			62	91	100	100		71	100	100	100	
C	34	59	70	83			18	33	53	69		7	26	49	70	
D	41	82	95	100			71	94	100	100		84	100	100	100	
E	34	58	73	84			47	75	97	98		55	96	99	100	
Coefficient γ_2																
A	40	76	90	96			54	85	95	100		59	94	98	100	
B	62	92	100	100			78	100	100	100		94	100	100	100	
C	43	74	87	96			49	85	92	100		45	86	97	99	
D	60	93	100	100			78	100	100	100		96	100	100	100	
E	62	91	98	99			77	97	99	100		96	100	100	100	
Coefficient γ_3																
A	20	39	55	70			25	42	54	70		22	44	62	76	
B	27	47	63	79			43	71	85	98		50	81	97	99	
C	18	33	55	70			14	35	45	53		17	28	40	54	
D	28	46	68	81			47	71	93	97		61	90	98	100	
E	15	40	55	67			34	55	76	93		40	78	94	98	

TABLE 12 - Continued

		<u>Sample Size 20</u>					<u>Sample Size 50</u>					<u>Sample Size 100</u>				
		<u>Intervals</u>					<u>Intervals</u>					<u>Intervals</u>				
Coefficient β_0																
Method		$\pm .05$	$\pm .10$	$\pm .15$	$\pm .20$		$\pm .05$	$\pm .10$	$\pm .15$	$\pm .20$		$\pm .05$	$\pm .10$	$\pm .15$	$\pm .20$	
Coefficient β_0	A	6	18	32	37		5	9	12	19		0	2	7	11	
	B	16	24	45	58		17	38	54	66		31	52	75	91	
	C	11	25	33	43		9	21	33	42		5	14	30	40	
	D	16	26	47	58		22	39	61	74		36	67	84	93	
	E	10	16	26	37		17	25	43	49		25	53	75	96	
Coefficient γ_0																
Coefficient γ_0	A	17	33	45	60		25	40	58	70		27	51	68	82	
	B	22	36	56	69		33	61	85	93		54	83	95	98	
	C	18	30	48	58		24	42	57	72		27	52	66	82	
	D	23	40	56	71		33	62	82	92		55	84	96	98	
	E	21	37	53	71		28	55	77	89		54	83	95	99	
Coefficient ρ_1																
Coefficient ρ_1	B	19	40	56	63		27	59	78	89		51	75	92	97	
	D	18	39	56	64		28	63	78	89		51	75	91	98	
	E	19	37	48	60		27	62	72	83		52	73	91	98	
Coefficient ρ_2																
Coefficient ρ_2	B	23	49	63	67		45	69	89	93		54	84	95	99	
	D	24	50	64	66		44	69	85	93		52	83	96	99	
	E	24	45	57	61		42	62	82	92		53	81	93	99	

TABLE 16

FREQUENCY OF FAILURE TO REJECT NULL HYPOTHESIS
STRONG SIMPLE AUTOREGRESSION

Null Hypothesis: Nonautoregressive Disturbances
Alternative Hypothesis: Simple Autoregression

Estimation Method	Sample Size	Test Using Estimated Value of ρ		Test Using Durbin-Watson Statistic ^a	
		Equation 1	Equation 2	Equation 1	Equation 2
B	20	20 ^b	18	..	..
D	20	19 ^b	18	53	45
D	50	2	0	2	0
D	100	0	0	0	0
E	20	20 ^b	18	..	..

Null Hypothesis: Nonautoregressive Disturbances
Alternative Hypothesis: Positive Simple Autoregression

Estimation Method	Sample Size	Test Using Estimated Value of ρ		Test Using Durbin-Watson Statistic ^a	
		Equation 1	Equation 2	Equation 1	Equation 2
B	20	14	11	..	..
D	20	14	11	44	34
D	50	2	0	2	0
D	100	0	0	0	0
E	20	16	16	..	..

Critical Value of ρ

<u>Sample Size</u>	<u>Two-Sided Test</u>	<u>One-Sided Test</u>
20	.43	.37
50	.29	.24
100	.20	.17

^a at the 5% level of significance

^b There was one case of rejection due to large negative value
of ρ .

TABLE 17

MEAN ESTIMATES OF THE COEFFICIENTS
MODERATE SIMPLE AUTOREGRESSION
 $\rho = .4$

True Value	Equation 1				Equation 2					
	β_0	β_1	β_2	β_3	ρ_1	γ_0	γ_1	γ_2	γ_3	ρ_1
	-25.00	2.000	.500	1.000	.400	-15.00	1.000	.500	1.000	.400
	Sample Size 20									
A	-23.56	1.947	.496	.941		-15.02	1.007	.514	.998	
B	-24.06	1.960	.493	.957	.298	-15.11	1.018	.520	.979	.300
C	-24.80	1.979	.504	.980		-15.06	.976	.509	1.027	
D	-24.65	1.990	.501	.984	.307	-15.14	.999	.516	.996	.297
E	-24.95	1.979	.506	.997	.305	-15.15	.999	.516	1.006	.291
	Sample Size 50									
A	-22.99	1.926	.493	.907		-14.99	1.005	.505	.993	
B	-23.86	1.954	.490	.944	.339	-14.95	1.032	.511	.957	.347
C	-24.30	1.969	.501	.963		-15.04	.968	.499	1.040	
D	-24.74	1.991	.499	.987	.351	-15.00	1.005	.505	.994	.336
E	-24.61	1.983	.498	.976	.344	-15.01	1.001	.505	.999	.333

Method

TABLE 17 - Continued

<u>Method</u>	<u>Equation 1</u>					<u>Equation 2</u>				
	β_0	β_1	β_2	β_3	ρ_1	<u>Sample Size 100</u>				ρ_1
						γ_0	γ_1	γ_2	γ_3	
A	-23.38	1.930	.497	.911		-15.14	1.003	.505	1.005	
B	-24.19	1.930	.493	.960	.370	-15.13	1.035	.513	.963	.384
C	-24.39	1.973	.505	.962		-15.11	.963	.497	1.051	
D	-24.99	2.000	.503	.997	.382	-15.12	1.004	.505	1.002	.370
E	-24.95	1.998	.503	.994	.380	-15.12	.998	.504	1.009	.366

