Multiple Time-Series Models

Patrick Brandt
University of Texas at Dallas

MULTIPLE TIME SERIES MODELS

Patrick T. Brandt
University of Texas at Dallas

John T. Williams
University of California, Riverside

1. INTRODUCTION TO MULTIPLE TIME SERIES MODELS

Many social science data problems are multivariate and dynamic in nature. For example, how is public sentiment about the president's job performance related to the aggregate economic performance of the country? Are arms expenditures by a series of countries related to each other or exogenous? Are the actions directed by country A toward country B related to the actions that country B directs toward country A? How is the percentage of Americans that identify with the major political parties—aggregate partisanship—related to their support for government policies? How are tax rates related to the proportion of political action committees that are organized by business? In each of these examples, it is possible to write down a single equation where one of the variables is the dependent variable and the other is the independent variable. But it is likely that in each of these examples there is simultaneity and that there exists a second equation with the roles of the independent and dependent variables reversed.

For the sample research questions noted above, both the variables are likely to be endogenous. One would expect that the same factors that explain changes in aggregate partisanship are endogenously and dynamically related to public support of government policies. Similarly, changes in tax rates for business are both a cause and a consequence of the lobbying efforts of corporate political action committees. Both these examples (addressed in more detail in Chapter 3) are ones where a researcher might posit two (or more) equations, one for each variable, and allow both the current and the past values of each variable in the model to affect each other.

Most social scientists learn to use regression relatively early in their statistical training. But single-equation regression models ignore the fact that for endogenous, dynamic relationships there is either explicitly or implicitly more than one regression equation. An analyst may choose to continue estimating a single regression and hope that statistical inferences are not too flawed, or he or she might decide to estimate a multiple-equation model using a variety of techniques developed in econometrics (e.g., seemingly unrelated regressions, autoregressive distributed lag [ADL] models, transfer function models). But even this accounts only for specification- and estimation-related issues such as serial correlation and endogeneity. Analysts must also confront the additional complication that the data are measured through time and have time periods as the units of analysis. A researcher looking at these questions could then choose to estimate a single- or multiple-equation model with some time series dynamics. This leads us to consider the relative merits of other avenues of analysis, including vector autoregression, error correction models, and (dynamic) factor analysis. Because time series data are much richer—that is, they contain more information than cross-sectional data—decisions about how to tackle multiple time series problems are crucial.

The need for these dynamic multiple-equation models stems from two very common realities in social science models. First, variables simultaneously influence one another, so both are referred to as endogenous variables. A multiple-equation system usually, but not always, has the same number of endogenous or dependent variables as equations.[1] Although the theoretical interest of an analyst may be on just a single equation, and this equation may be the only one estimated, statistics and econometric theory require that all equations be considered, otherwise inferences can be biased and inefficient. Second, when considering the relationships among multiple dependent variables, the unique or identified relationships for each equation of interest can be made only with reference to the system as a whole. Properly determining these relationships requires that information from all equations be used. Identification requires that there be enough exogenous variables, specified in the correct way, to be able to estimate any or all of the equations in a system. Estimation requires that exogenous variables from the entire system be used to provide the most unbiased and efficient estimates of the relationships among the variables as possible.[2]

In addressing each of these issues we expect that the dynamic relationships of the variables are of central interest. We would like to know how changes in one of the variables affects the others. It is possible that the relationship among the variables is endogenous in one of two senses. First, changes in one of the variables may have a delayed effect on another (so the effect is through the past values of one variable on the current or contemporaneous values of another). Alternatively, the relationship may be contemporaneous in that changes to the system of equations, known as shocks or innovations, may change both or several variables at the same time. This arises because the shocks to one variable are correlated with the shocks to another variable.

