DETECTING MACROECONOMIC IMPACTS ON AGRICULTURAL
PRICES AND EXPORT SALES:
A TIME SERIES FORECASTING APPROACH

by

Girard W. Bradshaw II

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of
Master of Science
in
Agricultural Economics

APPROVED:

David R. Orden, Chairman

William J. Taylor

Robert V. Foitz

June, 1988
Blacksburg, Virginia
LO 5655 V855 1989 B724 c.2
DETECTING MACROECONOMIC IMPACTS ON AGRICULTURAL PRICES AND EXPORT SALES:
A TIME SERIES FORECASTING APPROACH

by
Girard W. Bradshaw II
David R. Orden, Chairman
Agricultural Economics

(ABSTRACT)

The effect of movements in the real exchange rate on agricultural prices and agricultural export sales is assessed based on the principle of Granger causality. An out-of-sample forecasting procedure is used to conduct tests for Granger causality from the exchange rate to agricultural prices and export sales. Technical time series issues such as stationarity, the method of lag-length selection, in-sample versus out-of-sample tests for Granger causality, and long-range versus short-range forecasting are considered in relation to the outcome of Granger causality tests.

Theoretical and empirical studies are reviewed which indicate the importance of working with stationary data series when testing for Granger causality. Differing methods of lag-length selection are found to affect the outcome of both in-sample and out-of-sample tests for Granger causality. The usual in-sample tests for Granger causality are compared to out-of-sample tests; the results of the comparison reveal that the in-sample tests do not in general agree among themselves, nor do they agree with the out-of-sample tests' results. This indicates the
importance of searching the model space for the best specification before conducting Granger causality tests. Long-range forecasts are compared to the 1-step ahead forecasts used to test for Granger causality; these forecasts corroborate the out-of-sample tests for Granger causality in finding significant impacts from the exchange rate to agricultural export sales and agricultural prices.
Acknowledgements

The completion of this thesis marks both an end to an important phase in my life as well as a new beginning. The completion of a Master's degree is in some sense a milepost, a signal to the world that I have "mastered" something. I recall, however, the legend of the Delphic Oracle who told Socrates that he was the wisest of all men because he was the only one who had the knowledge that, in truth, he knew very little at all. In this sense, as I have proceeded with my formal education, I have indeed become wiser.

Above all, this is my favorite part of the thesis because for one nobody can tell me to rewrite it, and also because I get to express how I feel about the many people who share in my success. At the very top of my list comes my family. Since I was an undergraduate, my parents Lucy and Lee Bradshaw have seen me through many transitions, some good and some bad. Though I wish I could say the same for myself, their strength, their confidence in me, and their love for me has never wavered. The older I get the more I realize how insightful they are.
My Grandparents Rocco, Nancy, Ruth, and my late Grandfather Girard have shown me nothing but love and support throughout my life. They have taught me about life and what it takes to be a good person, to be respectful and caring, and to try and leave this world a little better than when you arrived. Each of them has held fast to this standard and has succeeded in it. I love them all dearly. The love of my wife has enhanced my life in ways I never dreamed possible. To say that her love and support have sustained me is not enough. She has worked tirelessly, without complaint, to support us, and I love and respect her more than mere words can express.

The saying goes that in college, you have some good professors and some bad professors, and only a few that you will remember for the rest of your life. For those professors who have changed my life, merely saying thank you seems hardly enough. To Nick Smith and Charlie Reed who gave me a love of philosophy, Aristophanes, and intellectual direction, I say only that each of you is a credit to your profession. I have seen how importantly each of you has figured into my life and the lives of my peers. You are appreciated and loved more than you know. I want to thank my major professor David Orden. I owe my successful and productive program in most part to him. He was always there with encouragement when I was down and sound advice when I needed it. I will remember most of all not the professional success I have had working with him, but that he is a good and dear friend. In the end, that is really something important. I also want to acknowledge and thank Leonard Shabman for giving me a chance to redeem myself in graduate school, and to Barbera Craig for lots of good
advice and a chance to show that I could be a good TA. Thanks are also due to
Bill Taylor and Bob Foutz who served on my thesis committee; their positive at-
titudes and the time they devoted to working with me made the process a fruitful
and enriching one.

Finally, I want to thank Lee and Eric Thunberg, Linda Kipps, Basu, David
Woddall-Gainey, John, Penny, and Lynn Furnival for being good friends.
TO MY GRANDPARENTS
# Table of Contents

**Chapter 1 - Introduction** ................................................................. 1  
1.1 The Time Series Approach to Forecasting .......................................... 1  
1.2 Objectives of this Study ................................................................. 7  
1.3 Time Series Analysis in Agricultural Economics .................................... 8  
1.4 Specification of Time Series Models ................................................. 11  
1.5 Outline of the Study ........................................................................ 13  

**Chapter 2 - Univariate Time Series Analysis and Modeling** ..................... 15  
2.1 Time Series, Stochastic Processes, and Stationarity ............................. 15  
2.2 Measures of Correlation .................................................................... 20  
2.3 Linear Model Representations .......................................................... 22  
2.4 Nonstationarity .................................................................................. 33  
2.5 Seasonality ....................................................................................... 39  
2.6 Specifying ARIMA Models ............................................................... 44  

**Chapter 3 - Vector Time Series Analysis and Modeling** .......................... 56  
3.1 Introduction ..................................................................................... 56
3.2 Linear Models for Vector Stochastic Processes ........................................... 58
3.3 Vector Autoregressive Models ................................................................. 62
3.4 Nonstationarity in the Multivariate Case ............................................... 71
3.5 Co-integrated Economic Variables ......................................................... 76

Chapter 4 - Detecting Macroeconomic Impacts on Agriculture ......................... 85
4.1 Data Description ......................................................................................... 85
4.2 Data Analysis and Tests of Time Series Properties .................................. 87
4.3 Evaluation of Forecasts ........................................................................... 105
4.4 Specification of Forecasting Models ......................................................... 107
4.5 Empirical Forecasting Results ................................................................. 121

Chapter 5 - Summary and Conclusions ......................................................... 136
5.1 Summary ................................................................................................. 136
5.2 Conclusions ............................................................................................. 138
5.3 Suggestions for Further Research ......................................................... 141

References ................................................................................................... 142
Chapter 1 - Introduction

We are born by accident into a purely random universe. Our lives are determined by entirely fortuitous combinations of genes. Whatever happens happens by chance. The concepts of cause and effect are fallacies. There are only seeming causes leading to apparent effects. Since nothing truly follows from anything else, we swim each day through seas of chaos, and nothing is predictable, not even the events of the very next instant.


1.1 The Time Series Approach to Forecasting

This study focuses on detecting the effects of real exchange rates on real agricultural prices and export quantities from a forecasting perspective. Almost fifteen years ago, the appearance of Schuh's (1974) paper on the effect of the exchange rate on the agricultural economy opened-up a new area of research into macroeconomic-agricultural linkages. Since 1974, the link between the macroe-
conomy and agriculture has been the topic of much research (e.g., Chambers and Just 1982; Schuh and Orden 1985). As a consequence, the importance of macroeconomic factors for agriculture has, in general, been well received. In spite of this literature, there still remain skeptics of the importance of macroeconomic factors for explaining the behavior of the agricultural economy (Hillman and Faminow 1986; Robinson 1987).

To further investigate the linkage between the macroeconomy and agriculture, this study will concentrate on the specification of time series forecasting models. The development and extension of time series analysis techniques can be credited to engineers, statisticians, operations researchers, and economists alike. Each field began research into time series techniques from a very different starting point with the resolution of very different problems in mind. Years of research have produced a body of knowledge that has considerable power for analyzing data, investigating theoretical relationships, and forecasting.

Though techniques for analyzing time series have been in existence since the turn of the century, the past twenty or so years have seen a considerable refinement and extension of the early techniques and a concomitant explosion in the literature on time series analysis. In economics in particular, the quantitative analysis of economic relationships has its historical foundation in regression analysis. To a degree, economists have come to view this more classical approach as being in some way separate from the time series approach, a distinction that is now breaking down.

Chapter 1 - Introduction
Economic forecasting is a field of endeavor in which this dichotomy had been quite pronounced until a few years ago. Historically, two approaches to forecasting are identified: (1) structural econometric/causal forecasting, and (2) extrapolative time series forecasting. A recent trend in the literature has been an exploration of how these two approaches can better be synthesized.¹

Structural econometric models are specified, as is well known, by appeal to prevailing economic theory. They consist of a set of dependent variables (the variables to be forecasted) and a set of independent variables which are used to "explain" or account for the variation in the dependent variables. These models aim to capture the structural relationships, identified from theoretical investigations, among the variables in the economy, often employing numerous over-identifying restrictions in the process. The popularity of large-scale simultaneous equation models of this type reached a peak in the 1960s and early 1970s. They continue to be widely used in commercial forecasting, and, to an extent, in research. However, in the late 1970s, forecasters using these models, in particular macroeconometric models, were confounded by their failure to accurately predict simultaneous high inflation and high unemployment levels (Lucas and Sargent 1979). This break-down in forecasting accuracy opened the door for simpler, less costly, and more accurate alternative forecasting models. Around the same time, the classic work of Box and Jenkins (1970) first appeared, offering time series

¹ The Structural Econometric Modeling Time Series Analysis (SEMTSA) approach of Zellner (1979), the book by Harvey (1981a), and papers by Davidson et al. (1978) and Hendry (1978) are examples of work which emphasize the need for more synthesis. Blanchard and Watson (1986), Bernanke (1986), Sims (1986) and Fackler (1988) have addressed the issue of the identification of dynamic simultaneous equation structural models and their relationship to multivariate time series models.
models as a systematic alternative to structural modeling for the expressed purpose of forecasting.

Time series models are built on the premise that a time-ordered random variable (or a vector of random variables) has a particular recurring statistical history which can be modelled and then exploited for the purpose of forecasting. The unique statistical history is used to project forward the likely path of the time series, thus generating an extrapolative forecast. Behind the idea of time series forecasting is the eclectic view that we may not know enough about the true structure of the economy to construct a detailed structural econometric model that will forecast well (see, for example, Sims 1980).

For illustrative purposes, we shall delineate two classes of time series models, those that do not allow for dynamic interactions among variables (univariate) and those that do (multivariate). Univariate time series models express the variation in a time series as a function of autoregressive terms (past own values) and moving average terms (contemporaneous and past errors):²

\[ X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q} \]  \hspace{1cm} (1.1.1)

Multivariate time series models, on the other hand, reflect the importance of "the influence of other observable variables known or suspected to be related to the series of interest" (Kling and Bessler 1985). The multivariate time series models to be used in this study will be vector autoregressions (VARs). A VAR

² In practice univariate autoregressions are often specified as alternatives to ARMA models as pure autoregressions are often easier to specify than ARMA models.
model does not impose *a priori* restrictions such as exogeneity or functional form as used in the identification of structural simultaneous equation models. Instead, a VAR is a reduced-form model in which interactions that are present in the data are allowed to emerge on their own. If \( X_t' = (X_{1t}, \ldots, X_{mt}) \) is a vector of variables that we wish to model with a VAR, under the conditions discussed in Chapter 3, \( X_t' \) has a vector autoregressive representation:

\[
\Phi(B)X_t = E_t
\]

(1.1.2)

where \( \Phi(B) \) is an \( m \times m \) infinite matrix function in the backshift operator, and \( E_t \) is an \( m \times 1 \) vector of well-behaved error terms. Each element of \( \Phi(B) \) is an infinite polynomial in the backshift operator which provides the structure of the autoregression. This infinite AR structure is then *approximated* by a finite autoregression for empirical estimation. Choosing the lag length in a VAR is an important issue in the empirical estimation of time series models, and will also be addressed in Chapter 3.

In practice, the question of whether a traditional structural econometric model or a time series model is better for a particular forecasting project turns on the validity of the prior information that we have. If a particular economic theory is "true," it would be unwise not to use that information. A univariate

---

3 The backshift operator, also called the lag operator, performs the following operation, \( B^kX_t = X_{t-k} \), on either a single random variable or a vector of variables.

4 One way to test the general credibility of identification restrictions would be to conduct a forecasting experiment where the structural model competes with various specifications of multivariate time series, and univariate time series models. Zellner (1982) argues that structural models which have excellent in-sample fit must still prove their ability to forecast well in order for them to be useful contributions to economic science.
ARIMA model incorporates no prior *economic* information thus may be a poor choice as a forecasting model in the face of a structural model imposing valid identification restrictions (Prothero and Wallis 1976). In fact, however, ARIMA models frequently out-perform structural econometric models in forecasting.\(^5\)

The multivariate time series approach asserts that the truth lies somewhere between the traditional simultaneous equation approach and the atheoretical univariate time series models. Through economic theory, variables can be identified that have a high prior probability of having an important effect on the variables to be forecasted, though we are not quite certain how these interrelationships are manifest, in, say, particular functional forms or exclusion restrictions. VAR models can be viewed as quasi-time series models in that they are specified using some a priori information from economic theory to guide the selection of the variables to be included in a multivariate specification. It is hoped that such a model will provide forecasts that are superior to a univariate model.

---

\(^5\) Ashley argues that one reason univariate models forecast well relative to structural models is that in many cases the forecasts of explanatory variables in structural models may be too poor that it will be useless to incorporate them in the forecasting model even if the theory supporting their relevance is valid. A well-known example is given in Nelson (1972) where it was demonstrated that a univariate ARIMA model could out-forecast the Federal Reserve-MIT-Penn quarterly model of the U.S. economy. See also Cooper (1972).
1.2 Objectives of this Study

The overall objective of this study is to detect exchange rate impacts on agricultural prices and export sales using time series forecasting models. Specific objectives include the following:

1. To review time series forecasting techniques and some of the recent technical issues that are associated with both univariate and multivariate time series analyses.

2. To test each agricultural price and export sales time series for stationarity and possibly for co-integration with the exchange rate, and to evaluate the impact of ignoring nonstationarity on forecasting accuracy.

3. To specify competing multivariate and univariate models for forecasting the agricultural prices and export sales variables and to assess their relative forecasting accuracy using a procedure that can evaluate the statistical significance of improvements in short-run forecasting accuracy, thereby performing out-of-sample tests for Granger causality.

4. To compare the usual in-sample Granger causality tests to the out-of-sample tests, and to evaluate the impact of using differing lag-selection methods on these results.
5. To heuristically evaluate long-range forecasts from the specified models and compare these results to those from the short-range forecasts.

1.3 Time Series Analysis in Agricultural Economics

In agricultural economics, one of the earliest comparisons of relative forecasting performance was conducted by Leuthold et al. (1970). Daily hog prices and quantities sold were forecasted using alternative specifications of an econometric model and a univariate ARIMA model. Each of these specifications was compared to forecasts from a simple random walk. The authors found that the atheoretical time series model compared favorably with the econometric model. Schmitz and Watts (1970) used a univariate ARIMA to forecast wheat yields in the largest wheat producing countries. The ARIMA forecasts were compared to a popular ad hoc method known as exponential smoothing. The results were mixed in this instance, with the ARIMA faring better in the U.S. case, while exponential smoothing predicted yields in Canada, Argentina, and Australia more accurately.

Following these early studies, the December 1970 issue of the American Journal of Agricultural Economics published a symposium on "New Models in Price Forecasting," reflecting, among agricultural economists, both an interest in

---

6 Exponential smoothing is a weighted moving average of the values in an observed time series with the weights chosen on some arbitrary basis (Nelson 1972), see also Chapter 2, section 2.5.
the new stochastic models popularized by Box and Jenkins and a strong interest in forecasting microeconomic phenomena. Many participants commented favorably on the use of time series models for agricultural price forecasting because of their good predictive performance (see e.g., Bieri and Schmitz 1970). However, after this early experimentation with time series modeling in agriculture, there came few articles on the subject reflecting the profession's strong connection with regression analysis which uses economic theory as a backbone.\footnote{A notable exception was Myers (1972). He introduced spectral analysis in conjunction with multiple regression for commodity forecasting. Spectral analysis is a branch of time series analysis which uses the time series' power spectrum to identify the historical patterns in the data for use in modeling. Spectral analysis is used in the \textit{frequency domain} of time series analysis. The models and techniques in this study are all in the \textit{time domain}. See Granger and Newbold (1986), chapter 2, for an introduction to spectral methods.}

Renewed interest in time series analysis within agricultural economics became apparent in the early 1980s. Brandt and Bessler (1981) investigated the performance of composite forecasting techniques using forecasts from econometric models, ARIMA models, and expert opinion methods. Their main conclusion was that composite forecasts out-perform any single model. Following their first study, Brandt and Bessler (1982) compared a univariate ARIMA model of hog prices, and a bivariate transfer-function model, and found, using the evaluation procedure in Ashley et al. (1980), that the improvement in going from the univariate to the bivariate model was not statistically significant. Spriggs (1981) compared the forecasts of a univariate ARIMA model with forecasts represented by futures prices, and found the futures prices to out-perform the ARIMA. Brandt and Bessler (1983) compared predictive accuracy among exponential smoothing, a univariate ARIMA, an econometric model, expert
opinion forecasts, and various composite forecasts. Their results indicated that quantitative methods generally out-performed the expert judgement, while composite forecasts performed best using a lowest mean square error criterion.

Recently, interest in multivariate time series models has turned to vector autoregressions. In the general economics literature, the PhD dissertation by Litterman (1976) and the pioneering article by Sims (1980) have led to increased uses of VAR techniques both as a tool for forecasting and policy analysis. In agricultural economics, Bessler (1984a) discusses the methodological aspects of fitting VARs and applies a VAR to evaluate dynamic economic relationships in the hog market. Bessler (1984b) uses a VAR model to investigate the relationship between relative prices and money in Brazil.

In terms of previous time series studies of exchange rate impacts on agriculture, Orden (1986) investigated the dynamic effects of macroeconomic shocks on U.S. agriculture using VARs. He found evidence that movements in the real exchange rate have substantial impacts on agricultural exports and real prices received by farmers. Orden's study used policy analysis techniques developed by Sims (1980, 1986), and employed Sims' lag length selection criterion.8

From a purely forecasting perspective, Bessler and Babula (1987) found the real exchange rate to have little impact on improving the accuracy of forecasts

8 Thornton and Batten (1985) found that, for money-income relationships, different lag structures can change the outcome of Granger causality tests. Therefore, we might place value on consideration of alternative test procedures for lag length selection, and such procedures will be discussed below.
of wheat exports. The real exchange rate did, however, have a notable effect on increasing the accuracy of forecasts of real wheat prices.⁹

1.4 Specification of Time Series Models

As discussed in section 1.3, multivariate time series models are not always able to out-perform univariate models in out-of-sample forecasting ability. This is puzzling as one might expect that a multivariate model should forecast at least as well as a univariate model because it theoretically encompasses that model.

The failure of multivariate models to out-forecast univariate models is likely the result of ignoring some very important characteristics of time series data. For example, Litterman (1984) has argued that aggregate economic data suffers from a low signal-to-noise ratio, meaning that the useful, systematic, variation is a much smaller proportion of the overall variation than is the purely random variation. This random noise overpowers the useful signal, i.e., the variation that can be used to explain movements in another variable. Litterman (1986) explains that the parameters in a VAR will likely fit both the useful signal as well as the noise, resulting in an overparameterized model. The random variation, however, is not useful for forecasting. The task of the forecaster, then, is to devise a way to filter out the random noise to reveal the variation that is systematic. The

⁹ Bessler and Babula use a decomposition of forecast-error variance to isolate the effect of the exchange rate on the wheat price in their study.
multivariate forecasting problem is thus a trade-off between oversimplification and overparameterization.

Different authors have proposed different ways to combat this overparameterization problem. Litterman (1976) and Doan, Litterman, and Sims (1983) propose a Bayesian approach, while Lutkepohl (1985) investigated the application of statistical decision criteria that result in the choice of a more parsimonious unrestricted VAR.\textsuperscript{10} Hsiao (1979) developed a procedure that also employs a statistical decision criterion, but allows for more realistic \textit{differing lag structures as well as exclusion restrictions to be imposed} in each equation in the system. Each of these approaches will be discussed in detail in Chapter 3.

Another reason given for the poor performance of VAR representations of economic time series is the lack of attention given to data issues inherent in time series econometrics. For example, if the time series to be modeled contains a deterministic trend or a unit root, or if the variance of the series is not constant throughout, the series is said to be \textit{nonstationary}. Many economic time series, especially macroeconomic series, are nonstationary in their levels (Wasserfallen 1986; Nelson and Plosser 1982). Because multiple time series theory relies on stationarity for its validity, the modeling of nonstationary series as if they are stationary can produce undesirable results, one of which can be poor forecasting performance. Another undesirable result of nonstationarity is the appearance of

\textsuperscript{10} An unrestricted VAR is one in which the lag lengths of all the included variables are the same, and each variable is allowed to dynamically interact with every other variable, i.e., there are no exclusion restrictions imposed.
spurious relationships when nonstationary variables are regressed on one another (Granger and Newbold 1974, 1986).

A related consideration arises when sets of random variables are being modeled, as in a VAR; attention must be paid not only to the stationarity of individual variables, but to possible equilibrium relationships among the variables. These relationships are manifest when two or more nonstationary variables have a linear combination that is stationary. Such variables are then said to be co-integrated. One of the consequences of ignoring co-integration is a loss in forecasting accuracy as important information concerning the underlying relationship among the variables being modeled is being ignored. In fact, when co-integration is present, the usual VAR representation (1.1.2) is inappropriate (Engle and Granger 1987).

1.5 Outline of the Study

The importance of the above issues to any time series modeling exercise makes it necessary to have a strategy for investigating the time series properties of the variables to be modeled before a particular model is chosen. Chapter 2 will provide an introduction to univariate time series analysis and modeling, including tests for stationarity and seasonality. In addition, procedures for modeling and estimating univariate time series are investigated. Chapter 3 contains an introduction to the analysis and modeling of vector stochastic processes, including
tests for co-integration. Chapter 4 contains a description of the data, the empirical results of the diagnostic tests, the orders of the estimated models, and tests for improved forecasting accuracy. Chapter 5 is a summary and conclusions chapter.
Chapter 2 - Univariate Time Series Analysis and Modeling

You look at where you’re going and where you are and it never makes sense, but then you look back at where you’ve been and a pattern seems to emerge. And if you project forward from that pattern, then sometimes you can come up with something.

Robert M. Pirsig, Zen and the Art of Motorcycle Maintenance.

2.1 Time Series, Stochastic Processes, and Stationarity

A time series is a sequence of observed time ordered values, such as yearly gross national product from 1957 to 1987.\(^\text{11}\) Given a series of such observations, \(X_1, X_2, \ldots, X_T\), it is assumed that these observations are one realization of time-

\(^{11}\) This is an example of a discrete time series because each observation occurs at an equidistant interval in time. By contrast, a continuous time series has observations at every point in time which results in a trace of values.
ordered random variables, also denoted by $X_1, X_2, \ldots, X_T$. The random variables are a subsequence of an infinite sequence of time-ordered random variables, $X_t, t \in (...) - 2, -1, 0, 1, 2, \ldots$, which is called a stochastic process (Judge et al. 1985). The purpose of time series modeling is to construct a model for the realization and in so doing approximate with this model the underlying stochastic process which, we assume, has generated that realization.

The relationship among any set of random variables is characterized by a joint distribution function. Since a stochastic process consists of a set of random variables, the elements of the stochastic process are characterized by a joint distribution function. The practitioner is relegated to working with the realization, $X_1, \ldots, X_T$, and each $X_t$ of this realization is assumed to be a random variable itself, theoretically capable of having an infinite number of different outcomes. This assumption allows us to treat the realization as if it were the stochastic process, therefore $X_1, \ldots, X_T$ is itself characterized by a joint distribution function as follows:

$$F(X_1, \ldots, X_T) = P\{X_1 \leq X_1^*, \ldots, X_T \leq X_T^*\} \quad (2.1.1)$$

Here, $P\{ \}$ denotes probability, and $X_t^*$ indicates a theoretically possible outcome of its associated random variable.

---

12 Granger and Newbold (1986) note that the relationship between the realized time series and the underlying stochastic process is analogous to the relationship between sample and population in statistics.

13 If $x$ and $y$ are discrete random variables, the function given by $f(x, y) = P(x = x', y = y')$ for each pair of values $(x', y')$ within the respective ranges of $x$ and $y$ is called the joint probability function or the joint probability distribution (Freund and Walpole 1985).
Now, suppose further that we know the joint distribution function in (2.1.1), to be a multivariate joint normal distribution. A joint normal distribution for a vector of random variables, \(X_t\), \(t = 1 \ldots T\) takes the form:

\[
F(X_t) = \frac{1}{(2\pi)^{1/2} |\Sigma|^{1/2}} \exp\left[-1/2(X_t - \mu_t)^\prime \Sigma^{-1} (X_t - \mu_t)\right]
\] (2.1.2)

Here, \(\Sigma\) is the \(T \times T\) positive definite covariance matrix, and \(\mu_t\) is a \(T \times 1\) vector of means corresponding to each random variable in \(X_t\). In order to calculate the joint distribution, we need to estimate both \(\mu_t\) and \(\Sigma\).