TABLE 18
FREQUENCY OF ESTIMATES IN INTERVALS AROUND TRUE PARAMETER VALUES
MODERATE SIMPLE AUTOREGRESSION

		<u>Sample Size 20</u>					<u>Sample Size 50</u>					<u>Sample Size 100</u>				
		<u>Intervals</u>					<u>Intervals</u>					<u>Intervals</u>				
		$\pm .05$	$\pm .10$	$\pm .15$	$\pm .20$		$\pm .05$	$\pm .10$	$\pm .15$	$\pm .20$		$\pm .05$	$\pm .10$	$\pm .15$	$\pm .20$	
Coefficient β_1																
Method																
A	36	65	76	92	92		27	63	91	98		30	75	97	100	
B	39	65	79	88	88		44	80	93	99		58	91	100	100	
C	32	67	78	93	93		51	83	94	99		63	92	100	100	
D	36	66	77	89	89		60	86	96	98		75	97	100	100	
E	29	51	64	73	73		36	59	75	85		46	77	90	96	
Coefficient β_2																
A	63	94	100	100	100		93	100	100	100		93	100	100	100	
B	67	89	99	100	100		88	100	100	100		95	100	100	100	
C	61	95	100	100	100		93	100	100	100		92	100	100	100	
D	66	92	99	100	100		89	100	100	100		92	100	100	100	
E	56	83	94	100	100		74	98	100	100		86	100	100	100	
Coefficient β_3																
A	18	39	61	73	73		27	50	66	79		26	50	78	90	
B	26	44	66	76	76		33	54	74	88		48	75	87	100	
C	24	42	65	79	79		32	61	77	87		40	71	85	98	
D	28	52	67	77	77		27	59	81	88		59	78	93	99	
E	26	41	59	65	65		19	46	63	76		42	70	83	91	

TABLE 18 - Continued

		<u>Sample Size 20</u>					<u>Sample Size 50</u>					<u>Sample Size 100</u>				
		<u>Intervals</u>					<u>Intervals</u>					<u>Intervals</u>				
Coefficient γ_1		$\pm.05$	$\pm.10$	$\pm.15$	$\pm.20$		$\pm.05$	$\pm.10$	$\pm.15$	$\pm.20$		$\pm.05$	$\pm.10$	$\pm.15$	$\pm.20$	
Method	A	45	69	83	99		60	93	98	100		77	96	100	100	
	B	41	66	85	96		53	89	98	100		60	97	100	100	
	C	40	67	84	90		58	82	97	99		63	90	97	99	
	D	38	66	86	96		60	89	98	100		71	98	100	100	
	E	22	40	63	81		40	70	93	99		55	85	99	100	
Coefficient γ_2																
Method	A	59	82	97	99		74	99	100	100		86	100	100	100	
	B	53	84	97	100		76	100	100	100		86	100	100	100	
	C	59	84	100	100		70	98	100	100		90	100	100	100	
	D	60	89	98	100		78	99	100	100		89	100	100	100	
	E	56	82	98	100		73	100	100	100		89	100	100	100	
Coefficient γ_3																
Method	A	19	38	57	71		38	65	85	89		43	80	92	96	
	B	26	44	56	69		36	64	82	89		44	77	91	99	
	C	19	40	54	72		43	67	82	93		44	68	84	95	
	D	27	45	62	74		37	64	83	93		56	75	96	98	
	E	24	42	48	60		34	61	79	89		34	66	88	96	

TABLE 18 - Continued

		Sample Size 20					Sample Size 50					Sample Size 100				
		<u>Intervals</u>					<u>Intervals</u>					<u>Intervals</u>				
Coefficient β_0																
Method		$\pm.05$	$\pm.10$	$\pm.15$	$\pm.20$		$\pm.05$	$\pm.10$	$\pm.15$	$\pm.20$		$\pm.05$	$\pm.10$	$\pm.15$	$\pm.20$	
A		16	24	36	51		12	19	35	51		14	31	44	59	
B		18	32	42	53		18	34	46	60		30	54	68	79	
C		14	26	34	51		16	40	63	73		33	54	73	84	
D		19	25	42	53		18	40	57	71		35	59	81	91	
E		17	36	41	48		15	28	40	48		25	43	56	65	
Coefficient γ_0																
A		28	50	65	77		43	74	87	95		61	84	98	100	
B		27	50	64	78		37	72	93	98		63	92	99	100	
C		30	52	64	78		43	76	88	94		64	84	98	100	
D		27	49	65	77		39	75	93	97		62	92	99	100	
E		14	48	63	77		45	80	93	97		62	92	99	100	
Coefficient ρ_1																
B		17	36	51	62		29	52	74	84		43	75	90	100	
D		16	36	51	64		33	58	76	90		43	75	92	100	
E		22	38	53	68		33	52	74	83		45	70	89	98	
Coefficient ρ_2																
B		12	26	40	53		29	52	73	81		42	78	90	93	
D		12	25	41	53		28	48	71	81		42	76	89	93	
E		12	29	47	56		28	49	69	81		41	74	88	91	

TABLE 20

FREQUENCY OF FAILURE TO REJECT NULL HYPOTHESIS
MODERATE SIMPLE AUTOREGRESSION

Null Hypothesis: Nonautoregressive Disturbances

Alternative Hypothesis: Simple Autoregression

Estimation Method	Sample Size	Test Using Estimated Value of ρ^a		Test Using Durbin-Watson Statistic ^b	
		Equation 1	Equation 2	Equation 1	Equation 2
E	20	72	61 ^c	..	...
D	20	70	60 ^c	96	93
D	50	29	36	52	46
D	100	0	6	2	8
E	20	70	55 ^c	..	...

