VECTOR AUTOREGRESSIONS FOR TESTING CAUSAL INFERENCES OF ECONOMIC THEORY

by

Tilak Abeysinghe

A thesis
presented to the University of Manitoba
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in
Department of Economics

Winnipeg, Manitoba

VECTOR AUTOREGRESSIONS FOR TESTING CAUSAL INFERENCES OF ECONOMIC THEORY

BY

TILAK ABEYSINGHE

A thesis submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

I hereby declare that I am the sole author of this thesis.

I authorize the University of Manitoba to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Tilak Abeysinghe

I further authorize the University of Manitoba to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Tilak Abeysinghe

ABSTRACT

Utilizing Granger's definition of cauasality, this study demonstrates that vector autoregressions can be used to test contradictory causal specifications of competing economic theories. A statistical procedure is developed for the specification of vector autoregressions which serve the above purpose. This specification entails the use of model selection criteria. Based on an evaluation exercise three model selection criteria--AIC, SBC, and PHI--are recommended for model specification.

Application of this procedure to a test, using Canadian data, of two models of fertility indicates that the relative income model performs better, as judged by model selection criteria, than the male-female wage model. This exercise also indicates that the causal relationship between fertility and labor force participation of married women is unidirectional from the former to the latter, is of relatively short duration, and is weakening over time. This causal relationship may be ignored in models designed to understand long run behaviour of these two variables; however, their joint specification, taking account of their lag structures, can yield better results compared to single specifications.

ACKNOWLEDGEMENTS

I would like to express my gratitude to Professor Richard Lobdell for his constant encouragement and supervision. My thanks are also due to Professors Norman Cameron and S. K. Singha for their constructive comments on this thesis. I must also thank Professor Cheng Hsiao, the external examiner, for reading the thesis.

The Department of Economics of the University of Manitoba, the University of Manitoba Graduate Fellowship Program, and the J. W. Dafoe Graduate Fellowship Program are sincerely thanked for the financial assistance provided.

My wife, Wimala Kumari, deserves a special thank for her patience and cooperation.

I am alone responsible for the errors that may lurk in this study.

Tilak Abeysinghe
Winnipeg
May 1987

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter I

INTRODUCTION

> all induction is blind, so long as the deduction of causal connections is left out of account; and all deduction is barren so long as it does not start from observation [J.N. Keynes, 1890: 164].

1.1 PROBLEMS WITH CONVENTIONAL ECONOMETRICS

The conventional approach[1] to econometric modelling has not been successful in a number of areas. Firstly, it is well acknowledged that, so far as forecasting is concerned, the economic-theory based econometric models have performed not better, worse on some occasions, than the "naive," "atheoretical," data-based (or statistical-theory based) time series models [Harvey, 1981a: 34]. This is notwithstanding the fact that the econometric models use larger information sets in contrast to the univariate time series models. Hendry and Richard [1983: 129] observe that when an econometric model is reduced to a univariate autoregressive moving average (ARMA) model [Zellner and Palm, 1974; Wallis, 1977] the resulting variance of the innovations of the ARMA model will be greater than or equal to the corresponding innovation variance of the econometric equation. Therefore, "fitting

[1] Some aspects of the conventional approach will be clear in the course of this presentation. A standard text book in econometrics in general does not deviate from the conventional approach.

worse than a univariate time-series model is a prima facie evidence of dynamic misspecification in an econometric equation."

The above conclusion of Hendry and Richard is attributable to the "narrow" foundation, the economic theory, upon which econometric equations are built. This accusation is not unwelcome by economic theorists. Hicks [1979: 2] expresses his judgement on economic theory as: "There are few economic 'laws' which can be regarded as at all firmely based." Blaug [1980: 160-162] gathers that "At any rate... few modern economists would claim that economics has so far produced more than one or two laws."[2] In a footnote to this statement Blaug draws attention to a remark by Samuelson:

> Samuelson... remarks that years of experience have taught him how treacherous are economic "laws" in economic life: e.g. Bowley's Law of constant relative wage share; Long's Law of constant population participation in the labor force; Pareto's Law of unchangeable inequality of incomes; Davidson's Law of constant private savings ratio; Colin Clark's Law of a 25 percent ceiling on government expenditure and taxation; Modigliani's Law of constant wealth-income ratio; Marx's Law of the falling rate of real wage and/or the falling rate of profit; everybody's Law of a constant capital-output ratio. If these be Laws Mother Nature is a criminal by nature.

Secondly, the economic discipline is rife with competing theories which are hardly encountered within a single econometric framework. In other words, conventional econometrics is preoccupied with "confirming theories (via 'correct'

[2] However, neither Hicks nor Blaug states what these firmly based economic laws are.

signs and magnitudes of estimated parameters, and high R^2 values) rather than evaluating alternative theories..." [Mizon, 1984: 134]. Though there is no simple answer to the question of theory choice in economics,[3] it is disturbing to see that econometrics has not been able to shed much light on the controversies surrounding competing theories. Blaug [1980: 261], for example, sees the status quo as: "In many areas of economics, different econometric studies reach conflicting conclusions and, given the available data, there are frequently no effective methods for deciding which conclusion is correct. In consequence, contradictory hypotheses continue to co-exist for decades or more." Tarascio and Caldwell [1979: 991] remark: "few debates in such fields as industrial organization or macroeconomics have been resolved on empirical grounds, though empirical testing is prolific."

Thirdly, even if econometric models are formulated to test a single theory, these models are so dependent on the theory that the role of the econometric model is reduced to determining "the orders of magnitude of a small number of unknown coefficients" pertaining to the theory [Hendry and Richard, 1983: 113]. An obvious subject of this criticism is the econometric "identification" procedure, put into effect by the Cowles Commission program [Hood and Koopmans, 1953], and which still remains in the toolkit of the practi-

---

[3] The question of theory choice plays a key role in methodological discussions in economics. See Tarascio and Caldwell [1979] for some of the issues involved.

tioner. Under this program the identification is achieved by imposing restrictions on the "structural" parameters based on a priori theoretical considerations. However, the very restrictions we impose to identify a structure could be the ones which underlie many of the controversies.

Sims [1980; see also Liu 1960] points to the "incredible" nature of the econometric identification procedure. He disagrees with the common claim of "overidentification" of simultaneous systems. The identification, he argues, is achieved not because there is more than enough or enough prior information to impose restrictions but because the structures are forced to be identified. Kalman [1982], attacking the identification from a system theoretic view point, argues that the econometric identification chooses one canonical form by imposing restrictions based on "prejudices" (data independent assumptions) which preempt the whole idea of identification. Leamer [1985], who is not much satisfied with Sims' alternative program, yet convinced by Sims' critique, argues that the weakness of the Cowles program is its complete commitment to the selection of the exogenous variables and identifying restrictions when only a partial commitment is possible.[4]

---

[4] Leamer's alternative is a Bayesian solution though he fully appreciates the problem of choosing prior ditributions especially within the context of simultaneous equation systems.

In summary, the conventional approach to econometrics has been to use economic theory to construct an econometric model, with little or no attention paid to data properties, for the purpose of testing the theory. Since this procedure necessarily involves imposing certain restrictions on the model on the basis of a priori theoretical considerations (assumptions), it may preempt the modeler's objective by producing an artifact (the econometric model) which does not shed much light on what the theorist attempts to understand.

1.2 SYNTHESIS FRAMEWORK

Economic data are mostly comprised of time series and two approaches are at hand for modelling these time series: one is economic theory based econometrics and the other is statistical theory based time series analysis. These two approaches have developed mostly independently of each other and have remained virtually disjoint until recently. There is now a growing concern for developing a combined approach which addresses some of the questions which conventional econometrics has failed to solve.[5]

In any descipline the necessity of a logical model, a theoretical abstraction, is undeniable. It is economic theory which provides insights about economic phenomena, brings

[5] For an illuminating piece on the synthesis approach which is now getting recognition as time series econometrics, see Hendry and Richard [1983] with discussions by Deistler, Engle, Granger, Kloek, Mizon, and Newbold. They all appear to be in agreement on the fruitfulness of a synthesis.

real world complexities into an understandable and manageable framework, and postulates causal links among variables that one observes within the nonexperimental setting of the economy or the world itself. Nevertheless, the economic theories developed so far are mostly formulated to fit into idealized states, static worlds. It is, however, promising to see that theories of intertemporal optimizing behavior are beginning to develop, which may help, in future, in the dynamic specification of econometric models. After surveying the approaches hitherto directed to this end, Trivedi [1984: 183] concludes that "there are very strong reasons for not imposing strong prior information in distributed lag analysis... It seems inevitable and not at all unreasonable that practicing econometricians should resort to data-based priors." Economic theory, therefore, has to be taken as a first approximation, not as an end itself, in econometric modelling. Time series modelling, on the other hand, provides dynamic structures and reveals many data properties, mostly ignored by economic theory, which need explanation. In summary a meaningful, purely time series approach is not possible without some initial economic theory, but an econometric model specified purely on the basis of economic theory may be misleading.

It is necessary to be more specific in order to highlight succinctly the issues involved and to arrive at the objectives of this study. For this purpose, following Hendry and

Richard [1983], I will concentrate on a "stationary stochastic world" though this assumption is not necessary.[6] In the presentation to follow immediately, the formulation of the data generation process (DGP) is drawn from Hendry and Richard although remaining arguments are my own, unless acknowledgement is made.

Let z(t)[7] be an observed vector of variables of dimension K, assumed to have come from a stationary stochastic process with a certain joint density function. Further, let Z(t) indicate the observation matrix available up to time t, with N denoting the sample size. It will be assumed now that there exist appropriate and adequate transformations, if necessary, to approximate the data generation process by a normal, linear equation system. Given a finite dimensional unknown parameter vector θ and the matrix of initial conditions, Z(0), the DGP or its approximation can be expressed as

$$z(t)|Z(t-1),\theta \sim N(\mu(t),\Sigma), \quad t=1,\ldots,N \tag{1.1}$$

where

$$\mu(t) = E[z(t)|Z(t-1),\theta] = \sum_{i=1}^{p} A_i z(t-i) \tag{1.2}$$

---

[6] It is generally assumed that nonstationary series can be transformed to a stationary one without loss of information.

[7] Since the computer SCRIPTing facility used to print this output does not allow letter subscripts (or letter superscripts) the conventional letter subscripts such as t will be inserted in parentheses as in z(t).

$$\Sigma = E[(z(t)-\mu(t))(z(t)-\mu(t))'|Z(t-1),\theta]. \tag{1.3}$$

In (1.2) A_i's are square (KxK) matrices of coefficients with p being the maximum lag length, which is assumed here to be known. The A_i's as well as Σ are functions of θ. In (1.3) $z(t)-\mu(t)$ is the innovation process which is white noise by construction. By defining $u(t)=z(t)-\mu(t)$ we arrive at the dynamic specification:

$$z(t) = \Sigma_{i=1}^{p} A_i z(t-i)+u(t). \tag{1.4}$$

This can compactly be written using the lag operator L ($L^j z(t)=z(t-j)$) as

$$A(L)z(t) = u(t), \tag{1.5}$$

where $A(L)=A_0L^0+A_1L^1+...+A_pL^p$ is a polynomial matrix in L with A_0 defined as an identity matrix (i.e., $A_0=I$). The z(t) process will be stationary if det[A(L)]=0 has all roots greater than unity in absolute value [Hannan, 1970: 326].

The system (1.4) is a <u>vector autoregression</u> (VAR) in which every variable depends on the lags of its own and of every other variable. The econometric formulation of this system usually allows for contemporaneous relations, i.e., $A_0 \neq I$, which gives rise to what is known as "simultaneity." However, the system (1.5) with A_0 being unrestricted in any form is not unique. The time series approach to the "identification" of this system is to set $A_0=I$ and leave the covariance matrix of u(t) unrestricted. An alternative way to do this is to transform A_0 to a triangular matrix, hence a diagonal covariance matrix; this, however, gives the appearence of a "causal-chain."

In practice the dimension of the vector z(t) can be unoperationally high and a large number of variables in z(t) may be remotely related to the particular phenomenon the investigator is trying to understand or explain. The role of economic theory becomes quite apparent here. That is, economic theory enables us to reduce the dimension of z(t) to a very much smaller set of variables of interest. The theory also postulates the causal links among these variables, classifies them into endogenous-exogenous groupings, and provides us with models[8] for further investigation.

For notational convenience let us redefine z(t) to be the reduced vector of variables (i.e., after integrating out the unnecessary variables) of dimension k ($k<K$). Using economic theoretic arguments, z(t) can be decomposed now as $z(t)=[y(t),x(t)]'$, where y(t) is a vector of endogenous variables and x(t) is a vector of exogenous variables. Thus, the econometric formulation corresponding to (1.5) (with z(t) redefined as above and $A_0 \neq I$) can be written as

---

[8] Following Hendry and Richard a model will be defined as a "reduced reparameterization of the data generation process" [p. 117].

$$\begin{bmatrix} B(L) & C(L) \\ D(L) & F(L) \end{bmatrix} \begin{bmatrix} y(t) \\ x(t) \end{bmatrix} = \begin{bmatrix} e_1(t) \\ e_2(t) \end{bmatrix}, \tag{1.6}$$

where e_1 and e_2 are serially independent error series. Further, by the assumption of exogeneity of $x(t)$, $D(L)=0$ and e_1 and e_2 are not contemporaneously correlated.

Since $D(L)$ is assumed to be zero (1.6) reduces to

$$B(L)y(t)+C(L)x(t) = e_1(t) \tag{1.7}$$

$$F(L)x(t) = e_2(t). \tag{1.8}$$

The subsystem (1.7) is the usual dynamic "structural" formulation[9] in econometrics and the subsystem (1.8) is the one which is ignored from econometric models. Expanding (1.7) and premultiplying it by B_0^{-1} (B_0 is assumed to be of full rank) we get the "reduced" form which expresses current endogenous variables as a function of lagged endogenous and current and lagged exogenous variables. Imposing a priori restrictions on (1.7) in order to obtain a unique correspondence between the reduced form and structural form parameters is called "econometric identification."

---

[9] To comply with the text book presentation, (1.7) can be written as $B_0y(t)+B^*(L)y(t)+C(L)x(t)=e_1(t)$, where $B^*(L)=B_1L^1...$ up to max lag, or $By(t)+\Gamma x(t)=e_1(t)$, where $B_0=B$ (beta) and $\Gamma=(B^*(L)\ C(L))$ and $x(t)$ in the latter case is redefined to include all the lagged endogenous and current and lagged exogenous variables which are called predetermined.

We can now focus on more specific questions. An apparent difference between the dynamic formulations of the structural form and the reduced form is that the structural form allows for what engineers call "instantaneous coupling" or contemporaneous relationships. An important question needs to be raised here: What theretical justification is there to call one form "structural" and the other "reduced" if the only difference between the two is the presence or absence of contemporaneous relations? If we agree that cause and effect necessarily involve a time lag, however short it may be, and given that the sampling interval is exactly the same as the causal lag, then the so-called reduced form must be the one which whould correctly reflect the structure of the decision making process of agents.

Granger [1982] consideres the issue of "apparent simultaneity" which gives rise to much of the contemporaneous relationships (or instantaneous coupling) observed among variables. He recognizes three causes, among possible others, which can account for instantaneous coupling: (1) the presence of overlapping variables which results from aggregation over space, agents, or commodities; e.g., M1 and M2 money supplies are two overlapping variables; (2) the sampling interval is longer than the minimum causal lag which creates a problem of aggregation over time; (3) missing causal variables. All of these, obviously, demand contemporaneous relationships in econometric models to take account of data problems, but such an allowance for contemporaneous relations in the model should not imply that the actual causal lag is instantaneous. Nevertheless, there are situations where the causal lag is so small that for all theoretical and practical reasons this lag could be ignored and the models can be constructed using contemporaneous relations alone. For example, a consumer going to buy a kilogram of beef in a supermarket will respond to price (i.e., decides to buy or not) almost instantaneously.

If the difference between structural and reduced forms does not lie in the presence or absence of contemporaneous relations in the model, then the issue is the meaning of "structural." Leamer [1985] has discussed the unsatisfactory nature of the text book definitions of the concept "structural," and points to the definition used by Koopmans and Bausch, and by Hurwikz, which has been resurrected by Sims [Sims, 1986] in consequence of Lucas' well known critique. According to this definition a parameter is structural if it is invariant under a certain class of interventions. Leamer goes on to point out further:

> the traditional distinction between the "structural" form and the "reduced" form of a simultaneous-equations system makes sense only if you are imagining a modification [intervention] that alters one of the "structural" equations and leaves all others unchanged. Such a modification alters the reduced-form parameters but leaves unchanged most of the "structural"-form parameters. But, as pointed out by Sims (1977[b]), there are many hypothetical modifications that can be expected to leave the reduced-form parameters unchanged and these parameters can equally well be regarded as structural. [p. 265]

Once the contemporaneous relations in (1.7) are allowed, we face the problem of econometric "identification" to which we have already refered in the previous section. An important related problem is the ignoring of (1.8) under the assumption of exogeneity. If Lucas' [1976] critique is taken seriously, then the parameters in (1.7) will no longer be "structural"; they depend on the underlying structural parameters and the parameters of (1.8). Moreover, there are serious concerns that even the exogeneity specifications imposed by the economic theory need also to be tested [Geweke, 1980a].

It is in light of these problems that Sims has favoured estimating an unconstrained VAR in which all the variables are considered endogenous. This formulation, though not structural from a conventional viewpoint, has a wide range of applications.[10] Forecasting is a well established application of models such as VAR in time series analysis. In response to the claims of Sargent and of Leamer that the VAR cannot be used for policy analysis, Sims [1986] has demostrated how such an exercise could be carried out using the VAR approach. The VAR procedure also allows us to formulate an encompassing framework to test competing theories without having to impose a priori restrictions favouring one theory over another.

---

[10] See Cooley and LeRoy [1985] for a summary and a critique of different intended uses of the VAR approach.

Sims' procedure, however, has been characterized by some [Cooley and LeRoy, 1985; Leamer, 1985; see also comments on Leamer by Eichenbaum] as atheoretic on the ground that it pays no regard to economic theory. This criticism may be relevant, but only partially. The VAR procedure cannot choose the relevant variables without utilizing a theory. Given the relevant variables, one can specify a VAR model without the use of economic theory. If one perceives this as a substitute for economic theory, then the above criticism has its relevance. Such a perception, however, misses an important difference between VAR procedure and economic theory. As we have already pointed out, the former is a statistical procedure which tries to capture the correlation structure of the variables within a given sampling period. The latter, on the other hand, has a much more general objective of explaining the underlying causal structure which gives rise to the observed correlations among the variables. Thus, the VAR procedure, in my opinion, is an important econometric tool, not an atheoretic substitute for economic theory.

If the objective of the VAR, as is of ours, is to formulate an encompassing framework for testing competing economic theories, we propose the following procedure. First construct the economic model of interest with the endogenous-exogenous classification of the variables. Then compare this model with competing models to identify common grounds. If there are variables which are considered exogenous by all the models, then these variables may be held exogenous in the VAR model, too.[11] This reduces the dimension of the VAR model and helps conseve degrees of freedom since we can limit the dimension of the VAR to the number of variables endogenous in all the economic models under consideration. Then, without imposing any further restrictions of any economic model under investigation, specify and estimate the VAR using a statistical approach (see Chapter 3). The VAR model thus estimated will show the interdependence among the variables which are endogenous and also the influence of the common exogenous variables on these endogenous variables. This provides a test of competing hypotheses within an encompassing framework.

1.3 OBJECTIVES OF THE STUDY

One important limitation of the VAR approach is the absence of a procedure for specification of the lag structure of the variables involved. As mentioned earlier, economic theory is of little help in dynamic specifications. A common practice followed by the users of VAR is to assume a common lag length for all the variables in the model. Thus, in (1.4) p is a common lag length assumed to be valid for all the k variables. However, in reality these variables are likely to have different lag lengths and a procedure must be developed

[11] This specification is slightly different from that given in (1.4) where all the variables are endogenous to the VAR model.

to estimate these lag lengths.

More specifically, since there are k variables in (1.4) there are (kxk) lag lengths to be estimated. These lag lengths may be expressed in terms of a lag length matrix P* which is of dimension (kxk) for (1.4). The typical element p_{ij} of P* is equal to the lag order of the polynomial $a_{ij}(L)$ of A(L) in (1.5). One objective of this study is to develop a statistical procedure to specify a VAR model so that the model will have an estimated lag length matrix P* instead of an assumed common lag length p.

The second objective of the study is an application of the above procedure to evaluate competing causal specifications of economic theories of fertility concernig the observed negative association between income and fertility.[12] The concept of causality to be utilized here is that of Granger's.

1.4 ORGANIZATION OF THE STUDY

Implicitly this study consists of two parts excluding the final chapter. Chapters one to three constitute the first part and chapters four and five the second. Although chapter one provides the background for the whole study, it bears more closely on chapters two and three. The first part is basically a theoretical exercise in time-series economet-

[12] The application of VAR procedure has been vary much limited to macroeconomics and to my knowledge there has been no such an exercise carried out in fertility economics.

rics culminating with the first objective being fulfilled. The second part is an empirical application as specified in the second objective.

Since Granger's concept of causality is the basic ingredient of the causality tests to be used in the study, chapter two is devoted to an examination of Granger's concept and the tests based on it. Having stated the relationship between Granger-causality conditions and the coefficients of a VAR, chapter three develops a statistical technique to specify VAR models. The chapter is concludes by demonstrating the application of the technique to a simulated experiment. Chapter 4 contains a critical evaluation of two economic theories of fertility while a test of these theories, using Canadian data, is the subject of chapter five. The final chapter provides a summary and conclusions of the study. The bibliography consists of three parts: the literature on causality and time series econometrics utilized to compose the first three chapters, fertility and related literature referenced in chapters four and five, and sources of data for chapter five.

Chapter II

TESTS OF CAUSALITY BASED ON GRANGER'S DEFINITION

It is not intended in this chapter to get involved in a discussion of the intricate issues surrounding the concept of causality. Taylor [1967] has attempted to summarize the philosophical definitions of "causality" since the time of Aristotle, while Hicks [1979] tries to clarify the usage of this concept in economics. The objective of this chapter is to examine the highly specific concept of causality proposed by Granger [1963, 1969, 1980, 1982], and to review the derived testing procedures. Granger's definition of causality has gained popularity in the last decade and a half because of its operational nature.

2.1 GRANGER'S DEFINITION OF CAUSALITY

Granger's concept of causality is built upon a predictability criterion proposed by Wiener. This definition relies on a <u>central</u> <u>axiom</u> that the future cannot cause the present or the past, whereas an allowance is made for future expectations to cause the present, given that these expectations are based on past information. The following is a restatement of this definition based on Granger [1982].

Suppose we are intererested in whether the vector series $X(t)$ of dimension p causes another vector series $Y(t)$ of dimension q. Let $U(t)$ be the universal information set as of time t and let $U(t)-X(t)$ denote all this information excluding $X(t)$. Further, let F and E denote the conditional distribution function and statistical expectation respectively.

<u>General</u> <u>Definition</u>:

$X(t)$ causes $Y(t+1)$ if

$$F[Y(t+1)|U(t)] \neq F[Y(t+1)|U(t)-X(t)] \tag{2.1}$$

That is, if the extra information in $X(t)$ alters the conditional distribution of $Y(t+1)$, then $X(t)$ is said to cause $Y(t+1)$.

In practice $U(t)$ is not known fully, hence we have to use a smaller information set, $I(t)$, which includes $X(t)$ and $Y(t)$. Granger reserves the term <u>prima</u> <u>facie</u> <u>cause</u> to refer to this situation in order to recognize the possibility that the assessment of causality may change when the information set is changed. Further, the specification of the whole distribution of $Y(t+1)$ may not be possible; therefore, one may define causality in terms of a necessary condition, i.e. causality in mean.

<u>An Operational Definition</u>:

$X(t)$ is said to be the <u>prima facie cause in mean</u> of $Y(t+1)$ with respect to the information set $I(t)$ if

$$\Delta(t)[Y|I] = E[Y(t+1)|I(t)] - E[Y(t+1)|I(t)-X(t)] \tag{2.2}$$

is not identically zero. If it is identically equal to zero, then $X(t)$ does not cause $Y(t+1)$.

The causality in mean can be associated with improved ability to forecast in a "least squares" sense. For this we assume that the above means are the optimum, unbiased least squares predictors of $Y(t+1)$ given the corresponding information sets.

Let

$$e(t)[Y|I-X] = Y(t) - E[Y(t)|I(t-1)-X(t-1)] \tag{2.3}$$

be the one-step forecast error of $Y(t)$ given the information set $I(t-1)-X(t-1)$. That is, $e(t)$ is that part of $Y(t)$ which cannot be predicted by $I(t-1)-X(t-1)$. $e(t)$ in (2.3) is a zero mean, white noise series.

Similarly let

$$e(t)[Y|I] = Y(t) - E[Y(t)|I(t-1)] \tag{2.4}$$

be the one-step forecast error given $I(t-1)$.

Now (2.3) - (2.4) yields

$$e(t)[Y|I-X] = e(t)[Y|I] + \Delta(t-1)[Y|I] \tag{2.5}$$

where $\Delta(t-1)$ (defined in 2.2 above) is a known quantity at time t-1 and is uncorrelated with $e(t)[Y|I]$, which follows from the definition of $e(t)[Y|I]$.[13]

If Σ_1, Σ_2, and Σ_3 denote the covariance matrices of the three quantities from left to right in (2.5) then we have

$$\Sigma_1 = \Sigma_2 + \Sigma_3 \tag{2.6}$$

If X(t) causes Y(t+1), then $\Delta(t-1)$ will not be zero (see eq. 2.2) and Σ_3 will be a positive definite matrix and so is

$$\Sigma_1 - \Sigma_2. \tag{2.7}$$

This implies that point forecasts of Y(t) can be improved if X(t) is used in the prediction, i.e. the forecast error variance of each Y(t) will be smaller if X(t) is used in the prediction.

Different aspects of this definition can easily be demonstrated in a bivariate setting. For this purpose let us redefine I(t) to consist of only two series, x(t) and y(t), which are now assumed to be covariance-stationary[14] scalar

---

[13] Granger observes that the decomposition of original forecast errors (eq. 2.5) shows that white noise residuals cannot always be interpreted as random shocks if there are ommitted causal variables.

[14] A time series is said to be covariance stationary (or wide-sense or weakly stationary) if its mean and variance do not change through time, and the covariance of the series at two time points will depend only on the distance

time series. Further, let $\bar{x}(t)$ and $\bar{y}(t)$ be the past values of $x(t)$ and $y(t)$ while $\bar{\bar{x}}(t)$ and $\bar{\bar{y}}(t)$ be their past and present values. Let σ^2 be the predictive error variance.

Definition 1:

x is said to cause y if $\sigma^2(y/\bar{y},\bar{x}) < \sigma^2(y/\bar{y})$. That is, x is said to cause y if y is better predicted by both past values of y and x than by past values of y alone.

Definition 2:

Feedback is said to occur if $\sigma^2(y/\bar{y},\bar{x}) < \sigma^2(y/\bar{y})$ and $\sigma^2(x/\bar{x},\bar{y}) < \sigma^2(x/\bar{x})$. That is, both x and y are causal to each other.

Definition 3:

x is said to cause y instantaneously if $\sigma^2(y/\bar{y},\bar{\bar{x}}) < \sigma^2(y/\bar{y},\bar{x})$. That is, if current x helps in the prediction of current y then instantaneous causality is said to occur. Instantaneous causality also implies instantaneous feedback in this definition.[15]

---

between these time points and not on time itself. Two time series, $x(t)$ and $y(t)$, will be jointly covariance stationary if, at lag k, $\mathrm{cov}[x(t),y(t-k)]$ is independent of time t for all k and means of x and y are finite. The joint stationarity implies individual stationarity.

[15] It is worth noting that instantaneous causality is not possible by definition. However, it could arise in practice when the minimum causal lag is less than the sampling interval and/or when there are missing causal variables and/or when there are overlapping variables (see Section 1.2).

2.2 RELEVANCE OF GRANGER'S DEFINITION

The apparently pure statistical nature of Granger's definition has led some [e.g. Pierce, 1977] to use it as a substitute for economic theory in the specification of causality, hence to criticism by others [Zellner, 1979; Conway et al., 1984; Leamer, 1985]. Although a predictability test could reduce the possibility of associating a spurious correlation with causality, still we need a theory to identify such associations as causality. After surveying the literature on causality, Zellner [1979] suggested that a better definition of causality would be that proposed by Fiegl [1953], i.e. predictability according to a law (or set of laws). By law, Zellner means an empirical generalization rather than a "pure" deductive conclusion [Zellner 1981].[16]

We can see, however, that this definition is not at variance with that of Granger's. Granger [1980] argues that his method can be viewed as "informal Bayesian", i.e. one may start with some prior beliefs about causality between variables, and change (adjust) these beliefs after a causality test. An important source of such prior beliefs is economic theory. Since the operational definition of causality uses a limited information set, a theory is essential in suggesting the relevent variables. If one can believe that the theory

---

[16] Zellner sees pure deduction, the propositions of which have no reference to the real world, as "denk spielen" [Zellner, 1981]. Hicks [1979, ch. 3] also argues that if a theory refers to facts at all, then deduction is a bridge between two inductions.

is "firmly based," then the causality imposed by the theory on the model can be taken for granted and there is no need to refer to "Granger causality." Thus the conventional tests of regression coefficints are only "conditional tests" since they take the causal specifications of a particular theory as correct [Granger, 1980]. However, the absence of "firmly based" theories and the presence of contradictory theories in economics demand "unconditional causality tests" in order to differentiate between such theories and to specify properly the econometric models. If we accept the proposition that the presence of causation implies predictability, then the causal implications of economic theories can be brought into a proper test by investigating their predictive performance using Granger's definition.

Another use of Granger's definition is to invoke predictive relations which may lead to discovery of new laws or theories. This is a form of what Zellner [1981] calls "reductive inference;" i.e. the recognition of unusual and surprising facts leading to the formulation of new theories. A well known example of this are the theories of consumption prompted by Kuznet's observations.

Despite these uses of Granger's definition, it is not an all encompassing definition of causality. Granger himself admitted that his definition may not be acceptable to philosophers [Granger and Newbold, 1977; Granger, 1980]. One important aspect of Granger's definition is its dependence on time in an asymmetric way: the assumption that time runs unidirectionally. Although recent developments in theoretical astrophysics may dispute this assumption, the relevence of time asymmetry (hence Granger's definition) in many areas including economics is hard to deny. Simon [1953], though his definition of causality is not the same as that of Granger's, argues that the necessary ingredients of a definition of causality are a functional relation and the asymmetry of the relationship. He continues:

> There is no necessary connection between the asymmetry of this relation and asymmetry in time, although an analysis of the causal structure of dynamic systems in econometrics and physics will show that lagged relations can generally be interpreted as causal relations.[73-74]

Hicks [1979] also emphasizes the same point by arguing that the work of economists is in time, "historical time;" hence in economics reference of cause and effect to time is hardly deniable. Geweke [1982, 1984] gathers arguments from other fields to strengthen the central axiom of Granger's definition. For example, he quotes Bunge, who, talking about relativity, argues that the cause must preceed or at most be simultaneous with the effect. In any case to avoid the possibility of other connotations of the word causality we will adopt hearafter the now common term "Granger causality" or "G-causality."

2.3 GRANGER-CAUSALITY TESTS IN A BIVARIATE SETTING

G-causality can be expressed in a number of equivalent ways and these have been summarized by Pierce and Haugh [1977; see also Skoog, 1976] under three theorems.[17] These equivalent conditions have led to the emergence of different test procedures which in practice have yielded contradictory results [see Feige and Pearce, 1979]. In this section we will review four of these tests which have been commonly used in applied work.

2.3.1 Sims Test

Let (x,y) be jointly covariance stationary (see footnote 14), a purely indeterministic process with mean zero.[18] Then Sims' [1972] Theorem 1 states that y does not cause x in Granger's sense if, and only if, in the vector moving average (Wold) representation

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} a(L) & b(L) \\ c(L) & d(L) \end{bmatrix} \begin{bmatrix} u(t) \\ v(t) \end{bmatrix} \tag{2.8}$$

---

[17] See Price [1979], Pierce and Haugh [1979], Evan and Wells [1983], and Layton [1984] for further discussions of these theorems.

[18] In practice x and y may be stationarity-induced variables through transformations. Purely indeterministic means that the best linear forecast p period ahead of the joint process conditional on the past values is going to be, as p is increased without limit, the unconditional mean of the process itself.

$b(L)$ can be chosen identically zero.[19] Here $a(L)$ etc. are polynomials in the lag operator L (i.e., $a(L) = \sum_{j=0}^{\infty} a_j L^j$) and u and v are mutually uncorrelated white noise innovation processes.

Sims' Theorem 2 states that y can be expressed as a distributed lag of current and past x's (with no future x's) with a disturbance process that is orthogonal to past, present, and future x's if, and only if, y does not cause x in Granger's sense.[20] Thus, when $b(L) = 0$ (2.8) reduces to

$$x(t) = a(L)u(t)$$

$$y(t) = c(L)u(t) + d(L)v(t)$$

and

$$\begin{aligned} y(t) &= c(L)a^{-1}(L)x(t) + d(L)v(t) \\ &= v(L)x(t) + f(t) \end{aligned} \tag{2.9}$$

where $v(L)$ is one-sided (no negative powers of L enter). As a result a regression of y on future, current, and past x's must result in zero regression coefficients for future x's. Hence, Sims' test is to write (2.9), with pre-determined po-

[19] Two versions of the proof of this theorem, one due to Sargent and the other due to Sims, have been provided by Sargent [1979, Ch. 11]. Given that $b(L)$ is zero, premultiplying (2.8) by the inverse (assuming its existence) of the polynomial matrix in (2.8) will yield the autoregressive representation of (2.8) with a lower triangular polynomial matrix. This will indicate the claim of Sims' theorem.

[20] Sims' Theorem 2 shows that Granger noncausality coincides with strict econometric exogeneity. For further discussions of this see Sims [1977b], Zellner [1979], Geweke [1981a], Granger [1981], Hendry and Richard [1983], and Wu [1983].

sitive integers p and q, as

$$y(t) = \sum_{j=-q}^{p} v(j)x(t-j) + f(t), \tag{2.10}$$

and to test the null hypothesis that $v(j) = 0$ for $j<0$.

If f(t) in (2.10) is serially independent, this test can be conducted by constructing the F-statistic from the restricted and unrestricted models in the usually way. However, f(t) is not serially independent (see 2.9) and this poses serious problems in the application of Sims' test. Sims suggests either to prewhiten the data series or to use a GLS procedure. But prewhitening requires considerable skill and Feige and Pearce [1979] find that the test results are quite sensitive to the choice of a prewhitening filter. The GLS procedure, on the other hand, demands restrictions on the variance-covariance matrix of the disturbance term. Under these circumstances the statistical testing has to be carried out using a test statistic such as likelihood ratio (LR), Wald (W), or LaGrange multiplier (LM) which converge in distribution to a chi-square variate asymptotically.[21] Apart from these difficulties one also has to specify the lead and lag lengths, q and p, in (2.10). The common practice is to set them arbitrarily; we will see later that this can lead to unsatisfactory results.

---

[21] See Evans and Savin [1982] for a study on a possible conflict among these test statistics due to the presence of a systematic numerical inequality $W \geq LR \geq LM$ under certain circumstances. See also Engle [1984] for a simple presentation and a survey of these principles.

2.3.2 Modified Sims Test

Geweke, Meese, and Dent [1983] have suggested a modification to Sims' test, which avoids the need of using prewhitening filters or GLS procedures required for the original Sims' test. In the modified test, (2.10) is written as

$$y(t) = \sum_{j=1}^{r} a(j)y(t-j) + \sum_{j=-q}^{p} v(j)x(t-j) + w(t) \tag{2.11}$$

where the disturbance term $w(t)$ is serially uncorrelated by construction. After determining r, q, and p (which remains as a problem), the null hypothesis that y does not cause x, i.e. $v(j) = 0$ for $j<0$, can be carried out using LR, W, or LM test statistics, as in Sims' test.

2.3.3 Granger Test

Assuming that the (x,y) process defined above has the vector autoregression representation

$$\begin{bmatrix} A(L) & B(L) \\ C(L) & D(L) \end{bmatrix} \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} u^*(t) \\ v^*(t) \end{bmatrix}, \tag{2.12}$$

where A(L) etc. are polynomials in the lag operator L (i.e. $A(L) = \sum_{j=0}^{\infty} A_j L^j$) and $A_0=D_0=1$ $B_0=C_0=0$ and $u^*(t)$ and $v^*(t)$ are white noise processes possibly mutually contemporaneously correlated, Granger [1969] proves that y does not cause x if, and only if, $B(L) = 0$. Further, it can be shown that whether or not $B(L) = 0$, the instantaneous causality (or instantaneous feedback) is present if, and only if, $\text{cov}(u^*(t), v^*(t)) \neq 0$ [see Skoog, 1976: 25-27].

Unlike Sims' result, which is, of course, mathematically more general, Granger's result implies an immediate statistical test. Thus, given that $B(L) = 0$, the null hypothesis that y does not cause x can be tested by regressing x on past x's and past y's and testing whether the regression coefficients of past y's are zero as a group.[22] This test can be carried out using LR, W, or LM test statistics as in Sims' test.[23]

Even in this test one has to specify the lag lengths of x and y. The common practice is to set them arbitrarily. However, the ommission of relevent lags may produce serial correlation in the error term. Further, unless one has strong reasons to believe in unidirectional causality, it becomes necessary to estimate the full model (2.12) to take account of feedback effects.

2.3.4 Pierce-Haugh Cross Correlation Test

Pierce and Haugh [1977] presented an alternative test procedure for G-causality based on work by Haugh [1976, original 1972]. Their approach is to characterise G-causality by means of the cross correlation function between the univariate innovation series of x(t) and y(t). The following is only a verbal presentation of their procedure; the theoreti-

---

[22] Sargent [1976] was the first to apply this test.

[23] Although it is possible to carry out a separate test on instantaneous G-causality the direction of causality cannot be established purely on statistical grounds.

cal basis of characterizing G-causality by means of univariate innovations is found in the above citation. The logic of using the univariate innovations is that they are the components of the original series which cannot be predicted by their own past; hence the association (contemporaneous and lagged) between the innovations of different series must reveal the causal pattern between the original series as is intuitively understood by Granger's definition.

The Pierce-Haugh procedure involves the following steps:

1. Transformation of the original series to induce stationarity.
2. Fitting univariate autoregressive moving average (ARMA) models to the transformed series using Box-Jenkins approach and estimation of the univariate innovations.
3. Estimation of the cross correlations between univariate innovation estimates and testing for causality.