Estimating means and variances for each \(X_t\) is a hopeless task with only the single realization, observations \(X_1, \ldots, X_T\), in hand. In order to calculate the covariance of, say, \(X_2\) and \(X_3\), we need the following information:

\[
\text{Cov}[X_2, X_3] = E[(X_2 - E[X_2])(X_3 - E[X_3])]
\] (2.1.3)

When there is only one realization of the underlying stochastic process, it is impossible to obtain a meaningful estimate of \(\mu_t\) for every \(t\); it is also impossible to estimate covariances without means. If it were possible to obtain several realizations of the same process over the same time period (e.g., if it were possible to stop the economy and go back to, say, 1950 and begin it again) we could obtain another realization of the process which generates our data. Each set of observations would be different because the process is stochastic, but would obey the

---

\(^{14}\) \(X_t\) will be used to represent a vector of random variables \(X_t = (X_{t1}, \ldots, X_{tn})\) as well as a series of observations on a single time series with the index running from 1 \(\ldots T\). The meaning should be clear from the context. See also the introduction to Chapter 3 for further clarification.
same probabilistic laws (Harvey 1981b). If we could sample the process \( m \) times, we could estimate the necessary mean values as follows:

\[
\hat{\mu}_t = \frac{1}{m} \sum_{j=1}^{m} X_{jt} \quad \text{for every } t = 1,2,\ldots,T. \quad (2.1.4)
\]

Unfortunately, for economic processes we are relegated to only one realization of a stochastic process as we obviously cannot stop the economy in order to return to some starting point.

In order to get around the constraint of having only one realization, it is necessary to make restrictive assumptions concerning the behavior of the mean and covariances of the underlying stochastic process. The first assumption to be made is that the process is *stationary*. A stochastic process is called stationary if the first two moments (the mean and covariances) of the joint distributions of its finite subsequences are finite and time invariant (Judge 1985):

\[
E[X_t] = \mu < \infty \quad \text{for all } t \quad (2.1.5a)
\]

\[
E[(X_t - \mu)^2] = \sigma^2 < \infty \quad \text{for all } t \quad (2.1.5b)
\]

\[
E[(X_t - \mu)(X_{t+k} - \mu)] = \gamma_k \quad \text{for all } t,k \quad (2.1.5c)
\]

Equations (2.1.5a) and (2.1.5b) imply that the process mean and variance are finite and do not change through time. Equation (2.1.5c) means that the covariance between any two values of the process depends solely on the *distance*
between these values in time and not on time itself. That is, covariances will be equal for pairs of observations that are equal distances apart in time. For example:

\[ \text{Cov}(X_2, X_5) = \text{Cov}(X_9, X_{12}) = \text{Cov}(X_n, X_{n+3}) = \gamma_3 \]  \hspace{1cm} (2.1.6)

A second assumption that must be imposed to work with empirical time series is an assumption which insures that the time average of the stochastic process is an unbiased estimator of the population mean, \( \mu \), and that the estimates of the process covariances, \( \gamma_k \), are consistent. This is the assumption that the process be \textit{ergodic}. Granger and Newbold (1986) offer the following heuristic definition of ergodicity: "What is required [for ergodicity] is that values of the process sufficiently far apart in time are almost uncorrelated, so that by averaging a series through time, one is continually adding new and useful information to the average" (Granger and Newbold 1986, p.5).\(^{16}\) Given the assumptions of stationarity and ergodicity, good estimates of the mean, variances and covariances of the underlying stochastic process can be found from a single empirical realization of the process.

\(^{15}\) This definition corresponds to what is known as \textit{weak stationarity} or \textit{covariance stationarity}. Another definition of stationarity is strong stationarity which requires that the \textit{joint distribution} of the process be invariant with respect to time, that is, \( P(X_{t-\cdots}, X_{t+h}) = P(X_{t+m-\cdots}, X_{t+m+h}) \) (see Nelson 1972).

\(^{16}\) Fuller (1976, p. 230) and Hannan (1970, p. 201) provide a more complete and rigorous account of the concept of ergodicity.
2.2 Measures of Correlation

The goal of forecasting is to say something about the future based on relationships that have been observed in the past. Consequently, we want to assess the linear association between the random variables in the stochastic process. We do this by calculating the covariation within the time series under examination, provided that (2.1.5a-c) hold. The estimated covariances, then, depend only on the time lag between the values, thus, the following notation will be used throughout:

\[ \gamma_0 = \text{variance of the process} = \text{Cov}(X_t, X_t) \]

\[ \gamma_k = \text{autocovariance at lag } k = \text{Cov}(X_t, X_{t+k}) \]

The set of values \( \{\gamma_k, \text{ for } k = 1, 2, \ldots\} \) is called the autocovariance function.\(^{17}\) Since the autocovariance depends on the unit of measurement of the underlying variables (i.e., dollars, bushels, tons), as well as the frequency of observation (i.e., years, quarters, months), dividing \( \gamma_k \) by \( \gamma_0 \) will normalize (make unit free) the autocovariance function. A normalized autocovariance function is called an autocorrelation function (ACF). The autocorrelation function of \( X_t \) is designated as follows:

\[ \text{ACF} \]

\(^{17}\) Covariances are called autocovariances because the relationship refers to two random variables within the same realization.
\[ \rho_k = \frac{\gamma_k}{\gamma_0} \quad \text{for all } k = 0, 1, 2, 3,... \]

Note that \( \rho_0 \) is by definition equal to unity. The autocorrelation function indicates the extent to which one value of the process is correlated with previous values. It can be thought of as the memory of the process, i.e., the extent to which the observation at time "t" depends on that at time "t-k" (Granger and Newbold 1986). An interesting feature of the ACF is that it is a symmetric function about the zero lag indicating that the autocorrelation at lag k of \( X_t \) is the same as the autocorrelation of \( X_{t-k} \) at lead k. In terms of covariances, \( \text{Cov}(X_n, X_{t-k}) = \text{Cov}(X_{t-k}, X_t) \). A plot of \( \rho_k \) against the lag length, k, also called the autocorrelation function, is sometimes called the correlogram.

Because the practitioner has access only to the sample data, the sample ACF (as opposed to the theoretical ACF) is calculated. Given observations \( X_t, t = 1, 2,..., T \), the sample autocorrelations are calculated as follows:

\[
\rho_k = \frac{\sum_{t=1}^{T-k} (X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^{T} (X_t - \bar{X})^2} \quad \text{for } k = 1, 2, 3,... \quad (2.2.1)
\]

In practice, it is the sample ACF, in large measure, that is the tool used for identifying the characteristics of the linear models chosen to approximate the underlying stochastic process.

Chapter 2 - Univariate Time Series Analysis and Modeling
2.3 Linear Model Representations

Introduction

In order to simplify the problem of having to infer the joint distribution from the sample observations, we abandon describing the joint distribution in favor of describing the generating mechanism of the sample observations. In a 1954 paper, Wold provided the justification for using linear models to approximate the generating mechanism any stationary stochastic process (Granger and Newbold 1986). His theorem states that any stationary stochastic process \( X_t \) can be decomposed uniquely into the sum of two separate, uncorrelated processes, specifically, \( X_t = D_t + Y_t \), where \( D_t \) is linearly deterministic and \( Y_t \) is an infinite moving average process that is purely nondeterministic.\(^{18}\) The upshot of Wold’s Theorem is that even if the \( X_t \) process is generated nonlinearly, a linear decomposition exists, and hence the process can be fully characterized by second moments. Although it may seem that the class of linear models is a restrictive class with which to approximate a stochastic process, Wold’s Theorem provides reassurance that a linear model will be a useful approximation to a process that may either be linear or not.

With the assumptions of stationarity and ergodicity, and following Wold’s Theorem, autoregressive moving average models can be built using the data realization at hand. Using the autocorrelation function, these models can be iden-
tified based on the fact that certain theoretical models yield certain theoretical autocorrelation functions. Hence, if we examine the sample autocorrelation function (2.2.1) we can work back to the type of model that would in theory produce such a function. The remainder of this section is devoted to investigating how particular linear models such as pure moving average, pure autoregressive, and mixed autoregressive moving average produce distinct signatures in terms of their covariance structures. These signatures are represented uniquely in the autocorrelation function. We can then attempt to match the sample ACF to the theoretical ACF and hence to the theoretical model to which it most likely corresponds.

Moving Average Models

A moving average model describes a structure in which contemporaneous and past stochastic error terms are the mechanism which generate the time series. The model is written as follows:

\[ X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} \]  

(2.3.1)

The moving average model is denoted by the shorthand MA(q), which is read "moving average model of order q." The \( \varepsilon_t \) are the random disturbances that drive the model, which are assumed to be generated by a white noise process. The term white noise is borrowed from the engineering literature and is a compact

19 Unless otherwise specified, it will be understood that the process \( X_t \) is a deviation from its mean and has no other deterministic components.
way of saying that the disturbances are a sequence of independent and identically distributed random variables with zero mean and constant variance.

A moving average model might be a useful approximation for a commodity market that receives news concerning harsh weather that is threatening a large portion of the supply. This random shock may be distributed over time as the market assesses the prospective damage, and the consequent impact on price.  

The MA(q) process is fully described by $q + 1$ parameters: $\sigma^2, \theta_1, \ldots, \theta_q$. A finite moving average process is always stationary, whereas other linear processes require some restrictions to be placed on their parameters to insure the process to be stationary. Somewhat akin to the stationarity conditions just mentioned, are the invertibility conditions for a MA process. This assures that the moving average process can be written as a purely autoregressive process.

To describe the invertibility conditions it is convenient to use the backshift operator (see footnote page 4). The backshift operator will be used to characterize large linear processes in compact form, and also to write the characteristic equation of the MA process (3.3.1) in a compact form. Once we express the

---

20 This example was borrowed from Granger and Newbold (1986, p. 24).

21 If we do not remove the mean prior to estimation, the MA(q) is described by $q + 2$ parameters, the additional parameter corresponding to the mean.

22 If a stochastic process $X_t$ can be written in the form $X_t = \sum_{j=0}^{q} \phi_j E_{t-j}$, then that process is said to belong to the general class of discrete linear processes, where $E_t$ is white noise and $\phi_j$ are parameters. From integral calculus, the condition $\sum_{j=0}^{q} \phi_j < \infty$ will insure that the process has finite variance. Since the MA(q) process has a finite number of terms in its sum, this condition will always hold making MA(q) always stationary (Nelson 1972).

23 The backshift operator can also be employed in polynomial form. The polynomial $a_0 X_t + a_1 X_{t-1} + a_2 X_{t-2} + \ldots + a_p X_{t-p}$ can be summarized in compact form as $a(B)X_t$ where $a(B) = a_0 + a_1 B + a_2 B^2 + \ldots + a_p B^p$.

24 The characteristic equation is a polynomial equation in the backshift operator which can be solved for values of $B$, which then imply restrictions of the coefficients to insure the invertibility of the model. See Box and Jenkins (1976).
characteristic equation, employing B as a dummy variable for which we can obtain numerical values, the invertibility conditions for the general MA(q) model can be expressed. For invertibility, the roots of the characteristic equation of the MA process, \( \theta(B) = 0 \), must all lie outside the unit circle; i.e., the solutions \( B_1, B_2, \ldots, B_q \) to the characteristic equation must all be greater than 1 in absolute value.

**Example: The MA(1) process**

The MA(1) process is the simplest moving average process and it is written as follows:\(^{25}\)

\[
X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}
\]

(2.3.2)

The variance is found by squaring the right hand side of (2.3.2):

\[
\gamma_0 = E[(\varepsilon_t + \theta_1 \varepsilon_{t-1})^2]
\]

\[
= E[(\varepsilon_t + \theta_1 \varepsilon_{t-1})(\varepsilon_t + \theta_1 \varepsilon_{t-1})]
\]

\[
= E[\varepsilon_t^2] + \theta_1 E[\varepsilon_t \varepsilon_{t-1}] + \theta_1 E[\varepsilon_{t-1} \varepsilon_t] + \theta_1^2 E[\varepsilon_{t-1}^2]
\]

Here, \( E[\varepsilon_t^2] \) and \( E[\varepsilon_{t-1}^2] \) denote the variance of the white noise process, and \( E[\varepsilon_t \varepsilon_{t-1}] \) denotes the autocovariance at lag 1 of the white noise process. By assumption, \( E[\varepsilon_t \varepsilon_{t-1}] = 0 \) and \( E[\varepsilon_t^2] = E[\varepsilon_{t-1}^2] = \sigma^2 \), so that:

\[
\gamma_0 = (1 + \theta_1^2) \sigma^2
\]

(2.3.3)

In like manner, the autocovariances can be found as follows:

\[
\gamma_1 = E[(X_t X_{t-1})]
\]

\[
\gamma_1 = E[(\varepsilon_t + \theta_1 \varepsilon_{t-1})(\varepsilon_{t-1} + \theta_1 \varepsilon_{t-2})]
\]

\[
= E[\varepsilon_t \varepsilon_{t-1}] + \theta_1 E[\varepsilon_t \varepsilon_{t-2}] + \theta_1 E[\varepsilon_{t-1} \varepsilon_{t-2}] + \theta_1^2 E[\varepsilon_{t-1} \varepsilon_{t-2}]
\]

\[
= \theta_1 \sigma^2
\]

(2.3.4)

\[
\gamma_k = 0 \quad \text{for } k = 2, 3, \ldots
\]

(2.3.5)

\(^{25}\) We violate the sign convention in (2.3.1) here for notational ease.
As expected, the variance and covariances are finite and time invariant, and consequently the process is stationary. From the above equations, the autocovariance function can be calculated:

\[ \rho_k = \frac{\theta_1}{1 + \theta_1^2} \quad \text{for } k = 1 \]  
(2.3.6a)

\[ \rho_k = 0 \quad \text{for } k = 2, 3, \ldots \]  
(2.3.6b)

For the MA(1) process, the ACF will have a spike at lag 1 and cutoff to zero thereafter. The value of the spike \( \rho_1 \) will be dependent upon the parameter value \( \theta_1 \).

Building on the method used for the MA(1) model, the ACF of an MA(q) process, can be found. The variance is:

\[ \gamma_0 = E[\epsilon_i^2] = (1 + \theta_1^2 + \ldots + \theta_q^2) \sigma^2 \]
(2.3.7)

The autocovariances are given by,

\[ \gamma_k = (\theta_k + \theta_1 \theta_{k+1} + \ldots + \theta_{q-k} \theta_q) \sigma^2 \quad \text{for } k = 1, 2, \ldots, q \]  
(2.3.8a)

\[ \gamma_k = 0 \quad \text{for } k > q \]  
(2.3.8b)

For example, suppose we want to calculate the autocovariance at lag 5 for, say, an MA(7) process. With \( k = 5 \) and \( q = 7 \), plugging into (2.3.8a) we obtain:

\[ \gamma_5 = (\theta_5 + \theta_1 \theta_6 + \theta_2 \theta_7) \sigma^2 \]

From (2.3.7) and (2.3.8a) and (2.3.8b) the ACF can be calculated for any MA(q) process.
\[
\rho_k = \frac{\theta_k + \theta_1 \theta_{k+1} + \ldots + \theta_{q-k} \theta_q}{1 + \theta_1^2 + \ldots + \theta_q^2} \quad \text{for } k = 1, 2, \ldots, q \tag{2.3.9a}
\]

\[
\rho_k = 0 \quad \text{for } k > q \tag{2.3.9b}
\]

Thus, the signature for the MA(q) process is a generalization of that of the MA(1) process. There are nontrivial values for the first \( k = 1, 2, \ldots, q \) autocorrelations, and then a cutoff thereafter. Thus the order of the process also represents the memory of the process. By examining the ACF of the sample data, an ACF which displays insignificant values for lags \( q^* + 1, q^* + 2, \ldots \) indicates a MA(q*) process.

**Autoregressive Models**

The autoregressive model is denoted AR(p), where \( p \) represents the order of the process. This model explains the process as a combination of lagged own values, and a white noise error term. It is written as follows:

\[
X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \epsilon_t \tag{2.3.10}
\]

An autoregressive model does not automatically satisfy the stationarity conditions as does the moving average model (see footnote 21). Stationarity conditions for the autoregressive model are in the form of restrictions on the values of the parameters, \( \phi_1, \phi_2, \ldots, \phi_p \). The stationarity conditions for the general AR(p) model are described in terms of the characteristic equation of the process.

Chapter 2 - Univariate Time Series Analysis and Modeling
If the roots of the characteristic equation, $\phi(B) = 0$, all lie outside the unit circle, the process is said to be stationary.\footnote{Harvey (1981b) provides a complete account of the stationarity conditions and how they are derived for all AR models.} For example, for a first order process, $AR(1)$, if $|\phi_1| < 1$ the model is stationary.

**Example: $AR(1)$ process**

The $AR(1)$ process is written as follows:

$$X_t = \phi_1 X_{t-1} + \varepsilon_t \tag{2.3.11}$$

It should also be noted that this process can be written as a moving average process of infinite order if $X_{t-1}$ is substituted out recursively ad infinitum with its $AR(1)$ form. The following result is obtained as a consequence:

$$X_t = \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \phi_1^3 \varepsilon_{t-3} + ... \tag{2.3.12}$$

If the stationarity condition for the $AR(1)$ process as defined above holds, this MA sequence will converge and hence be stationary (see Harvey 1981b, p. 29 for a proof). The variance is calculated by squaring the right hand side of (2.3.11):

$$\gamma_0 = E[(\phi_1 X_{t-1} + \varepsilon_t)^2]$$

$$= E[\phi_1^2 X_{t-1}^2 + 2\phi_1 X_{t-1} \varepsilon_t + \varepsilon_t^2]$$

$$= \phi_1^2 E[X_{t-1}^2] + 2\phi_1 E[X_{t-1} \varepsilon_t] + E[\varepsilon_t^2]$$

Because the error term associated with $X_{t-1}$ is $\varepsilon_{t-1}$, which is uncorrelated with $\varepsilon_t$ by assumption ($\varepsilon_t$ is white noise), the middle term is zero, leaving:

$$\gamma_0 = \phi_1^2 \varphi_0 + \sigma_\varepsilon^2 \tag{2.3.13}$$

which gives, upon combining terms,

$$\gamma_0 = \frac{\sigma_\varepsilon^2}{1 - \phi_1^2} \tag{2.3.14}$$

Calculating the covariances:

$$\gamma_1 = E[X_{t-1}X_t] = E[X_{t-1}(\phi_1 X_{t-1} + \varepsilon_t)]$$

$$= E[\phi_1 X_{t-1}^2 + X_{t-1} \varepsilon_t]$$
\[ \gamma_0 = \phi_1 \gamma_0 \]

Substituting for \( \gamma_0 \), the autocovariance at lag 1 is:

\[ \gamma_1 = \phi_1 \frac{\sigma^2}{1 - \phi_1^2} \] \hspace{1cm} (2.3.15)

Similarly, the autocovariance at lag 2 can be calculated:

\[ \gamma_2 = E[X_{t-2}X_t] \]

In order to calculate this quantity, substitute for \( X_t \) in terms of \( X_{t-2} \). This provides a convenient form for the calculation of \( \gamma_2 \):

\[ \gamma_2 = E[X_{t-2}(\phi_1^2X_{t-2} + \phi_1 \epsilon_{t-1} + \epsilon_t)] \]

\[ = \phi_1^2E[X_{t-2}^2] + \phi_1E[X_{t-2}\epsilon_{t-1}] + E[X_{t-2}\epsilon_t] \]

\[ \gamma_2 = \phi_1^2\gamma_0 = \frac{\phi_1^2\sigma^2}{1 - \phi_1^2} \] \hspace{1cm} (2.3.16)

A pattern can now be seen emerging, and in fact the covariance for any k-lag displacement can be written in a general form,

\[ \gamma_k = \phi_1^k \gamma_0 = \phi_1^k \frac{\sigma^2}{1 - \phi_1^2} \] \hspace{1cm} (2.3.17)

This equation provides an intuitive justification for the stationarity restriction. Unless \( |\phi_1| < 1 \), it will be the case that at longer lags there will be higher correlation, and this is not an appealing concept. Intuitively, events in the distant past should have less of an effect on the current value of the series than recent events.

The autocorrelation function can be found as follows, directly from (2.3.14) and (2.3.17):

\[ \rho_k = \frac{\gamma_k}{\gamma_0} = \phi_1^k \frac{\sigma^2}{1 - \phi_1^2} \frac{1 - \phi_1^2}{\sigma^2} \]

\[ \rho_k = \phi_1^k \text{ for } k = 1, 2, \ldots \] \hspace{1cm} (2.3.18)

This ACF, unlike that for the MA(1) process, has an infinite memory that decays exponentially (decays because \( \phi_1 < 1 \)) as \( k \to \infty \). This means that the ACF damps out exponentially for an autoregressive process, and the speed of this damping depends on the parameter values.\(^2\)

\(^2\) If the parameter \( \phi_1 \) is strictly positive, the decay will resemble a damped exponential, however, if \( \phi_1 \) is negative the decay will resemble a damped sine wave. This can readily be seen as even powers of \( k \) in (2.3.18) will produce a positive autocorrelation, whereas for odd values of \( k \) the autocorrelations will be
Building on the procedure and pattern for an AR(1) process, the autocovariances for an AR(p) process may be found. For the AR(1) process \( \gamma_0 = \phi_1 \gamma_1 + \sigma^2 \), equation (2.3.14), \( \gamma_1 = \phi_1 \gamma_0 \), and \( \gamma_2 = \phi_1 \gamma_1 \). Therefore, the AR(p) will have autocovariances as follows:

\[
\begin{align*}
\gamma_0 &= \phi_1 \gamma_1 + \ldots + \phi_p \gamma_p + \sigma^2 \\
\gamma_1 &= \phi_1 \gamma_0 + \ldots + \phi_p \gamma_{p-1} \\
&\vdots \\
\gamma_p &= \phi_1 \gamma_{p-1} + \ldots + \phi_p \gamma_0
\end{align*}
\]

(2.3.19a) (2.3.19b) (2.3.19c)

Given parameters \( \phi_1 \ldots \phi_p \) and \( \sigma^2 \), the above equations can be solved for the unknown \( \gamma_0 \ldots \gamma_p \). Because \( \phi_1 \ldots \phi_p \) and \( \sigma^2 \) are unknown, estimates can be substituted which are derived from sample data, and consequently estimates of the unknown \( \gamma_0 \ldots \gamma_p \) can be estimated from (2.3.19a-c). Note that if the final p equations are divided through by \( \gamma_0 \), we obtain equations which are expressed in terms of the autocorrelations and hence can be solved for the autocorrelations. Such a system of equations in terms of the autocorrelations is called the Yule-Walker system (see Box and Jenkins 1976, p.55).

The ACF of any higher order AR processes will then be solutions of the Yule-Walker equations plotted against lags 1,2,...,k. The ACF of any AR process

\text{negative. This yields an ACF that alternates in sign as it damps out exponentially resembling a damped sine wave.}
will then resemble a damped exponential or sine wave making it virtually impossible to distinguish among AR processes of higher orders by viewing the sample ACF alone. There is, however, a "tool" that can be employed to help the practitioner distinguish among various AR processes. This "tool" is called the partial autocorrelation function which will be introduced in section 2.6.

**Autoregressive Moving Average Models**

There are instances in which stationary stochastic processes cannot be modeled parsimoniously (i.e., with as few parameters as possible) as only pure autoregressive or pure moving average. Consequently, models can be constructed that have both autoregressive and moving average components. The general form of the ARMA(p,q) process is expressed as follows:

\[ X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \ldots - \theta_q \epsilon_{t-q} \]  

(2.3.20)

The stationarity of a mixed process depends on its autoregressive part only, i.e., that the roots of the equation \( \phi(B) = 0 \) all lie outside the unit circle. The invertibility of a mixed process depends solely on the moving average part of the process.\(^{28}\) The conditions for invertibility, then, are the same for the ARMA(p,q) as for the MA(q), i.e., the roots of the characteristic equation \( \theta(B) = 0 \) must all lie outside the unit circle.

---

\(^{28}\) For a proof, see Harvey (1981b) p. 37-39.
Example: ARMA(1,1) process

The ARMA(1,1) process is written in the following form:

\[ X_t = \phi_1 X_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1} \]  

(2.3.21)

The ARMA(1,1) process can be represented as either a purely autoregressive process of infinite order by successively eliminating \( \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots \), etc, or as a purely moving average process of infinite order by successively eliminating \( X_{t-1}, X_{t-2}, \ldots \), etc. This results, in either case, in many more parameters to be estimated than in the simpler ARMA model, and hence, an inefficient use of the data.