A central concern in translating a theory into an empirically estimable model (i.e., one where we can estimate the parameters and make inferences about them) is that we may not know the *structure* or equation(s) that correctly represents a model. That is, suppose that some (multivariate) probability density $f(y|\beta)$ describes the observed data y in terms of some parameters β. This density would identify a unique structure or set of equations if there is no other set of values β that produce the same density.[3] As social scientists, we do not know for sure if the equations we write down for the specification of a model are the correct ones in many circumstances. A consequence of this is that many disputes over the interpretation of models and their parameters are really based not on the properties of the models per se but rather on the disagreements about the structures or equations used to represent those models.[4] Our goal in this chapter is to illuminate some of the choices faced by social scientists in building theory that accompanies modeling of multiple time series. We will first highlight some of the major choices that analysts must make and then describe the implications of these choices.

Social science theories are built in several stages. First, the researcher must identify the main variables and relationships to be explained with a theory. Here, researchers look at the main theory (or theories) that informs their empirical questions and specifies the relevant variables and relationships to be modeled. Even when there are competing theories present, this stage presents few problems, assuming that relevant time series can be measured for the variables of interest.

Once the main aspects and variables of a theory are determined, the researcher begins the critical phase of selecting the functional form or mathematical structure of the model. It is at this stage that many different models of the same underlying theory or theories begin to emerge. At this stage, we need to make decisions about how theories are translated into equations. This stage will also require that the equations be identified or that there be sufficient restrictions on the equations we specify to ensure that a unique set of parameters can be estimated and interpreted. To do this we need both information about the data, the equations, and the a priori beliefs of the researcher to determine whether or not a model is formally identified.[5]

The third stage of theorizing and model building is fitting the specified model to data and interpreting the results. We note that this step is in one sense noncontroversial because there is wide consensus about how models should be fit and what criteria should be used to evaluate them (unbiasedness, efficiency, minimum mean-squared error, consistency, etc.) A more relevant issue, though, is the determination of the dynamic properties of the specified model from its estimates. Because we are focusing on models of multiple time series, we need to be concerned with methods that can be used to do this.

Finally, once the model has been fit and interpreted, there is often a need to revisit the earlier steps to evaluate the impact of specification decisions and look again at confirmed or disconfirmed aspects of the theory.

The most critical aspects of this model-building process are the specification of the functional form of the system (Stage 2). The remainder of the process, to a large degree, depends on the decisions made in this stage. Models that are specified with missing variables or incorrect dynamics will suffer from the same problems as ordinary least squares (OLS) models—bias and inefficiency. In addition, failing to include a relationship or a factor that is part of the multivariate system can lead to simultaneity biases. Note that these parameter ills then induce problems with interpretation and hypothesis testing.

Standard simultaneous equation models and univariate time series models are commonly proposed to address these types of questions. Although these kinds of models have much to offer, they also have limitations. We next discuss the general decisions that researchers face and highlight some of the trade-offs in the choice of different models. We present the four main approaches that are typically used to model univariate and multiple time series data: autoregressive integrated moving average (ARIMA) models, simultaneous or structural equation (SEQ) systems, error correction models (ECMs), and vector autoregression (VAR). In the remainder of the text we use simultaneous and structural equation models interchangeably. We discuss how each of the main approaches to modeling dynamic simultaneous relationships forces the researcher to make certain choices about the relationships that may or may not be clearly specified in a theory and the empirical representation and statistical model they specify.

Critical to our presentation is the following: The differences among these methods have less to do with *technique* and more to do with *approach*. All these methods employ some version of linear regression (OLS, generalized least squares, multistage least squares, etc.) or maximum-likelihood methods for estimation. What differ among these methods are the assumptions and building blocks that are the basis for inferences and interpretation.

1.1 Simultaneous Equation Approach

A first approach for building a multiple-equation time series model would be to work in the simultaneous equation (SEQ) paradigm. SEQ models are present in the multiple disciplines of the social sciences. The SEQ paradigm was largely developed by the Cowles Commission in the 1940s and 1950s at Yale. The Commission's early goal was to develop a methodological paradigm for modeling the economy using econometrics. So the researchers there worked to adapt existing econometric methods to the study of large-scale, multiequation models of the economy. In this case, the early Cowles model was an empirical representation of standard Keynesian macroeconomic theory.