Null Hypothesis: Nonautoregressive Disturbances

Alternative Hypothesis: Positive Simple Autoregression

Estimation Method	Sample Size	Test Using Estimated Value of ρ^a		Test Using Durbin-Watson Statistic ^b	
		Equation 1	Equation 2	Equation 1	Equation 2
B	20	59	55	..	..
D	20	59	56	92	89
D	50	20	24	66	36
D	100	0	6	0	6
E	20	56	48	..	..

^aThe critical values of ρ are the same as those shown earlier in table 16.

^bat the 5% level of significance

^cThere was one case of rejection due to large negative value of ρ .

TABLE 21

MEAN ESTIMATES OF THE COEFFICIENTS
VECTOR AUTOREGRESSION

True Value	<u>Equation 1</u>					<u>Equation 2</u>					
	β_0	β_1	β_2	β_3	ρ_{11}	ρ_{12}	γ_0	γ_1	γ_2	γ_3	ρ_{21} ρ_{22}
	-25.00	2.000	.500	1.000	.500	.300	-15.00	1.000	.500	1.000	.500 .300
	<u>Sample Size 20</u>										
A	-22.70	1.892	.483	.873			-14.89	1.057	.506	.920	
B	-22.50	1.903	.485	.899	.436		-14.44	1.072	.499	.896	.505
C	-24.74	1.988	.506	.984			-15.00	.949	.486	1.056	
D	-24.42	1.975	.505	.964	.458		-14.61	1.019	.488	.948	.469
E	-23.79	1.928	.499	.932	.434		-16.64	1.010	.490	.957	.463
F	-23.80	1.951	.463	.957	.338	.128	-15.37	1.020	.517	.997	.240 .453
	<u>Sample Size 50</u>										
A	-20.92	1.834	.476	.790			-14.84	1.067	.503	.914	
B	-21.99	1.875	.477	.846	.543		-14.43	1.099	.502	.842	.608
C	-23.80	1.958	.499	.942			-14.97	.941	.482	1.072	
D	-24.55	1.908	.507	.956	.557		-14.53	1.025	.486	.942	.577
E	-23.80	1.950	.499	.930	.549		-14.50	1.048	.491	.910	.586
F	-24.67	1.988	.487	.989	.458	.270	-14.91	.994	.500	.987	.424 .292

Method

TABLE 21 - Continued

<u>Equation 1</u>						<u>Equation 2</u>						
<u>Method</u>	<u>Sample Size 100</u>											
	β_0	β_1	β_2	β_3	ρ_{11}	ρ_{12}	γ_0	γ_1	γ_2	γ_3	ρ_{22}	ρ_{21}
A	-21.10	1.828	.478	.801			-14.88	1.061	.499	.925		
B	-22.42	1.884	.480	.874	.594		-14.71	1.104	.507	.855	.644	
C	-23.58	1.945	.497	.935			-14.75	.925	.472	1.060		
D	-24.65	1.981	.510	.969	.601		-14.67	1.014	.482	.968	.608	
E	-23.97	1.951	.501	.950	.604		-14.70	1.056	.494	.908	.624	
F	-24.95	2.002	.494	1.004	.468	.311	-14.84	.992	.498	.993	.459	.317

Mean Estimates of the Components of the Final Estimates
of the Matrix R From Method F

<u>Sample Size</u>	<u>Preliminary</u>				<u>Correction</u>			
	ρ_{11}	ρ_{12}	ρ_{22}	ρ_{21}	ρ_{11}	ρ_{12}	ρ_{22}	ρ_{21}
20	.263	.083	.189	.348	.075	.045	.051	.105
50	.406	.248	.331	.290	.052	.022	.093	.002
100	.405	.338	.396	.350	.063	-.027	.063	-.033

TABLE 23

FREQUENCY OF ESTIMATES IN INTERVALS AROUND TRUE PARAMETER VALUES
VECTOR AUTOREGRESSION^a

		<u>Sample Size 20</u>					<u>Sample Size 50</u>					<u>Sample Size 100</u>				
		<u>Intervals</u>					<u>Intervals</u>					<u>Intervals</u>				
Coefficient β_1		$\pm .05$	$\pm .10$	$\pm .15$	$\pm .20$		$\pm .05$	$\pm .10$	$\pm .15$	$\pm .20$		$\pm .05$	$\pm .10$	$\pm .15$	$\pm .20$	
Method																
A	20	44	63	89			2	12	45	70		1	2	34	75	
B	22	52	69	87			7	25	73	94		1	31	85	99	
C	39	70	83	94			44	71	93	100		44	80	97	99	
D	38	67	85	97			54	88	98	100		70	99	100	100	
E	22	37	56	76			36	72	86	94		49	83	98	100	
F	36	64	76	84			44	72	96	100		44	72	96	100	
Coefficient β_2																
A	67	92	100	100			75	98	100	100		90	100	100	100	
B	68	95	95	100			84	100	100	100		95	100	100	100	
C	64	90	99	100			83	99	100	100		95	100	100	100	
D	70	93	99	99			92	100	100	100		96	100	100	100	
E	60	83	98	98			89	99	100	100		98	100	100	100	
F	60	92	96	100			84	100	100	100		96	100	100	100	
Coefficient β_3																
A	19	29	54	70			5	8	24	44		2	5	29	54	
B	16	33	61	74			7	25	44	72		7	30	69	92	
C	33	55	72	80			28	52	73	91		35	65	86	95	
D	22	47	65	83			43	67	85	98		58	87	100	100	
E	15	36	53	66			28	49	71	85		43	66	95	100	
F	28	48	72	80			48	68	80	88		64	88	100	100	