The moments of the ARMA(1,1) process can be calculated directly. The interesting dimensions of this model are present in its covariance structure which incorporates features of both AR and MA processes. The variance is calculated by squaring the right hand side of (2.3.21):

\[ \gamma_0 = \text{Var}[X_t] = E[(\phi_1 X_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1})^2] \]

\[ = E[(\phi_1^2 X_{t-1}^2 + \varepsilon_t^2 + \theta_1^2 \varepsilon_{t-1}^2 - 2 \phi_1 \varepsilon_t X_{t-1} - 2 \phi_1 \theta_1 X_{t-1} \varepsilon_{t-1} - 2 \theta_1 \varepsilon_t \varepsilon_{t-1})] \]

\[ = \phi_1^2 \gamma_0 + \sigma_\varepsilon^2 + \theta_1^2 \sigma_\varepsilon^2 - 2 \phi_1 \theta_1 \sigma_\varepsilon^2 \]

\[ \gamma_0 = \frac{(1 + \theta_1^2 - 2 \phi_1 \theta_1) \sigma_\varepsilon^2}{1 - \phi_1^2} \]  

(2.3.22)

The autocovariance at lag 1 is calculated similarly:

\[ \gamma_1 = E[X_t X_{t-1}] = E[X_{t-1}(\phi_1 X_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1})] \]

\[ = \phi_1 E[X_{t-1}] + E[X_{t-1} \varepsilon_t] - \theta_1 E[X_{t-1} \varepsilon_{t-1}] \]

\[ = \phi_1 \gamma_0 - \theta_1 \sigma_\varepsilon^2 \]

Substituting for \( \gamma_0 \) and rearranging:

\[ \gamma_1 = \phi_1 \left[ \frac{1 + \theta_1^2 - 2 \theta_1 \phi_1}{1 - \phi_1^2} \right] \sigma_\varepsilon^2 - \theta_1 \sigma_\varepsilon^2 \]

\[ = \left[ \frac{\phi_1 + \theta_1^2 \phi_1 - 2 \theta_1 \phi_1^2}{1 - \phi_1^2} \right] \sigma_\varepsilon^2 - \theta_1 \sigma_\varepsilon^2 \]

\[ \gamma_1 = \left[ \frac{(1 - \phi_1 \theta_1)(\phi_1 - \theta_1)}{1 - \phi_1^2} \right] \sigma_\varepsilon^2 \]  

(2.3.23)

The autocovariance at lag 2 is as follows:

\[ \gamma_2 = E[X_t X_{t-2}] + E[X_{t-2}(\phi_1 X_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1})] \]
\[ \gamma_2 = \phi_1 \gamma_1 \]  
Similarly, for all lags greater than 2:

\[ \gamma_k = \phi_1^{k-1} \gamma_1 \]  

(2.3.25)

To compare the ARMA(1,1) with the AR(1) and the MA(1) the ACF is calculated:

\[ \rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{(1 - \phi_1 \theta_1)(\phi_1 - \theta_1)}{1 + \theta_1^2 - 2\theta_1 \phi_1} \]  

(2.3.26a)

\[ \rho_k = \phi_1 \rho_{k-1} \quad \text{for} \quad k \geq 2 \]  

(2.3.26b)

Thus the ACF begins at a starting value equal to \( \rho_1 \) which is a function of both \( \phi_1 \) and \( \theta_1 \), and then decays geometrically from this value as \( \phi_1 \). This reflects that the MA part of the process has a memory of only one period while the AR part has an infinite memory.

This pattern extends to higher order ARMA(p,q) processes. The general pattern for the autocorrelation function is that the first q autocorrelations are functions of both the moving average and autoregressive parameters. Higher order autocorrelations are given strictly by the autoregressive portion of the process.

2.4 Nonstationarity

Introduction

The linear models previously introduced have been constructed under the assumption of stationarity, which allows the time averaging of the series for the purpose of estimating the mean, and hence allows for estimating the various covariances of interest. Again, stationarity requires that the mean and variance
of the series be finite and time invariant, and the covariance between any two values of the process be dependent solely on the distance between these values in time and not on time itself.

In practice, the levels of many economic time series appear to be nonstationary (Nelson and Plosser 1982; Wasserfallen 1986). Hence, in order to apply linear models such as ARMA models, a time series must be tested for and possibly transformed to stationarity. A useful starting point is to examine a time plot of the raw data series. If the series exhibits fluctuations that are more violent for a particular segment of the series than for others, the series very likely is variance nonstationary, i.e. there is not a constant variance throughout the series (condition (2.1.5b) is violated). The most common method for inducing variance stationarity in a series is to take the natural logarithm of each observation. This transformation will reduce the swings of the levels which constitute the variance nonstationarity and often yield a series that is a good approximation to one having constant variance.

Nonstationarity in Mean, Trends, and Integrated Processes

A more insidious form of nonstationarity, however, is nonstationarity in the mean of the series. In this case, the series shows no propensity to return to, or move around, a particular fixed level. When a series has this lack of affinity for a mean, and the movement seems to be in a particular direction, the series is often said to exhibit a "trend". In this study, "trend" will be reserved for a deterministic functional dependence on time. For example, consider a series that has two parts,
a deterministic linear trend and a residual representing the stationary component which includes all of the interesting variation that we wish to model:

\[ X_t = \alpha + \beta t + \varepsilon_t \]  

(2.4.1)

By estimating (2.4.1) as a linear regression model, the residual \( \varepsilon_t \) is often treated as a stationary series with well defined variance, covariances, and autocorrelations (Nelson 1984). This would mean that \( \varepsilon_t \) could be modeled as an ARMA process after the trend is removed from \( X_t \). The function of time need not be linear, however, thus we consider the more general representation:

\[ X_t = f(t) + \varepsilon_t \]  

(2.4.2)

The relation in (2.4.2) is called a trend stationary process (TSP); \( X_t \) is stationary around the trend function. It is important to note that (2.4.2) is just one hypothesis concerning the manifestation of nonstationarity, and indeed there are problems with it. Even if we could know for sure that the variable \( X_t \) is a TSP, there is little chance that the actual functional form could ever be accurately specified. If the nonstationarity is not correctly modeled, the residuals in (2.4.2) will not be stationary. In addition, over the course of a time series, we may observe local upward trends followed by local downward trends. Consequently a global OLS trend line will offer a poor representation of the nonstationarity.

An alternative hypothesis about the way nonstationarity in mean arises was introduced by Box and Jenkins (1970). They view nonstationarity not as a manifestation of deterministic functions of time, but as the accumulation of ran-
dom shocks. In this case, the first differences of the series are stationary. This kind of process takes the form:

\[ X_t = X_{t-1} + D + \epsilon_t \]  

(2.4.3)

In this relation the series \( \epsilon_t \) is a stationary process, and \( D \) is the fixed mean of the first differences which is often called the drift parameter. The level of the series at any given time \( t \) is equal to the previous level plus the drift plus the random shock. The series is cumulative, or additive, in its level. This additivity exhibits itself as an apparent trend. Equation (2.4.3) is consequently said to belong to the difference stationary class of processes (DSP).

DSPs are also called "integrated" processes, the word "integrated" reflecting the additive nature of the series. The following three definitions will be useful in following sections (from Granger 1986, p.216):

**Definition 2.4.1a.** If a time series \( Z_t \) needs no differencing to become stationary, it is called integrated of order zero which is denoted \( Z_t \sim I(0) \).

**Definition 2.4.1b.** If a time series \( Z_t \) must be differenced \( d \) times to become \( I(0) \), it is called integrated of order \( d \) which is denoted \( Z_t \sim I(d) \).

**Definition 2.4.1c.** Let \( \Delta^b \) represent \( b \) applications of the difference operator. If \( Z_t \sim I(d) \) then the \( b^{th} \) differenced series is \( \Delta^b Z_t \sim I(d - b) \).

---

29 Difference requires subtracting previous values of the observations from one another. First differencing, for example, means the subtraction of the first observation from the second, the second from the third, etc. Second order differencing takes the result from the first difference and repeats the procedure of subtracting the preceding value. Note that each time the series is differenced, the number of observations is decreased by one.

30 The simplest member of this class of processes is the random walk where \( \epsilon_t \) would be a white noise process, and the drift would be zero.
Testing for Unit Roots

Dickey (1975), Fuller (1976), and Dickey and Fuller (1979, 1981) have developed a series of tests (henceforth DF tests) for discriminating between the hypotheses that a given time series belongs to the DSP class against the alternative that it belongs to the TSP class. Their tests only entertain a DSP that is I(1). The procedure is to perform OLS on the model:

\[ X_t = \alpha + \beta t + \rho X_{t-1} + \varepsilon_t \]  

(2.4.4)

The null hypothesis for the first test, \( \tau_\rho \), is that \( \rho = 1 \), or that \( X_t \) contains a unit root and is nonstationary, with an alternative model that the series is generated by a stationary autoregression (\( \rho < 1 \)) with drift. In other words, to test this hypothesis we run OLS on (2.4.4) and exclude the time parameter term, \( \beta t \). The null hypothesis for the second test, \( \tau_\nu \), is again that \( \rho = 1 \), with an alternative that the series is generated by a stationary autoregression around a linear time trend (drift plus a time parameter). For this hypothesis test, the time parameter \( \beta t \) is included in the estimated equation. Fuller (1976 p.373) has provided critical values for \( \tau_\rho \) and \( \tau_\nu \), both of which are "t-ratios", \( (\rho - 1)/\sigma_\rho \), that follow non-standard distributions. The rejection regions are given by small values of \( \tau_\rho \) or \( \tau_\nu \).

Dickey and Fuller also describe two likelihood ratio tests for the joint null hypothesis of a simple random walk. In the first of these tests, \( \Phi_1 \), the null hypothesis is \( (\alpha, \rho) = (0,1) \) in a model that is assumed not to include a time parameter. In the second test, \( \Phi_2 \), the null hypothesis is \( (\alpha, \beta, \rho) = (0,0,1) \) in a model
that may have a linear time trend. Finally, they describe a likelihood ratio test for the joint-null hypothesis of a random walk with drift, \( (\alpha, \beta, \rho) = (\alpha, 0, 1) \), in a model that again includes a time parameter. The rejection regions are for large values of the test statistics, and critical values are found in Dickey and Fuller (1981 p.1069). Again, each of these test statistics follows a nonstandard distribution.

A question that arises with the DF test is whether it is appropriate to model \( X_t \) as AR(1) (or a random walk) as the error, \( \varepsilon \), in (2.4.4) may not be empirical white noise. For example, if there is evidence of moving average behavior a higher order autoregression may be needed to approximate the dynamics of the \( X_t \) process. Consequently, a more general model is often fit. This results in an augmented Dickey-Fuller test (ADF) based on the model:

\[
X_t = \alpha + \beta t + \rho X_{t-1} + \sum_{i=1}^{p} \phi_i \Delta X_{t-i} + \varepsilon_t
\]  

(2.4.5)

where lags of \( \Delta X_t \) are added until \( \varepsilon \), is made empirical white noise.\(^{31}\) The hypotheses to be tested are the same for this specification as for model (2.4.4). Fuller (1976) and Dickey and Fuller (1981) show that their tests \( \tau_\mu, \tau_\phi, \Phi_1, \Phi_2, \) and \( \Phi_3 \) also apply for higher order autoregressions.

A somewhat less formal test for stationarity requires that the sample ACF of the time series die out quickly to insignificance for a stationary series. A useful procedure is to employ both the sample ACF and a Dickey-Fuller test (2.4.5) as

\(^{31}\) See section 2.6 on diagnostic checking for tests of the white noise hypothesis.
tests for stationarity. Nelson and Kang (1984) have shown that many economic time series can be considered members of the DSP class, hence differencing should remove the nonstationarity and one can proceed with the fitting of an ARMA model.\textsuperscript{32}

2.5 Seasonality

Introduction

Economic time series often exhibit a tendency to be relatively high or low in particular parts of the sample period, the typical example being the level of retail sales at Christmas time. If such a tendency is observed consistently from year to year, it is called seasonal. Seasonal variation contributes a large percentage of the total variation in some economic time series, and there is a controversy among econometricians as to how one should deal with seasonal series. It is often argued that this seasonal variation is not in itself of interest as it may not contain any economically meaningful information (Granger and Watson 1984). Whether one believes this to be true may depend on the series in question, however, assuming this is an accurate account of seasonality we may want to adjust the series to remove the variation that is directly attributable to seasonal factors.\textsuperscript{33}

\footnotetext{32}{An ARMA model fitted to a $d$-differenced series is known as an ARIMA($p,d,q$) model. The "I" in ARIMA stands for "integrated" because the first-differences can be summed or integrated to get back to the original pre-differenced series. This is useful to note because with forecasting problems, the practitioner wants to make forecasts in terms of the levels of the series, rather than the differences.}

\footnotetext{33}{Granger (1978) also makes the argument that two time series which are not related could both have sea-}
Another approach to seasonality is to try and model its effect on a particular variable. By choosing this approach one attempts to understand not only why a particular variable displays seasonal variation, but precisely how that seasonality is generated. If we understand the causes of seasonality and have a plausible modeling strategy, modeling the seasonal component directly may be the best approach.\(^4\) It is generally recognized, however, that we do not fully understand how to model seasonality directly, though there are many hypotheses (Granger 1978). However, though we may not understand how to model the seasonal variation, it is observable, and there exist techniques for filtering out the seasonality from a time series.\(^5\)

**Identifying a Seasonal Pattern**

A simple and direct way to detect seasonality is to examine the sample ACF of the series in question. Large, significant autocorrelations will appear and persist at the seasonal lags (12, 24, 36,... for monthly data, and 4, 8, 12,... for quarterly data) if seasonal patterns exist.\(^6\) Nonstationarity in the data can often mask seasonal patterns in the levels, and an examination of an appropriately differenced (or detrended) series may prove more useful. This is the approach that will

---

\(^4\) The Box and Jenkins ARIMA approach favors assuming a particular structure for the seasonality and including it directly in the model. However, as Pierce (1984) points out, the Box and Jenkins approach is neither the only way nor the most widely accepted procedure.

\(^5\) For those series that have unstable seasonal patterns resulting from, say, weather, it is possible that neither a modeling approach nor an adjustment approach will properly account for (or remove) such effects.

\(^6\) Sims (1974) notes that in some cases detecting seasonality in this way may lead one to think none exists when, however, it is masked in some way or is evolving through time.
be used in this study. Sims (1974) advocates using spectral techniques as a more powerful alternative to viewing the ACF, but this approach is beyond the scope of this work.

**Seasonal Adjustment Techniques**

The aim of seasonal adjustment procedures is to remove some of the excess variation at the seasonal lags, thereby "smoothing" the observed series. If such a task is performed well, the autocorrelations at the seasonal lags will not appear as anomalous. If such a procedure over-corrects for seasonality, negative autocorrelations may be observed at the seasonal lags thereby replacing one kind of seasonality (high positive correlations), with another kind of seasonal pattern (negative correlation).

Many methods for seasonal adjustment now exist, and we shall consider three widely used methods as alternatives. The first method to be considered is the regression approach which removes seasonal means. This familiar approach is carried out, in this instance for monthly data, by performing OLS on the model:

$$X_t = \delta_0 + \delta_2 Y_2 + ... + \delta_{12} Y_{12} + W_t$$  \hspace{1cm} (2.5.1)

where $Y_i$ are dummy variables which take on values of 1 in month $i$ and values of 0 otherwise. The residual $W_t$ is then treated as the seasonally adjusted series. The advantage to such an approach is that hypotheses can be tested about the nature of the seasonality by examining the estimated coefficients. An obvious
drawback, however, is that the errors will probably (we hope) be estimable as some ARMA process, meaning that equation (2.5.1) will suffer serial correlation in \( W_r \). Because of this, it may be wise to include the seasonal dummies directly in the equations we estimate rather than deseasonalize prior to estimation.

A second method often used to deseasonalize data is the method of exponential smoothing. This method expresses the smoothed time series as an exponentially-declining weighted function of past values of the raw time series:

\[
\overline{X}_t = \gamma (1 - \gamma) X_{t-1} + \gamma (1 - \gamma)^2 X_{t-2} + \gamma (1 - \gamma)^3 X_{t-3} + \ldots, \quad 0 < \gamma < 1
\] (2.5.2)

where \( \overline{X}_t \) is the smoothed series. With some algebraic manipulation (see Granger and Newbold 1986 p.166), the smoothed series can be calculated using the following algorithm:

\[
\overline{X}_t = \gamma X_t + (1 - \gamma) \overline{X}_{t-1}, \quad 0 < \gamma < 1
\] (2.5.3)

Obviously, this algorithm is not self-starting in that there is no initial value for \( \overline{X}_1 \). Granger and Newbold (1986) suggest starting off with \( \overline{X}_1 = X_1 \), then (2.5.3) can be used to recursively calculate the rest of the smoothed series. This smoothed series is then used to estimate the seasonal component of the series, which is assumed to be either an additive or multiplicative component of the overall series. If the data are expressed in log form we naturally assume the seasonal is additive, and the following algorithm can be used to find the seasonal factors:

\[
S_t = \beta (X_t - \overline{X}_t) + (1 - \beta) S_{t-1}
\] (2.5.4)
where $\beta$ is the seasonal smoothing weight, and $l$ is the seasonal lag, $l = 12$ for monthly data. The deseasonalized level, $\bar{X}'_t$, then, is estimated with the following equation:

$$\bar{X}'_t = \phi(X_t - S_{t-l}) + (1 - \phi)\bar{X}_{t-1}'$$  \hspace{1cm} (2.5.5)

The "ad-hocness" of the exponential smoothing approach is due to the arbitrary selection of the smoothing weights, $\gamma, \beta, \phi$. Values of the parameters close to zero provide the most smoothing, and values close to 1 the least." Again, the algorithms (2.5.4) and (2.5.5) require starting-up values (see Granger and Newbold (1986) p.167 for details). Exponential smoothing also requires assumptions about the trend properties of the data; in light of the discussion in section 2.4 it is probably useful to remove the trend or difference the data prior to smoothing seasonal variation.\textsuperscript{38}

The last method we shall discuss is the official method of seasonal adjustment developed and used by the U.S. Bureau of the Census. This method is often called a ratio-to-moving average method. The first step is to take a two-sided moving average of the original series (again a starting value problem arises) divide the original series by the smoothed series to get an initial estimate of the seasonal component, and then adjust this estimate so that the sum of the seasonally adjusted series is equal to the sum of the original series. This official method is much like the exponential smoothing method described above. In

\textsuperscript{37} Doan and Litterman (1987) suggest values around 0.3.

\textsuperscript{38} For a detailed discussion of the various methods available for exponential smoothing see Gardner (1985).
many ways, however, it is more complex as it is capable of dealing with (usually removing) outliers, and correcting for trading day variation. The use of moving averages, and the removal of outliers makes the procedure non-linear (Judge et al. 1985). This method is used widely, especially by the U.S. government.

2.6 Specifying ARIMA Models

Introduction

The fitting of ARIMA models to time series data is often referred to as an art or an acquired skill, so that no adequate treatment of the techniques involved in fitting ARIMA models to sample data can be given in a summary. There exist many comprehensive introductions to fitting ARIMA models (e.g., Pankratz 1983; McCleary and Hay 1983; Vanadale 1983; Granger and Newbold 1986), and the interested reader is referred there. In this section, a cursory introduction to the methodology of univariate time series modeling is provided. The steps involved in modeling are identification, estimation, and diagnostic checking, and these steps are repeated until the diagnostic checks yield white noise residuals indicating that the history of the process has been adequately captured. This modeling procedure is due to Box and Jenkins (1976).

Identification

After the appropriate order of differencing and the applicability of any
other stationarity transformations are determined and applied to the sample data, the second step in the identification process can begin. The sample autocorrelations are estimated using the formula in (2.2.1), and are plotted against lags \( k = 1,2,\ldots,T/4 \). Autocorrelations at lags greater than \( T/4 \) are calculated with progressively fewer and fewer observations, and consequently may be poor estimates of the population autocorrelations.\(^{39}\)

For a pure MA(q) process, the sample autocorrelations should become statistically insignificant after lag \( q \), since they approximate the theoretical autocorrelations that are equal to zero beyond lag \( q \). Because the sample ACF is an estimate of the theoretical ACF, a guide to the statistical significance of the estimates is required. An approximate expression for the variance of an autocorrelation, assuming that the generating process is Gaussian and stationary, is given by Bartlett (1946):

\[
\text{Var}(\rho_k) \approx \frac{1}{T} \sum_{s=-\infty}^{+\infty} \left[ \rho_s^2 + \rho_{s+k} \rho_{s-k} - 4 \rho_k \rho_s \rho_{s-k} + 2 \rho_s^2 \rho_k^2 \right]
\]  \( (2.6.1) \)

Because the population autocorrelations are unknown, the sample autocorrelations are substituted into a variant of (2.6.1):

\[
\text{Var}(r_k) \approx \frac{1}{T} \left[ 1 + 2 \sum_{i=1}^{q} r_i^2 \right] \quad k > q
\]  \( (2.6.2) \)

\(^{39}\) The stopping point \( T/4 \) is a common rule of thumb.
This formula is called the large sample standard error of \( r_k \), and it gives a bound on the statistical significance of autocorrelations at lag \( k \) beyond some value \( q \) after which the theoretical ACF is assumed to have died out. Theoretically, one could calculate confidence bounds for every \( r_k \) using the formula (2.6.2), a rather cumbersome task. Computer programs often estimate standard errors for groups of autocorrelations such that \( q \) in (2.6.2) takes on the successive values of 0,12,24,...,T/4.

A result of Anderson (1942), that in moderately large samples the generating process can be assumed to be a normal process, is useful because we then can construct confidence bands of plus or minus two standard deviations about zero to provide a rough guide to the significance of the sample autocorrelations (Granger and Newbold 1986). Following this approach we calculate \( 1.96/\sqrt{T} \) which yields (asymptotically) a 95% confidence interval inside which a sample autocorrelation is approximately zero.\(^{40}\)

As noted earlier, the ACF is not particularly useful for distinguishing among various pure autoregressive models. However, there is a function that can help to distinguish among the various orders of AR processes. This function, which can be estimated from the sample autocorrelations, is called the partial autocorrelation function (PACF). The concept behind using the PACF to select among various AR orders involves considering the implications of adding an additional lagged value of the process to an existing autoregression. For example, an autoregression of order \( p - 1 \) is written as follows:

\(^{40}\) In the limit the standard error of any correlation coefficient is equal to \( 1/\sqrt{T} \).
\[ X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_{p-1} X_{t-p-1} \]  \hspace{1cm} (2.6.3)

By adding one more lag of \( X_t \), \( X_{t-p} \), to (2.6.3), we obtain:

\[ X_t = \phi_1 X_{t-1} + \ldots + \phi_{p-1} X_{t-p-1} + \phi_p X_{t-p} \]  \hspace{1cm} (2.6.4)

The coefficient \( \phi_p \) measures the excess correlation not accounted for by the first \( p-1 \) coefficients. It can then be said that \( \phi_p \) measures the partial effect that \( X_{t-p} \) has in "explaining" the variation in \( X_t \) given all the lagged \( X_t \) up to \( X_{t-p-1} \) (Vanadacele). The highest order AR coefficient in these regressions of \( X_t \) on its lagged values is then defined as the partial autocorrelation at lag \( p \). Therefore, to calculate the PACF over lags \( k = 1, 2, \ldots, p \), we could run \( p \) regression equations, each with one more lagged \( X_t \) variable on the RHS than the previous equation, with the coefficient on the largest lag in each equation being the partial autocorrelation at that lag. Partial autocorrelations will be denoted \( \rho_{kk} \). Sample partial autocorrelations will be denoted \( r_{kk} \). The same 95% confidence bands which were plotted around the ACF can be used for testing the significance of the sample partial autocorrelations. Using this tool we expect the PACF for an AR(p) process to cutoff after lag \( p \). Therefore, the PACF for an AR(p) process behaves similarly to the ACF for an MA(q) process. We should expect that an AR process will have an ACF that decays exponentially and a PACF that cuts off after lag \( p \). On the other hand, an MA process will have a PACF that decays exponentially and an ACF that cuts off after lag \( q \).
Durbin (1960) has provided a recursive formula for the calculation of the sample partial autocorrelations from the sample autocorrelations (Box and Jenkins 1976):

\[
r_{k+1} = \frac{r_{k+1} - \sum_{j=1}^{k} r_{k-j}r_{k+1-j}}{1 - \sum_{j=1}^{k} r_j^2} \quad \text{for } k = 0, 1, 2, 3, \ldots \quad (2.6.5)
\]

This formula provides an alternative to the recursive regressions for estimating the PACF.

A further difficulty in identifying univariate time series models comes in identifying the mixed ARMA process. Because the mixed process has both an infinite MA and an infinite AR representation both the ACF and PACF will tail off to zero slowly. However, it is also known that in the mixed process the first \( q \) autocorrelations are functions of both MA and AR parameters, which give rather irregular patterns, so that after lag \( q \) a more regular decay to zero should be observed. A similar pattern exists for the first \( p \) partial autocorrelations. These expected patterns can be used as guides in identification. However, because we are working with estimates, the sample ACF and PACF will not always give exact guides, or look precisely as we might expect. This is where experience helps to guide the model builder. The iterative identification-estimation-diagnostic checking process recognizes the uncertainty at the identification stage, and by
repeating the process and discarding models that do not have white noise residuals it is hoped that the iterations converge on a model that is satisfactory.

**Estimation**

The identification stage will produce one or more tentative models which need to be estimated. For the pure AR(p) model, OLS is easily applied. However, for MA(q) and ARMA(p,q) models, the presence of MA terms implies that a nonlinear estimation method is necessary.

In this section, maximum likelihood estimates will be derived. These estimates will be consistent and asymptotically efficient provided that the sample is reasonably large (> 50). In constructing an ARIMA model, we assume that the data observations are outcomes of the random variable being modelled with the ARIMA and its unknown parameters. Thus, we can claim that the observations are drawn from the joint probability distribution \( P(X_t | \phi, \theta, \sigma^2) \) where \( X_t \) denotes the sequence of observations on the random variable \( X_t \). The method of maximum likelihood poses the following question: Given the observations on the random variable \( X_n \), what are the values of the parameters that are most likely to have generated them? The next step is to devise a function that measures "likelihood". The likelihood function, \( L(\phi, \theta, \sigma^2 | X_t) \), which takes the same form as the probability density for the random variable \( X_t \) is a measure of this likelihood. To obtain the form of the likelihood function, we must assume a particular distributional form for the probability function, \( P(X_t | \phi, \theta, \sigma^2) \). A useful assumption is that the disturbances in the ARIMA model we wish to estimate are
\[ p(\varepsilon_i \mid \sigma_s^2) = \frac{1}{\sqrt{2\pi\sigma_s^2}} \exp \left( -\frac{\varepsilon_i^2}{2\sigma_s^2} \right) \]  

\[ (2.6.6) \]

Because the disturbances are assumed to be white noise and hence are independent, their joint distribution is a product of each disturbance's marginal distribution (2.6.6) as follows:

\[ P(\varepsilon_1, \ldots, \varepsilon_T \mid \sigma_s^2) = (2\pi)^{-T/2} (\sigma_s^2)^{-T/2} \exp \left( -\frac{1}{2\sigma_s^2} \sum_{i=1}^{T} \varepsilon_i^2 \right) \]  

\[ (2.6.7) \]

The Conditional Likelihood for an ARMA process

Assume that we have \( T \) observations on the random variable \( X_t \) and that \( X_t \) is a stationary time series. Assume as well that there exist one or more tentatively identified ARMA(p,q) models that are to be estimated from the sample data. Rewriting (2.3.21), we can express the general ARMA model as follows:

\[ \varepsilon_t = X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} \]  

\[ (2.6.8) \]

It can be seen that the observations cannot simply be substituted into (2.6.8) to obtain values for the \( \varepsilon_t \) disturbances even if we knew the parameter values. This
is because at time \( t = 1 \) there exist no values for observations \( X_{t-1}, X_{t-2}, \ldots, X_{t-p} \). Neither, then, can values for \( \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \varepsilon_{t-q} \) be estimated. This is an initial value problem. Suppose, however, that \( p \) values of the \( X_t \) series, say, \( X_t^* \), and \( q \) values of the \( \varepsilon_t \) series, say, \( \varepsilon_t^* \), were available corresponding to the needed starting values. If this were the case, values of \( \varepsilon_t \) for \( t = 1 \ldots T \) could be calculated from (2.6.8), and these values would be conditional on the starting values that were used.