Model building with SEQs is based on taking the representation of a single theory or approach and rendering it into a set of equations. Using a single theory to specify the relationships among several variables leads to choices about which variables are exogenous to the system and which are endogenous. The exogenous variables are those that are determined outside the system or are considered fixed (at a point in time or in the past). Those that are determined inside the system and are the dependent variables of the equations are endogenous. The result is a *single* structural system of equations that express the relationships among the variables. The reason there is such a focus on a single theory is that multiple theories may lead to different, typically non-nested specifications of the structural equations (good examples include Zellner, 1971; Zellner & Palm, 2004).

Consider the earlier research question about how aggregate partisanship is related to public support of government policies. A simultaneous equation representation of the relationships among these variables would have two equations, one for each variable. Each endogenous variable would be a function of the other and (possibly) past values of each variable. To estimate such a system, one would need to rewrite the system as a reduced form set of equations where each endogenous variable is a function of predetermined or exogenous values. Unspecified in this modeling approach is how decisions about the number of past values influence the system of equations or how the system of equations will be identified. Typically (vague) appeals are made to "theory" and hypothesis test results for the inclusion or exclusion of variables.

Several issues arise in constructing SEQ models in this manner. First, alternative theories must be nested within a common structure to be compared. If the models cannot be nested (because of nonlinearity or different specifications), then no single structure can be used to compare different models. Second, the models require that choices be made about the inclusion or exclusion of different variables and lagged values to ensure identification. Two methods are common here: restricting "predetermined" or lagged endogenous variables as exogenous variables and the classification of variables as either endogenous or exogenous. Here, "theory" is used to restrict the parameter space of the model parameters. Often hypothesis tests are used to determine the exclusion of variables, but this then induces pretest biases in the final models, because the exclusion of variables based on tests typically leads to overconfident results.

As argued by Sims (1980), such exclusion restrictions are often not theoretically justified and are often not well supported by empirical analysis. One consequence of this is that extra lagged values are included or excluded from an SEQ model leading to incorrect dynamic specifications. Even if the models have white noise or serially uncorrelated residuals, the specification search for these models may be incorrect and may imply the wrong dynamic specification, because it has incorrectly restricted the parameter space.

Finally, these models often perform poorly at forecasting and policy analysis. Alternative, simpler models typically will outperform complex, multiple-equation simultaneous equation models.

1.2 ARIMA Approach

Another approach to multiple time series models starts from a time series perspective. One could address multiple time series as a collection of univariate series. In this vein, the researcher would use the standard ‘‘Box-Jenkins approach’’ or ARIMA models for each of the series (Box & Jenkins, 1970). Once the dynamics are known, one could begin to build a model where some of the variables are included as pulse, intervention, or other exogenous effects on the right-hand side of an ARIMA model.

The Box-Jenkins approach is oriented toward forecasting and describing the behavior of a time series (Granger & Newbold, 1986). The general Box-Jenkins approach defines a class of models—in this case ARIMA models—to describe a time series. One then fits a series of ARIMA models to each of the series with the goal being to choose the most parsimonious model with uncorrelated residuals. This approach requires that we designate some of the variables as endogenous and others as exogenous for the fitting of the model. The Box-Jenkins approach is particularly successful at forecasting—in fact, Box-Jenkins-style models will typically outperform SEQ models in forecast performance. The main reason for this is parsimony: The models are built by exploiting the parsimony principle and allowing the data to speak as much as possible.

A Box-Jenkins model for the aggregate partisanship percentage and the percentage of the public supporting the government policies over time would be constructed as follows. Suppose we are most interested in predicting public support. One would first construct a univariate ARIMA model of the public support dynamics. Next, once the model is determined, one would add the aggregate partisanship covariate to see if it improved the fit of the public support model. One would examine various specifications of the partisanship variable (including contemporaneous values and various lags). Hypothesis tests would be used to determine the best specification.

One might reverse the roles of the variables as well and fit a model to the partisanship measure.