If we redefine $u(t)$ and $v(t)$ as the univariate innovations of the stationary series $x(t)$ and $y(t)$ (u and v here are different from those in 2.8) then the cross correlation at lag k can be estimated as

$$\hat{r}(k) = \sum_t \hat{u}(t-k)\hat{v}(t)/[\sum_t \hat{u}^2(t)\sum_t \hat{v}^2(t)]^{\frac{1}{2}}, \quad k=0,\pm 1,\ldots \qquad (2.13)$$

where ^ indicates the estimated values. Thus, significant cross correlations for $k>0$ imply that x causes y, because the past x-innovations in (2.13) are significantly correlated with current y-innovation. Other causality events can be stated in a similar way.

Note that if the true innovations, $u(t)$ and $v(t)$, were known, then we can estimate the sample cross correlations $r(k)$ between $u(t)$ and $v(t)$. Haugh [1976] has shown that under the null hypothesis of series independence the cross correlation estimates $\hat{r}(k)$ and $r(k)$ have the same asymptotic distribution, that is, they are independently normally distributed with zero means and variance N^{-1}, where N is the sample size. Therefore, the overall independence of the two series can be tested using the statistic

$$S_1 = N \sum_{k=-m}^{m} \hat{r}^2(k), \tag{2.14}$$

which has a chi-square distribution with $2m+1$ degrees of freedom, where m is a pre-specified integer. Further, Pierce [1977] has proposed the statistic

$$S_2 = N \sum_{k=1}^{m} \hat{r}^2(k), \tag{2.15}$$

which has a chi-square distribution with m degrees of freedom, to test for unidirectional causality.

Since the univariate innovations are white noise the cross correlations and the regression coefficients of these innovations are proportional to each other. Pierce and Haugh advocated cross correlation analysis since it provides a symmetric procedure to examine the causality between the two series.

Pierce [1977] applied the cross correlation analysis to test for causality between a number of monetary variables and found rather weak relationships thus refuting the strong relationships postulated by certain economic theories. This has led others to reexamine the whole procedure. Schwert [1979], for example, has shown that even if the true innovations were known, the distributed lag structure between innovations can be substantially different from that of the original series. Nelson [1979] has indicated that the weak relationship found between the innovation series is consistent with the strong relationship between the original variables because the cross correlation procedure substitutes an estimation of an unlimited number of correlations for tests based on a small number of parameters. Further remarks on the theoretical properties of the procedure are found in Sims [1977a], Granger [1977], Schewert [1979, Appendix B], Pierce [1977], and Price [1979]. Finally, as in Sims' test, the required pre-whitening procedure is not simple.[24]

2.3.5 Small Sampling Properties of Causality Tests

As reported in previous sections, causality tests have not always produced compatible results in applied work and this is quite likely a result of the difference in relative size and power of different tests in small samples. Unfortunate-

---

[24] Some important issues related to prewhitening are discussed in Chatfield and Prothero [1973], Newbold and Granger [1974], Prothero and Wallis [1976], Davis, Triggs, and Newbold [1977], and LJung and Box [1978].

ly, all the test statistics discussed above have known distributions only asymptotically and only under the null hypothesis of no causality. Therefore, a comparision of the tests when the alternative is true is not possible. Geweke [1981b], and Geweke et al. [1983] have used an indirect method called "approximate slopes of the tests", which is also asymptotic, to compare the tests under the alternative. In any case, the small sampling properties of these tests have to be evaluated, given the current state of knowledge, only through sampling experiments.

The findings of three such experiments, Nelson and Schwert [1981], Guilkey and Salemi [1982], and Geweke et al. [1983], show that, in sample sizes ranging from 50 to 200, the Granger test and the modified Sims test are more powerful than the Sims and the ARMA residual tests (cross correlation as well as regression tests).[25] Nelson and Schwert find that the ARMA residual tests are less powerful, even if the true innovations were known. Sims test appears to be quite sensitive to the methods used to correct for serial correlation. Another interesting finding of these studies is that the overparameterization of the regression models, the

---

[25] Only Nelson and Schwert, among the above three groups, have investigated the ARMA residual tests. In another study Geweke [1981b] has compared the cross correlation test with the modified Sims test as tests of series independence and finds that the latter performed better. Battaglia and Carlucci [1982] also have conducted a simulation experiment on causality tests and find mixed results. They caution against the inadequate treatment of serial correlation in the regression models.

practice arbitrarily followed in applied work to avoid the possibility of omitting relevent lags, leads to loss of power.

In addition to the power of the Granger and the modified Sims tests, they are easy to implement since only OLS estimation is required. Of the two, the Granger test is the simpler which also tends to save more degrees of freedom. Guilkey and Salemi also find that the Granger test consistently outperformed the modified Sims test by small amounts; thus they strongly recommend the Granger test to the practitioner. However, in order to retain more power due care must be given to proper parameterization. Another advantage of the Granger test is that it can easily be extended to a multivariate setting, which will be the subject matter of the next section.

2.4 GRANGER-CAUSALITY TESTS IN A MULTIVARIATE SETTING

The bivariate G-causality tests discussed above are essentially single equation tests though they warrent estimating two seperate equations for x and y to test for feedback. However, as can be seen from models (2.8) and (2.12), a more appropriate method would be to estimate the system simultaneously within the vector time series framework.[26] The rele-

[26] Caines and Chan [1975] is an early example which uses the vector time series approach to test for G-causality. Unlike Sims, Caines and Chan test the hypothesis that y does not cause x, which they call "x is feedback free," by constructing the likelihood ratio statistic from the constrained and unconstrained vector ARMA models.

vence of this approach becomes much clear when there are more than two variables involved.

Application of G-causality to multivariate settings is hampered by the complexities involved in vector time series modelling.[27] Skoog [1976] has extended the bivariate propositions to the multivariate setting and investigated the theoretical aspects of bivariate causality in a trivariate system. Hsiao [1982] has further elaborated the trivariate system (which can also be a tri-vector system) and set forth the conditions under which direct, indirect, and spurious causality could occur. Rather than reiterating the symbolic definitions of Hsiao, we will explain these definitions verbally and then characterize them in a vector autoregression.

If the G-causal relationship between two variables is invariant with a change of the information set, then _direct causality_ between them is said to occur.[28] As understood by the term itself, _indirect causality_ between two variables occurs through some other variable(s). The distinction between direct and indirect causalities in practice requires the assumption that the largest information set one considers includes all the relevant variables. Hsiao has defined

---

[27] Multiple time series modelling has received considerable attention in recent years [see Newbold, 1981 and 1984 for references]. This will definitely encourage studies in multivariate G-causality tests.

[28] It is worth recalling Granger's term _prima facie cause_ (Section 2.1) which he has reserved for situations in which causality changes with changing information sets.

two types of spurious causality, one of which had already been pointed out by Granger [1969]. Type I spurious causality is said to be present if a variable is found to be causal only in a larger information set, but not in a smaller information set. This will be illustrated shortly by means of an example. Type II spurious causality could occur due to the ommission of a causal variable which is, in fact, the cause of the correlation between the two variables under investigation. An example given by Granger [1980] is "lightening and thunder"; the former appears to be the cause of the latter whereas both are caused by a third variable, atmospheric conditions. This type of spurious causality could be detected by moving to a larger information set.

The above causality relations can easily be characterized by zero constraints on a vector autoregression.[29] Consider the trivariate (or the trivector) covariance stationary system $X = (X_1, X_2, X_3)'$ which is assumed to have the vector AR representation $A(L)X(t) = u(t)$ with $A(0) = I$, i.e.,

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} a_{11}(L) & a_{12}(L) & a_{13}(L) \\ a_{21}(L) & a_{22}(L) & a_{23}(L) \\ a_{31}(L) & a_{32}(L) & a_{33}(L) \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}, \tag{2.16}$$

---

[29] Skoog [1976, p.51] observes that the vector autoregression is superior to vector moving average representation in characterizing multivariate causality. Kang [1981] has presented the necessary and sufficient conditions for G-causality in vector ARMA models.

where $a_{ij}(L) = \sum_{l=1}^{\infty} a_{ijl} L^{l}$ and (u_1, u_2, u_3) are white noise processes possibly mutually contemporaneouly correlated such that $\mathrm{cov}(u_1, u_2, u_3) = \Sigma$. (The subscript t has been ommited for convenience.) Then the following G-causality characterization is possible:

1. X_i does not cause X_j directly if, and only if, $a_{ji}(L) = 0$, $(i,j = 1,2,3)$.
2. X_3 does not cause X_1 if, and only if, $A(L)$ is lower block triangular.
3. X_3 causes X_1 indirectly (but not directly) if, and only if, $a_{13}(L) = 0$ and $a_{12}(L)$ and $a_{23}(L)$ are not zero for some lags.
4. There is type I spurious causality from X_3 to X_1 if, and only if, $a_{32}(L) = 0$, $a_{13}(L) \neq 0$ for some lags and there exists a nonzero $c(L) = c_0 + c_1 L + c_2 L^2 + \ldots$ such that $[-a_{12}(L) - a_{13}(L)] = c(L)[1 - a_{22}(L) - a_{23}(L)]$.
5. There is type II spurious causality from X_3 to X_1 if, and only if, $a_{13}(L) = 0$, $a_{23}(L) = 0$ and $a_{12}(L)$ and $a_{32}(L)$ are not zero for some lags.[30]

---

[30] Proposition (1) is the same as Skoog's [1976] Proposition (6), condition (i) (p. 50) except that proposition (1) above allows room for the possibility of indirect causality. Proposition (2) is a direct extension of the bivariate theorems proven by Granger [1969], Sims [1972], Caines and Chan [1975], Skoog [1976], and Pierce and Haugh [1977]. The proof of propositions (3) to (5) are found in Hsiao [1982].

Since proposition (4) may not be straightforward, we illustrate its meaning with the following example. Consider the following trivariate AR(1) structure which satisfies the conditions of proposition (4):

$$\begin{bmatrix} 1-aL & -bL & -cL \\ 0 & -bL(1-eL)^{-1} & -cL(1-eL)^{-1} \\ 0 & 0 & 1-dL \end{bmatrix} \begin{bmatrix} y \\ x \\ z \end{bmatrix} = \begin{bmatrix} u \\ v \\ w \end{bmatrix}, \qquad (2.17)$$

where u, v, and w are the innovations.

Notice that $[-bL\ -cL] = (1-eL)[-bL(1-eL)^{-1}\ -cL(1-eL)^{-1}]$, thus c(L) in proposition (4) is equal to (1-eL). In (2.17) y and x can be treated separately since z depends only on its own innovations. Solving x equation for x and substituting into y equation, we would be able to verify that y does not depend on z. An alternative way to see this is to transform (2.17) into MA form by inverting the polynomial matrix. The conditions in proposition (4) will guarentee that the matrix position (1,3) of this inverse will always be zero and since u, v, and w can always be chosen to be orthogonal to each other, the MA representation removes the dependence of y on x.

Thus although y appears to depend on both x and z in (2.17) the dependence on z is spurious because this dependence disapeasars in the smaller information set consisting of y and z. However, checking the last condition in proposition (4) would be difficult in practice due to rounding errors and other difficulties: therefore, a roundabout way to test for type I spurious causality is to use smaller information sets in a larger system.

As in the bivariate single-equation case, there is no unique testing procedure for G-causality in the multivariate case. The likelihood ratio test can be conducted by constructing the restricted and unrestricted vector models as was done by Caines and Chan [1975] (see footnote 26). Tjøstheim [1981] has suggested a LaGrange multiplier test. However, before carrying out any of these tests, one has to specify the lag lengths of the variables involved. A procedure for lag length specification and testing for G-causality in a multivariate setting will be developed in the next chapter.

2.5 SUMMARY

According to Granger's definition a variable X is said to cause another variable Y if Y can be prediced better by including X in Y's predictive information set than without it. This is a highly operational concept compared to many other notions of causality. Therefore, the conflicting causal implications of competing economic theories can be brought into an encompassing test using Granager's definition.

The causality tests currently in use are essentially single-equation methods, which can handle only two variables at a time. Among the four tests discussed in this chapter--Sims test, modified Sims test, Granger test, and cross correlation test--the Granger test has been found, by others, to be promissing both in terms of performance in simulation exercises and its simplicity. Further, since it deals with a single equation of a VAR model, the extension of the procedure to VAR is straightforward.

The use of VAR for causality testing entails important features which a single-equation cannot handle; that is, the ability of VAR to take account of spurious causality. It is shown in Section 2.4 that four types of causality events--direct, indirect, Type I spurious, and Type II spurious causality-- can be examined mostly as zero constraints on the parameters of the model.

Chapter III

VECTOR AUTOREGRESSION AND TESTING FOR GRANGER-CAUSALITY

3.1 INTRODUCTION

Vector autoregression (VAR) models have already appeared in chapters 1 and 2, but for the sake of completeness of this chapter the model (1.5) is reintroduced below:

$$A(L)z(t) = u(t), \tag{3.1}$$

where z(t) is a (kx1) vector of random variables with zero means, u(t) is a (kx1) vector of random shocks, or one-step-ahead forecast errors, which are white noise with zero means and

$$\begin{aligned} E[u(t)u(s)'] &= \Sigma, \ t=s \\ &= 0, \ t\neq s. \end{aligned} \tag{3.2}$$

That is, the components of u(t) are univariate white noise processes, uncorrelated with each other at different time points, but possibly cross correlated at common time points (i.e. contemporaneously correlated). Therefore, Σ need not necessarily be diagonal. Further, A(L) is the (kxk) polynomial matrix of maximum order p with $A_0=I$; i.e. $A(L) = A_0L^0 + A_1L^1 + \ldots + A_pL^p$, where L is the lag operator. The process in (3.1) will be stationary if the roots of det[A(L)]=0 all have modulus greater than one [Hannan, 1970: 326]. A typical element of A(L) may be written as

$$\sum_{l=0}^{p_{ij}} a_{ijl} L^{l}.$$

Thus $p = \max_{ij} p_{ij}$, and the lag length matrix is $P^* = (p_{ij})$.

One disadvantage of using VAR models instead of VARMA (vector autoregressive moving average) models is that the former may not be as parsimonious[31] as the latter. VARMA models, on the other hand, pose two problems: lack of uniqueness (see any standard text), and difficulty of estimation [Tiao and Box, 1979]. Tiao and Box have demonstrated the inadequacy of the conditional likelihood approximation to the exact likelihood (see Section 3.3.2) when the MA (moving average) component is present. Therefore, VAR models possess compensating advantages compared to VARMA models.

Although in vector time series modelling every variable is taken to depend on every other variable, the model can easily be extended to include exogenous variables. However, when competing theories refer to the same set of variables with conflicting endogenous-exogenous categorizations or when the exogeneity implied by the theory is not firmly based, then one has to consider the complete set of variables as endogenous to avoid the possibility of inflicting

[31] A model is said to be parsimonious if it contains relatively few parameters. The meaning of this is well embedded in Milton Friedman's statement: "A hypothesis is important if it 'explains' much by little..." [See Harvey, 1981a: 5-6].

spurious restrictions. In this chapter all variables will be considered endogenous and the extension of the VAR model to include exogeneous variables is straightforward.

A crucial problem in multiple time series modelling is the determination of the lag orders of the variables in the model, for which economic theory is of little help. In the application of the causality tests discussed in the previous chapter, the common practice has been to specify the lag orders arbitrarily. We have already mentioned in Section 2.2.5 that overparameterization leads to loss of power. Thornton and Batten [1985] have conducted an extensive investigation on money-income causality by using different lag length specifications in the Granger test (See Section 2.2.3). They find that the arbitrary lag length specifications such as (4,4), (8,8), etc. on the dependent and the independent variables yield contradictory results. They also find that the longer lag length for the dependent variable to account for autocorrelation and the shorter lag length for the independent variable to conserve degrees of freedom does not work well either. However, the lag lengths specified by three "model selection criteria" have, in general, provided compatible results.

The objective of this chapter is to present a sequential procedure to estimate the lag length matrix P* and then to evaluate the performance of this procedure in sampling experiments. A discussion of existing procedures is given in

Section 3.2, and Section 3.3 presents the sequential model building strategy. The final specification of the model thus reached can be used for causal inference. A theoretical evaluation of the model selection criteria, which play an important role in our modelling strategy, is contained in Section 3.4. The sampling experiment and the results are presented in Section 3.5 and the last section provides a summary of the chapter.

3.2 EXISTING MODEL BUILDING PROCEDURES

3.2.1 Single Time Series Methods

A number of statistical procedures of model building are present both in standard regression and time series analyses.[32] In the time series literature the classical hypothesis testing procedure has been used in AR modelling for a long time [see Priestly, 1981a: 370]. The addition or deletion of a lag under this procedure depends on a (conventionally) pre-determined level of significance. In practice, however, the nominal and the actual significance levels may

[32] Three important reviews on model selection procedures are Hocking [1976] Amemiya [1980], and Judge et al. [1980]. Using the standard regression model with nonstochastic regressors these studies, when taken together, have discussed, among other things, (1) the classical hypothesis testing framework, (2) model selection criteria, and (3) biased estimation procedures, basically the Stein rule. In general, they were critical about the first as a model selection procedure. Amemiya favours the use of a selection criterion such as PC or AIC (see the text). Judge et al. support the Stein rule estimator. Hocking is of the opinion that due to the difficulties of a rigorous mathematical analysis of these techniques, the use of a particular method has to be established by practical considerations (i.e., through simulation exercises).

differ and the choice of an optimal level of significance remains unresolved.

Another approach to AR modelling is the residual variance plot. The residual variance is expected to decrease as more and more lags are added and then fluctuate when the true order is reached. This method requires a skillful guess of the order which minimizes the residual variance since the plot does not itself indicate a unique choice.

The Box-Jenkins approach to univariate ARMA modelling is well known but the procedure is highly "judgemental" [Zellner and Palm, 1974]. Model selection critera such as R^2 or Theil's $\bar{R}^2$ are hardly popular in the time series literature. Even within the regression framework their weaknesses are well known [see Hocking, 1976; Judge et al., 1980; Granger and Newbold 1974]. The model selection criteria to be discussed later are a response to the weaknesses of the traditional methods which, as viewed by Akaike [1974], are inappropriate to a situation such as time series modelling which involves multiple decision making.

3.2.2 Multiple Time Series Methods

Unlike the single series methods, multiple series modelling strategies have shown a retarded development due to their complexity. The procedures hitherto developed to specify VARMA models share three basic features found in univariate Box-Jenkins approach: 1. tentative specification, 2. estimation, and 3. diagnostic checking. Among these, the tentative specification is the most laborious phase. At least four types of specification strategies may be found in the literature.[33]

First type: These procedures use univariate residual cross correlations to identify the full model [Haugh and Box, 1977; Granger and Newbold, 1977]. The procedure, however, suffers from the weaknesses discussed in Section 2.3.4 and its extension to systems larger than bivariate for simultaneous accounting of the relationships is difficult.

Second type: Tiao and Box [1979] have extended the univarate Box-Jenkins procedure to the multivariate case. In this procedure the sample auto and cross correlation matrices are used to identify MA structures and partial cross correlation matrices derived from a stepwise autoregression are used to identify AR structures. (This latter strategy dates back to Quenouille, 1957.)

Both the first and the second type procedures use the hypothesis testing framework in the specification process. An acknowledged drawback of these procedures is that they rely heavily on the model builder's skill and judgement, and that judgemental requirement is rather decisive.

---

[33] These four groups do not include the statespace modelling procedure suggested by Akaike [1976].

<u>Third type</u>: These procedures use univariate model specifications to suggest a structure for the full model [Zellner and Palm, 1974; Wallis, 1977]. Since the univariate model building depends on Box-Jenkins methodology, skill and judgemental requirements play a very important role in these procedures.

<u>Fourth type</u>: The modelling strategy here is to specify single equations (any number of variables) to identify the full model [Hsiao, 1979; Abeysinghe, 1982]. These procedures differ from the previous ones in a distinct way: they are designed to specify VAR models only, not mixed processes. By this manner they try to avoid the difficulties of VARMA modelling mentioned in Section 3.1. Further, they differ from Sims [1980] type (somewhat arbitrary) determination of a common lag order for all the variables and the Tiao-Box type stepwise autoregression.

The problem with assuming a common lag order (p) for all the variables is the severe loss of degrees of freedom. In a VAR with a common lag order the number of parameters to be estimated grows with the square of the number of variables, thereby exhausting the degrees of freedom very quickly. Sims has discussed the problems associated with the loss of degrees of freedom. Instead of an arbitrary determination of p, it may be estimated using a model selection criterion such as AIC. However, Nickelburg's [1985] small sampling experiments on the use of model selection criteria to estimate p show mixed and unsatisfactory results.

Although stepwise autoregression is designed to estimate a lag length matrix, P*, this is approached through a series of VAR fits with common lag orders p=1,2,..,P, where P is the largest order considered. Hypothesis testing is used at each step to determine P*. In small samples this procedure also suffers from the problems of severe loss of degrees of freedom, thereby biasing the hypothesis testing procedure at each step. Further, since the significance of the coefficients is determined by t values, misleading P* is likely to appear, given the small sample size usually encountered, due to distorted t values resulting through the effect of multicollinearity among the lag variables.

The common feature of the procedures suggested by Hsiao [1979] and by Abeysinghe [1982], is that the variables are admitted into the equation sequentially (for details see next section). Hsiao ignores the covariance matrix Σ in the first stage and specifies each equation by combining Granger's definition of causality and Akaike's final prediction error (FPE) criterion. The effect of Σ is taken into account at the stage of estimating the full model. Diagnostic checking is carried out by overfitting and underfitting the full model using the hypothesis testing framework.[34]

---

[34] Hsiao's procedure has been applied by Hsiao [1979, 1981, 1982], Caines et al. [1981], McMillin and Fackler [1984], and Thornton and Batten [1985].

Abeysinghe [1982], on the other hand, specifies the VAR model with a diagonal Σ and a triangular A_0 matrix, by combining model selection criteria (Akaike's AIC, BIC and Schwarz's SBC) and the hypothesis testing framework. It is important to note, however, that Hsiao has presented his procedure in a more general framework, whereas Abeysinghe has confined his to a bivariate system.

Since both the model selection criteria and the hypothesis framework have their own weaknesses, incorporating both in the model specification makes it difficult to attribute the poor performance of the procedure to either of them. The model selection criteria, however, provide a non-arbitrary and less judgemental methodology for model building. Therefore, they will constitute a major ingredient of the specifiction procedure presented below.

3.3 SEQUENTIAL MODEL BUILDING STRATEGY FOR VAR MODELS

The procedure presented below stems from the fourth type of model building procedures discussed in the previous section, and tends towards Hsiao's methodology. The procedure is presented here in an open framework so that a specific format could be reached after evaluating its performance in sampling experiments.

This procedure shares three stages of modelling (tentative specification, estimation, and diagnostic checking) common to all the procedures discussed in Section 3.2.2.

Each stage of modelling will be presented in the three subsections to follow. The model selection criteria (MSC's; see Section 3.4) alone are used in the entire specification process and the choice of a particular MSC is left to be determined later on the basis of their performance in small samples in relation to the specification procedure used.[35]

3.3.1 Tentative Specification

The following zero mean bivariate process of (x,y) is used for illustrative purposes:

$$x(t) = a(L)x(t) + b(L)y(t) + u(t) \tag{3.3a}$$

$$y(t) = c(L)x(t) + d(L)y(t) + v(t), \tag{3.3b}$$

where a(L) etc. are polynomials in the lag operator L and u and v are zero mean white noise innovations with (2x2) covariance matrix Σ which is assumed to be nondiagonal.

Since it is possible to obtain consistent parameter estimates by applying OLS to each equation separately (see Section 3.3.2), i.e. by ignoring Σ, the tentative specification proceeds by specifying each equation of the system separately. The lag lengths are chosen by minimizing a chosen MSC by varying the lag length p between zero and P,

[35] Although Hsiao chooses FPE by linking it to Granger's definition of causality, we will defer this decision to a later stage. The use of an MSC as a test of G-causality differs from using the hypothesis testing framework, which is commonly applied for testing for G-causality, in that the significance levels determined by the MSC's are non-arbitrary. However, the choice of an MSC in the presence of many with similar attributes is difficult.

where P is the highest order considered arbitrarily.

The most effective way to specify (3.3a) is to consider all the possible lag combinations between zero and P for x and y and choose that combination which minimizes the chosen MSC. This amounts to calculating $(P+1)^2$ MSC's, and for a k variable system this number rises to $(P+1)^k$ MSC's; i.e. an exponetial increase of the computational burden.

The computational burden of a complete search can be reduced to a great extent by determining the lag orders of a(L) and b(L) in (3.3a) sequentially. That is, determine the order of a(L) first by regressing x on lag x by increasing the lag order (p) from 0 to P; given the chosen order for a(L), then determine the order of b(L) by varying p of y from 0 to P. Thus, the number of MSC's to be calculated is reduced to 2(P+1).[36]

An immediate question is what determines the sequential order of the variables entering the equation. It may be argued that x has to be regressed on its own past first. The justification for this follows from Granger's definition of causality. That is, the secondary variables are relevant insofar as they can supply additional information not already supplied by one's own past. However, the same line of argument cannot be extended to determine the sequential order of the secondary variables. This may be determined arbitrarily

[36] The exact number is actually 2P+1 since the zero lag for y is the already calculated lag length for x.

or by using prior information.[37] If these variables are orthogonal to each other, then an arbitrary decision will be perfectly acceptable. Since orthogonality is very unlikely the entire sequential process is subject to what may be called a "proxy effect."

The proxy effect is the possibility that the lags of the first entered variables act as proxies for the later entered variables and thereby weaken or nullify the influence of the later entered variables. This possibility may be illustrated by rewriting the process (3.3a and b) as a bivariate AR(1) model. Thus,

$$\begin{bmatrix} 1-aL & -bL \\ -cL & 1-dL \end{bmatrix} \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} u(t) \\ v(t) \end{bmatrix} \qquad (3.4)$$

Now using the relation $\mathrm{adj}A = \det A . A^{-1}$ and pre-multipling both sides of (3.4) by the adjoint of the polynomial matrix we get

---

[37] Caines et al. [1981] applying Hsiao's procedure to a trivariate system, added the variable with the lowest FPE from bivariate equations first. In other words, the sequential order is determined on the basis of what may be called "minimum of minima" or "minmin" principle. This obviously increases the calculation burden. For a k variable equation the number of FPE's required to be calculated at univariate, bivariate, trivariate,..., and k-variate stages are respectively P+1, (k-1)P, (k-2)P,...,1. Thus the total number of FPE's required is 2P[k(k-1)+1]+1 which is larger than the kP+1 requirement if the sequential order is arbitrarily determined. The estimation of submodels, however, would be necessary to test for spurious causal relations.

$$[(1-aL)(1-dL)-bcL^2]\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} 1-dL & bL \\ cL & 1-aL \end{bmatrix}\begin{bmatrix} u(t) \\ v(t) \end{bmatrix}. \qquad (3.5)$$

Each equation in (3.5) is a univariate ARMA(2,1) model with an identical autoregressive part [Zellner and Palm, 1974; Palm, 1977]. This also means that each variable can be expressed as an AR(∞) process. Thus, by autoregressing x into infinite past it is possible to capture the effect of all other related variables.

The proxy effect arising from the sequential procedure may be minimized or possibly eliminated by choosing an MSC with a larger penalty term (Section 3.4). It is expected that the sampling experiments will shed some light on this problem.

It is now possible to outline the steps of the tentative specification stage [see Hsiao, 1979].

1. First regress x and y in (3.3a and b), all the variables in the full model, on their own past by varying the lag orders from 0 to P and estimate the orders of a(L) and d(L) by locating the minima of the chosen MSC. If a minimum occurs at P or close to P the value of P has to be increased.
2. Now take x and keep its own lag order determined in step (1) fixed and add y to the equation. Determine the lag order of b(L) by minimizing the MSC over y-

lags from 0 to P. When there are more than two variables in the model continue the process by admitting variables sequentially until all the variables are exhausted.

3. Repeat step (2) for equation (3.3b) and determine the order of c(L). Continue this process for all the equations in larger models.

4. Combine the single equation specifications to identify the full model. The joint estimation then may be carried out as described in the following section.

3.3.2 Estimation

In the univariate (scalar) case the AR model of order p can be estimated by the LS method utilizing the last N-p obsevations, where N is the sample size. Given a serially independent innovation process the Mann-Wold theorem [Harvey, 1981a: 48; Priestly, 1981a: 346] shows that the LS estimates are consistent and asymptotically normally distributed.

If the innovation process is Gaussian the exact likelihood function can be written as the product of the conditional likelihood of the last N-p observations (given the first p observations) and the likelihood of first p observations [Harvey, 1981b: 12-17]. The LS estimation now becomes equivalent to the conditional ML estimation. It can be shown [Harvey, 1981b: 122] that the large sample difference between the exact ML and the regression (LS) results are very minimal. Monte Carlo studies have indicated that for an AR(1) process sample size as small as 20 observations produce only neglegible difference in estimates by the two methods. However, the difference tends to increase in small samples as the order of the AR process increases and when the roots of the AR polynomial get closer to unity [Harvey, 1981b: 135-136].

The estimation methods of univariate AR(p) models can easily be extended to VAR(p) models, where p is now a common lag order for all the variables.[38] The estimation procedure of model (3.1) will be illustrated by means of a VAR(1) model; the extension to a VAR(p) model is straightforward. Moreover, it is well known that an AR(p) (vector or scalar) process can always be represented as an AR(1) process [Priestly, 1981a: 797].

The first order VAR(1) model will be written as

$$z(t) = Az(t-1) + u(t), \tag{3.6}$$

in which it is assumed that u(t) has a multivariate normal distribution with zero mean and variance Σ.

---

[38] Although our objective is not to assume a common p, for ease of presentation of the estimation method we assume existence of a VAR(p) model instead of a VAR(P*) model. For different methods of estimation see Priestly [1981a: 346-52].

Given the vector $z(t-1)$, $z(t)$ is distributed as a multivariate normal with mean $Az(t-1)$ and the covariance matrix Σ. Thus, the conditional likelihood, given $z(1)$, is

$$L_c(A,\Sigma/z) \propto |\Sigma|^{-\frac{N-1}{2}} \exp\{-(1/2)\mathrm{tr}\Sigma^{-1}S(A)\}, \tag{3.7}$$

where

$$S(A) = [z(t)-Az(t-1)]'[z(t)-Az(t-1)].$$

The exact likelihood function can be derived by multiplying (3.7) by the unconditional density function of $z(1)$. From (3.6) we can see that $z(1)$ is multivariate normal with zero mean (this becomes obvious when (3.6) is written as $(I-AL)z(t) = u(t)$) and $\mathrm{cov}(z(1)) = V = AVA'+\Sigma$ [see Harvey, 1981b: 49]. Thus the distribution of the vector $z(1)$ is

$$f[z(1)] \propto |V|^{-\frac{1}{2}} \exp\{-(1/2)z(1)'V^{-1}z(1)\} \tag{3.8}$$

Multiplying (3.7) by (3.8) and taking the log gives the exact log likelihood function as

$$\begin{aligned} \log L = \text{const.} &-(1/2)\log|V| - \tfrac{N-1}{2}\log|\Sigma| - (1/2)\mathrm{tr}\Sigma^{-1}S(A) \\ &-(1/2)z'(1)V^{-1}z(1). \end{aligned} \tag{3.9}$$

The matrix V depends on A, and the derivatives of $\log|V|$ with respect to A are complicated functions. Box and Jenkins [1970: 277] have shown (for the univariate case) that for large N, $\log|V|$ can be ignored, hence to this order of approximation the exact ML estimates of A are given by minimizing the last two terms of (3.9). However, if we assume that z(1) is fixed then (3.9) reduces to (3.7) in log form and maximization of this conditional likelihood is equivalent to the minimization of

$$S = \sum_{t=2}^{N} [z(t)-Az(t-1)]'\Sigma^{-1}[z(t)-Az(t-1)]. \tag{3.10}$$

The minimization of S in (3.10) is equivalent to the multivariate LS criterion. Since the system in (3.10) resembles a seemingly unrelated regression with the same set of variables on the right hand side, the OLS estimation of each equation separately provides fully efficient and consistent estimates even if Σ is not diagonal. However, if the lag orders of the right hand side variables are different, which is similar to the case of constraints across equations, the OLS on each equation still provide consistent parameter estimates but they will be less efficient than the joint parameter estimates [Nelson, 1976]. Nelson has derived the asymptotic results for an AR(1) model and shown that the gain in efficiency is greater, the larger the contemporaneous correlation and the greater the difference between two AR parameters of the two equations.

Asymptotically efficient estimators for the joint system can be obtained in two steps. Since consistent estimates are possible from OLS applied to each equation separately, in the first stage OLS can be used to estimate Σ from the single equation residuals. In the second stage, the system parameters can be estimated jointly given the estimate of Σ, using the multivariate Gauss-Newton scheme. Σ is then reestimated from the joint system.

3.3.3 Diagnostic Checking

Diagnostic checking can be carried out by overfitting and underfitting. Instead of using the hypothesis testing framework, we propose to specify a number of P*'s, including the one identified from single equations, and to examine which P* minimizes the chosen MSC from the joint estimation. The MSC now has to be redefined to suit the joint model.

For example, if the lag length matrix identified from the single equation specifications is

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

then we can specify a number of P*'s in the neighborhood of the above matrix such as

$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

for diagnostic checking. An objective way to choose P*'s is given in Section 3.5.4 with a complete demonstration of the recommended diagnostic checking procedure.

3.3.4 Some Remarks on the Procedure

The above sequential procedure has a number of attractive features. It is designed to estimate the lag length matrix P*. This would be behaviourally more meaningful and statistically more efficient as compared with a common lag order. The estimation procedure is straightforward, simple and less costly; only OLS and seemingly unrelated regressions are necessary. Thus the extra cost required at the tentative specification stage may be offset by the less costly estimation procedure. Skill and judgemental requirements are minimal since the entire procedure relies on MSC's to perform a better job. The MSC's also remove the arbitrary significance level requirement of the classical hypothesis testing framework.

The sections to follow will evaluate this procedure in detail.

3.4 EVALUATION: MODEL SELECTION CRITERIA

The performance of the specification procedure described in the previous section depends crucially on the performance of the chosen MSC (model selection criteria). Unfortunately a large number of such criteria has emerged[39] and there is no sound ground for choosing one for practical use. An important characteristic of the MSC's is that they try to trade

---

[39] Hocking [1976], Amemiya [1980], Judge et al. [1980], and Priestly [1981a, 1981b] provide a good secondary source to many of these criteria.

off the bias associated with a parsimoneous parameterization against the inefficiency arising from overparameterization (see Appendix A). However, very little is known about the sampling properties of these criteria under different model structures, though some criteria are reported to be quite useful in practice [see Akaike, 1974]. Many of these criteria are designed for linear relationships and their final form is a function of the error sum of squares, thus all of them are related. In this section we demonstrate how these criteria are related to each other with reference to a univariate AR model. In our sequential specification, as the number of variables entering a single equation grows, the lag or the parameter count variable , p, has to be set equal to the exact number of regression coefficients estimated.[40]

Akaike [1974] has developed a general framework for model selection using the Kullback-Leibler information measure and defined the now well known AIC criterion (Akaike Information Criterion) as

$$\mathrm{AIC}(p) = -2\log(\text{maximum likelihood}) + 2p, \tag{3.11}$$

where p is the number of independently adjusted parameters. An estimate of the model size, p, is obtained by minimizing AIC(p) over p=0,1,...,P, where P is the largest model size

[40] Because of scripting difficulties of letter subscripts, and for notational clarity $\sigma^2(p)$, i.e. σ^2 with subscript p, will be denoted by σ_1^2 and $\sigma^2(p-1)$ by $\sigma^2{}_1$ throughout this section unless otherwise specified. Here σ_1^2 and $\sigma^2{}_1$ refer to the vaiances of p and p-1 parameter models, respectively.

considered. The second term in (3.11) is regarded as a penalty imposed on the maximum likelihood estimation of the model size (ML estimation of the model size would be the largest model considered). Thus (3.11) provides a solution to the bias-variance trade-off mentioned above. The AIC may be viewed as an extension of the criteria based on mean square prediction error (MSPE) suggested by Akaike [1969] and others.

Following Atkinson (1981) we can write (3.11) in a general form as

$$-\log(\text{maximum likelihood}) + \alpha p/2, \tag{3.12}$$

where α is either a constant (mostly 2, as in 3.11) or a function of the sample size. In the following presentation the values of α of each criterion will be derived explicitly in order to evaluate and compare the asymptotic properties of these criteria.

For a linear regression, including AR models, (3.12) becomes

$$(N/2)\log\sigma^2 + RSS(p)/2\sigma^2 + \alpha p/2 + \text{const.}, \tag{3.13}$$

where σ^2 is the error variance and RSS(p) is residual sum of squares of the p-parameter model. If σ^2 is taken as the common variance for all models or if it is assumed to be known, then (3.13) yields a general criterion for linear models [see also Geweke and Meese, 1981],

$$C(p,\alpha) = RSS(p) + \alpha p\sigma^2, \ p=0,1,\ldots,P, \tag{3.14}$$

which is to be minimized over p.

If σ^2 in (3.14) is replaced by its unbiased estimate, $\hat{\sigma}^2$, and α by 2 we get Mallow's [1973] criterion, MC,

$$MC = \tilde{\sigma}_1^2 + (2p/N)\hat{\sigma}^2, \tag{3.15}$$

where $\tilde{\sigma}_1^2 = RSS(p)/N$. A monotonic transformation of (3.15) yields Mallow's $C_p = [RSS(p)/\hat{\sigma}^2]+2p-N$ [Amemiya, 1980]. This criterion involves the practical difficulty of estimating σ^2.

Another criterion which could directly be derived from (3.14) is Akaike's [1969, 1970] FPE (Final Prediction Error) or Amemiya's [1980] PC (Prediction Criterion). Amemiya derived PC by taking the expectation of the MSPE for the subset model and this, in effect, is similar to replacing σ^2 in (3.14) by $\hat{\sigma}_1^2 = RSS(p)/(N-p)$ and α by 2. Thus we get

$$FPE(p) = PC(p) = (1 + p/N)\hat{\sigma}_1^2. \tag{3.16}$$

If $\hat{\sigma}_1^2$ is replaced by the ML estimate $\tilde{\sigma}_1^2 = \hat{\sigma}_1^2(N-p)/N$, then (3.16) becomes

$$FPE(p) = (1 + p/N)(1 - p/N)^{-1}\tilde{\sigma}_1^2, \tag{3.17}$$

and the binomial expansion yields

$$\begin{aligned} FPE(p) &= \tilde{\sigma}_1^2[1 + 2p/N] + O(N^{-2}) \\ &= \tilde{\sigma}_1^2[1 + 2p/N], \text{ for large } N, \end{aligned} \tag{3.18}$$

with $\alpha = 2$ being expressed explicitly.