Given starting values, the conditional log-likelihood function for parameter values \( \phi, \theta, \) and \( \sigma^2_e \) would be:

\[
\ell(\phi, \theta, \sigma^2_e) = -T \ln \sigma_e - \frac{S(\phi, \theta)}{\sigma^2_e} \tag{2.6.9}
\]

Here, \( S(\phi, \theta) \) is the conditional sum of squares function which is:

\[
S(\phi, \theta) = \sum_{t=1}^{T} \varepsilon_t(\phi, \theta | X_t, X_t^*, \varepsilon_t^*)^2 \tag{2.6.10}
\]

For a fixed value of \( \sigma^2_e \) the log-likelihood function (2.6.9) depends only on the conditional sum of squares (2.6.10), thus we need only examine the conditional sum of squares function to determine the behavior of the conditional likelihood.

**Example: MA(1)**

For the MA(1) model (2.3.3), the conditional sum of squares function can be constructed:

\[
S(\theta_1) = \sum_{t=1}^{T} \left[ \hat{\varepsilon}_t(\theta_1) \right]^2 = \sum_{t=1}^{T} \left[ X_t - \theta_1 \hat{\varepsilon}_{t-1}(\theta_1) \right]^2 \tag{2.6.11a}
\]
The next question then is how to evaluate \( \hat{e}_t(\theta_1) \) for each \( t = 1 \ldots T \). The first disturbance is expressed as follows:

\[
\hat{e}_1(\theta_1) = X_1 - \theta_1 \hat{e}_0(\theta_1) \tag{2.6.12a}
\]

and in general,

\[
\hat{e}_t(\theta_1) = X_t - \theta_1 \hat{e}_{t-1}(\theta_1) \quad 1 < t \leq T \tag{2.6.12b}
\]

Because (2.6.12a-b) are recursive, there is an immediate impasse because there is no observation on \( \hat{e}_0 \) from which to calculate \( \hat{e}_1 \) (the starting value problem). In practice, because \( \hat{e}_t \sim \mathcal{N}(0, \sigma^2_t) \) it is safe to assume a zero value for the starting values, as that is their expected value.\(^{41}\) For the MA(1) model, we need only set this first \( \hat{e}_0 \) to zero to get the recursive calculation started. For the MA(q) model we would need to set \( \hat{e}_0, \ldots, \hat{e}_{q-1} = 0 \) to begin the calculation.

Continuing with the MA(1) example, with a starting value of 0:

\[
\hat{e}_1(\theta_1) = X_1
\]

\[
\hat{e}_2(\theta_1) = X_2 - \theta_1 \hat{e}_1(\theta_1)
= X_2 - \theta_1 X_1
\]

\[
\hat{e}_3(\theta_1) = X_3 - \theta_1 \hat{e}_2(\theta_1)
= X_3 - \theta_1 X_2 - \theta_1^2 X_1
\]

\[\ldots\ldots\ldots\ldots\ldots\]

Using this pattern, equation (2.5.11) can be rewritten as follows:

\[
S(\theta_1) = \sum_{t=1}^{T} (X_t - \theta_1 X_{t-1} - \ldots - \theta_1^{t-1} X_1)^2 \tag{2.6.13}
\]

Because (2.6.13) is non-linear in \( \theta_1 \), the first order conditions will be non-linear and in general an iterative numerical optimization algorithm must be used to locate the value of \( \theta_1 \) for which \( S(\theta_1) \) is a minimum. In this simple example a grid search can be performed over the interval (-1,1) for the value of \( \hat{\theta}_1 \) that will minimize the conditional sum of squares. For more complex objective functions, an algorithm such as Gauss-Newton can be employed.\(^{42}\)

\[\text{---}
\]

\(^{41}\) There are alternative methods for obtaining the starting values, and the methods used vary among computer packages. For a more comprehensive discussion on starting values see Box and Jenkins 1976.

\(^{42}\) See Harvey 1981 and Judge et al. 1985 for a more comprehensive treatment of nonlinear optimization.
Diagnostic Checking

Diagnostic checking is the stage in the model building strategy in which the suitability of the estimated model is assessed. A natural test of model adequacy is to see if the residuals of the estimated model are white noise by inspecting the ACF and PACF of the residual series. If both the ACF and PACF display no significant values, then we can be somewhat confident that the model has been properly specified. However, because the residuals are estimates, and because the standard error bands are not exact, there is likely to be ambiguity in the ACF and PACF plots of the residuals even if they are truely empirical white noise. Because of this possible ambiguity, another test for white noise residuals can be employed. This test is called the Ljung-Box Q-Test and it checks the significance of a set of sample autocorrelations (Ljung and Box 1978). For the set of autocorrelations corresponding to lags \( k = 1, 2, ..., M \), the appropriate test statistic is:

\[
Q = T(T + 2) \sum_{k=1}^{M} (T - k)^{-1} r_k^2(e)
\]  

(2.6.14)

Here, \( M \) is the number of residual autocorrelations calculated and summed. This statistic is distributed Chi-Square with \( M-p-q \) degrees of freedom, with high values of \( Q \) leading to the rejection of model adequacy. The Q-Statistic test avoids the problem of the standard errors calculated by \( 1.96/\sqrt{T} \) understating the
significance of sample autocorrelations at low lags (Granger and Newbold 1986).\footnote{As noted by Davies and Newbold (1979) this statistic may also have low power and should not be the sole guide to model adequacy.}

Another common diagnostic check is called model "overfitting". In this case a model which is tentatively identified and seems to most likely have white noise residuals is tested against alternative specifications with either an extra MA term or an extra AR term, but never both. The overfitted models are estimated, and the significance of the coefficients on the added terms is tested.

It is interesting to note why the parameterization should only be increased in one direction or the other and not both directions simultaneously. Suppose that the model which is estimated is an AR(1) model, and, though we do not know it, it is the true model. Given an AR(1) specification the following is also true:

\[ X_t - \phi_1 X_{t-1} = \varepsilon_t \quad (2.6.15a) \]

And as well:

\[ X_{t-1} - \phi_1 X_{t-2} = \varepsilon_{t-1} \quad (2.6.15b) \]

Then, subtracting (2.6.15b) from (2.6.15a) we get:

\[ X_t = (1 + \phi_1) X_{t-1} - \phi_1 X_{t-2} + \varepsilon_t - \varepsilon_{t-1} \quad (2.6.16) \]
Equation (2.6.16) is an ARMA(2,1) model which is implied by the AR(1), and we would expect this model to have significant coefficients and display a good fit simply because it is equivalent to the AR(1) which is the true model.\(^4\) This parameter redundancy might entice the practitioner to accept a more complex model when a simpler one is the true model. The philosophy adopted at the identification stage which takes us from the simplest models up to more general models is designed to avoid this pitfall.

When a model passes the requisite diagnostic checks, the model building procedure terminates. It is useful to employ all these diagnostic techniques in concert. If the model fails on one or more of the diagnostic checks, the practitioner returns to the identification stage and begins again.

---

\(^4\) This example is from Nelson 1972.
Chapter 3 - Vector Time Series Analysis and Modeling

My vantage point permitted a clear view of the depths of the pit below: a desolation bathed with the tears of its tormented crew, who moved about the circle of the pit at about the pace of a litany procession. Silent and weeping, they wound round and round it ... they came on backwards staring backwards at their loins, for to look before them was forbidden.


3.1 Introduction

In this chapter, an extension of univariate time series analysis to vector (or multivariate) time series analysis will be considered. If we believe that the movements of other variables can have an important effect on the variable we are attempting to forecast, it seems appropriate to want to include such variables in
the forecasting model. To this end, vector time series analysis is concerned with sets of random variables. Let \( X'_t = (X_{1t}, X_{2t}, \ldots, X_{mt}) \) denote an \( m \)-dimensional, discrete-time, vector stochastic process. On each of the \( m \) random variables in the vector, there are \( T \) observations. Often, \( X_t \) will be used to represent the entire collection of random variables, \( X_{1t}, X_{2t}, \ldots, X_{mt} \); in addition, \( X_t \) also denotes the vector of observations associated with time \( t \). The specific meaning should be clear from the context.

The \( m \) components of the vector are chosen based on an appeal to economic theory which might suggest variables that contain information pertinent to the investigation. Apart from this, however, the philosophy of the modeling procedure for the multivariate case is basically the same as that for the univariate case. The idea in the univariate case is to use a linear model to approximate the history of the time series variable as captured in autoregressive and moving average components. If the error term is reduced to white noise, the regularities in the data have been captured so that only purely uncorrelated random shocks remain.

In the multivariate case, a vector of time-ordered random variables will be modeled such that each component, \( X_{it}, i = 1, 2, \ldots, m \), can be expressed as a linear combination of its own past as captured in autoregressive and moving average terms, contemporaneous and past values of the other series in the vector, and contemporaneous and past values of error terms corresponding to the other variables in the vector. When the disturbance term in each equation is reduced to white noise, it can be stated confidently that the vector stochastic process has been approximated sufficiently well.
3.2 Linear Models for Vector Stochastic Processes

Stationary Vector Stochastic Processes

As in the case for univariate time series, where we need to work with a stationary process, in the multivariate case we require stationarity of the vector process. Given a discrete vector stochastic process of dimension $m$, $X_t' = (X_{1t}, X_{2t}, \ldots, X_{mt})$ $t = 1\ldots T$, if the first two moments of the process are finite and time invariant the process is weakly stationary and hence can be described using these first two moments. A vector stochastic process $X_t$ is weakly stationary if each of the following hold:

$$E[X_t] = \mu < \infty \text{ for all } t \quad (3.2.1a)$$

$$E[(X_t - \mu)(X_t - \mu)'] = V < \infty \text{ for all } t \quad (3.2.1b)$$

$$E[(X_t - \mu)(X_{t+k} - \mu)'] = \Gamma(k) \text{ for all } t \text{ and } k \quad (3.2.1c)$$

Here, $X_t = (X_{1t}, X_{2t}, \ldots, X_{mt})$ and $\mu = (\mu_1, \mu_2, \ldots, \mu_m)$. $V$ is the $m \times m$ covariance matrix of the process. This symmetric matrix contains the variances of each random variable in $X_t$ on its main diagonal and the contemporaneous covariances between the pairs of series in $X_t$ on its off-diagonal elements. $\Gamma(k)$ is the $m \times m$ autocovariance matrix at lag $k$ which has on its main diagonal the autocovariance at lag $k$ of each series in the process. The off-diagonal elements of each autocovariance matrix contain the cross-covariances at lag $k$ between each possi-
ble pair of the series in the vector $X_t$. Note also that there are $k$ of these autocovariance matrices, one for each lag 1...$k$.

The first two conditions, (3.2.1a,b), require that each series is individually stationary in mean and has a constant, finite variance. The third condition, (3.2.1c), requires that all covariances, whether intraseries (as an autocovariance), or interseries (across every pair of the $m$ variables in the vector process), be independent of $t$ and depend only on the time displacement $k$. Again the implicit assumption of ergodicity is assumed to hold (see Hannan 1970).

**Example: Bivariate Stochastic Process**

As an illustration, consider the following stationary bivariate stochastic process, $X'_t = (X'_{1t}, X'_{2t})$. Suppose each component has $t = 1,2,...,T$ observations.

Again assuming that each observation is a deviation from the mean of the series, the covariance matrix can be derived as follows:

$$V = E[(X_t)(X'_t)] = E\begin{bmatrix} X_{1t} & X_{2t} \\ X_{2t} & X_{1t} \end{bmatrix}$$

(3.2.2)

$$V = E\begin{bmatrix} X_{1t}X_{1t} & X_{1t}X_{2t} \\ X_{2t}X_{1t} & X_{2t}X_{2t} \end{bmatrix}$$

(3.2.3)

Along the main diagonal there are the variances of each series in the bivariate relationship, $E[X_{1t}X_{1t}]$ and $E[X_{2t}X_{2t}]$. The off diagonal elements are the contemporaneous cross-covariances. Note also that $\Gamma(1), \Gamma(2),..., \Gamma(k)$, each a $2 \times 2$ matrix, can be estimated. Each matrix has the autocovariance at lag $k$ on its main diagonal, and the cross-covariances at lag $k$ are the off-diagonal elements.

---

45 If the additional assumption is made that the vector process is joint-normally distributed, then equations (3.2.1a-c) are sufficient to describe strict or wide sense stationarity as any normally distributed vector of random variables can be described fully by its first and second moments.

46 Recall that autocovariances are symmetric about the zero lag, where $\gamma_k = \gamma_{-k}$. Cross-covariances, denoted $\gamma_{xy}(k)$ are not symmetric about $k = 0$ so that in general, $\gamma_{xy}(k) \neq \gamma_{yx}(-k)$ because one variable may be a leading indicator of another.
Multivariate Wold Representation

In the univariate case, Wold's theorem was used to justify the construction of a linear model as a useful approximation to the underlying stochastic process which is assumed to have generated the time series under investigation. Wold's theorem demonstrates that whether or not a stochastic process is in reality a linear generating process, we can construct a linear process that will serve as a useful approximation. Analogously, a vector stochastic process, if it is purely nondeterministic and if (3.2.1a-c) hold, has for each element, the following linear vector moving average (VMA) representation (Hannan 1970):

$$X_t = \Theta(B)E_t$$  \hspace{1cm} (3.2.4)

The operator, $\Theta(B)$, is an $m \times m$ matrix in the backshift operator B. Every element in the matrix $\Theta(B)$ is an infinite polynomial in the backshift operator, such that, $\theta_q(B) = \sum_{k=0}^{\infty} \theta_{q,k} B^k$.

**Example: Bivariate Stochastic Process**

The bivariate process, $X'_t = (X_{1t}, X_{2t})$, has a multivariate Wold representation as follows:

$$\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} \theta_{11}(B) & \theta_{12}(B) \\ \theta_{21}(B) & \theta_{22}(B) \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$  \hspace{1cm} (3.2.5)

Each element of a vector stochastic process can be modeled as a linear combination of its own contemporaneous and past disturbances and the contemporaneous and past disturbances of all other components in the vector. Note also that the vector $E'_t = (\epsilon_{1t}, \epsilon_{2t}, ..., \epsilon_{mt})$ is a multivariate white noise process with mean zero, all auto- and cross-correlations equal to zero, and finite variance.
However, this VMA(\infty) representation is not a useful form. If the matrix \Theta(B) can be approximated by the product of two matrices, A^{-1}(B) and B(B) each involving only finite order polynomials in B, the following general class of linear models can be considered as an approximation of the stochastic process (Granger and Newbold 1986):

\[ A(B)X_t = B(B)E_t \quad (3.2.6) \]

Here, A(B) and B(B), both \(m \times m\), are matrix functions in the backshift operator that have elements which follow the structure,

\[ A_{ij}(B) = \sum_{k=0}^{p_{ij}} A_{ij,k}B^k \quad \text{and} \quad B_{ij}(B) = \sum_{k=0}^{q_{ij}} B_{ij,k}B^k \]

Equation (3.2.6) is called a multivariate ARMA(p,q)\(^{47}\) process where \(p = [p_{ij}]\) and \(q = [q_{ij}]\) are each \(m \times m\) matrices representing the lag lengths on the autoregressive and moving average terms respectively.

**Example: Bivariate Process**

Consider again the stationary bivariate process \(X_t' = (X_{1t}, X_{2t})\), written as a VARMA process (3.2.6):

\[
\begin{bmatrix}
A_{11}(B) & A_{12}(B) \\
A_{21}(B) & A_{22}(B)
\end{bmatrix}
\begin{bmatrix}
X_{1t} \\
X_{2t}
\end{bmatrix}
= \begin{bmatrix}
B_{11}(B) & B_{12}(B) \\
B_{21}(B) & B_{22}(B)
\end{bmatrix}
\begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix} \quad (3.2.7)
\]

Suppose we know the values for the \(p\) and \(q\) matrices:

\[ P = \begin{bmatrix}
p_{11} = 2 & p_{12} = 2 \\
p_{21} = 2 & p_{22} = 2
\end{bmatrix} \quad Q = \begin{bmatrix}
q_{11} = 1 & q_{12} = 1 \\
q_{21} = 1 & q_{22} = 1
\end{bmatrix} \quad (3.2.8) \]

---

\(^{47}\) It is also called a vector ARMA(p,q) or VARMA(p,q) process.
Suppose also that the normalization $A_{11,0} = A_{22,0} = 1$ is imposed. The system of equations (3.2.8) can then be written in the more familiar form:

\[
X_{1t} = A_{11,1}X_{1,t-1} + A_{11,2}X_{1,t-2} + A_{12,0}X_{2t} + A_{12,1}X_{2,t-1} + A_{12,2}X_{2,t-2} + B_{11,0}\varepsilon_{1t} + B_{11,1}\varepsilon_{1,t-1} + B_{12,0}\varepsilon_{2t} + B_{12,1}\varepsilon_{2,t-1}
\]

\[
X_{2t} = A_{22,1}X_{2,t-1} + A_{22,2}X_{2,t-2} + A_{21,0}X_{1t} + A_{21,1}X_{1,t-1} + A_{21,2}X_{1,t-2} + B_{21,0}\varepsilon_{1t} + B_{21,1}\varepsilon_{1,t-1} + B_{22,0}\varepsilon_{2t} + B_{22,1}\varepsilon_{2,t-1}
\]

(3.2.9a) (3.2.9b)

When the roots of the characteristic equation $|A(B)| = 0$ all lie outside the unit circle the process is stationary. When the roots of the characteristic equation $|B(B)| = 0$ lie outside the unit circle, the process is invertible. Both $A(B)$ and $B(B)$ are assumed to possess full rank and consequently to be invertible.

3.3 Vector Autoregressive Models

**Introduction**

Because the interactions among series are often difficult to identify, and because the presence of moving average terms makes a nonlinear routine a necessity for estimation, the VARMA models of the previous section are often abandoned in favor of the less parsimonious but more easily specified and estimated vector autoregressive models, or VAR($p$). Here, $p$ is a matrix of autoregressive lag lengths as in (3.2.8). If the moving average portion of a VARMA model is invertible, then there exists an infinite autoregressive representation of these parameters. This infinite autoregressive representation can then be approximated by a finite autoregression. The tradeoff here is that a more parsimonious model that is correctly identified will generally provide better forecasts because it is a closer approximation to the underlying process, and conse-
quently makes more efficient use of the data. However, if the more parsimonious (e.g., VARMA(p,q)) model is more difficult to specify correctly, it is conceivable that a less parsimonious model, such as a VAR, that can be specified correctly may forecast better than a questionably specified albeit more parsimonious VARMA model.

A direct and intuitive way to achieve a VAR representation is to take the inverse of the VMA(∞) representation provided by Wold's theorem (equation 3.2.4), which gives an infinite autoregressive representation that itself can be approximated:

\[
\begin{bmatrix}
\phi_{11}(B) & \phi_{12}(B) & \cdots & \phi_{1m}(B) \\
\phi_{21}(B) & \phi_{22}(B) & \cdots & \phi_{2m}(B) \\
\vdots & \vdots & \ddots & \vdots \\
\phi_{m1}(B) & \phi_{m2}(B) & \cdots & \phi_{mm}(B)
\end{bmatrix}
\begin{bmatrix}
X_{1t} \\
X_{2t} \\
\vdots \\
X_{mt}
\end{bmatrix}
= 
\begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t} \\
\vdots \\
\varepsilon_{mt}
\end{bmatrix}
\]  

(3.3.1)

Each of the \( \phi_{ij}(B) \) are polynomials in the backshift operator of finite order (the order being the approximating lag length which is chosen as part of the identification procedure). Here, \( \varepsilon_{1t}, \ldots, \varepsilon_{mt} \) are white noise disturbances with mean zero and \( m \times m \) covariance matrix \( \Sigma < \infty \). The equation (3.3.1) can be written in compact form as:

\[
\Phi(B)X_t = E_t
\]  

(3.3.2)

where \( \Phi(B) \) is a finite matrix function in the backshift operator.

Chapter 3 - Vector Time Series Analysis and Modeling 63
Choosing Lag Length in a VAR

It was noted in Chapter 1 that VAR modeling is a trade-off between oversimplification and overparameterization, and the modeler must devise a way to filter the noise from the signal in such a way that a model is parsimonious, but not too simple that it fails to capture the interactions in the data.

Litterman (1976) proposes a way of filtering the noise from the signal through the use of Bayesian priors. He suggests that it is reasonable to expect that coefficients on long lags are more likely to be zero than coefficients on shorter lags and proposes priors centered on a random walk. By specifying Normal prior distributions about zero with smaller standard deviations for the coefficients as lag length increases (a "tighter" prior), long lags are allowed to enter the equations at the margin only if there is strong evidence for such relationships in the data (see Litterman 1984, and Doan, Litterman, and Sims 1983 for discussions of Bayesian VAR models for forecasting). Bessler (1986) uses a nonsymmetric random walk prior (i.e., the prior on cross variable effects is different from the prior on own lags) on the coefficients in a VAR model for forecasts of the U.S. hog market. He finds that the Bayesian VAR with this prior outforecasts a univariate autoregression, an unrestricted VAR, and a Bayesian VAR with a symmetric prior.

An alternative to the Bayesian procedure for obtaining a more parsimonious VAR representation is to use a multivariate statistical decision cri-

---

44 A prior is an informed belief that the modeler brings to the modeling exercise. The priors appear in the form of probability distributions on the coefficients.

49 The random walk prior is justified for many macroeconomic and financial variables (e.g., Nelson and Plosser 1982).
terion for the choice of lag length. Lutkepohl (1985) has investigated the use of 12 such statistical decision rules in a monte carlo simulation. He found several of them to be quite accurate in choosing the correct lag length. Lutkepohl's results indicate that the Bayesian Information Criterion of Schwartz (1978) and the criterion of Hannan and Quinn (1979) are the most accurate given a moderately large sample size. These decision criteria are applicable in both a univariate and a multivariate lag selection problem, the multivariate being the general case.

A third alternative is a procedure Hsiao (1979) developed to help overcome the overparameterization problem by allowing for more realistic differing lag structures in each equation of the system. Hsiao's procedure uses the Final Prediction Error criterion of Akaike (1971), though any of a number of available statistical rules could be used as the underlying decision criterion in his procedure (see Judge et al. 1985 p. 675). Though Hsiao's procedure is not without fault, it is a useful procedure for modeling restricted VARs. Having fewer parameters in a VAR allows the remaining parameters to be estimated with more degrees of freedom, hence more accurately. Hsiao's method for reducing the number of parameters to be estimated in a VAR is closer to the time series philosophy of allowing the data to determine the model specification than a Bayesian procedure which forces the modeler to choose a prior to impose. In this study we restrict

---

50 Webb (1985) applies the Akaike Information Criterion (AIC) within his own procedure to the choice of lag length in a VAR. He notes a consistent improvement in the forecasting accuracy of his specification using the AIC over an unrestricted VAR.

51 Doan, Litterman, and Sims (1983) propose the quasi-Bayesian approach of using the data to select an optimal prior.
our attention to statistical decision procedures for choosing lag length in unrestricted VARs, and to Hsiao's procedure for specifying restricted VARs.

**Choosing Lag Length in an Unrestricted VAR**

Non-Bayesian procedures for selecting the lag-order of an unrestricted VAR can be characterized as falling, generally, into one of two categories: (1) Sequential testing procedures, and (2) Criterion function optimization.

**Sequential Testing Procedures**

These procedures are a traditional method for choosing the order of a lagged process, such as an autoregression or a distributed lag. The test which applies in this case is the likelihood ratio (LR) test in which an $\text{AR}(p)$ specification is tested against an $\text{AR}(x)$ specification where $x$ denotes some maximum lag length to which a succession of the restricted models, with lags $p < x$ are compared. The following statistic is used (Judge et al. 1985):

$$\Lambda = T(\ln |\hat{\Sigma}_p| - \ln |\hat{\Sigma}_x|) \quad (3.3.3)$$

where $\hat{\Sigma}_j$ is the estimated covariance matrix from a model with $j$ lags ($j = p, x$). The statistic in (3.3.3) is distributed asymptotically Chi-square with $m(\kappa - p)$ degrees of freedom. Here, $m$ denotes the number of equations in the system, $x$ the maximal lag length, and $p$ the estimated lag length. In general, for small samples, the actual distribution of the test statistic in (3.3.3) is not well approximated by the asymptotic Chi-square distribution. Because of this, Lutkepohl
(1985) notes the poor performance of this method in approximating the lag lengths in his simulation study.

