21 - 1

~~20~~ 21

MODELS WITH LAGGED VARIABLES

~~20~~21.1 INTRODUCTION

FN

This chapter begins our introduction to the analysis of economic time series. By most views, this field has become synonymous with empirical macroeconomics and the analysis of financial markets.[1] In this and the next chapter, we will consider a number of models and topics in which time and relationships through time play an explicit part in the formulation. Consider the **dynamic regression model** KT

$$y_t = \beta_1 + \beta_2 x_t + \beta_3 x_{t-1} + \gamma y_{t-1} + \varepsilon_t. \qquad (\cancel{20}21\text{-}1)$$

Models of this form specifically include as right-hand-side variables previous as well as contemporaneous values of the regressors. It is also in this context that lagged values of the dependent variable appear as a consequence of the theoretical basis of the model rather than as a computational means of removing autocorrelation. There are several reasons lagged effects might appear in an empirical model:

- In modeling the response of economic variables to policy stimuli, it is expected that there will be possibly long lags between policy changes and their impacts. The length of lag between changes in monetary policy and its impact on important economic variables such as output and investment has been a subject of analysis for several decades.
- Either the dependent variable or one of the independent variables is based on expectations. **Expectations** about economic events are usually formed by aggregating new information and past experience. Thus, we might write the expectation of a future value of variable x, formed this period, as KT

$$x_t = E_t[x^*_{t+1} \mid z_t, x_{t-1}, x_{t-2}, \ldots] = g(z_t, x_{t-1}, x_{t-2}, \ldots).$$

[1] The literature in this area has grown at an impressive rate, and, more so than in any other area, it has become impossible to provide comprehensive surveys in general textbooks such as this one. Fortunately, specialized volumes have been produced that can fill this need at any level. Harvey (1990) has been in wide use for some time. Among the many other books, three very useful works are Enders (2003), which presents the basics of time-series analysis at an introductory level with several very detailed applications; Hamilton (1994), which gives a relatively technical but quite comprehensive survey of the field; and Lutkepohl (2005), which provides an extremely detailed treatment of the topics presented at the end of this chapter. Hamilton also surveys a number of the applications in the contemporary literature. Two references that are focused on financial econometrics are Mills (1993) and Tsay (2005). There are also a number of important references that are primarily limited to forecasting, including Diebold (1998a, 2003) and Granger and Newbold (1996). A survey of research in many areas of time-series analysis is Engle and McFadden (1994). An extensive, fairly advanced treatise that analyzes in great depth all the issues we touch on in this chapter is Hendry (1995). Finally, Patterson (2000) surveys most of the practical issues in time series and presents a large variety of useful and very detailed applications.

21-2

AU: Confirm x-ref to Example 15.1 is OK

For example, forecasts of prices and income enter demand equations and consumption equations. (See Example 15.1 for an influential application.)

- Certain economic decisions are explicitly driven by a history of related activities. For example, energy demand by individuals is clearly a function not only of current prices and income, but also the accumulated stocks of energy using capital. Even energy demand in the macroeconomy behaves in this fashion—the stock of automobiles and its attendant demand for gasoline is clearly driven by past prices of gasoline and automobiles. Other classic examples are the dynamic relationship between investment decisions and past appropriation decisions and the consumption of addictive goods such as cigarettes and theater performances.

We begin with a general discussion of models containing **lagged variables.** In Section 20.2, we consider some methodological issues in the specification of dynamic regressions. In Sections 20.3 and 20.4, we describe a general dynamic model that encompasses some of the extensions and more formal models for time-series data that are presented in Chapter 21. Section 20.5 takes a closer look at some of issues in model specification. Finally, Section 20.6 considers systems of dynamic equations. ~~These are largely extensions of the models that we examined at the end of Chapter 13. But the interpretation is rather different here.~~ This chapter is generally not about methods of estimation. OLS and GMM estimation are usually routine in this context. Because we are examining time-series data, conventional assumptions including ergodicity and stationarity will be made at the outset. In particular, in the general framework, we will assume that the multivariate stochastic process $(y_t, \mathbf{x}_t, \varepsilon_t)$ are a **stationary** and ergodic process. As such, without further analysis, we will invoke the theorems discussed in Chapters 4, ~~15, 16~~, and ~~19~~ that support least squares and GMM as appropriate estimate techniques in this context. In most of what follows, in fact, in practical terms, the dynamic regression model can be treated as a linear regression model and estimated by conventional methods (e.g., ordinary least squares or instrumental variables if ε_t is autocorrelated). As noted, we will generally not return to the issue of estimation and inference theory except where new results are needed, such as in the discussion of nonstationary processes.

20.2 DYNAMIC REGRESSION MODELS

In some settings, economic agents respond not only to current values of independent variables but to past values as well. When effects persist over time, an appropriate model will include lagged variables. Example 20.1 illustrates a familiar case.

Example 20.1 A Structural Model of the Demand for Gasoline

Drivers demand gasoline not for direct consumption but as fuel for cars to provide a source of energy for transportation. Per capita demand for gasoline in any period, G/Pop, is determined partly by the current price, P_g, and per capita income, Y/Pop, which influence how intensively the existing stock of gasoline using "capital," K, is used and partly by the size and composition of the stock of cars and other vehicles. The capital stock is determined, in turn, by income, Y/Pop; prices of the equipment such as new and used cars, Pnc and Puc; the price of alternative modes of transportation such as public transportation, Ppt; and past prices of gasoline as they influence forecasts of future gasoline prices. A structural model of

21-3

these effects might appear as follows:

per capita demand: $G_t/Pop_t = \alpha + \beta Pg_t + \delta Y_t/Pop_t + \gamma K_t + u_t,$

stock of vehicles: $K_t = (1-\Delta)K_{t-1} + I_t$, $\Delta =$ depreciation rate,

investment in new vehicles: $I_t = \theta Y_t/Pop_t + \phi E_t[Pg_{t+1}] + \lambda_1 Pnc_t + \lambda_2 Puc_t + \lambda_3 Ppt_t$

expected price of gasoline: $E_t[Pg_{t+1}] = w_0 Pg_t + w_1 Pg_{t-1} + w_2 Pg_{t-2}.$

The capital stock is the sum of all past investments, so it is evident that not only current income and prices, but all past values, play a role in determining K. When income or the price of gasoline changes, the immediate effect will be to cause drivers to use their vehicles more or less intensively. But, over time, vehicles are added to the capital stock, and some cars are replaced with more or less efficient ones. These changes take some time, so the full impact of income and price changes will not be felt for several periods. Two episodes in the recent history have shown this effect clearly. For well over a decade following the 1973 oil shock, drivers gradually replaced their large, fuel-inefficient cars with smaller, less-fuel-intensive models. In the late 1990s in the United States, this process has visibly worked in reverse. As American drivers have become accustomed to steadily rising incomes and steadily falling real gasoline prices, the downsized, efficient coupes and sedans of the 1980s have yielded the highways to a tide of ever-larger, six- and eight-cylinder sport utility vehicles, whose size and power can reasonably be characterized as astonishing.

20.2.1 LAGGED EFFECTS IN A DYNAMIC MODEL

The general form of a dynamic regression model is

$$y_t = \alpha + \sum_{i=0}^{\infty} \beta_i x_{t-i} + \varepsilon_t. \tag{20-2}$$

In this model, a one-time change in x at any point in time will affect $E[y_s \mid x_t, x_{t-1}, \ldots]$ in every period thereafter. When it is believed that the duration of the lagged effects is extremely long—for example, in the analysis of monetary policy—**infinite lag** models that have effects that gradually fade over time are quite common. But models are often constructed in which changes in x cease to have any influence after a fairly small number of periods. We shall consider these **finite lag** models first.

Marginal effects in the static classical regression model are one-time events. The response of y to a change in x is assumed to be immediate and to be complete at the end of the period of measurement. In a dynamic model, the counterpart to a marginal effect is the effect of a one-time change in x_t on the **equilibrium** of y_t. If the level of x_t has been unchanged from, say, $\bar{x}$ for many periods prior to time t, then the equilibrium value of $E[y_t \mid x_t, x_{t-1}, \ldots]$ (assuming that it exists) will be

$$\bar{y} = \alpha + \sum_{i=0}^{\infty} \beta_i \bar{x} = \alpha + \bar{x}\sum_{i=0}^{\infty} \beta_i, \tag{20-3}$$

where $\bar{x}$ is the permanent value of x_t. For this value to be finite, we require that

$$\left| \sum_{i=0}^{\infty} \beta_i \right| < \infty. \tag{20-4}$$

Consider the effect of a unit change in $\bar{x}$ occurring in period s. To focus ideas, consider the earlier example of demand for gasoline and suppose that x_t is the unit price. Prior to the oil shock, demand had reached an equilibrium consistent with accumulated habits,

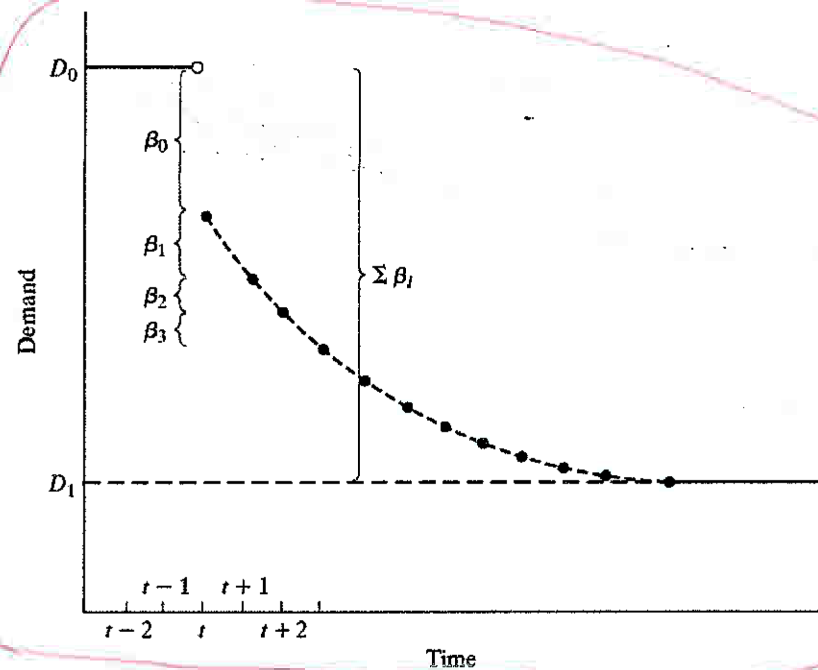

FIGURE ~~20~~21.1 Lagged Adjustment.

FIG 21.1

experience with stable real prices, and the accumulated stocks of vehicles. Now suppose that the price of gasoline, Pg, rises permanently from $\overline{Pg}$ to $\overline{Pg}+1$ in period s. The path to the new equilibrium might appear as shown in Figure ~~20~~21.1. The short-run effect is the one that occurs in the same period as the change in x. This effect is β_0 in the figure.

DEFINITION ~~20~~21.1 Impact Multiplier
β_0 = impact multiplier = short-run multiplier.

DEFINITION ~~20~~21.2 Cumulated Effect
The accumulated effect τ periods later of an impulse at time t is $\beta_\tau = \sum_{i=0}^{\tau} \beta_i$.

In Figure ~~20~~21.1, we see that the total effect of a price change in period t after three periods have elapsed will be $\beta_0 + \beta_1 + \beta_2 + \beta_3$.

The difference between the old equilibrium D_0 and the new one D_1 is the sum of the individual period effects. The **long-run multiplier** is this total effect.