There are limitations to this approach when building a multivariate time series model. First, it ignores the fact that some of the variables in the model may help to proxy dynamics in the others. If this is the case, then the suggested procedure may lead to severe overfitting of the data, because the standard Box-Jenkins approach is to filter or explain most of a variable's variance using its own past values. Second, this approach leads us to focus on the dynamics of the variables first rather than on the general relationships in the system. Third, because this is generally done in separate equations—one for each of the variables—we expect that unless the equations are perfectly independent, there will be inefficiency in the estimates. Finally, unless the variables are causally related in a specific way, treating the many variables in separate ARIMA models will lead to inefficient estimates. The reason is that if the residuals for the several variables are contemporaneously correlated (i.e., at the same time), then the estimates will be inefficient. Only when the results of each equation are explicit of the others can we use a sequence of independent equations to model the results.

1.3 Error Correction or LSE Approach

Error correction models are a specialized case of ARIMA regression and simultaneous equation models. They are commonly referred to as the London School of Economics or "LSE" approach because they have been advocated by economists there (Pagan, 1987). The basic building block of an ECM is an autoregressive distributed lag (ADL) specification for two or more variables with provisions for the (possible) long-run relationships among the variables.

The ECM approach differs from ARIMA models in that the long-run relationships—typically, stochastic and deterministic trends—are directly modeled. In ARIMA models, these long-run components, trends or unit roots, are "differenced" from the data to create a stationary data series that can be modeled as an ARIMA process. The ECM approach instead explains the long-run components in two or more data series as a function of each other. The ECM approach uses the long-run component in two or more of the series being modeled to derive a common (stochastic trend) representation that is shared among the series. An ECM uses this common representation to produce a model that has a common long-run component for the variables and a short-run component known as an error correction mechanism that describes how each variable varies or equilibrates around the common long-run component.

The ECMs can be applied to both stationary (mean reverting) and nonstationary (unit root) data. For stationary data, the ECM allows one to estimate a common or equilibrium level for the variables and how each varies around the equilibrium. This model is equivalent to an ADL model, which is an ARIMA model with exogenous variables. For nonstationary or trending data, the ECM modeler starts with a specialized set of data series—two or more series that have unit roots or are integrated to order one.[6] This determination is made using unit root tests (such as the augmented Dickey-Fuller [ADF] test). Once the series are found to be unit roots, a specialized estimation technique is used to estimate both the long- and short-run relationships in the data. For bivariate relationships, a one- or two-step ECM estimation procedure can be used. For multivariate time series (typically with unit roots), the vector ECM (VECM), described in detail in Johansen (1995), is used. The first step in this process is the determination of the common stochastic trend processes in the data. Then, once these long-run trends have been estimated, the short-run dynamics around the long-run trend are estimated using a regression model.

Because the ECM and its multivariate version, the VECM, are based on describing the long- and short-run components of a multivariate time series regression model, researchers can test for a variety of relationships among the common long- and short-run dynamics and how they are related across the various series. For nonstationary data, the representation of the ECM ensures that there is a particular form of "causal" relationship among the series (Engle & Granger, 1987). This causal relationship is known as a "Granger causal" relationship, where the past values of one series must (linearly) predict the current values of the other series. This means that the trend in the two integrated series is "driven" or predicted by the changes in one of the variables. As such, these models are a specialized case of the simultaneous equation models, because they impose and estimate a common trend structure across the series.

Consider again the model of the relationship between aggregate partisanship and the percentage of the public that supports government policies over time. Some argue that these variables are nonstationary or unit roots because they are the sum of a series of accumulated events or shocks that persist (see the discussion and references in Chapter 3 for more details). If this is the case, then an ECM may be appropriate. The ECM can then be used to evaluate how the short- and long-run relationships of these two variables are related and which variable Granger causes the other. The ECM will allow one to apply hypothesis tests to determine the long- and short-run structures of the relationship among the series. The ECM representation of the relationship will provide more information about the dynamics, but we will have to estimate the cointegration relationship and the short-run dynamics.