Sims (1980) has suggested a correction for the test statistic in (3.3.3) which, in small samples, improves the properties of the test. The correction involves subtracting from T in (3.3.3) the number of variables in each unrestricted equation in a VAR, say, k. The LR statistic is then computed as if the sample size is T-k.

**Criterion Function Optimization**

Most procedures for choosing lag length are of this type. An example of the kind of criterion function that is being considered is the Final Prediction Error of Akaike (1971). By minimizing the 1-step ahead prediction error, the lag length j is chosen such that:

\[
FPE(j) = \left[ \frac{T + jm + 1}{T - jm - 1} \right]^m |\hat{\Sigma}_j|
\]  

(3.3.4)

is a minimum among lag lengths \(j \leq x\) (Lutkepohl 1985). Here, T is the sample size, j is the lag length, and m is the number of variables in the system.

Two other selection criterion often considered are Schwartz's Bayesian Information Criterion (BIC) (Schwartz 1978), and the Hannan-Quinn criterion (see Hannan and Quinn 1979). These two selection rules take the form:

\[
BIC(j) = \ln |\hat{\Sigma}_j| + \frac{m^2 j \ln T}{T}
\]

(3.3.5)
\[ HQ(j) = \ln |\hat{\Sigma}_j| + \frac{2jm^2 \ln \ln T}{T} \]  

(3.3.6)

In each case the object is to choose the lag length \( j \) such that the functions (3.3.4), (3.3.5), or (3.3.6) are minimized. The HQ and BIC criteria can be interpreted as being of the same family in the sense that the first term in each, \( \ln |\hat{\Sigma}_j| \), decreases with increasing AR order, whereas the second term in each imposes a "penalty" by increasing with increasing AR order. Therefore, for a minimum to be reached, both components must be optimally balanced (Lutkepohl 1986). It can be shown that both HQ and BIC will provide consistent estimates of the true lag length, and do a reasonably good job approximating the asymptotic distribution of the true process in moderately sized samples. The FPE criterion will asymptotically overestimate the true order of the process with a positive probability (Judge et al. 1985; Lutkepohl 1985). Lutkepohl provides evidence in his monte carlo study that in small to moderate samples, from among the 12 criterion and sequential tests examined, the BIC and the Hannan-Quinn criteria selected the correct order most often, and consequently the resulting estimated VARs provided superior forecasts.

The lag length suggested by any one of these criteria applies for each variable in the VAR. Hence, each variable enters the VAR with the same lag length. Intuitively, there is no reason to expect that this should be the case. Because some variables may be leading indicators of others, there are likely to be equations in which some variables do not enter at all. By ignoring these possible exclusion restrictions, inefficient parameter estimates will result (Judge et al.
1981), and it is very likely that poor forecasts will be produced. Some authors have suggested eliminating statistically insignificant parameters after estimating the unrestricted VAR (Taio and Box 1981). Although practically appealing, high multicollinearity in an autoregression may distort the "t-tests" by biasing them toward not rejecting the null hypothesis of no relationship. An alternative procedure would be to test the joint hypotheses that groups of coefficients are insignificant using the familiar F-Test (Orden 1982). The next section describes a step-by-step method designed specifically for specifying a restricted VAR.

**Specifying Restricted VARs**

Hsiao (1979) has developed a sequential procedure which will select the lag lengths of each of the variables in a VAR separately, allowing for differing lag structures among the involved variables. The steps involved in the procedure are reiterated here for the bivariate case. The selection of the lag lengths in each of the successive steps requires the application of the univariate versions of one of the criterion functions from the previous section. Hsiao suggests the use of the FPE criterion. Suppose we wish to model the bivariate system $X' = (X_1, X_2)$ with Hsiao's procedure. The appropriate criterion function is then:

$$FPE(j, h) = \left[ \frac{T + j + h + 1}{T - j - h - 1} \right]^{\hat{\sigma}^2_{j,h}}$$  \hspace{1cm} (3.3.7)

Here $\hat{\sigma}^2_{j,h}$ is the ML estimate of the residual variance when the following model is fit to the data:
\[ X_{1t} = a(B)^j X_{1t} + b(B)^h X_{2t} + \hat{e}_t \]  

(3.3.8)

The values of \( j \) and \( h \) are the lag lengths of \( X_1 \) and \( X_2 \) in (3.3.8) respectively. The first step is to choose an output of the system, i.e., a variable which is likely to be affected by lagged values of the other. To illustrate the procedure, suppose we hypothesize that \( X_2 \) determines the movement in \( X_1 \), so that \( X_1 \) is the output of the system. We proceed as follows:

1. Using the FPE in (3.3.7) select the order of a univariate autoregression on the output variable \( X_{1n} \), varying the lag length on the output variable, \( j \), from 0 to \( j^{\max} \). Call the optimal lag length \( j^* \). The variable \( h \) representing the lag length of the control variable, \( X_{2n} \), is set to zero in this step.

2. Determine the lag order \( h^* \) of the control variable, \( X_{2n} \), assuming that the order of \( X_{1n} \) is that chosen in step 1, i.e., \( j^* \).

3. Next, Hsiao recommends checking whether upon the inclusion of \( X_2 \), some of the lags of \( X_1 \), are made unnecessary, i.e. that the univariate autoregression on \( X_{1n} \), picked up the effects of lagged \( X_{2n} \). To do this, set the lags of \( X_{2n} \) at the value chosen in step 2, \( h^* \), and vary the lags of \( X_{1n} \) from zero to \( j^* \). This new lag length may or may not equal \( j^* \).
4. Lastly, compare the lowest FPEs of steps 1 and 3. If the smallest FPE is from the model fit in step 1, a univariate autoregression for $X_{1t}$ is used. If the converse is true, the optimal model for $X_{1t}$ is one including $h^*$ lags of $X_{2t}$.

If a model is chosen that has lags of $X_{2t}$, the second equation of the VAR must be specified. To do this, steps 1-4 are repeated but reversing the places of $X_{1t}$ and $X_{2t}$ by first choosing a univariate autoregression on $X_{2t}$. If lags of $X_{1t}$ enter this second equation, the relationship is said to include feedback.

Finally, Hsiao's sequential procedure to specify a VAR may introduce bias, so that diagnostic checks should be conducted by over- and underfitting the tentatively chosen models using LR tests.

### 3.4 Nonstationarity in the Multivariate Case

**Introduction**

Linear vector time series models, such as VARs, are only applicable to stationary vector time series (Judge et al. 1985). A vector time series $X'_t = (X_{1t}, X_{2t}, \ldots, X_{mt})$, is stationary when each series is individually stationary in mean and variance. In addition, all covariances, whether intraseries (an autocovariance) or interseries (across every pair of the m variables in the vector process) must be independent of $t$ and depend only on the time displacement between observations (see equations 3.2.1a-c).
As discussed in Chapter 2, there is reason to believe that many economic time series are nonstationary. Hence, the practitioner is faced with the problem of how to apply the theory of vector linear models to nonstationary time series. Extrapolating from the univariate case, a practical solution would seem to be to examine the univariate time series properties of each series in the vector to be modeled and use appropriate transformations to reduce each individual series to stationarity. Such an approach is advocated in some articles on VAR modeling (e.g., Hsiao 1979). Hsiao logged and differenced each series in the bivariate money-income relationship before proceeding with specifying a VAR. However, differencing of the individual series has been criticized by others (e.g. Taio and Box 1981; Lutkepohl 1982). The difficulty noted is that while each individual series may be nonstationary, "for vector time series, linear combinations of the components of \( [X_t] \) may often be stationary, and simultaneous differencing of all series can lead to unnecessary complications in model fitting" (Taio and Box 1981 p.804). This phenomenon of linear combinations of nonstationary series being stationary has been termed co-integration (Granger 1980; Granger and Weiss 1983; Engle and Granger 1986; Engle and Yoo 1987). Essentially, if there exist linear combinations of the individual nonstationary series that are stationary, differencing each series individually will result in a system that is overdifferenced. If this is the case, the system will no longer have a multivariate linear time series representation with an invertible moving average (Engle and Yoo 1987). Intuitively, if a system is co-integrated, estimating a model in differences ignores the equilibrium relationships among the nonstationary variables that contain im-
portant information. Modelling the co-integration restrictions, then, should help a model produce forecasts that are more accurate than a model in which the restrictions are ignored (Engle and Yoo 1987).

A related issue concerns the appearance of spurious relationships when nonstationary variables are regressed on each other (Granger and Newbold 1974, 1986). Phillips (1986) has demonstrated, using large-sample theory, that when non-stationary series are regressed upon one another the t-tests of significance are biased toward rejecting the null hypothesis of no relationship. In related work, Phillips and Durlauf (1986) demonstrate that the asymptotic theory for non-stationary multiple time series departs significantly from classical theory. In the case of non-stationary series, the asymptotic distribution of the usual test criteria is nuisance parameter dependent, meaning that classical test statistic distributions are no longer applicable.

**Estimation and Inference with Integrated Processes**

For integrated processes of order greater than zero (i.e. nonstationary series) the use of statistical techniques which assume stationarity can give incorrect results in the multivariate case as well as in the univariate case. To illustrate the issues involved, consider the static regression:

\[ Y_t = \alpha + \beta' X_t + \epsilon_t \]  \hspace{1cm} (3.4.8)

where \( \beta' \) is a vector of coefficients and \( X_t \) is a vector of regressors. Suppose that \( Y_t \) and the \( X_t \)'s are each \( \sim I(1) \). Rearranging (3.4.8):
\( \epsilon_t = Y_t - x - \beta'X_t \)  \hspace{2cm} (3.4.9)

In general the linear combination in (3.4.9) will yield \( \epsilon_t \sim I(1) \) because \( \epsilon_t \) is a linear combination of I(1) series. Hence the residuals will be nonstationary.\(^{52}\) Phillips (1986) and Phillips and Durlauf (1986) investigate the effect of using integrated processes in static multivariate regressions such as (3.3.8), and in multiple time series regressions such as VARs. They conclude, using large sample theory, that the distributions associated with the usual inferential statistics do not follow the same distributions that they would under stationarity. For the case of static multiple regression, Phillips proves that the coefficients of the regression do not converge in probability to constants as the sample size goes to infinity, as is the case when the variables are stationary; that is, the variables have no limiting distribution (Phillips 1986a; Banerjee et al. 1986). Phillips also shows that the distributions of the t-ratios diverge as the sample size goes to infinity. This means that no asymptotically correct critical values exist for conventional significance tests. For critical values from conventional asymptotics, the rejection rate for the null hypothesis will increase with sample size (Phillips 1986, p.318). These theoretical results confirm the monte carlo evidence in Granger and Newbold (1974, 1986). The bias toward wrongly rejecting is their concept of spurious regressions.\(^{53}\)

Granger and Newbold illustrate the problem by regressing independent random

\(^{52}\) For the case where \( Y_t \sim I(1) \) and \( X_t \sim I(0) \) the residual in (3.4.9) will be I(1) as well.

\(^{53}\) Yule (1926) was the first to formally investigate this phenomenon, often called "spurious" or "nonsense" correlations. Yule examined the correlations between unrelated series. When the series were stationary, no correlation was observed, as expected. For I(1) series, the correlation distribution indicated a high degree of linear association, and for I(2) series the most often encountered correlations were \( \pm 1 \).
walks on one another. They note that using the usual t-test (designed under the maintained hypothesis that the variables involved are stationary) at a 5-percent level of significance will, on average, lead to wrongly rejecting the null hypothesis 75-percent of the time. Where the number of independent variables is greater than one, Granger and Newbold (1974) report a bias in F-tests toward wrongly rejecting the joint null that all coefficients are zero from 76-percent to 96-percent of the time, with the rejection rate increasing with the number of included variables.44

Dynamic multivariate time series regressions with integrated processes, as opposed to static regressions, are investigated by Phillips and Durlauf (1986). They find that, in this case, OLS does provide consistent estimates of the regression coefficients. However, these estimates are not asymptotically normally distributed. An important result is that the limiting covariance matrices for the estimated coefficients have distributions that depend on the number of variables in the system. These nuisance parameter dependencies invalidate the usual classical significance tests. New statistical tests must therefore be devised which are free of nuisance parameter dependencies.

Thus, nonstationarity in a dynamic multivariate regression, such as a VAR, can cause serious problems for statistical inference. In order to avoid being fooled by spurious relationships, making invalid conclusions based on the application

44 To underscore their results, Granger and Newbold work with statistically independent variables. However, the distributional results proved by Phillips (1986) also apply to correlated time series. The crucial results are that the coefficients do not converge to constants and that the distributions of the test statistics diverge as the sample size increases to infinity.
of the wrong asymptotic theory, or generating poor forecasts, one must insure that the series involved are stationary.

3.5 Co-integrated Economic Variables

Co-integration and a Notion of Equilibrium

In a static multivariate regression, finding an ARIMA representation for the residuals and/or differencing the included variables should eliminate the occurrence of invalid conclusions on the basis of classical inferential techniques. For dynamic time series regressions with integrated processes, consideration of co-integration plays a vital role in deciding what to do about nonstationarity.

Consider two time series, \( X_{1t} \) and \( X_{2t} \), each of which is I(1). Given the following regression equation:

\[ X_{1t} = \beta X_{2t} + \epsilon_t \]  \hspace{1cm} (3.5.1)

the residual series will probably also be non-stationary, specifically I(1), because it is a linear combination of I(1) variables. There is, however, a special case where there exists a constant, say, \( \zeta \), such that two I(1) series such as \( X_{1t} \) and \( X_{2t} \) will have a unique linear combination:

\[ z_t = X_{1t} - \zeta X_{2t} \]  \hspace{1cm} (3.5.2)
that is stationary (more precisely \( z_t \sim I(0) \)). If such a linear combination exists, \( X_{1t}, X_{2t} \) are termed co-integrated, with the co-integrating constant, \( \zeta \), denoted \( (X_{1t}, X_{2t}) \sim CI(1,1) \). More generally, for any pair of series, \( X_{1t} \) and \( X_{2t} \), both \( \sim I(d) \), if there exists a linear combination (3.5.2) such that \( z_t \sim I(d - b) \) with \( b > 0 \), the pair \( X_{1t}, X_{2t} \) are termed co-integrated and this is denoted \( (X_{1t}, X_{2t}) \sim CI(d, b) \). In the two-variable case, \( \zeta \) will be a unique co-integrating constant. For vectors of time series with \( m > 2 \) there may be multiple vectors of co-integrating constants that yield a stationary linear combination (see Engle and Granger 1987).

Because of the temporal dissimilarity of \( z_t \) from its components in (3.5.2), \( X_{1t} \) and \( X_{2t} \) must be related in a unique fashion for co-integration to occur.\(^{55}\) A popular heuristic analogy is that if there exist two series, \( X_{1t} \) and \( X_{2t} \) each of which has a strong seasonal component, any linear combination of the two series should yield a third series that also has a strong seasonal component. However, it could be the case that the seasonal components of \( X_{1t} \) and \( X_{2t} \) have identical shape yielding a certain linear combination that has no seasonal component. Effectively, the seasonal components cancel out (Engle and Granger 1987).

Co-integration has an interpretation that is quite useful to economists. By rearranging (3.5.11), and suppressing the term \( z_t \):

\[
X_{1t} = \zeta X_{2t} \tag{3.5.3}
\]

\(^{55}\) Spectral analysis gives the following explanation: \( X_{1t} \) and \( X_{2t} \) are each dominated by low-frequency or long-wave components whereas \( z_t \) is not. This implies that these low-frequency components cancel out in the linear combination \( X_{1t} - \zeta X_{2t} \) yielding a \( z_t \sim I(0) \) (Granger 1986).
The $X_{1t}X_{2t}$ relationship can be considered an *equilibrium* relationship. This notion of equilibrium is more likely to be encountered in the context of a natural science, as it is analogous to the equilibrium state achieved, say, in a chemical reaction. Granger(1986) provides the following explanation: "The term equilibrium is used in many ways by economists. Here [in co-integrated systems] the term is not used to imply anything about the behavior of economic agents, but rather describes the tendency of an economic system to move towards a particular region of the outcome space." The variable $z_t$ in (3.5.2), then, can be said to measure departures from equilibrium. This notion of equilibrium stems from the fact that, separately, $X_{1t}, X_{2t}$ each move about unbounded because neither has finite variance or constant mean. If, however, (3.5.2) holds with $z_t \sim I(0)$ for all $t$, then it can be said that in the long run $X_{1t}$ and $X_{2t}$ must be moving together with their unique linear combination being bounded by finite mean and variance.\(^{56}\) Granger (1986) cites prices and wages, the money supply and prices, government income and expenditure (perhaps only at the state or local level), and the imports and exports of a country as pairs of variables that may be cointegrated.

A more general definition of co-integration can now be considered where there is a vector $X'_t = (X_{1t}, X_{2t}, ..., X_{mt})$ of time series to be modeled. Suppose that each series $X_{jt}$ is $\sim I(1)$, but there exists a vector of constants $\alpha$ such that:

$$z_t = \alpha'X_t \sim I(0) \quad (3.5.4)$$

---

\(^{56}\) It can be shown that if two variables are co-integrated and each has a deterministic trend, the co-integrating parameter will eliminate the linear combination of the trend functions. This means that the trend functions must be of the same type, e.g., both second degree polynomials in $t$ (See Granger(1986)).
If \( \alpha' \) exists, it will be called the co-integrating vector (Granger and Newbold 1986). There may be \( r \) different vectors \( \alpha' \) for which (3.5.4) will hold. In this case, \( \alpha' \) will be a matrix, and \( r \) will be the order of co-integration or the co-integrating rank (because \( r \) would be the rank of the \( \alpha' \) matrix where each row of \( \alpha' \) would be an independent co-integrating vector).

**Testing for Co-integration**

Because we are interested to see whether \( z_t \) in (3.5.2) is I(0) or I(1), it seems plausible to apply the stationarity tests of chapter 2 to these residuals.\(^{18}\) Under the null hypothesis of the DF test, \( z_t \) is treated as the series which has the unit root. This corresponds, for the present purpose, to a null hypothesis of no co-integration. This is because a unit root in \( z_t \) would mean that there is not a linear combination of I(1) variables that is I(0). The test of \( z_t \) for I(1), therefore, seems to be a straightforward application of the DF or ADF test. Unfortunately, this is not the case. In order to first obtain the residuals, \( z_t \) in (3.5.2), we must estimate by OLS the following equation called the "co-integrating regression":

\[
X_{1t} = \hat{c} + \hat{\xi}X_{2t} + \hat{\varepsilon}_t
\]  

\((3.5.5)\)

\(^{57}\) It should be noted that testing a system for co-integration can be a useful procedure in itself, independent of its purpose as a first stage in specifying a proper vector time series model. Because a co-integrated system can be thought of as a system in long-run equilibrium, tests for co-integration can be used as a check of the consistency of certain economic theories that posit that equilibrium relationships will be observed in the time series data. For some examples of this type of work see Hall 1986; Jenkinson 1986; and Ballibe and Selover 1987.

\(^{58}\) It is useful to first note that each individual time series must have an order of integration that is equal to the other's for co-integration to make sense. To insure this, one should test each series for I(1) by using the usual DF procedure outlined in chapter 3. Also, a plot of the ACF and ACF of first differences may provide additional insight. If such a pre-test shows that the series have different orders of integration, a formal co-integration test is unnecessary.
However, the co-integrating regression yields both an estimate of the co-integrating parameter, $\hat{\xi}$, and the residual series, $\hat{e}_t$.\(^9\) Since $\hat{e}_t$ can be obtained only after first obtaining $\hat{\xi}$, it has a dependency on the estimate of the co-integrating parameter. In the unit root test on the $\hat{e}_t$ series from the co-integrating regression, the large sample behavior of the "t-statistic" has nuisance parameter dependencies which stem from this dependency. These are manifest as a dependency of the "t-statistic" on the number of variables in the co-integrating regression (Engle and Yoo 1987). The critical values in Fuller (1976) and in Dickey and Fuller (1981) used for the usual DF and ADF unit root tests do not apply for the cointegration test because they do not account for these nuisance parameter dependencies. New critical values, dependent on the number of variables in the vector time series, are provided in tables in Engle and Yoo (1987).\(^6\) Engle and Granger suggest the following variation on the forms of (2.4.4) and (2.4.5). The "co-integration DF test" is a test of the "t-statistic" on $\delta$ in the following regression:

$$\Delta \hat{e}_t = \delta \hat{e}_{t-1} + \hat{u}_t$$

The "co-integration ADF test" is a test of the "t-statistic" on $\delta$ in:

\(^9\) Stock (1987) has shown that if $Y_t$ and $X_t$ are cointegrated, OLS estimates of $\xi$ are highly efficient and super consistent; that is, as $T\to\infty$ $\hat{\xi}$ will converge to its true value twice as rapidly as would be the case for a usual OLS parameter estimate in a similar, stationary, regression.

\(^6\) For the cointegration test, we can ignore the trend functions and the hypothesis becomes $I(1)$ vs. $I(0)$ with the relevant statistic being the t-ratio on the parameter $\rho$. 

Chapter 3 - Vector Time Series Analysis and Modeling
\[ \Delta \hat{\xi}_t = \delta \hat{\xi}_{t-1} + \sum_{j=1}^{p} \gamma_j \Delta \hat{\xi}_{t-j} + u_t \]  

(3.5.7)

For both (3.5.6) and (3.5.7) the hypothesis to be tested is that the residual series from the co-integrating regression, \( \hat{\xi}_t \sim I(1) \). The null hypothesis in (3.5.6) corresponds to \( \delta = 0 \), or no co-integration. Notice that if \( \delta = 0 \) in (3.5.6) we are left with \( \Delta \hat{\xi}_t = u_t \), or a random walk which is indeed \( I(1) \). In equation (3.5.7), if \( \delta \) is zero we are left with an autoregression in the differences of \( \hat{\xi}_t \) which implies that the raw series \( \hat{\xi}_t \sim I(1) \). If, however, the null hypothesis of either test is rejected, we can infer that the system is co-integrated.

Another test for co-integration proposed by Engle and Granger which is also carried out under the null hypothesis of non-cointegration is a variation on the Durbin-Watson test (D-W) for serial correlation (Durbin and Watson 1950). Here, the D-W statistic from the co-integrating regression is compared to critical values obtained by Monte-Carlo simulation (see Engle and Yoo (1987) for critical values).\(^4\) If the D-W statistic is significantly greater than zero, this implies that \( \hat{\xi} \) does contain a unit root, and hence the system is not co-integrated. Engle and Yoo (1987) show that if \( \hat{\xi} \) does contain a unit root, the D-W statistic will be greater than zero in its probability limit.

Engle and Granger (1987) claim that the D-W test is somewhat more powerful than the ADF test, however, the ADF test has more stable critical val-

\(^4\) The test does not use the same critical values as the usual D-W test because it is being used for a different purpose.
ues. Granger (1986) notes that for samples of 100 observations, the power of the ADF test seems satisfactory.

**Modelling Co-integrated Systems**

The co-integration relationship is a unique relationship and indeed is a special case to be encountered in the modeling of vector time series. It is now clear that by indiscriminately differencing variables in a vector time series to induce stationarity in each individual series, important information, i.e., that the series move together in the long run, can be lost if the series are co-integrated. Therefore the next question becomes: How to incorporate this long-run information in the modeling process?

**Granger's Representation Theorem**

Granger (1983) and Engle and Granger (1987) state and prove a theorem that shows that a co-integrated system, where the co-integrating rank (order) is \( r \), and \( d = b = 1 \), has two significant and useful representations:

1. (1) There exists a vector autoregressive representation in levels,

\[ A^1 \text{AR}(1) \text{ representation} \]

2. (2) There exists a cointegration representation in differences,

\[ A^1 I(1) \text{ representation} \]

---

63 Engle and Yoo (1987) note that even for large samples (\( T = 100 \)) the critical values of the D-W test for co-integration are very unstable, hence they recommend against using this test as anything more than a very rough first approximation.

64 These same tests can be applied to vectors of time series with \( m > 2 \). In these cases we are interested in finding the vector \( a' \) of co-integrating constants. Also, by changing the dependent variable in the co-integrating regression several such \( a' \) vectors can be tested for. It should also be noted that other tests for co-integrated time series have been developed such as those in Phillips and Ouliaris (1987) and Stock and Watson (1987a).

65 The following is an abbreviated statement of the theorem useful for the purposes of this study. For a full statement and a proof see Engle and Granger (1987).
\( \Phi(B)X_t = E_t \) \hspace{1cm} (3.5.8)

with the following restrictions: \( \Phi(1) \) has rank \( r \), and \( \Phi(0) = I_m \) where \( I_m \) is the identity matrix.

2. (2) There exists an error correction representation with \( z_t = \alpha'X_t \) an \( r \times 1 \) vector of stationary random variables:

\[ \Phi^*(B)(1 - B)X_t = -\gamma z_{t-1} + E_t \] \hspace{1cm} (3.5.9)

with \( \Phi^*(0) = I_m \).

**Example: Bivariate Error-Correction Model**

A bivariate co-integrated system has an error-correction model representation:

\[ \Delta X_t = -\xi_1 z_{t-1} + \alpha(B)\Delta X_t + \beta(B)\Delta Y_t + \epsilon_{1t} \] \hspace{1cm} (3.5.10a)

\[ \Delta Y_t = -\xi_2 z_{t-1} + \gamma(B)\Delta X_t + \phi(B)\Delta Y_t + \epsilon_{2t} \] \hspace{1cm} (3.5.10b)

with \( |\xi_1| + |\xi_2| \neq 0 \).