TABLE 23 - Continued

		Sample Size 20					Sample Size 50					Sample Size 100				
		<u>Intervals</u>					<u>Intervals</u>					<u>Intervals</u>				
Coefficient γ_1		$\pm .05$	$\pm .10$	$\pm .15$	$\pm .20$		$\pm .05$	$\pm .10$	$\pm .15$	$\pm .20$		$\pm .05$	$\pm .10$	$\pm .15$	$\pm .20$	
Method																
A		40	69	86	96		34	66	93	100		38	82	99	100	
B		34	62	84	98		14	45	88	100		2	44	95	100	
C		39	69	85	93		46	75	88	95		39	69	91	98	
D		48	81	96	98		54	88	98	100		78	100	100	100	
E		24	63	80	87		43	69	92	97		42	86	95	100	
F		52	76	88	96		76	96	100	100		88	100	100	100	
Coefficient γ_2																
A		59	87	99	100		70	97	100	100		90	100	100	100	
B		66	93	99	100		90	100	100	100		97	100	100	100	
C		53	84	99	100		69	93	100	100		75	97	100	100	
D		58	92	100	100		86	100	100	100		91	100	100	100	
E		64	94	100	100		87	100	100	100		94	100	100	100	
F		52	84	96	100		88	100	100	100		92	100	100	100	
Coefficient γ_3																
A		17	37	54	77		22	42	56	78		32	58	76	89	
B		19	34	48	66		13	22	49	67		5	21	50	86	
C		22	50	62	74		24	49	65	80		34	58	77	88	
D		17	41	69	85		29	56	80	95		44	79	99	100	
E		16	31	46	68		24	44	61	74		25	52	78	94	
F		28	48	64	72		44	84	96	96		56	84	96	100	

TABLE 23 - Continued

		<u>Sample Size 20</u>					<u>Sample Size 50</u>					<u>Sample Size 100</u>				
		<u>Intervals</u>					<u>Intervals</u>					<u>Intervals</u>				
<u>Coefficient</u>	<u>Method</u>	$\pm.05$	$\pm.10$	$\pm.15$	$\pm.20$		$\pm.05$	$\pm.10$	$\pm.15$	$\pm.20$		$\pm.05$	$\pm.10$	$\pm.15$	$\pm.20$	
Coefficient β_0	A	9	19	25	35		1	3	4	8		0	1	1	2	
	B	8	18	29	36		5	10	15	23		2	3	11	24	
	C	16	24	35	51		23	37	46	57		13	36	56	70	
	D	16	31	39	46		21	37	61	76		37	67	82	90	
	E	9	19	23	30		8	24	37	51		22	46	69	75	
	F	24	28	36	40		16	40	56	68		52	72	80	88	
Coefficient γ_0	A	20	46	62	76		26	55	71	83		40	70	94	100	
	B	28	57	72	85		34	62	82	96		50	86	99	100	
	C	24	45	62	71		29	55	69	79		44	70	91	100	
	D	30	52	70	78		37	61	80	94		46	87	98	100	
	E	23	49	69	78		35	60	82	95		47	85	98	100	
	F	12	24	48	68		44	64	84	84		48	76	88	96	

TABLE 23 - Continued

		<u>Sample Size 20</u>					<u>Sample Size 50</u>					<u>Sample Size 100</u>				
		<u>Intervals</u>					<u>Intervals</u>					<u>Intervals</u>				
Coefficient ρ_{11}		$\pm .05$	$\pm .10$	$\pm .15$	$\pm .20$		$\pm .05$	$\pm .10$	$\pm .15$	$\pm .20$		$\pm .05$	$\pm .10$	$\pm .15$	$\pm .20$	
Method																
B	15	36	50	61			16	38	59	75		13	39	65	87	
D	14	35	54	63			23	39	61	78		18	36	63	83	
E	18	32	46	56			18	40	60	77		16	38	62	86	
F	24	36	36	44			44	60	72	80		40	72	84	92	
Coefficient ρ_{22}																
B	14	32	49	59			14	33	52	67		11	24	49	70	
D	11	29	44	57			18	31	52	78		19	45	58	81	
E	15	34	47	56			16	35	48	62		15	33	53	72	
F	20	28	32	52			36	56	76	88		40	68	88	88	
Coefficient ρ_{12}																
F	12	28	32	48			28	44	76	96		64	84	88	100	
Coefficient ρ_{21}																
F	12	32	40	52			16	52	72	84		56	84	92	92	

^aOnly 25 replications were performed for Method F, in contrast to 100 replications for the other 5 methods. Thus, the frequencies reported for Method F are the actual frequencies found in 25 replications multiplied by 4.

TABLE 25

FREQUENCY OF FAILURE TO REJECT NULL HYPOTHESIS
VECTOR AUTOREGRESSION

Null Hypothesis: Nonautoregressive Disturbances
Alternative Hypothesis: Vector Autoregression

<u>Sample Size</u>	<u>F-Test^a</u>
20	52
50	0
100	0

Null Hypothesis: Simple Autoregressive Disturbances
Alternative Hypothesis: Vector Autoregression

<u>Sample Size</u>	<u>F-Test^a</u>
20	52
50	16
100	0

Null Hypothesis: Nonautoregressive Disturbances
Alternative Hypothesis: Simple Autoregression

<u>Sample Size</u>	<u>Test Using Estimated Value of ρ^b</u>		<u>Test Using Durbin-Watson Statistic^c</u>		<u>F-Test^a</u>
	<u>Equation 1</u>	<u>Equation 2</u>	<u>Equation 1</u>	<u>Equation 2</u>	
20	37	42	78	79	68
50	4	6	9	16	0
100	0	0	0	0	0

Null Hypothesis: Nonautoregressive Disturbances
Alternative Hypothesis: Positive Simple Autoregression

<u>Sample Size</u>	<u>Test Using Estimated Value of ρ^b</u>		<u>Test Using Durbin-Watson Statistic^c</u>	
	<u>Equation 1</u>	<u>Equation 2</u>	<u>Equation 1</u>	<u>Equation 2</u>
20	28	30	67	75
50	3	4	4	8
100	0	0	0	0

^aat the 5% level of significance; the frequencies reported are the actual frequencies (25 cases) multiplied by 4

^bThe critical values of ρ are the same as those shown earlier in table 16.

^cat the 5 % level of significance

CHAPTER V

FOOTNOTES

¹R.L. Basmann, Exact Finite Sample Distributions for Some Econometric Estimators and Test Statistics: A Survey and Appraisal," in Frontiers of Quantitative Economics, Vol. II, ed. M.D. Intriligator and D.A. Kendrick (Amsterdam: North-Holland Publishing Company, 1974).