Inference and model formulation using ECMs, although well developed (e.g., Banerjee, Dolado, Galbraith, & Hendry, 1993), can be quite complicated for nonstationary or unit root data. Many economic variables (e.g., consumption, gross domestic product, government spending) will have unit roots. This is an important reason for considering these models, because they allow one to look at the short- and long-run dynamics of these variables correctly. However, frequentist inference using nonstationary data in these models is complex. This is because of the nonstandard distributions and the complications of computing dynamic analyses when there are unit root variables in the model. This means that hypothesis tests for the presence of error correction relationships and the number of error correction relationships and tests concerning the model parameters often have nonstandard distributions that must be simulated or analyzed using nonstandard test statistic tables (Cromwell, et al., 1994, Lutkepohl, 2004). Further, the causal structure of ECMs may be no easier to determine. Sims, Stock, and Watson (1990) note that such inference in models with multiple unit roots is difficult.

1.4 Vector Autoregression Approach

A final approach to modeling multivariate time series is the VAR model. VAR modelers do not assume to know the correct structure of the underlying relationship that generated the multiple time series. Instead, they focus on the underlying correlation and dynamic structure of the time series.

The VAR approach starts by focusing on the interrelated dynamics of the series. It asks the following questions (in contrast to the SEQ approach):

1. What is to say that some lagged variables would not be in each equation? Does restricting the dynamics for identification make sense?
2. What impacts do each of the variables have on each other over time?
3. If a variable affects one equation in the system of equations, what is to say that it does (or does not) affect another?
4. Are rational expectations—the idea that a variable is best predicted by its immediate past value plus a random component—present? In this case, the past is of little predictive value, and policy makers and analysts are interested in how the random components—innovations or policy shocks—are translated into outcomes. In this framework, the shocks are themselves exogenous variables.

These are all critiques of the standard (i.e., Cowles Commission) approach to simultaneous equation models. The main difference in the VAR approach is that it is built on creating a complete dynamic specification of the series in a system of equations. The basic idea of VAR modeling is built on the insights of the Wold decomposition theorem (Hamilton, 1994, pp. 108–109; Wold, 1954). Wold showed that every dynamic time series could be partitioned into a set of deterministic and stochastic components.

All these critiques point toward understanding dynamics. In response to these critiques, Sims (1972, 1980) pioneered the VAR methodology, building on the idea of dynamic decomposition of the variable in the system. Sims rejected the use of standard simultaneous equation models for three reasons:

1. Identification restrictions on parameters used in SEQ models are typically not based on theory and thus may lead to incorrect conclusions about the structure of the models and the estimates.
2. SEQ models are often based on tenuous assumptions about the exogeneity and endogeneity of the variables. Because the true lag lengths of the variables are not known a priori, identification is then based on possibly specious assumptions about exogeneity. The formal identification of a dynamic simultaneous equation model requires that the exact true lag length be known for each variable; otherwise, identification assumptions may not hold (Hatanaka, 1975).
3. If the variables in the model are themselves policy projections, additional identification problems will be present because of temporal restrictions. This is the rational expectations critique: Models are typically treated as though ceteris paribus claims will be true. In fact, they are not, then we need to be able to assess the probabilistic implications of different paths for the variables.

The Sims-proposed method for addressing the tenuous identification problems of the SEQ approach is to focus on the dynamic specification of the reduced form model. This is in contrast to the SEQ approach, which focuses on the identification choices in the model specification. Sims' approach is to ensure that the modeling approach to multiple time series provides a complete characterization of the dynamics of several series. This is done using a multivariate autoregressive model to account for the dynamics of all the variables.[7]

The VAR model proposed by Sims is a multivariate autoregressive model where each variable is regressed on its past values and the past values of the other variables in the system. Model building in VAR models then depends on the selection of the appropriate variables (based on theory). The specification of the dynamic structure proceeds based on testing for the appropriate lag length using the sample data. Sims (1980) argues that one of the critical contributions of the VAR approach is that it can serve to define the "battleground" of empirical debates about multiple time series data. It does this by providing a model of the dynamic and empirical regularities of a set of related time series. From this point, one is able to refine and develop the empirical model to evaluate theoretical debates.