When both dependent and independent variables form a multivariate normal distribution, the expectation of the MSPE for the subset model, after dropping the terms involving only N, yields another criterion named SP [see Hocking, 1976, Amemiya, 1980]:

$$\begin{aligned} SP(p) &= (N-p)^{-1}\hat{\sigma}_1^2 \\ &= \tilde{\sigma}_1^2 N^{-1}(1 - p/N)^{-2}, \end{aligned} \tag{3.19}$$

where $\hat{\sigma}_1^2$ and $\tilde{\sigma}_1^2$ are defined as in FPE above. The binomial expansion of (3.19) yields

$$\begin{aligned} SP(p) &= \tilde{\sigma}_1^2 N^{-1}(1 + 2p/N) + O(N^{-2}) \\ &= FPE(p)/N, \text{ for large N}, \end{aligned} \tag{3.20}$$

and obviously $\alpha = 2$ in SP(p), too.

If σ^2 in (3.13) is replaced by the ML estimate $\tilde{\sigma}_1^2$ and α by 2 we get

$$AIC(p) = \log\tilde{\sigma}_1^2 + 2p/N. \tag{3.21}$$

The Taylor series expansion of the logarithm of (3.17) yields

$$AIC(p) \doteq N\log[FPE(p)], \text{ for large N}. \tag{3.22}$$

Sawa [1976] has arrived at a new form of information criterion using the same Kullback-Leibler measure, but departing from Akaike concerning the idea of the true model. Sawa measures the distance between what he calls the pseudo-true parameter values and the postulated parameter values. His criterion, named as BIC, is defined as

$$BIC(p) = N\log\tilde{\sigma}_p^2+2(p+2)\{\hat{\sigma}^2/\tilde{\sigma}_p^2-2[\hat{\sigma}^2/\tilde{\sigma}_p^2]\}, \tag{3.23}$$

where $\hat{\sigma}^2$ is an independent estimate of true variance. If $\hat{\sigma}^2 \neq \tilde{\sigma}_p^2$ then

$$BIC(p) = N\log\tilde{\sigma}_p^2+2(p+1) = AIC(p+1), \tag{3.24}$$

with 1 added to p for the independent estimate of σ^2. Though $\alpha = 2$, BIC tends to favour more parsimoneous models than AIC as seen from the behaviour of the ratio $\hat{\sigma}^2/\tilde{\sigma}_p^2$, which increases and approaches unity as p is increased in large samples. Chow [1981] has also suggested a correction to AIC, but both Sawa's and Chow's corrections lead to practical problems of estimating σ^2.

Parzen [1974] has suggested a criterion which measures, as he describes, the integrated relative MSE of the spectral estimate for each p. Parzen has shown that this leads to the calculation of what he calls CAT (Criterion Autoregressive Transfer function) defined as

$$CAT(p) = \sum_{j=0}^{p}(N-j)N^{-2}\tilde{\sigma}_j^{-2} - (N-p)N^{-1}\tilde{\sigma}_p^{-2} \tag{3.25}$$

where $\tilde{\sigma}_p^2$ is defined as before. The value of α imbeded in this criterion can be found indirectly by taking CAT(p) -

CAT(p-1) and comparing it with FPE(p) - FPE(p-1). Using the large N definition of FPE given in (3.16) we get

$$\mathrm{FPE}(p)-\mathrm{FPE}(p-1) = (1+2p/N)[\tilde{\sigma}_p^2-\tilde{\sigma}_{p-1}^2]+(2/N)\tilde{\sigma}_{p-1}^2.$$

Around the true order of the model $\tilde{\sigma}_p^2 \mp \tilde{\sigma}_{p-1}^2$, hence the above is approximately equal to $(2/N)\tilde{\sigma}_p^2$. Similarly,

$$\begin{aligned}\mathrm{CAT}(p)-\mathrm{CAT}(p-1) &= \sum_{j=0}^{p}(N-j)N^{-2}\tilde{\sigma}_j^{-2}-(N-p)N^{-1}\tilde{\sigma}_p^{-2} \\ &\quad -[\sum_{j=0}^{p-1}(N-j)N^{-2}\tilde{\sigma}_j^{-2} - (N-p+1)N^{-1}\tilde{\sigma}_{p-1}^{-2}] \\ &= (N-p)N^{-2}\tilde{\sigma}_p^{-2}- (N-p)N^{-1}\tilde{\sigma}_p^{-2}+ (N-p+1)N^{-1}\tilde{\sigma}_{p-1}^{-2} \\ &\mp (2/N - p/N^2)\tilde{\sigma}_p^{-2},\end{aligned}$$

for $\tilde{\sigma}_p^2 \mp \tilde{\sigma}_{p-1}^2$ around the true order of the model. For large N we get $\mathrm{CAT}(p) - \mathrm{CAT}(p-1) \mp (2/N)\tilde{\sigma}_p^{-2}$, thus $\alpha = 2$ in CAT, too.

Schwarz [1978] developed a Bayesian criterion for model selection by evaluating the leading terms of the asymptotic expansion of the posterior density. Thus he proposed to choose the model which maximizes

$$\mathrm{SBC}(p) = \log(\text{maximum likelihood}) - 1/2(\log N)p \qquad (3.26)$$

over p. Here we use SBC to abbreviate Schwarz' Bayesian Criterion. For the linear model considered so far, (3.26) becomes

$$\mathrm{SBC}(p) = \log\tilde{\sigma}_p^2 + (p\log N)/N, \qquad (3.27)$$

which is to be minimized over $p=1,\ldots,P$. Unlike the selection criteria discussed above, α in SBC is a function of the sample size, i.e., $\alpha(N) = \log N$.

Chow [1981] has, however, shown that (3.26) is a poor approximation to the posterior probability for small N. Akaike [1978, 1979] has also developed a Bayesian modification to AIC, which he calls BIC. This criterion is very similar to SBC except that Akaike's BIC places a higher penalty on large models. Geweke and Meese [1981] have introduced a modification to SBC, but their criterion involves the difficulty of estimating σ^2.

The last criterion to be mentioned in this section is that of Hannan and Quinn [1979]. They have suggested a modification to AIC for reasons to be discussed in the next section. Their modified AIC is defined as

$$\mathrm{PHI}(p) = \phi(p) = \log\tilde{\sigma}_p^2 + (2pc\log\log N)/N, \qquad (3.28)$$

where c is an arbitraty constant to be chosen greater than one. The α in this criterion is equal to 2cloglogN.

3.4.1 Statistical Properties of the Selection Criteria

The model selection criteria considered in the previous section can be categorized into two groups: (1) those with constant α, i.e., $\alpha = 2$, and (2) those with α a function of the sample size, N. In the first group are MC (or C_p), FPE (=PC), SP, AIC, BIC, and CAT. The SBC and PHI are in the second group. This division helps to understand the asymptotic properties of the criteria in the two groups.

3.4.1.1 Asymptotic Properties

Akaike [1970] has shown that the FPE criterion does not provide a consistent estimate of the true order of an autoregression and suggested a modification to FPE in order to achieve consistency. Shibata [1976] has derived the asymptotic distribution of the order estimated by AIC for an AR(p_0) model, where p_0 is the true order, and found that this distribution depends only on p_0 and P, P being the largest model order considered, and not on the values of the autoregressive parameters. This investigation also shows that AIC overestimates the true order asymptotically with a non-zero probability, though the probability of over-fitting dies out fairly quickly.

Bhansali and Downham [1977] have further investigated the FPE criterion by expressing (3.18) in a general form by replacing 2 by α. They have shown that a consistent estimate of the model order can be obtained by arbitrarily increasing α. They have further pointed out that the choice of P is not crucial provided that $P>p_0$. Hannan and Quinn [1979] have shown that a consistent estimate of equation order can be obtained by defining an increasing function of N, $\alpha(N)$, such that $\alpha(N)/N$ tends to zero as N tends to infinity. They also showed that $\alpha(N)$ attains its lower bound when $\alpha(N)$ = loglogN. Thus SBC and PHI are consistent estimators of the equation order while others with $\alpha = 2$ are not. A further study of the consistency of differnt criteria has been provided by Geweke and Meese [1981]. Using exogenous regressors they reaffirmed the inconsistency of those criteria with constant α.

3.4.1.2 Small Sample Properties

The inconsistency of AIC and other criteria with fixed α need not be considered a defect of these criteria [Akaike, 1970; Hannan and Quinn, 1979]. On the one hand, consistency is an aysmptotic property whereas in practice we have to deal with finite small samples. The finite order AR models, on the other hand, may be an approximation to an infinite order AR or some other process. Therefore, these criteria choose only the best approximation to such a process in a finite sample.

Unfortunately an analytical study of small sampling properties of these criteria is virtually impossible. In consequence, we have to rely on simulation methods to understand their properties in small samples. A number of such studies already exist designed with reference to a particular problem, e.g., scalar autoregression or scalar ARMA models. Shibata[1976] found that though the asymptotic distribution of the order estimated by AIC does not depend on the size of the AR parameters, in small samples they affect the distribution. It has also been found that SBC tends to underestimate the true order in small samples as the parameter values become smaller. The objective of Hannan and Quinn's modification to AIC is to reduce the effect of underestimation found in SBC while retaining its consistency property.

As a practical guidance, Bhansali and Downham [1977] suggest to use different α's in (3.18) to find whether the chosen order shifts with incresing α. If the order does not shift with increasing α we have more confidence in the chosen model, whereas sharp changes in the order chosen is an indication to be more careful in the model choice. However,the choice of a particular α remains unanswered. Therefore, instead of changing α arbitrarily one could use a number of different criteria with different α's (or with different penalty terms) to see if they yield the same model order.

3.4.2 Model Selection Criteria and the Hypothesis Testing Framework

Using the selection criteria to choose a model is equivalent to using the hypothesis testing framework with test statistics such as F or the likelihood ratio with critical points (hence significance levels) determined objectively. For example, if AIC is used to choose between two standard regression models, one nested in the other, we choose model 1 with p parameters instead of model 2 with p+q parameters if $AIC_1 < AIC_2$. This decision rule is equivalent to choosing model 1 if [Sawa, 1978]

$$F < [\exp(2q/N)-1].(N-p-q)/q, \tag{3.29}$$

where $F = (N-p-q)(\tilde{\sigma}_1{}^2-\tilde{\sigma}_2{}^2)/q\tilde{\sigma}_2{}^2$, where $\hat{\sigma}^2$'s are the ML estimates of the error variances of the smaller and the larger models respectively. The right hand side of (3.29) is the critical point associated with minimum AIC rule. As compared to conventional 5% or 1% level of significance, the significance level associated with the minimum AIC rule varies depending on N, p, and q.

Sawa has tabulated the critical points and the levels of significance associated with minAIC and minBIC for some chosen N, p, and (q=1). His table shows that minAIC as well as minBIC usually allows significance levels higher than the conventional levels. For moderate values of p (e.g., p=3) the significance level for minAIC varies from about 30 percent to 16 percent as the degrees of freedom increase from 10 to 1000.[41] (The corresponding figures for minBIC are approximately 10 percent and 16 percent.) The link between the other selection criteria and the hypothesis testing framework can also be demonstrated, but such an attempt is beyond the scope of this chapter (see Sawa [1978], Amemiya [1980], and Judge et al. [1980] for more information).

---

[41] The general contention of the critiques of the classical hypothesis testing framework is that the significance level should decrease with increasing sample size (Mizon, 1984: 139).

3.5 EVALUATION: SMALL SAMPLE PERFORMANCE

In this section we carry out two sampling experiments: the first is designed to assess the performance of the model selection criteria in small samples at the tentative specification stage, and the second to evaluate the overall model building procedure.

3.5.1 Experiment 1

Experiment 1 is designed to pursue the following objectives.

1. The distributional pattern of the model selection criteria over different parameter structures.
2. The ability of the model selection criteria to pick the correct lag when the order of the autoregression is finite.
3. The sensitivity of the selection criteria to the parameter size.
4. The possibility of finding Granger-noncausality when in fact it is present.
5. The sensitivity of the lag distributions to the sequential order of the regressor variables: an evaluation of the proxy effect.
6. The sensitivity of the model selection criteria to the presence of contemporaneous correlation.
7. The performance of the model selection criteria under finite and infinite order autoregressions.

3.5.1.1 Experimental Design

Two types of experiments are designed below: (1) the true model is a finite order VAR, and (2) the true model order is infinite. In the latter case all the model selection criteria provide biased estimates of the lag order. Both types of experiments are confined to bivariate models.

For the finite order VAR we use the following first order model:

$$z(t) = \Phi z(t-1) + v(t) \tag{3.30a}$$

which can be expanded as

$$\begin{bmatrix} y(t) \\ x(t) \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} y(t-1) \\ x(t-1) \end{bmatrix} + \begin{bmatrix} v_1(t) \\ v_2(t) \end{bmatrix} \tag{3.30b}$$

where v(t) is a vector of normal random variables with zero mean and covariance matrix Σ_v such that

$$\Sigma_v = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

There is no contemporaneous correlation (see Section 3.1) if Σ_v is diagonal. However, since we are interested in allowing for the presence of contemporaneous correlation in our models we can write (3.30a and b) in a convenient form (for data generation) with a diagonal covariance matrix as

$$Az(t) = Bz(t-1) + u(t) \tag{3.31}$$

where A is chosen to be upper triangular (for uniqueness) and $u(t) = [u_1(t), u_2(t)]'$ is a vector of normal random variables with zero mean and diagonal covariance matrix Σ_u.

By premultiplying (3.31) by A^{-1} we get (3.30a) such that $A^{-1}B=\Phi$, $A^{-1}u(t)=v(t)$, and $A^{-1}\Sigma_u A^{-1\prime}=\Sigma_v$.

The parameters of the model (3.31), hence of (3.30), are specified as follows.

$$\begin{bmatrix} 1 & -a_{12} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y(t) \\ x(t) \end{bmatrix} = \begin{bmatrix} .5 & b_{12} \\ b_{21} & .5 \end{bmatrix} \begin{bmatrix} y(t-1) \\ x(t-1) \end{bmatrix} + \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix}, \tag{3.32}$$

and

$$\Sigma_u = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The valuses of a_{12}, b_{12}, and b_{21} are specified so as to obtain some desired values for ϕ_{12} and ϕ_{21} which enable us to evaluate the objectives specified at the begining. These parameter values are given in Table 3.1.

In Structure 1 X and Y are independent, while in Structures 2-5 there is a unidirectional relation from X to Y, i.e. one-way G-causality. Structures 6 and 7 represent feedback systems. The feedback is introduced to Structure 3, which has a slightly weak unidirectional relation, to see how feedback is likely to impair the performance of the mod-

TABLE 3.1

Parameter specification for VAR(1) model

STRUCTURE	a_{12}	b_{12}	b_{21}	ϕ_{11}	ϕ_{22}	ϕ_{12}	ϕ_{21}	σ_1^2	σ_2^2	σ_{12}
1	0	0	0	.5	.5	0	0	1.00	1	0
2	.1	.1	0	.5	.5	.15	0	1.01	1	.1
3	.6	.1	0	.5	.5	.40	0	1.36	1	.6
4	.1	.6	0	.5	.5	.65	0	1.01	1	.1
5	.6	.6	0	.5	.5	.90	0	1.36	1	.6
6	.6	.1	.1	.56	.5	.40	.1	1.36	1	.6
7	.6	.1	.3	.68	.5	.40	.3	1.36	1	.6

el selection criteria in weak relationships. The roots of all the determinantal lag polynomials are real and greater than unity in absolute value.

In order to obtain an infinite order VAR we use the following first order VMA (vector moving average) model.

$$\begin{bmatrix} y(t) \\ x(t) \end{bmatrix} = \begin{bmatrix} 1+\theta_{11}L & \theta_{12}L \\ \theta_{21}L & 1+\theta_{22}L \end{bmatrix} \begin{bmatrix} v_1(t) \\ v_2(t) \end{bmatrix}, \tag{3.33}$$

where $v(t)=[v_1(t),v_2(t)]'$ is a vector of normal random variables with zero mean and covariance matrix Σ .

The corresponding autoregressive model for (3.33), assuming invertibility, is then

$$\begin{bmatrix} 1+\theta_{11}L & \theta_{12}L \\ \theta_{21}L & 1+\theta_{22}L \end{bmatrix}^{-1} \begin{bmatrix} y(t) \\ x(t) \end{bmatrix} = \begin{bmatrix} v_1(t) \\ v_2(t) \end{bmatrix}$$

or

$$\begin{bmatrix} \phi_{11}(L) & \phi_{12}(L) \\ \phi_{21}(L) & \phi_{22}(L) \end{bmatrix} \begin{bmatrix} y(t) \\ x(t) \end{bmatrix} = \begin{bmatrix} v_1(t) \\ v_2(t) \end{bmatrix} \tag{3.34}$$

where

$$\begin{bmatrix} \phi_{11}(L) & \phi_{12}(L) \\ \phi_{21}(L) & \phi_{22}(L) \end{bmatrix} = \begin{bmatrix} 1+\theta_{11}L & \theta_{12}L \\ \theta_{21}L & 1+\theta_{22}L \end{bmatrix}^{-1}$$

$$= D^{-1} \begin{bmatrix} 1+\theta_{22}L & -\theta_{12}L \\ -\theta_{21}L & 1+\theta_{11}L \end{bmatrix}$$

and
$$D = (1+\theta_{11}L)(1+\theta_{22})-\theta_{12}\theta_{21}L^2.$$

Setting $\theta_{21}=0$ to obtain a unidirectional relation we get $\phi_{21}(L)=0$ and

$$\phi_{11}(L) = (1+\theta_{11}L)^{-1}=(1+\theta_{11}L+\theta_{11}^2L^2+\ldots) \tag{3.35}$$

$$\phi_{22}(L) = (1+\theta_{22}L)^{-1}=(1+\theta_{22}L+\theta_{22}^2L^2+\ldots) \tag{3.36}$$

$$\begin{aligned} \phi_{12}(L) &= -\theta_{12}L(1+\theta_{11}L)^{-1}(1+\theta_{22}L)^{-1} \\ &= -\theta_{12}L[\sum_{n=0}^{\infty} [\sum_{j=0}^{n} \theta_{11}^{n-j}\, \theta_{22}^{j}]L^n]. \end{aligned} \tag{3.37}$$

The analytical solution to the feedback system where $\theta_{21}\neq 0$ is more complicated than the above case.

The parameter specifications for the model (3.33) are given in Table 3.2. Here the coefficient matrix Θ is set equal to the Φ matrix of Table 3.1. For all the structures in this table the covariance matrix is specified as

$$\Sigma_V = \begin{bmatrix} 1 & .6 \\ .6 & 1 \end{bmatrix}. \tag{3.38}$$

TABLE 3.2

Parameter specification for VMA(1) model

	STRUCTURE				
PARAMETER	8	9	10	11	12
θ_{11}	.50	.50	.50	.50	.50
θ_{22}	.50	.50	.50	.50	.50
θ_{12}	0	.15	.40	.65	.90
θ_{21}	0	0	0	0	0

In order to generate $v(t)=(v_1(t),v_2(t))'$ with the covariance matrix given in (3.38) we use the Cholesky Decomposition of Σ_V, i.e., factorize Σ_V as $\Sigma_V=CC'$, where C is lower triangular, and then transform a Normal(0,I) vector $u(t)=(u_1(t),u_2(t))'$ as $v=Cu$ to obtain v with the above covariance matrix. The matrix C is equal to

$$\begin{bmatrix} 1 & 0 \\ .6 & .8 \end{bmatrix}$$

The experiments were conducted using sample sizes 50 and 100, each with 100 replications. The normal random deviates with zero mean and unit variance were generated by calling the SAS subroutine RANNOR. This subroutine uses the Box-Muller transformation to transform uniform random numbers generated by the subroutine RANUNI which uses a 'primemodulus multiplicative generator' to generate uniform random numbers between 0 and 1. First the initial conditions were set to zero and then $u_1(t)$ and $u_2(t)$ were generated for $t=1,2,\ldots,N+100$. The first 90-99 observations were then discarded to minimize the effect of the arbitrary initial conditions. Enough data were generated to keep N constant for all the lag variables. The random number generator was restarted, but with different seeds, for each structure, and was continuously run for the 100 replications. This introduces a quasi-independence to the experiments over the structures and the replications.

After generating all the variables for a given structure, we follow the sequential model fitting procedure presented in Section 3.4. Since we are interested in the performance of the model selection criteria in this process, it is sufficient to specify only the first equation in (3.30) which in an unspecified form can be written as

$$y(t) = \sum_{i=1}^{\infty}\alpha(i)y(t-i)+\sum_{i=1}^{\infty}\beta(i)x(t-i)+e(t). \tag{3.39}$$

First we form the lag distribution of y alone by varying the lag length between 0 and P, and then form the lag distribution of x over the same lag range given some prespecified lag lengths for y. Since the value of P is unknown, it is generally advocated to increase P as N increases [Akaike, 1971; Geweke and Meese, 1981]. Thus the values chosen are P=5 for N=50 and P=10 N=100.[42] The SAS procedure RSQUARE

[42] Omoto, Nakagawa, and Akaike [1972] have advocated to keep P below N/5k, where k is the number of variables in the

plus a DATA step was used to calculate the chosen model selection criteria and then the SAS procedures RANK and FREQ were used to form the frequency distributions utilizing the minimum criterion values. CP and BIC were calculated using the true variances.(A sample computer program used to generate data and calculate the model selection criteria is given in Appendix B.)

3.5.1.2 Results of the Sampling Experiment 1

All the tables related to this section are given at the end of the section. The distribution of lag lengths estimated by the chosen model selection criteria in samples of size 50 are shown in Table 3.3. In this table the analysis is confined to the Structures from 1 to 5 only. Table 3.3 as well as Table 3.4 provides some preliminary results which help reduce the amount of numerical work in the subsequent exercises. We make the following observations from Table 3.3

1. A sample of size 50 relative to the VAR(1) model appears to be large enough to confirm the asymptotic properties established in Section 3.3.1. That is, the distribution of CP, SP, CAT, FPE, and AIC are very similar and there is virtually no difference between FPE and AIC. However, the distribution of BIC remains somewhat different from those in the same group (i.e.

---

model (k=2 in our case). Geweke and Meese [1981] have set P equal to N/lnN. We, however, adhered to the former rule of thumb in our choice of P.

those with α=2, see Section 3.4.1).

2. The frequency of picking the correct lag order for y (i.e. lag y=1) remains quite high throughout all the structures, except Structure 5, and this is remarkable in the case of SBC. We have to note that ϕ_{11}=.5, the coefficient of y(t-1), is a reasonably high value, within the range zero to one, which should be picked up by any good model selection criterion. Nevertheless, there is a likelihood that the lag length estimate of y will be greater than the true lag due to the possibility that lag y's attempt to capture some of the effect of x's, since y enters the equation first (i.e. the proxy effect). Our results, however, show that the frequencies for lag y>1 are small and the pattern of these frequencies reflect the typical attenuating behaviour of the model selection criteria rather than any effect of the sequential procedure.
3. The distributions of lag x given lag y = i (i=1,2,3,4) reveal an interesting feature: the lag x distribution does not depend on the pre-specification (or the misspecification) of lag y, i.e. no proxy effect is present. However, there must be an upper limit to the (mis)specification of lag y without affecting the specification of lag x because a VAR model can be reduced to a univariate AR(∞) process (Section 3.2.2).

4. The frequency of picking the correct lag order for x given a lag y depends on the size of ϕ_{12}. The implication is that when ϕ_{12} is small, in Structure 2 for example, the frequency of concluding, at least tentatively, that X does not G-cause Y is rather high (see frequencies at lag x=0). This frequency is quite high in the case of SBC, which has a tendency to underfit the model order. Expectedly, this frequency goes to zero as ϕ_{12} increases.
5. Structures 2-4 were rerun by setting the contemporaneous covariance (σ_{12}) to zero to assess the impact of its presence or absence on the performance of model selection criteria. These results are reported in Table 3.4. Comparison of Table 3.3 and 3.4 shows that the lag distributions are very much the same in Structures 2 and 4 where the contemporaneous covariance is very small (σ_{12}=.1). However, the frequency of picking the correct lag for y improves in Structures 3 and 5 when the contemporaneous covariace of .6 is set to zero; this is especially remarkable in Structure 5. We have already mentoned in Secton 3.3.2 that the single equation estimates are less efficient when cross equation correlation is present and large. The above finding shows that it also affects the lag length specification in our sequential procedure. This appears to be more serious when the lagged correlation is large (Structure 5, ϕ_{12}=.9). This speci-

fication error is expected since that the variance estimate entering the model selection criteria at a given lag length from a single equation is larger than that from the joint estimation. The implication of this finding is that the diagnostic checking stage has to be devised carefully so as to uncover possible misspecifications that could occur at the tentative specification stage.

Our next step is to assess the consistency of the above findings over all the different structures given in Tables 3.1 and 3.2 as the sample size as well as the maximum lag order (P) increases. However, because of the distributional similarity of CP, SP, CAT, FPE, and AIC in the sample size 50 and because of their asymptotic convergence, we decided to retain only one of them, namely AIC, as a representative of the group; AIC is the most well known among them. We also decided to drop BIC since it requires knowledge of the true variance or an estimate of it. Thus we proceed with AIC, SBC, and PHI in the rest of the analysis. Further, the distribution of lag x|lag y is now confined only to three cases, i.e. lag y=1,4,8.

The simulation results of the autoregressive structures given in Table 3.1 are reported in Table 3.5. We can observe from this table that the previous findings are still maintained. In particular we can make the following remarks.

1. The frequency of picking the correct lag for y does not show an improvement at sample size 100. This is obviously because of the dependence of the lag distributions on P which increases from 5 to 10 as N increases from 50 to 100.
2. The distribution of lag x given a lag y does not change even if lag y is misspecified to be 8, i.e. no proxy effect is apparent yet.
3. The frequency of picking the correct lag order for x given a lag y improves as the sample size increases even if ϕ_{12} is small. At the same time the frequency of erroneously concluding that G-causality (lag x=0) is absent goes down with the increase in N. The lag x distribution given lag y=1 corresponding to Structure 2 at sample size 200 (not reported here) showed a substantial reduction in the frequencies at lag x=0; the percentages of reduction for AIC, SBC, and PHI are respectively 56, 73, and 63. The same exercise for Structure 3 reduced the lag x=0 frequencies to zero.
4. The introduction of feedback to Structure 3, i.e. Structures 6 and 7, has not changed the previous results. The frequencies at lag y=1 have, however, increased because of the increase in ϕ_{11} in Structures 6 and 7 (see Table 3.1).

The simulation results of the moving average structures (i.e. the infinite order autoregressive structures) given in Table 3.2 are reported in Table 3.6. Here we are more interested in analysing the possibility of finding G-noncausality when, in fact, G-causality is present since the other aspects, such as lag length specification, do depend on N and P as well as on the parameter zise. Any finite parameterization here is only an approximation to the true autoregression. It is, however, useful to report all the observations that we could make from Table 3.6.

1. The lag y distributions across the structures 8 through 12 are similar, though not strictly compatible as in previous tables. AIC in general tends to pick a longer lag at both N=50 and N=100; the lags with highest frequencies range from 1 to 3. SBC and PHI are more likely to pick lag y=1 for N=50 and lag y=2 for N=100. This parsimony, given the sample size, is quite consistent with the theoretical structure for y, which declines geometrically from 0.5, the highest cofficient, at lag 1 to 0.0625 at lag 4.
2. The distribution of lag x given a lag y shows a changing pattern, except in Structure 8, as lag y and θ_{12} change. This pattern can be observed easily by reading through the columns under N=100. For lower values of θ_{12} (e.g., .15 and .4) the lag x distribution tends to move towards the lower end of the lag

scale as lag y is increased from 1 to 4 and 4 to 8. This pattern reverses when θ_{12} is high (e.g. .65 and .9). That is, the lag x distribution tends to move toward the upper end of the lag scale with increasing lag y. The behaviour of the lag x distribution when θ_{12} is small appears to be a result of the proxy effect. However, it is difficult to understand, without an analytical solution, why the pattern reverses when θ_{12} rises. Moreover, this distributional pattern changes when σ_{12} is set to zero (see Table 3.7). It is not clear, therefore, whether it is due to the proxy effect or contemporaneous correlation or both.

3. In Structure 8 where X does not cause Y, the frequency of picking the correct lag for x (lag x=0) given a lag y is quite high. And this distribution does not depend on the lag y specification. On the other hand, the frequency of erroneously concluding that there is no G-causality (i.e. picking lag x=0) is quite high for small values of θ_{12} (e.g. θ_{12}=.15), given θ_{11} and θ_{22}, though, this frequency goes down with increase in sample size. For large values of θ_{12} (e.g. θ_{12}=.65) the possibility of such a conclusion is virtually zero at both sample sizes for all criteria.

4. Structures 8 through 12 were reanalysed by setting σ_{12} to zero (Table 3.7) and it was found that the distributional patterns did not always remain the same, indicating the influence of contemporaneous correlation on the results of Table 3.6.

3.5.1.3 Conclusions from Experment 1

In contrast to Nickelburg's [1985] study in which he finds unsatisfactory performance of a number of model selection criteria in estimating a common lag order for VAR, we find they perform quite satisfactorily in small samples if the model order is finite and if the sequential procedure given in Section 3.3 is used for order specification. Since there is no correct finite lag order for infinite order models, due care must be given to the possibility of finding Granger-noncausality when, in fact, such causality is present.

As it is unknown whether the model order is finite or infinite it is always advisable to estimate a lag order by a mix of model selection criteria such as AIC, SBC, and PHI since they are designed to balance the bias-variance trade-off in small samples. The use of these three criteria is particularly important because they have different emphases on the bias-variance trade-off.

If the three criteria disagree with each other, especially regarding the presence or absence of G-causality, this is an indication that the parameter values are small. However, the probability of reaching an incorrect conclusion can be substantially reduced by increasing the sample size even if the parameter values are small. In this regard AIC is less likely to reach an incorrect conclusion of 'no G-causality.' Thus, Hsiao's choice of FPE (which is equivalent to AIC in large samples) for detecting G-causality can be maintained.

Nevertheless, when the parsimony of the model order becomes more important, i.e. where variables with low causal strength could be ignored, SBC or PHI could be used.

Since the presence of contemporaneous correlation is often the case, due to data aggregation and omitted variables, specification error is likely even if the parameter values are high. Therefore, diagnostic checking has to be carried out carefully using the joint models as demonstrated in the next sampling experiment.[43]

---

[43] Experiment 2 follows Tables 3.3-3.7.

TABLE 3.3

Distribution of lag length estimates in 100 replications
(VAR(1) Structures 1-5, N=50)

STRUCTURE 1: X, Y independent, $\phi_{12}=0$, $\sigma_{12}=0$

Lag y	CP	SP	CAT	FPE	AIC	BIC	SBC	PHI
0	2	2	5	2	2	2	2	2
1	76	76	71	74	74	80	95	84
2	10	10	10	10	10	9	3	8
3	3	4	4	4	3	1	0	3
4	8	7	8	8	7	7	0	3
5	1	1	2	2	4	1	0	0
Lag x	given lag y = 1							
0	74	74	73	71	71	82	95	85
1	9	9	9	9	9	7	3	6
2	4	2	2	2	2	4	1	2
3	2	4	5	6	6	1	0	1
4	7	6	6	6	6	5	1	5
5	4	5	5	6	6	1	0	1
Lag x	given lag y = 2							
0	73	72	72	68	68	81	97	83
1	8	10	10	10	10	8	2	8
2	2	2	1	2	2	4	1	2
3	4	4	4	6	6	2	0	3
4	9	8	8	7	7	4	0	3
5	4	4	5	7	7	1	0	1
Lag x	given lag y = 3							
0	76	74	73	70	69	87	95	84
1	7	9	10	10	10	7	5	6
2	2	1	1	1	1	2	0	2
3	5	4	3	5	5	1	0	3
4	7	7	7	8	9	3	0	4
5	3	5	6	6	6	0	0	1
Lag x	given lag y = 4							
0	75	74	72	70	69	86	94	81
1	9	10	11	10	11	7	5	6
2	4	2	2	3	3	3	1	2
3	4	4	3	3	3	1	0	4
4	6	6	6	8	8	3	0	5
5	2	4	6	6	6	0	0	2

Table 3.3 contd.

STRUCTURE 2: X causes Y, ϕ_{12}=.15, σ_{12}=.1

Lag y	CP	SP	CAT	FPE	AIC	BIC	SBC	PHI
0	1	1	2	1	1	1	1	1
1	81	80	78	77	77	83	95	87
2	9	12	12	12	11	7	2	7
3	3	3	3	4	4	3	1	2
4	6	4	4	4	5	6	1	3
5	0	0	1	2	2	0	0	0
Lag x	given lag y = 1							
0	48	49	47	47	47	54	73	61
1	31	29	27	27	27	29	24	26
2	7	5	5	5	5	7	3	3
3	5	7	8	8	8	3	0	4
4	6	6	8	8	8	6	0	4
5	3	4	5	5	5	1	0	2
Lag x	given lag y = 2							
0	48	50	46	46	46	52	74	57
1	29	27	26	26	25	30	22	29
2	5	4	4	4	4	8	2	2
3	7	6	8	8	8	5	2	5
4	9	10	10	10	11	4	0	5
5	2	3	6	6	6	1	0	2
Lag x	given lag y = 3							
0	49	50	47	44	44	54	77	58
1	31	28	27	27	27	31	19	28
2	4	5	7	7	7	7	2	2
3	6	5	5	7	7	4	2	5
4	7	8	9	9	9	3	0	5
5	3	4	5	6	6	1	0	2
Lag x	given lag y = 4							
0	47	47	45	41	41	57	72	54
1	30	29	29	29	26	32	23	30
2	4	6	5	5	6	6	4	6
3	8	6	6	7	7	3	1	3
4	7	7	8	9	9	2	0	5
5	4	5	7	9	11	0	0	2

Table 3.3 contd.

STRUCTURE 3: X causes Y, ϕ_{12}=.4, σ_{12}=.6

Lag y	CP	SP	CAT	FPE	AIC	BIC	SBC	PHI
0	0	0	0	0	0	0	0	0
1	66	67	66	65	65	68	97	82
2	14	17	17	16	16	16	3	11
3	9	9	9	9	9	9	0	1
4	5	5	6	6	6	5	0	4
5	4	2	2	4	4	2	0	2
Lag x	given lag y = 1							
0	19	18	18	16	16	19	36	21
1	53	47	46	46	46	56	56	59
2	15	16	16	18	18	14	5	14
3	7	8	9	9	9	6	1	3
4	4	4	4	4	4	4	2	2
5	2	7	7	7	1	1	0	1
Lag x	given lag y = 2							
0	15	17	16	16	16	20	37	22
1	53	47	46	44	44	59	52	54
2	16	17	17	17	17	13	9	17
3	7	8	9	9	9	4	1	4
4	7	6	7	8	8	3	1	2
5	2	5	5	6	6	1	0	1
Lag x	given lag y = 3							
0	16	18	17	16	16	21	36	20
1	53	48	47	46	46	57	53	55
2	15	15	17	17	17	12	9	14
3	10	10	10	10	10	8	1	7
4	4	4	4	4	4	2	1	2
5	2	5	5	7	7	0	0	2
Lag x	given lag y = 4							
0	17	15	14	14	14	21	34	22
1	53	51	50	48	47	60	53	50
2	14	16	16	14	13	13	11	16
3	6	8	8	9	9	5	1	4
4	6	6	6	7	7	1	1	6
5	4	4	6	8	10	0	0	2

Table 3.3 contd.

STRUCTURE 4: X causes Y, ϕ_{12}=.65, σ_{12}=.1

Lag y	CP	SP	CAT	FPE	AIC	BIC	SBC	PHI
0	0	0	0	0	0	0	0	0
1	59	74	71	71	71	61	93	84
2	14	12	12	12	12	14	6	7
3	11	6	7	7	7	12	1	5
4	6	5	6	5	5	6	0	3
5	10	3	4	5	5	7	0	1
Lag x	given lag y = 1							
0	1	1	1	1	1	1	1	1
1	72	70	69	68	68	76	90	79
2	12	11	11	11	11	13	6	9
3	5	7	7	8	8	4	1	4
4	5	6	7	7	7	4	2	4
5	5	5	5	5	5	2	0	3
Lag x	given lag y = 2							
0	1	1	1	1	1	1	1	1
1	71	70	67	64	63	78	86	79
2	9	9	10	11	11	10	7	6
3	7	9	9	9	9	5	4	7
4	8	6	7	9	10	4	2	4
5	4	5	6	6	6	2	0	3
Lag x	given lag y = 3							
0	1	1	1	1	1	1	1	1
1	69	74	69	66	65	83	87	82
2	11	8	7	8	8	10	8	5
3	7	7	9	10	10	5	4	8
4	8	5	7	8	8	0	0	2
5	4	5	7	7	8	1	0	2
Lag x	given lag y = 4							
0	1	1	1	1	1	1	1	1
1	69	73	70	64	61	85	87	81
2	8	6	6	8	9	9	9	7
3	8	9	8	7	8	4	3	7
4	5	5	5	6	6	1	0	3
5	9	6	10	14	15	0	0	1

Table 3.3 concluded

STRUCTURE 5: X causes Y, $\phi_{12}=.9$, $\sigma_{12}=.6$

Lag y	CP	SP	CAT	FPE	AIC	BIC	SBC	PHI
0	0	0	0	0	0	0	0	0
1	29	42	40	39	39	33	71	49
2	36	37	38	38	38	37	26	34
3	11	9	9	9	9	9	1	8
4	15	9	9	10	10	14	2	8
5	9	3	4	4	4	7	0	1
Lag x	given lag y = 1							
0	0	0	0	0	0	0	0	0
1	71	68	65	63	63	76	92	83
2	13	15	16	15	15	13	6	11
3	9	8	9	9	9	8	1	4
4	5	5	5	6	6	2	1	1
5	2	4	5	7	7	1	0	1
Lag x	given lag y = 2							
0	0	0	0	0	0	0	1	0
1	70	61	58	58	58	77	91	77
2	11	18	17	17	17	13	5	12
3	10	10	11	11	11	6	2	7
4	6	7	8	8	8	3	1	3
5	3	4	6	6	6	1	0	1
Lag x	given lag y = 3							
0	0	0	0	0	0	0	2	0
1	67	65	64	61	60	78	89	71
2	17	18	18	18	18	14	7	17
3	8	9	9	10	11	4	1	6
4	5	5	5	5	5	4	0	4
5	3	3	4	6	6	0	1	2
Lag x	given lag y = 4							
0	0	0	0	0	0	0	0	0
1	70	65	61	60	60	80	87	73
2	18	21	19	19	19	15	10	19
3	5	5	7	8	8	2	1	2
4	5	6	6	6	6	3	0	4
5	2	3	7	7	7	0	0	2

TABLE 3.4

Distribution of lag length estimates in 100 replications
(VAR(1) Structures 1-5, N=50, σ_{12}=0)

STRUCTURE 2: X causes Y, ϕ_{12}=.15

Lag y	CP	SP	CAT	FPE	AIC	BIC	SBC	PHI
0	2	2	3	2	2	2	2	2
1	80	81	78	78	78	81	94	86
2	8	9	10	11	11	8	2	6
3	4	2	2	2	2	3	1	2
4	6	5	5	5	5	6	1	4
5	0	1	2	2	2	0	0	0
Lag x	given lag y = 1							
0	50	50	45	43	43	52	74	59
1	29	29	29	29	29	29	23	27
2	6	6	6	6	6	7	2	3
3	5	5	7	8	8	5	1	4
4	7	7	8	8	8	6	0	5
5	3	3	5	6	6	1	0	2
Lag x	given lag y = 2							
0	47	48	44	42	42	54	74	57
1	30	30	29	29	29	29	21	28
2	5	3	3	4	4	7	3	2
3	6	7	8	8	8	4	2	5
4	9	8	10	10	9	5	0	6
5	3	4	6	7	8	1	0	2
Lag x	given lag y = 3							
0	48	47	45	43	43	57	76	57
1	33	30	30	29	29	30	19	28
2	4	5	5	5	5	6	3	2
3	6	7	6	7	7	4	2	5
4	7	8	8	9	9	3	0	6
5	2	3	6	7	7	0	0	2
Lag x	given lag y = 4							
0	47	46	44	41	40	58	73	54
1	31	32	29	28	29	29	21	30
2	5	5	6	5	4	7	4	3
3	6	6	6	7	7	3	2	4
4	7	7	7	9	10	3	0	6
5	4	4	8	10	10	0	0	3

TABLE 3.4 contd.