²Ray C. Fair, "A Comparison of Alternative Estimators of Macroeconomic Models," International Economic Review 14 (1973): 261-77.

³Michael D. Hurd, "Small-Sample Estimation of a Structural Equation with Autocorrelated Errors," Journal of the American Statistical Association 67(1972): 567-73.

⁴S. Goldfeld and R. Quandt, Nonlinear Methods in Econometrics (Amsterdam: North-Holland Publishing Company, 1972), Chapter 7.

⁵David K. Guilkey, "Estimation of Systems of Regression Equations with First-Order Vector Autoregressive Errors" (Ph.D. dissertation, University of North Carolina, 1973), Chapter II.

⁶Franklin M. Fisher, The Identification Problem in Econometrics (New York: McGraw-Hill Book Company, 1966), pp. 168-75.

⁷Only 25 replications were made with Hatanaka's 2SLS(RA)-VECTOR estimator in the third experiment.

⁸The GENERATE command from the MIDAS computer program was used. Documentation may be found in William A. Ericson and Daniel J. Fox, Simulation with Midas (Ann Arbor, Michigan: Statistical Research Laboratory, The University of Michigan, 1976). The manual states that "an algorithm by Brent is used to generate pseudo-random numbers from the normal distribution with mean 0 and variance 1," p. 66, and refers to Richard P. Brent, "A Gaussian Pseudo-random Number Generator (Algorithm 488)," Communications of the Association for Computing Machinery (December, 1974): 704-6. 150 values were drawn for $\varepsilon_1(t)$ and $\varepsilon_2(t)$ per sample. From these values and the autoregressive specifications, values for $u_1(t)$ and $u_2(t)$ were obtained. A subset of the last 100 values of $u_1(t)$ and $u_2(t)$ were subsequently used: the first 20 values in the subset for samples of size 20, the first 50 values in the subset for samples of size 50, and all 100 values in the subset for samples of size 100.

⁹The GENERATE command from the MIDAS computer program was again used. 100 values were drawn for $X_1(t)$ and $X_2(t)$; these same values were used in each replication. The first 20 values were used for samples of size 20, the first 50 values were used for samples of size 50, and all 100 values were used for samples of size 100. The values that were used are contained in appendix B.

¹⁰R.L. Basmann, "A Note on the Exact Finite Sample Frequency Functions of Generalized Classical Linear Estimators in Two Leading Overidentified Cases," Journal of the American Statistical Association 56 (1961): 619-36.

¹¹Arnold Zellner, An Introduction to Bayesian Inference in Econometrics (New York: John Wiley and Sons, Inc., 1971), pp. 276-87.

¹²Method F was left out of the first two experiments with simple autoregression because it is known that this method will be inefficient relative to method D. An interesting result to derive would be to determine the degree of method F's inefficiency; however, to do so accurately would require more than 25 or even 100 replications.

¹³Basmann, "A Note on the Exact Finite Sample Frequency Functions," p. 625.

¹⁴Ray C. Fair, A Short-Run Forecasting Model for the United States Economy (Lexington, Massachusetts: Heath Lexington Books, 1971).

¹⁵A similar conclusion was reached by both Cragg and Summers from the evidence of sampling experiments with simultaneous equation estimation in the context of nonautocorrelated disturbances. See J.G. Cragg, "On the Relative Small-Sample Properties of Several Structural-Equation Estimators," Econometrica 35 (1967): 109; R.M. Summers, "A Capital Intensive Approach to the Small Sample Properties of Various Simultaneous Equation Estimators," Econometrica 33 (1965): 1-41.

¹⁶Goldfeld and Quandt, Nonlinear Methods, Chapter 7.

¹⁷Hurd, "Small-Sample Estimation," pp. 567-73.

¹⁸Guilkey, "Estimation of Systems," Chapter II.

CHAPTER VI

CONCLUSION

The investigation of the estimation of simultaneous equation models with lagged endogenous variables and autoregressive disturbances, the purpose of this dissertation, is now complete. In chapter I, the logic and justifications underlying the first-order simple and vector autoregressive schemes are sketched. The rationale for the specification of autoregressive disturbances is generally weak since it relies on intuition rather than theory. If the assumption of autoregressive disturbances is accepted, however, then the popular simple autoregressive scheme is a specification both plausible and easy to interpret. The vector autoregressive scheme is more elaborate, but strains one's ability to explain the forces behind the specification. Chapter II presents a survey of the estimation methods developed to handle the compound econometric problem properly. For clarity and reference, appendix A contains two charts stressing the relationships among some familiar estimation methods in addition to those presented in the survey. Chapters III and IV record the reestimation of two simultaneous equation models with lagged endogenous variables that were originally estimated under the assumption of nonautocorrelated disturbances. The reestimation of the models involves methods included in the survey and designed under the assumption of autoregressive disturbances. For Teigen's model

the assumption of nonautoregression is rejected in favor of the assumption of simple autoregression; for Perry's model, the assumptions of both nonautoregression and simple autoregression are rejected in favor of vector autoregression. The new estimates of the models generally confirm the original conclusions of these studies. The new estimates of Perry's model have very little effect on the shape and location of the Phillips curve. In contrast, there are more noticeable differences between the estimates of the demand and supply elasticities derived from the original and new estimates of Teigen's model. The new elasticity estimates, however, are still within the range found by other studies of the money demand and supply functions. Chapter V contains the format and results of a set of sampling experiments conducted to learn more about the relative performance of various estimation methods in the context of the compound econometric problem considered here. The main results of the sampling experiments suggests that the methods designed specifically for this estimation problem are superior only when a high degree of autoregression is present and the sample contains a sufficient number of observations. Other results indicate that a) the use of the t distribution is usually a reasonable procedure to follow in the tests of hypotheses and the construction of interval estimates; and b) when the available sample is undersized, it may be wise to use an estimation method that corrects for autoregression, but disregards simultaneity.