The logic of the VAR approach can be applied to the aggregate partisanship and public support of government policies example. Instead of starting with a set of structural (e.g., SEQ), dynamic (e.g., ARIMA or ECM), or a priori causal relationships (e.g., ECM) among the variables, a VAR model begins by assuming that the reduced form dynamics are of central interest. Thus, rather than impose possibly structural or dynamic restrictions on the relationships among the two series, a VAR would have two equations, one for each variable. Each variable would be regressed on its past values and the past values of the other variable. The resulting residuals (after checking for serial correlation) would be exogenous shocks or innovations. One could look at the responses of each equation to see how these "surprises" in each variable affect the observed system. After accounting for these (historical) dynamics, one could engage in inferences about the Granger causal relationships between the two variables and try to determine the endogenous structure and dynamics of the two series.

Building multivariate time series models according to the VAR methodology does not then depend on a single theory. Instead, multiple theories can be compared explicitly and evaluated (using hypothesis testing) without the identification assumptions that would be made in the specification of alternative simultaneous equation models. Because the variables in the VAR model are not segmented a priori into endogenous and exogenous variables, we are less likely to violate the model specification and incorrectly induce simultaneity biases by incorrectly specifying a variable as exogenous when it is really endogenous.

The key distinction between VAR and SEQ models is the treatment of identification assumptions. In the SEQ model, these are taken as fixed, invariant, and specified by theory. In the VAR approach, such zero-order restrictions (e.g., excluding variables from some equations or omitting some lagged values of some variables from some equations) are seen as unlikely to be true. Thus, in an effort to eliminate biases from these incorrect restrictions, VAR models are able to consider trading off these biases for some inefficiency. The biases in the SEQ model estimates are the result of omitting lagged values that should be included in the models. Under Sims' logic, often some lags of some variables are (incorrectly) excluded to identify the SEQ model. These incorrect restrictions, which lead to the omission of relevant lagged variables, produce omitted variable bias. The solution to mitigating this bias is to include all possible lags (which may be more than necessary). In this case, the goal is to reduce bias at the cost of efficiency.

The key identification assumption in a VAR model is how the contemporaneous effects of each variable are related to each other. Because the VAR model is specified in terms of the lagged values of the variables in the system on each other, identification concerns the specification of the residual or contemporaneous covariance matrix of the residuals alone. The benefit of this is that it allows one to separate the interpretation of the model dynamics from the identification. This allows researchers to explicitly look at how identification decisions are related to the path of the variables' dynamics.

VAR modelers also have a different conception of the interplay of data and models. The goal of a VAR model is to provide a probability model of the dynamics and correlations among the data (Sims, 1980). Thus, VAR models are considered best when based on a simple, unbiased specification that accounts for the uncertainty about the dynamics and the model. To do this pretest biases must be avoided (Pagan, 1987). Thus, unlike the "specify-estimate-test-respecify" logic of classical approaches, SEQ models, and ARIMA models, VAR models employ few hypothesis tests to justify their specification. This leads to a less biased representation of the model and its dynamics rather than the false sense of precision that can accompany other modeling strategies. That is, once we have entered this cycle of specification testing, the resulting inferences are a function of the test procedure and are less certain than the reported test statistics and associated levels of significance or reported P values would lead us to believe.

1.5 Comparison and Summary

This brief review of possible approaches to multiple time series models has been intended to connect other approaches to the VAR methodology (see, e.g., Pagan, 1987; Sims, 1996). The ARIMA, ECM, and SEQ methods are special cases of the more general VAR methodology. Freeman, Williams, and Lin (1989) presented a basic comparison of the VAR and SEQ approaches. Here, we have discussed these and other models with the goal of comparing how dynamics are modeled and how inferences are made.