STRUCTURE 3: X causes Y, $\phi_{12}=.4$

Lag y	CP	SP	CAT	FPE	AIC	BIC	SBC	PHI
0	0	0	0	0	0	0	0	0
1	72	79	76	75	75	76	96	86
2	6	6	6	7	7	6	2	5
3	11	8	8	8	8	11	1	7
4	7	5	6	6	6	5	1	2
5	4	2	4	4	4	2	0	0
Lag x	given lag y = 1							
0	4	4	4	4	4	4	9	7
1	69	67	65	63	63	75	80	75
2	10	11	11	12	12	10	6	6
3	6	6	7	8	8	5	2	5
4	6	6	7	7	7	5	3	5
5	5	6	6	6	6	1	0	2
Lag x	given lag y = 2							
0	4	4	4	4	4	4	9	6
1	68	66	63	63	63	76	78	74
2	8	9	8	8	8	9	7	6
3	7	9	9	9	9	4	4	5
4	8	7	9	9	9	5	2	5
5	5	5	7	7	7	2	0	4
Lag x	given lag y = 3							
0	4	4	4	3	3	4	9	6
1	66	66	65	63	61	78	79	75
2	10	9	8	8	8	11	8	8
3	7	8	8	9	9	5	4	5
4	8	8	9	10	11	1	0	4
5	5	5	6	7	8	1	0	2
Lag x	given lag y = 4							
0	4	4	3	3	3	5	10	7
1	65	67	62	60	60	80	78	75
2	9	7	7	6	6	11	9	7
3	6	7	7	8	8	4	3	4
4	8	7	10	11	11	0	0	3
5	8	8	11	12	12	0	0	4

TABLE 3.4 contd.

STRUCTURE 4: X causes Y, ϕ_{12}=.65

Lag y	CP	SP	CAT	FPE	AIC	BIC	SBC	PHI
0	0	0	0	0	0	0	0	0
1	57	76	74	73	73	63	94	86
2	13	10	11	10	10	14	4	6
3	11	6	7	9	9	12	2	5
4	8	5	5	5	5	6	0	2
5	11	3	3	3	3	5	0	1
Lag x	given lag y = 1							
0	1	1	1	1	1	1	1	1
1	70	72	68	67	67	75	88	78
2	12	11	10	10	10	13	6	9
3	6	6	8	8	8	5	2	5
4	6	5	7	8	8	4	3	4
5	5	5	6	6	6	2	0	3
Lag x	given lag y = 2							
0	1	1	1	1	1	1	1	1
1	69	72	65	64	62	79	86	79
2	11	10	10	10	10	10	7	6
3	7	7	8	9	10	4	4	6
4	8	5	9	9	9	4	2	5
5	4	5	7	7	8	2	0	3
Lag x	given lag y = 3							
0	1	1	1	1	1	1	1	1
1	69	70	69	64	64	80	87	81
2	10	11	10	9	8	13	8	7
3	9	8	8	10	11	5	4	7
4	6	5	6	8	8	0	0	1
5	5	5	6	8	8	1	0	3
Lag x	given lag y = 4							
0	1	1	1	1	1	1	1	1
1	72	72	68	68	66	82	87	82
2	7	7	7	6	7	12	8	5
3	9	9	8	9	9	4	4	7
4	3	5	5	5	5	1	0	2
5	8	6	11	11	12	0	0	3

TABLE 3.4 concluded

STRUCTURE 5: X causes Y, ϕ_{12}=.9

Lag y	CP	SP	CAT	FPE	AIC	BIC	SBC	PHI
0	0	0	0	0	0	0	0	0
1	40	71	71	71	71	42	92	82
2	15	14	14	14	14	16	5	10
3	19	7	7	6	6	18	3	5
4	11	5	5	6	6	9	0	3
5	15	3	3	3	3	15	0	0
Lag x	given lag y = 1							
0	0	0	0	0	0	0	0	0
1	69	69	68	68	67	78	90	80
2	14	13	12	12	12	11	5	8
3	7	8	8	8	9	5	2	5
4	5	5	6	6	6	4	3	4
5	5	5	6	6	6	2	0	3
Lag x	given lag y = 2							
0	0	0	0	0	0	0	0	0
1	70	71	64	63	63	78	86	80
2	11	11	13	12	12	11	6	8
3	7	8	10	9	9	6	6	6
4	9	6	7	9	9	3	2	4
5	3	4	6	7	7	2	0	2
Lag x	given lag y = 3							
0	0	0	0	0	0	0	0	0
1	72	73	70	69	67	80	88	81
2	10	10	9	10	10	11	9	9
3	8	9	11	11	10	6	3	6
4	7	4	5	5	7	2	0	1
5	3	4	5	5	6	1	0	3
Lag x	given lag y = 4							
0	0	0	0	0	0	0	0	0
1	74	75	72	70	68	81	87	81
2	9	7	7	8	8	12	8	8
3	9	10	11	10	10	5	5	7
4	5	5	5	5	6	2	0	2
5	3	3	5	7	8	0	0	2

TABLE 3.5

Distribution of lag length estimates in 100 replications
(VAR(1) Structures 1-7, N=50 100)

STRUCTURE 1: X, Y independent, $\phi_{12}=0$, $\sigma_{12}=0$

	AIC		SBC		PHI	
Lag y	N=50	N=100	N=50	N=100	N=50	N=100
0	2	0	2	0	2	0
1	74	71	95	94	84	88
2	10	11	3	4	8	9
3	3	7	0	1	3	1
4	7	5	0	0	3	1
5-10	4	6	0	1	0	1
Lag x	given lag y = 1					
0	71	72	95	97	85	89
1	9	10	3	3	6	8
2	2	6	1	0	2	0
3	6	1	0	0	1	1
4	6	3	1	0	5	1
5-10	6	8	0	0	1	1
Lag x	given lag y = 4					
0	69	71	94	95	81	87
1	11	9	5	4	6	7
2	3	6	1	0	2	2
3	3	0	0	1	4	1
4	8	5	0	0	5	1
5-10	6	9	0	0	2	2
Lag x	given lag y = 8					
0	-	66	-	93	-	87
1	-	10	-	4	-	6
2	-	8	-	1	-	2
3	-	0	-	2	-	1
4	-	4	-	0	-	1
5-10	-	12	-	0	-	3

Maximum lag order considered for N=50 is 5.

TABLE 3.5 contd.

STRUCTURE 2: X causes y, $\phi_{12}=.15$, $\sigma_{12}=.1$

	AIC		SBC		PHI	
Lag y	N=50	N=100	N=50	N=100	N=50	N=100
0	1	0	1	0	1	0
1	77	70	95	95	87	85
2	11	12	2	2	7	9
3	4	7	1	1	2	2
4	5	5	1	1	3	3
5-10	2	6	0	1	0	1
Lag x	given lag y = 1					
0	47	30	73	64	61	47
1	27	43	24	35	26	42
2	5	8	3	0	3	6
3	8	2	0	1	4	1
4	8	5	0	0	4	1
5-10	5	12	0	0	2	3
Lag x	given lag y = 4					
0	41	35	72	67	54	49
1	26	38	23	32	30	40
2	6	8	4	0	6	7
3	7	2	1	1	3	1
4	9	4	0	0	5	1
5-10	11	13	0	0	2	2
Lag x	given lag y = 8					
0	-	31	-	63	-	49
1	-	39	-	36	-	39
2	-	9	-	0	-	7
3	-	3	-	1	-	3
4	-	5	-	0	-	0
5-10	-	13	-	0	-	2

TABLE 3.5 contd.

STRUCTURE 3: X causes Y, ϕ_{12}=.4, σ_{12}=.6

	AIC		SBC		PHI	
Lag y	N=50	N=100	N=50	N=100	N=50	N=100
0	0	0	0	0	0	0
1	65	69	97	96	82	86
2	16	12	3	3	11	9
3	9	7	0	1	1	3
4	6	4	0	0	4	1
5-10	4	8	0	0	2	1
Lag x	given lag y = 1					
0	16	5	36	13	21	6
1	46	64	56	85	59	83
2	18	15	5	1	14	8
3	9	4	1	1	3	1
4	4	3	2	0	2	0
5-10	7	9	0	0	1	2
Lag x	given lag y = 4					
0	14	4	34	13	22	6
1	47	65	53	81	50	79
2	13	10	11	5	16	10
3	9	7	1	0	4	2
4	7	3	1	0	6	1
5-10	10	11	0	1	2	2
Lag x	given lag y = 8					
0	-	4	-	12	-	6
1	-	61	-	78	-	78
2	-	10	-	8	-	9
3	-	8	-	1	-	3
4	-	3	-	0	-	2
5-10	-	14	-	1	-	2

TABLE 3.5 contd.

STRUCTURE 4: X causes Y, ϕ_{12}=.65, σ_{12}=.1

	AIC		SBC		PHI	
Lag y	N=50	N=100	N=50	N=100	N=50	N=100
0	0	0	0	0	0	0
1	71	66	93	95	84	86
2	12	14	6	3	7	7
3	7	11	1	1	5	4
4	5	4	0	1	3	2
5-10	5	5	0	0	1	1
Lag x	given lag y = 1					
0	1	0	1	0	1	0
1	68	73	90	96	79	86
2	11	9	6	2	9	9
3	8	2	1	2	4	1
4	7	5	2	0	4	2
5-10	5	11	0	0	3	2
Lag x	given lag y = 4					
0	1	0	1	0	1	0
1	61	67	87	96	81	88
2	9	10	9	3	7	8
3	8	3	3	1	7	1
4	6	6	0	0	3	2
5-10	15	14	0	0	1	1
Lag x	given lag y = 8					
0	-	0	-	0	-	0
1	-	61	-	95	-	86
2	-	11	-	3	-	8
3	-	4	-	2	-	1
4	-	7	-	0	-	3
5-10	-	17	-	0	-	2

TABLE 3.5 contd.

STRUCTURE 5: X causes Y, $\phi_{12}=.9, \sigma_{12}=.6$

	AIC		SBC		PHI	
Lag y	N=50	N=100	N=50	N=100	N=50	N=100
0	0	0	0	0	0	0
1	39	16	71	53	49	37
2	38	57	26	42	34	51
3	9	15	1	4	8	8
4	10	5	2	1	8	3
5-10	4	7	0	0	1	1
Lag x	given lag y = 1					
0	0	0	0	0	0	0
1	63	68	92	97	83	84
2	15	15	6	2	11	11
3	9	3	1	1	4	2
4	6	3	1	0	1	1
5-10	7	11	0	0	1	2
Lag x	given lag y = 4					
0	0	0	0	0	0	0
1	60	63	87	94	73	84
2	19	14	10	6	19	8
3	8	6	1	0	2	4
4	6	3	0	0	4	1
5-10	7	14	0	0	2	1
Lag x	given lag y = 8					
0	-	0	-	0	-	0
1	-	60	-	94	-	86
2	-	15	-	5	-	7
3	-	4	-	0	-	3
4	-	6	-	0	-	2
5-10	-	15	-	1	-	2

TABLE 3.5 contd.

STRUCTURE 6: Feedback, $\phi_{12}=.4$, $\phi_{21}=.1$, $\sigma_{12}=.6$

	AIC		SBC		PHI	
Lag y	N=50	N=100	N=50	N=100	N=50	N=100
0	0	0	0	0	0	0
1	76	74	98	99	88	92
2	8	7	2	1	5	5
3	6	7	0	0	2	1
4	6	4	0	0	4	1
5-10	4	8	0	0	1	1
Lag x	given lag y = 1					
0	15	5	37	13	24	5
1	49	67	56	84	56	85
2	19	13	4	3	13	7
3	6	2	1	0	2	1
4	4	2	2	0	4	0
5-10	7	11	0	0	1	2
Lag x	given lag y = 4					
0	14	4	36	13	23	6
1	45	62	50	80	53	79
2	15	11	12	6	15	9
3	9	7	1	0	3	3
4	7	3	1	0	3	1
5-10	10	13	0	1	3	2
Lag x	given lag y = 8					
0	-	4	-	12	-	6
1	-	60	-	78	-	78
2	-	10	-	8	-	9
3	-	7	-	1	-	3
4	-	3	-	0	-	2
5-10	-	16	-	1	-	2

TABLE 3.5 concluded

STRUCTURE 7: Feedback, ϕ_{12}=.4, ϕ_{21}=.3, σ_{12}=.6

	AIC		SBC		PHI	
Lag y	N=50	N=100	N=50	N=100	N=50	N=100
0	0	0	0	0	0	0
1	79	69	98	99	92	90
2	9	11	1	0	5	6
3	4	5	1	1	2	1
4	5	5	0	0	0	3
5-10	3	10	0	0	1	0
Lag x	given lag y = 1					
0	14	5	39	9	25	5
1	48	68	54	88	56	84
2	24	13	5	3	11	9
3	5	3	1	0	2	0
4	2	4	1	0	3	0
5-10	7	7	0	0	3	1
Lag x	given lag y = 4					
0	12	4	37	11	22	6
1	50	66	51	82	51	77
2	14	7	8	6	16	10
3	8	7	2	1	4	2
4	9	3	1	0	3	2
5-10	7	13	1	0	4	3
Lag x	given lag y = 8					
0	-	3	-	11	-	6
1	-	66	-	81	-	79
2	-	8	-	7	-	8
3	-	7	-	0	-	3
4	-	3	-	0	-	2
5-10	-	13	-	1	-	2

TABLE 3.6

Distribution of lag length estimates in 100 replications
(VMA(1) Structures 8-12, N=50 100, σ_{12}=.6)

STRUCTURE 8: X does not cause Y, θ_{12}=0

	AIC		SBC		PHI	
Lag y	N=50	N=100	N=50	N=100	N=50	N=100
0	2	0	3	0	2	0
1	37	15	63	60	51	35
2	35	38	28	32	33	42
3	13	17	5	8	8	16
4	8	10	1	0	3	3
5-10	5	20	1	0	3	4
Lag x	given lag y = 1					
0	69	61	92	86	81	73
1	6	4	4	6	6	8
2	11	19	3	6	7	10
3	9	6	0	1	4	5
4	1	2	1	0	2	2
5-10	4	8	0	1	0	2
Lag x	given lag y = 4					
0	71	70	94	92	88	83
1	6	12	3	6	5	7
2	5	4	3	1	4	5
3	4	3	0	0	1	2
4	8	3	0	0	2	0
5-10	6	8	0	1	0	3
Lag x	given lag y = 8					
0	-	67	-	92	-	83
1	-	10	-	6	-	7
2	-	5	-	1	-	5
3	-	6	-	0	-	2
4	-	2	-	0	-	0
5-10	-	10	-	1	-	3

Maximum lag order considered for N=50 is 5.

TABLE 3.6 contd.

STRUCTURE 9: X causes Y, θ_{12}=.15

	AIC		SBC		PHI	
Lag y	N=50	N=100	N=50	N=100	N=50	N=100
0	1	0	1	0	1	0
1	28	6	56	33	40	17
2	35	35	32	44	36	45
3	16	25	9	17	15	23
4	13	14	1	5	5	9
5-10	7	18	1	1	3	6
Lag x	given lag y = 1					
0	38	13	79	59	60	35
1	11	7	5	8	8	9
2	22	40	12	24	17	34
3	15	9	3	7	10	13
4	7	11	1	1	2	6
5-10	7	20	0	1	3	3
Lag x	given lag y = 4					
0	57	42	88	82	77	66
1	14	17	5	8	10	14
2	5	13	6	5	8	10
3	9	7	0	2	0	4
4	9	3	1	2	3	2
5-10	6	18	0	1	2	4
Lag x	given lag y = 8					
0	-	45	-	80	-	65
1	-	19	-	10	-	13
2	-	10	-	5	-	10
3	-	5	-	3	-	6
4	-	4	-	1	-	2
5-10	-	17	-	1	-	4

TABLE 3.6 contd.

STRUCTURE 10: X causes Y, $\theta_{12}=.4$

	AIC		SBC		PHI	
Lag y	N=50	N=100	N=50	N=100	N=50	N=100
0	1	0	1	0	1	0
1	23	4	50	26	35	12
2	30	21	32	45	33	34
3	24	28	14	24	21	29
4	15	19	2	5	6	13
5-10	7	27	1	0	4	12
Lag x	given lag y = 1					
0	2	0	15	0	8	0
1	8	0	20	7	12	1
2	34	25	36	56	34	39
3	30	31	22	26	28	35
4	15	16	5	8	12	15
5-10	11	28	2	3	6	10
Lag x	given lag y = 4					
0	11	2	36	11	20	4
1	14	3	21	27	21	15
2	19	15	18	31	19	26
3	26	24	16	18	21	26
4	13	22	8	5	8	14
5-10	17	34	1	8	11	15
Lag x	given lag y = 8					
0	-	2	-	13	-	4
1	-	5	-	24	-	17
2	-	13	-	35	-	25
3	-	20	-	18	-	25
4	-	21	-	6	-	14
5-10	-	39	-	4	-	15

TABLE 3.6 contd.

STRUCTURE 11: X causes Y, θ_{12}=.65

	AIC		SBC		PHI	
Lag y	N=50	N=100	N=50	N=100	N=50	N=100
0	2	0	2	0	2	0
1	28	9	58	35	42	21
2	30	26	26	45	28	40
3	23	29	12	15	21	26
4	5	15	1	4	4	9
5-10	12	19	1	1	3	4
Lag x	given lag y = 1					
0	0	0	0	0	0	0
1	2	0	9	0	4	0
2	20	10	44	31	32	16
3	38	35	33	43	40	42
4	20	22	11	19	17	24
5-10	20	33	3	7	7	18
Lag x	given lag y = 4					
0	0	0	2	0	0	0
1	1	0	14	2	5	0
2	17	3	34	24	24	10
3	36	24	36	45	43	38
4	19	28	10	19	14	29
5-10	27	45	4	10	14	23
Lag x	given lag y = 8					
0	-	0	-	0	-	0
1	-	0	-	1	-	0
2	-	3	-	27	-	10
3	-	22	-	42	-	32
4	-	27	-	21	-	32
5-10	-	48	-	9	-	26

TABLE 3.6 concluded

STRUCTURE 12: X causes Y, θ_{12}=.9

	AIC		SBC		PHI	
Lag y	N=50	N=100	N=50	N=100	N=50	N=100
0	3	0	4	0	3	0
1	42	17	74	55	56	36
2	20	32	19	36	21	41
3	22	23	3	8	15	17
4	6	10	0	0	4	4
5-10	7	17	0	1	1	2
Lag x	given lag y = 1					
0	0	0	0	0	0	0
1	0	0	4	0	2	0
2	7	1	29	12	14	8
3	41	24	46	49	46	43
4	26	30	17	27	23	27
5-10	26	45	4	12	15	22
Lag x	given lag y = 4					
0	0	0	0	0	0	0
1	0	0	3	0	0	0
2	9	1	0	10	16	4
3	38	17	48	48	48	35
4	28	30	13	28	21	35
5-10	25	52	6	14	15	26
Lag x	given lag y = 8					
0	-	0	-	0	-	0
1	-	0	-	0	-	0
2	-	0	-	12	-	4
3	-	17	-	45	-	29
4	-	26	-	30	-	37
5-10	-	59	-	13	-	30

TABLE 3.7

Distribution of lag length estimates in 100 replications
(VMA(1) Structures 8-12, N=50 100, σ_{12}=0)

STRUCTURE 8: X does not cause Y, θ_{12}=0

	AIC		SBC		PHI	
Lag y	N=50	N=100	N=50	N=100	N=50	N=100
0	1	0	1	0	1	0
1	27	14	62	47	45	26
2	46	42	31	46	41	51
3	12	20	4	5	7	16
4	12	10	2	2	5	4
5-10	2	14	0	0	1	3
Lag x	given lag y = 1					
0	70	71	94	93	85	87
1	11	9	3	6	5	8
2	4	5	2	0	4	1
3	4	2	0	1	1	1
4	7	4	1	0	4	2
5-10	4	9	0	0	1	1
Lag x	given lag y = 4					
0	69	67	94	94	83	86
1	11	10	2	3	5	8
2	6	7	3	2	6	3
3	3	1	1	1	1	1
4	6	4	0	0	3	1
5-10	5	11	0	0	2	1
Lag x	given lag y = 8					
0	-	70	-	91	-	86
1	-	7	-	6	-	8
2	-	9	-	2	-	4
3	-	1	-	1	-	1
4	-	3	-	0	-	0
5-10	-	10	-	0	-	1

Maximum lag order considered for N=50 is 5.

TABLE 3.7 contd.

STRUCTURE 9: X causes Y, θ_{12}=.15

	AIC		SBC		PHI	
Lag y	N=50	N=100	N=50	N=100	N=50	N=100
0	2	0	2	0	2	0
1	35	15	65	54	47	33
2	33	46	27	40	36	45
3	15	18	5	4	6	15
4	11	7	1	2	7	4
5-10	4	13	0	0	2	3
Lag x	given lag y = 1					
0	53	37	85	86	70	59
1	21	20	12	8	16	17
2	9	20	2	5	5	17
3	7	8	0	1	5	3
4	6	5	1	0	3	2
5-10	4	10	0	0	1	2
Lag x	given lag y = 4					
0	52	39	87	86	70	61
1	19	12	9	9	16	16
2	10	13	2	4	6	13
3	6	10	1	1	4	4
4	4	9	0	0	1	1
5-10	9	17	1	0	3	5
Lag x	given lag y = 8					
0	-	35	-	85	-	59
1	-	14	-	10	-	19
2	-	13	-	4	-	11
3	-	12	-	1	-	5
4	-	8	-	0	-	1
5-10	-	18	-	0	-	5

TABLE 3.7 contd.

STRUCTURE 10: X causes Y, $\theta_{12}=.4$

	AIC		SBC		PHI	
Lag y	N=50	N=100	N=50	N=100	N=50	N=100
0	5	0	6	0	6	0
1	41	31	74	71	56	47
2	28	34	17	25	24	39
3	12	15	1	4	9	8
4	10	6	2	0	4	3
5-10	4	9	0	0	1	3
Lag x	given lag y = 1					
0	9	0	36	7	18	2
1	19	5	23	24	22	14
2	30	33	32	50	35	45
3	22	24	7	16	16	25
4	11	20	1	2	6	10
5-10	9	18	1	1	3	4
Lag x	given lag y = 4					
0	7	0	36	11	15	2
1	10	2	24	18	20	9
2	17	16	24	36	27	29
3	24	22	13	25	18	24
4	21	21	1	6	10	21
5-10	21	39	2	4	10	15
Lag x	given lag y = 8					
0	-	1	-	10	-	3
1	-	2	-	14	-	7
2	-	18	-	38	-	30
3	-	23	-	27	-	26
4	-	21	-	7	-	16
5-10	-	35	-	4	-	18

TABLE 3.7 contd.

STRUCTURE 11: X causes Y, θ_{12}=.65

	AIC		SBC		PHI	
Lag y	N=50	N=100	N=50	N=100	N=50	N=100
0	9	1	12	2	12	2
1	53	44	74	82	68	72
2	19	29	13	13	16	21
3	8	9	1	3	2	4
4	6	3	0	0	2	0
5-10	5	12	0	0	0	1
Lag x	given lag y = 1					
0	0	0	4	0	0	0
1	3	0	21	5	12	1
2	32	18	41	43	42	29
3	35	29	29	36	31	39
4	19	25	3	13	10	22
5-10	11	28	2	3	5	9
Lag x	given lag y = 4					
0	0	0	4	0	0	0
1	2	0	13	1	6	1
2	17	3	28	20	25	8
3	21	24	37	43	30	36
4	28	23	12	23	21	24
5-10	32	50	6	13	18	31
Lag x	given lag y = 8					
0	-	0	-	0	-	0
1	-	1	-	2	-	1
2	-	1	-	25	-	8
3	-	19	-	42	-	37
4	-	27	-	17	-	21
5-10	-	52	-	14	-	33

TABLE 3.7 concluded

STRUCTURE 12: X causes Y, θ_{12}=.9

	AIC		SBC		PHI	
Lag y	N=50	N=100	N=50	N=100	N=50	N=100
0	19	10	24	11	22	11
1	52	54	69	79	64	72
2	15	18	7	8	11	13
3	3	7	0	2	2	3
4	4	0	0	0	0	0
5-10	7	11	0	0	1	1
Lag x	given lag y = 1					
0	0	0	0	0	0	0
1	1	0	7	0	2	0
2	16	5	40	22	26	10
3	38	28	37	48	41	44
4	31	32	11	24	22	32
5-10	14	35	5	6	9	14
Lag x	given lag y = 4					
0	0	0	0	0	0	0
1	0	0	4	0	0	0
2	7	1	27	8	16	2
3	26	11	39	47	30	29
4	33	25	18	26	30	31
5-10	34	63	12	19	24	38
Lag x	given lag y = 8					
0	-	0	-	0	-	0
1	-	0	-	0	-	0
2	-	1	-	7	-	1
3	-	10	-	48	-	29
4	-	25	-	26	-	30
5-10	-	64	-	19	-	40

3.5.2 Experiment 2

This experiment is designed to gain further insight into the modelling strategy presented in Section 3.3. It also provides a comprehensive demonstration of all the stages, including diagnostic checking, of the model building procedure. The experiment is conducted using single realizations of sample sizes 50, 100, and 200; thus replications are avoided. For the purpose of comparison two alternative estimates of the lag length matrix will also be obtained: one by using AIC to estimate the maximum lag order of the whole system and the other by using the multivariate Box-Jenkins procedure advocated by Tiao and Box [1981]. A sample computer program usable for system estimation is provided in Appenix B, Section B.3.

3.5.2.1 Model Descriptions

The model we examine is a trivariate AR(2) one with the following structure imposed on it.

$$\begin{bmatrix} x(t) \\ y(t) \\ z(t) \end{bmatrix} = \begin{bmatrix} .5L+.15L^2 & -.5L & -.2L \\ 0 & .6L+.15L^2 & .1L+.5L^2 \\ 0 & 0 & .7L+.15L^2 \end{bmatrix} \begin{bmatrix} x(t) \\ y(t) \\ z(t) \end{bmatrix} + \begin{bmatrix} v_1(t) \\ v_2(t) \\ v_3(t) \end{bmatrix}, \tag{3.40a}$$

where v is assumed to be a normal random vector with zero mean and covariance

$$\Sigma = \begin{bmatrix} 1 & .5 & .2 \\ .5 & 1 & .4 \\ .2 & .4 & 1 \end{bmatrix}. \tag{3.40b}$$

This model is designed to capture a number of features commonly encountered in real applications. These features may be illustrated by relating the variables (x,y,z) to the subject matter of the next chapter, i.e. determinants of fertility. Thus we may consider x, y, and z as:

x: variable (or vector of variables) of interest (e.g. a fertility measure)

y: intermediate or proximate variable(s) (e.g. contraception)

z: causal variable(s) (socioeconomic variables, e.g. years of education)

In the model we have postulated that $(dx/dy) < 0$, $(dy/dz) > 0$, and $(dx/dz) < 0$. There is indirect relation between x and z working through y while the direct influence of z on x is weak. Further, we have introduced a response delay of y on x: the influence becomes strong only at the second lag. The influence of own past is assumed to be positive but rapidly dampened. The covariance matrix Σ reflects the contemporaneous correlation structure of the model. All the roots of the determinantal polynomial of (3.40a), i.e. roots of $|A(L)|=0$, were chosen to be greater than one in absolute value. The smallest root is 1.15 and the largest, in absolute value, is 5.81.

As in the previous experiment a vector $u=(u_1\ u_2\ u_3)'$ with zero mean and covariance matrix I was generated by calling SAS subroutine RANNOR and then the vector v was obtained by setting v=Cu, where C is the lower triangular matrix obtained by Cholesky decomposition of Σ. The matrix C is approximately equal to

$$\begin{bmatrix} 1 & 0 & 0 \\ .5 & .8660 & 0 \\ .2 & .3464 & .9165 \end{bmatrix}.$$

The data were generated for t=1,2,...,N+100, where N is the sample size, and then first 99 observations were discarded, mimicking the previous experiment. However, unlike the previous experiment we did not generate enough data to keep N constant as the lag length is increased from 0 to P. Instead, after obtaining a sample of size N, the initial conditions for t=-P+1,...,-1,0 were set to zero as in any other time series application. The value chosen for P is 10 for all sample sizes.

3.5.2.2 Tentative Specification

The tentative specification of the lag length matrix, P*, is carried out using AIC, SBC, and PHI. The sequential order of the variables to enter an equation at the bivariate stage is determined on the basis of the "minimum of minima" or "min-min" principle (see note under Table 3.8). We find that this choice leads to better results compared to an arbitrary choice.

TABLE 3.8

AIC values for x-equation, N=50

	AIC values			
	Univariate	Bivariate		Trivariate
Lag:(p)	x(p)	y(p)|x(1)	z(p)|x(1)	z(p)|x(1),y(1)
0	138.85	56.95	56.95	37.78
1	56.95*	37.78*	52.67*	37.48*
2	57.31	39.72	53.65	38.99
3	60.22	39.47	55.34	40.66
4	60.77	40.64	57.33	41.55
5	62.56	42.20	55.35	43.24
6	64.32	43.90	57.06	44.07
7	66.28	45.68	59.05	45.51
8	68.26	47.49	60.72	47.48
9	68.82	49.46	62.19	49.45
10	70.78	51.28	64.19	51.53

* minimum AIC values.
Note: The minAIC in the third column is less than the minAIC in the fourth column. Therefore, y enters the equation first at the bivariate stage.

For brevity of presentation we do not report all the criterion values calculated in the specification process. However, for illustrative purpose we have reported in Table 3.8 the AIC values resulting in the specification of the x-equation when the sample size is 50. The lag orders chosen by the three criteria are reported in Table 3.9. Although these criteria lead to unique choices of the lag orders it is possible to utilize a second choice for the purpose of diagnostic checking. The second choice would be the lag related to the second minimum if it occurs relatively close to the first minimum. These choices are indicated by double numbers such as 1|2 in Table 3.9.

We can easily compose the lag length matrices, P*'s, identified by each criterion by reading down the columns of AIC, SBC, and PHI under each equation in Table 3.9. Further, by combining the first and the second choices of lags we can form a set of P*'s for diagnostic checking. By this manner it is possible to avoid or minimize the use of arbitrary choices of P*'s for diagnostic checking. Since SBC has a strong tendency to underfit in small samples it would not be necessary to choose a P* with the lag orders below those of SBC for diagnostic checking. The P*'s corresponding to each criterion read from Table 3.9 are given below. The P*'s chosen from this set for diagnostic checking are given in Table 3.10.

<u>Sample size 50</u>:

$$P^*(AIC) = \begin{array}{c} \begin{array}{ccc} x & y & z \end{array} \\ \begin{bmatrix} 1|2 & 1 & 1 \\ 0 & 1|2 & 4|2 \\ 0 & 0 & 1|2 \end{bmatrix} \end{array} \quad P^*(SBC) = \begin{array}{c} \begin{array}{ccc} x & y & z \end{array} \\ \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \end{array} \quad P^*(PHI) = \begin{array}{c} \begin{array}{ccc} x & y & z \end{array} \\ \begin{bmatrix} 1 & 0 & 0|1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \end{array}$$

TABLE 3.9

Lag orders chosen by AIC, SBC, and PHI at the tentative specification stage

Sample size 50												
Specification	x-equation				y-equation				z-equation			
stage		AIC	SBC	PHI		AIC	SBC	PHI		AIC	SBC	PHI
Univariate	x	1|2	1	1	y	1|2	1	1	z	1|2	1	1
Bivariate 1	y*	1	1	1	x	1	0	1	x	0	0	0
2	z	1	1	1	z*	4|2	2	2	y	0	0	0
Trivariate	z	1	0	0|1	x	0	0	0	-	-	-	-
Sample size 100												
Univariate	x	2	1|2	2	y	2	1|2	2	z	2	1	1|2
Bivariate 1	y*	4|3	1	3|1	x	5	1	1	x	0	0	0
2	z	1	1	1	z*	2	2	2	y	0	0	0
Trivariate	z	1	1	1	x	0	0	0	-	-	-	-
Sample size 200												
Univariate	x	2	1	1|2	y	9|2	1|2	2	z	2	1	1|2
Bivariate 1	y*	3	1	3	x	5	2	2	x	0	0	0
2	z	3	3	3	z*	3	2	3|2	y	0	0	0
Trivariate	z	1	1	1	x	0	0	0	-	-	-	-

Notes:

1. Each successive stage of specification is conditional on the lag choice of the previous stage(s).
2. x, y and z inside boxes are the specification variables at the corresponding stage.
3. * indicates the variable which received sequential priority on the basis of minmin principle at the bivariate stage.
4. The double numbers such as 1|2 show the first choice| second choice options. The second choice is based on the second minimum value if it is relatively close to the minimum value.

<u>Sample size 100</u>:

$$P^*(AIC)=\begin{array}{c} \begin{array}{ccc} x & y & z \end{array} \\ \begin{bmatrix} 2 & 4|3 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 2 \end{bmatrix} \end{array} \quad P^*(SBC)=\begin{array}{c} \begin{array}{ccc} x & y & z \end{array} \\ \begin{bmatrix} 1|2 & 1 & 1 \\ 0 & 1|2 & 2 \\ 0 & 0 & 1 \end{bmatrix} \end{array} \quad P^*(PHI)=\begin{array}{c} \begin{array}{ccc} x & y & z \end{array} \\ \begin{bmatrix} 2 & 3|1 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 1|2 \end{bmatrix} \end{array}$$

<u>Sample size 200</u>:

$$P^*(AIC)=\begin{array}{c} \begin{array}{ccc} x & y & z \end{array} \\ \begin{bmatrix} 2 & 3 & 1 \\ 0 & 9|2 & 3 \\ 0 & 0 & 2 \end{bmatrix} \end{array} \quad P^*(SBC)=\begin{array}{c} \begin{array}{ccc} x & y & z \end{array} \\ \begin{bmatrix} 1 & 3 & 1 \\ 0 & 1|2 & 2 \\ 0 & 0 & 1 \end{bmatrix} \end{array} \quad P^*(PHI)=\begin{array}{c} \begin{array}{ccc} x & y & z \end{array} \\ \begin{bmatrix} 1|2 & 3 & 1 \\ 0 & 2 & 3|2 \\ 0 & 0 & 1|2 \end{bmatrix} \end{array}$$

We can make the following observations from the above exercise. When the sample size is 50 all three criteria tend to ignore the lags with small coefficients, though a second choice may be possible. As the sample size increases to 100 and 200 we observe mixed results: AIC does not underfit and it may lead to profligate parameterization, SBC still underfits, and PHI shows mixed possibilities. Thus in practice it may be difficult to arrive at the true P* with a single criterion if the parameter values are small. However, since we have three P*'s corresponding to each criterion we may be able to arrive at the true P* by considering the consistency of the first and second choices made by these criteria. This choice becomes further possible since we know the overfitting and underfitting tendencies of each criterion.

It is also important to note that an arbitrary determination of the sequential order of the variables may lead to unsatisfactory results. We can observe from Table 3.9 that if x were added to y-equation before z then there is a high possibility of choosing an incorrect nonzero lag for x (see bivariate results). However, when the sequential order is determined on the basis of the "minmin" principle z enters the equation before x and the lag order chosen for x then always becomes zero.

3.5.2.3 Diagnostic Checking and Final Specification

Having obtained a set of P*'s for diagnostic checking we can now proceed to system estimation. The system estimation is carried out using the "seemingly unrelated regression" (SUR) procedure (see Section 3.3.2). The covariance matrix Σ is replaced by the ML estimate $(\hat{\Sigma})$ to obtain SUR estimates of AR parameters. Then SUR residuals are used to reestimate Σ. The MSC's now have to be redefined to suit the joint estimation. For the joint system the maximum log likelihood may be approximated by $-(N/2)\log|\hat{\Sigma}|$. Accordingly the three MSC's may be redefined as

$$AIC(P^*) = N\log|\hat{\Sigma}| + 2p^* \tag{3.41}$$

$$SBC(P^*) = N\log|\hat{\Sigma}| + (\log N)p^* \tag{3.42}$$

$$PHI(P^*) = N\log|\hat{\Sigma}| + 2(\log\log N)p^*, \tag{3.43}$$

where each criterion is defined for a given P*, and p* is the number of AR parameters corresponding to P* (i.e. the sum of all the elements of P*). These criterion values calculated for chosen P*'s are given in Table 3.10. The minimum criterion values are marked with asterisks.

Table 3.10 reveals quite intersting and promissing results. When the sample size is 50 none of the criteria choses the true lag length matrix: AIC choses P*2 while SBC and PHI choose P*3. We can easily observe that P*3 is more parsimonious than P*2; thus SBC and PHI have ignored all the small coefficient lags (see the true model). AIC, on the other hand, has made a better choice by retaining the small coefficient lag at matrix position (1,3). We will soon see that this choice is much better than those of the two alternative procedures to be considered.

When the sample size is increased to 100 the three criteria unanimously choose the true P*. However, at sample size 200 SBC and PHI choose the true P* while AIC chooses P*1, which is less parsimonious than the true one. This shows AIC's tendency to overparameterize in large samples (see Section 3.4.1). We may note that while none of the relevant lags are missing from AIC's choice at sample size 200, the minimum AIC value is relatively very close to the AIC of the true P*.