For case (1), the VAR representation, it is the co-integration which makes \( \Phi(1) \) singular. For \( r = 1 \) this matrix is of rank 1. This system is not invertible, and consequently there can be no meaningful VMA representation. From a purely empirical standpoint, a VAR in levels can be estimated, but it will not impose any of the co-integration restrictions, and hence should be less efficient than a representation that imposes the co-integrating restrictions.
For case (2), the error correction representation, the co-integration is captured, in the bivariate case, uniquely through the $z_{t-1}$ term, when, $z_{t-1} = X_{1,t-1} - \xi X_{2,t-1}$. This representation captures the co-integration in terms of the levels of the co-integrated variables (Engle and Granger 1987). Intuitively, the levels enter the equation as last period's departure from long-run equilibrium. As can clearly be seen, if the error-correcting term is dropped from equation (3.5.9), a VAR in differences is left. Specifying a VAR in differences, then, if the variables are co-integrated, is a specification error because the error-correcting term would incorrectly (as demonstrated in Granger's Representation Theorem) be excluded from the equation.

The estimation of a nonstationary, possibly co-integrated system can be thought of as a two stage process. In stage one the system is tested for co-integration, and in stage two, assuming that co-integration is found, an error correcting process of the form (3.5.9) can be specified and estimated. If co-integration is not found, a VAR in differences is appropriate.

The model selection process is a process that, to a large degree, depends on the information we can extract from the data concerning its time series properties. In the next chapter, we employ the data analysis techniques discussed above to help in specifying univariate models for forecasting agricultural prices, export sales, and bivariate models with prices and the exchange rate as well as export sales and the exchange rate. The relative forecasting accuracy of the bivariate vs. univariate models will then be assessed.
Chapter 4 - Detecting Macroeconomic Impacts on Agriculture

There are two things you are better off not watching in the making: sausages and econometric estimates.

Ed Leamer, American Economic Review.

4.1 Data Description

The price data used herein are average monthly cash prices of No.1 Hard Red Winter Wheat at Kansas City, No.2 Yellow Corn at Chicago, and No.1 Yellow Soybeans at the Illinois Processor, each is deflated by the U.S. CPI. The CPI was taken from various issues of the Survey of Current Business.
price data were obtained from the crops section of the Economic Research Service, USDA.

The wheat, corn, and soybean export sales data were computed by Fred Ruppel of Texas A&M University from weekly data available through the Export Sales Reporting Division, Foreign Agricultural Service of the USDA. We investigate export sales rather than the often used export shipments because export sales are the economic variables we expect to be most responsive to changes in exchange rates. Export shipments are better characterized as logistic variables, depending heavily on such other factors as transportation costs, the availability of freighter space, planned and unplanned delays in shipping dates, and other such operational factors (see Ruppel 1987).

The sample period of 1975:7 to 1986:12 was chosen for two reasons. First, in order to observe exchange rate effects over a period of largely flexible market-determined rates, a post 1973 sample period was necessary. Second, in the period prior to the middle of 1975 there existed great instability in US grain export sales. Direct and lingering effects from the oil shocks of 1972-73 probably contributed significantly to this instability, as well as the shift of exchange rate regimes, the growth and linking of world capital markets, and low real interest rates (Orden 1987). As a consequence, the estimation period begins in 1975:7 at a time when, to a large degree, stability returned to the export sales markets.66

66 As a result of this instability, the export sales data prior to 1975:7 included many negative entries, making the use of the logarithmic transformation impossible. In the period after 1975:7 two negative values appear (one each in the wheat sales and corn sales series). This is the result of an enormous cancellation of wheat and corn sales to Afghanistan following the Soviet Union’s invasion of that country. In order to deal with these two negative observations we have assigned a relatively small positive value in place of the negative value such that the logarithm of that entry will result in the lowest logged value of the series over the sample period without substantially altering the existing smoothness of the series. By assigning an even
The exchange rate used in this study is the total agricultural, real, trade-weighted exchange rate calculated by the Demand and Trade Section of the USDA. This overall index is calculated as follows: The weights for the index are average value shares of US commercial agricultural exports from 1976-78. The current real exchange rate for each country is computed by taking the ratio of the same period CPI in the US to that of the country in question and multiplying by the period average spot rate. These weighted changes are then summed into a total which is the real index.

4.2 Data Analysis and Tests of Time Series Properties

An examination of the time plots of each of the raw series suggested that each be expressed in natural logarithms to account for apparently nonstationary variances.67 The time plots of the logged series appear in figures 1 through 7. The plots of the exchange rate, wheat price, corn price, and soybean price all indicate possible nonstationarity in mean. The plots of wheat sales, corn sales, and soybean sales reveal a relatively more stable sample mean and hence these data series are probably stationary.

smaller positive value instead, we perhaps could have more faithfully represented the Afghanistan cancellation as a true outlier. It is our contention, however, that this cancellation is truly out of place in the series involved and could, if represented faithfully, substantially distort the results (McCleary and Hay 1983).

67 Henceforth when a variable is referred to as "wheat price" or "corn sales" it should be understood that it is the logarithm of the wheat price and corn sales, and similarly for all variables.
Figure 1. Time plot of logged wheat price

Figure 2. Time plot of logged corn price
Figure 3. Time plot of logged soybean price

Figure 4. Time plot of logged wheat sales

Chapter 4 - Detecting Macroeconomic Impacts on Agriculture
Figure 5. Time plot of logged corn sales

Figure 6. Time plot of logged soybeans sales

Chapter 4 - Detecting Macroeconomic Impacts on Agriculture
Figure 7. Time plot of logged agricultural exchange rate
Estimates of the autocorrelations and partial autocorrelations of the levels and first differences of each series provide a useful starting point for the evaluation of possible nonstationarities, and are reported in tables 1 and 2. For the exchange rate and each of the agricultural prices, the autocorrelations of the levels are large at low lags and decay slowly. Autocorrelations through lag 24 are significant for the exchange rate, wheat price, and soybean price. For the corn price, autocorrelations through lag 17 are significant.\(^4\) By comparison, the autocorrelations of the first differences of these series are insignificant after lag 1, indicating the presence of at most one unit root in the data generating processes of the exchange rate, wheat price, corn price, and soybean price. For the sales variables, the autocorrelations of the levels die out by at most the third lag, indicating that these variables are probably stationary in their levels.

As the autocorrelations of the series to be modeled indicate possible nonstationarity for some of the variables, formal tests for unit roots were conducted. For each of the variables tested, the simple DF regression showed signs of serial correlation in the residuals according to the correlogram and the Ljung-Box Q-Statistic (see Ljung and Box 1978).\(^5\) To remedy this lack of fit, higher order autoregressions were estimated. For the wheat price a two-lag model proves sufficient to eliminate serial correlation in the residuals of the resulting model, and for all other variables a one-lag model appears sufficient.

\(^4\) Following the discussion in section 2.6, approximate 95% confidence bounds inside which estimated autocorrelations can be regarded as zero are equal to \(1.96/\sqrt{136} = \pm .17\).

\(^5\) The null hypothesis is no serial correlation, consequently model adequacy is rejected for large values of this statistic.
<table>
<thead>
<tr>
<th>Lag</th>
<th>Wheat Price ACF</th>
<th>Wheat Price PACF</th>
<th>Corn Price ACF</th>
<th>Corn Price PACF</th>
<th>Soybean Price ACF</th>
<th>Soybean Price PACF</th>
<th>Wheat Sales ACF</th>
<th>Wheat Sales PACF</th>
<th>Corn Sales ACF</th>
<th>Corn Sales PACF</th>
<th>Soybean Sales ACF</th>
<th>Soybean Sales PACF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.96</td>
<td>.96</td>
<td>.94</td>
<td>.94</td>
<td>.96</td>
<td>.96</td>
<td>.51</td>
<td>.51</td>
<td>.43</td>
<td>.43</td>
<td>.36</td>
<td>.36</td>
</tr>
<tr>
<td>2</td>
<td>.90</td>
<td>-.22</td>
<td>.86</td>
<td>-.21</td>
<td>.90</td>
<td>-.26</td>
<td>.27</td>
<td>.02</td>
<td>.28</td>
<td>.11</td>
<td>.23</td>
<td>.12</td>
</tr>
<tr>
<td>3</td>
<td>.84</td>
<td>-.02</td>
<td>.76</td>
<td>-.12</td>
<td>.84</td>
<td>-.02</td>
<td>.16</td>
<td>.02</td>
<td>.18</td>
<td>.03</td>
<td>.06</td>
<td>-.07</td>
</tr>
<tr>
<td>4</td>
<td>.78</td>
<td>.00</td>
<td>.68</td>
<td>.05</td>
<td>.79</td>
<td>.13</td>
<td>.10</td>
<td>.01</td>
<td>.14</td>
<td>.04</td>
<td>.02</td>
<td>-.01</td>
</tr>
<tr>
<td>5</td>
<td>.72</td>
<td>.02</td>
<td>.61</td>
<td>.10</td>
<td>.75</td>
<td>.03</td>
<td>-.05</td>
<td>-.14</td>
<td>.04</td>
<td>-.07</td>
<td>-.08</td>
<td>-.09</td>
</tr>
<tr>
<td>6</td>
<td>.72</td>
<td>.02</td>
<td>.56</td>
<td>.02</td>
<td>.72</td>
<td>.06</td>
<td>-.02</td>
<td>.07</td>
<td>.04</td>
<td>.03</td>
<td>.09</td>
<td>.17</td>
</tr>
<tr>
<td>7</td>
<td>.63</td>
<td>.05</td>
<td>.53</td>
<td>.10</td>
<td>.69</td>
<td>-.00</td>
<td>.08</td>
<td>.11</td>
<td>.05</td>
<td>.03</td>
<td>-.05</td>
<td>-.12</td>
</tr>
<tr>
<td>8</td>
<td>.59</td>
<td>.10</td>
<td>.50</td>
<td>-.04</td>
<td>.66</td>
<td>.03</td>
<td>.04</td>
<td>-.05</td>
<td>.00</td>
<td>-.04</td>
<td>-.05</td>
<td>-.05</td>
</tr>
<tr>
<td>9</td>
<td>.56</td>
<td>-.04</td>
<td>.47</td>
<td>.01</td>
<td>.64</td>
<td>.03</td>
<td>.11</td>
<td>.13</td>
<td>-.03</td>
<td>-.03</td>
<td>.06</td>
<td>.15</td>
</tr>
<tr>
<td>10</td>
<td>.53</td>
<td>-.01</td>
<td>.45</td>
<td>.02</td>
<td>.62</td>
<td>-.05</td>
<td>.14</td>
<td>.03</td>
<td>.03</td>
<td>.06</td>
<td>.13</td>
<td>.09</td>
</tr>
<tr>
<td>11</td>
<td>.50</td>
<td>-.04</td>
<td>.42</td>
<td>-.04</td>
<td>.59</td>
<td>-.04</td>
<td>.23</td>
<td>.14</td>
<td>.10</td>
<td>.10</td>
<td>.19</td>
<td>.15</td>
</tr>
<tr>
<td>12</td>
<td>.46</td>
<td>-.07</td>
<td>.39</td>
<td>-.08</td>
<td>.55</td>
<td>-.03</td>
<td>.25</td>
<td>.13</td>
<td>.24</td>
<td>.22</td>
<td>.24</td>
<td>.08</td>
</tr>
<tr>
<td>13</td>
<td>.43</td>
<td>.01</td>
<td>.34</td>
<td>-.04</td>
<td>.52</td>
<td>-.01</td>
<td>.19</td>
<td>-.04</td>
<td>.08</td>
<td>-.14</td>
<td>.06</td>
<td>-.12</td>
</tr>
<tr>
<td>14</td>
<td>.34</td>
<td>.03</td>
<td>.29</td>
<td>-.01</td>
<td>.49</td>
<td>.07</td>
<td>.05</td>
<td>-.11</td>
<td>.17</td>
<td>.12</td>
<td>.06</td>
<td>.08</td>
</tr>
<tr>
<td>15</td>
<td>.36</td>
<td>-.04</td>
<td>.24</td>
<td>-.05</td>
<td>.47</td>
<td>-.03</td>
<td>.08</td>
<td>.12</td>
<td>.25</td>
<td>.17</td>
<td>.07</td>
<td>.05</td>
</tr>
<tr>
<td>16</td>
<td>.33</td>
<td>.01</td>
<td>.21</td>
<td>.05</td>
<td>.45</td>
<td>.04</td>
<td>-.01</td>
<td>-.12</td>
<td>.14</td>
<td>-.09</td>
<td>.02</td>
<td>-.03</td>
</tr>
<tr>
<td>17</td>
<td>.30</td>
<td>-.03</td>
<td>.18</td>
<td>.01</td>
<td>.43</td>
<td>-.02</td>
<td>-.06</td>
<td>.01</td>
<td>.06</td>
<td>-.04</td>
<td>-.12</td>
<td>-.16</td>
</tr>
<tr>
<td>18</td>
<td>.27</td>
<td>.00</td>
<td>.16</td>
<td>-.01</td>
<td>.42</td>
<td>.08</td>
<td>-.05</td>
<td>-.01</td>
<td>.11</td>
<td>.06</td>
<td>.07</td>
<td>.17</td>
</tr>
<tr>
<td>19</td>
<td>.25</td>
<td>.01</td>
<td>.14</td>
<td>-.01</td>
<td>.41</td>
<td>.03</td>
<td>-.07</td>
<td>-.14</td>
<td>.06</td>
<td>-.03</td>
<td>-.14</td>
<td>-.15</td>
</tr>
<tr>
<td>20</td>
<td>.23</td>
<td>-.03</td>
<td>.13</td>
<td>.01</td>
<td>.40</td>
<td>-.02</td>
<td>-.09</td>
<td>-.01</td>
<td>-.02</td>
<td>-.03</td>
<td>-.01</td>
<td>.04</td>
</tr>
<tr>
<td>21</td>
<td>.21</td>
<td>.08</td>
<td>.12</td>
<td>-.00</td>
<td>.39</td>
<td>-.07</td>
<td>-.03</td>
<td>.05</td>
<td>-.05</td>
<td>-.05</td>
<td>.11</td>
<td>.13</td>
</tr>
<tr>
<td>22</td>
<td>.20</td>
<td>.02</td>
<td>.10</td>
<td>-.01</td>
<td>.36</td>
<td>-.11</td>
<td>.08</td>
<td>.04</td>
<td>.00</td>
<td>.02</td>
<td>.12</td>
<td>-.04</td>
</tr>
<tr>
<td>23</td>
<td>.20</td>
<td>.06</td>
<td>.09</td>
<td>-.01</td>
<td>.33</td>
<td>-.00</td>
<td>.12</td>
<td>.09</td>
<td>-.09</td>
<td>-.11</td>
<td>.03</td>
<td>-.00</td>
</tr>
<tr>
<td>24</td>
<td>.20</td>
<td>.02</td>
<td>.07</td>
<td>.03</td>
<td>.31</td>
<td>.10</td>
<td>.16</td>
<td>.02</td>
<td>-.06</td>
<td>-.02</td>
<td>.01</td>
<td>-.15</td>
</tr>
</tbody>
</table>

Note: Standard errors are approximately 1.96/SQRT (136) = ± 0.17; where 136 is sample size.
| Lag | Wheat Price ACF | Wheat Price PACF | Corn Price ACF | Corn Price PACF | Soybean Price ACF | Soybean Price PACF | Wheat Sales ACF | Wheat Sales PACF | Corn Sales ACF | Corn Sales PACF | Soybean Sales ACF | Soybean Sales PACF |
|-----|----------------|------------------|---------------|----------------|------------------|------------------|------------------|----------------|----------------|----------------|----------------|------------------|------------------|
| 1   | .34            | .43              | .37           | -.27           | -.37             | -.37             | -.37             | -.37           | -.37           | -.37           | -.37           | -.37             | -.37             |
| 2   | -.06           | -.20             | .12           | -.08           | .06              | -.09             | -.12             | -.21           | -.04           | -.21           | .36             | -.14             | -.14             |
| 3   | -.13           | -.04             | -.14          | -.20           | -.18             | -.21             | -.05             | -.16           | -.06           | -.18           | -.11           | -.18             | -.08             |
| 4   | -.09           | -.05             | -.18          | -.04           | -.16             | -.02             | .09              | -.00           | .06            | -.06           | .05             | -.08             | -.05             |
| 5   | .02            | .06              | -.08          | .06            | -.13             | -.06             | -.18             | -.21           | -.10           | -.16           | -.22           | -.30             | -.20             |
| 6   | -.03           | -.10             | -.12          | -.16           | -.06             | -.03             | -.06             | -.22           | .01            | -.14           | .25             | .01              | -.11             |
| 7   | -.10           | -.07             | -.01          | -.04           | -.07             | -.08             | .14              | -.02           | .04            | -.06           | -.11           | -.07             | -.07             |
| 8   | -.06           | .00              | -.17          | -.13           | -.05             | -.05             | -.11             | -.20           | -.01           | -.07           | -.09           | -.25             | -.14             |
| 9   | -.02           | -.02             | -.06          | .06            | .06              | .09              | .03              | -.09           | -.09           | -.16           | .04             | -.14             | -.10             |
| 10  | .04            | .03              | .07           | .07            | .01              | -.08             | -.05             | -.18           | .00            | -.18           | -.00           | -.19             | -.09             |
| 11  | .10            | .07              | .18           | .09            | -.01             | -.04             | .06              | -.15           | -.06           | -.27           | .01             | -.10             | -.10             |
| 12  | .07            | .02              | .18           | .02            | -.05             | -.03             | .07              | .01            | .27            | .09            | .18             | .09              | -.09             |
| 13  | .00            | -.02             | .04           | -.07           | -.14             | -.15             | .09              | .08            | -.21           | -.16           | -.14           | -.12             | -.12             |
| 14  | -.03           | -.00             | -.08          | -.07           | -.12             | -.04             | -.17             | -.15           | .00            | -.22           | -.01           | -.07             | -.07             |
| 15  | -.03           | -.01             | -.11          | .01            | -.07             | -.04             | .13              | .10            | .17            | .05            | .05             | -.00             | -.00             |
| 16  | .04            | .05              | -.03          | .03            | -.02             | -.05             | -.05             | -.04           | -.04           | .01            | .07             | .12              | -.12             |
| 17  | -.01           | -.06             | -.08          | -.13           | -.09             | -.15             | -.34             | -.01           | -.11           | -.10           | -.26           | -.21             | -.21             |
| 18  | -.04           | .01              | -.11          | -.06           | -.05             | -.05             | .02              | .10            | .08            | .01            | .31             | .11              | .11              |
| 19  | .03            | .06              | .05           | .09            | -.01             | -.02             | -.01             | -.02           | .03            | -.01           | -.26           | -.08             | -.08             |
| 20  | -.04           | -.10             | .00           | .02            | .10              | .04              | -.09             | -.08           | .04            | .02            | -.01           | -.17             | -.17             |
| 21  | -.06           | -.03             | .04           | -.07           | .15              | .13              | -.05             | -.07           | -.03           | .09            | .02             | -.02             | -.02             |
| 22  | -.06           | -.04             | .06           | -.01           | .08              | -.03             | .06              | -.13           | .13            | .09            | .09             | -.01             | -.01             |
| 23  | -.09           | -.08             | .02           | -.04           | -.09             | -.16             | .01              | .05            | -.12           | .00            | -.06           | .13              | .13              |
| 24  | -.04           | -.02             | -.06          | -.08           | -.10             | -.01             | .07              | -.01           | -.00           | .09            | -.04           | -.04             | -.04             |

Note: Standard errors of the estimates are $1.96/\sqrt{135} = \pm 0.17$; where 135 is sample size.
The results of the tests for unit roots are summarized in table 3. There is evidence for the presence of unit roots in the exchange rate, wheat price, corn price, and soybean price. The evidence for the presence of a unit root is most conclusive for the exchange rate and the wheat price series. In these two cases, none of the test statistics recommended by Dickey and Fuller (see Chapter 2, section 2.4) suggests rejection of the null hypothesis, and we therefore conclude that each series contains a unit root.

For the corn price, the value of the test statistic $\tau_\mu$ suggests that we do not reject the null hypothesis of a unit root. The statistic $\tau_\tau$ provides further evidence of a unit root, though at a smaller level of confidence. The value of the statistic $\Phi_1$ indicates that the joint-null hypothesis $(\alpha, \rho) = (0,1)$ -- see equation (2.4.5) -- is not rejected, and the value of the statistic $\Phi_2$ indicates that the joint null hypothesis $(\alpha, \beta, \rho) = (0,0,1)$ is also not rejected. Lastly, the joint null hypothesis $(\alpha, \beta, \rho) = (\alpha,0,1)$ is rejected at the 10-percent level based on the value of $\Phi_3$, though this does not provide convincing evidence against the presence of a unit root when considered together with the other tests. We therefore conclude that the corn price is nonstationary as a result of the presence of a unit root.

The unit root tests on the soybean price are least conclusive. The statistic $\tau_\mu$ falls in the rejection region at the 10-percent level, and the statistic $\tau_\tau$ falls in the rejection region at the 1-percent level. The statistic $\Phi_1$ does not lie in the rejection region. The statistics $\Phi_2$ and $\Phi_3$ fall in rejection regions at the 5-percent and the 1-percent levels respectively.\footnote{It is critical to remember that a rejection of the joint null which includes a restriction that $\rho = 1$ does not necessarily imply that we are rejecting that particular restriction.} To sort out the conflicting evidence on the
Table 3. Dickey-Fuller tests for unit roots

<table>
<thead>
<tr>
<th>test</th>
<th>wheat price(2)</th>
<th>corn price(1)</th>
<th>soybean price(1)</th>
<th>wheat sales(1)</th>
<th>corn sales(1)</th>
<th>soybean sales(1)</th>
<th>exchange rate(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tu</td>
<td>-2.28</td>
<td>-2.30</td>
<td>-2.38</td>
<td>-4.95</td>
<td>-5.95</td>
<td>-5.71</td>
<td>1.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tt</td>
<td>-2.81</td>
<td>-3.25</td>
<td>-4.33</td>
<td></td>
<td></td>
<td></td>
<td>-0.89</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ph11</td>
<td>3.39</td>
<td>3.21</td>
<td>3.15</td>
<td>12.37</td>
<td>17.65</td>
<td>16.34</td>
<td>2.07</td>
</tr>
<tr>
<td>Ph12</td>
<td>3.77</td>
<td>4.08</td>
<td>8.14</td>
<td></td>
<td></td>
<td></td>
<td>3.63</td>
</tr>
<tr>
<td>Ph13</td>
<td>4.30</td>
<td>5.58</td>
<td>9.36</td>
<td></td>
<td></td>
<td></td>
<td>4.09</td>
</tr>
</tbody>
</table>

Note: The sample period for these tests was 1974:1-1985:4 for the wheat price, corn price, and soybean price, and 1975:8-1985:2 for the wheat sales, corn sales, and soybean sales. Because these tests are asymptotically valid tests, the larger available data set for the price variables was used.

The number in parentheses beside each variable name indicates the number of own lags in the Dickey-Fuller regression. Critical values of the test statistics are from Fuller (1976, p.373) and Dickey and Fuller (1981, p.1063).

a - reject at 1-percent level
b - reject at 5-percent level
c - reject at 10-percent level
results of the unit root tests for soybean price, we consider the empirical power of these unit root tests. Dickey and Fuller (1981) rank the tests, denoted by their corresponding statistics, on the basis of their power as follows: $\Phi_1 > \Phi_2 > \Phi_3$ and $\Phi_3 > \tau_*$. For soybean price, the most powerful test provides an indication that we should not reject the null of a unit root.\textsuperscript{71} We therefore conclude that the proper form for modeling the soybean price series is in first differences.

Unlike the evidence for the exchange rate and price series, the results of the unit root tests on the sales variables conclusively reject the presence of unit roots in each of the series tested. Because the time plots of the sales series reveal no steady increase or decrease in the mean of the series (as was the case for the exchange rate and the prices) we do not employ the tests $\tau_1$, $\Phi_2$, or $\Phi_3$ which have linear trends as alternative hypotheses. For each sales variable, the results of unit root tests are consistent. In each case we reject the presence of a unit root at the 1-percent level.\textsuperscript{72}

As mentioned in Chapter 2, autocorrelations are useful for evaluating seasonality. To get a more accurate picture of seasonal tendencies in the data, plots of the ACFs and PACFs of the first differences of the exchange rate and the agricultural prices, and the levels of the sales variables are provided in figures 8

\textsuperscript{71} Consistent with this conclusion, Bradshaw and Orden (1988) found that forecasting models for the soybean price series specified in differences outperform models of the series specified in levels in an out-of-sample competition.