Table 1.1 presents a summary of each of these methods, which extends the summary initially presented in Freeman Williams, and Lin (1989). The table shows the main methodological differences in the specification of time series models.

The critical point is that VAR models are a generalization of the other approaches. Each of the other modeling approaches focuses on some feature of time series data that *may be* true in practice. However, from the standpoint of model formation and theory testing, the more general approach is a VAR model.

TABLE 1.1
Comparison of Time Series Modeling Approaches

	ARIMA	*Error Correction*	*SEQ*	*VAR*
Model Building				
Specification	Single theory focused on univariate series	Long- and short-run trends and dynamics based on results of tests for cointegration and unit roots	Single theory with assumptions about endo-geneity and exogeneity	Recognition of multiple theories by including variables as endogenous
Estimation	Maximum likelihood; OLS	Johansen procedure; one- or two-step procedure	Higher-order OLS and maximum-likelihood methods; corrections for heteroscedasticity and serial correlation; tests for overidentification and orthogonality	OLS and tests for lag length
Methodological Conventions				
Hypothesis testing	Analysis of individual coefficients	Tests of cointegrating relationships; short-term dynamics	Analysis of individual coefficients; goodness of fit	Analysis of significance of blocks of coefficients; tests for exogeneity
Dynamic analysis	Dynamic multipliers; intervention analysis	Analysis of cointegrating vector; impulse responses	Simulation; deduction of model dynamics	Forecasts, model projections, decomposition of forecast error variation, impulse responses

Why then do we advocate the VAR approach? First, we are not trying to rule out structural models. In fact, such structural models can and will do a good job in some cases when the restrictions in those models are true. They will provide better inferences, summaries of dynamics, representations of relationships, and other measures. Second, using a step-by-step approach to structural model building *that is explicit in its treatment of assumptions about dynamics* will be to produce a better model based on the parsimony principle.

When are these structural equation models a poor choice, and when should they be replaced by VAR models? There are three situations, all of which are when these models fail. First, unless we know or test the precise structure of the relationships among the variables in our models, SEQ models will be misspecified. Second, for policy or counterfactual analysis, unless the models are correct, we may make incorrect inferences. Finally, if one of our goals is to characterize uncertainty and dynamics, then VAR models will typically be superior because they are less likely to be overly precise via ad hoc pretesting of the models under consideration.

Second, there are times when one would prefer to use an ECM. These are when one wants to isolate the long- and short-run behaviors of several series simultaneously, or when trending or unit root variables are present in a multiple time series models. In these cases, one is actually estimating a VAR model with a set of restrictions or assumptions (which can be tested) about how the long-run behavior of the multiple time series model evolves. Our contention in this book is that to understand and apply the ECM and VECM models, one should have a solid basis in the more general, unrestricted VAR approach. We return to the relationship between VAR and ECM models in the next chapter.

The next chapter outlines the mathematical details of a VAR model. We then discuss how this model is used for inference about the relationships in multivariate time series data.

2. BASIC VECTOR AUTOREGRESSION MODELS

Vector autoregression (VAR) models are not a statistical technique or methodology. Rather, VAR models are an approach to modeling dynamics among a set of (endogenous) variables. This approach focuses on the dynamics of multiple time series and typically employs multiple regression and multivariate seeming unrelated regression models. The central focus is on the data and their dynamics. A central tenet of VAR models is the idea that restrictions on the data and parameters in the model should be viewed skeptically.

How skeptically? Consider some hypothetical time series data that have a rich dynamic and correlated structure. Consider a "view" of these data: With perfect knowledge you can see these data and know their rich dynamics. Now consider closing one eye. With one eye closed you only see part of the dynamics of the data (the left or the right depending on which one you closed) and may lack the full perception of the depth of these data. VAR modeling is an effort to force you to keep your eyes open rather than incorrectly closing your eyes or occluding your vision with incorrect assumptions.

What then is a VAR model? Simply, it is an interdependent reduced form dynamic model. For each endogenous variable in the system of equations,