From these somewhat limited simulation results, we may conclude that in small samples the choice of AIC is preferable while in large samples SBC and PHI are acceptable. How-

TABLE 3.10

System estimated AIC, SBC, and PHI for chosen lag length matrices, P*'s

Sample size 50				
Lag length matrix		AIC	SBC	PHI
P*1 =	$\begin{matrix} 1 & 1 & 1 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{matrix}$	16.5177	33.7259	23.0707
P*2 =	$\begin{matrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{matrix}$	16.1062*	29.4903	21.2029
P*3 =	$\begin{matrix} 1 & 1 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{matrix}$	16.3350	27.8072*	20.7037*
P*4 = (True)	$\begin{matrix} 2 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 2 \end{matrix}$	18.6016	35.8098	25.1546
Sample size 100				
P*1 =	$\begin{matrix} 2 & 3 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 2 \end{matrix}$	-1.7911	29.4707	10.8612
P*2 =	$\begin{matrix} 2 & 4 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 2 \end{matrix}$	-0.0686	33.7986	13.6380
P*3 =	$\begin{matrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{matrix}$	-1.4349	16.8013	5.9456
P*4 =	$\begin{matrix} 2 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 1 \end{matrix}$	-2.7895	20.6570	6.6997
P*5 = (True)	$\begin{matrix} 2 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 2 \end{matrix}$	-7.0157*	13.8255*	1.4190*

continued on the next page.

Sample size 200				
P*1 =	2 3 1 0 2 3 0 0 2	-48.0195*	-5.1415	-30.6674
P*2 =	1 1 1 0 1 2 0 0 1	-35.5639	-12.4756	-26.2204
P*3 =	1 3 1 0 2 2 0 0 1	-41.9318	-8.9487	-28.5840
P*4 = (True)	2 1 1 0 2 2 0 0 2	-46.4208	-13.4372*	-33.0726*

* attached to numbers indicates the minimum values.

ever, the use of more than one criterion is always recommended since they are likely to indicate possible misspecifications.

3.5.2.4 Two Alternative Procedures

In order to compare our procedure with two alternatives, we report in Table 3.11 the choices made by AIC when it is used to estimate the maximum lag order (p) of the system as a whole and the choices indicated by the Tiao-Box symbolic representation of the partial correlation matrices. In addition to the visual inspection of the symbolic partial correlation matrices , Tiao and Box use a chi-square statistic to determine p, but we did not calculate this statistic. In this symbolic presentation + and - signs are used to indicate the coefficients which are greater, in absolute value, than 2 times the standard error (sign indicates the sign of the coefficient) and . indicates values in between.

TABLE 3.11

Maximum lag order of the VAR estimated by AIC and Tiao-Box symbolic partial correlation matrices

Sample size 50

AIC: p = 1

Partial correlations

Var/Lag	1	2	3	4	5-10
x	+ - .	. . .	. . .	. . .	. . .
y	. + .	. . .	. . .	. . +	. . .
z	. . +	. . .	. . .	. - .	. . .

Sample size 100

AIC: p = 1

Partial correlations

Var/Lag	1	2	3-6	7	8-10
x	+ - .	. . .	. . .	. . .	. . .
y	. + .	. . +	. . .	+ . .	. . .
z	. . +	. . .	. . .	. . .	. . .

Sample size 200

AIC: p = 2

Partial correlations

Var/Lag	1	2	3	4	5	6	7-10
x	+ - -	. - .	. . .	. . .	. . .	. . .	. . .
y	. + +	. . +	. + -	. . .	- . .	. . .	. . .
z	. . +	. . .	. . .	. . .	. . .	+ . .	. . .

+ is >2.S.E., - is <-2.S.E., and . between

From this table we can see that AIC picks the correct maximum lag order (i.e. p=2) only when the sample size is

200. This reaffirms Nickelburg's findings (see Section 3.2.2). On the other hand arriving at the true P* by visual inspection of the symbolic matrices is rather difficult. For example, when the sample size is 50 we have two possibilities: one is p=1 and the other is p=4. If the associated lag length matrices are denoted by P*(1) and P*(4) then we have

$$P^*(1) = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad P^*(4) = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 4 \\ 0 & 4 & 1 \end{bmatrix}.$$

It is easy to see that neither of these matrices approxmimate the true one. (Compare this with AIC's choice in sample size 50 in our procedure.) An important lag is missing in P*(1), i.e. matrix position (2,3), while in P*(4) there is a false fourth order lag occuring at matrix position (3,2). We can find similar problems at sample sizes 100 and 200, too. Such distortions are quite likely to be a result of the multicollinearity problem we mentioned in Section 3.2.2.

3.5.2.5 Efficiency Gain of Using SUR Estimation

For the sake of completeness of this experiment we add one more table reporting the standard errors of estimates from OLS and SUR estimations and the final estimates of Σ corresponding to the true model (Table 3.12).

TABLE 3.12

Standard errors of estimates from OLS and SUR estmation and final estimates of Σ (True model)

x-equation						
	N=50		N=100		N=200	
Var	OLS	SUR	OLS	SUR	OLS	SUR
x(t-1)	12.5849	11.9329	8.4618	8.1065	5.6421	5.3308
x(t-2)	12.4841	11.6465	8.2747	7.7909	5.2869	4.8932
y(t-1)	15.3923	15.0006	9.6586	9.3785	5.8311	5.5861
z(t-1)	16.5243	16.4343	11.0313	11.0160	6.9362	6.8796
y-equation						
y(t-1)	14.7744	13.7179	9.7089	9.0296	6.9307	6.2927
y(t-2)	14.5936	13.3396	9.1974	8.4588	6.3272	5.6780
z(t-1)	16.8377	16.3364	10.6828	10.3449	8.0055	7.6152
z(t-2)	17.4802	16.5679	11.1354	10.4181	8.3996	7.6402
z-equation						
z(t-1)	14.4308	14.3556	9.9920	9.9913	7.0977	7.0670
z(t-2)	14.6183	14.5058	10.0392	10.0382	7.1235	7.0816
Residual covariance matrix from SURE						
x	1.88 .38 -.17		1.55 .38 .02		1.30 .45 .13	
y	.90 .25		.80 .23		.91 .30	
z	.77		.80		.83	

Note: Standard errors are expressed as S.E.x10^2

We can observe from this table that the gain in efficiency of using SUR is rather marginal. (Note that SUR estimates are more efficient asymptotically.) The low gain in efficiency may be attributable to the correlation among the lag explanatory variables and to the small cross equation correlation estimates (not reported). (The cross equation correlation estimates reach true values as the sample size is increased.) The estimate of Σ reported in this table is a result of only one iteration: further iterations appeared unnecessary considering the closeness of OLS and SUR estimates of Σ. The estimate of Σ improves (i.e. is closer to true Σ) as the sample size increases.

3.6 SUMMARY AND CONCLUSIONS

In this chapter we have proposed a sequential model building strategy for estimating the lag length matrix of a vector autoregression. The major ingredients of the procedure have come from Hsiao [1979] and Abeysinghe [1982]. It is aimed at serving a dual purpose: the specification of the VAR model itself and the derivation of causal implications from the parameter estimates resulting from the final specification. It avoids, therefore, the need for a secondary test for G-causality though one could do so if it is required to conform to a conventional test procedure.

The model specification is carried out in three stages similar to the Box-Jenkins model building methodology. These three stages are, however, not mutually exclusive.

1. The tentative specification of the model is done by specifying each equation of the model seperately. The variables are added to each equation sequentially. A chosen model selection criterion will determine the lag order of each variable entering the equation. The individually specified equations are combined to identify the full model.

2. The estimation procedure is basically least squares. The estimation of single equations at the tentative specification stage requires only OLS while the joint model may be estimated using a seemingly unrelated regression procedure.

3. The diagnostic checking is carried out by specifying a number of lag length matrices (P*'s) in the neighbourhood of the tentatively identified one and then by examining the values of the chosen model selection criterion calculated from the jointly estimated model. The model which minimizes the joint criterion value will be retained.

We have evaluated the performance of the procedure at length. In order to aid the choice of a model selection criterion we carried out an analytical exercise as well as a sampling experiment. These evaluations led us to choose AIC, SBC, and PHI for practical use. These were chosen not because of their superior performance, but because they represent two groups of model selection criteria on the one hand, and they place three differnt emphases on the bias-variance trade off on the other. The use of these three together, instead of one, is likely to work as a check against misspecifications. Moreover, they can be used to arrive at a set of lag length matrices for diagnostic checking as we did in experiment 2.

The sampling experiments show that the performance of the model selection criteria is quite satisfactory (as compared to Nickelburg [1985], for example) when the parameter values are moderately high. These criteria tend not to agree with each other when the parameter values are small, especially regarding the presence or absence of G-causality, at the tentative specification stage. To remedy this carefull diagnostic checking is required. If sample size could be increased (e.g., from 50 to 100), these criteria perform better even if the parameter values are small.

An important finding is that the presence of high contemporaneous correlation is likely to cause misspecifications at the tentative specification stage. Such misspecifications have to be uncovered at the diagnostic checking stage. Another aspect we were interested in was the proxy effect resulting from the sequential specification. The evidence from the VAR(∞) model on the proxy effect is not conclusive enough, and it is virtually absent in the VAR(1) experiments. Our diagnostic checking procedure appears to be quite successful in eliminating such effects.

Chapter IV

A REVIEW OF ECONOMIC THEORIES OF FERTILIY

"Are children inferior goods?" is a challenging question that emerged when neoclassical price theory was applied by a group of Chicago-Columbia economists to explain parents' demand for children. Although this exercise has enhanced economists' awareness of economic demographic relationships the answers given to the above question have not convinced the majority of the profession.[44] On the other hand, there are a number of alternative economic theories which can also be used to explain the observed negative association between income and fertility. The most closely related alternative to the above theory is the Pennsylvania model proposed by Easterlin, Pollak, and Wachter [1980]. The Pennsylvania model may also be taken as representative of alternative views on income-fertility relatonship. The objective of this chapter is to provide a critical evaluation of these two theories.

---

[44] Heckman and Willis [1975], two major contributors to the price theoretic approach, write, for example: "A number of explanations for the puzzeling inconsistency of the 'income effect' on fertility have been advanced, but it is probably accurate to say that none has been universally accepted."

4.1 CHICAGO-COLUMBIA APPROACH

The Chicago-Columbia (C-C) approach to human fertility clearly represents a seperate school by virtue of the unique nature of the thoughts of the adherents. It was an attempt to extend the boundaries of neoclassical price theory to the terrain of fertility which has been in the domain of demographers and sociologists. Gary Becker pioneered the task in 1960,[45] and the development of the theories of allocation of time and human capital in the 1960's helped delineate the conceptual framework of the C-C economic model of fertility. This conceptual framework is variously known as the "household production model," "new theory of consumer behavior," or "new home economics."

The publication of two supplements to the Journal of Political Economy in early 1970's, which appeared as a National Bureau of Economic Research (NBER) volume in 1974 [Schultz, T.W., 1974] exemplified the formal application of the household production model to fertility.

4.1.1 Basic Framework

According to the household production model, the household derives utility not directly from market goods as postulated by traditional theory, but from a set of household produced non-marketable "basic commodities" (Becker's term, 1965) such as health, prestige, companionship, and child services

[45] The idea of applying microeconomic theory to fertility goes back at least to Leibenstein [1954].

(a vector of attributes of children) the inputs of which are goods bought in the market place and time of the household members. A given household production technology will determine the properties of each production functon.

In applications of household production model to fertility, children are assumed to be consumer durables and each family[46] is assumed to maximize a lifetime family utility function of the form [see Becker, 1981, Ch. 5]

$$U = U(N,Z), \tag{4.1}$$

where N is the quantity of children and Z is a composit commodity representing all other basic commidities.[47] This utility function has to be maximized subject to a budget constraint of the form

$$\pi_N N + \pi_Z Z = I, \tag{4.2}$$

where I is "full income" (i.e., the sum of the nonearned income, e.g., property income, of the family plus the potential wage income that could be earned if the total time of the husband and wife were allocated to market work) of the family and π_N and π_Z are "shadow prices," which have to be calculated using the explicit and implicit prices of market

[46] Becker [1981], in his A Treatise on the Family, has used the terms household and family almost synonymously (see his Ch. 2).

[47] The appropriate variable for N in (4.1) is child services [Willis, 1974]. Keely [1975] points out that N is a poor proxy for child services unless other attributes related to child quality are held constant.

goods and time inputs used in the production of N and Z. The production functions of N and Z are assumed to have constant returns to scale in order that the shadow prices do not depend on the levels of N and Z. Since the cost of own time, household technology, and efficiency in household production are assumed to be different among families, these shadow prices are expected to differ from family to family.

The optimal solutions resulting from maximizing (4.1) subject to (4.2) are functions of the shadow prices and full income; taste is assumed to be homogeneous or invariant. The usual predictions of consumer choice theory are expected to hold here as well. That is, a rise in the shadow price of children will reduce the demand for children, *ceteris paribus*, and a rise in full income will increase the demand for children, *ceteris paribus*, if children are a normal good.[48]

4.1.2 Methodological Issues

The C-C model has stimulated a great deal of research. The results have led T.P. Schultz [1981: 186] to conclude: "Cross sectional studies of individual countries at all levels of development have confirmed the qualitative predictions of this rudimentary demand theory of fertility." Becker [1981: 95] also claims that this model is applicable not only to modern developed nations of today but also to

[48] When child quality is taken as an argument of the utility function (4.1) the budget constraint (4.2) takes a nonlinear form: See Willis [1974], DeTray [1974], Becker and Lewis [1974], and Becker [1981].

Western countries of a few centuries ago, as well as to developing countries of this century.

The C-C model has generated a fair amount of criticism as well as alternative theories. Although some have responeded to these criticisms quite positively [T.P. Schultz, 1981] others have continued to reproduce the model in its original spirit [Becker, 1981]. Kelley [1976] characterizes this latter attitude of the school as showing "a mild intolerance of other approaches."

The basic methodology of the C-C school is the Chicago approach to positive economics; that is, a model is evaluated only on the basis of its predictive ability while the realism of assumptions and the explanatory power of the model are considered less important. Ben Porath [1982], who appears more inclined toward the C-C approach than to any other, comments on Becker's Treatise as follows.

> Indeed my main bone of contention with Becker has to do with what one might call the "deductive illusion." Becker's style is to start with a specific theoretical formulation; certain implications are then deduced, and they are confronted with some facts or data which provide a "test" of the theory. The life history of better ideas is often the reverse. A phenomenon is observed, a theory is formulated within some framework to account for what is already known, additional implications of the theory are deduced, this includes examinaton of new evidence, and the outcome generally leads to a reformulation of the theory. [pp. 59-60]

Further, the C-C approach to theorizing is mathematics oriented with the use of very abstract concepts. For Samuelson [1976] the appearence of Easterlin's theory of American baby boom is a

> relief from the rather sterile verbalization by which economists have tended to describe fertility decisions in terms of the jargan of indifference curves, thereby tending to intimidate non-economists who have not mis-spent their youth in mastering the intricacies of the modern utility theory. [p. 244]

As we shall see in the next section, one cannot ignore the crucial role played by assumptions if a model is to explain real world phenomena. Abstraction, on the other hand, though required, should not lead one to think that the exercise is mere "denk spielen" (see Section 4.1.4).

4.1.2.1 Assumptions

The C-C model rests upon a large body of assumptions. While some of them appear to be reasonable others need reconsideration. The most salient assumptions are:

1. Stable preferences
2. Maximizing behavior
3. Equilibrium
4. Existence of a family utility function
5. One-period planning
6. Known and fixed shadow prices
7. Constant returns to scale
8. Nonjointness in production
9. Mother's time intensity of children

The first three assumptions are basic to neoclassical price theory and the rest are specific to the household production model. Each of these assumptions has received some attention in the literature. In this section we reevaluate these assumptions, especially to remind the users of the model where further thinking is required.

1. Stable preferences: In the new household production model, tastes are assumed to be stable over time and invariant across the society. Probably the strongest statement made so far about tastes is that of Stigler and Becker [1977]: Like the Rocky Mountains, they say, tastes are the same for every one all the time.

Utility maximization with stable preferences is the basic postulate of traditional theory, too, but the traditional theory does not rule out the possibility of variable tastes (over time as well as across households). However, this theory assumes that tastes, if variable, are totally unrelated to the other variables in the model. Accordingly taste is left out, among other things, as the residual variation unexplained by prices and income.[49]

---

[49] Becker has not been radically different from this approach when he wrote in 1973, together with Michael: "For economists to rest a large part of their theory of choice on differences in taste is disturbing since they admittedly have no useful theory of formation of taste, nor can they rely a well developed theory of taste from any other discipline in social sciences since none exists." [Michael and Becker, 1973: 380]

Not all economists are in favor of the assumption of stable or invariant tastes. Taste formation through interdependent preferences, habit perpetuation, and so on, have been recognized since the time of classical economists. A few, only a few, noteworthy attempts in the postwar era to incorporate taste into the formal model are: Duesenberry [1949], Pollak [1970, 1976, 1978] in the nondemographic literature, and Easterlin [1973, 1978], Leibentein [1974, 1975], and Easterlin et al. [1980] in the economic demographic literature.

In developed countries the average family size has fallen from large to medium to small.[50] The preferred average family size today, regardless of the budget constraint, is very unlikely to be of large parity. Evidence gathered by Easterlin [1978: 67] indicates that the ideal or the preferred family size of many adult Americans concentrate in the two to four children range. Even if allowance is made for the possibility that the ideals tend to be dominated by observed behavior, large parity may not effectively enter the preference map. The shift in average family size may have been caused by factors other than a change in preferencs, but eventually preferences also shift to accept the observed as a norm.

---

[50] Each parity size may be defined as more than six, four to six, and zero to three children, respectively [Leibenstein, 1975: 21].

Easterlin et al. [1980] point out that since controlled experiments are not possible in this matter it is difficult to disentangle price and income effects from taste effects. But the introduction of habit formation into demand models not only improves R^2, but also changes the estimates of price and income elasticities and marginal budget shares. Further, they argue, with reference to the C-C model, that assessing technological differences faced by different households is no easier than that of tastes.[51] They continue to argue: "Indeed economists do not have very satisfactory theories of systematic differences in technologies available..., and the problems posed by household technologies are substantially more difficult because the outputs are often not measured directly." Therefore, to ignore taste on the ground that there is no theory of taste formation and to incorporate technology on the ground that there is one, is unacceptable.

Furthermore, the C-C models with such a restrictive set of assumptions are not capable of determining a priori the direction of fertility change in response to change in explanatory variables. Willis [1974: 39] writes: "Since the issues to be resolved in order to derive hypotheses about the signs ... depend on the nature of family tastes, economic theory as such has very little to say." Such expres-

[51] In the C-C model, one factor that contributes to differences in the costs of household production (cross sectionally and over time) is the household production technology.

sions indicate the need for explicit treatment of tastes in the model.

2. Maximizing behavior: It is assumed there exists a constraint-concerned calculating (rational) household (or family) which always tries to maximize its own utility. Heckman and Willis [1975] try to justify this assumption as follows:

> [I]t may strain the credibility of the reader to suppose that behavior is in fact governed by the complex calculations implied by our model... We shall simply assert that it is plausible to imagine that "rules of thumb" or "behavioral norms" which emerge to guide decision making in complex situations tend to be perpetuated to the extent that they approximate optimal decisions." [p. 115]

The most outstanding figure who questions the maximizing postulate not only in relation to fertility but also in general as a behavioral assumption, is Leibenstein [1976, 1981]. He argues that active decision making is less frequent compared to passive decision making. People work mostly according to a routine and they do not change their behavior unless a significant change in pressure, caused by some force, is observed. As pressure keeps changing the individuals change their behavior by looking at the behavior of others in similar circumstances, or according to some authoritative advice, or by making partial calculations. They do full calculations only as the last resort and only at points of significant pressure change. Leibenstein, thus, does not rule out maximizing behaviour. What he claimes is that routine behavior is more prevalent than maximizing behavior. This is so because the total cost (monetary as well as psychic) involved in decision making is high.

However, child bearing, like marriage, is an important decision that a couple has to make. It is likely that some make full calculations, at the onset of the marriage, as to the total number of children they are going to have. But some may do these calculations only for the last child. But in a one-period planing model, the best approximation one could have is to make all the calculations at the begining.

The more important point, which has been emphasised by many, is, however, a different one. That is, the individual or household maximizing behavior does not necessarilly result in optimal solutions for the group (e.g., the Prisoner's dilemma). Even Bernard Shaw has pointed out the potential conflict between parental decisions as to family size and the collective interest as to total population size [see Archives, 1983]. Arthur [1982] captures the essence of these arguments thus: "lack of information, nonconvexities, externalities, indivisibilities, increasing returns, and prisoner's dilemmas all call for the use of rules, rights, agreements, hierarchies, organizational institutions- in short structure." Thus, quite contrary to the argument made by Heckman and Willis, "behavioral norms" arise as a solution to the non-optimality of individual solutions rather than as a result of individual solutions. Obviously this is a different type of rationality. At times individuals or individual households may follow a routine defined by society, still the solutions may be optimal. This is a question about the macrofoundation of microeconomics.

3. Equilibrium: Although equilibrium analysis is well accepted among fertility theorists, the stability of the equilibrium is at question. The point to make here is essentially related to the previous discussion of maximizing behavior. For the purpose of analysis the optimal solution resulting from individual maximizing behavior may be called "goal equilibrium" while the equilibrium resulting through non-personal forces (e.g., the market) may be called "non-goal equlibrium."[52]

The underlying assumptions of the C-C theory of fertility ensures a non-zero goal-equilibrium level of children for each family. This is so because the theory holds that the marginal benefit (or utility) of an extra child declines with the increase in the number of children and the parents choose to have children up to a point where the marginal cost of having a child (usually assumed to be fixed above zero) equals the marginal benefit of that child. There is no guarantee, however, that this equilibrium will occur at or above the replacement level of fertility. The experience of developed nations is that there is a significant percentage

[52] These terms are due to Chaing, Alpha C. Chaing [1984] Fundamental Methods of Mathematical Economics, Third ed. (New York: McGraw-Hill).

of women who never marry [see Davis and Oever, 1982: 509] and mostly have no children at all and still another significant portion, though married, have no children or have only one child.[53] Unless an offsetting effect comes from those who have more than two children, population growth should decline to levels below zero [Caldwell, 1982: 264]. Under these circumstances there is no guarantee that industrialized nations will reach a stable "nongoal-equilibrium." As Kingsley Davis [1983] speculates, we could expect the emergence of a new set of "behavioural norms" which guarantee the attainment of optimal and stable "nongoal-equilibrium" level of fertility for the nation as a whole.

4. Well behaved family utility function: Utility function which is well-behaved, i.e. the conformity with assumptions such as continuity, strict quasi-concavity, nondecreasing form etc., is required for mathematical tractability and to yield unique solutions.

The assumption of a family utility function involves the problem of aggregating the preferences of individual members. Becker [1981] considers "altruism" as a unifying code of preferences in a family. Thus the presence of an altruistic head, who is a benevolent dictator according to Becker's descriptions, constrains selfish members from selfish behaviour. Becker has embodied this argument in his Rot-

[53] In Canada, in 1981, 23 percent of ever married women who were 45 years of age and above had less than two children [1981 Census of Canada, Catalogue 92-906].

ten Kid Theorem [Ch. 8], which indicates that the selfish acts of a member do find recourse from the altruistic head who decides the slice of the income pie each member receives.

However, lack of information and disagreements over resource allocation among family members may limit altruism as the basis of a family utility function. Leibenstein [1976] advocates the use of the individual as the unit of analysis instead of the family or the household. An important implication of using the individual as the unit of analysis is that fertility decisions are subject to husband-wife conflicts, which may not be uncommon. An understanding of the mechanism of conflict resolution, which Becker attempts through altruism, is a much needed research area, not only for defining the suitable unit of analysis, but also for understanding fertility, divorce, and other family related decisions.

5. One-period planning: This assumption indicates that all child bearing, child rearing, and consumption decisions are made by the couple at the outset of their marriage. This assumption closes the door for sequential decision making. Such an assumption is plausible if the couple is well informed about their future, i.e., if there is no uncertainty about events as yet not experienced. Given the difficulties of constructing dynamic models, a one-period planning model still may provide quite useful insights.

<u>6-8</u>. <u>Known and fixed shadow prices</u>, <u>constant returns to scale</u>, <u>and nonjointness in production</u>: All these assumptions are closely related. The commodity shadow prices are equal to the marginal costs of producing commodities, and under constant returns to scale, the marginal cost is constant and equal to the average cost. As a result the shadow price of a commodity does not depend on the commodity output level produced. Under the assumption of nonjointness each commodity appearing in the utility function is produced by a seperate production process; therefore the commodity shadow prices do not depend on the output levels of other commodities. As a result, the commodity shadow prices are exogenously given to a household and commodity demand functions satisfy all conditions of the traditional demand functions.

The important question is how reasonable are these assumptions in reality. The cost of rearing the second child appears to be less than the first and subsequent costs may be approximately the same (Leibenstein, 1975: fn. 16]. Therefore, both increasing and constant returns to scale may be present. Pollak and Wachter [1975] have shown that even if the assumption of constant returns to scale is retained, the independence of relative shadow prices from output levels depends on the assumption of nonjointness. However, they argue that jointness in household production is the rule not the exception. Thus if these assumptions are more often violated than not, then shadow prices do depend on the commodity output bundle a household chooses and therefore are not exogenous.

Pollak and Wachter therefore recommend, instead of shadow prices, the use of directly observable market goods prices, wage rates, and nonlabor incomes in commodity demand functions. These functions, however, cannot be interpreted as demand functions in the traditional sense because the conditions such as Slutsky are now not applicable.

9. Mother's time intensity of children: This assumption will be discussed in the next section.

4.1.3 Research Problems: Fertility Decline and Income Growth

We can now return to the opening question of this chapter and examine the explanations provided by the C-C model to the observed negative association between fertility and income. The main line of argument of the C-C approach is that the effective price of children does not remain constant as income grows; rather it increases. Two price-augmenting effects associated with income, recognized by the C-C approach, are the time-cost of the mother, and child quality.

4.1.3.1 Time Cost

With the introduction of the notion of the opportunity cost of time by Mincer [1962, 1963], the time cost became an important component of the price of a child. However, the time cost of a child is measured by the female wage rate under the assumption that chidren are intensive in mother's time. Thus it is argued that since the female wage rate has been rising over time, the shadow price of a child must have been rising also, and this acted to reduce desired fertility, despite the rise in family incomes. In cross section analyses, it is argued that the rich men tend to have wives with higher wage potentials [Mincer, 1963] or higher values of time [Willis, 1974], hence rich families face a higher price of children. And this price effect more than offsets the income effect on fertility. Becker [1981: 98] states that two thirds of the total child cost (price) in the U.S. is due to this time cost as measured by the female wage rate.

One problem of using the female wage rate to represent the wife's time cost is the presence of mothers not working in the labor market.[54] In every developed country, except the communist block, a large proportion of married women (perhaps as large as fifty percent) in the medium and high income groups are not in the labor market. Japan is a typical example where the level of fertility has declined quite rapidly in a short period whereas the labor force participation of married women remains low. The question is, what is the appropriate measure of time cost of nonworking mothers.

---

[54] The term "nonworking mothers" will be used to indicate those mothers who do not work in the labor market.

The C-C model answers this question by assuming that a mother chooses to allocate her total time to nonmarket activities because her marginal productivity, hence the shadow value of her time, of nonmarket activities is greater than the market wage rate which she could command [T.P Schultz, 1981, Ch. 4; Becker, 1981: 26].[55] This is a questionable assertion. If a mother (or a woman) does not participate in the labor market for cultural reasons, as is the case for many medium and high income families of different cultures, then the value of time argument is irrelevant in explaining the fertility decline of these groups [Leibentein, 1974].

Even in industrialised developed countries, where (alleged) cultural biases are less apparent today, it was the emerging factory system which replaced the household as a production unit and pushed many mothers out of the labor market and deprived them of their economic power [Davis, 1984]. Possessed with large families these mothers may have stayed out of labor market by necessity rather than by choice. It is interesting to see that, later, having a non-working wife became a part of husband's status under this nontraditional system which is now regarded as traditional by many. Therefore, both the necessity and value system ap-

---

[55] In the case of working mothers, they choose a certain number of hours to work in the market (assuming that they have the option of this choice) by equating the market wage rate to the marginal productivity of the nonmarket activities. Thus they work in the market up to a point where the wage rate exceeds the nonmarket marginal productivity of time.

peared to have played a role in determining the working status of a wife.

Even if it is granted that some wives choose not to work in the labor market because of marginal productivity considerations alone, an important empirical question remains. That is, how to impute values for time allocated to nonmarket activities? One approach to this problem is to use the mother's education as a proxy for the value of her time. However, education can influence many other things such as taste, fecundity through consumption, infant and child deaths. Since all these variables affect fertility, one cannot claim that the full effect of education on fertility results from the value of time alone.

T.P. Schultz [1981, Ch.4] has suggested the use of a wage function with explanatory variables such as education and training. First one estimates the wage function for a group of working mothers and then, assuming the regression coefficients are the same for all mothers, predicts wages for nonworking mothers using their values of the explanatory variables. These estimates are in turn taken as the imputed values of time for nonworking mothers. The success of this procedure depends on many things. For example, if these two groups differ in many aspects other than the chosen set of explanatory variables then the imputed wages would be misleading. Furthermore, since nonworking mothers are hypothesised to have higher nonmarket marginal productivity (value of time) than the market wage rate they could command, the use of the wage rate of working mothers in the calculation of the regression coefficients biases the estimates.[56]

Other than the two points discussed above in relation to the time cost argument--the presence of a large segment of nonworking mothers who may not work for reasons other than marginal productivity considerations and the difficulty of measuring their value of time if they do consider marginal productivity in choosing home or work--there is a third point which needs consideration. That is, the assumption that children are intensive in mother's time. There is no doubt that children are intensive in someone's time, but whose time is an open question.

Evidence appears to indicate that children are not so intensive in mother's time. Some estimates show that the time requirement of a mother for an extra child varies between 5 and 20 hours per weak, depending on the age of the child, birth order, and the method of computing extra hours [Weller, 1977: 506-7]. The upper limit, 20 hours, of these estimates is only a half of a forty-hour work week. Moreover, in less developed countries where the extended family is still prevalent, grandparents and older siblings often provide the time for the young. Mothers, in fact, spend a substantial amount of time in agricultural work and other

[56] In empirical work one has to be careful in mixing these two samples because of the heteroskedasticity between the two groups arisng from the wage imputation procedure.

nonchild activities. In developed countries, where the extended family has now almost dissappeared, obligatory schooling of children has reduced mother's time intensity of school-age children substantially [Davis, 1984]. The expansion of the day care industry and baby sitting facilities may largely reduce the mother's time intensity of preschool-age children.[57]

If children are less intensive in mother's time than postulated by the theory, one has to explain the observed negative association between the female wage rate and fertility. In empirical applications of the C-C model very few scholars have calculated the price of a child by incorporating the actual time cost and the cost of market purchased goods.[58] The use of the female wage rate in regressions of fertility is based on the theoretical implication that the greater the mother's time intensity of child rearing, the stronger the negative association between fertility and the female wage rate. Empirical studies mostly find a strong negative association. Is this evidence against the argument that children are less intensive in mother's time than postulated? The answer would be, not necessarily. Although

---

[57] Easterlin et al. [1980] point out that time intensity "results from the interplay of tastes, technology, and the budget constraints,... [t]here is little evidence to support the notion that the technology requires child-rearing to be time intensive independent of prices and tastes." [p. 117]

[58] Lindert [1977, 1980] is an exception who tries to measure the relative cost of a child by fixing the input bundle.

greater time intensity implies a stronger association between the two, the reverse need not hold. Rising female wages due to market forces and falling fertility due to forces unrelated to female wages are likely possibilities. In the rest of this section we will investigate different possibilities of the relationship among fertility, labor force participation of wives, and the female wage rate and in Section 4.2.1 we will examine causes of fertility decline other than rising female wages.

A fruitful discussion of fertility decline and its relation to the female wage rate cannot be conducted in isolation from the labor force participation (LFP) of women, married women in particular. Extensive research has been done, mostly using cross section data, on causality between fertility and LFP of women. The relationship between the two tends to be negative, especially if women are employed in modern labor markets.[59]

Four types of causal relationships have been assigned to the negative association between LFP of women and fertility. Firstly, it is argued that the observed negative association between LFP of (married) women and fertility is not a causal relation, but rather a result of other factors, mainly the female wage rate [Mincer, 1963]. Becker [1981: 245] considers that falling births, rising divorces, rising LFP of

---

[59] See Weller [1977] for a summary of research findings on this relationship across nations.

(married) women and a number of other demogrphic changes are caused by the rising earning power of women. This is the view taken by other advocates of the C-C model [see T.P. Schultz, 1981, Ch. 6]. There are studies, however, which find that even after controlling for other factors, the relationship between fertility and female employment persists [Weller, 1977].[60]

Secondly, it has been argued that falling fertility, among other things, has paved the way for rising LFP of married women [Davis, 1984]. Davis from an evolutionary perspective draws attention to motivational factors of LFP of married women. The basic thrust of Davis' argument is that, despite the attraction provided by rising female wages and increasing employment opportunities for women, there is a large segment of married women not working in the labor market even in 1980's. He therefore concludes that motivational factors play an important role in pushing married women into the labor market. One of these motivational factors is smaller family size achieved at early years of marriage.[61] Davis argues:

> A woman who is at home full time and yet has only one to three children who are at school most of the day has less than a full-time vocation. Put

---

[60] Esterlin and a number of others have emphasised "relative income" as a major common cause of fertility and LFP of married women [see Section 4.2.1].

[61] For example, a steady rise in LFP of married women in the U.S. is evidenced only since 1930's and by this time the average family size had already reached two to three children per family [see Davis].

> in economic terms, the oppotunity costs of nonemployment are very high in relation to the familial needs being satisfied. [p.408]

Though Davis considers reduced fertility a cause of rising LFP of married women he does not discuss the causes of fertility decline. However, his analysis strongly implies that the secular decline in fertility is a consequence of forces which were at work prior to the onset of rising female wages.

Another motivational factor which Davis emphasises is rising separations and divorces, which also show a "lead" relationship with LFP of married women. A career, Davis argues, provides a hedge against rising separations and divorces. Even in the absence of separations and divorces, a career could establish independence and self-worth of a wife. Such a change in attitudes may have an important bearing on rising education levels and the women's movement. Thus, the female wage rate is only one explanatory variable of LFP of married women.

Thirdly, it is argued that rising LFP of women causes fertility decline. There may be a number of factors which may explain rising LFP of women (see previous arguments). Fertility, however, cannot be one of them. Furthermore, other explanatory factors must not show a lag relationship with LFP of women, and the latter with fertility.

Whatever reasons may be behind the rising LFP of women, the question is why it relates negatively to fertility. One explanation is "role incompatibility." That is, given that the role as a mother is incompatible with the role as a worker in the market, the acceptance of employment implies some or full sacrifice of the other. Another argument is the opportunity cost of time, measured in terms of foregone income, if time were devoted to children. (More on this point later.) Another way employment could affect fertility negatively is through a change of tastes. Weller [1977: 509] argues that "employment may alter the taste matrix in such a way that the females's tastes for child bearing vis-a-vis other activities are diminished." Employment may also increase the influence a wife has over family decisions and provides an opportunity to exercise her desire to maintain low fertility. Another mechanism through which employment may affect fertility is increased age at marriage. Women may delay marriage until a career is established.

Fourthly, some have argued that fertility and LFP of women affect each other [see Weller, 1977]. All the arguments made regarding the first three cases become relevant in explaining mutual causation. Weller points out, however, that sequentiallity or simultaneity of joint influence is not well understood at this time.

The above discussion of the relationship between fertility and LFP of (married) women shows a diversity of opinion.

One of the deficiencies of research done on this subject is the inappropriate inferences made on causation based upon cross-sectional data. Weller [1977:512] points out that "because this is a causal relationship being evaluated, the analysis should be longitudinal."

Before closing this subsection one more point has to be emphasised with regard to the argument that rising female wages increase the opportunity cost of time devoted to children by the mother, therefore mothers choose market work at the expense of children. A puzzeling question is why mothers should consider only the opportunity cost of children in their choice, but not the opportunity cost of work (in terms of forgone number of children and other child-related aspects). In other words, since both work and children offer their own rewards (monetary as well as psychic), choice of one implies sacrifice of the other, assuming the incompatibility of the roles as a mother and a worker.[62] In the C-C model mothers must choose market work in order for the model to work. One way to justify this is to argue that the rewards offered by work are greater than those of chidlren. This may be so if one considers monetary rewards only. The question is, can we ignore psychic rewards, net of psychic costs. My opinion is that there is something other than the

[62] Research on role incompatibility shows that the average family size is the lowest for wage earners and the highest for the unpaid family workers. The self-employed category falls in between. In general the wage earners have the least flexible time arrangements for work [Ostry and Zaidi, 1979: 55, note 31].

'choice' of above type that leads mothers to choose work. Mothers choose work, instead of a large family, mostly out of necessity and partly because of changes in tastes and attitudes (see Section 4.2.1).

4.1.3.2 Child Quality

The other price (of child) augmenting effect of the C-C model is the interaction between child quantity and quality. The major advocate of this approach is Becker [Becker, 1960, 1981; Becker and Lewis, 1974]. Becker [1981: 102] thinks that the most important cause behind the negative association between income and fertility (secular as well as cross sectional) is the price augmenting effect of the interaction between quantity and quality of children.[63]

Quality per child (Q) is another commodity in the utility function of the family. Maximizing this utility function subject to a nonlinear budget constraint, Becker shows that the shadow price of a child depends on Q (i.e., $\pi_N = p_c Q$) and the shadow price of Q depends on the number of children, N (i.e., $\pi_Q = p_c N$) where p_c is defined as "constant cost of a unit of quality." Becker defines π_N as the "quality expenditure per child" and π_Q as the "expenditure on N children per unit of quality." Since each shadow price depends multiplicatively on the quantity of the other commodity, any

---

[63] This claim is somewhat contradictory to his earlier statement [Becker, 1981: 98] that two thirds of the child cost in the U.S. is due to the mother's time cost.

small exogenous change in one quantity can produce, according to Becker, a larger change in equilibrium shadow prices through the interaction effect even if these two commodities are not close substitutes. "Consequently," Becker argues, "economic development can have significant negative effects on fertility even when the 'true' income elasticity of demand for children is positive and sizable." [p. 112].[64]

The first question one would ask is, what is quality? Becker [1981: 95] defines child quality as "expenditure on children in education, training, and attractiveness."[65] If quality is expenditure, it is difficult to conceive how expenditure could have a "price" [Arthur, 1982]. If Becker's definition of π_Q is rewritten as "expenditure on N children per unit of <u>expenditure</u>," where the underlined word has replaced "quality," we get a circular statement which does not carry any meaning. Furthermore, child quality is more suggestive of child attributes, or "child outputs" than of "child inputs" [Lindert, 1980]. A better term for Becker's child quality would be resource intensity of a child [T.P. Schultz, 1981].

---

[64] Becker [1960] argued that contraceptive knowledge was a major cause for high income families to have less children. Later, however, he abandoned this argument, see Becker [1981: 99].

[65] Becker later redefines child quality as "adult wealth of children" [Ch. 6, p.114].