The implication the sampling experiments have for Teigen's and Perry's models is that the new estimates may not be better

estimates of these models. Teigen's sample contains roughly 50 observations and the disturbances contain a moderate degree of simple autoregression. The results of experiment II indicate that the accuracy of the 2SLS and 2SLS-AUTO methods is comparable in this situation. A complicating factor for the applicability of the experiment's results to Teigen's model is the presence of nonlinearities, which are taken into account in the reestimation. Perry's sample also contains about 50 observations, and the disturbances are assumed to follow the vector autoregressive scheme. The results of experiment III suggest that 2SLS(RA)-VECTOR estimates should be far superior to Perry's OLS estimates, given the vector autoregressive scheme specified in the experiment. The estimates of the elements of the R matrix, however, reveal a pattern that would be difficult to compare with that in the experiment or to identify as a high or low degree of vector autoregression. It is thus even more dangerous than usual to apply the result from a sampling experiment to a specific case; therefore, the superiority of the new estimates of Perry's model is uncertain.

The results and findings of this investigation may offer researchers the following guidelines in the estimation of simultaneous equation models with lagged endogenous variables and autoregressive disturbances:

1. When faced with an undersized sample, the use of Fair's adaptation of the 2SLS-AUTO method is not recommended. This contradicts the suggestion made in an econometric textbook and in the documentation for a computer program developed specifically for

economic research.¹ These sources refer the reader directly to Fair's article when faced with the compound estimation problem.²

2. If the sample size is no problem, then an estimation method accounting for both simultaneity and simple autoregression, such as Amemiya's 2SLS-AUTO method, is cautiously suggested. When the researcher has 10 to 15 years of quarterly data, often the case, then the 2SLS-AUTO method performs as well or better than the other methods, whether the model has simple or vector autoregressive disturbances.³ This method is also easy to use. On the other hand, there seems to be little or no potential gain from employing a method designed for vector autoregression, which is also more costly to use. Besides, the assumption of vector autoregressive disturbances raises questions that would be difficult for the model builder to answer. When the sample is comprised of time series data, therefore, consideration of the vector autoregressive scheme should not be automatic, as is the practice for the simple autoregressive scheme.
3. The penalty in estimation accuracy for misspecifying the autoregressive scheme is small when the researcher has 10 to 15 (or fewer) years of quarterly data. Perhaps what is called for is a reduction in the time allocated by researchers to the consideration of autoregressive specifications for their models.

Although this study addresses a number of relevant issues, others are left untouched. More extensive sampling experiments may resolve unanswered questions concerning the performances of the alternative estimators in various situations. Comparisons of the

forecasting accuracy of the alternative estimators, rather than their accuracy in estimating the model's parameters, may be an interesting area to probe. If the appropriate techniques are developed, the vector autoregressive scheme may become a rich source of information for pinpointing the locations of specification errors in econometric models. Knowledge that a factor outside the model seems to affect certain equations at one period in one direction and then affect other equations one period later in the same or opposite direction may aid in the identification of the omitted variables.

CHAPTER VI

FOOTNOTES

¹Robert S. Pindyck and Daniel L. Rubinfeld, Econometric Models and Economic Forecasts (New York: McGraw-Hill Book Company, 1976), p. 147; Malcolm Grieg and Robin Morley, Time Series Processor (Vancouver, British Columbia: Computing Centre, The University of British Columbia, 1974).

²Ray C. Fair, "The Estimation of Simultaneous Equation Models with Lagged Endogenous Variables and First Order Serially Correlated Errors," Econometrica 38 (1970): 507-16.

³Guilkey found a parallel result when $\rho_{11} = \rho_{22} > \rho_{12} = \rho_{21}$ and $T=20$. David K. Guilkey, "Estimation of Systems of Regression Equations with First-Order Vector Autoregressive Errors" (Ph.D. dissertation, University of North Carolina, 1973).

APPENDIX A

The charts below could be called family trees of estimators. The estimators in chapter II's survey and some other familiar estimators are divided into the least squares and maximum likelihood families, and then organized to show how these estimators are related or have descended from other estimators.

LEAST SQUARES ESTIMATORS

<u>Specification for Disturbances</u>	<u>Type of Equation System</u>	<u>Limited-Information</u>	<u>Full-Information</u>
Nonautoregression ($R=0$)	Seemingly Unrelated Regressions	OLS	Zellner's Method (or Aitken's GLS)
	Simultaneous Equations	2SLS ^a	3SLS
Simple Autoregression (R is Diagonal)	Seemingly Unrelated Regressions	Cochrane-Orcutt or Hildreth-Lu Methods	Parks' Method ^b
	Simultaneous Equations	Amemiya's Method ^c	Extension of Amemiya's Method
Vector Autoregression (R is General)	Seemingly Unrelated Regressions	Guilkey's Method with Φ Diagonal	Guilkey's Method ^d
	Simultaneous Equations	Dhrymes' Method with Φ Diagonal	Dhrymes' Method ^e

MAXIMUM LIKELIHOOD ESTIMATORS

<u>Specification for Disturbances</u>	<u>Type of Equation System</u>	<u>Limited-Information</u>	<u>Full-Information</u>
Nonautoregression ($R=0$)	Seemingly Unrelated Regressions	ML	ML ^f
	Simultaneous Equations	LIML	FIML
Simple Autoregression (R is Diagonal)	Seemingly Unrelated Regressions	ML ^g	ML Counter-part to Parks' Method
	Simultaneous Equations	Sargan's Method with R and ϕ Diagonal	Sargan's Method with R Diagonal
Vector Autoregression (R is General)	Seemingly Unrelated Regressions	ML Counter-part to Guilkey's Method with ϕ Diagonal	ML Counter-part to Guilkey's Method
	Simultaneous Equations	Sargan's Method with ϕ Diagonal	Sargan's Method ^h

^aThe 2SLS and 3SLS estimators have been modified for under-sized samples; examples are the use of principal components and the Brundy and Jorgenson idea.