KT

DEFINITION 20.3 Equilibrium Multiplier
$\beta = \sum_{i=0}^{\infty} \beta_i$ = *equilibrium multiplier = long-run multiplier.*

Because the lag coefficients are regression coefficients, their scale is determined by the scales of the variables in the model. As such, it is often useful to define the

$$\textbf{lag weights: } w_i = \frac{\beta_i}{\sum_{j=0}^{\infty} \beta_j}, \tag{20-5}$$

so that $\sum_{i=0}^{\infty} w_i = 1$, and to rewrite the model as

$$y_t = \alpha + \beta \sum_{i=0}^{\infty} w_i x_{t-i} + \varepsilon_t. \tag{20-6}$$

(Note the equation for the expected price in Example 20.1.) Two useful statistics, based on the lag weights, that characterize the period of adjustment to a new equilibrium are the **median lag** = smallest q^* such that $\sum_{i=0}^{q^*} w_i \geq 0.5$ and the **mean lag** = $\sum_{i=0}^{\infty} i w_i$.[2]

KT

FN 2

20.2.2 THE LAG AND DIFFERENCE OPERATORS

KT

A convenient device for manipulating lagged variables is the **lag operator**,

$$L x_t = x_{t-1}.$$

Some basic results are $La = a$ if a is a constant and $L(Lx_t) = L^2 x_t = x_{t-2}$. Thus, $L^p x_t = x_{t-p}$, $L^q(L^p x_t) = L^{p+q} x_t = x_{t-p-q}$, and $(L^p + L^q)x_t = x_{t-p} + x_{t-q}$. By convention, $L^0 x_t = 1x_t = x_t$. A related operation is the first difference,

$$\Delta x_t = x_t - x_{t-1}.$$

Obviously, $\Delta x_t = (1 - L)x_t$ and $x_t = x_{t-1} + \Delta x_t$. These two operations can be usefully combined, for example, as in

$$\Delta^2 x_t = (1 - L)^2 x_t = (1 - 2L + L^2)x_t = x_t - 2x_{t-1} + x_{t-2}.$$

Note that

$$(1 - L)^2 x_t = (1 - L)(1 - L)x_t = (1 - L)(x_t - x_{t-1}) = (x_t - x_{t-1}) - (x_{t-1} - x_{t-2}).$$

The dynamic regression model can be written

$$y_t = \alpha + \sum_{i=0}^{\infty} \beta_i L^i x_t + \varepsilon_t = \alpha + B(L)x_t + \varepsilon_t,$$

[2]If the lag coefficients do not all have the same sign, then these results may not be meaningful. In some contexts, lag coefficients with different signs may be taken as an indication that there is a flaw in the specification of the model.

21-6

where $B(L)$ is a polynomial in L, $B(L) = \beta_0 + \beta_1 L + \beta_2 L^2 + \cdots$. A **polynomial in the lag operator** that reappears in many contexts is

$$A(L) = 1 + aL + (aL)^2 + (aL)^3 + \cdots = \sum_{i=0}^{\infty} (aL)^i.$$

If $|a| < 1$, then

$$A(L) = \frac{1}{1 - aL}.$$

A **distributed lag** model in the form

$$y_t = \alpha + \beta \sum_{i=0}^{\infty} \gamma^i L^i x_t + \varepsilon_t$$

can be written

$$y_t = \alpha + \beta(1 - \gamma L)^{-1} x_t + \varepsilon_t,$$

if $|\gamma| < 1$. This form is called the **moving-average form** or **distributed lag form.** If we multiply through by $(1 - \gamma L)$ and collect terms, then we obtain the **autoregressive form,**

$$y_t = \alpha(1 - \gamma) + \beta x_t + \gamma y_{t-1} + (1 - \gamma L)\varepsilon_t.$$

In more general terms, consider the pth order **autoregressive model,**

$$y_t = \alpha + \beta x_t + \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + \cdots + \gamma_p y_{t-p} + \varepsilon_t,$$

which may be written

$$C(L) y_t = \alpha + \beta x_t + \varepsilon_t,$$

where

$$C(L) = (1 - \gamma_1 L - \gamma_2 L^2 - \cdots - \gamma_p L^p).$$

Can this equation be "inverted" so that y_t is written as a function only of current and past values of x_t and ε_t? By successively substituting the corresponding autoregressive equation for y_{t-1} in that for y_t, then likewise for y_{t-2} and so on, it would appear so. However, it is also clear that the resulting distributed lag form will have an infinite number of coefficients. Formally, the operation just described amounts to writing

$$y_t = [C(L)]^{-1}(\alpha + \beta x_t + \varepsilon_t) = A(L)(\alpha + \beta x_t + \varepsilon_t).$$

It will be of interest to be able to solve for the elements of $A(L)$ (see, for example, Section ~~20~~.6.6). By this arrangement, it follows that $C(L)A(L) = 1$ where

21

$$A(L) = (\alpha_0 L^0 + \alpha_1 L + \alpha_2 L^2 + \cdots).$$

plus

By collecting like powers of L in

$$(1 - \gamma_1 L - \gamma_2 L^2 - \cdots - \gamma_p L^p)(\alpha_0 L^0 + \alpha_1 L + \alpha_2 L^2 + \cdots) = 1,$$

plus

21-7

we find that a recursive solution for the α coefficients is

$$
\begin{aligned}
&L^0\colon\ \alpha_0 && = 1\\
&L^1\colon\ \alpha_1 - \gamma_1\alpha_0 && = 0\\
&L^2\colon\ \alpha_2 - \gamma_1\alpha_1 - \gamma_2\alpha_0 && = 0\\
&L^3\colon\ \alpha_3 - \gamma_1\alpha_2 - \gamma_2\alpha_1 - \gamma_3\alpha_0 && = 0\\
&L^4\colon\ \alpha_4 - \gamma_1\alpha_3 - \gamma_2\alpha_2 - \gamma_3\alpha_1 - \gamma_4\alpha_0 && = 0\\
&\ldots\\
&L^p\colon\ \alpha_p - \gamma_1\alpha_{p-1} - \gamma_2\alpha_{p-2} - \cdots - \gamma_p\alpha_0 && = 0
\end{aligned}
\qquad (\text{~~20~~}21\text{-}7)
$$

and, thereafter,

$$L^q\colon\ \alpha_q - \gamma_1\alpha_{q-1} - \gamma_2\alpha_{q-2} - \cdots - \gamma_p\alpha_{q-p} = 0.$$

After a set of $p-1$ starting values, the α coefficients obey the same difference equation as y_t does in the dynamic equation. One problem remains. For the given set of values, the preceding gives no assurance that the solution for α_q does not ultimately explode. The preceding equation system is not necessarily stable for all values of γ_j (although it certainly is for some). If the system is stable in this sense, then the polynomial $C(L)$ is said to be **invertible.** The necessary conditions are precisely those discussed in Section ~~20~~21.4.3, so we will defer completion of this discussion until then.

KT

Finally, two useful results are

$$B(1) = \beta_0 1^0 + \beta_1 1^1 + \beta_2 1^2 + \cdots = \beta = \text{long-run multiplier},$$

and

$$B'(1) = [dB(L)/dL]_{|L=1} = \sum_{i=0}^{\infty} i\beta_i.$$

It follows that $B'(1)/B(1) =$ mean lag.

~~20~~21.2.3 SPECIFICATION SEARCH FOR THE LAG LENGTH

Various procedures have been suggested for determining the appropriate lag length in a dynamic model such as

$$y_t = \alpha + \sum_{i=0}^{p} \beta_i x_{t-i} + \varepsilon_t. \qquad (\text{~~20~~}21\text{-}8)$$

One must be careful about a purely significance based specification search. Let us suppose that there is an appropriate, "true" value of $p > 0$ that we seek. A **simple-to-general approach** to finding the right lag length would depart from a model with only the current value of the independent variable in the regression and add deeper lags until a simple t test suggested that the last one added is statistically insignificant. The problem with such an approach is that at any level at which the number of included lagged variables is less than p, the estimator of the coefficient vector is biased and inconsistent. [See the omitted variable formula (7-4).] The asymptotic covariance matrix is biased as well, so statistical inference on this basis is unlikely to be successful. A **general-to-simple approach** would begin from a model that contains more than p lagged values—it

KT KT KT KT

4-10

21-8

is assumed that although the precise value of p is unknown, the analyst can posit a maintained value that should be larger than p. Least squares or instrumental variables regression of y on a constant and $(p+d)$ lagged values of x consistently estimates $\boldsymbol{\theta} = [\alpha, \beta_0, \beta_1, \ldots, \beta_p, 0, 0, \ldots]$.

Because models with lagged values are often used for forecasting, researchers have tended to look for measures that have produced better results for assessing "out of sample" prediction properties. The adjusted R^2 [see Section 3.5.1] is one possibility. Others include the Akaike (1973) information criterion, AIC(p),

$$\text{AIC}(p) = \ln \frac{\mathbf{e}'\mathbf{e}}{T} + \frac{2p}{T}, \tag{20-9}$$

and Schwarz's criterion, SC(p):

$$\text{SC}(p) = \text{AIC}(p) + \left(\frac{p}{T}\right)(\ln T - 2). \tag{20-10}$$

(See Section ~~1.4~~ 5.10.1.) If some maximum P is known, then $p < P$ can be chosen to minimize AIC(p) or SC(p).[3] An alternative approach, also based on a known P, is to do sequential F tests on the last $P > p$ coefficients, stopping when the test rejects the hypothesis that the coefficients are jointly zero. Each of these approaches has its flaws and virtues. The Akaike information criterion retains a positive probability of leading to overfitting even as $T \to \infty$. In contrast, SC(p) has been seen to lead to underfitting in some finite sample cases. They do avoid, however, the inference problems of sequential estimators. The sequential F tests require successive revision of the significance level to be appropriate, but they do have a statistical underpinning.[4]

AU: Term "infinite lag model" was KT on msp 21-3. Here also?

20.3 SIMPLE DISTRIBUTED LAG MODELS

Before examining some very general specifications of the dynamic regression, we briefly consider an **infinite lag model,** which emerges from a simple model of expectations.

There are cases in which the distributed lag models the accumulation of information. The formation of expectations is an example. In these instances, intuition suggests that the most recent past will receive the greatest weight and that the influence of past observations will fade uniformly with the passage of time. The geometric lag model is often used for these settings. The general form of the model is

$$\begin{aligned} y_t &= \alpha + \beta \sum_{i=0}^{\infty} (1-\lambda)\lambda^i x_{t-i} + \varepsilon_t, \quad 0 < \lambda < 1, \\ &= \alpha + \beta B(L) x_t + \varepsilon_t, \end{aligned} \tag{20-11}$$

where

$$B(L) = (1-\lambda)(1 + \lambda L + \lambda^2 L^2 + \lambda^3 L^3 + \cdots) = \frac{1-\lambda}{1-\lambda L}.$$

[3] For further discussion and some alternative measures, see Geweke and Meese (1981), Amemiya (1985, pp. 146–147), Diebold (1998, pp. 85–91), and Judge et al. (1985, pp. 353–355).

[4] See Pagano and Hartley (1981) and Trivedi and Pagan (1979).

21-9

KT

The lag coefficients are $\beta_i = \beta(1-\lambda)\lambda^i$. The model incorporates **infinite lags**, but it assigns arbitrarily small weights to the distant past. The lag weights decline geometrically;

$$w_i = (1-\lambda)\lambda^i, \quad 0 \le w_i < 1.$$

AU: Both terms "mean lag" "median lag" were KTs on msp 21-3. Here also?

The **mean lag** is

$$\overline{w} = \frac{B'(1)}{B(1)} = \frac{\lambda}{1-\lambda}.$$

The **median lag** is p^* such that $\sum_{i=0}^{p^*-1} w_i = 0.5$. We can solve for p^* by using the result

$$\sum_{i=0}^{p} \lambda^i = \frac{1-\lambda^{p+1}}{1-\lambda}.$$

Thus,

$$p^* = \frac{\ln 0.5}{\ln \lambda} - 1.$$

The impact multiplier is $\beta(1-\lambda)$. The long-run multiplier is $\beta\sum_{i=0}^{\infty}(1-\lambda)\lambda^i = \beta$. The equilibrium value of y_t would be found by fixing x_t at $\overline{x}$ and ε_t at zero in (~~20~~21-11), which produces $\overline{y} = \alpha + \beta\overline{x}$.

The geometric lag model can be motivated with an economic model of expectations. We begin with a regression in an expectations variable such as an expected future price based on information available at time t, $x^*_{t+1|t}$, and perhaps a second regressor, w_t,

$$y_t = \alpha + \beta x^*_{t+1|t} + \delta w_t + \varepsilon_t,$$

and a mechanism for the formation of the expectation,

$$x^*_{t+1|t} = \lambda x^*_{t|t-1} + (1-\lambda)x_t = \lambda L x^*_{t+1|t} + (1-\lambda)x_t. \qquad (\cancel{20}\,21\text{-}12)$$

The currently formed expectation is a weighted average of the expectation in the previous period and the most recent observation. The parameter λ is the adjustment coefficient. If λ equals 1, then the current datum is ignored and expectations are never revised. A value of zero characterizes a strict pragmatist who forgets the past immediately. The expectation variable can be written as

$$x^*_{t+1|t} = \frac{1-\lambda}{1-\lambda L}x_t = (1-\lambda)[x_t + \lambda x_{t-1} + \lambda^2 x_{t-2} + \cdots]. \qquad (\cancel{20}\,21\text{-}13)$$

Inserting (~~20~~21-13) into (~~20~~21-12) produces the geometric distributed lag model,

$$y_t = \alpha + \beta(1-\lambda)[x_t + \lambda x_{t-1} + \lambda^2 x_{t-2} + \cdots] + \delta w_t + \varepsilon_t.$$

The geometric lag model can be estimated by nonlinear least squares. Rewrite it as

$$y_t = \alpha + \gamma z_t(\lambda) + \delta w_t + \varepsilon_t, \quad \gamma = \beta(1-\lambda). \qquad (\cancel{20}\,21\text{-}14)$$

The constructed variable $z_t(\lambda)$ obeys the recursion $z_t(\lambda) = x_t + \lambda z_{t-1}(\lambda)$. For the first observation, we use $z_1(\lambda) = x^*_{1|0} = x_1/(1-\lambda)$. If the sample is moderately long, then assuming that x_t was in long-run equilibrium, although it is an approximation, will not unduly affect the results. One can then scan over the range of λ from zero to one to locate the value that minimizes the sum of squares. Once the minimum is located, an estimate of the asymptotic covariance matrix of the estimators of $(\alpha, \gamma, \delta, \lambda)$ can be

21-10

found using ~~(11-12)~~ 7-15 and Theorem ~~11.2~~ 7.2. For the regression function $h_t(\text{data} \mid \alpha, \gamma, \delta, \lambda)$, $x_{t1}^0 = 1, x_{t2}^0 = z_t(\lambda)$, and $x_{t3}^0 = w_t$. The derivative with respect to λ can be computed by using the recursion $d_t(\lambda) = \partial z_t(\lambda)/\partial\lambda = z_{t-1}(\lambda) + \lambda \partial z_{t-1}(\lambda)/\partial\lambda$. If $z_1 = x_1/(1-\lambda)$, then $d_1(\lambda) = z_1/(1-\lambda)$. Then, $x_{t4}^0 = d_t(\lambda)$. Finally, we estimate β from the relationship $\beta = \gamma/(1-\lambda)$ and use the delta method to estimate the asymptotic standard error.