Apart from the definitional issues there are a number of other problematic areas in the child quality argument. Duesenberry [1960] pointed out that parents do not necessarily choose a level of child quality independent of their own standard of living. Even if they do, it is likely that they try to achieve some target quality level for their children. Since taste is assumed to be homogeneous over income classes, the target quality level may correspond to some average observed by all income groups (e.g. Bachelor's degree or Master's degree). In this case Becker's argument implies that poor families must have fewer children because they will be more constrained by income when they try to achieve the target quality for their children.

To illustrate the above line of argument assume the annual income of a poor couple is $5,000 while that of a rich couple is $25,000. Further, assume that the subsistence expenditure (expenditure on necessary food, clothing and shelter) for both families is $3,000. The poor family is now left with $2,000 to spend on child quality while the rich family is left with $22,000 to spend on child quality. Even if the rich family spends three or four times more on child quality than the poor family could afford, the rich family still could have more children than the poor family. However, if taste is endogenous then the above result can easily be reversed or made to be consistent with the observed negative relationship between family size and income. We explore this possibility in Section 4.2.

According to Becker any small exogenous change in child quality is sufficient to cause a chain reaction to augment the price of children. How does this exogenous change come about? Becker refers to some examples such as public schooling programms to raise Q or family planning programs to lower N. However, both N and Q are parents' choice variables. Whatever program the government or some other agency introduces, the resulting N and Q reflect parents' choices, therefore π_N and π_Q are not really exogenous. Empirical testing of the model is now made doubly difficult [T.P. Schultz, 1981].[66] In light of these problems, Rosenzweig and Wolpin [1980] have attempted to test the interaction hypothesis using the occurence of twins in India as an exogenous increase in N, and they find that this rise in N lowers child quality as measured by years of schooling. This exercise, however, highlights the extreme difficulty of empirical testing of the interaction hypothesis.

4.1.4 Other Research Problems

An important research problem arising from the C-C model is the identification of empirical analogues for the concepts used in the model. The entire model rests on abstract quantities. The arguments of the utility function are attributes such as child services, child quality, health etc., and the budget constraint involves the calculation of shadow prices. Hannon [1982], commenting on Becker's [1981] <u>Treatise on the</u>

---

[66] See the discussion on assumptions 6-8 in Section 4.1.2.1.

Family, writes: "Given the high degree of abstraction and the crucial role of unobservables (utility functions, endowments, efficiencies, heritabilities, prices of children, and so forth) I wonder whether *any* data could be shown convincingly to be inconsistent with the theory." [p. 71] In other words, these concepts leave an opportunity for users of the model to interpret and measure these concepts as they wish in order to obtain affirmative results which confirm the predictions of the theory [Leibenstein, 1980].

Another deficiency of the C-C model is its neglect of the supply side of fertility. This is particularly important in situations where the supply factors of fertility play a more important role than the demand factors. A notable advancement made by Easterlin is the inclusion of the supply side in the C-C model (see next section).

4.2 PENNSYLVANIA MODEL

A major competitor of the Chicago-Columbia model is the model proposed by Easterlin, Pollak, and Wachter [1980] of the University of Pennsylvania. The major ingredients of this model come from the synthesis framework developed by Easterlin [1978]. An overview of the Pennsylvania model (P-model) has been provided by the authors themselves. In the following discussion we compare and contrast the P-model with the C-C model so as to highlight the differences and departures.

The P-model combines the demand side of fertility emphasised by the C-C model and the supply side stressed by social demographers. Apart from supply side considerations, the P-model differes from the C-C model in a number of distinctive ways.

As with C-C model, the P-model is formulated in a static framework (one-period planning), assuming the existence of a family utility function which the family seeks to maximize. However, the objective function of the C-C model has been modified by excluding the child quality variable, by replacing the quantity of children by completed family size (i.e. births-deaths) and by inserting the following additional variables: infant deaths, frequency of coitus, a vector of practices such as lactation, length of time over which fertility regulation is practiced, intensity of the fertility regulation practices, and social averages representing living standards and family size of the previous generation. These social averages are introduced to endogenize preferences. The preference ordering of the other variables in the utility function is conditional upon these averages. The social averages represent the aspiration levels of a family.

Similar to the C-C model there are constraints of household technology, time, and budget. In addition, the P-model has three more constraints: a births function, a deaths function, and an identity representing the completed family size as the difference between births and deaths (infant and child). All these constraints can be reduced to four basic constraints: budget constraint, technological constraint, a births function, and a deaths function.

Unlike the C-C model, constant returns to scale is not a necessary assumption of the P-model. Furthermore, unperceived jointness in household production plays an important role in the P-model; that is, the unperceived effect of consumption on fecundity and infant mortality. The time inputs of the P-model include those of all household members, not just husband and wife. A further difference between the two models is that the optimal solutions for the household decision variables (basic commodities, market purchased goods, time inputs, births, completed family size given deaths, and the other arguments in the utility function except the social averages) are not viewed as demand functions of shadow variables. Rather they are functions of directly observable variables which the household takes as given: goods prices, wage rates, nonlabor income, household technology, births function, deaths function, cost function of fertility regulation, family's reproductive span, and the social averages.

A few more points of clarification are in order. Although births is a choice variable in the P-model, infant and child deaths are not. However, Easterlin et al. admit the possibility that the family could influence the deaths variable through patterns of consumption.[67] Further the P-model's

[67] Note that the deaths variable does not enter the C-C mod-

fertility variables are marital fertility and completed family size, not desired fertility. Since the family's reproductive span (the difference between age at marriage and age at permanent sterility) is given in the P-model, age at marriage is not a choice variable in the model. The age at marriage is determined by forces outside the model. This may be so for an individual family, but if the (social) goal of fertility control is population control then the age at marriage is an important variable to study.

In summary the most distinctive features of the P-model are: endogenous preferences, unperceived jointness in household production, and the supply side of fertility. These are notable improvements over the C-C model, but further improvement is required in areas such as family utility function, coordination between micro and macro equilibria, and dynamic decision making. Apart from these, there are a few specific aspects of the P-model which require mention.

The P-model is quite general, and this appears to have been achieved at the expense of predictions. Easterlin et al. have not derived elasticities of marital fertility and completed family size with respect to the predetermined variables. Undoubtedly these elasticities would take very complicated forms and a priori prediction of their signs likely to be impossible without additional assumptions. This

---

el though in the application of the C-C model to less developed countires, infant and child mortality is usually added to the regression equation of fertility.

by itself, however, cannot be considered a serious drawback because even with such restrictive assumptions the C-C model fails to indicate unambigously a priori the signs of many elasticity coefficients. De Tray [1974], for example, concludes that "no a priori conclusions on fertility behavior can be drawn from the [C-C] model." Willis [1974] also finds similar difficulties.

A further difficulty with the P-model is that its empirical application is severely hampered by a lack of data. Easterlin et al. try to get around this problem by incorporating special assumptions to break the model into four special cases [see Easterlin et al., 1980 Section 2.3]. For this purpose they incorporate the concepts of desired fertility and natural fertility and compare them with actual fertility. Desired fertility is the level of fertility a family could realize in a "perfect contraceptive society," i.e. a society without contraceptive costs (economic costs and preference drawbacks). Natural fertility, on the other hand, is the level of fertility that results when no deliberate attempt is made to control fertility. It is consistent with social and biological controls which produce a fertility outcome below the biological maximum.

According to these special cases, natural fertility is the fertility outcome at low levels of economic development, and for lower strata of socioeconomic classes. This is an important departure from the C-C model because, as Easterlin et al. point out, application of demand models to a natural fertility setting may produce misleading results. For example, the income variable of the demand model may show a positive effect on fertility, but the income is more likely to have worked through the births and deaths functions, not through choice. Therefore, the demand model may "fit," but for the wrong reasons [Easterlin et al.: 84]. The relevance of natural fertility declines at higher levels of economic development and for higher strata of socioeconomic classes, and the demand model, as espoused by Easterlin et al., becomes the relevant model.[68]

4.2.1 Endogenous Tastes and Fertility Decline

Two different hypotheses have been offerd to explain the negative association between income and fertility, one is the price or cost (of children) effect emphasised by the C-C model and the other is the endogenous taste effect emphasised by a number of other fertility models. The advocates of the former hypothesis exclude a priori the latter hypothesis by the assumption of homogenous preferences. The advocates of the latter hypothesis, while not ignoring the former hypothesis, argue that, given the large income differentials and substantial income growth, it is implausible

---

[68] Natural fertility becomes less relevant to traditional societies if these societies practice social controls, not covered by the natural fertility concept, on population size. Outmigration is one such example. See Leibenstein [1980: 136] for a list of population control methods practiced by traditional societies.

for the cost of children, as measured by quality of children or mother's time cost, to explain the negative association under the assumption of homogenous tastes [Leibenstein, 1975: 3]. The P-model is an attempt to bring these two hypotheses, with some modifications, into a single framework, though the P-model lays emphasis on the latter hypothesis.

Formation of taste and its influence on fertility has been variously expressed since the time of classical economists. The classical economists, like modern economists, held the view that a rise in real incomes will raise fertility and hence population. However, Malthus saw that the rich postpone their marriages to reduce births in order to maintain higher living standards [Malthus, 1959]. Bastiat argued, very much in a modern spirit, that not only the rich but all groups of society will practice preventive checks to maintain rising living standards [Archives, 1981]. Marx defined the value of labor power as consisiting of three components--physiological, skill, and social needs--and argued that reproduction is possible only if the workers are able to satisfy all these needs [Abeysinghe, 1986]. These ideas have been carried in to the present era under the concepts of relative income, relative economic status, socioeconomic status, and aspiration levels. [Easterlin, 1973; Leibenstein, 1975; Easterlin et al. 1980 are a few examples.] A common theme that runs through these exercises may be summarized in the words of Bastiat: "Once natural wants are satisfied, others arise that are artificial at the begining, if you will, but which in their turn become natural through the force of habit, and when they are satisfied, others arise, and still others, with no discernible end." [p.337]

Easterlin et al. focus on two forms of taste formation: the "socialization" effect and the "intrafamily" effect. Thus, both consumption and family size decisions are influenced by present and past consumption habits and family size patterns of other families, including one's own parent family.[69]

A well known formulation of endogenous tastes is Easterlin's relative income hypothesis, which is used to explain both the fertility and LFP decisions of women.[70] An interesting feature of the relative income hypothesis is that it is designed to explain both the secular and cyclical (baby boom) behaviour of fertility.

Relative income, as originally formulated [Easterlin, 1968], refers to the current income of young adults (e.g., age group 15-24) relative to the income of their parents. This measures the ability of young adults to maintain a life

---

[69] Since habit perpetuation cannot be specified in a one-period planning model, Easterlin et al. consider only the average consumption and family size of other families. As a further simplification, they disregard the influence of current consumption and family size of other families and focus on the lag influence i.e., the influence of the previous generation or cohort.

[70] See Wachter [1975, 1977] for further work on this hypothesis.

style to which they have become accustomed prior to their marriage. This is the "intrafamily" version mentioned above. The "socialization" version can easily be incorporated into this specification by allowing (with a lag) the life styles of "important others" to influence desired standard of living of young couples. Thus, relative income is current income divided by desired income; the desired income is determined by a desired standard of living.

According to the relative income hypothesis, the post World War Two baby boom is a result of increased relative income while the subsequent fall in fertility and growth of LFP of married women is a result of falling relative incomes. For example, the entry of baby boomers into the labor market reduced their current wages relative to their desired living standards which they had acquired during the previous prosporous period. As a result fertility fell and LFP of married women increased.

A question raised [T.P. Schultz, 1981] in relation to relative income hypothesis is: Would there be another upturn of fertility in the near future? The argument is that the effect of baby boomers on relative incomes has now dissappeared and the adult children of baby boomers are likely to have lower aspiration levels relative to the wages that they are going to receive, therefore causing an upturn of fertility. This same argument implies a withdrawal of married women from the labor market. This argument is based on the original formulation of the relative income hypothesis, i.e., the intrafamily version. However, if we incorporate the socialization version, for which Leibenstein [1975] has provided a rigorous analysis, then the upturn of fertility depends on how "status expenditure" patterns of different social-status groups change as income grows.[71] For example, if "status goods"[72] become relatively cheap or if "interstatus income compression"[73] ceases, then demand for children may increase [Leibenstein, 1975: 31].

In summary, the two important variables in the endogenous taste model are relative income (current income of young adults relative to their desired level of income which is required to maintain a particular life style) and (lagged) average size of other families, especially that of status groups to which a young couple belong or expect to belong. Both variables are expected to exert a positive influence on fertility.

---

[71] Withdrawal of married women from labor force is a complicated issue. Once set in motion, LFP of married women may develop its own momentum and dynamic quite independent of relative income effects.

[72] Status goods are those consumption goods required to maintain a life style of a particular social-status group.

[73] Interstatus income compression makes emulation and shift to higher social status groups easier.

Chapter V

AN EMPIRICAL TEST OF FERTILITY MODELS

In this chapter we will test, using Canadian data, the causal relationships specified by the two models of fertility discussed in the previous chapter. The theoretical variables entering this analysis are desired fertility, labor force participation rate (LFPR) of married women, full income (see Section 5.1.3), price of children, relative income, and average fertility of the society. Mutual causation of the first two variables is of particular interest; therefore they will be the endogenous variables of the VAR specification. The last four variables are taken to be exogenously determined, as postulated by the theories. Since we are using aggregate (national) data, average fertility of the society will simply be lagged fertility itself. These lag effects are allowed to emerge from the VAR specification.

Empirical tests based on time series data of these fertility models are severely limited by data deficiencies. Firstly, the theoretically most suitable data series either do not exist or are not available for comparable and satisfactory time lengths. Secondly, tests based on proxies may be of questionable quality. Even for proxies, sufficiently long annual data series of the same length are not readily available. Attempts to remedy these deficiencies are presented in the next section. The statistical analysis is reported in Section 5.2.[74]

5.1 SPECIFICATON OF VARIABLES

5.1.1 Fertility

Following Wachter [1975], observed fertility, $F(t)$, may be expressed as:

$$F(t) = D(t) + [S(t) - D(t)]P(t) \tag{5.1}$$

where $D(t)$ = Desired fertility,

$S(t)$ = Supply of fertility or natural fertility

$P(t)$ = Failure rate of birth controls.

It will be assumed, on average, $S(t) \geq D(t)$. If the failure rate is zero ($P(t)=0$), observed fertility is equal to desired fertility. If $P(t)=1$, i.e., no effective fertility control, then observed fertility is equal to natural fertility. For $0<P(t)<1$ there exists what may be termed excess fertility. We assume, in developed countries including Canada, that observed fertility is dominated by desired fertility; therefore, we may ignore the second term on the right hand side of (5.1).[75]

---

[74] Quarterly data series of recent origin would provide a large sample size, but these avoid the baby boom period which provides an important test of any theory of fertility.

[75] In any event data on $S(t)$ and $P(t)$ are very difficult to compile within a time series context. Further, note that empirical studies of demand models which use $F(t)$ as the dependent variable implicitly assume that the second term

Different measures of F(t) are available. The most widely available is the crude birth rate (CBR); that is the number of births in a given year per thousand of mid-year population. This is not a satisfactory measure because it ignores the age-sex composition of a population. A change in the CBR may be a result of changing age-sex composition rather than a change in family size itself. A much better measure is the age-specific fertility rate; that is the number of births in a given year per thousand females of a given child bearing age. A summary measure, the total fertility rate (TFR), is derived by summing the age specific fertility rates. Statistics Canada has published unbroken data series of age-specific fertility rates and the TFR since 1921. We will use the TFR in our analysis. TFR for selected years are given in Table 5.1 and the full series is graphed in Figure 5.1. The hump in the graph shows the postwar baby boom. The decline of fertility after the baby boom appears to have begun in 1959.

5.1.2 Labor Force Participation of Married Women

An annual series of labor force participation rate (LFPR) of married women is available, from published sources, only since 1959. The source of these data is the Labor Force Survey which has been conducted since 1945. The LFPR of married women shows a precipitous growth and the participation rate of any other female age group would be highly inappropriate

---

on the rignt hand side of (5.1) is less important.

as a proxy for the LFPR of married women (Table 5.1). Therefore, it is desirable to estimate LFPR of married women using employment data for the years between 1945 and 1959. Data on female employment by marital status are available from the Labor Force Survey since 1945. Estimates of LFPR of married women were obtained by dividing employment data by the population of married females. (See Table 5.1 for the source of population data.)

For 1951 our estimate of LFPR of married women is 11.0 percent while the actual figure is 11.2 percent. We can, therefore, expect the error margin of our estimates to be small. These estimates show a slight decline in participation rates between 1945 and 1948 (a decline from 12.1 to 10.4; see Figure 5.1). This is consistent with the postwar experience of males returning to work and wives to housework [Phillips and Phillips, 1983:30-31]

Before turning to other data series it may be useful to make some observations on Canadian fertility and LFP of women. The TFR had been declining during a period (till late 1930's) in which LFPR of married women had remained very low. A steady growth of LFPR of married women is observable only since 1950's. During this period fertility goes from a baby boom to a bust. Unlike for married women, the participation of young females in the labor market has always remained relatively high.

TABLE 5.1

Total fertility rate and labor force participation rates of women, Canada, selected years

		Participation Rates		
Year	TFR	married	25 & over	24 & under
1921	3536	2.2	13.9	23.8
1931	3200	2.4	15.8	25.3
1941	2832	3.5	17.8	35.5
1951	3503	11.2	18.4	40.8
1959	3935	17.9	23.0	-
1961	3840	20.7	25.5	39.4
1971	2187	33.0	35.4	50.8
1981	1704	50.6	48.1	63.2
1985	1669	54.7	51.6	64.6

Source: Statistics Canada, Vital Statistics, Cat. 84-202; The participation rate of married women in the Canadian labor force, Cat. 71-522; Historical Labor Force Statistics, Cat. 71-201; Population estimates by marital status, age and sex, Cat. 91-203; Historical Statistics of Canada, 2nd ed.

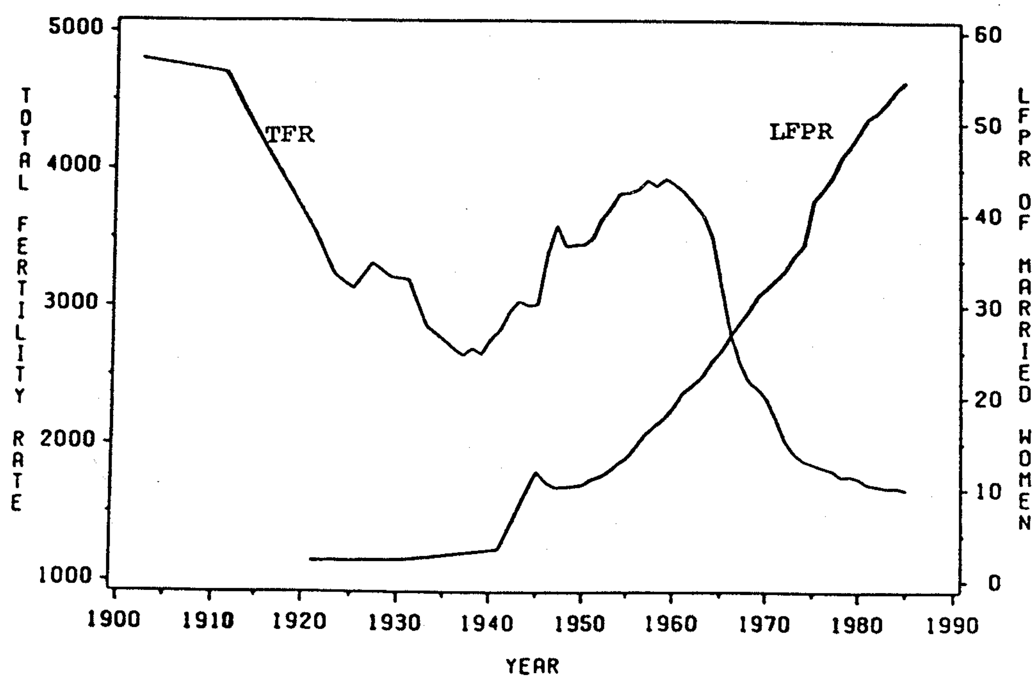

Figure 5.1: Total fertility rate and labor force participation rate of married women

5.1.3 Income

The theoretical measure of family income is "full income," the sum of non-labor income and potential labor income of members of the family, basically husband and wife. The potential labor income is obtained if the full time of family members (e.g., 24 hours a day) is devoted to market work. Full income thus does not depend on actual hours of work. Measured family income, on the other hand, is endogenous to the actual hours worked by family members as well as to family size, the family decision variables; therefore, meausered family income cannot be used in this analysis.

To overcome this difficulty, the income variable is mostly measured by the husband's actual income under the assumption that the labor force participation and hours of work of the husband is largely exogenously determined. The total annual earnings of the husband still may contain endogenously determined components, i.e., he may work extra hours to supplement family income, or his income may contain family related allowances. A widely used measure, therefore, is the male hourly wage rate or weekly or monthly salary which is determined in the market place.

However, an unbroken time series of male wage rates is not available. Historical Statistics of Canada (Leacy, 1983)[76] contains a short series, from 1934 to 1969 with some

[76] Since Historical Statistics of Canada is a well known volume its reference will be exluded from later citations.

missing valuses, of wages of male workers in manufacturing industries, but this does not serve our purpose. Time series of wage data are available by industry as averages for both sexes. These averages are, however, biased downward because of a generally lower average female wage rate. A better proxy for the male wage rate would be the hourly earnings of hourly rated wage-earners in the construction industry. This is so because construction workers have been almost one hundred percent males. Furthermore, the hourly male wage rate of manufacturing mentioned above is very similar to the hourly wage rate of construction industry during the overlapping period of the two series:

Year		1945	1950	1955	1960	1965
Manufac. male wage rate	$	.74	1.14	1.57	1.93	2.33
Construc. wage rate	$	.74	1.06	1.52	1.94	2.53

Construction wage data from 1945 to 1970 can be obtained from Historical Statistics of Canada, and from the Bank of Canada Review. We use this series as a proxy for the male wage rate. This series, extended by manufacturing male wage series for the period 1934-1944 is graphed in Figure 5.2.

The decline in the real wage rate after 1978 (Figure 5.1) is not peculiar to the construction industry. Statistics Canada has recently begun to compile data on male and female annual earnings separately.[77] These data, provided for selected years, show a decline in real average annual earn-

[77] Statistics Canada, <u>Earnings of Men and Women</u>, Cat. 13-217.

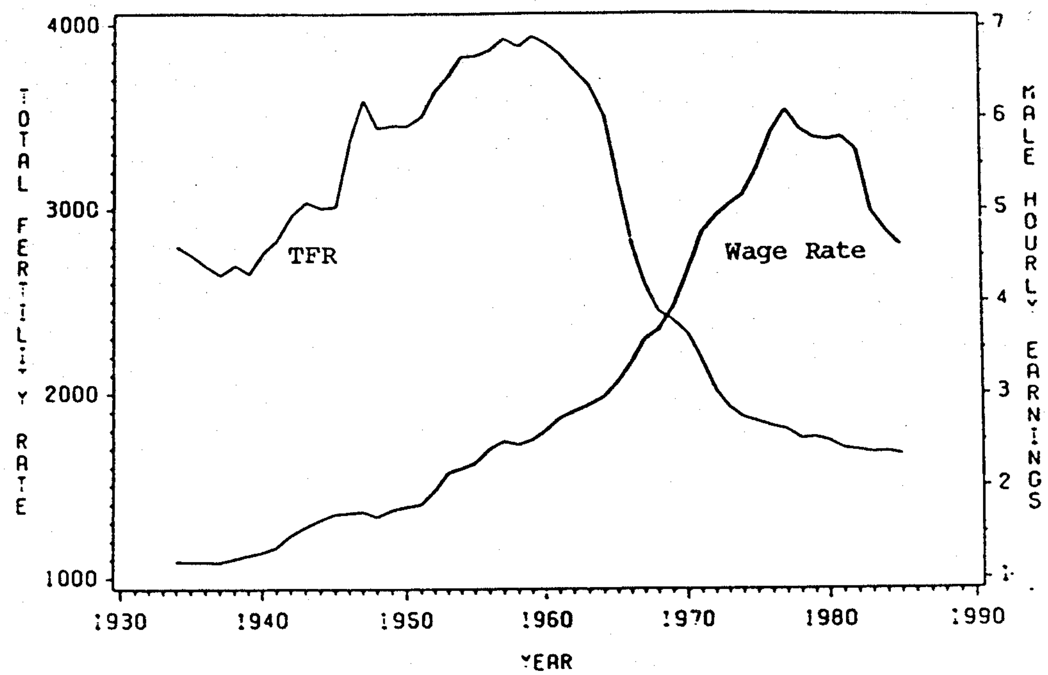

Figure 5.2: TFR, and construction wage rate in 1971 dollars as a proxy for male real wage rate

ings (1984 dollars) of males from $22,700 in 1979 to $20,935 in 1984. The striking feature is that the average earnings of women does not show such a decline though the growth has slowed down or become mildly negative. (Figure 5.3 depicts this behaviour.)

5.1.4 Price of Children

Theoretically, the price of children is the shadow price of a child. This involves an estimation of the real cost of the goods and time input bundle used in bearing and rearing a child. Since such a measure is difficult to compile, empirical studies use the female wage rate as a measure of the price of children (see Section 4.1.3). Obtaining a long time series of female wage rate also poses problems. Female wage data between 1956 and 1975 by selected office occupation of five major cities--Halifax, Montreal, Toronto, Winnipeg, and Vancouver-- are directly available from Historical Statistics of Canada.[78]

Since the female labor force is heavily and increasingly employed in clerical occupations [Phillips and Phillips, 1983:48-49], the average wage rate of female office clerks and typists may be taken to represent the average female wage rate. Of four occupational categories, office clerk (junior and senior) and typist (junior and senior), the average weekly salaries of junior clerks and junior typists were almost the same in all the five cities. These salaries also represented the median of the four groups. Therefore, the average weekly salary of junior office clerks was taken to represent average female wage rate. (The choice between junior clerks and typists was arbitrary.) Furthermore, of five cities, the Montreal salaries were mostly in the median group, therefore Montreal was taken as the representative city.[79] Data after 1975 were taken from the Department of Labor survey mentioned above. The survey was not published in 1978; this missing value was interpolated. For 1985, the

---

[78] The source of these data is Wage rates, Salaries and Hours of Labor published annualy by the federal Department of Labor.

[79] Winnipeg and Halifax stand below the median while Toronto and Vancouver above the median. This pattern, however, seems to be changing in the 1980's.

Canadian average was used. Data prior to 1956 are the average weekly salaries of females in manufacturing industries published in Historical Statistics of Canada. Although the amalgamation of female salaries of manufacturing with that of junior office clerks may appear unusual, these two series are not very different during the overlapping period. Some of these averages for manufacturing industry and Montreal junior clerks are respectively: 1956 ($39, $38), 1958 ($42, $43), and 1960 ($44, $43). The female wage series thus constructed is graphed in Figure 5.3.

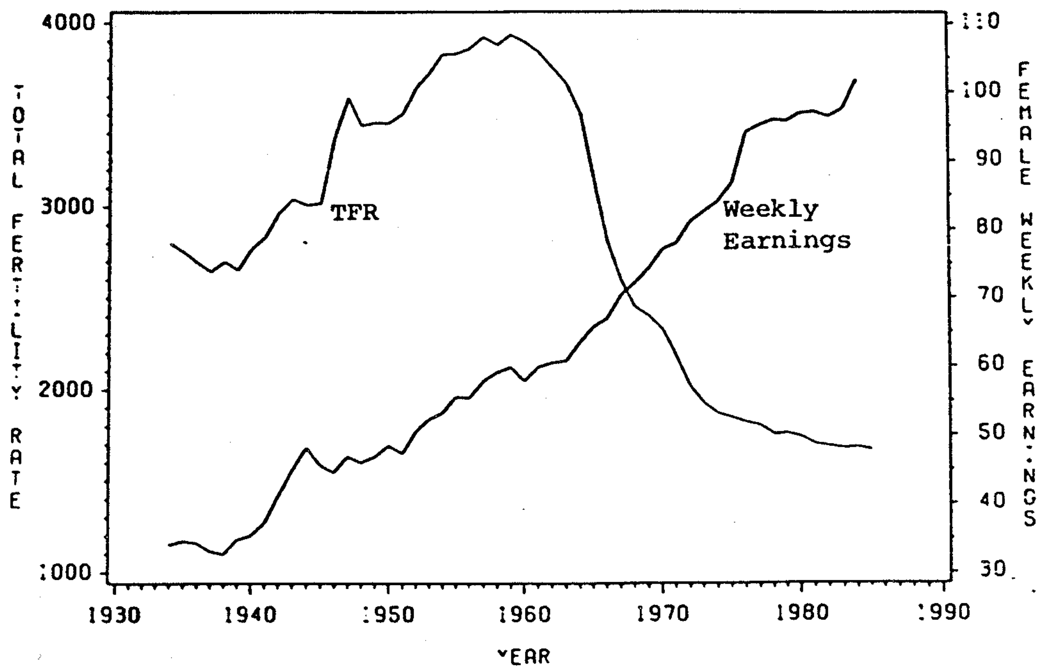

Figure 5.3: TFR, and female weekly salary in 1971 dollars of Montreal city junior office clerks

5.1.5 Relative Income

The last series to construct is relative income. Relative income refers more specifically to the current earnings of young adults relative to their desired income or desired standard of living.[80] Since the desired standard of living, hence the desired income, depends on one's own accustomed way of living as well as the average standard of living of other age groups or socioeconomic groups, construction of a time series of relative income is severely handicapped by data deficiencies. Furthermore, the income measures, in the absence of full income, must not be endogenous to the family decision variables, such as hours of work, family size, and the number of family participants in the labor market.

For illustrative purpose it is possible to construct a few measures of relative income using Consumer Finance Survey data.[81] Annual income of male heads of families relative to the income of heads of age group 45-54 (the highest family income group) are given in Table 5.2.[82] This measure of relative income shows a mild secular decline and a cyclical

---

[80] Desired income here is not an unlimited quantity as one might think. It is a historically and socially determined quantity.

[81] Statistics Canada has conducted Consumer Finance Survey at two to four year time intervals beginning in 1951 and annually since 1971.

[82] Wachter [1977] introduced this measure. He uses family income as well as income of individual males. One difficulty of constructing these measures is the change of age grouping in certain years. We had to combine comparable income groups to derive data since 1951 (see note under Tables 5.2 and 5.3).

TABLE 5.2

Median income of male heads of families, by age group, relative to age group 45-54 (%)

Year*	≤24	25-34	35-44	45-54	55-64	≥65
1951	76.0	83.1	93.0	100	86.1	53.6
1954	78.1	86.4	90.1	100	93.3	56.4
1957	75.1	87.5	89.6	100	85.7	49.9
1959	74.9	89.2	96.5	100	91.1	50.1
1961	66.1	88.5	99.7	100	86.1	39.4
1965	65.2	91.1	99.6	100	79.7	39.4
1967	77.8	88.7	93.5	100	85.5	43.5
1969	73.0	88.6	94.7	100	85.5	41.4
1972	65.1	86.9	94.5	100	83.2	41.1
1974	66.5	84.7	94.2	100	85.2	41.9
1976	64.1	84.0	92.4	100	78.0	37.2
1978	63.8	82.1	92.4	100	81.0	38.7
1980	67.0	84.4	94.1	100	86.5	43.9
1985	56.5	80.4	94.6	100	86.8	-

* For years 1951-1959 family income is used.
Source: Based on Statistics Canada, Income Distributions, Cat. 13-529, 13-521, 13-528, 13-534, 13-544, and 13-207 various issues.

movement.

The above measure of relative income is likely to be endogenous due to the influence of hours of work and family size on income. A fall in relative income may induce more work, which in turn would affect relative income. This effect may be reduced if we consider the relative income of young individual males. Table 5.3 shows the median income of individual males in general (A), and those whose major source of income is wages and salaries (B) relative to the income of age group 35-44 (the highest individual income group). The secular decline in relative incomes is well pronounced in this table. To visualize the cyclical movement the relative income of group A (age group 24 & under) is graphed in Figure 5.4. It is interesting to see how both fertility and relative income follow similar cyclical pattern. Relative income shows a slight increase in mid 1970's, but with more fluctuations, and fertility seems to be levelling off in this period.

TABLE 5.3

Median income of young individual males relative to age group 35-44 (%)

	A: Total Income*		B: Wages & Salaries	
Year	≤24	25-34	20-24	25-34
1951	44.8	90.4	-	-
1954	47.5	96.8	-	-
1957	52.5	99.9	-	-
1959	43.1	90.7	-	-
1961	39.1	89.2	-	88.5
1965	31.0	89.8	-	89.4
1967	30.2	89.2	57.4	90.5
1969	22.7	93.6	49.5	88.9
1972	28.5	89.6	46.4	85.1
1974	31.2	95.0	48.2	83.7
1976	29.9	86.6	46.5	84.0
1978	28.9	82.3	46.5	84.4
1980	27.0	82.0	44.7	83.4
1982	26.9	84.2	42.6	80.1
1985	25.0	78.0	35.2	76.7

* After 1967 income data are for single, divorced and separated males.
Source: See Table 5.2.

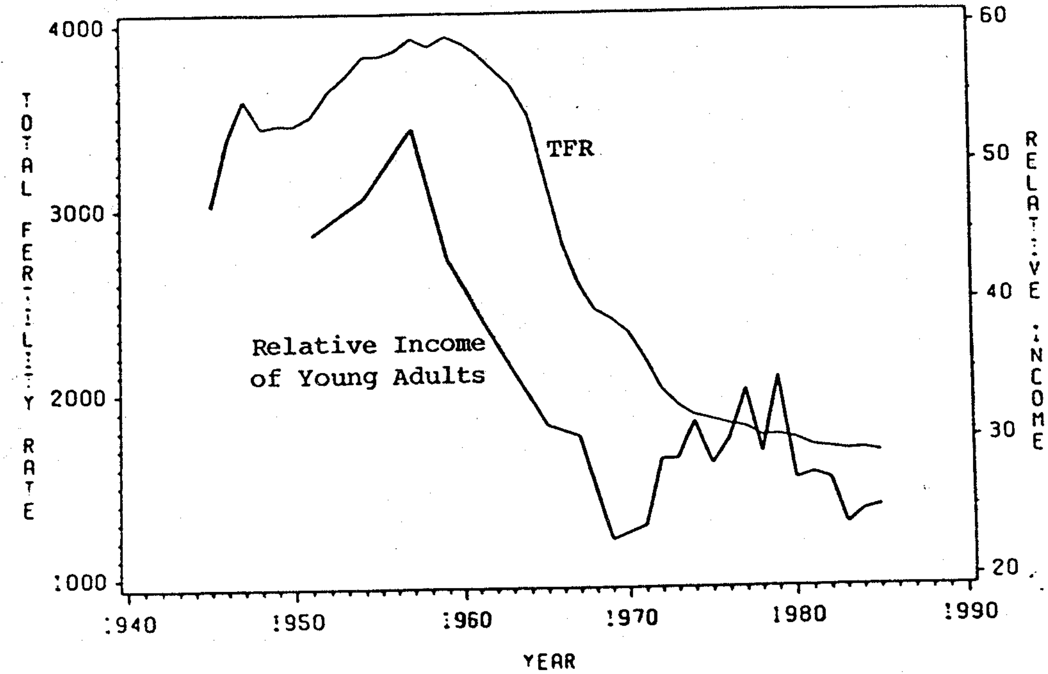

Figure 5.4: TFR, and median income of young individual males (24 and under) relative to age group 35-44

A problem with the above measures of relative income is that long and unbroken time series are not available. Different proxies have been used for reltive income in time series exercises. One is Wachter's [1972] relative wage. Wachter [1975] uses a simple moving average of the past ten years of average real hourly wage rate of U.S manufacturing production workers as a mesure of desired wage rate. Before doing this, he multiplies the wage rate by the rate of employment (i.e. 1 - unemployment rate) to take account of the possibility of having a job. The resulting wage rate is interpreted as an expected wage rate.

Wachter's relative wage rate, however, does not reflect fully the concept of relative income. It captures the effect of past standards of living, but not the contemporaneous effect of the socioeconomic environment in which young families live. Our preliminary investigations showed that the explanatory power of this relative wage is very low.[83] Another proxy suggested by Wachter [1977], which he uses to analyse the labor force participation of women, is the proportion of young adults aged 16-34 in the labor force. Given other things the same, a larger supply of young adults in the labor market would depress their wages relative to other age groups, and therefore increase the labor force participation of women. When applied to fertility we would expect a decline in fertility. However, if wage rates change for reasons other than the proportion of young adults in the labor force, the above proxy will not be appropriate as a measure of relative income.

Because of these difficulties, we propose to measure relative income in a different way. Desired income of young families at time t may be expressed as

$$y^*(t) = y(t) + B(t), \tag{5.1}$$

---

[83] A regression of log of TFR on log of Wachter's relative wage rate yielded (without a correction for autocorrelation) a significant coefficient with expected sign, but an adjusted R^2 of 0.07. The better fit of Wacher's model is due mainly to the presence of lag fertility in his equaton.

where y^* is desired income, y is full income, and B is personal borrowing for consumption which includes housing as well. If current full income is equal to desired income, then B will be zero. This formulation is based on the observation that consumption indebtedness, on average, of young families is much higher than that of older families. According to a recent Family Expenditure Survey [Statitics Canada, 1984], 80 percent of heads of families with no mortgage were at least 45 years old, and 30 percent were aged at least 65. On the other hand, only 38 percent of family heads with a mortgage were older than 44, and only 3 percent were aged 65 or older. Other than mortgage there are borrowings for the purchase of other consumer durables.

Dividing both sides of (5.1) by $y^*(t)$ we get:

$$y(t)/y^*(t) + B(t)/y^*(t) = 1. \tag{5.2}$$

It follows from (5.2) that, if relative income, $y(t)/y^*(t)$, of young adults is falling over time, their borrowings relative to desired income, $B(t)/y^*(t)$, must be rising. However, given that lifetime net borrowings (borrowings-savings) have to be zero, if not negative,[84] young adults are constrained to reduce total consumption, given lifetime full income. If families aim at maintaining per capita consumption unaffected by the above constraint, they can reduce total consump-

[84] In fact, with the decline of the extended family, saving for old age security has become very important. Therefore savings during working age may be viewed as a necessity rather than a surplus.

tion only by reducing family size. The other dimension of this argument is that while families might try to reduce total consumption by reducing family size, they may try to earn additional income by sending other family members (wives in particular) into the labor market.[85]

Some families may avoid debt by economizing on per capita consumption and restricting family size. This behaviour, however, does not contradict the above argument since the income constraint is operating in a similar way. These families, instead of adjusting current income toward desired income by borrowing, adjust desired income toward current income by curtailing consumption.