\textsuperscript{72} Stock and Watson (1987b) also suggest checking the difference specifications for quadratic trends as well, however, we follow Nelson and Plosser (1982) who take the position that for a log-differenced series to have a deterministic trend would imply that rates of change are ever increasing ($\beta > 0$) or ever decreasing ($\beta < 0$), a curious hypothesis for most economic variables, except, perhaps, a controlled variable such as the money supply.
Figure 16. Autocorrelation function of wheat sales
Figure 17. Partial autocorrelation function of wheat sales
Figure 18. Autocorrelation function of corn sales
Figure 19. Partial autocorrelation function of corn sales
through 21. The exchange rate shows no spikes at any of the seasonal lags as might be expected. The wheat price and the soybean price show no significant spikes at the seasonal lags either. The corn price shows a seasonal spike at lag 12 that is barely significant (.188). However, there is no such spike at lag 24 (in fact the autocorrelation there is negative) indicating that seasonal adjustment is not warranted. In the case of wheat sales, a significant seasonal spike appears at lag 12, though at lag 24 the autocorrelation is not significant. The evidence of seasonality is weak in this case as well. For the corn sales and soybean sales there are also spikes at lag 12 though not at other seasonal lags. The lag 24 autocorrelation for the corn sales is negative (-0.06) and for soybean sales it is virtually zero (0.01). In sum there is slight but not compelling evidence of seasonal variation.\footnote{The dashed lines in figures 8-21 are 95-percent confidence bands (see Chapter 2 section 2.6).}

The next step in the data analysis is to test for co-integration between the I(1) series, specifically between each of the prices and the exchange rate.\footnote{Bessler and Babula (1987) in investigating the same wheat price and wheat sales variables as in this study seasonally adjust before examining the autocorrelations. The analysis here suggests that seasonal filtering may be unnecessary.} To conduct the co-integration test, the exchange rate is regressed on each of the commodity prices. The residuals from these co-integrating regressions are then tested for unit roots using the tables in Engle and Yoo (1987). The results for the co-integration tests are found in table 4. For each agricultural price-exchange rate system, a simple DF regression showed no signs of serial correlation in the

\footnote{Because the sales variables uniformly rejected the unit root hypothesis it is not necessary to test these series for co-integration with the exchange rate (see Chapter 4).}
Table 4. Tests for co-integration in bivariate systems

<table>
<thead>
<tr>
<th>test</th>
<th>wheat price exchange rate</th>
<th>corn price exchange rate</th>
<th>soybean price exchange rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>DF</td>
<td>-2.29</td>
<td>-2.40</td>
<td>-2.30</td>
</tr>
</tbody>
</table>

Note: Critical values were interpolated from Engle and Yoo (1987, p.157) for a sample size of 115.
residuals, allowing us to use the more powerful DF co-integration test. In each case, we do not reject the null hypothesis of *no co-integration*.

Based on the foregoing data analysis, we maintain that the appropriate time series specifications for the exchange rate, wheat price, corn price, and soybean price series are in first differences. The export sales variables, however, will be modeled in their levels.

The importance of modeling stationary series when forecasting, is demonstrated in a recent empirical study by Bradshaw and Orden (1988). The authors used the same agricultural price data as in this study. Their study finds that when careful attention is paid to possible nonstationarity in the data by testing for unit roots, and, if appropriate, modeling univariate and bivariate models in differences, better forecasting models can be constructed than when unit roots are ignored. In addition, models which simply included linear time trends and ignored the *form* of the nonstationarity (i.e., that the data series were I(1) and not trending deterministically) produced poorer forecasts for both agricultural prices and crop specific exchange rates than the models specified in differences. Such results were observed in comparisons between *univariate* levels plus trend models and differences models as well as between *bivariate* VARs in levels with trend and bivariate VARs in differences. These results argue for testing for unit roots *before* constructing forecasting models, as in many instances ignoring nonstationarity can cost the forecaster substantial losses in accuracy.
4.3 Evaluation of Forecasts

Introduction

Work by Ashley et al. (1980), Ashley (1981), and Bessler and Babula (1987) has focused on using forecasting ability to aid in determining the validity of cross-variable interactions based on the principle of Granger causality (see Granger 1980). The basic definition of Granger causality is this: Given two stochastic time series $X_t$ and $Y_t$, $Y_t$ is said to cause $X_t$ if and only if an optimal forecasting model for $X_t$ using past values of $X_t$ and $Y_t$ performs better than one using only past values of $X_t$.

Thus, Granger causality is a statement concerning forecasting performance.

Assessing Granger causality by testing the significance of blocks of cross-lags in VAR models has been standard procedure for some time. Ashley (1981) comments on this procedure by noting that in small samples misleading results can obtain because the same sample data are used to estimate the causal relationship as well as to evaluate its significance. Consequently, an out-of-sample comparison of nested models seems a more natural evaluative technique, closer in spirit to the definition of Granger causality than in-sample tests.

Until recently, there existed no statistical procedure for testing the significance of differences in the out-of-sample performance of two competing models. Because the forecasts are typically of the same variable, they undoubtedly would

---

76 This definition appears in Ashley (1981, p.651).

77 See also Ashley et al. (1980) p. 1156-57.
not be independent of one another, making the use of conventional hypothesis tests inappropriate. Ashley et al. (1980) outline the following out-of-sample test for Granger causality based on 1-step ahead prediction errors. The 1-step ahead errors should be used when employing the test of the significance of a decrease in MSE because 1-step ahead forecast errors have the desirable property of being white noise for optimal linear predictors. Typically, forecast errors of more than 1-step ahead are not white noise and therefore cannot be considered independent of one another.

Let \( \varepsilon_t^u \) represent a set of 1-step ahead postsample forecast errors of a univariate model for a particular variable, and \( \varepsilon_t^b \) be a set of 1-step ahead forecast errors from a bivariate model. Following Ashley's (1981) notation, we can then define the following linear combinations of these series:

\[
\text{DIF}_t = \varepsilon_t^u - \varepsilon_t^b
\]

\[
\text{SUM}_t = \varepsilon_t^u + \varepsilon_t^b
\]

The following regression equation can then be estimated:

\[
\text{DIF}_t = \alpha_1 + \alpha_2 (\text{SUM}_t - \overline{\text{SUM}}) + v_t
\]

where \( \overline{\text{SUM}} \) is the sample mean of \( \text{SUM}_t \), \( t = 1, \ldots, T \).

Ashley et al. (1980, p.1154) show that \( \alpha_1 \) is the difference in mean forecast error, and \( \alpha_2 \) is proportional to the difference in forecast error variance between the univariate and bivariate models. Testing the significance of the decrease in
the mean square forecasting error (in going from the univariate to the bivariate model) can be accomplished by testing the following hypothesis:

\[ H_0: \alpha_1 = \alpha_2 = 0 \]  \hspace{1cm} (4.4.3a)

\[ H_A: \alpha_1 > 0 \text{ and/or } \alpha_2 > 0 \]  \hspace{1cm} (4.4.3b)

If we reject the joint null hypothesis \( H_0 \) we can conclude that the bivariate model outperforms the univariate model. If either of the two estimated coefficients is significantly negative than we cannot conclude that the bivariate model is an improvement. If one coefficient is negative but not significant, a one-tailed t-test can be used on the other coefficient, and an F-test can be employed if the two coefficients are positive.\(^7\)

4.4 Specification of Forecasting Models

**Specification of Univariate Forecasting Models**

For the present investigation four types of univariate time series forecasting models will be specified. First, following the Box-Jenkins approach outlined in Chapter 2, ARIMA models will be constructed for each of the six agricultural

---

\(^7\) The authors contend that the F-test is *four-tailed* because it does not take into account the signs of the estimated coefficients. Thus one should report a significance level equal to one-fourth of that provided by the standard F-tables. In addition, if the sample mean of the forecast errors from either models is negative, the series must be multiplied by -1 before running regression 4.4.2, see Brandt and Bessler (1983) on this point.
price and export sales variables. Second, three univariate autoregressive models will be specified for each of these series. The orders of the autoregressive models will be chosen by minimizing the Bayesian Information Criterion (BIC), the Hannan-Quinn criterion (HQ), and the Final Prediction Error criterion (FPE).\textsuperscript{79} The sample mean and a simple random walk will be used as benchmarks to evaluate the absolute quality of the various univariate forecasting models. The estimation period for all of the models, both univariate and multivariate, is from 1975 (the month depending on the number of lags in the equation or system) to 1985:2. A 22-month postsample period of available data, from 1985:3 to 1986:12, was held out over which to evaluate the forecasting accuracy of each of the models.

Table 5 contains the final specifications for the four univariate forecasting models for each agricultural variable. Among the ARIMA models of the sales variables, there are no autoregressive terms in the corn model, one in the wheat model, and two in the soybean model. Moving average terms were found to improve the forecast accuracy of all of the sales variables. The ARIMA models for the prices, in contrast, were found to be low-order autoregressions. The BIC and HQ criteria agree in every case on the selection of the orders of approximating autoregressions and thus are listed together in table 5. The FPE criterion selects higher order autoregressions than the BIC and HQ criteria.\textsuperscript{80} Finally, inclusion

\textsuperscript{79} These are the univariate analogs of the criteria introduced in Chapter 3. For a complete discussion see Judge et al. 1985.

\textsuperscript{80} This is true in general. In assessing Granger causality by in-sample hypothesis testing this bias-efficiency tradeoff might play a critical role in the outcome. The lower-order BiC and HQ models are likely to have biased coefficients but more precise variance estimates, while the higher-order FPE models are more likely to have coefficients that are unbiased but inefficiently estimated.
<table>
<thead>
<tr>
<th>Variable Forecasted</th>
<th>ARIMA</th>
<th>BIC/HQ</th>
<th>FPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wheat Sales</td>
<td>(1,0,1)</td>
<td>(1,0)</td>
<td>(10,0)</td>
</tr>
<tr>
<td>Corn Sales</td>
<td>(0,0,1)</td>
<td>(1,0)</td>
<td>(12,0)</td>
</tr>
<tr>
<td>Soybean Sales</td>
<td>(2,0,1)</td>
<td>(2,0)</td>
<td>(10,0)</td>
</tr>
<tr>
<td>Wheat Price</td>
<td>(2,1,0)</td>
<td>(1,1)</td>
<td>(7,1)</td>
</tr>
<tr>
<td>Corn Price</td>
<td>(3,1,0)</td>
<td>(1,1)</td>
<td>(6,1)</td>
</tr>
<tr>
<td>Soybean Price</td>
<td>(3,1,0)</td>
<td>(1,1)</td>
<td>(3,1)</td>
</tr>
</tbody>
</table>

a - indicates that the constant term was excluded from the specification because it increased the postsample mean square error of the model.
or exclusion of constant terms in the specifications was determined solely on the basis of whether a given model performed better out-of-sample with or without the constant term.

**Specification of Multivariate Forecasting Models**

The VAR specifications for each price-exchange rate and each sales-exchange rate system include three unrestricted forms (UVAR), with lag lengths chosen by the multivariate versions of the BIC, HQ, and FPE criteria introduced in Chapter 3. In addition, restricted VARs were specified (RVAR) following Hsiao's (1979) procedure outlined in Chapter 3. The difference between the unrestricted forms chosen by the statistical criteria and the restricted forms from Hsiao's procedure is that the restricted forms allow for differing lag structures for each variable in each equation. Table 6 summarizes the lag structures for the various multivariate models.

The multivariate BIC and HQ criteria again agree on the orders of lags selected and are listed together in Table 6. The FPE criterion chooses much more heavily parameterized UVARs. Hsiao's procedure chose tentative models with a large number of lag restrictions compared to the FPE UVAR. Both cross-variable effects of exchange rates on prices and prices on exchange rates are suggested by Hsiao's procedure. The feedback relationships from the prices and sales to the exchange rate indicate that there may be a third variable driving both variables in each bivariate system. It is likely that the exchange rate is a transmission
Table 6. Orders of multivariate forecasting models

<table>
<thead>
<tr>
<th>Variable</th>
<th>BIC/HQ UVAR</th>
<th>FPE UVAR</th>
<th>Hsiao RVAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wheat Sales</td>
<td>1</td>
<td></td>
<td>a</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>9</td>
<td>10-10/9-1</td>
</tr>
<tr>
<td>Corn Sales</td>
<td>1</td>
<td></td>
<td>a</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>10</td>
<td>12-10/10-1</td>
</tr>
<tr>
<td>Soybean Sales</td>
<td>1</td>
<td></td>
<td>a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9</td>
<td>10-9/0-1</td>
</tr>
<tr>
<td>Wheat Price</td>
<td>1</td>
<td></td>
<td>a</td>
</tr>
<tr>
<td>Corn Price</td>
<td>1</td>
<td></td>
<td>7-10/10-1</td>
</tr>
<tr>
<td>Soybean Price</td>
<td>1</td>
<td></td>
<td>6-10/1-1</td>
</tr>
</tbody>
</table>

Note: The notation for the lag structure on the RVARs is as follows, a(B)-b(B)/c(B)-d(B):

- a(B) = Own lags of forecasted variable
- b(B) = Lags of exchange rate in forecasted variable equation.
- c(B) = Lags of forecasted variable in exchange rate equation.
- d(B) = Own lags of exchange rate.

a - Indicates that the constant term was excluded from the specification because it increased the postsample mean square error of the model.
mechanism of other macroeconomic shocks which influence the agricultural prices and sales through other transmission routes as well.

Tables 7 through 18 contain likelihood ratio (LR) statistics for overfit and underfit alternative models to the tentative models chosen by Hsiao’s procedure. For the overfit models we test for extra own-lags and extra lags of the exchange rate in the price and sales equations. Also, extra own-lags were considered for the exchange rate equation as Hsiao’s tentative model suggested only one lag. For the underfitting, we checked the significance of all feedback relationships (i.e., dependence of the exchange rate on the agricultural prices and sales variables). In addition, tests for the significance of the exchange rate in the price and sales equations were conducted.

For wheat sales, in table 7, the overfit model 3 does prove superior to the tentative model, and therefore we reject the tentative model in favor of alternative model 3. The underfitting procedure (table 8) for the most part validates the tentative model, though the statistic on model 2 could be considered too low to ignore. Hsaio (1979) notes that the FPE criterion is more liberal on including variables than the conventional significance levels of 5- and 10-percent, and as a consequence we are less strict about the underfitting p-values. The p-value of 17-percent could be interpreted to suggest that there is no Granger causality from the exchange rate to wheat sales. Since this hypothesis is the focus of our investigation, however, we will let the univariate model “argue for itself” in the out-of-sample competition to follow.
<table>
<thead>
<tr>
<th>polynomial</th>
<th>tentative model</th>
<th>alternate model 1</th>
<th>alternate model 2</th>
<th>alternate model 3</th>
<th>alternate model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>a(B)</td>
<td>9</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>b(B)</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>c(B)</td>
<td>9</td>
<td>10</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>d(B)</td>
<td>1</td>
<td>10</td>
<td>5</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>DF</td>
<td>11</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>LR</td>
<td>4.14</td>
<td>1.14</td>
<td>4.14</td>
<td>1.46</td>
<td></td>
</tr>
<tr>
<td>P-Value</td>
<td>.966</td>
<td>.888</td>
<td>.042</td>
<td>.918</td>
<td></td>
</tr>
</tbody>
</table>
Table 8. LR tests for underfitting wheat sales/exchange rate VAR.

<table>
<thead>
<tr>
<th>Polynomial</th>
<th>tentative model</th>
<th>alternate model 1</th>
<th>alternate model 2</th>
<th>alternate model 3</th>
<th>alternate model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>a(B)</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>b(B)</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>c(B)</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>d(B)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>DF</td>
<td></td>
<td>21.7</td>
<td>14.0</td>
<td>36.3</td>
<td>0.013</td>
</tr>
<tr>
<td>LR</td>
<td></td>
<td>21.7</td>
<td>14.0</td>
<td>36.3</td>
<td>0.013</td>
</tr>
<tr>
<td>P-Value</td>
<td></td>
<td>0.010</td>
<td>0.172</td>
<td>0.013</td>
<td>0.172</td>
</tr>
</tbody>
</table>
For the corn sales-exchange rate RVAR the overfit models 3 and 4 in table 9 seem to argue for more parameterization of both corn sales and the exchange rate in the first equation. Consequently the parameterization 12-10/10-1 was estimated. The LR statistic was 9.79 with 6 degrees of freedom and a p-value of 13.4-percent. On this basis we will include these extra lags in the specification. The underfitting exercise (table 10) indicates that the chosen model is adequate.

The soybean sales-exchange rate RVAR has a tentative model with no feedback. Overfitting (table 11) indicates that the tentative structure is viable. The underfitting results shown in table 12, however, reveal that a univariate structure may be appropriate as we do not reject that the lags on b(B) are zero. Again, we will let the univariate structure argue for itself in the out-of-sample competition.

Tables 13 and 14 contain the LR statistics for overfitting and underfitting the wheat price-exchange rate model. The statistics for the overfit models validate the tentative model. The underfitting suggests a univariate structure with alternative models 2 and 3 solidly in favor of no causality. We will, however, provisionally accept the tentative model.

The LR statistics for the overfitting of the corn price-exchange rate RVAR (table 15) indicate that more lags may be required in the feedback relationship. Alternative model 3 was estimated and found to be no improvement in out-of-sample performance over the tentative model. The underfitting procedure (table 16) suggested the elimination of the feedback relationship of corn price to the
### Table 9. LR tests for overfitting corn sales/exchange rate RVAR.

<table>
<thead>
<tr>
<th>polynomial model</th>
<th>tentative model</th>
<th>alternate model 1</th>
<th>alternate model 2</th>
<th>alternate model 3</th>
<th>alternate model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>a(B)</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>b(B)</td>
<td>6</td>
<td>10</td>
<td>6</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>c(B)</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>d(B)</td>
<td>1</td>
<td>10</td>
<td>5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>DF</td>
<td>13</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>LR</td>
<td>7.99</td>
<td>2.22</td>
<td>7.99</td>
<td>10.0</td>
<td></td>
</tr>
<tr>
<td>P-Value</td>
<td>.844</td>
<td>.696</td>
<td>.092</td>
<td>.007</td>
<td></td>
</tr>
</tbody>
</table>

### Table 10. LR tests for underfitting corn sales/exchange rate RVAR.

<table>
<thead>
<tr>
<th>polynomial model</th>
<th>tentative model</th>
<th>alternate model 1</th>
<th>alternate model 2</th>
<th>alternate model 3</th>
<th>alternate model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>a(B)</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>b(B)</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>c(B)</td>
<td>10</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>d(B)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>DF</td>
<td>10</td>
<td>6</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR</td>
<td>22.1</td>
<td>11.9</td>
<td>34.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-Value</td>
<td>.015</td>
<td>.063</td>
<td>.005</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 11. LR tests for overfitting soybean sales/exchange rate RVAR.

<table>
<thead>
<tr>
<th>polynomial</th>
<th>tentative model</th>
<th>alternate model 1</th>
<th>alternate model 2</th>
<th>alternate model 3</th>
<th>alternate model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>a(B)</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>b(B)</td>
<td>9</td>
<td>10</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>c(B)</td>
<td>0</td>
<td>10</td>
<td>5</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>d(B)</td>
<td>1</td>
<td>10</td>
<td>1</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>DF</td>
<td>20</td>
<td>5</td>
<td>4</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>LR</td>
<td>18.8</td>
<td>6.32</td>
<td>2.04</td>
<td>14.8</td>
<td></td>
</tr>
<tr>
<td>P-Value</td>
<td>.534</td>
<td>.277</td>
<td>.728</td>
<td>.140</td>
<td></td>
</tr>
</tbody>
</table>

Table 12. LR tests for underfitting soybean sales/exchange rate RVAR.

<table>
<thead>
<tr>
<th>polynomial</th>
<th>tentative model</th>
<th>alternate model 1</th>
<th>alternate model 2</th>
<th>alternate model 3</th>
<th>alternate model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>a(B)</td>
<td>10</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b(B)</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c(B)</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d(B)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DF</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR</td>
<td>7.93</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-Value</td>
<td>.535</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 13. LR tests for overfitting wheat price/exchange rate RVAR.

<table>
<thead>
<tr>
<th>polynomial</th>
<th>tentative model</th>
<th>alternate model 1</th>
<th>alternate model 2</th>
<th>alternate model 3</th>
<th>alternate model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>a(B)</td>
<td>7</td>
<td>10</td>
<td>7</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>b(B)</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>c(B)</td>
<td>10</td>
<td>10</td>
<td>12</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>d(B)</td>
<td>1</td>
<td>10</td>
<td>1</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>DF</td>
<td>12</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>LR</td>
<td>3.14</td>
<td>.903</td>
<td>2.14</td>
<td>2.64</td>
<td></td>
</tr>
<tr>
<td>P-Value</td>
<td>.964</td>
<td>.637</td>
<td>.709</td>
<td>.450</td>
<td></td>
</tr>
</tbody>
</table>

Table 14. LR tests for underfitting wheat price/exchange rate RVAR.

<table>
<thead>
<tr>
<th>polynomial</th>
<th>tentative model</th>
<th>alternate model 1</th>
<th>alternate model 2</th>
<th>alternate model 3</th>
<th>alternate model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>a(B)</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>---</td>
</tr>
<tr>
<td>b(B)</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>---</td>
</tr>
<tr>
<td>c(B)</td>
<td>10</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>---</td>
</tr>
<tr>
<td>d(B)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>---</td>
</tr>
<tr>
<td>DF</td>
<td>10</td>
<td>10</td>
<td>20</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>LR</td>
<td>12.6</td>
<td>3.52</td>
<td>16.2</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>P-Value</td>
<td>.246</td>
<td>.966</td>
<td>.706</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Polynomial</td>
<td>Tentative Model</td>
<td>Alternate Model 1</td>
<td>Alternate Model 2</td>
<td>Alternate Model 3</td>
<td>Alternate Model 4</td>
</tr>
<tr>
<td>------------</td>
<td>----------------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>a(B)</td>
<td>6</td>
<td>10</td>
<td>10</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>b(B)</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>c(B)</td>
<td>1</td>
<td>10</td>
<td>1</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>d(B)</td>
<td>1</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>DF</td>
<td></td>
<td>22</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>LR</td>
<td>11.0</td>
<td>5.10</td>
<td>7.58</td>
<td>6.45</td>
<td></td>
</tr>
<tr>
<td>P-Value</td>
<td>.975</td>
<td>.277</td>
<td>.108</td>
<td>.168</td>
<td></td>
</tr>
</tbody>
</table>

**Table 15. LR tests for overfitting corn price/exchange rate RVAR.**

<table>
<thead>
<tr>
<th>Polynomial</th>
<th>Tentative Model</th>
<th>Alternate Model 1</th>
<th>Alternate Model 2</th>
<th>Alternate Model 3</th>
<th>Alternate Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>a(B)</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>---</td>
</tr>
<tr>
<td>b(B)</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>---</td>
</tr>
<tr>
<td>c(B)</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>---</td>
</tr>
<tr>
<td>d(B)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>---</td>
</tr>
<tr>
<td>DF</td>
<td></td>
<td>1</td>
<td>10</td>
<td>11</td>
<td>---</td>
</tr>
<tr>
<td>LR</td>
<td>.934</td>
<td>17.4</td>
<td>18.3</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>P-Value</td>
<td>.334</td>
<td>.087</td>
<td>.075</td>
<td>---</td>
<td></td>
</tr>
</tbody>
</table>

**Table 16. LR tests for underfitting corn price/exchange rate RVAR.**
### Table 17. LR tests for overfitting soybean price/exchange rate RVAR.

<table>
<thead>
<tr>
<th>polynomial</th>
<th>tentative model</th>
<th>alternate model 1</th>
<th>alternate model 2</th>
<th>alternate model 3</th>
<th>alternate model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>a(B)</td>
<td>3</td>
<td>10</td>
<td>6</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>b(B)</td>
<td>10</td>
<td>10</td>
<td>6</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>c(B)</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>d(B)</td>
<td>1</td>
<td>10</td>
<td>1</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>DF</td>
<td>16</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>LR</td>
<td>6.91</td>
<td>4.60</td>
<td>6.16</td>
<td>1.87</td>
<td></td>
</tr>
<tr>
<td>P-Value</td>
<td>.975</td>
<td>.203</td>
<td>.187</td>
<td>.600</td>
<td></td>
</tr>
</tbody>
</table>

### Table 18. LR tests for underfitting soybean price/exchange rate RVAR.

<table>
<thead>
<tr>
<th>polynomial</th>
<th>tentative model</th>
<th>alternate model 1</th>
<th>alternate model 2</th>
<th>alternate model 3</th>
<th>alternate model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>a(B)</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>---</td>
</tr>
<tr>
<td>b(B)</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>---</td>
</tr>
<tr>
<td>c(B)</td>
<td>10</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>---</td>
</tr>
<tr>
<td>d(B)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>---</td>
</tr>
<tr>
<td>DF</td>
<td>10</td>
<td>10</td>
<td>20</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>LR</td>
<td>9.31</td>
<td>6.04</td>
<td>15.5</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>P-Value</td>
<td>.503</td>
<td>.812</td>
<td>.747</td>
<td>---</td>
<td></td>
</tr>
</tbody>
</table>
exchange rate. This model was also estimated and found to be no significant improvement out-of-sample. Consequently, the tentative model will be retained.

The final RVAR, this for the soybean price-exchange rate system, seems to require no additional variables according to the test results in table 17. Table 18 indicates that a univariate model is preferred, and again we shall wait for the out-of-sample competition to assess the validity of the causal relationship.

Hsiao’s procedure of diagnostic checking by over- and underfitting provides tentative in-sample evidence of Granger causality from the exchange rate to corn sales and corn price, and to wheat sales. No Granger causality is found from the exchange rate to the soybean sales, wheat price, and soybean price variables. This evidence will be investigated further in the next section.

4.5 Empirical Forecasting Results

Introduction

In this section the postsample MSE of the 1-step ahead forecasts from the estimated models is reported, and three issues are addressed. The first issue concerns the effect of alternative model specification procedures on out-of-sample tests for Granger causality. The second issue involves comparing the more commonly used in-sample tests for Granger causality with the out-of-sample tests described in Section 4.3. Finally, the issue of long-range versus short-range forecasting in relation to detecting Granger causality will be discussed.
Out-Of-Sample Tests for Granger Causality

Each of the univariate and multivariate equations whose orders were listed in tables 5 and 6 were used to produce a series of 1-step ahead out-of-sample forecasts of the log-levels of the agricultural price and export sales variables over the period 1985:3 to 1986:12. These forecasts were compared to actual data held out for this postsample period, and for each model the mean square error of the 1-step ahead forecasts (MSE) was calculated. These postsample mean square errors are shown in table 19 for each variable forecast and each method employed. For comparison, column 1 of table 19 lists the MSEs for a simple random walk (naive) of 1-step ahead forecasts. Column 2 shows the MSE of forecasts represented by the extrapolation of the sample mean.