For purposes of estimating long- and short-run elasticities, researchers often use a different form of the geometric lag model. The **partial adjustment** model describes the *desired* level of y_t,

$$y_t^* = \alpha + \beta x_t + \delta w_t + \varepsilon_t,$$

and an *adjustment equation*,

$$y_t - y_{t-1} = (1-\lambda)(y_t^* - y_{t-1}).$$

If we solve the second equation for y_t and insert the first expression for y_t^*, then we obtain

$$\begin{aligned} y_t &= \alpha(1-\lambda) + \beta(1-\lambda)x_t + \delta(1-\lambda)w_t + \lambda y_{t-1} + (1-\lambda)\varepsilon_t \\ &= \alpha' + \beta' x_t + \delta' w_t + \lambda y_{t-1} + \varepsilon_t'. \end{aligned}$$

This formulation offers a number of significant practical advantages. It is intrinsically linear in the parameters (unrestricted), and its disturbance is nonautocorrelated if ε_t was to begin with. As such, the parameters of this model can be estimated consistently and efficiently by ordinary least squares. In this revised formulation, the short-run multipliers for x_t and w_t are β' and δ'. The long-run effects are $\beta = \beta'/(1-\lambda)$ and $\delta = \delta'/(1-\lambda)$. With the variables in logs, these effects are the short- and long-run elasticities.

Example ~~20~~21.2 Expectations-Augmented Phillips Curve

In Example ~~19~~20.3, we estimated an expectations-augmented Phillips curve of the form

$$\Delta p_t - E[\Delta p_t \mid \Psi_{t-1}] = \beta[u_t - u^*] + \varepsilon_t.$$

Our model assumed a particularly simple model of expectations, $E[\Delta p_t \mid \Psi_{t-1}] = \Delta p_{t-1}$. The least squares results for this equation were

$$\begin{aligned} \Delta p_t - \Delta p_{t-1} &= \underset{(0.7405)}{0.49189} - \underset{(0.1257)}{0.090136}\,u_t + e_t \qquad R^2 = 0.002561,\ T = 201. \end{aligned}$$

202

The implied estimate of the natural rate of unemployment is $-(0.49189/-0.090136)$ or about 5.46 percent. Suppose we allow expectations to be formulated less pragmatically with the expectations model in (~~20~~21-12). For this setting, this would be

$$E[\Delta p_t \mid \Psi_{t-1}] = \lambda E[\Delta p_{t-1} \mid \Psi_{t-2}] + (1-\lambda)\Delta p_{t-1}.$$

The strict pragmatist has $\lambda = 0.0$. Using the method set out earlier, we would compute this for different values of λ, recompute the dependent variable in the regression, and locate the value of λ which produces the lowest sum of squares. Figure ~~20~~21.2 shows the sum of squares for the values of λ ranging from 0.0 to 1.0.

FIG 21.2

The minimum value of the sum of squares occurs at $\lambda = 0.66$. The least squares regression results are

$$\Delta p_t - \widehat{\Delta p_{t-1}} = \underset{(0.6617)}{1.69453} - \underset{(0.11125)}{0.30427}\,u_t + e_t \qquad T = 201.$$

202

21 - 11

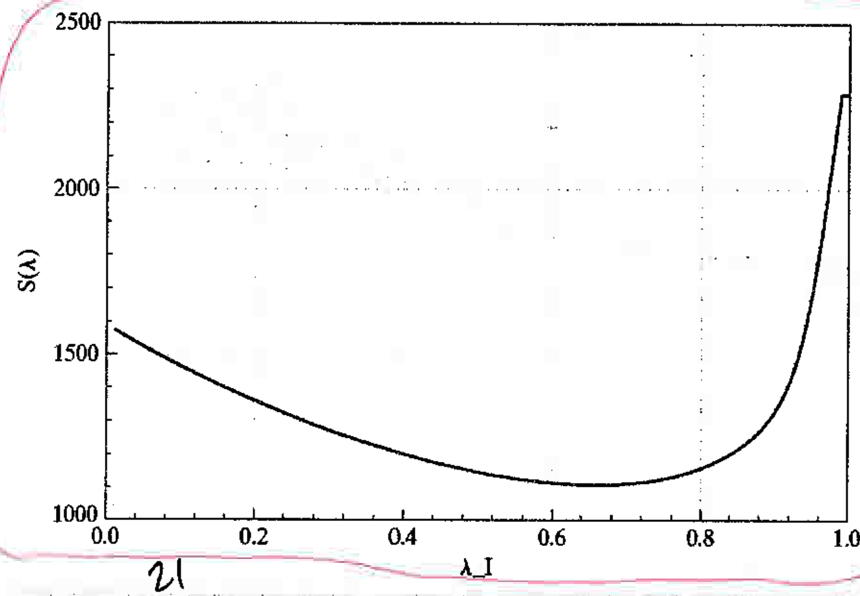

FIGURE 20.2 Residuals Sums of Squares for Phillips Curve Estimates.

The estimated standard errors are computed using the method described earlier for the nonlinear regression. The extra variable described in the paragraph after (20-14) accounts for the estimated λ. The estimated asymptotic covariance matrix is then computed using $(\mathbf{e}'\mathbf{e}/201)[\mathbf{W}'\mathbf{W}]^{-1}$ where $w_1 = 1, w_2 = u_t$ and $w_3 = \partial \Delta p_{t-1}/\partial\lambda$. The estimated standard error for λ is 0.04610. Because this is highly statistically significantly different from zero ($t = 14.315$), we would reject the simple model. Finally, the implied estimate of the natural rate of unemployment is $-(-1.69453/0.30427)$ or about 5.57 percent. The estimated asymptotic covariance of the slope and constant term is -0.0720293, so, using this value and the estimated standard errors given earlier and the delta method, we obtain an estimated standard error for this estimate of 0.5467. Thus, a confidence interval for the natural rate of unemployment based on these results would be (4.49 percent, 6.64 percent), which is in line with our prior expectations. There are two things to note about these results. First, because the dependent variables are different, we cannot compare the R^2s of the models with $\lambda = 0.00$ and $\lambda = 0.66$. But, the sum of squares for the two models can be compared (they are 1592.32 and 1112.89), so the second model fits far better. One of the payoffs is the much narrower confidence interval for the natural rate. The counterpart to the one given earlier when $\lambda = 0.00$ is (1.13%, 9.79%). No doubt the model could be improved still further by expanding the equation. (This is considered in the exercises.)

Example 20.3 Price and Income Elasticities of Demand for Gasoline

We have extended the gasoline demand equation estimated in Examples 19.2 and 19.6 to allow for dynamic effects. Table 20.1 presents estimates of three distributed lag models for gasoline consumption. The unrestricted model allows five years of adjustment in the price and income effects. The expectations model includes the same distributed lag (λ) on price and income but different long-run multipliers (β_{Pg} and β_I). [Note, for this formulation, that the extra regressor used in computing the asymptotic covariance matrix is $d_t(\lambda) = \beta_{Pg} d_{\text{price}}(\lambda) + \beta_I d_{\text{income}}(\lambda)$.] Finally, the partial adjustment model implies lagged effects for all the variables in the model. To facilitate comparison, the constant and the first four slope coefficients in the partial adjustment model have been divided by the estimate of $(1 - \lambda)$. The implied long- and short-run price and income elasticities are shown in Table 20.2.

TB 21.1

TB 21.2

21-12

TABLE ~~20~~21.1 Estimated Distributed Lag Models

		Expectations		*Partial Adjustment*	
Coefficient	*Unrestricted*	*Estimated*	*Derived*	*Estimated*	*Derived*
Constant	−28.5512	−16.1867		−4.9489	
ln *Pnc*	0.01738	−0.1050		−0.1429	
ln *Puc*	0.07602	0.02815		0.09435	
ln *Ppt*	0.04770	0.2550		0.03243	
Trend	−0.02297	0.02064		−0.004029	
ln *Pg*	−0.08282	−0.06702*	−0.06702*	−0.07627	−0.07627
ln *Pg*[−1]	−0.07152		−0.06233		−0.06116
ln *Pg*[−2]	0.03669		−0.05797		−0.04904
ln *Pg*[−3]	−0.04814		−0.05391		−0.03933
ln *Pg*[−4]	0.02958		−0.05013		−0.03153
ln *Pg*[−5]	−0.1481		−0.04663		−0.02529
ln Income	1.1074	0.04372*	0.04372*	0.3135	0.3135
ln Income[−1]	0.3776		0.04066		0.2514
ln Income[−2]	−0.01255		0.03781		0.2016
ln Income[−3]	−0.03919		0.03517		0.1616
ln Income[−4]	0.2737		0.03270		0.1296
ln Income[−5]	0.09350		0.03042		0.1039
Zt(Price)	—	−0.06702			
Zt(Income)	—	0.04372			
ln (*G*/*Pop*)[−1]	—			0.80188	
β	—	−0.9574			
γ	—	0.6245			
λ	—	0.9300		0.80188	
$e'e$	0.01565356	0.03911383		.01151860	
T	47	52		51	

*Estimated directly

TABLE ~~20~~21.2 Estimated Elasticities

	Short-Run		*Long-Run*	
	Price	*Income*	*Price*	*Income*
Unrestricted model	−0.08282	1.1074	−0.2843	1.8004
Expectations model	−0.06702	0.04372	−0.9574	0.6246
Partial adjustment model	−0.07628	0.3135	−0.3850	1.5823

~~20~~21.4 AUTOREGRESSIVE DISTRIBUTED LAG MODELS

Both the finite lag models and the geometric lag model impose strong, possibly incorrect restrictions on the lagged response of the dependent variable to changes in an independent variable. A very general compromise that also provides a useful platform KT for studying a number of interesting methodological issues is the **autoregressive distributed lag (ARDL)** model. KT

$$y_t = \mu + \sum_{i=1}^{p} \gamma_i y_{t-i} + \sum_{j=0}^{r} \beta_j x_{t-j} + \delta w_t + \varepsilon_t, \qquad (~~20~~21\text{-}15)$$

21-13

22

in which ε_t is assumed to be serially uncorrelated and homoscedastic (we will relax both these assumptions in Chapter ~~21~~). We can write this more compactly as

$$C(L)y_t = \mu + B(L)x_t + \delta w_t + \varepsilon_t$$

by defining polynomials in the lag operator,

$$C(L) = 1 - \gamma_1 L - \gamma_2 L^2 - \cdots - \gamma_p L^p,$$

and

$$B(L) = \beta_0 + \beta_1 L + \beta_2 L^2 + \cdots + \beta_r L^r.$$

The model in this form is denoted ARDL(p, r) to indicate the orders of the two polynomials in L. The partial adjustment model estimated in the previous section is the special case in which p equals 1 and r equals 0. A number of other special cases are also interesting, including the familiar model of **autocorrelation** ($p = 1, r = 1, \beta_1 = -\gamma_1\beta_0$), the classical regression model ($p = 0, r = 0$), and so on.

21 ~~20~~.4.1 ESTIMATION OF THE ARDL MODEL

Save for the presence of the stochastic right-hand-side variables, the ARDL is a linear model with a classical disturbance. As such, ordinary least squares is the efficient estimator. The lagged dependent variable does present a complication, but we considered this in Chapter ~~19~~. Absent any obvious violations of the assumptions there, least squares continues to be the estimator of choice. Conventional testing procedures are, as before, asymptotically valid as well. Thus, for testing linear restrictions, the Wald statistic can be used, although the F statistic is generally preferable in finite samples because of its more conservative critical values.