If the above argument is accepted, then the relative income measure may be written as,

$$y(t)/y^*(t) = y(t)/[y(t)+B(t)]. \tag{5.3}$$

Since full income is difficult to measure and measured income is endogenous to hours of work, we use the following procedure to estimate (5.3). Assuming that the major source of income of young adults is labor income (i.e., ignoring property income) we write the right hand side of (5.3), after dropping t, for the nation as a whole as

$$WH/[WH + B], \tag{5.4}$$

---

[85] Thus measured consumption divided by measured income (average propensity to consume) may remain constant over time, whereas desired consumption relative to full income may not.

where W is the wage rate and H is either the actual or the potential number of hours worked per year. If H is actual, then WH is measured income; if H is potential, then WH is full income. Since measuring potential H[86] is difficult we may write (5.4) with actual H as

$$W/[W+B/H] = W/[W+(B/X)(X/H)], \tag{5.5}$$

where B/H is borrowing per hour worked and X is a population measure. If we could find an appropriate population measure such that X/H is a constant (a) then the endogenous part of (5.5) will be eliminated. Thus (5.5) reduces to

$$RI = W/[W+aB/X], \tag{5.6}$$

where RI stands for relative income. Annual data on actual H for the total work force are available from the Labor Force Survey since 1975. We find that the ratio of population aged 15 and over to H is approximately constant (0.05) between 1975 and 1985. Assuming this constancy holds for the three decades prior to 1975, we use a=0.05 and population aged 15 and over as X in (5.6).

---

[86] Potential H may be estimated as a product of trend of average number of hours worked per week, trend of average number of weeks worked per year, and peak to peak trend of total employment. To be more appropriate, the average hours and weeks must be for male full time workers and total employment must be for young adult labor force. This procedure was suggested by Professor Norman Cameron. However, suitable data for our entire sampling period is difficult to obtain.

According to available data there are two components of B: consumer and mortgage credit flows. Financial Flow Accounts provide data on consumer credit flows after 1961. Although this item is listed under the "persons and unincorporated business" sector, consumer credit is defined as "credit extended to persons for the purchase of consumer durables and other personal consumption." Data on consumer credit outstanding prior to 1961 are available from Bank of Canada Suppliments to Statistical Summary. Flow of credit prior to 1961 was simply taken as the difference of outstanding credit. Residential mortgage credit flows are available as two components: mortgage credit for new residential construction and for exisiting residential property. These data were obtained from the Bank of Canada Review, and Canadian Housing Statistics, and Housing in Canada reports published by Central (later Canada) Mortgage and Housing Corporation (CMHC). Data for 1945 and 1946 were extrapolated. W in (5.6) is hourly wage rate of manufacturing industry multiplied by employment rate (i.e., W is the expected wage rate).[87]

Relatve income measured by (5.6) is graphed in Figure 5.6. The unusual rise in relative income in 1982 (an unusual year in many respects) is the result of a sharp decline in consumer credit flow in that year. Considering this as an

---

[87] Converting these data series to constant dollar terms is not necessary since such a conversion cancel out in (5.6).

unimportant outlier a correction could have been introduced; however, we did not make such a correction.

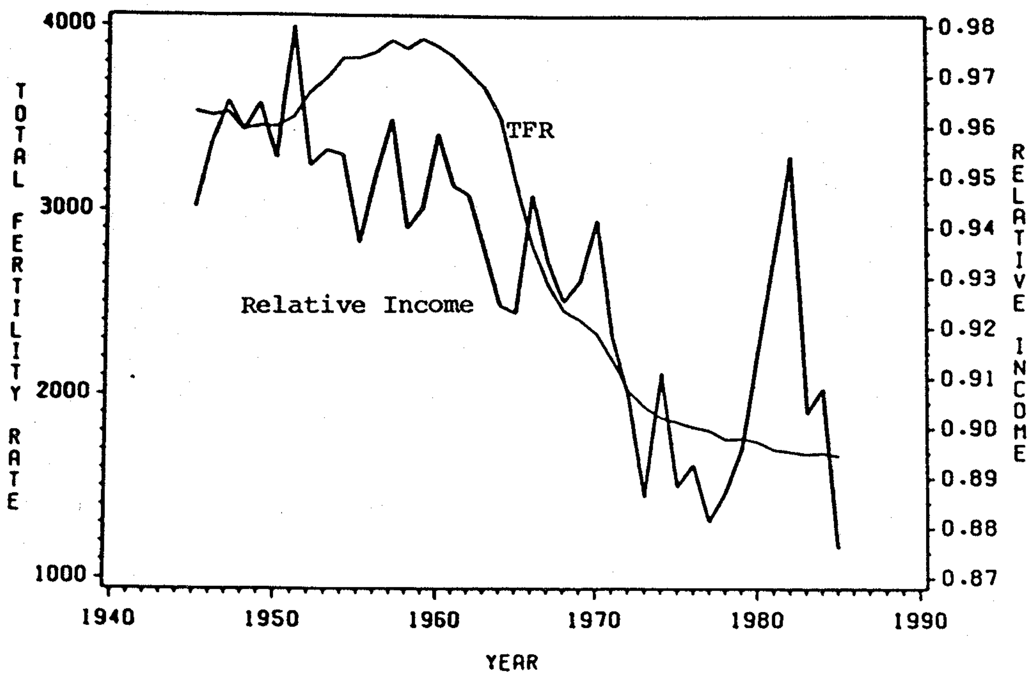

Figure 5.5: TFR, and relative income measured by (5.6)

5.2 STATISTICAL ANALYSIS

The variables used in this analysis will be symbolized as follows. All the variables are in logarithmic form.[88]

F: Total fertility rate

---

[88] Logarithmic transformation is widely used in econometric applications. This, in fact, has a number of advantages. It reduces skewness, which is generally quite high in economic data series. Further, most economic time series are likely to have an increasing trend in both mean and variance as in $X(t) = T(t)Z(t)$, where $Z(t)$ is a stationary series and $T(t)$ is the trend function. The log transformation yields $\log X(t) = \log T(t) + \log Z(t)$, which has a trend in mean only. This can be modelled explicitly or removed by differencing or by detrending.

L: Labor force participation rate of married women

Wm: Male real wage rate (hourly) x employment rate

Wf: Female real wage rate (weekly) x employment rate

RI: Relative income as defined in equation (5.6).

The multiplicative factor, employment rate, is used following other studies [Butz and Ward, 1979; Wachter, 1975]. In short, fertility is expected to depend positively on lagged fertility, negatively on LFPR of married women, positively on male wage rate (assuming the positive income effect outweighs the negative substitution effect), negatively on female wage rate (assuming the negative substitution effect outweighs the positive income effect), and positively on relative income. The LFPR of married women is expected to depend positively on lagged L, negatively on fertility, negatively on male wage rate, positively on female wage rate, and negatively on relative income.

Some commonly encountered single equation regression results related to fertility are reported in Table 5.4. The upper portion of the table does not involve a correction for autocorrelation, but conclusions drawn from this type of exercise are not rare. Butz and Ward [1979], for example, conclude from regressions with DW statistic ranging from .74 to 1.53, that a male income and female wage model can explain both cyclical and secular movements of U.S. fertility. However, if they were to correct for autocorrelation they might observe different results. For example, a correction for

TABLE 5.4

Some standard regression results

Dependent variable: F

Explana. Variable	Regression No. 1	2	3	4
Intercept	2.42	13.66	8.52	13.55
	(2.99)	(9.03)	(98.03)	(8.51)
Wm	-0.88	0.11	-	0.11
	(-2.64)	(0.44)	-	(0.44)
Wf	0.82	-1.41	-	-1.38
	(1.57)	(-3.29)	-	(-3.05)
RI	-	-	8.78	0.35
	-	-	(7.76)	(0.26)
Period	1934-85	1945-85	1945-85	1945-85
DF	48	38	39	37
$\bar{R}^2$	.46	.84	.61	.84
Root MSE	0.225	0.136	0.212	0.137
DW	.05	.15	.61	.15
	After correcting for first order autocorr.			
Intercept	8.13	10.56	8.09	10.37
	(11.50)	(10.86)	(97.61)	(9.59)
Wm	-0.23	-0.22	-	-0.25
	(-1.51)	(-1.29)	-	(-1.35)
Wf	-0.02	-0.59	-	-0.54
	(-0.11)	(-2.18)	-	(-1.76)
RI	-	-	2.86	0.22
	-	-	(3.14)	(0.45)
DF	47	37	38	36
$\bar{R}^2$	.15	.53	.21	.53
Root MSE	0.047	0.049	0.107	0.050

t-ratios are in parentheses.

first order autocorrelation changes our results quite significantly (Table 5.4). Further, as may be seen in this table the explanatory power of the male-female wage model is lower when the sampling period is lengthened to 1934 (regression 1). In other words, when the sampling period is shortened (regression 2) to reduce the impact of the fertility cycle depicted in Figure 5.1 the male-female wage model performs better as measured by R^2. As seen in the lower portion of the table, the income effect is statistically insignificant and not positive. Assessing the impact of the sampling period on the relative income model (regression 3) is not possible since data prior to 1945 are not available. When the two models are combined (regression 4), the coefficients other than that of the intercept become insignificant.

There are a number of weaknesses in the above exercise. As mentioned before the upper portion of Table 5.4 ignores autocorrelation, hence is of doubtful validity. The lower portion of the table is based on the assumption that the autocorrelation is of first order. If it is of a different order, then inferences based on the lower portion of the table are also questionable. This analysis also ignores an important interacting variable, LFP of married women. Further, one has to be cautious when correlating variables with a strong time trend. If the trend in the dependent variable is caused by the trend in the independent variable, then the levels of variables may be used for analysis. However, since we are not certain that we have not left out a variable which may be the cause of a common trend, correlating levels of variables may not be a profitable exercise. Some variables may develop built-in movements; LFP of married women, for example, may become independent of the initial push factors and may develop its own momentum, perhaps due to a change in tastes. The analysis to be carried out next (VAR analysis) is designed to overcome these difficulties.

5.2.1 Vector autoregression Analysis

We use the methodology of VAR modelling given in Chapter 3 (experiment 2) to construct a bivariate autoregression of F and L with Wm, Wf, and RI as exogenous variables. It can be seen from Figures 5.6-5.9 that the first differences of F and L do not necessarily yield stationary series; therefore the analysis was carried out using second differences of all the series.[89]

Lags up to 10 were specified and the resulting AIC, SBC, and PHI values from single equation specifications are given in Tables 5.5 and 5.6. No lags were specified for exogenous variables. Optimal lag specifications which emerged from these two tables are given below.

F equation				L equation			
	AIC	SBC	PHI		AIC	SBC	PHI
Own lag	1	1	1	Own lag	1	1	1
Lag of L	0	0	0	Lag of F	3	0	3

---

[89] We analysed both first differenced and second differenced series and found compatible results in most cases. However, there is a risk of misleading results by a mechanical application of model selection criteria to nonstationary series [Hsiao, 1981].

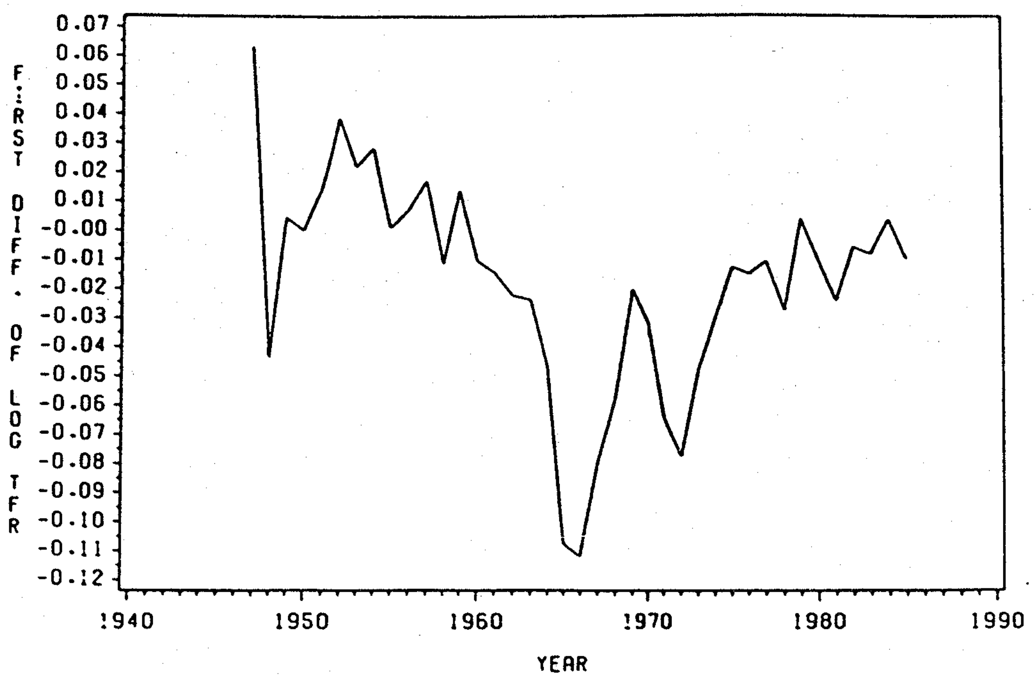

Figure 5.6: First difference of log of total fertility rate

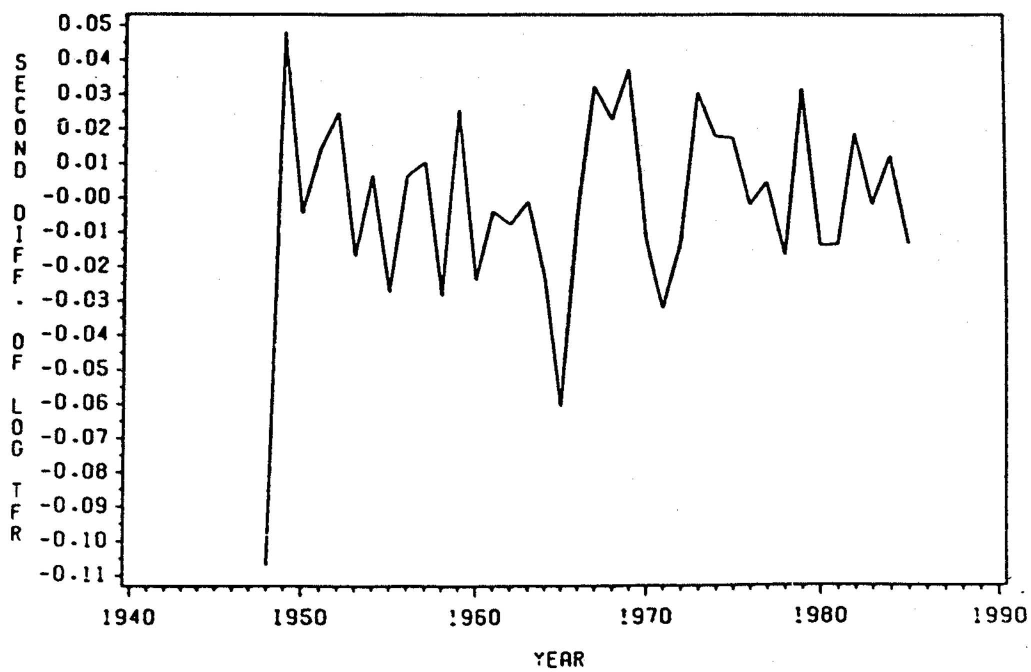

Figure 5.7: Second difference of log of total fertility rate

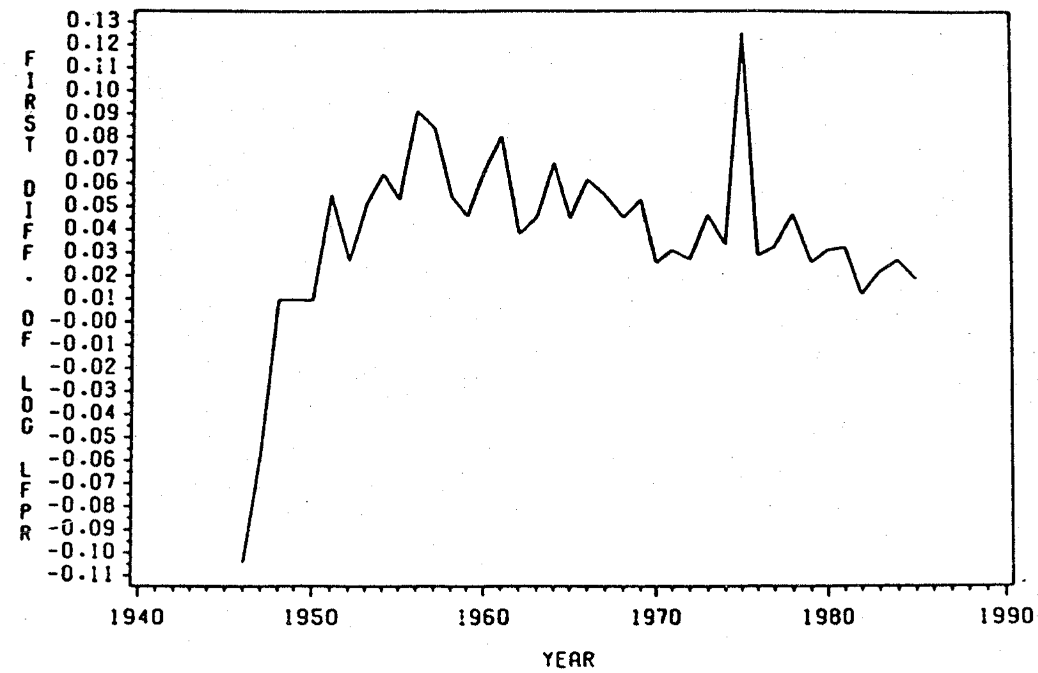

Figure 5.8: First difference of log of LFPR of married women

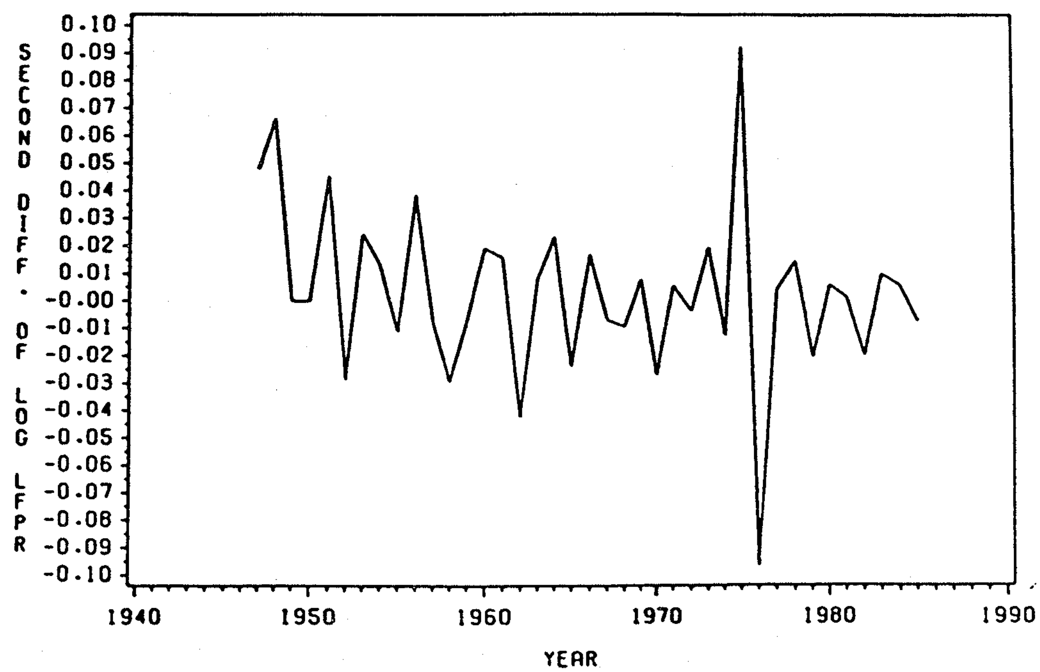

Figure 5.9: Second difference of log of LFPR of married women

TABLE 5.5

Choice of own optimal lag by fitting one dimensional autoregressions to F and L

	F			L		
Own lag+	AIC	SBC	PHI	AIC	SBC	PHI
0	-264.1	-264.1	-264.1	-259.6	-259.6	-259.6
1	-272.6*	-269.2*	-271.4*	-265.9*	-264.2*	-265.3*
2	-270.9	-265.9	-269.1	-264.2	-260.9	-263.0
3	-269.8	-263.2	-267.4	-264.1	-259.1	-262.3

+ Lags 4-10 are not reported for brevity.
* Minimum of 10 lags.

TABLE 5.6

Choice of cross lags given own lag=1 and other exogenous variables in each equation

Lag of	F equation			L equation		
F or L+	AIC	SBC	PHI	AIC	SBC	PHI
0	-269.5*	-261.2*	-266.5*	-267.2	-258.9*	-264.2
1	-267.9	-257.9	-264.3	-267.3	-257.3	-263.7
2	-266.1	-254.4	-261.9	-265.8	-254.1	-261.6
3	-264.4	-251.1	-259.6	-269.9*	-256.6	-265.2*
4	-263.7	-248.7	-258.3	-267.9	-253.0	-262.6

+ Lags 5-10 are not reported for brevity.
* Minimum of 10 lags.

From these optimal lags we obtain two lag length matrices (P*'s), i.e.,

$$P^*1 = \begin{array}{cc} F & L \\ \end{array} \begin{array}{c} \\ \end{array} \quad \begin{array}{c|c} & \begin{array}{cc} F & L \end{array} \\ \end{array}$$

$$P^*1 = \begin{array}{c} \begin{array}{cc} F & L \end{array} \\ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ F \\ L \end{array} \qquad P^*2 = \begin{array}{c} \begin{array}{cc} F & L \end{array} \\ \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ F \\ L \end{array}$$

In addition to the above, we arbitrarily consider the following set of P*'s to recognize the possibility that the model selection criteria may have been too parsimonious in the choice of cross lags in Table 5.6, since our sample size is relatively small (41 years). These P*'s are:

$$P^{*}3 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \qquad P^{*}4 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

$$P^{*}5 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \qquad P^{*}6 = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$$

Joint models corresponding to each of these P*'s were then estimated by incorporating the following combinations of exogenous variables:

Exogenous variables

Set	Wm	Wf	RI
A	x	x	–
B	–	–	x
C	–	x	x
D	x	x	x

where (x) indicates the presence of the variable in the specification and (–) indicates absence. Thus there are 24 joint specifications to consider. AIC, SBC, and PHI values resulting from joint estimation of each of these 24 specifications are given in Table 5.7. The joint estimation was carried out using the seemingly unrelated regression procedure described in Chapter 3.

TABLE 5.7

Model selection criteria from 24 joint specifications

Lag matrix	Exo. set**	AIC	SBC	PHI
P*1	A	-546.13	-536.15	-542.55
	B	-544.98	-536.66*	-541.99
	C	-544.55	-534.57	-540.97
	D	-544.50	-532.86	-540.32
P*2	A	-549.52	-534.55	-544.15
	B	-549.83*	-536.52+	-545.05*
	C	-549.30	-534.33	-543.93
	D	-548.89	-532.25	-542.92
P*3	A	-544.72	-531.41	-539.95
	B	-543.60	-531.95	-539.42
	C	-543.66	-530.35	-538.88
	D	-543.15	-528.18	-537.78
P*4	A	-546.33	-534.68	-542.15
	B	-545.56	-535.58	-541.98
	C	-545.06	-533.42	-540.89
	D	-544.75	-531.44	-539.97
P*5	A	-544.63	-532.98	-540.45
	B	-543.05	-533.07	-539.47
	C	-543.25	-531.61	-539.08
	D	-543.00	-529.69	-538.23
P*6	A	-541.50	-526.53	-536.13
	B	-539.57	-526.26	-534.80
	C	-539.88	-524.91	-534.51
	D	-539.94	-523.33	-534.00

** Exogenous variable set:
A: Wm, Wf; B: RI; C: Wf, RI; D: Wm, Wf, RI.
* Minimum criterion values
+ Second minimum of SBC

The optimal choice of all three criteria (indicated by * in Table 5.7) is the relative income model (set B). However, AIC and PHI choose P*2 indicating a unidirectional G-causality from fertility to LFP of married women, while SBC choses P*1 indicating no G-causality between the two variables. (Note that the difference of SBC values of set B for P*1 and P*2 are neglegible). This difference in the choice of P* by the three criteria reflects SBC's tendency to choose parsimonious models. We observed from both first differenced and second differenced series that the choice between the male-female wage model (set A) and the relative income model (set B) is rather a delicate one. For this reason we estimated both models with both P*1 and P*2 for further examination. Since one lag of second differenced series implies two lags in levels, the joint models were estimated in levels by specifying the lag orders corresponding to P*1 and P*2. One lag of F from F equation was dropped due to insignificant size of coefficients when both lags were present. The results are given in Table 5.8.

Keeping in mind the data deficiencies discussed in Section 5.1, we can draw the following conclusions from this analysis.

<u>Fertility</u>:

1. Elasticities of fertility (F) with respect to female wage rate (price of children) and relative income are statistically significant (using standard levels of

TABLE 5.8

Joint estimates for male-female wage and relative income models with two causal specification bet. F and L

(1) No G-causality between F and L

	Male-female wage model			Relative income model		
Variable	Coeff.	S.E.	Pr>|t|	Coeff.	S.E.	Pr>|t|
F equation						
Intercept	12.953	1.348	.0001	8.456	0.203	.0001
F(t-1)	0.058	0.015	.0004	0.008	0.024	.7360
Wm	0.067	0.224	.7645	-	-	-
Wf	-1.342	0.381	.0012	-	-	-
RI	-	-	-	8.779	1.096	.0001
L equation						
Intercept	-6.178	1.128	.0001	1.454	0.164	.0001
L(t-1)	0.084	0.051	.1086	0.235	0.115	.0485
L(t-2)	-0.076	0.039	.0570	0.119	0.092	.2039
Wm	0.024	0.180	.8956	-	-	-
Wf	2.236	0.326	.0001	-	-	-
RI	-	-	-	-9.129	1.828	.0001
	(2) Unidirectional G-causality from F to L					
F equation						
Intercept	13.042	1.348	.0001	8.143	0.215	.0001
F(t-1)	0.051	0.015	.0016	0.049	0.025	.0639
Wm	0.073	0.224	.7461	-	-	-
Wf	-1.351	0.381	.0011	-	-	-
RI	-	-	-	8.749	1.096	.0001
L equation						
Intercept	4.696	0.845	.0001	2.492	0.044	.0001
L(t-1)	1.529	0.250	.0001	1.533	0.261	.0001
L(t-2)	-0.515	0.233	.0344	-0.662	0.260	.0160
F(t-1)	-0.484	0.079	.0001	-0.486	0.082	.0001
F(t-2)	0.182	0.077	.0245	0.225	0.086	.0131
F(t-3)	0.007	0.006	.2961	0.004	0.007	.5467
F(t-4)	-0.008	0.007	.2221	-0.010	0.008	.2089
F(t-5)	0.010	0.004	.0339	0.011	0.005	.0430
Wm	0.076	0.076	.3254	-	-	-
Wf	-0.594	0.227	.0137	-	-	-
RI	-	-	-	0.646	0.371	.0914

significance such as 1, 5, and 10 percent) and have expected signs. However, relative income elasticities are much larger than female wage elasticities. That is, a 1 percent increase in female wage rate leads to a 1.3 percent decrease in fertility whereas a 1 percent decrease in relative income leads to a 9 percent decline in fertility (see Table 5.8 (1) and (2)).

2. The income effect, as measured by the male wage rate, is positive as expected but statistically insignificant. This is a common finding of regression studies of fertility. This indicates the difficulty of explaining the negative association between fertility and family income by a male-female wage model as posited by the Chicago-Columbia theory of fertility. The strength of the relative income model lies in its focus on a wider range of effects, instead of a narrower focus on the price of children, which counteracts the effect of rising incomes on fertility. This can be expected theoretically [see Leibenstein, 1975] even if one does not agree with our measure of relative income. This strength is further enhanced, in a statistical sense, by its ability to capture all these effects in a single variable. This is well reflected in the choice made by model selection criteria in our exercise. While parsimony is an important aspect of model selection criteria, they do not necessirily choose parsimonious models if a less par-

simonious model has greater predictive power. This is evident from AIC's and PHI's choice of the relative income model of Table 5.8(2) despite the presence of more parameters.

3. Lag fertility, which takes account of the average level of fertility of society in the Pennsylvania model, appears to exert a significant positive influence on current fertility (except in the fifth column of Table 5.8(1)). However, this effect is very small; see the coefficients of F(t-1) of F equation.

<u>LFP of married women</u>

4. If it is assumed that there is no direct causal relationship between F and L (Table 5.8(1)), then female wage rate and relative income elasticities have expected signs and are statistically significant. As in the case of fertility, relative income elasticity is much larger, in absolute value, than is female wage elasticity. Male income elasticity has a postive sign, but it is highly insignificant. Two lags of L together exert a positive and significant influence on current L indicating the built-in movements.

5. Unidirectional causality from F to L, detected by model selection criteria (Table 5.8(2)), indicates that current L depends on lagged fertility up to five years. It is interesting that this coincides with the pre-school age of children. The sum of these lagged

fertility coefficients is approximately -0.3, which indicates that a 1 percent decrease in fertility leads to 0.3 percent increase in LFPR of married women.

6. Introduction of lagged F into L equation reversed the signs of female wage rate and relative income, a result unacceptable theoretically, and has reduced the absolute size of these elasticities quite significantly. It is difficult to explain the incorrect sign except by arguing that it may be a statistical artifact. However, there is an explanation for the reduction of the size of the elasticies.

Unidirectional causality from F to L appears to reflect short run behaviour. Regardless of the size of the female wage rate or the level of relative income, many mothers may withdraw from the labor market for about five years to raise children. In the long run, we could expect these women to return to the labor market.[90] Thus in the long run there may not be a causal relation between F and L. As a crude test of this we estimated L equation using data five years apart. (Such a test is crude because there remain only eight observations.) Then we introduced two lags

---

[90] Female labor force participation by age has a saddle shaped pattern, very high in young ages, low in child bearing ages, and high in late ages. This saddle is, however, slowly disappearing [Phillips and Phillips, 1983:36-38].

of F into L equation and observed that female wage rate and relative income elasticities retain the expected signs and remain statistically significant. Thus the models in Table 5.8(1) may be considered as an approximation to long run behaviour.

Another interesting aspect to consider is that the short run influence of F on L may be weakening over time. The LFPR of mothers with pre-school age children show a dramatic increase in recent years. For example, LFPR of mothers with children aged 0-5 has increased from 37 percent in 1978 to 52 percent in 1984.[91] To shed more light on this, we divided our sample period into two equal parts (1946-1965 and 1966 to 1985) and jointly estimated F and L equations as specified in Table 5.8(2). From this exercise we observed that for the period 1946-1965 the sum of the coefficients of five F lags was approximately -0.3 while female wage and relative income elasticities were insignificant and had incorrect (theoretically) signs. For the period 1966-1985, the sum of the five lag coefficients for the two models were -0.17 and -0.10, a noticable reduction, in absolute value, from the pre-1965 level. Further, the female wage and relative income elasticities had the expected signs, but were still insignificant.

---

[91] Statistics Canada, Family Characteristics and Labor Force Activity: Annual averages 1977-1984, Cat. 71-533.

5.3 SUMMARY AND CONCLUSIONS

It is difficult to obtain or compile data of reasonable length for time series analysis of fertility models. An attempt was made to overcome these difficulties by using proxies and the best possible estimates of relevant data series. Annual data since 1945 were used for analysis. The fertility measure is the total fertility rate. Income and price of children variables of the Chicago-Columbia model were approximated, as is usually done, by male and female wage rates. The male wage rate was proxied by the hourly wage rate of construction workers. Relative income of the Pennsylvania model was estimated by incorporating borrowings for consumption into wage income. This is a novel feature added to this analysis.

Vector autoregression analysis was carried out using second differences of logarithms of the data series. The total fertility rate (F) and the labor force participation rate of married women (L) were the endogenous variables whose mutual causation was to be understood, while male and female wages and relative income were the exogenous variables. After choosing an optimal joint specification for F and L using AIC, SBC, and PHI model selection criteria, the joint models were reestimated using the levels of the variables.

The analysis revealed a number of interesting results which are presented in the previous section. From these results the following conclusions emerge. In the short run there is a causal relation between fertility and labor force participation of married women. This relation appears to be unidirectional from F to L, and not from L to F. This short run relationship is weakening over time. Therefore, a changing parameter model would be appropriate for short run specifications. In the long run, there appears to be no causal relation between F and L. Although it is possible that declining family size until the 1940's paved the way for wives into the labor market, such a long lead lag relation seems to have dissappeared after the 1940's. Thus the long run behaviour of F and L in the present era are likely to be governed by such factors as female wage rates and the relative income. According to our model selection criteria, relative income model explains the data better in terms of predictive variance. The long run behaviour of F and L may be approximated, in the absence of a long series, by a joint model similar to that of Table 5.8(1).

Chapter VI

SUMMARY AND CONCLUSIONS

Vector autoregression (VAR) analysis is an important addition to the tool kit of econometrics. Although in a typical VAR model every variable is taken to be endogenous, depending on the type of application intended, a VAR model can be specified by including both endogenous and exogenous variables as in a dynamic simultaneous equation system. This study demonstrates how VAR analysis could be used to formulate an encompassing framework for evaluating contradictory causal implications of competing economic thories. The concept of causality incorporated herein is that of Granger's.

One limitation of VAR modelling is the difficulty of specifying lag orders of the variables entering a model. Since economic theory is of little help in this respect, one has to rely on statistical procedures to accomplish this task. The practice of assuming a common lag order for all the variables and determining this arbitrarily or through a model selection criterion, though simple, does not yield satisfactory results. By elaborating a procedure proposed and applied by Hsiao, and by incorporating the features of similar work of Abeysinghe, this study developes a procedure for VAR specification which serves a dual purpose: specification of VAR models with variable lag lengths, and then to test causal inferences of competing economic theories based on this specification.

The specification procedure follows the three stages of Box-Jenkins modelling, i.e, tentative specification, estimation, and diagnostic checking. These three stges are not mutually exclusive. The entire specification is carried out using model selection criteria. At the tentative specification stage, each equation of a VAR model is specified separately by admitting variables to each equation sequentially. The chosen lag orders are the ones which minimize a chosen model selection criterion. The sequential order of the variables is determined on the basis of a "minimum of minima" principle. Only ordinary least squares estimation is required at this stage. After identifying a set of lag length matrices for the system as a whole, the system estimation and diagnostic checking (final specification) are carried out using seemingly unrelated regression procedures and system model selection criteria.

In order to aid the choice of a particular model selection criterion, an anlytical as well as an experimental evaluation of a number of model selection criteria was carried out.[92] Based on these results, three model selection criteria--AIC, SBC, and PHI--were chosen for model specifi-

---

[92] A detailed listing of results and conclusions of these exercises are presented in Chapter 3.

cation. Use of these three criteria together aids diagnostic checking and is likely to reduce the possibility of misleading results. So far as causal inferences (based on Granger's definition) are concerned, AIC appeares to perform better. Thus Hsiao's choice of FPE criterion, which is equivalent to AIC in large samples, can be maintained. However, AIC tends to profligate parameterization.

The above procedure is then used to test, using Canadian data, two economic models of fertility, the Chicago-Columbia model and the Pennsylvania model, both of which attempt to explain the observed negative association between income and fertility. The Chicago-Columbia model assumes invariant tastes and explains falling fertility as the result of the rising price of children. The price effect is mostly measured by the female wage rate and the income effect by the male wage rate. The Pennsylvania model considers taste an endogenous variable, and explains fertility movements by relative incomes and society's average level of fertility. Although the labor force paricipation (LFP) of married women does not appear among these variables, empirical studies abound which either take fertility as a determinant of LFP of married women or LFP of married women as a determinant of fertility, in addition to the above variables specified by the two models.

Statistical analysis carried out in Chapter 5 indicates that the causal relationship between fertility and LFP of married women is unidirectional, from fertility to LFP of married women. Further, this relationship appears to exist primarily in the short run--i.e., the presence of pre-school age children deter mothers from work temporarily--and is weakening over time. In the long run, both fertility and LFP of married women seem to be jointly determined by such exogenous factors as the female wage rate and relative income. However, joint specification of fertility and LFP of married women by taking account of their lag structures is likely to yield better results compared to single specifications. Optimal choice made by model selection criteria favors the relative income model, and relative income elasticities are found to be much larger, in absolute value, than the female wage elasticities.

There is further work to be done to improve the VAR specification procedure. If the lag order chosen by a model selection criterion is five, then our procedure retains all five lags, assuming that the influence of lags diminish as lag order increases. However, it is possible that only the fifth lag is relevant while the first four are irrelevant. In our empirical exercise we used t-ratios to aid this decision. However, a better procedure should be developed for determining the relevence of intermediate lags. Unquestionably, further research is required to compile better and longer data series for more definitive empirical analysis.

Appendix A

BIAS-VARIANCE TRADE-OFF

Given a set of possible regressors the concern for choosing a particular set (a parsimonious set) into the regression arises out of the possible bias-variance trade-off. Since these results are already well established a summary statement will suffice here (see Hocking, 1976; Judge et.al., 1980, ch. 11).

Consider the standard normal linear regression model

$$y = X\beta + e = X_1\beta_1 + X_2\beta_2 + e, \qquad (A.1)$$

where y is a (Nx1) vector of observations on the dependant variable, X is a (Nxk) matrix of observations on k explanatory variables partitioned as $X=(X_1,X_2)$ with $k_1+k_2=k$, β is (kx1) vector of parameters partitioned conformably as $\beta=(\beta_1,\beta_2)'$, and e is (Nx1) vector of normal random errors with a zero mean vector and a covariance matrix $\sigma^2 I$.

Let $b=(b_1,b_2)'$ be the least squares (maximum likelihood) estimate of $\beta=(\beta_1,\beta_2)'$ and $\hat{\sigma}^2$ be the unbiased estimate of σ^2. Further, let $\breve{b}_1$ be the subset least squares estimate of β_1 when X_2 is excluded from model (A.1) and let $\breve{\sigma}^2$ be the corresponding estimate of σ^2. Then we can derive,

$$b = (b_1, b_2)' = (X'X)^{-1}X'y$$

$$= \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}^{-1} \begin{bmatrix} X_1'y \\ X_2'y \end{bmatrix}$$

$$\hat{\sigma}^2 = y'My/(N-k)$$

$$\tilde{b}_1 = (X_1'X_1)^{-1}(X_1'y)$$

$$\tilde{\sigma}^2 = y'M_1y/(N-k_1),$$

where $M = I-X(X'X)^{-1}X'$ and $M_1 = I-X_1(X_1'X_1)^{-1}X_1'$.

If model (A.1) is true, it is well known that b and $\hat{\sigma}^2$ are BLUE. Thus b $N(\beta,(X'X)^{-1}\sigma^2)$, and $(N-k)\hat{\sigma}^2/\sigma^2$ is distributed as a central chi-square with N-k degrees of freedom.