In general, only for soybean sales is the naive model competitive with the univariate models, forecasting virtually as well as the BIC/HQ UAR model. Among the univariate models, none dominates across all of the variables forecasted. The BIC/HQ method produces the lowest MSE for the wheat sales and soybean sales variables, the FPE method has the lowest MSE for the corn price and corn sales variables, and the ARIMA model is best for the wheat price and soybean price variables.\(^8\)

As for the multivariate methods, Hsiao's restricted VARs were uniformly the preferred models in terms of producing the lowest MSE among all of the multivariate methods investigated. This provides strong evidence in favor of using restricted forms of VARs.

---

\(^8\) The MSEs for the ARIMA method and the FPE method forecasts of the soybean price are the same. This is because both methods chose an AR(3) specification.
Table 19. Postsample mean square error of 1-step ahead forecasts, all values E-03.

<table>
<thead>
<tr>
<th>Variable</th>
<th>NAIVE</th>
<th>MEAN</th>
<th>BIC/HQ UAR</th>
<th>ARIMA</th>
<th>FPE UAR</th>
<th>BIC/HQ VAR</th>
<th>FPE VAR</th>
<th>HSA10 RVAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wheat Sales</td>
<td>141.34</td>
<td>148.92</td>
<td>83.543</td>
<td>84.490</td>
<td>94.235</td>
<td>71.135</td>
<td>102.96</td>
<td>69.408</td>
</tr>
<tr>
<td>Corn Sales</td>
<td>198.35</td>
<td>409.79</td>
<td>197.75</td>
<td>298.69</td>
<td>193.08</td>
<td>201.24</td>
<td>237.14</td>
<td>167.22</td>
</tr>
<tr>
<td>Soy Sales</td>
<td>173.60</td>
<td>245.83</td>
<td>176.54</td>
<td>201.69</td>
<td>190.23</td>
<td>172.73</td>
<td>279.25</td>
<td>123.24</td>
</tr>
<tr>
<td>Wheat Price</td>
<td>3.3081</td>
<td>275.48</td>
<td>2.5138</td>
<td>2.2964</td>
<td>2.3689</td>
<td>2.6175</td>
<td>2.8388</td>
<td>2.0897</td>
</tr>
<tr>
<td>Soy Price</td>
<td>0.8555</td>
<td>290.33</td>
<td>0.7243</td>
<td>0.6821</td>
<td>0.6821</td>
<td>0.8347</td>
<td>1.6108</td>
<td>0.7899</td>
</tr>
</tbody>
</table>
On an absolute scale, comparing the estimated models with the random walk and sample mean forecasts, none of the methods produced MSEs across all variables forecasted that were uniformly worse than either the naive or sample mean methods. Only one model, the FPE UVAR on the soybean price-exchange rate system, is worse than either the naive or sample mean forecasts of the price series, while seven of the models for the sales variables are worse than the corresponding forecasts from the naive or sample mean methods. One general conclusion seems to be that the price series are less difficult to forecast than the sales series as most price models have MSEs well below both benchmarks. Soybean sales could be considered to be the most difficult sales variable to forecast accurately as it has the most models with MSEs above the naive and mean benchmarks.

Model Specification and Out-of-Sample Tests for Granger Causality

The first issue that is to be investigated here concerns how we interpret the exchange rate mattering (in a Granger causal sense) when different procedures are employed for lag length selection. Table 20 shows tests for the significance of decreases in MSE going from univariate BIC/HQ to bivariate BIC/HQ chosen models.\(^{22}\) Column 4 contains the marginal significance levels where such a relationship exists. In only one case is the decrease in MSE significant at a conventional significance level, the case of the wheat sales. These results are

---

\(^{22}\) In tables 20-23 the statistical test developed in Ashley et al. (1980) is referred to as the AGS test after the authors of the paper, Ashley, Granger, and Schmalensee.
Table 20. Results of AGS tests for the significance of differences in out of sample mean square error, BIC selected models.

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>BIC/HQ</th>
<th>BIC/HQ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UAR</td>
<td>UVAR</td>
</tr>
<tr>
<td>Wheat Sales</td>
<td>83.543</td>
<td>71.135</td>
</tr>
<tr>
<td>Corn Sales</td>
<td>197.75</td>
<td>201.24</td>
</tr>
<tr>
<td>Soy Sales</td>
<td>178.54</td>
<td>172.73</td>
</tr>
<tr>
<td>Wheat Price</td>
<td>2.6138</td>
<td>2.6175</td>
</tr>
<tr>
<td>Corn Price</td>
<td>4.6578</td>
<td>4.8197</td>
</tr>
<tr>
<td>Soy Price</td>
<td>0.7243</td>
<td>0.8347</td>
</tr>
</tbody>
</table>

Significance Level: 0.0045, 0.2546, ---

Chapter 4 - Detecting Macroeconomic Impacts on Agriculture
consistent with the results in Bradshaw and Orden (1988) on the commodity price forecasts.\textsuperscript{83}

Table 21 contains the results of the same experiment for the FPE selected models, both univariate and bivariate. In this case only the forecasts of the corn price from the bivariate model are lower MSE than the corresponding univariate forecasts, though the decrease in MSE is significant only at the 22-percent level.

Table 22 shows a comparison of the forecasts from the ARIMA method with those of Hsiao’s RVAR procedure. We pair these procedures because they are each inherently more flexible than the statistical decision procedures, involving more diagnostic checking. The RVAR dominates the ARIMA in all but the soybean price case. The decrease in MSE from the ARIMA to the RVAR models is significant at conventional levels in the cases of the wheat sales, corn sales, soybean sales, and corn price variables. Forecasts of the wheat price from the RVAR model are also statistically more accurate than forecasts from the ARIMA model at the 14-percent level, offering slightly less confidence in the improvement than conventional significance levels.

Table 23 shows a comparison of forecast MSE from the best univariate model and the best bivariate model for each of the agricultural variables. As noted earlier, the BIC criterion gives the best univariate model in two cases, the FPE in two cases, the ARIMA in 1 case, and the ARIMA/FPE give the same best forecasts in one case, while the RVAR specification gives the best bivariate model in all cases. The superiority of the bivariate model is statistically signif-

\textsuperscript{83} Bradshaw and Orden (1988) did not investigate forecasts of the sales variables.
Table 21. Results of AGS tests for the significance of differences in out of sample mean square error, FPE selected models.

<table>
<thead>
<tr>
<th>VARIABLE FORECASTED</th>
<th>FPE UAR</th>
<th>FPE RVAR</th>
<th>SIGNIFICANCE LEVEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wheat Sales</td>
<td>94.235</td>
<td>102.96</td>
<td>---</td>
</tr>
<tr>
<td>Corn Sales</td>
<td>193.08</td>
<td>237.14</td>
<td>---</td>
</tr>
<tr>
<td>Soy Sales</td>
<td>190.23</td>
<td>279.25</td>
<td>---</td>
</tr>
<tr>
<td>Wheat Price</td>
<td>2.3669</td>
<td>2.8388</td>
<td>---</td>
</tr>
<tr>
<td>Corn Price</td>
<td>4.2216</td>
<td>4.0348</td>
<td>0.2169</td>
</tr>
<tr>
<td>Soy Price</td>
<td>0.6821</td>
<td>1.6108</td>
<td>---</td>
</tr>
</tbody>
</table>

Table 22. Results of AGS tests for the significance of differences in out of sample mean square error, ARIMA and Hsiao's RVAR.

<table>
<thead>
<tr>
<th>VARIABLE FORECASTED</th>
<th>ARIMA</th>
<th>HSAIO'S RVAR</th>
<th>SIGNIFICANCE LEVEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wheat Sales</td>
<td>84.490</td>
<td>69.408</td>
<td>0.0594</td>
</tr>
<tr>
<td>Corn Sales</td>
<td>298.69</td>
<td>167.22</td>
<td>9.33E-05</td>
</tr>
<tr>
<td>Soy Sales</td>
<td>201.69</td>
<td>123.24</td>
<td>0.0984</td>
</tr>
<tr>
<td>Wheat Price</td>
<td>2.2964</td>
<td>2.0897</td>
<td>0.1372</td>
</tr>
<tr>
<td>Corn Price</td>
<td>4.3501</td>
<td>2.6679</td>
<td>2.76E-03</td>
</tr>
<tr>
<td>Soy Price</td>
<td>0.6821</td>
<td>0.7899</td>
<td>---</td>
</tr>
</tbody>
</table>
Table 23. Results of AGS tests for the significance of differences in out of sample mean square error, best univariate and best multivariate models.

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>BEST UNIVARIATE</th>
<th>BEST MULTIVARIATE</th>
<th>SIGNIFICANCE LEVEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wheat Sales</td>
<td>83.543 (BIC)</td>
<td>69.408 (RVAR)</td>
<td>0.0645</td>
</tr>
<tr>
<td>Corn Sales</td>
<td>193.08 (FPE)</td>
<td>167.22 (RVAR)</td>
<td>0.1344</td>
</tr>
<tr>
<td>Soy Sales</td>
<td>178.54 (BIC)</td>
<td>123.24 (RVAR)</td>
<td>0.1024</td>
</tr>
<tr>
<td>Wheat Price</td>
<td>2.2964 (ARIMA)</td>
<td>2.0897 (RVAR)</td>
<td>0.1372</td>
</tr>
<tr>
<td>Corn Price</td>
<td>4.2218 (FPE)</td>
<td>2.6679 (RVAR)</td>
<td>5.64E-03</td>
</tr>
<tr>
<td>Soy Price</td>
<td>0.6821 (ARIMA)</td>
<td>0.7899 (RVAR)</td>
<td>---</td>
</tr>
</tbody>
</table>
icant, at conventional levels, in the case of wheat sales, soybean sales (marginally), and corn price. At lower levels of confidence, 13.4-percent for the corn sales, and 13.7-percent for the wheat price, the bivariate models are also superior to the univariate models.

A comparison among tables 20-23 illustrates the importance of model specification in evaluating Granger causality from the exchange rate to agricultural price and export sales variables. The soybean price remains the only variable across which all of the methods find consistently, and in this case there is no evidence in support of Granger causality. For wheat sales, the exchange rate matters in a significant way in three of the four comparisons. Only in the case of the FPE selected models is the multivariate model inferior.

For corn sales, the evidence is more conflicting. For the BIC/HQ and FPE selected models, the MSEs of the multivariate models are higher than those of the univariate models, indicating that the exchange rate does not help forecast corn sales in a Granger sense. In contrast, the RVAR is significantly better than the ARIMA model and marginally better than the best univariate model (at the 13.4-percent significance level). For the soybean sales the RVAR again outforecasts the ARIMA and the best univariate model indicating a Granger causal relationship. However, had we only examined the FPE and BIC/HQ multivariate models versus their univariate counterparts, we would not have reached this conclusion.

For the wheat price and corn price variables, the results across tables 20-23 are equally inconsistent. In two cases, for the FPE and BIC/HQ selected models,
the exchange rate does not help forecast the wheat price. However, the results in tables 22 and 23 indicate that the evidence is marginally (13.7-percent level) in favor of the exchange rate Granger causing the wheat price. For the corn price, the results from table 20 indicate that the bivariate model is no better than the univariate model. The evidence is weak for the causal relationship according to the FPE univariate versus multivariate comparison, showing a significance level of 22-percent. In contrast, however, a very significant Granger causal relationship from the exchange rate to the corn price is observed when the RVAR multivariate model is compared to the ARIMA and the best univariate model.

These results reveal the difficulty involved in testing for Granger causality, and highlight the importance of finding "optimal" forecasting models, whether univariate or bivariate, with which to conduct the tests. The evidence offered here indicates that without an exhaustive search of the model space it is difficult to have confidence in the results of a given test for Granger causality.

Comparing In-Sample to Out-of-Sample Tests for Granger Causality

Table 24 (columns 1, 2, and 3) contains marginal significance levels for tests of the null hypothesis that lags of the exchange rate are jointly zero in the sales and price equations of the three VAR selection methods. An asterisk next to the marginal significance level indicates possible evidence in favor of Granger causality. Column 4 of table 24 contains the results from the out-of-sample AGS tests for Granger causality. Again, an asterisk indicates evidence in favor of Granger causality.
Table 24. In-Sample versus Out-of-Sample tests for Granger causality.

<table>
<thead>
<tr>
<th>Variable</th>
<th>BIC/HQ</th>
<th>FPE</th>
<th>RVAR</th>
<th>AGS TEST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wheat Sales</td>
<td>.106</td>
<td>.367</td>
<td>.172</td>
<td>.065</td>
</tr>
<tr>
<td>Corn Sales</td>
<td>.621</td>
<td>.127</td>
<td>.063</td>
<td>.134</td>
</tr>
<tr>
<td>Soybean Sales</td>
<td>.920</td>
<td>.230</td>
<td>.635</td>
<td>.102</td>
</tr>
<tr>
<td>Wheat Price</td>
<td>.231</td>
<td>.849</td>
<td>.966</td>
<td>.137</td>
</tr>
<tr>
<td>Corn Price</td>
<td>.111</td>
<td>.409</td>
<td>.067</td>
<td>5.6E-03</td>
</tr>
<tr>
<td>Soybean Price</td>
<td>.428</td>
<td>.799</td>
<td>.812</td>
<td>--</td>
</tr>
</tbody>
</table>

* - Indicates a Granger causal relationship, i.e. a rejection of the null hypothesis that all lags of the exchange rate are jointly zero.
Comparing the results among the in-sample tests for Granger causality we find that the three multivariate methods do not agree unanimously on Granger causality for any of the six variables. This parallels the conclusions reached by the out-of-sample tests insofar as the method of specification plays a vital role in the outcome of the Granger causality tests. However, for the wheat sales, corn sales, and corn price variables, Granger causality is indicated by at least one of the multivariate methods.

Some additional conclusions regarding testing for Granger causality are revealed by comparing in-sample and out-of-sample test procedures. In particular, the out-of-sample test finds evidence of Granger causality in all but the case of the soybean price at levels of significance that are equal to or less than those from the in-sample tests. Thus, the in-sample tests (taken together) agree with the out-of-sample tests in the case of the wheat sales, corn sales, corn price, and soybean price variables, but disagree in the cases of the soybean sales and wheat price. Comparing each in-sample test individually to the out-of-sample conclusions we find much less agreement. The BIC/HQ method agrees for wheat sales, corn price, and soybean price. The FPE method agrees only in the case of corn sales, and the RVAR underfitting agrees in only two cases, the corn sales and the corn price.

Deciding which model selection procedure is best for the in-sample tests is more difficult than that for the out-of-sample test where the best forecasting model is the obvious choice. This difficulty, as well as the fact that the out-of-
sample test is more faithful to the definition of Granger causality, suggests the merit of testing for Granger causality out-of-sample.

**Long-Run vs. Short-Run Forecasts and Granger Causality**

In this study 1-step ahead forecast errors were used to estimate mean square errors and the significance of increases in forecasting accuracy. As a practical matter, however, market participants and policy analysts are often interested in forecasts over a longer horizon. For the multivariate models in this study, the use of forecasts longer than 1-step ahead means that forecasts of the exchange rate come into play in determining forecasts of the agricultural prices and export sales in the bivariate models.

Table 25 contains the postsample mean square errors of 1-22 step ahead forecasts of each of the six agricultural variables. Among the univariate models, the FPE selected models are lowest MSE for the wheat sales and corn sales variables, the ARIMA is the best model for the soybean sales variable, and the BIC/HQ selected models are best for each of the price variables. These results are consistent with those from the 1-step ahead forecasts which indicated that no one univariate method dominated the others. Each of the best univariate models is lower MSE than the sample mean benchmark.

For the multivariate models, the RVAR is uniformly superior for the three sales variables, however, the dominance over the other multivariate methods observed in the 1-step ahead comparison in table 19 does not extend to the price variables in the long range comparison. For the wheat price variable, the
Table 25. Postsample mean square error of 1-22 step ahead forecasts, all values E-03.

<table>
<thead>
<tr>
<th>Variable</th>
<th>MEAN</th>
<th>BIC/HQ UAR</th>
<th>ARIMA</th>
<th>FPE UAR</th>
<th>Method</th>
<th>BIC/HQ VAR</th>
<th>FPE VAR</th>
<th>HSAIO RVAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wheat Sales</td>
<td>148.92</td>
<td>125.84</td>
<td>129.59</td>
<td>114.48</td>
<td></td>
<td>124.99</td>
<td>140.56</td>
<td>93.987</td>
</tr>
<tr>
<td>Corn Sales</td>
<td>409.79</td>
<td>285.40</td>
<td>409.31</td>
<td>277.25</td>
<td></td>
<td>304.13</td>
<td>265.60</td>
<td>254.12</td>
</tr>
<tr>
<td>Soy Sales</td>
<td>245.83</td>
<td>244.91</td>
<td>230.77</td>
<td>238.09</td>
<td></td>
<td>246.26</td>
<td>265.90</td>
<td>237.09</td>
</tr>
<tr>
<td>Corn Price</td>
<td>315.86</td>
<td>54.919</td>
<td>66.100</td>
<td>70.990</td>
<td></td>
<td>47.981</td>
<td>39.309</td>
<td>55.438</td>
</tr>
<tr>
<td>Soy Price</td>
<td>290.33</td>
<td>4.9802</td>
<td>11.044</td>
<td>11.044</td>
<td></td>
<td>3.4687</td>
<td>1.9564</td>
<td>4.8525</td>
</tr>
</tbody>
</table>
BIC/HQ multivariate model is best, whereas the FPE selection method produces the best multivariate model for the corn price and the soybean price variables.  

Comparing the lowest MSE univariate models to the lowest MSE multivariate models, in every case the multivariate model has lower MSE than the univariate model. Again there may be some question as to the significance of these relationships. However, given that these results do not conflict with those of the 1-step ahead forecasts (except that in the long-range comparison the exchange rate seems to help forecast soybean price), the preponderence of the evidence points to the exchange rate having an effect on the agricultural price and export sales variables.

---

There is some question as to the significance of these relationships especially for the corn sales, soybean sales, wheat price, and corn price variables. Nevertheless, the result that the RVAR is uniformly superior to all UVAR specifications does not hold for the longer range forecasts of these agricultural variables.
Chapter 5 - Summary and Conclusions

*If you torture the data long enough, nature will confess.*

*R. Coase, quoted in Leamer, *American Economic Review*.*

5.1 Summary

The first specific objective of this study was to provide a review of time series forecasting techniques and some of the recent, associated, technical issues. Chapter 2 introduced some basic concepts in time series analysis as well as an introduction to the specification of ARIMA models. Chapter 2 also stressed the importance of working with stationary series, and tests for unit roots were investigated as a means for assessing stationarity. Chapter 3 introduced concepts in multivariate time series analysis, the theory of vector autoregressive models, and practical techniques for specifying restricted and unrestricted VARs. The prob-
lem of how to address multivariate nonstationarity was also investigated, and tests for co-integrated time series were described.

The second objective was to assess the time series properties of each of the data series. The prices and the exchange rate were found to be $I(1)$ processes, whereas the export sales series were found to be stationary in their levels. No co-integration was found between the agricultural prices and the exchange rate.

The third objective was to specify univariate and multivariate time series forecasting models for agricultural prices and export sales. Chapter 4 contains the orders of the estimated models, comparisons of forecasting accuracy among all models, and an evaluation of the various competing univariate and bivariate models for price and sales forecasts. A procedure due to Ashley et al. (1980) was used to test the statistical significance of improvements in forecasting accuracy in going from univariate to bivariate models. As the bivariate models included the exchange rate as an explanatory variable, this was a means to testing for Granger causality.

The fourth objective was to compare the usual in-sample Granger causality tests to the AGS Granger causality tests and to investigate the impact of lag-selection methods on the results from the both methods. It was found that using different lag selection methods yielded conflicting conclusions about the usefulness of the exchange rate in forecasting agricultural price and sales variables. This underscores the importance of searching among alternative models for the best predictor. The out-of-sample evaluation procedure used herein was com-
pared to various in-sample methods; these two methods also yielded conflicting conclusions.

For the fifth objective, evidence from a long range forecasting experiment using the same models was compared to the results from the short range forecasts. In general the long-range methods yielded similar conclusions i.e., the exchange rate is an important factor to consider when modeling agricultural prices and export sales.

5.2 Conclusions

The conclusions offered in chapter 4 highlight the impact of technical time series issues on the detection of Granger causality from the exchange rate to agricultural prices and export sales. Our results point to three important technical issues that should be considered when conducting a test for Granger causality.

The first issue concerns the impact of nonstationarity on testing for Granger causality. Chapters 2 and 3 presented a review of the theoretical results on stationarity and its impact on forecasting models. The theory reviewed indicated that working with stationary data series should enhance predictive ability out-of-sample. Earlier empirical research by the author (Bradshaw and Orden (1988)) revealed that ignoring unit root nonstationarity in the data may lead to sub-optimal forecasting models and consequently to misleading conclusions on the significance of exchange rate impacts on agricultural prices.
A second issue brought out by the present investigation concerns the effects of lag-length selection on forecasting performance. Our results parallel those from earlier studies (e.g., Thornton and Batten 1985) which have found that the outcomes of in-sample test for Granger causality can be influenced by the lag selection technique employed for specifying an unrestricted VAR. In this study, the evidence suggests that both in-sample and out-of-sample tests for Granger causality are affected by the choice of lag selection criteria.

A third issue concerns the debate over using in-sample versus out-of-sample tests for Granger causality. Some evidence has suggested that in-sample tests for Granger causality are not realistic tests of the hypothesis. Again, we find that tests of Granger causality are sensitive to whether the test conducted is an in-sample or an out-of-sample test. These two methods only agreed in one of six cases in which the hypothesis of Granger causality was tested.

In sum, the influence of these three technical issues on testing for Granger causality underscores the importance of careful specification of time series models with respect to stationarity, lag-length selection, and the type of test (in-sample or out-of-sample) used. The definition of Granger causality calls for optimal forecasting models, and our research indicates the difficulty involved in finding such models. Care must be taken to properly prepare the data with regard to its stationarity, and a thorough search of the model space must be made in order to have a reasonable level of confidence in the results. Otherwise, arguments such as the ones above could be offered as reasons for doubting the quality of the results. In addition, that in-sample tests of Granger causality offer different con-
clusions than out-of-sample tests indicates that in-sample tests, which are farther in spirit from the out-of-sample test implied by Granger’s definition of causality, may not be justified.

In this study the above cautions were exercised and consequently we have a reasonable degree of confidence in the results. Comparing our best univariate and best bivariate models, tests for the significance of the Granger causal relationships indicate causality, at reasonable levels of significance, from the exchange rate to agricultural export sales in each of the three cases studied. This is not too surprising if we believe that purchasing power parity holds only as a long-run condition. In such a case, movements in the exchange rate would have real effects, and these effects should be observed on a variable such as export sales.

The importance of model specification is highlighted again by comparing these results to a previous study by Bessler and Babula (1987). They used the same wheat export sales and wheat price variables as the present study, and found that the exchange rate did not help to forecast wheat sales, but did help to forecast wheat price, a rather anomalous result. If movements in the real exchange rate affect real commodity prices, we should expect to see concomitant influence on the level of export sales of the commodity price.

In this study, the evidence on Granger causality from the exchange rate to agricultural prices is less consistent then for export sales. In one case, corn price, the exchange rate has a very significant impact on the accuracy of forecasts. For the wheat price, the evidence is marginal (13.7-percent level) as to the impact of
the exchange rate, and in the case of the soybean price, however, the exchange rate does not improve the accuracy of the forecasts. This wide spectrum of results makes a consistent interpretation difficult. However, the evidence from the long range forecasts seems to support the view that the exchange rate does help forecast the agricultural prices. Thus the preponderence of the evidence seems to be in favor of the result that the exchange rate does influence agricultural prices.

5.3 Suggestions for Further Research

A next step to building on the results in this study would be to develop models that include sales, prices, and exchange rates to detect interactions among all three variables. In addition, there is some evidence of the impact of monetary policy on the agricultural sector with the exchange rate as the transmission mechanism (Chambers and Just 1982). If this is the case, including a money variable in addition to or instead of the exchange rate may provide a stronger link from the macroeconomy to the agricultural sector. This seems even more plausible in light of the feedback results revealed in applying Hsiao’s procedure.

Finally, the models here are simply reduced form forecasting models, that offer little about the structure of macroeconomic-agriculture linkages other than being able to describe such things as Granger causality. The next step is a more articulate identification of macroeconomic impacts on agriculture using dynamic time series models.
References


Bessler, D.A. (1986), Forecasting Vector Autoregressions with Bayesian Priors, American Journal of Agricultural Economics, 3, 144-151.


Litterman, R.B. (1984), Specifying Vector Autoregressions for Macroeconomic Forecasting, Staff Report 92, Federal Reserve Bank of Minneapolis, Research Dept.


Lucas, R.E. and T.J. Sargent (1979), Beyond Keynesian Macroeconomics, in *Rational Expectations and Econometric Practice*, Minneapolis: University of Minnesota Press.


VITA


GIRARD ENTERED THE MASTER'S PROGRAM IN AGRICULTURAL ECONOMICS AT VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY IN 1986 AND RECEIVED AN MS DEGREE IN JUNE OF 1988. HE PLANS TO CONTINUE HIS EDUCATION AND WILL ENTER A PHD PROGRAM IN ECONOMICS AT THE UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL IN SEPTEMBER 1988.