20

One subtle complication in the model has attracted a large amount of attention in the recent literature. If $C(1) = 0$, then the model is actually inestimable. This fact is evident in the distributed lag form, which includes a term $\mu/C(1)$. If the equivalent condition $\Sigma_i \gamma_i = 1$ holds, then the stochastic difference equation is unstable and a host of other problems arise as well. This implication suggests that one might be interested in testing this specification as a hypothesis in the context of the model. This restriction might seem to be a simple linear constraint on the alternative (unrestricted) model in (~~20~~-15). Under the null hypothesis, however, the conventional test statistics do not have the familiar distributions. The formal derivation is complicated [in the extreme, see Dickey and Fuller (1979) for an example], but intuition should suggest the reason. Under the null hypothesis, the difference equation is explosive, so our assumptions about well behaved data cannot be met. Consider a simple ARDL(1, 0) example and simplify it even further with $B(L) = 0$. Then,

21

$$y_t = \mu + \gamma y_{t-1} + \varepsilon_t.$$

If γ equals 1, then

$$y_t = \mu + y_{t-1} + \varepsilon_t.$$

Assuming we start the time series at time $t = 1$,

$$y_t = t\mu + \Sigma_s \varepsilon_s = t\mu + v_t.$$

21-14

The conditional mean in this **random walk with drift** model is increasing without limit, so the unconditional mean does not exist. The conditional mean of the disturbance, v_t, is zero, but its conditional variance is $t\sigma^2$, which shows a peculiar type of heteroscedasticity. Consider least squares estimation of μ with $m = (\mathbf{t}'\mathbf{y})/(\mathbf{t}'\mathbf{t})$, where $\mathbf{t} = [1, 2, 3, \ldots, T]$. Then $E[m] = \mu + E[(\mathbf{t}'\mathbf{t})^{-1}(\mathbf{t}'\mathbf{v})] = \mu$, but

$$\mathrm{Var}[m] = \frac{\sigma^2 \sum_{t=1}^{T} t^3}{\left(\sum_{t=1}^{T} t^2\right)^2} = \frac{O(T^4)}{[O(T^3)]^2} = O\left(\frac{1}{T^2}\right).$$

So, the variance of this estimator is an order of magnitude smaller than we are used to seeing in regression models. Not only is m mean square consistent, it is **superconsistent.** As such, without doing a formal derivation, we conclude that there is something "unusual" about this estimator and that the "usual" testing procedures whose distributions build on the distribution of $\sqrt{T}(m - \mu)$ will not be appropriate; the variance of this normalized statistic converges to zero.

This result does not mean that the hypothesis $\gamma = 1$ is not testable in this model. In fact, the appropriate test statistic is the conventional one that we have computed for comparable tests before. But the appropriate critical values against which to measure those statistics are quite different. We will return to this issue in our discussion of the Dickey–Fuller test in Section ~~22~~ 23.2.4.

~~20~~ 21.4.2 COMPUTATION OF THE LAG WEIGHTS IN THE ARDL MODEL

The distributed lag form of the ARDL model is

$$\begin{aligned} y_t &= \frac{\mu}{C(L)} + \frac{B(L)}{C(L)} x_t + \frac{1}{C(L)} \delta w_t + \frac{1}{C(L)} \varepsilon_t \\ &= \frac{\mu}{1 - \gamma_1 - \cdots - \gamma_p} + \sum_{j=0}^{\infty} \alpha_j x_{t-j} + \delta \sum_{l=0}^{\infty} \theta_l w_{t-l} + \sum_{l=0}^{\infty} \theta_l \varepsilon_{t-l}. \end{aligned}$$

This model provides a method of approximating a very general lag structure. In Jorgenson's (1966) study, in which he labeled this model a **rational lag** model, he demonstrated that essentially any desired shape for the lag distribution could be produced with relatively few parameters.[5]

The lag coefficients on $x_t, x_{t-1}, \ldots$, in the ARDL model are the individual terms in the ratio of polynomials that appear in the distributed lag form. We denote these as coefficients

$$\alpha_0, \alpha_1, \alpha_2, \ldots = \text{ the coefficient on } 1, L, L^2, \ldots \text{ in } \frac{B(L)}{C(L)}. \qquad \text{(~~20~~ 21-16)}$$

A convenient way to compute these coefficients is to write (~~20~~ 21-16) as $A(L)C(L) = B(L)$. Then we can just equate coefficients on the powers of L. Example ~~20~~ 21.4 demonstrates the procedure.

[5] A long literature, highlighted by Griliches (1967), Dhrymes (1971), Nerlove (1972), Maddala (1977a), and Harvey (1990), describes estimation of models of this sort.

21-15

The long-run effect in a rational lag model is $\sum_{i=0}^{\infty}\alpha_i$. This result is easy to compute because it is simply

$$\sum_{i=0}^{\infty}\alpha_i = \frac{B(1)}{C(1)}.$$

A standard error for the long-run effect can be computed using the delta method.

21 ~~20~~.4.3 STABILITY OF A DYNAMIC EQUATION

KT

In the geometric lag model, we found that a **stability** condition $|\lambda| < 1$ was necessary for the model to be well behaved. Similarly, in the AR(1) model, the autocorrelation parameter ρ must be restricted to $|\rho| < 1$ for the same reason. The dynamic model in 21 (~~20~~-15) must also be restricted, but in ways that are less obvious. Consider once again the question of whether there exists an equilibrium value of y_t.

21 In (~~20~~-15), suppose that x_t is fixed at some value $\overline{x}$, w_t is fixed at zero, and the disturbances ε_t are fixed at their expectation of zero. Would y_t converge to an equilibrium? The relevant dynamic equation is

$$y_t = \overline{\alpha} + \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + \cdots + \gamma_p y_{t-p},$$

where $\overline{\alpha} = \mu + B(1)\overline{x}$. If y_t converges to an equilibrium, then, that equilibrium is

$$\overline{y} = \frac{\mu + B(1)\overline{x}}{C(1)} = \frac{\overline{\alpha}}{C(1)}.$$

KT

Stability of a dynamic equation hinges on the **characteristic equation** for the autoregressive part of the model. The roots of the characteristic equation,

$$C(z) = 1 - \gamma_1 z - \gamma_2 z^2 - \cdots - \gamma_p z^p = 0, \qquad \text{(21 ~~20~~-17)}$$

must be greater than one in absolute value for the model to be stable. To take a simple example, the characteristic equation for the first-order models we have examined thus far is

$$C(z) = 1 - \lambda z = 0.$$

The single root of this equation is $z = 1/\lambda$, which is greater than one in absolute value if $|\lambda|$ is less than one. The roots of a more general characteristic equation are the reciprocals of the characteristic roots of the matrix

$$\mathbf{C} = \begin{bmatrix} \gamma_1 & \gamma_2 & \gamma_3 & \cdots & \gamma_{p-1} & \gamma_p \\ 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ & & & \ldots & & \\ 0 & 0 & 0 & \ldots & 1 & 0 \end{bmatrix}. \qquad \text{(21 ~~20~~-18)}$$

Because the matrix is asymmetric, its roots may include complex pairs. The reciprocal of the complex number $a + bi$ is $a/M - (b/M)i$, where $M = a^2 + b^2$ and $i^2 = -1$. We thus require that M be less than 1.

The case of $z = 1$, the unit root case, is often of special interest. If one of the roots of $C(z) = 0$ is 1, then it follows that $\sum_{i=1}^{p}\gamma_i = 1$. This assumption would appear to be a simple hypothesis to test in the framework of the ARDL model. Instead, we find the explosive case that we examined in Section 20.4.1, so the hypothesis is more complicated than it first appears. To reiterate, under the null hypothesis that $C(1) = 0$, it is not possible for the standard F statistic to have a central F distribution because of the behavior of the variables in the model. We will return to this case shortly.

The **univariate autoregression,** KT

$$y_t = \mu + \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + \cdots + \gamma_p y_{t-p} + \varepsilon_t,$$

can be augmented with the $p - 1$ equations

$$y_{t-1} = y_{t-1},$$

$$y_{t-2} = y_{t-2},$$

and so on to give a **vector autoregression, VAR** (to be considered in the next section):

$$\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{C}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t,$$

where $\mathbf{y}_t$ has p elements, $\boldsymbol{\varepsilon}_t = (\varepsilon_t, 0, \ldots)'$, and $\boldsymbol{\mu} = (\mu, 0, 0, \ldots)'$. It will ultimately not be relevant to the solution, so we will let ε_t equal its expected value of zero. Now, by successive substitution, we obtain

$$\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{C}\boldsymbol{\mu} + \mathbf{C}^2\boldsymbol{\mu} + \cdots,$$

which may or may not converge. Write $\mathbf{C}$ in the spectral form $\mathbf{C} = \mathbf{P\Lambda Q}$, where $\mathbf{QP} = \mathbf{I}$ and $\boldsymbol{\Lambda}$ is a diagonal matrix of the characteristic roots. (Note that the characteristic roots in $\boldsymbol{\Lambda}$ and vectors in $\mathbf{P}$ and $\mathbf{Q}$ may be complex.) We then obtain

$$\mathbf{y}_t = \left[\sum_{i=0}^{\infty} \mathbf{P}\boldsymbol{\Lambda}^i\mathbf{Q}\right]\boldsymbol{\mu}. \tag{20-19}$$

If all the roots of $\mathbf{C}$ are less than one in absolute value, then this vector will converge to the equilibrium

$$\mathbf{y}_\infty = (\mathbf{I} - \mathbf{C})^{-1}\boldsymbol{\mu}.$$

Nonexplosion of the powers of the roots of $\mathbf{C}$ is equivalent to $|\lambda_p| < 1$, or $|1/\lambda_p| > 1$, which was our original requirement. Note finally that because $\boldsymbol{\mu}$ is a multiple of the first column of $\mathbf{I}_p$, it must be the case that each element in the first column of $(\mathbf{I} - \mathbf{C})^{-1}$ is the same. At equilibrium, therefore, we must have $y_t = y_{t-1} = \cdots = y_\infty$.

Example 20.4 A Rational Lag Model

F5.2 → Appendix Table F5.1 lists quarterly data on a number of macroeconomic variables including consumption and real GDP for the U.S. economy for the years 1950 to 2000, a total of 204 quarters. The model

$$c_t = \delta + \beta_0 y_t + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-3} + \gamma_1 c_{t-1} + \gamma_2 c_{t-2} + \gamma_3 c_{t-3} + \varepsilon_t$$

is estimated using the logarithms of real consumption and real GDP, denoted c_t and y_t. Ordinary least squares estimates of the parameters of the ARDL(3,3) model are

$$c_t = 0.7233c_{t-1} + 0.3914c_{t-2} - 0.2337c_{t-3}$$
$$+ 0.5651y_t - 0.3909y_{t-1} - 0.2379y_{t-2} + 0.1902y_{t-3} + e_t.$$

21-17

TABLE ~~20~~21.3 Lag Coefficients in a Rational Lag Model

Lag	0	1	2	3	4	5	6	7
ARDL	0.565	0.018	−0.004	0.062	0.039	0.054	0.039	0.041
Unrestricted	0.954	−0.090	−0.063	0.100	−0.024	0.057	−0.112	0.236

(A full set of quarterly dummy variables is omitted.) The Durbin–Watson statistic is 1.78597, so remaining autocorrelation seems unlikely to be a consideration. The lag coefficients are given by the equality

$$(\alpha_0+\alpha_1 L+\alpha_2 L^2+\cdots)(1-\gamma_1 L-\gamma_2 L^2-\gamma_3 L^3)=(\beta_0+\beta_1 L+\beta_2 L^2+\beta_3 L^3).$$

Note that $A(L)$ is an infinite polynomial. The lag coefficients are

$$
\begin{aligned}
&1:\quad \alpha_0 && =\beta_0 \text{ (which will always be the case)},\\
&L^1:\ -\alpha_0\gamma_1+\alpha_1 && =\beta_1 \text{ or } \alpha_1=\beta_1+\alpha_0\gamma_1,\\
&L^2:\ -\alpha_0\gamma_2-\alpha_1\gamma_1+\alpha_2 && =\beta_2 \text{ or } \alpha_2=\beta_2+\alpha_0\gamma_2+\alpha_1\gamma_1,\\
&L^3:\ -\alpha_0\gamma_3-\alpha_1\gamma_2-\alpha_2\gamma_1+\alpha_3 && =\beta_3 \text{ or } \alpha_3=\beta_3+\alpha_0\gamma_3+\alpha_1\gamma_2+\alpha_2\gamma_1,\\
&L^4:\ -\alpha_1\gamma_3-\alpha_2\gamma_2-\alpha_3\gamma_1+\alpha_4 && =0 \text{ or } \alpha_4=\gamma_1\alpha_3+\gamma_2\alpha_2+\gamma_3\alpha_1,\\
&L^j:\ -\alpha_{j-3}\gamma_3-\alpha_{j-2}\gamma_2-\alpha_{j-1}\gamma_1+\alpha_j && =0 \text{ or } \alpha_j=\gamma_1\alpha_{j-1}+\gamma_2\alpha_{j-2}+\gamma_3\alpha_{j-3},\quad j=5,6,\ldots,
\end{aligned}
$$

and so on. From the fourth term onward, the series of lag coefficients follows the recursion $\alpha_j=\gamma_1\alpha_{j-1}+\gamma_2\alpha_{j-2}+\gamma_3\alpha_{j-3}$, which is the same as the autoregressive part of the ARDL model. The series of lag weights follows the same difference equation as the current and lagged values of y_t after r initial values, where r is the order of the DL part of the ARDL model. The three characteristic roots of the **C** matrix are 0.8631, −0.5949, and 0.4551. Because all are less than one, we conclude that the stochastic difference equation is stable.