^bR.W. Parks, "Efficient Estimation of a System of Regression Equations When Disturbances are Both Serially and Contemporaneously Correlated," Journal of the American Statistical Association 62 (1967): 500-9.

^cSee supra, pp. 24 -27. Both the limited and full-information versions of this estimator have been modified for undersized samples. Fair suggests using either 1) a subset of the variables in the augmented reduced form (see pages 28-30 in the text) or 2) the Brundy and Jorgenson idea (see pages 36-38 in the text) to obtain fitted values for the endogenous variables appearing as explanatory variables. Dhrymes, Berner, and Cummins suggest estimating the reduced form and using a dynamic simulation to obtain the fitted values (see pages 30-31 in the text).

^dDavid K. Guilkey, Estimation of Systems of Regression Equations with First-Order Vector Autoregressive Errors," (Ph.D. dissertation, University of North Carolina, 1973, Chapter V.

^eSee pages 31-36 in the text; Hatanaka's residual adjustment innovation was devised in order to achieve efficiency without iteration in the least squares estimation of models with autoregressive disturbances and lagged dependent or endogenous variables acting as explanatory variables. Hatanaka's variation can be used in conjunction with any of the limited or full-information least squares estimators of equation systems (simultaneous equation or seemingly unrelated regressions) with simple or vector autoregressive disturbances.

^fJan Kmenta, Elements of Econometrics (New York: Macmillan Company, 1971), p. 526

^gKmenta, Elements, pp. 284-5.

^hSee pages 22-24 in the text.

APPENDIX B

The following are the values of the exogeneous variables used in the sampling experiments reported in chapter V.

<u>Index</u>	<u>X₁</u>	<u>X₂</u>	<u>Index</u>	<u>X₁</u>	<u>X₂</u>
1	11.23790	12.82440	40	7.63621	11.91650
2	6.78703	10.55580	41	8.88254	13.98040
3	8.13231	15.05180	42	10.08680	10.52600
4	9.84447	9.31937	43	9.84860	8.53525
5	10.99670	9.93914	44	8.49561	10.06440
6	8.78374	8.89821	45	8.65538	9.13652
7	7.39328	11.10020	46	10.50620	10.72330
8	11.81300	8.81532	47	14.17670	8.34658
9	12.09200	8.88236	48	9.15826	8.09318
10	13.87150	10.88350	49	10.77040	9.21584
11	7.12801	13.18810	50	13.93970	9.47083
12	12.77210	4.63257	51	9.47083	8.27633
13	8.79974	7.91461	52	7.99486	15.01270
14	10.66010	9.24818	53	10.32570	10.37430
15	14.61850	10.71360	54	8.49789	11.63430
16	11.83460	13.15850	55	11.32330	9.58203
17	8.44862	6.77363	56	10.17120	8.20637
18	10.34320	10.51270	57	8.52275	13.15990
19	11.69470	11.01630	58	9.18003	10.73400
20	6.80635	11.24050	59	9.72237	3.95689
21	11.88760	6.30854	60	9.31752	9.72957
22	8.94684	8.55931	61	12.84590	12.76130
23	8.25112	8.94481	62	12.29180	12.84820
24	12.02700	12.28480	63	7.70325	10.59380
25	6.43125	13.12030	64	9.40716	11.06940
26	5.66116	13.14210	65	8.80869	9.55737
27	9.84437	7.63310	66	8.79495	13.24220
28	9.14978	9.37900	67	7.24960	10.01580
29	6.79882	13.37350	69	13.00900	10.14780
30	6.56172	10.95680	70	7.53760	7.72182
31	11.77190	11.90830	71	9.11757	12.13060
32	10.20430	13.22070	72	12.28070	8.50314
33	7.98113	8.41475	73	8.08212	9.12959
34	10.87250	11.63980	74	11.07010	7.19188
35	10.41970	13.07370	75	8.18168	9.88751
36	9.10784	10.16010	76	9.74406	11.26970
37	8.20722	12.62680	77	10.70340	10.60470
38	12.45520	11.88260	78	8.69045	9.52762
39	10.78870	9.44528	79	10.09900	10.62190

<u>Index</u>	<u>X₁</u>	<u>X₂</u>
80	11.42700	10.32100
81	10.58930	10.35590
82	9.71065	9.20290
83	8.11979	9.22264
84	10.61360	8.76183
85	8.20736	10.33620
86	10.67630	12.56290
87	8.49137	6.78492
88	11.81880	7.46775
89	9.04828	8.54076
90	10.53530	7.47863
91	11.34880	10.05030
92	9.71325	10.19230
93	7.68015	7.97059
94	11.40230	6.45464
95	9.04266	8.89524
96	5.71814	8.83114
97	12.22150	10.46860
98	12.47270	6.74372
99	6.65500	5.84570
100	7.73649	10.35330

BIBLIOGRAPHY

- Amemiya, Takeshi. "Specification Analysis in the Estimation of Parameters of a Simultaneous Equation Model with Autoregressive Errors." Econometrica 34 (1966): 285-306.
- Basmann, R.L. "A Note on the Exact Finite Sample Frequency Functions of Generalized Classical Linear Estimators in Two Leading Overidentified Cases." Journal of the American Statistical Association 56 (1961): 619-36.
- _____. "Exact Finite Sample Distributions for Some Econometric Estimators and Test Statistics: A Survey and Appraisal." In Frontiers of Quantitative Economics, Vol. II, pp. 209-88. Edited by M.D. Intrilligator and D.A. Kendrick. Amsterdam: North-Holland Publishing Company, 1974.
- Brundy, J.M., and Jorgenson, D.W. "Efficient Estimation of Simultaneous Equations by Instrumental Variables." Review of Economics and Statistics 53 (1971): 207-24.
- Cragg, J.G. "On the Relative Small-Sample Properties of Several Structural-Equation Estimators." Econometrica 35 (1967): 89-110.
- Dhrymes, Phoebus J. "Full Information Estimation of Dynamic Simultaneous Equations Models with Autoregressive Errors." Discussion paper No. 203, University of Pennsylvania, 1971.
- Dhrymes, Phoebus J., and Erlat, H. "Asymptotic Properties of Full Information Estimators in Dynamic Autoregressive Simultaneous Equation Models." Journal of Econometrics 2 (1974): 247-59.
- Dhrymes, Phoebus J., and Taylor, J.B. "On an Efficient Two-Step Estimator for Dynamic Simultaneous Equations Models with Autoregressive Errors." International Economic Review 17 (1976): 362-76.
- Dhrymes, Phoebus J.; Berner, R.; and Cummins, D. "Limited Information Estimation of Simultaneous Equations Models with Lagged Endogenous Variables and Autocorrelated Errors." Discussion paper No. 183, University of Pennsylvania, 1970.