As for the subset estimators we can show that $\tilde{b}_1$ is normally distributed with

$$E(\tilde{b}_1) = \beta_1 + A\beta_2,$$

where $A = (X_1'X_1)^{-1}X_1'X_2$, and

$$Var(\tilde{b}_1) = (X_1'X_1)^{-1}\sigma^2. \quad \text{Further,}$$

$$MSE(\tilde{b}_1) = (X_1'X_1)^{-1}\sigma^2+A\beta_2\beta_2'A'.$$

The random variable $(N-k_1)\tilde{\sigma}^2/\sigma^2$ is distributed as a non-central chi-square with

$$E(\tilde{\sigma}^2) = [\sigma^2+\beta_2'X_2'M_1X_2\beta_2]/(N-k_1).$$

Thus the following results can be established.

1. The subset estimator $\tilde{b}_1$ is biased unless $\beta_2=0$ or X_1 and X_2 are orthogonal.
2. The matrix $Var(b_1)-Var(\tilde{b}_1)$ is positive semi-definite whether or not $\beta_2=0$. That is, the LS estimate of β_1 when both X_1 and X_2 are included has precision less than or equal to the LS estimate of β_1 when X_2 is excluded.
3. Whether the gain in precision is offset or not by the bias, i.e. whether $Var(b_1)-MSE(\tilde{b}_1)$ is negative or positive definite, depends on whether the matrix $Var(b_2)-\beta_2\beta_2'$ is negative or positive definite.
4. Unless $\beta_2=0$ the subset estimator $\tilde{\sigma}^2$ is biased upward.

Appendix B

A SAMPLE OF SAS PROGRAMS USED IN SIMULATION

B.1 CREATING DATA SET

The following program generates normal random numbers with zero mean and unit variance by calling the subroutine RANNOR. The variables X and Y are then generated according to the parameter specification of the structure 2. The data set is then retained for repeated use in experiments with sample sizes 50 and 100.

```
// JOB ',,I=40,T=50'
// EXEC SASV5,SIZE=3840K,OPTIONS='S=80'
//OUT DD DSN=ABEYSIN.OUT.BASE,DISP=(NEW,CATLG),
//    SPACE=(TRK,(50,5),RLSE),
//    VOL=SER=WORK04,UNIT=DISK
*-----------------------------------------------*
|                                               |
|        GENERATING DATA FOR STRUCTURE 2        |
|                                               |
*-----------------------------------------------*;
DATA OUT.BASE;
   SEED1=123247;SEED2=125549;
   E=0;N=0;X=0;Y=0; /*INITIALIZATION*/
   SUB1=0;
   SUB2=0;
   DO I=1 TO 10099;
      Y0=1;
      X1=X;Y1=Y;    /* LAGGED VARIABLES */
      CALL RANNOR(SEED1,E);
      CALL RANNOR(SEED2,N);
      X = .5*X1 +E;
      Y = .1*X + .5*Y1 + .1*X1 + N;
      Y2=LAG2(Y);Y3=LAG3(Y);Y4=LAG4(Y);Y5=LAG5(Y);Y6=LAG6(Y);
      Y7=LAG7(Y);Y8=LAG8(Y);Y9=LAG9(Y);Y10=LAG10(Y);
      X2=LAG2(X);X3=LAG3(X);X4=LAG4(X);X5=LAG5(X);X6=LAG6(X);
      X7=LAG7(X);X8=LAG8(X);X9=LAG9(X);X10=LAG10(X);
      DROP SEED1 SEED2 E N;
      IF MOD(I,50)=0 THEN SUB1=SUB1+1; /*REPLICATION FOR N=50 */
      IF MOD(I,100)=0 THEN SUB2=SUB2+1; /*REPLICATION FOR N=100*/
```

```
      IF I≥100 THEN OUTPUT OUT.BASE;
   END;
```

B.2 GENERATING FREQUENCY DISTRIBUTIONS

The following is a portion of the program used for generating frequency distributions of the chosen model selection criteria under structure 2 with sample size 50. Procedure RSQUARE plus a DATA step calculate the chosen model selection criteria. Procedures RANK and FREQ generate the frequency distributions. Those frequencies under rank 1 correspond to the minimum values of the model selection criteria.

```
// JOB
// EXEC SASV5,SIZE=1280K,OPTIONS='NOGRAPHICS NOSOURCE'
//OUT DD DSN=ABEYSIN.OUT.BASE,DISP=SHR
*-------------------------------------------------*
|                                                 |
| FITTING MODELS AND FORMING FREQ. DISTRIBUTIONS  |
|                                                 |
*-------------------------------------------------*;
TITLE 'STRUCTURE 2 WITH N=50';
DATA BASE1;
   SET OUT.BASE;
   IF I<5100;
PROC RSQUARE DATA=BASE1 NOINT Σ=1.005 NOPRINT OUTEST=EST
   SSE MSE CP SP JP AIC BIC SBC;
   MODEL Y=Y1;
   MODEL Y=Y1 X1/INCLUDE=2 STOP=3;
   MODEL Y=Y1 X1 X2/INCLUDE=3 STOP=4;
   MODEL Y=Y1 X1-X3/INCLUDE=4 STOP=5;
   MODEL Y=Y1 X1-X4/INCLUDE=5 STOP=6;
   MODEL Y=Y1 X1-X5/INCLUDE=6 STOP=7;
   BY SUB1;
* ADDING HANNAN PHI AND PARZEN CAT;
DATA EST1;
   SET EST;
   BY SUB1;
   N=50;
   DO J=1 TO 100;
      IF SUB1=J+1 THEN DO;
         PHI=N*LOG(_SSE_/N)+2*_P_*LOG(LOG(N));
         SUM+1/_MSE_;
         CAT=SUM/N-1/_MSE_;
         KEEP SUB _P_ _CP_ _SP_ _JP_ _AIC_
              _BIC_ _SBC_ PHI CAT;
```

```
      END;
   END;
PROC RANK DATA=EST1 OUT=RKND1;
   VAR _CP_ _SP_ _JP_ _AIC_ _BIC_ _SBC_ PHI CAT;
   BY SUB1;
PROC FREQ DATA=RKND1;
   TABLES _P_*(_CP_ _SP_ _JP_ _AIC_ _BIC_ _SBC_ PHI CAT)/
      NOPERCENT NOROW NOCOL;
   TITLE2 'AR Y1 REG X1-X5 MODELS';
```

B.3 SYSTEM ESTIMATION

The following piece of program may be used for system estimation. It is written for the true model given in Section 3.5.4. It uses SAS procedure SYSLIN SUR to estimate the full model using seemingly unrelated regression (SUR) procedure. Then SIMLIN is used to generate the residuals form SUR estimates and CORR is used to obtain sum of squares and cross products (SSCP) matrix from residuals. A DATA step plus the MTRIX procedure are used to calculate the covariance matrix of residuals, AC, SBC, and PHI values.

```
*-------------------------------------------*
|             SYSTEM ESTIMATION             |
|  SEEMINGLY UNRELATED REGRESSION PROCEDURE |
|       AIC SBC AND PHI FOR THE SYSTEM      |
*-------------------------------------------*;
TITLE 'AR SYSTEM ESTIMATION N=50';
DATA BASE3;
   .
   .
*                 ;
PROC SYSLIN SUR DATA=BASE3 NODFS OUTEST=EST1; /* NOPRINT */
   X: MODEL X=X1-X2 Y1 Z1/NOINT;
   Y: MODEL Y=Y1-Y2 Z1-Z2/NOINT;
   Z: MODEL Z=Z1-Z2/NOINT;
PROC PRINT DATA=EST1;
TITLE2 'LAG LENGTH MATRIX P*4 TRUE MODEL';
*                 ;
PROC SIMLIN EST=EST1 DATA=BASE3 TYPE='SUR' NORED;
   ENDOGENOUS X Y Z;
   EXOGENOUS X1-X2 Y1-Y2 Z1-Z2;
   OUTPUT OUT=RES1 R=RX RY RZ;
*PROC PRINT DATA=RES1;
```

```
*   VAR X Y Z RX RY RZ;
PROC CORR DATA=RES1 PEARSON NOPRINT NOCORR
          VARDEF=N SSCP OUT=SSCP1(TYPE=SSCP);
   VAR RX RY RZ;
*PROC PRINT DATA=SSCP1;
*  TITLE2 'SUR RESIDUAL SSCP MATRIX';
DATA SSCPA;
   SET SSCP1;
   IF _TYPE_='SSCP';
   KEEP RX RY RZ;
*PROC PRINT DATA=SSCPA;
*   TITLE2 'SUR RESIDUAL SSCP MATRIX';
PROC MATRIX PRINT;
   FETCH SSCP DATA=SSCPA; /* INPUT SSCPA TO MATRIX SSCP */
   N=50; /* SAMPLE SIZE */
   P=10; /* NUMBER OF AR PARAMETERS */
   COV=SSCP#/N; /* ML ESTIMATE OF COV MATRIX */
   D1=DET(COV); /* DETERMINANT */
   AIC=N*LOG(D1)+2*P;
   SBC=N*LOG(D1)+LOG(N)*P;
   PHI=N*LOG(D1)+2*LOG(LOG(N))*P;
TITLE3 'SSCP COV DETERMINANT AIC SBC AND PHI';
```

BIBLIOGRAPHY

CAUSALITY AND TIME SERIES ECONOMETRICS

Abeysinghe, T. (1982) "Feedback effect of money supply on monetary base: causality analysis using Sri Lankan data," Unpublished M.A. thesis, Faculty of Economics, Thammasat University, Bangkok.

Akaike, H. (1969) "Fitting autoregressive models for prediction," Annals of the Institute of Statistical Mathematics 21, 243-47.

Akaike, H. (1970) "Statistical predictor identification," Annals of the Institute of Statistical Mathematics 22, 203-17.

Akaike, H. (1971) "Autoregressive model fitting for control," Annals of the Institute of Statistical Mathematics 23, 163-80.

Akaike, H. (1974) "A new look at the statistical model identification," IEEE Transactions on Automatic control AC-19, 716-22.

Akaike, H. (1976) "Canonical correlation analysis of time series and the use of an information criterion," in Mehra, R.K. and Lainiotis, D.G. eds., System Identification: Advances and Case Studies (New York: Academic Press).

Akaike, H. (1978) "A Bayesian analysis of the minimum AIC procedure," Annals of the Institute of Statistical Mathematics 30A, 9-14.

Akaike, H. (1979) "A Bayesian extension of the minimum AIC procedure of autoregressive model fitting," Biometrica 66, 237-42.

Akaike, H. (1981) "Likelihood of a model and information criteria," Journal of Econometrics 16, 3-20.

Amemiya, T. (1980) "Selection of regressors," International Economic Review 21, 331-54.

Ashley, R.; Granger, C.W.J.; and Schmalensee, R. (1980) "Advertizing and aggregate consumption: An analysis of causality," Econometrica 48, 1149-167.

Atkinson, A.C. (1981) "Likelihood ratios, posterior odds and information criteria," Journal of Econometrics 16, 15-20.

Battaglia, F. and Carlucci, F. (1982) "A simulation study of causality tests," in Anderson, O.D. ed., Time Series Analysis: Theory and Practice 1 (New York: North-Holland) 655-99.

Bessler, D.A. and Kling, J.L. (1984) "A note on tests of Granger Causality," Applied Economics 16, 335-42.

Bhansali, R.J. and Downham, D.Y. (1977) "Some properties of the order of an autoregressive model selected by a generalization of Akaike's FPE criterion," Biometrica 64, 547-51.

Blaug, M. (1980) The Methodology of Economics (New York: Cambridge University Press).

Box, G.E.P and Jenkins, G.M. (1977) Time Series Analysis: Forecasting and Control 2nd ed. (San Fransisco: Holden-Day).

Caines, P.E. and Chan, C.W. (1975) "Feedback between stationary stochastic processes," IEEE Transactions on Automatic Control AC-20, 498-508.

Caines, P.E.; Keng, C.W.K.; and Sethi, S. (1981) "Causality analysis and multivariate autoregressive modelling with an application to supermarket sales analysis," Journal of Economic Dynamic and Control 3, 267-98.

Chatfield, C. and Prothero, D.L. (1973) "Box-Jenkins seasonal forecasting: Problems in a case study," Journal of the Royal Statistical Society A136, 295-336.

Chow, G.C. (1981) "A comparison of the information and posterior probability criteria for model selection," Journal of Econometrics 16, 21-23.

Conway, R. et.al. (1984) "The impossibility of causality testing," Agricultural Economic Research 36, 1-19.

Cooley, T.F. and LeRoy, S.F. (1985) "Atheoretical macroeconometrics," Journal of Monetary Economics 16, 283-308.

Damsleth, E. (1982) "Estimating for moving average models: why does it fail?," in Anderson, O.D. ed., Time Series Analysis: Theory and Practice 2 (Amsterdam: North-Holland) 5-6.

Davies, N.; Triggs, C.M.; and Newbold, P. (1977) "Significance levels of the Box-Pierce Portmanteau statistic in finite samples," Biometrica 64, 517-22.

Eichenbaum, M. (1985) "A comment on 'Vector autoregression for causal inference' by Leamer," in Brunner, K. and Metzler, A.H. eds., Understanding Monetary Regiems Carnegie Rochester Conference Series on Public Policy 22 (Amsterdam: North-Holland) 305-16.

Engle, R.F. (1984) "Wald, Likelihood ratio, and Lagrange Multiplier tests in Econometrics," in Griliches, Z. and Intriligator, M. eds., Handbook of Econometrics Vol. 2 (Amsterdam: North-Holland) 776-826.

Engle, R.F.; Hendry, D.F.; and Richard, S.F. (1983) "Exogeneity," Econometrica 51, 277-304.

Evans, G.B.A. and Savin, N.E. (1982) "Conflict among testing procedures in a linear regression model with lagged dependent variables," in Hilderbrand, W. ed., Advances in Econometrics (New York: Cambridge Univ. Press) 263-83.

Evans, L. and Wells, G. (1983) "Pierce and Haugh on characterization of causality: a reexamination," Journal of Econometrics 22, 331-35.

Feigl, H. (1953) "Notes on causality," in Feigl, H. and Brodbeck, M. eds., Readings in the Philosophy of Science (New York: Appleton-Century-Crofts Inc.).

Fiege, E.L. and Pearce, D.K. (1979) "The causal relationship between money and income: some caveates for time series analysis," Review of Economics and Statistics 61, 521-33.

Geweke, J. (1978) "Testing the exogeneity specification in the complete dynamic simultaneous equation model," Journal of Econometrics 7, 163-86.

Geweke, J. (1981a) "A comparison of tests of the independence of two covariance stationary time series," Journal of the American Statistical Association 76, 363-73.

Geweke, J. (1981b) "The approximate slopes of econometric tests," Econometrica 49, 1427-442.

Geweke, J. (1982a) "Causality, exogeneity, and inference," in Hiderbrand, W. ed., Advances in Econometrics (New York: Cambridge Univ. Press) 209-235.

Geweke, J. (1982b) "Measurement of linear dependence and feedback between multiple time series," Journal of the American Statistical Association 77, 304-13. (with discussions)

Geweke, J. (1984) "Inference and causality in economic time series models," in Griliches, Z. and Intriligator, M. eds., Handbook of Econometrics Vol 2 (Amsterdam: North-Holland) 1101-144.

Geweke, J. and Meese, R. (1981) "Estimating regression models of finite but unknown order," International Economic Review 22, 55-70.

Geweke, J.; Meese, R.; and Dent, W. (1983) "Comparing alternative tests of causality in temporal systems," Journal of Econometrics 21, 161-94.

Granger, C.W.J. (1963) "Economic processes involving feedback," Information and Control 6, 28-48.

Granger, C.W.J. (1969) "Investigating causal relations by econometric and cross spectral methods," Econometrica 37, 424-38.

Granger, C.W.J. (1980) "Testing for causality: A personal viewpoint," Journal of Economic Dynamics and Control 2, 329-52.

Granger, C.W.J. (1981) "Some properties of time series data and their use in econometric model specification," Journal of Econometrics 16, 121-30.

Granger, C.W.J. (1982) "Generating mechanisms, models, and causality," in Hilderbrand, W. ed., Advances in Econometrics (New York: Cambridge Univ. Press) 237-53.

Granger, C.W.J. and Hatanaka, M. (1964) Spectral Analysis of Economic Time Series (Princeton: Princeton Univ. Press).

Granger, C.W.J. and Newbold, P. (1977a) "The time series approach to econometric model building," in New Methods in Business Cycle Research: Proceedings from a Conferenc Federal Reserve Bank of Minneapolis, 7-21.

Granger, C.W.J. and Newbold, P. (1977b) Forecasting Economic Time Series (New York: Academic Press)

Greenberg, E. and Webster, C.E.Jr. (1983) Advanced Econometrics: A Bridge to the Literature (New York: John Wiley & Sons).

Guilkey, D.K. and Salemy, M.K. (1982) "Small sampling properties of three tests for Granger-causal ordering in a bivariate stochastic system," The Review of Economics and Statistics 64, 668-80.

Hannan, E.J. (1970) Multiple Time Series (New York: John Wiley & Sons).

Hannan, E.J. (1976) "The identification and parameterization of ARMAX and state space forms," Econometrica 44, 713-23.

Hannan, E.J. and Quinn, B.G. (1979) "The determination of the order of an autoregression," Journal of the Royal Statistical Society B 41, 190-95.

Harvey, A.C. (1981a) The Econometric Analysis of Time Series (New York: John Wiley & Sons).

Harvey, A.C. (1981b) Time Series Models (Oxford: Philip Allen Publishers Ltd.).

Haugh, L.D. (1976) "Checking the independence of two covariance stationary time series: a univariate residual cross correlation approach," Journal of the American Statistical Association 71, 378-85.

Haugh, L.D. and Box, G.E.P. (1977) "Identification of dynamic regression (distributed lag) models connecting two time series," Journal of the American Statistical Association 72, 121-30.

Hendry, D.F. (1984) "Monte Carlo experimentation in econometrics," in Griliches, Z. and Intriligator, M. eds., Handbook of Econometrics Vol 2 (Amsterdam: North-Holland) 939-76.

Hendry, D.F.; Pagan, A.R.; and Sargan, J.D. (1984) "Dynamic specification," in Griliches, Z. and Intriligator, M. eds., Handbook of Econometrics Vol 2 (Amsterdam: North-Holland) 1025-100.

Hendry, D.F. and Richard, J.F. (1983) "The econometric analysis of time series," International Statistical Review 51, 111-63. (with discussions)

Hicks, J.R. (1979) Causality in Economics (Oxford: Blackwell).

Hocking, R.R. (1976) "The analysis and selection of variables in linear regression," Biometrica 32, 1-50.

Hood, W.C. and Koopmans, T.C. eds. (1953) Studies in Econometric Method by Cowles Commission Reserch Staff members (New York: John Wiley & Sons).

Hsiao, C. (1979) "Autoregressive modelling of Canadian money and income data," Journal of the American Statistical Association 74, 553-60.

Hsiao, C. (1981) "Autoregressive modelling and money income causality detection," Journal of Monetary Economics 7, 85-106.

Hsiao, C. (1982) "Time series modelling and causal ordering of Canadian money income and interest rates," in Anderson, O.D. ed., Time Series Analysis: Theory and Practice 1 (New York: North-Holland) 671-99.

Intriligator, M.D. (1983) "Economic and econometric models," in Griliches, Z. and Intriligator, M. eds., Handbook of Econometrics Vol 1 (Amsterdam: North-Holland) 181-221.

Jones, R.H. (1975) "Fitting autoregressions," Journal of the American Statistical Association 70, 590-92.

Judge, G.G.; Griffiths, W.E.; Hill, R.C.; and Lee, T. (1980) The Theory and Practice of Econometrics (New York: John Wiley & Sons).

Kalman, R.E. (1982) "Identifiability and problems of model selection," Advances in Econometrics (new York: Cambridge University Press) 169-207.

Kang, H. (1981) "Necessary and sufficient conditions for causality testing in multivariate ARMA models," Journal of Time Series Analysis 2, 95-101.

Kendall, Sir M.; Stuart, A.; and Ord, J.K. (1977) The Advanced Theory of Statistics Vol 3, 4th ed. (New York: Macmillan Publishing Co. Inc.).

Klein, L.R. (1982) "Economic theoretic restrictions in econometrics," in Chow, G.C. and Corsi, P. eds., Evaluating the Reliability of Macro-economic Models (New York: John Wiley & Sons) 23-42. (with discussion by L.L. Pasinetti)

Korisha, S.G. (1983) "Causal implications: the linkage between time series and econometric modelling," Journal of Forcasting 2, 151-68.

Layton, A.P. (1984) "A further note on the detection of Granger instantaneous causality," Journal of Time Series Analysis 5, 15-18.

Leamer, E.E. (1983) "Model choice and specification analysis," in Griliches, Z. and Intriligator, M. eds., Handbook of Econometrics (Amsterdam: North-Holland) 286-330.

Leamer, E.E. (1985) "Vector autoregressions for causal inference," in Brunner, K. and Metzler, A.H. eds., Understanding Monetary Regiems, Carnegie Rochester Conference Series on Public Policy, 22 (Amsterdam: North-Holland) 255-304.

Liu, T.C. (1960) "Underidentification, structural estimation, and forecasting," Econometrica 28, 855-65.

LJung, G.M. and Box, G.E. (1978) "On measure of lack of fit in time series models," Biometrica 65, 297-303.

Lucas, R.E.,Jr. (1976) "Macro-economic policy evaluation: a critique," in Brunner, K. and Metzler, A.H. eds., The Phillips Curve and Labor Markets, Carnegie-Rochester Conference Series on Public Policy 1 (Amsterdam: North-Holland) 19-46.

Lutkepohl, H. (1982) "The impact of omitted variables on the structure of multiple time series: Quenouille's data revisited," in Anderson, O.D. ed., Time Series Analysis: Theory and Practice 2 (Amsterdam: North-Holland) 143-59.

McMillin, W.D. and Fackler, J.S. (1984 "Monetary vs. credit aggregates: an evaluation of monetary policy targets," Southern Economic Journal 50, 711-23.

Mizon, G.E. (1984) "The encompassing approach in econometrics," in Hendry, D.F. and Wallis, K.F. eds., Econometrics and Quantitative Economics (New York: Basil Blackwell Inc.) 135-72.

Morgen, B.J.T. (1984) Elements of Simulation (New York: Chapman and Hall).

Nelson, C.R. (1976) "Gain in efficiency from joint estimation of systems of autoregressive-moving average processes," Journal of Econometrics 4, 331-48.

Nelson, C.R. (1979) "Discussion of the Zellner and Schwert papers," in Brunner, K. and Metzler, A.H. eds., Three Aspects of Policy and Policy Making (Amsterdam: North-Holland 97-102.

Nelson C.R. and Schwert, G.W. (1982) "Tests for predictive relationships between time series variables: a Monte-Carlo investigation," Journal of the American Statistical Association 77, 11-18.

Nelson, H.L.J. and Granger, C.W.J. (1979) "Experience with using the Box-Cox transformation when forecasting economic time series," Journal of Econometrics 10, 57-59.

Nerlove, M.; Grether, D.M.; and Carvallio, J.L. (1979) Analysis of Economic Time Series: A Synthesis (New York: Academic Press.

Newbold, P. (1981) "Some recent developments in time series analysis," International Statistical Review 49, 53-66.

Newbold, P. (1982) "Causality testing in economcis," in Anderson, O.D. ed., Time Series Analysis: Theory and Practice 1 (Amsterdam: North-Holland 71-16.

Newbold, P. (1984) "Some recent developments in time series analysis II," International Statistical Review 52, 183-92.

Newbold, P. and Granger, C.W.J. (1974) "Spurious regressions in econometrics," Journal of Econometrics 2, 110-20.

Nichelburg, G. (1985)) "Small-sample properties of dimensionality statistic for fitting VAR models to aggregate economic data," Journal of Econometrics 28, 183-92.

Omoto, T; Nakagawa, T.: and Akaike, H. (1972) "Statistical approach to computer control of cement rotary kilns," Autometrica 8, 35-48.

Pagano, M. (1981) "On fitting distributed lag models subject to polynomial restrictions," Journal of Econometrics 16, 171-98.

Palm, F. (1977 "On univariate time series methods and simultaneous equation models," Journal of Econometrics 5, 379-388.

Palm, F. and Zellner, A. (1980) "Large sample estimation and testing procedures for dynamic equation systems," Journal of Econometrics 12, 251-83.

Parzen, E. (1974) "Some recent advances in time series analysis," IEEE Transactions on Automatic Control AC-19, 723-30.

Parzen, E. (1977) "Multiple time series modelling: determining the order of approximating autoregressive schemes," in Krishnaiah, P. ed., Multivariate Analysis IV (Amsterdam: North-Holland) 283-95.

Pierce, D.A. (1977) "Relationships--and the lack thereof--between economic time series, with special reference to money and interest rates," Journal of the American Statistical Association 72, 11-22. (with discussions)

Pierce, D.A. and Haugh, L.D. (1977) "Causality in temporal systems: characterization and a survey," Journal of Econometrics 5, 265-93.

Pierce, D.A. and Haugh, L.D. (1979) "The characterization of instantaneous causality: a comment," Journal of Econometrics 10, 257-59.

Plosser, C.T. and Schwert, G.W. (1978) "Money, income and sunspots: measuring economic relationships and effects of differencing," Journal of Monetary Economics 4, 637-60.

Price, M.J. (1979) "Instantaneous causality: a correction," Journal of Econometrics 10, 253-56.

Priestly, M.B. (1981a) Spectral Analysis and Time Series Vol 1 Univariate Series (New York: Academic Press).

Priestly, M.B. (1981b) Spectral Analysis and Time Series Vol 2 Multivariate Series, Prediction and Control (New York: Academic Press.

Prothero, D.L. and Wallis, K.F. (1976) "Modelling macroeconomic time series," Journal of the Royal Statistical Society A 139, 468-500.

Quenouille, M.H. (1957) Analysis of Multiple Time Series (New York: Hafner).

Sargent, T.J. (1976) "A classical macroeconomic model for the United States," Journal of Political Economy 84, 207-237.

Sargent, T.J. (1979) Macroeconomic Theory (New York: Academic Press).

Sargent, T.J. and Sims, C.A. (1977) "Business cycle modelling without pretending to have too much a priory economic theory," in Sims, C.A. ed., New Methods in Business Cycle Research: Proceedings from a Conference, Fedaral Reserve Bank of Minneapolis, 45-109.

SAS User's Guide (1985) Version 5 Edition, SAS Institute Inc, Cary, North Carolina.

Sawa, T. (1978) "Information criteria for discriminating among alternative regression models," Econometrica 46, 1273-291.

Schwarz, G. (1978) "Estimating the dimension of a model," Annals of Statistics 6, 461-64.

Schwert, G.W. (1979) "Tests of causality: the message in the innovations," in Brunner, K. and Metzler, A.H. eds., Three Aspects of Policy and Policy Making (Amsterdam: North-Holland) 55-96.

Sheehan, R.G. and Grien, R. (1982) "Sunspots and cycles: a test of causation," Southern Economic Journal 48, 775-77.

Shibata, R. (1976) "Slection of the order of an autoregressive model by Akaike's information criterion," Biometrica 63, 117-26.

Skoog, G.R. (1976) "Causality characterizations: bivariate, trivariate and multivariate prepositions," Staff Report No. 14, Fedaral Reserve Bank of Minneapolis.

Simon, H.A. (1953) "Causal ordering and indentifiability," in Hood, W.C. and Koopmans, T.C. eds., Studies in Econometric Method (New York: John Wiley & Sons) 49-74.

Sims, C.A. (1972) "Money, income, and causality," American Economic Review 62, 540-52.

Sims, C. (1977a) "Comment to Pierce," Journal of the Americal Statistical Association 72, 23-24.

Sims, C.A. (1977b) "Exogeneity and causal ordering in macroeconomic models," in Sims, C.A. ed., New Methods in Business Cycle Research: Proceeding from a Conference Fedaral Reserve Bank of Minneapolis, 23-43.

Sims, C.A. (1979) "A comment on the papers by Zellner and Schwert," in Brunner, K. and Metzler, A.H. eds., Three Aspects of Policy and Policy Making (Amsterdam: North-Holland) 103-108.

Sims, C.A. (1980) "Macroeconomics and reality," Econometrica 48, 1-48.

Sims, C.A. (1986) "Are forecasting models usable for policy analysis?," Quarterly Review Fedaral Reserve Bank of Minneapolis, 2-16.

Tarascio, V.J. and Caldwell, B. (1979) "Theory choice in econometrics: philosophy and practice," Journal of Economic Issues 13, 983-1006.

Taylor, R. (1967) "Causation," The Encyclopedia of Philosophy Vol 2 (New York: Macmillan Co. and the Free Press) 56-66.

Tiao, G.C. and Box, G.E.P. (1981) "Modelling multiple time series with applications," Journal of the American Statistical Association 76, 802-16.

Thornton, D.L. and Batten, D.S. (1985) "Lag-length selection and tests of Granger causality between money and income," Journal of Money, Credit, and Banking 17, 164-78.

Tjøstheim, D. (1981) "Granger causality in multiple time series," Journal of Econometrics 17, 157-76.

Trivedi, P.K. (1984) "Uncertain prior information and distributed lag analysis," in Hendry, D.F. and Wallis, K.F. eds., Econometrics and Quantitative Economics (New York: Basil Blackwell Inc. 173-210.

Wallis, K.F. (1977) "Multiple time series analysis and the final form of econometric models," Econometrics 45, 1481-497.

Wu, D.M. (1983) "Tests of causality, predeterminedness, and exogeneity," International Economic Review 24, 547-58.

Zellner, A. (1979) "Causality and econometrics," in Brunner, K. and Metzler, A.H. eds., Three Aspects of Policy and Policy Making (Amsterdam: North-Holland) 9-54. (with discussions)

Zellner, A. (1981) "Philosophy and objectives of econometrics," in Currie, D.; Nobay, R.; and Peel, D. eds., Macroeconomic Analysis: Essays in Macroeconomics and Econometrics (London: Croom Helan)

Zellner, A. and Palm, F. (1974) "Time series analysis and simultaneous equation econometric models," Journal of Econometrics 2, 17-54.

FERTILITY AND RELATED LITERATURE

Abeysinghe, T. (1986) "Rising needs and falling family size: Implications from Marx on demographic transition," Review of Social Economy 44, 281-93.

Archives (1981) "Frederic Bastiat on population and economic progress," Population and Development Review 7, 335-44.

Archives (1983) "Bernard Show on 'The population question'," Population and Development Review 9, 145-52.

Arthur, B. (1982) "Review: 'Garry S. Becker, A Treatise on the Family," Population and Development Review 8, 393-97.

Becker, G.S. (1960) "An economic analysis of fertility," in Demographic and Economic Change in Developed Countries, Universities National Bureau Conference Series No. 11 (Princeton: Princeton Univ. Press) 209-31.

Becker, G.S. (1965) "A theory of the allocation of time," Economic Journal 75, 493-517.

Becker, G.S. (1981) A Treatise on the Family (Cambridge: Harvard Univ. Press).

Becker, G.S. and Lewis, H.G. (1974) "Interaction between quantity and quality of children," in Schultz, T.W. ed., Economics of the Family NBER (Chicago: Chicago Univ. Press) 81-90.

Ben-Porath, Y. (1982) "Economics and the Family- match or mismatch? A review of Becker's A Treatise on the Family," Journal of Economic Literature 20, 52-64.

Butz, W.P. and Ward, M.P. (1979) "The emergence of countercyclical U.S. fertility," American Economic Review 69,

Caldwell, J.C. (1982) Theory of Fertility Decline (New York: Academic Press).

Carliner, G.; Robinson, C.; and Tomes, N. (1980) "Female labor supply and fertility in Canada," Canadian Journal of Economics 13, 46-64.

Conger, D.J. and Campbell, J.M.Jr. (1978) "Simultaneity in the birth rate equation: The effect of education, labor force participation, income and health," Econometrica 46, 631-41.

Davis, K. (1984) "Wives and work: The sex role revolution and its consequences," Population and Development Review 10, 397-418.

Davis, K. and Oever, van den P. (1982) "Demographic foundations of new sex roles," Population and Development Review 8, 495-511.

De Tray, D.N. (1974) "Child quality and the demand for children," in Schultz, T.W. ed., Economics of the Family NBER (Chaicago: Chicago Univ. Press) 502-22.

Duesenberry, J. (1949) Income, Savings, and the Theory of Consumer Behavior, (Cambridge: Harvard Univ. Press).

Easterlin, R.A. (1968) Population, Labor Force, and Long Swings in Economic Growth: The American Experience (New York: Columbia Univ. Press).

Easterlin, R.A. (1973) "Relative economic status and the American fertility swing," in Sheldon, E.B. ed., Family Economic Behavior: Problems and Prospects (Philadelphia: J.B. Lippincott) 170-223.

Easterlin, R.A. (1978) "The economics and sociology of fertility: A synthesis," in Tilly, C. ed., Historical Studies of Changing Fertility (Princeton: Princeton Univ. Press) 57-133.

Easterlin, R.A.; Pollak, R.A.; and Wachter, M.L. (1980) "Toward a more general economic model of fertility determination: endogenous preferences and natural fertility," in Easterlin, R.A. ed., Population and Economic Change in Developing Countries NBER (Chicago: Chicago Univ. Press) 81-150.

Heckman, J.J. and Willis, R.J. (1975) "Estimation of a stochastic model of reproduction: An econometric approach," in Terleckyj, N.E. ed. Household Production and Consumption (New York: NBER) 99-138.

Keely, M.C. (1975) "A comment on Leibenstein's 'An interpretation of the economic theory of fertility'," Journal of Economic Literature 13, 461-68.

Kelley, A.C. (1976) "Review of 'Economics of the Family: Marriage, Children and Human Capital' edited by T.W. Schultz," Journal of Economic Literature 14, 516-20.

Keynes, J.N. (1890) The Scope and Method of Political Economy, (London: MacMillan & Co.).

Leibenstein, H. (1954) A Theory of Economic-Demographic Development (Princeton: Princeton Univ. Press).

Leibenstein, H. (1974) "An interpretation of the economic theory of fertility: Promissing path or blind alley?," Journal of Economic Literature 12, 457-79.

Leibenstein, H. (1975) "The economic theory of fertility decline," Quarterly Journal of Economics 89, 1-31.

Leibenstein, H. (1976) Beyond Economic Man (Cambridge: Harvard Univ. Press).

Leibenstein, H. (1980) "Comments on 'Toward a more general model of fertility determination' by Easterlin et al.," in Easterlin, R.A. ed., Population and Economic Change of Developing Countris NBER (Chicago: Chicago Univ. Press) 135-140.

Leibenstein, H. (1981) "Economic decision theory and fertility behavior," Population and Development Review 7, 381-400.

Lindert, P.H. (1977) "American fertility patterns since the civil war," in Lee, R.D. ed., Population Patterns in the Past (New York: Academic Press) 229-76.

Lindert, P.H. (1980) "Child costs and economic development," Easterlin, R.A. ed., Population and Economic Change in Developing Countries NBER (Chicago: Chicago Univ. Press) 5-69.

Malthus, R.T. (1959) Population: The First Essay Ann Arbor Paperbacks, The Univ of Michigan Press.

Michael, R. and Becker, G.S. (1973) "On the new theory of consumer behavior," Swidish Journal of Economics 75, 378-96.

Mincer, J. (1962) "Labor force participation of married women," in _Aspects of Labor Economics_ NBER (Princeton: Princeton Univ. Press) 63-105.

Mincer, J. (1963) "Market prices, opportunity costs and income effects," in Christ, C.F. ed., _Measurement in Economics_ (Stanford: Stanford Univ. Press) 67-82.

Ostry, S. and Zaidi, M.A. (1979) _Labor Economics In Canada_, Third edition (Toronto: MacMillan of Canada).

Phillips, P. and Phillips, E. (1983) _Women and Work_, (Toronto: James Lorimer & Co.).

Pollak, R.A. (1970) "Habit formation and dynamic demand functions," _Journal of Political Economy_ 78, 745-63.

Pollak, R.A. (1976) "Interdependent Preferences," _American Economic Review_ 66, 309-20.

Pollak, R.A. (1978) "Endogenous tastes in demand and welfare analysis," _American Economic Review_ 68, 374-79.

Pollak, R.A. and Wachter, M.L. (1975) "The relevance of the household production function and its implications for the allocation of time," _Journal of Political Economy_ 83, 255-77.

Samuelson, P.A. (1976) "An economist's non-linear model of self-generated fertilitiy waves," _Population Studies_ 30, 243-47.

Schultz, T.P. (1981) _Economics of Population_ (Mass.: Addison-Wesley).

Schultz, T.W. ed.,(1974) _Economics of the Family_ NBER (Chicago: Chicago Univ. Press) 3-22.

Stigler, G.J. and Becker, G.S. (1976) "De gustibus non est dispuntandum," _American Economic Review_ 67, 76-90.

Wachter, M.L. (1972) "A labor supply model for secondary workers," _The Review of Economics and Statistics_, 54, 141-151.

Wachter, M.L. (1975) "A time series fertility equation: The potential for a baby boom in the 1980's," _International Economic Review_ 16, 609-24.

Wachter, M.L. (1977) "Intermediate swings in labor force participation," _Brookings Papers on Economic Activity_ 2, 545-76.

Weller, R.H. (1977) "Demographic correlates of woman's participation in economic activities," International Population Conference, Mexico Vol. 3 (Belgium: IUSSP) 497-516.

Willis, R.J. (1974) "Economic theory of fertility behavior," in Schultz, T.W. ed., Economics of the Family NBER (Chicago: Chicago Univ. Press) 25-75.

SOURCES OF DATA

Bank of Canada, Statistical Summary Supplements, Bank of Canada Review (Various issues) Ottawa.

Central (Canada) Mortage and Housing Corporation, Housing in Canada, Canadian Housing Statistics (Various issues).

Leacy, F.H. ed. (1983) Historical Statistics of Canada 2nd ed. (Ottwa: Statistics Canada).

Skoulas, N. (1974) Determinants of the Participation Rate of Married Women in the Canadian Labor Force: An Econometric Analysis, (Ottwa: Statistics Canada) Cat. No. 71-522.

Statistics Canada or Dominion Bureau of Statistics, Distribution of Non-farm Incomes in Canada by Size, Cat. Nos. 13-503, 13-504, 13-512, 13-517, 13-521, 13-528; Income Distributions by Size in Canada, Cat Nos. 13-534, 13-544, 13-207(various issues); Income Distributions (1951-1965), Cat. No. 13-529; Family Expenditure in Canada, (1984) Cat. No. 62-555; Canadian Statistical Review, (various issues) Cat. No. 11-003; Population Estimates by Marital Status, Age, and Sex (various issues) Cat. No. 91-203; Earnings of Men and Women, (1984) Cat. No. 13-217; The Labour Force, (various issues) Cat. No. 71-001; Family Characteristics and Labor Force Activity (1977-1984) Cat. No. 71-533; Vital Statistics, (various issues) Cat. No. 84-202; Historical Labor Force Statistics (1985) Cat. No. 71-201; Financial Flow Accounts, Cat. Nos. 13-563, 13-214.