TB 21.3

The first seven lag coefficients of the estimated ARDL model are listed in Table ~~20~~21.3 with the first seven coefficients in an unrestricted lag model. The coefficients from the ARDL model only vaguely resemble those from the unrestricted model, but the erratic swings of the latter are prevented by the smooth equation from the distributed lag model. The estimated long-term effects (with standard errors in parentheses) from the two models are 1.0634 (0.00791) from the ARDL model and 1.0570 (0.002135) from the unrestricted model. Surprisingly, in view of the large and highly significant estimated coefficients, the lagged effects fall off essentially to zero after the initial impact.

~~20~~21.4.4 FORECASTING

KT

Consider, first, a **one-period-ahead forecast** of y_t in the ARDL(p, r) model. It will be convenient to collect the terms in μ, x_t, w_t, and so on in a single term,

$$\mu_t=\mu+\sum_{j=0}^{r}\beta_j x_{t-j}+\delta w_t.$$

Now, the ARDL model is just

$$y_t=\mu_t+\gamma_1 y_{t-1}+\cdots+\gamma_p y_{t-p}+\varepsilon_t.$$

Conditioned on the full set of information available up to time T and on forecasts of the exogenous variables, the one-period-ahead forecast of y_t would be

$$\hat{y}_{T+1|T}=\hat{\mu}_{T+1|T}+\gamma_1 y_T+\cdots+\gamma_p y_{T-p+1}+\hat{\varepsilon}_{T+1|T}.$$

21-18

To form a prediction interval, we will be interested in the variance of the forecast error,

$$e_{T+1|T} = \hat{y}_{T+1|T} - y_{T+1}.$$

This error will arise from three sources. First, in forecasting μ_t, there will be two sources of error. The parameters, μ, δ, and $\beta_0, \ldots, \beta_r$ will have been estimated, so $\hat{\mu}_{T+1|T}$ will differ from μ_{T+1} because of the sampling variation in these estimators. Second, if the exogenous variables, x_{T+1} and w_{T+1} have been forecasted, then to the extent that these forecasts are themselves imperfect, yet another source of error to the forecast will result. Finally, although we will forecast ε_{T+1} with its expectation of zero, we would not assume that the actual realization will be zero, so this step will be a third source of error. In principle, an estimate of the forecast variance, $\text{Var}[e_{T+1|T}]$, would account for all three sources of error. In practice, handling the second of these errors is largely intractable, while the first is merely extremely difficult. [See Harvey (1990) and Hamilton (1994, especially Section 11.7) for useful discussion. McCullough (1996) presents results that suggest that "intractable" may be too pessimistic.] For the moment, we will concentrate on the third source and return to the other issues briefly at the end of the section.

Ignoring for the moment the variation in $\hat{\mu}_{T+1|T}$—that is, assuming that the parameters are known and the exogenous variables are forecasted perfectly—the variance of the forecast error will be simply

$$\text{Var}[e_{T+1|T} \,|\, x_{T+1}, w_{T+1}, \mu, \boldsymbol{\beta}, \delta, y_T, \ldots] = \text{Var}[\varepsilon_{T+1}] = \sigma^2,$$

so at least within these assumptions, forming the forecast and computing the forecast variance are straightforward. Also, at this first step, given the data used for the forecast, the first part of the variance is also tractable. Let $\mathbf{z}_{T+1} = [1, x_{T+1}, x_T, \ldots, x_{T-r+1}, w_T, y_T, y_{T-1}, \ldots, y_{T-p+1}]$, and let $\hat{\boldsymbol{\theta}}$ denote the full estimated parameter vector. Then we would use

$$\text{Est. Var}[e_{T+1|T} \,|\, \mathbf{z}_{T+1}] = s^2 + \mathbf{z}'_{T+1}\{\text{Est. Asy. Var}[\hat{\boldsymbol{\theta}}]\}\mathbf{z}_{T+1}.$$

Now, consider forecasting further out beyond the sample period:

$$\hat{y}_{T+2|T} = \hat{\mu}_{T+2|T} + \gamma_1 \hat{y}_{T+1|T} + \cdots + \gamma_p y_{T-p+2} + \hat{\varepsilon}_{T+2|T}.$$

Note that for period $T+1$, the forecasted y_{T+1} is used. Making the substitution for $\hat{y}_{T+1|T}$, we have

$$\hat{y}_{T+2|T} = \hat{\mu}_{T+2|T} + \gamma_1(\hat{\mu}_{T+1|T} + \gamma_1 y_T + \cdots + \gamma_p y_{T-p+1} + \hat{\varepsilon}_{T+1|T}) + \cdots + \gamma_p y_{T-p+2} + \hat{\varepsilon}_{T+2|T},$$

and, likewise, for subsequent periods. Our method will be simplified considerably if we use the device we constructed in the previous section. For the first forecast period, write the forecast with the previous p lagged values as

$$\begin{bmatrix} \hat{y}_{T+1|T} \\ y_T \\ y_{T-1} \\ \vdots \end{bmatrix} = \begin{bmatrix} \hat{\mu}_{T+1|T} \\ 0 \\ 0 \\ \vdots \end{bmatrix} + \begin{bmatrix} \gamma_1 & \gamma_2 & \cdots & \gamma_p \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} y_T \\ y_{T-1} \\ y_{T-2} \\ \vdots \end{bmatrix} + \begin{bmatrix} \hat{\varepsilon}_{T+1|T} \\ 0 \\ 0 \\ \vdots \end{bmatrix}.$$

The coefficient matrix on the right-hand side is $\mathbf{C}$, which we defined in (~~20~~21-18). To maintain the thread of the discussion, we will continue to use the notation $\hat{\mu}_{T+1|T}$ for the forecast of the deterministic part of the model, although for the present, we are

21-19

assuming that this value, as well as $\mathbf{C}$, is known with certainty. With this modification, then, our forecast is the top element of the vector of forecasts,

$$\hat{\mathbf{y}}_{T+1|T} = \hat{\boldsymbol{\mu}}_{T+1|T} + \mathbf{C}\mathbf{y}_T + \hat{\boldsymbol{\varepsilon}}_{T+1|T}.$$

We are assuming that everything on the right-hand side is known except the period $T+1$ disturbance, so the covariance matrix for this $p+1$ vector is

$$E[(\hat{\mathbf{y}}_{T+1|T} - \mathbf{y}_{T+1})(\hat{\mathbf{y}}_{T+1|T} - \mathbf{y}_{T+1})'] = \begin{bmatrix} \sigma^2 & 0 & \cdots \\ 0 & 0 & \vdots \\ \vdots & \cdots & \ddots \end{bmatrix},$$

and the forecast variance for $\hat{y}_{T+1|T}$ is just the upper left element, σ^2.

Now, extend this notation to forecasting out to periods $T+2$, $T+3$, and so on:

$$\begin{aligned}\hat{\mathbf{y}}_{T+2|T} &= \hat{\boldsymbol{\mu}}_{T+2|T} + \mathbf{C}\hat{\mathbf{y}}_{T+1|T} + \hat{\boldsymbol{\varepsilon}}_{T+2|T} \\ &= \hat{\boldsymbol{\mu}}_{T+2|T} + \mathbf{C}\hat{\boldsymbol{\mu}}_{T+1|T} + \mathbf{C}^2\mathbf{y}_T + \hat{\boldsymbol{\varepsilon}}_{T+2|T} + \mathbf{C}\hat{\boldsymbol{\varepsilon}}_{T+1|T}.\end{aligned}$$

Once again, the only unknowns are the disturbances, so the forecast variance for this two-period-ahead forecasted vector is

$$\text{Var}[\hat{\boldsymbol{\varepsilon}}_{T+2|T} + \mathbf{C}\hat{\boldsymbol{\varepsilon}}_{T+1|T}] = \begin{bmatrix} \sigma^2 & 0 & \cdots \\ 0 & 0 & \vdots \\ \vdots & \cdots & \ddots \end{bmatrix} + \mathbf{C}\begin{bmatrix} \sigma^2 & 0 & \cdots \\ 0 & 0 & \vdots \\ \vdots & \cdots & \ddots \end{bmatrix}\mathbf{C}'.$$

Thus, the forecast variance for the two-step-ahead forecast is $\sigma^2[1 + \Psi(1)_{11}]$, where $\Psi(1)_{11}$ is the $(1, 1)$ element of $\Psi(1) = \mathbf{C}\mathbf{j}\mathbf{j}'\mathbf{C}'$, where $\mathbf{j}' = [\sigma, 0, \ldots, 0]$. By extending this device to a forecast F periods beyond the sample period, we obtain

$$\hat{\mathbf{y}}_{T+F|T} = \sum_{f=1}^{F} \mathbf{C}^{f-1}\hat{\boldsymbol{\mu}}_{T+F-(f-1)|T} + \mathbf{C}^F\mathbf{y}_T + \sum_{f=1}^{F} \mathbf{C}^{f-1}\hat{\boldsymbol{\varepsilon}}_{T+F-(f-1)|T}. \qquad \text{(21-20)}$$

This equation shows how to compute the forecasts, which is reasonably simple. We also obtain our expression for the conditional forecast variance,

$$\text{Conditional Var}[\hat{y}_{T+F|T}] = \sigma^2[1 + \Psi(1)_{11} + \Psi(2)_{11} + \cdots + \Psi(F-1)_{11}], \qquad \text{(21-21)}$$

where $\Psi(i) = \mathbf{C}^i\mathbf{j}\mathbf{j}'\mathbf{C}^{i\prime}$.

The general form of the F-period-ahead forecast shows how the forecasts will behave as the forecast period extends further out beyond the sample period. If the equation is stable—that is, if all roots of the matrix $\mathbf{C}$ are less than one in absolute value—then $\mathbf{C}^F$ will converge to zero, and because the forecasted disturbances are zero, the forecast will be dominated by the sum in the first term. If we suppose, in addition, that the forecasts of the exogenous variables are just the period $T+1$ forecasted values and not revised, then, as we found at the end of the previous section, the forecast will ultimately converge to

$$\lim_{F\to\infty} \hat{\mathbf{y}}_{T+F|T} \mid \hat{\boldsymbol{\mu}}_{T+1|T} = [\mathbf{I} - \mathbf{C}]^{-1}\hat{\boldsymbol{\mu}}_{T+1|T}.$$

To account fully for all sources of variation in the forecasts, we would have to revise the forecast variance to include the variation in the forecasts of the exogenous variables and the variation in the parameter estimates. As noted, the first of these is likely to be intractable. For the second, this revision will be extremely difficult, the more so when we also account for the matrix $\mathbf{C}$, as well as the vector $\boldsymbol{\mu}$, being built up from the estimated parameters. The level of difficulty in this case falls from impossible to merely extremely difficult. In principle, what is required is

$$\text{Est. Conditional Var}[\hat{y}_{T+F|T}] = \sigma^2[1 + \Psi(1)_{11} + \Psi(2)_{11} + \cdots + \Psi(F-1)_{11}] \\ + \mathbf{g}'\text{Est. Asy. Var}[\hat{\mu}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}]\mathbf{g},$$

where

$$\mathbf{g} = \frac{\partial \hat{y}_{T+F}}{\partial[\hat{\mu}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}]}.$$

[See Hamilton (1994, Appendix to Chapter 11) for formal derivation.]