- Drhymes, Phoebus J.; Berner, R.; and Cummins, D. "A Comparison of Some Limited Information Estimators for Dynamic Simultaneous Equations Models with Autocorrelated Errors." Econometrica 42 (1974): 311-32.
- Durbin, J. "Testing for Serial Correlation in Least-Squares Regressions When Some of the Regressors are Lagged Dependent Variables." Econometrica 38 (1970): 410-21.
- Ericson, William A., and Fox, Daniel J. Simulation with Midas. Ann Arbor: Statistical Research Laboratory, The University of Michigan, 1976.
- Fair, Ray C. "The Estimation of Simultaneous Equation Models with Lagged Endogenous Variables and First-Order Serially Correlated Errors." Econometrica 38 (1970): 507-16.
- _____. A Short-Run Forecasting Model for the United States Economy. Lexington, Massachusetts: Heath Lexington Books, 1971.
- _____. "Efficient Estimation of Simultaneous Equations with Autoregressive Errors by Instrumental Variables." Review of Economics and Statistics 54 (1972): 444-9.
- _____. "A Comparison of Alternative Estimators of Macroeconomic Models." International Economic Review 14 (1973): 261-77.
- Fisher, Franklin M. The Identification Problem in Econometrics. New York: McGraw-Hill Book Company, 1966.
- Gandolfo, Giancarlo. Mathematical Methods and Models in Economic Dynamics. Amsterdam: North-Holland Publishing Company, 1971.
- Gibson, William E. "Demand and Supply Functions for Money in the United States: Theory and Measurement." Econometrica 40 (1972): 361-70.
- _____. "Demand and Supply Functions for Money: A Comment." Econometrica 44 (1976): 387-9.
- Goldfeld, S., and Quandt, R. Nonlinear Methods in Econometrics. Amsterdam: North-Holland Publishing Company, 1972.
- Greig, Malcolm, and Morley, Robin. Time Series Processor. Vancouver, British Columbia: Computing Centre, The University of British Columbia, 1974.
- Griliches, Zvi. "A Note on Serial Correlation Bias in Estimates on Distributed Lags." Econometrica 27 (1961): 27.

Guilkey, David K. "Estimation of Systems of Regression Equations with First-Order Vector Autoregressive Errors." Ph.D. dissertation, University of North Carolina, 1973.

_____. "A Test for the Presence of First-Order Vector Autoregressive Errors When Lagged Endogenous Variables are Present." Econometrica 43 (1975): 711-17.

Hatanaka, Michio. "An Efficient Two-Step Estimator for the Dynamic Adjustment Model with Vector Autoregressive Errors." Journal of Econometrics 2 (1974): 199-220

_____. "Several Efficient Two-Step Estimators for the Dynamic Simultaneous Equations Model with Autoregressive Disturbances." Journal of Econometrics 4 (1976): 189-204.

Hendry, David F. "Maximum Likelihood Estimation of Systems of Simultaneous Regression Equations with Errors Generated by a Vector Autoregressive Process." International Economic Review 12 (1971): 257-72.

_____. "Correction." International Economic Review 15 (1972): 260.

Hurd, Michael D. "Small Sample Estimation of a Structural Equation with Autocorrelated Errors." Journal of the American Statistical Association 67 (1972): 567-73.

Kelejian, Harry H. "Two-Stage Least Squares and Econometric Systems Linear in Parameters but Nonlinear in the Endogenous Variables." Journal of the American Statistical Association 66 (1971): 573-4.

Kmenta, Jan. Elements of Econometrics. New York: Macmillan Company, 1971.

Koutsoyiannis, A. Theory of Econometrics. New York: Macmillan Company, 1973.

Maddala, G.S. "Generalized Least Squares with an Estimated Covariance Matrix." Econometrica 39 (1971): 23-33.

Nerlove, M., and Wallis, K.F. "Use of the Durbin-Watson Statistic in Inappropriate Situations." Econometrica 34 (1966): 235-8.

Oberhofer, W. and Kmenta, J. "A Generalized Procedure for Obtaining Maximum Likelihood Estimates in Generalized Regression Models." Econometrica 42 (1974): 579-90.

- Parks, R.W. "Efficient Estimation of a System of Regression Equations When Disturbances are Both Serially and Contemporaneously Correlated." Journal of the American Statistical Association 62 (1967): 500-9.
- Perry, George L. Unemployment, Money Wage Rates, and Inflation. Cambridge, Massachusetts: The M.I.T. Press, 1966.
- Pindyck, Robert S., and Rubinfeld, Daniel L. Econometric Methods and Economic Forecasts. New York: McGraw-Hill Book Company, 1976.
- Sargan, J.D. "The Maximum Likelihood Estimation of Economic Relationships." Econometrica 29 (1961): 414-26.
- Summers, R.M. "A Capital Intensive Approach to the Small Sample Properties of Various Simultaneous Equation Estimators." Econometrica 33 (1965): 1-41.
- Teigen, Ronald L. "Demand and Supply Functions for Money in the United States: Some Structural Estimates." Econometrica 32 (1964): 476-509.
- _____. "Demand and Supply Functions for Money: Another Look at Theory and Measurement." Econometrica 44 (1976): 377-85.
- Zellner, Arnold. An Introduction to Bayesian Inference in Econometrics. New York: John Wiley and Sons, Incorporated, 1971.