One possibility is to use the bootstrap method. For this application, bootstrapping would involve sampling new sets of disturbances from the estimated distribution of ε_t, and then repeatedly rebuilding the within-sample time series of observations on y_t by using

$$\hat{y}_t = \hat{\mu}_t + \gamma_1 y_{t-1} + \cdots + \gamma_p y_{t-p} + e_{bt}(m),$$

where $e_{bt}(m)$ is the estimated "bootstrapped" disturbance in period t during replication m. The process is repeated M times, with new parameter estimates and a new forecast generated in each replication. The variance of these forecasts produces the estimated forecast variance.[6]

20.5 METHODOLOGICAL ISSUES IN THE ANALYSIS OF DYNAMIC MODELS

20.5.1 AN ERROR CORRECTION MODEL

Consider the ARDL(1, 1) model, which has become a workhorse of the modern literature on time-series analysis. By defining the first differences $\Delta y_t = y_t - y_{t-1}$ and $\Delta x_t = x_t - x_{t-1}$ we can rearrange

$$y_t = \mu + \gamma_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t$$

to obtain

$$\Delta y_t = \mu + \beta_0 \Delta x_t + (\gamma_1 - 1)(y_{t-1} - \theta x_{t-1}) + \varepsilon_t, \qquad \textbf{(20-22)}$$

where $\theta = -(\beta_0 + \beta_1)/(\gamma_1 - 1)$. This form of the model is in the **error correction** form. In this form, we have an **equilibrium relationship,** $\Delta y_t = \mu + \beta_0 \Delta x_t + \varepsilon_t$, and the **equilibrium error,** $(\gamma_1 - 1)(y_{t-1} - \theta x_{t-1})$, which account for the deviation of the pair of variables from that equilibrium. The model states that the change in y_t from the previous period consists of the change associated with movement with x_t along the

[6] Bernard and Veall (1987) give an application of this technique. See, also, McCullough (1996).

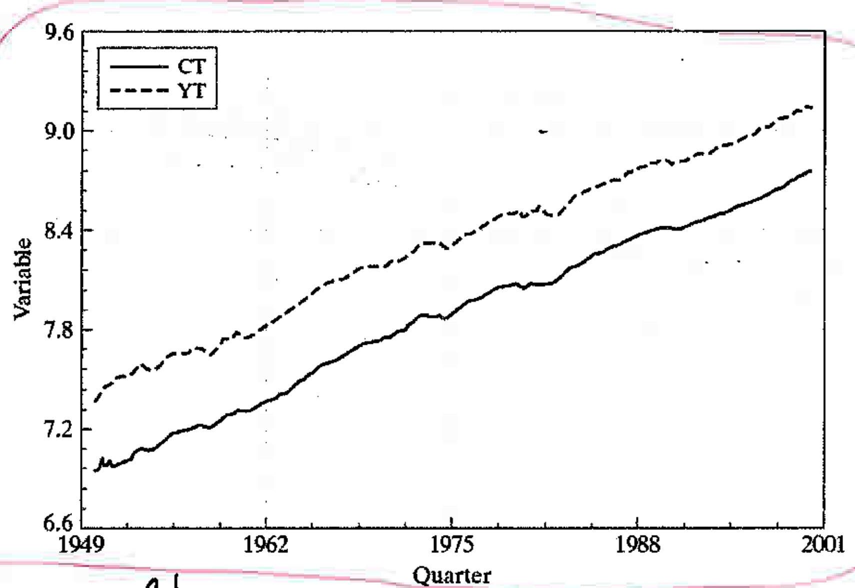

FIGURE 20.3 Consumption and Income Data.

long-run equilibrium path plus a part $(\gamma_1 - 1)$ of the deviation $(y_{t-1} - \theta x_{t-1})$ from the equilibrium. With a model in logs, this relationship would be in proportional terms.

23 It is useful at this juncture to jump ahead a bit—we will return to this topic in some detail in Chapter 21—and explore why the error correction form might be such a useful formulation of this simple model. Consider the logged consumption and income data plotted in Figure 20.3. It is obvious on inspection of the figure that a simple regression of the log of consumption on the log of income would suggest a highly significant relationship; in fact, the simple linear regression produces a slope of 1.0567 with a t ratio of 440.5 (!) and an R^2 of 0.99896. The disturbing result of a line of literature in econometrics that begins with Granger and Newbold (1974) and continues to the present is that this seemingly obvious and powerful relationship might be entirely spurious. Equally obvious from the figure is that both c_t and y_t are trending variables. If, in fact, both variables unconditionally were random walks with drift of the sort that we met at the end of Section 20.4.1—that is, $c_t = t\mu_c + v_t$ and likewise for y_t—then we would almost certainly observe a figure such as 20.3 and compelling regression results such as those, *even if there were no relationship at all*. In addition, there is ample evidence in the recent literature that low-frequency (infrequently observed, aggregated over long periods) flow variables such as consumption and output are, indeed, often well described as random walks. In such data, the ARDL(1, 1) model might appear to be entirely appropriate even if it is not. So, how is one to distinguish between the spurious regression and a genuine relationship as shown in the ARDL(1, 1)? The first difference of consumption produces $\Delta c_t = \mu_c + v_t - v_{t-1}$. If the random walk proposition is indeed correct, then the spurious appearance of regression will not survive the first differencing, whereas if there is a relationship between c_t and y_t, then it will be preserved in the error correction model. We will return to this issue in Chapter 21, when we examine the issue of integration and cointegration of economic variables.

FIG 21.3

21

21

21

23

21 -22

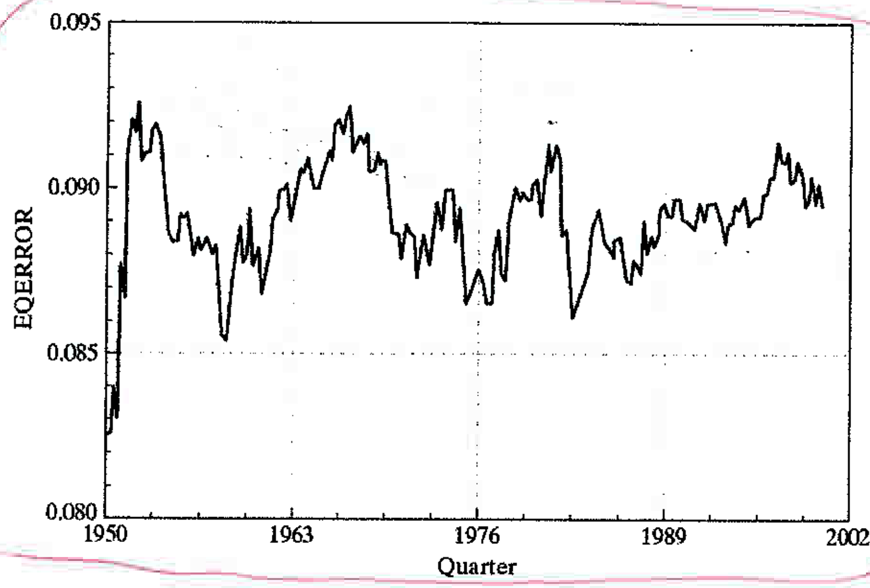

FIGURE 20.4 Consumption–Income Equilibrium Errors.

Example 20.5 An Error Correction Model for Consumption

The error correction model is a nonlinear regression model, although in fact it is intrinsically linear and can be deduced simply from the unrestricted form directly above it. Because the parameter θ is actually of some interest, it might be more convenient to use nonlinear least squares and fit the second form directly. (The model is intrinsically linear, so the nonlinear least squares estimates will be identical to the derived linear least squares estimates.) The logs of consumption and income data in Appendix Table F5.1 are plotted in Figure 20.3. Not surprisingly, the two variables are drifting upward together.

The estimated error correction model, with estimated standard errors in parentheses, is

$$c_t - c_{t-1} = \underset{(0.02899)}{-0.08533} + (\underset{(0.03029)}{0.90458} - 1)[c_{t-1} - \underset{(0.01052)}{1.06034}y_{t-1}] + \underset{(0.05090)}{0.58421}(y_t - y_{t-1}).$$

FIG 21.4

The estimated equilibrium errors are shown in Figure 20.4. Note that they are all positive, but that in each period, the adjustment is in the opposite direction. Thus (according to this model), when consumption is below its equilibrium value, the adjustment is upward, as might be expected.

20.5.2 AUTOCORRELATION

The disturbance in the error correction model is assumed to be nonautocorrelated. As we saw in Chapter 19, autocorrelation in a model can be induced by misspecification. An orthodox view of the modeling process might state, in fact, that this misspecification is the *only* source of autocorrelation. Although admittedly a bit optimistic in its implication, this misspecification does raise an interesting methodological question. Consider once again the simplest model of autocorrelation from Chapter 19 (with a small change in notation to make it consistent with the present discussion),

$$y_t = \beta x_t + v_t, \qquad v_t = \rho v_{t-1} + \varepsilon_t, \tag{20-23}$$

where ε_t is nonautocorrelated. As we found earlier, this model can be written as

$$y_t - \rho y_{t-1} = \beta(x_t - \rho x_{t-1}) + \varepsilon_t, \tag{~~20~~ 21-24}$$

or

$$y_t = \rho y_{t-1} + \beta x_t - \beta\rho x_{t-1} + \varepsilon_t. \tag{~~20~~ 21-25}$$

This model is an ARDL(1, 1) model in which $\beta_1 = -\gamma_1\beta_0$. Thus, we can view (~~20~~ 21-25) as a restricted version of

$$y_t = \gamma_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t. \tag{~~20~~ 21-26}$$

21

The crucial point here is that the (nonlinear) restriction on (~~20~~ 21-26) is testable, so there is no compelling reason to proceed to (20-23) first without establishing that the restriction is in fact consistent with the data. The upshot is that the AR(1) disturbance model, as a general proposition, is a testable restriction on a simpler, linear model, not necessarily a structure unto itself.

Now, let us take this argument to its logical conclusion. The AR(p) disturbance model,

$$v_t = \rho_1 v_{t-1} + \cdots + \rho_p v_{t-p} + \varepsilon_t,$$

or $R(L)v_t = \varepsilon_t$, can be written in its moving average form as

$$v_t = \frac{\varepsilon_t}{R(L)}.$$

[Recall, in the AR(1) model, that $\varepsilon_t = u_t + \rho u_{t-1} + \rho^2 u_{t-2} + \cdots$.] The regression model with this AR(p) disturbance is, therefore,

$$y_t = \beta x_t + \frac{\varepsilon_t}{R(L)}.$$

But consider instead the ARDL(p, p) model

$$C(L)y_t = \beta B(L)x_t + \varepsilon_t.$$

These coefficients are the same model if $B(L) = C(L)$. The implication is that *any model with an AR(p) disturbance can be interpreted as a nonlinearly restricted version of an ARDL(p, p) model.*

The preceding discussion is a rather orthodox view of autocorrelation. It is predicated on the AR(p) model. Researchers have found that a more involved model for the process generating ε_t is sometimes called for. If the time-series structure of ε_t is not autoregressive, much of the preceding analysis will become intractable. As such, there remains room for disagreement with the strong conclusions. We will turn to models whose disturbances are mixtures of autoregressive and moving-average terms, which would be beyond the reach of this apparatus, in Chapter ~~21~~ 22.

21 ~~20~~.5.3 SPECIFICATION ANALYSIS

The usual explanation of autocorrelation is serial correlation in omitted variables. The preceding discussion and our results in Chapter ~~19~~ 20 suggest another candidate: misspecification of what would otherwise be an unrestricted ARDL model. Thus, upon finding evidence of autocorrelation on the basis of a Durbin–Watson statistic or an LM statistic,

21-24

we might find that relaxing the nonlinear restrictions on the ARDL model is a preferable next step to "correcting" for the autocorrelation by imposing the restrictions and refitting the model by FGLS. Because an ARDL(p, r) model with AR disturbances, even with $p = 0$, is implicitly an ARDL$(p+d, r+d)$ model, where d is usually one, the approach suggested is just to add additional lags of the dependent variable to the model. Thus, one might even ask why we would ever use the familiar FGLS procedures. [See, e.g., Mizon (1995).] The payoff is that the restrictions imposed by the FGLS procedure produce a more efficient estimator than other methods. If the restrictions are in fact appropriate, then not imposing them amounts to not using information.

A related question now arises, apart from the issue of autocorrelation. In the context of the ARDL model, how should one do the specification search? (This question is not specific to the ARDL or even to the time-series setting.) Is it better to start with a small model and expand it until conventional fit measures indicate that additional variables are no longer improving the model, or is it better to start with a large model and pare away variables that conventional statistics suggest are superfluous? The first strategy, going from a *simple model to a general model*, is likely to be problematic, because the statistics computed for the narrower model are biased and inconsistent if the hypothesis is incorrect. Consider, for example, an LM test for autocorrelation in a model from which important variables have been omitted. The results are biased in favor of a finding of autocorrelation. The alternative approach is to proceed from a *general model to a simple one*. Thus, one might overfit the model and then subject it to whatever battery of tests are appropriate to produce the correct specification at the end of the procedure. In this instance, the estimates and test statistics computed from the overfit model, although inefficient, are not generally systematically biased. (We have encountered this issue at several points.)

The latter approach is common in modern analysis, but some words of caution are needed. The procedure routinely leads to overfitting the model. A typical time-series analysis might involve specifying a model with deep lags on all the variables and then paring away the model as conventional statistics indicate. The danger is that the resulting model might have an autoregressive structure with peculiar holes in it that would be hard to justify with any theory. Thus, a model for quarterly data that includes lags of 2, 3, 6, and 9 on the dependent variable would look suspiciously like the end result of a computer-driven fishing trip and, moreover, might not survive even moderate changes in the estimation sample. [As Hendry (1995) notes, a model in which the largest and most significant lag coefficient occurs at the last lag is surely misspecified.]

21 ~~20~~.6 VECTOR AUTOREGRESSIONS

The preceding discussions can be extended to sets of variables. The resulting autoregressive model is

$$\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\Gamma}_1 \mathbf{y}_{t-1} + \cdots + \boldsymbol{\Gamma}_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \tag{21 ~~20~~-27}$$

where $\boldsymbol{\varepsilon}_t$ is a vector of nonautocorrelated disturbances (innovations) with zero means and contemporaneous covariance matrix $E[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'] = \boldsymbol{\Omega}$. This equation system is a vector

autoregression, or VAR. Equation (20-27) may also be written as

$$\Gamma(L)\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t$$

where $\Gamma(L)$ is a matrix of polynomials in the lag operator. The individual equations are

$$y_{mt} = \mu_m + \sum_{j=1}^{p}(\Gamma_j)_{m1}y_{1,t-j} + \sum_{j=1}^{p}(\Gamma_j)_{m2}y_{2,t-j} + \cdots + \sum_{j=1}^{p}(\Gamma_j)_{mM}y_{M,t-j} + \varepsilon_{mt},$$

where $(\Gamma_j)_{ml}$ indicates the (m, l) element of Γ_j.

VARs have been used primarily in macroeconomics. Early in their development, it was argued by some authors [e.g., Sims (1980), Litterman (1979, 1986)] that VARs would forecast better than the sort of structural equation models discussed in Chapter 13. One could argue that as long as μ includes the current observations on the (truly) relevant exogenous variables, the VAR is simply an overfit reduced form of some simultaneous equations model. [See Hamilton (1994, pp. 326-327).] The overfitting results from the possible inclusion of more lags than would be appropriate in the original model. (See Example 20.7 for a detailed discussion of one such model.) On the other hand, one of the virtues of the VAR is that it obviates a decision as to what contemporaneous variables are exogenous; it has only lagged (predetermined) variables on the right-hand side, and all variables are endogenous.

The motivation behind VARs in macroeconomics runs deeper than the statistical issues.[7] The large structural equations models of the 1950s and 1960s were built on a theoretical foundation that has not proved satisfactory. That the forecasting performance of VARs surpassed that of large structural models-some of the later counterparts to Klein's Model I ran to hundreds of equations-signaled to researchers a more fundamental problem with the underlying methodology. The Keynesian style systems of equations describe a structural model of decisions (consumption, investment) that seem loosely to mimic individual behavior; see Keynes's formulation of the consumption function in Example 1.1 that is, perhaps, the canonical example. In the end, however, these decision rules are fundamentally ad hoc, and there is little basis on which to assume that they would aggregate to the macroeconomic level anyway. On a more practical level, the high inflation and high unemployment experienced in the 1970s were very badly predicted by the Keynesian paradigm. From the point of view of the underlying paradigm, the most troubling criticism of the structural modeling approach comes in the form of "the Lucas critique" (1976), in which the author argued that the *parameters* of the "decision rules" embodied in the systems of structural equations would not remain stable when economic policies changed, even if the rules themselves were appropriate. Thus, the paradigm underlying the systems of equations approach to macroeconomic modeling is arguably fundamentally flawed. More recent research has reformulated the basic equations of macroeconomic models in terms of a microeconomic optimization foundation and has, at the same time, been much less ambitious in specifying the interrelationships among economic variables.

The preceding arguments have drawn researchers to less structured equation systems for forecasting. Thus, it is not just the form of the equations that has changed. The

[7] An extremely readable, nontechnical discussion of the paradigm shift in macroeconomic forecasting is given in Diebold (2003). See also Stock and Watson (2001).

21-26

variables in the equations have changed as well; the VAR is not just the reduced form of some structural model. For purposes of analyzing and forecasting macroeconomic activity and tracing the effects of policy changes and external stimuli on the economy, researchers have found that simple, small-scale VARs without a possibly flawed theoretical foundation have proved as good as or better than large-scale structural equation systems. In addition to forecasting, VARs have been used for two primary functions: testing Granger causality and studying the effects of policy through impulse response characteristics.

21 ~~20~~.6.1 MODEL FORMS

To simplify things for the present, we note that the pth order VAR can be written as a first-order VAR as follows:

$$\begin{pmatrix} \mathbf{y}_t \\ \mathbf{y}_{t-1} \\ \cdots \\ \mathbf{y}_{t-p+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu} \\ \mathbf{0} \\ \cdots \\ \mathbf{0} \end{pmatrix} + \begin{bmatrix} \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_2 & \cdots & \boldsymbol{\Gamma}_p \\ \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{pmatrix} \mathbf{y}_{t-1} \\ \mathbf{y}_{t-2} \\ \cdots \\ \mathbf{y}_{t-p} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_t \\ \mathbf{0} \\ \cdots \\ \mathbf{0} \end{pmatrix}.$$

[See, e.g., (21 ~~20~~-18).] This means that we do not lose any generality in casting the treatment in terms of a first-order model

$$\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\Gamma}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t.$$

In Example 15.10, we examined Dahlberg and Johansson's model for municipal finances in Sweden, in which $\mathbf{y}_t = [\Delta S_t, \Delta R_t, \Delta G_t]'$, where S_t is spending, R_t is receipts, G_t is grants from the central government, and $p = 3$. We will continue that application in Example 20.7.

Au: Confirm x-ref to Example 20.7 is OK

In principle, the VAR model is a seemingly unrelated regressions model—indeed, a particularly simple one because each equation has the same set of regressors. This is the traditional form of the model as originally proposed, for example, by Sims (1980). The VAR may also be viewed as the reduced form of a simultaneous equations model; the corresponding structure would then be

$$\boldsymbol{\Theta}\mathbf{y}_t = \boldsymbol{\alpha} + \boldsymbol{\Psi}\mathbf{y}_{t-1} + \boldsymbol{\omega}_t,$$

where $\boldsymbol{\Theta}$ is a nonsingular matrix and $\text{Var}[\boldsymbol{\omega}_t] = \boldsymbol{\Sigma}$. In one of Cecchetti and Rich's (2001) formulations, for example, $\mathbf{y}_t = [\Delta y_t, \Delta\pi_t]'$ where y_t is the log of aggregate real output, π_t is the inflation rate from time $t-1$ to time t, $\boldsymbol{\Theta} = \begin{bmatrix} 1 & -\theta_{12} \\ -\theta_{21} & 1 \end{bmatrix}$, and $p = 8$. (We will examine their model in Section 21 ~~20~~.6.8.) In this form, we have a conventional simultaneous equations model, which we analyzed in detail in Chapter 13. As we saw, for such a model to be identified—that is, estimable—certain restrictions must be placed on the structural coefficients. The reason for this is that ultimately, only the original VAR form, now the reduced form, is estimated from the data; the structural parameters must be deduced from these coefficients. In this model, to deduce these structural parameters, they must be extracted from the reduced form parameters, $\boldsymbol{\Gamma} = \boldsymbol{\Theta}^{-1}\boldsymbol{\Psi}$, $\boldsymbol{\mu} = \boldsymbol{\Theta}^{-1}\boldsymbol{\alpha}$, and $\boldsymbol{\Omega} = \boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Theta}^{-1\prime}$. ~~We analyzed this issue in detail in Section 13.3. The results would be the same here.~~ In Cecchetti and Rich's application, certain restrictions were placed on the lag coefficients in order to secure identification.

Au: Confirm x-ref to Chap. 13 is OK

21-27

20.6.2 ESTIMATION

21

Au: Confirm x-ref to Sec 10.2.2 is OK

In the form of (20-27)—that is, without autocorrelation of the disturbances—VARs are particularly simple to estimate. Although the equation system can be exceedingly large, it is, in fact, a seemingly unrelated regressions model with identical regressors. As such, the equations should be estimated separately by ordinary least squares. (See Section 10.2.2 for discussion of SUR systems with identical regressors.) The disturbance covariance matrix can then be estimated with average sums of squares or cross-products of the least squares residuals. If the disturbances are normally distributed, then these least squares estimators are also maximum likelihood. If not, then OLS remains an efficient GMM estimator. The extension to instrumental variables and GMM is a bit more complicated, as the model now contains multiple equations (see Section 15.6.3), 13 but since the equations are all linear, the necessary extensions are at least relatively straightforward. GMM estimation of the VAR system is a special case of the model discussed in Section 18.6.3. (We will examine an application in Example 20.7.) 21

13

The proliferation of parameters in VARs has been cited as a major disadvantage of their use. Consider, for example, a VAR involving five variables and three lags. Each Γ has 25 unconstrained elements, and there are three of them, for a total of 75 free parameters, plus any others in μ, plus $5(6)/2 = 15$ free parameters in Ω. On the other hand, each single equation has only 25 parameters, and at least given sufficient degrees of freedom—there's the rub—a linear regression with 25 parameters is simple work. Moreover, applications rarely involve even as many as four variables, so the model-size issue may well be exaggerated.

21 20.6.3 TESTING PROCEDURES

Formal testing in the VAR setting usually centers either on determining the appropriate lag length (a specification search) or on whether certain blocks of zeros in the coefficient matrices are zero (a simple linear restriction on the collection of slope parameters). Both types of hypotheses may be treated as sets of linear restrictions on the elements in $\gamma = \text{vec}[\mu, \Gamma_1, \Gamma_2, \ldots, \Gamma_p]$.

We begin by assuming that the disturbances have a joint normal distribution. Let $\mathbf{W}$ be the $M \times M$ residual covariance matrix based on a restricted model, and let $\mathbf{W}^*$ be its counterpart when the model is unrestricted. Then the likelihood ratio statistic,

$$\lambda = T(\ln|\mathbf{W}| - \ln|\mathbf{W}^*|),$$

can be used to test the hypothesis. The statistic would have a limiting chi-squared distribution with degrees of freedom equal to the number of restrictions. In principle, one might base a specification search for the right lag length on this calculation. The procedure would be to test down from, say, lag q to lag p. The *general-to-simple* principle discussed in Section 20.5.3 would be to set the maximum lag length and test down from it until deletion of the last set of lags leads to a significant loss of fit. At each step at which the alternative lag model has excess terms, the estimators of the superfluous coefficient matrices would have probability limits of zero and the likelihood function would (again, asymptotically) resemble that of the model with the correct number of lags. Formally, suppose the appropriate lag length is p but the model is fit with $q \geq p + 1$ lagged terms.

21-28

Then, under the null hypothesis,

$$\lambda_q = T[\ln|\mathbf{W}(\boldsymbol{\mu}, \boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_{q-1})| - \ln|\mathbf{W}^*(\boldsymbol{\mu}, \boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_q)|] \xrightarrow{d} \chi^2[M^2].$$

The same approach would be used to test other restrictions. Thus, the Granger causality test noted in Section 20.6.5 would fit the model with and without certain blocks of zeros in the coefficient matrices, then refer the value of λ once again to the chi-squared distribution.

21

For specification searches for the right lag, the suggested procedure may be less effective than one based on the information criteria suggested for other linear models (see Section 7.4). Lutkepohl (2005, pp. 128–135) suggests an alternative approach based on the minimizing functions of the information criteria we have considered earlier;

5.10.1

$$\lambda^* = \ln(|\mathbf{W}|) + (pM^2 + M)\text{IC}(T)/T,$$

where T is the sample size, p is the number of lags, M is the number of equations, and $\text{IC}(T) = 2$ for the Akaike information criterion and $\ln T$ for the Schwarz (Bayesian) information criterion. We should note that this is not a test statistic; it is a diagnostic tool that we are using to conduct a specification search. Also, as in all such cases, the testing procedure should be from a larger model to a smaller one to avoid the misspecification problems induced by a lag length that is smaller than the appropriate one.

The preceding has relied heavily on the normality assumption. Because most recent applications of these techniques have either treated the least squares estimators as robust (distribution-free) estimators, or used GMM (as we did in Chapter 15), it is necessary to consider a different approach that does not depend on normality. An alternative approach that should be robust to variations in the underlying distributions is the Wald statistic. [See Lutkepohl (2005, pp. 93–95).] The full set of coefficients in the model may be arrayed in a single coefficient vector, $\boldsymbol{\gamma}$. Let $\mathbf{c}$ be the sample estimator of $\boldsymbol{\gamma}$ and let $\mathbf{V}$ denote the estimated asymptotic covariance matrix. Then, the hypothesis in question (lag length, or other linear restriction) can be cast in the form $\mathbf{R}\boldsymbol{\gamma} - \mathbf{q} = \mathbf{0}$. The Wald statistic for testing the null hypothesis is 13

$$W = (\mathbf{Rc} - \mathbf{q})'[\mathbf{RVR}']^{-1}(\mathbf{Rc} - \mathbf{q}).$$

Under the null hypothesis, this statistic has a limiting chi-squared distribution with degrees of freedom equal to J, the number of restrictions (rows in $\mathbf{R}$). For the specification search for the appropriate lag length (or the Granger causality test discussed in the next section), the null hypothesis will be that a certain subvector of $\boldsymbol{\gamma}$, say $\boldsymbol{\gamma}_0$, equals zero. In this case, the statistic will be

$$W_0 = \mathbf{c}_0'\mathbf{V}_{00}^{-1}\mathbf{c}_0,$$

where $\mathbf{V}_{00}$ denotes the corresponding submatrix of $\mathbf{V}$.

Because time-series data sets are often only moderately long, use of the limiting distribution for the test statistic may be a bit optimistic. Also, the Wald statistic does not account for the fact that the asymptotic covariance matrix is estimated using a finite sample. In our analysis of the classical linear regression model, we accommodated these considerations by using the F distribution instead of the limiting chi-squared. (See Section 5.4.) The adjustment made was to refer W/J to the $F[J, T - K]$ distribution. This produces a more conservative test—the corresponding critical values of JF converge

5.5

21-29

~~of~~ to those of the chi-squared *from above*. A remaining complication is to decide what degrees of freedom to use for the denominator. It might seem natural to use MT minus the number of parameters, which would be correct if the restrictions are imposed on all equations simultaneously, because there are that many "observations." In testing for causality, as in Section ~~20~~21.6.5 below, Lutkepohl (2005, p. 95) argues that MT is excessive, because the restrictions are not imposed on all equations. When the causality test involves testing for zero restrictions within a single equation, the appropriate degrees of freedom would be $T - Mp - 1$ for that one equation.

~~20~~21.6.4 EXOGENEITY

In the classical regression model with nonstochastic regressors, there is no ambiguity about which is the independent or conditioning or "exogenous" variable in the model

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t. \qquad (\cancel{20}21\text{-}28)$$

This is the kind of characterization that might apply in an experimental situation in which the analyst is choosing the values of x_t. But, the case of nonstochastic regressors has little to do with the sort of modeling that will be of interest in this and the next chapter. There is no basis for the narrow assumption of nonstochastic regressors, and, in fact, in most of the analysis that we have done to this point, we have left this assumption far behind. With stochastic regressor(s), the regression relationship such as the preceding one becomes a conditional mean in a bivariate distribution. In this more realistic setting, what constitutes an "exogenous" variable becomes ambiguous. Assuming that the regression relationship is linear, (~~20~~21-28) can be written (trivially) as

$$y_t = E[y_t \mid x_t] + (y_t - E[y_t \mid x_t]),$$

where the familiar moment condition $E[x_t\varepsilon_t] = 0$ follows by construction. But, this form of the model is no more the "correct" equation than would be

$$x_t = \delta_1 + \delta_2 y_t + \omega_t,$$

which is (we assume)

$$x_t = E[x_t \mid y_t] + (x_t - E[x_t \mid y_t]),$$

and now, $E[y_t\omega_t] = 0$. Both equations are correctly specified in the context of the bivariate distribution, so there is nothing to define one variable or the other as "exogenous." This might seem puzzling, but it is, in fact, at the heart of the matter when one considers modeling in a world in which variables are jointly determined. The definition of exogeneity depends on the analyst's understanding of the world they are modeling, and, in the final analysis, on the purpose to which the model is to be put.

The methodological platform on which this discussion rests is the classic paper by Engle, Hendry, and Richard (1983), where they point out that exogeneity is not an absolute concept at all; it is defined in the context of the model. The central idea, which will be very useful to us here, is that we define a variable (set of variables) as exogenous *in the context of our model* if the joint density may be written

$$f(y_t, x_t) = f(y_t \mid \boldsymbol{\beta}, x_t) \times f(x_t \mid \boldsymbol{\theta})$$

where the parameters in the conditional distribution do not appear in and are functionally unrelated to those in the marginal distribution of x_t. By this arrangement, we

21-30

can think of "autonomous variation" of the parameters of interest, β. The parameters in the conditional model for $y_t \mid x_t$ can be analyzed as if they could vary independently of those in the marginal distribution of x_t. If this condition does not hold, then we cannot think of variation of those parameters without linking that variation to some effect in the marginal distribution of x_t. In this case, it makes little sense to think of x_t as somehow being determined "outside" the (conditional) model. (We considered this issue in Section 13.8 in the context of a simultaneous equations model.)

AU: Confirm x-ref to Sec 13.8 is OK

A second form of exogeneity we will consider is **strong exogeneity**, which is sometimes called **Granger noncausality**. Granger noncausality can be superficially defined by the assumption

AU: KT "Granger noncausality" is not in Chap. list

$$E[y_t \mid y_{t-1}, x_{t-1}, x_{t-2}, \ldots] = E[y_t \mid y_{t-1}].$$

That is, lagged values of x_t do not provide information about the conditional mean of y_t once lagged values of y_t, itself, are accounted for. We will consider this issue at the end of this chapter. For the present, we note that most of the models we will examine will explicitly fail this assumption.

To put this back in the context of our model, we will be assuming that in the model

$$y_t = \beta_1 + \beta_2 x_t + \beta_3 x_{t-1} + \gamma y_{t-1} + \varepsilon_t,$$

and the extensions that we will consider, x_t is weakly exogenous—we can meaningfully estimate the parameters of the regression equation independently of the marginal distribution of x_t, but we will allow for Granger causality between x_t and y_t, thus generally not assuming strong exogeneity.

~~20~~ 21.6.5 TESTING FOR GRANGER CAUSALITY

Causality in the sense defined by Granger (1969) and Sims (1972) is inferred when lagged values of a variable, say, x_t, have explanatory power in a regression of a variable y_t on lagged values of y_t and x_t. ~~(See Section 12.2.2)~~ The VAR can be used to test the hypothesis.[8] Tests of the restrictions can be based on simple F tests in the single equations of the VAR model. That the unrestricted equations have identical regressors means that these tests can be based on the results of simple OLS estimates. The notion can be extended in a system of equations to attempt to ascertain if a given variable is weakly exogenous to the system. If lagged values of a variable x_t have no explanatory power for *any* of the variables in a system, then we would view x_t as weakly exogenous to the system. Once again, this specification can be tested with a likelihood ratio test as described later—the restriction will be to put "holes" in one or more Γ matrices—or with a form of F test constructed by stacking the equations.

FN 8

FN 9

Example ~~20~~ 21.6 Granger Causality[9]

All but one of the major recessions in the U.S. economy since World War II have been preceded by large increases in the price of crude oil. Does movement of the price of oil *cause* movements in U.S. GDP in the Granger sense? Let $\mathbf{y}_t = [\text{GDP, crude oil price}]_t'$. Then,

[8] See Geweke, Meese, and Dent (1983), Sims (1980), and Stock and Watson (2001).

[9] This example is adapted from Hamilton (1994, pp. 307–308).

21-31

a simple VAR would be

$$\mathbf{y}_t = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \alpha_1 & \alpha_2 \\ \beta_1 & \beta_2 \end{bmatrix} \mathbf{y}_{t-1} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}.$$

To assert a causal relationship between oil prices and GDP, we must find that α_2 is not zero; previous movements in oil prices do help explain movements in GDP even in the presence of the lagged value of GDP. Consistent with our earlier discussion, this fact, in itself, is not sufficient to assert a causal relationship. We would also have to demonstrate that there were no other intervening explanations that would explain movements in oil prices *and* GDP. (We will examine a more extensive application in Example ~~20~~ 21.7.)

14

To establish the general result, it will prove useful to write the VAR in the multivariate regression format we used in Section ~~16~~.9.3.b. Partition the two data vectors $\mathbf{y}_t$ and $\mathbf{x}_t$ into $[\mathbf{y}_{1t}, \mathbf{y}_{2t}]$ and $[\mathbf{x}_{1t}, \mathbf{x}_{2t}]$. Consistent with our earlier discussion, $\mathbf{x}_1$ is lagged values of $\mathbf{y}_1$ and $\mathbf{x}_2$ is lagged values of $\mathbf{y}_2$. The VAR with this partitioning would be

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Gamma}_{11} & \boldsymbol{\Gamma}_{12} \\ \boldsymbol{\Gamma}_{21} & \boldsymbol{\Gamma}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{bmatrix}, \qquad \text{Var}\begin{bmatrix} \boldsymbol{\varepsilon}_{1t} \\ \boldsymbol{\varepsilon}_{2t} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}.$$

We would still obtain the unrestricted maximum likelihood estimates by least squares regressions. For testing Granger causality, the hypothesis $\boldsymbol{\Gamma}_{12} = \mathbf{0}$ is of interest. (See Example ~~20~~.6.) For testing the hypothesis of interest, $\boldsymbol{\Gamma}_{12} = \mathbf{0}$, the second set of equations is irrelevant. For testing for Granger causality in the VAR model, only the restricted equations are relevant. The hypothesis can be tested using the likelihood ratio statistic. For the present application, testing means computing

21

$\mathbf{S}_{11}$ = residual covariance matrix when current values of $\mathbf{y}_1$ are regressed on values of both $\mathbf{x}_1$ and $\mathbf{x}_2$,

$\mathbf{S}_{11}(0)$ = residual covariance matrix when current values of $\mathbf{y}_1$ are regressed only on values of $\mathbf{x}_1$.

The likelihood ratio statistic is then

$$\lambda = T(\ln|\mathbf{S}_{11}(0)| - \ln|\mathbf{S}_{11}|).$$

Au: Confirm x-ref to Example 20.7 is OK

The number of degrees of freedom is the number of zero restrictions.

The fact that this test is wedded to the normal distribution limits its generality. The Wald test or its transformation to an approximate F statistic as described in Section ~~20~~.6.3 is an alternative that should be more generally applicable. When the equation system is fit by GMM, as in Example 20.7, the simplicity of the likelihood ratio test is lost. The Wald statistic remains usable, however. Another possibility is to use the GMM counterpart to the likelihood ratio statistic (see Section ~~15~~.5.2) based on the GMM criterion functions. This is just the difference in the GMM criteria. Fitting both restricted and unrestricted models in this framework may be burdensome, but having set up the GMM estimator for the (larger) unrestricted model, imposing the zero restrictions of the smaller model should require only a minor modification.

21 13 22 22

There is a complication in these causality tests. The VAR can be motivated by the Wold representation theorem (see Section ~~21~~.2.5, Theorem ~~21~~.1), although with assumed nonautocorrelated disturbances, the motivation is incomplete. On the other hand, there is no formal theory behind the formulation. As such, the causality tests are predicated on a model that may, in fact, be missing either intervening variables or

21-32

additional lagged effects that should be present but are not. For the first of these, the problem is that a finding of causal effects might equally well result from the omission of a variable that is correlated with both (or all) of the left-hand-side variables.

~~20~~21.6.6 IMPULSE RESPONSE FUNCTIONS

Any VAR can be written as a first-order model by augmenting it, if necessary, with additional identity equations. For example, the model

$$\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\Gamma}_1\mathbf{y}_{t-1} + \boldsymbol{\Gamma}_2\mathbf{y}_{t-2} + \mathbf{v}_t$$

can be written

$$\begin{bmatrix}\mathbf{y}_t\\ \mathbf{y}_{t-1}\end{bmatrix} = \begin{bmatrix}\boldsymbol{\mu}\\ \mathbf{0}\end{bmatrix} + \begin{bmatrix}\boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_2\\ \mathbf{I} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{y}_{t-1}\\ \mathbf{y}_{t-2}\end{bmatrix} + \begin{bmatrix}\mathbf{v}_t\\ \mathbf{0}\end{bmatrix},$$

which is a first-order model. We can study the dynamic characteristics of the model in either form, but the second is more convenient, as will soon be apparent.

As we analyzed earlier, in the model

$$\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\Gamma}\mathbf{y}_{t-1} + \mathbf{v}_t,$$

dynamic stability is achieved if the characteristic roots of $\boldsymbol{\Gamma}$ have modulus less than one. (The roots may be complex, because $\boldsymbol{\Gamma}$ need not be symmetric. See Section ~~20~~21.4.3.) ~~for the case of a single equation and Section 13.9 for analysis of essentially this model in a simultaneous equations context.)~~

Assuming that the equation system is stable, the equilibrium is found by obtaining the final form of the system. We can do this step by repeated substitution, or more simply by using the lag operator to write

$$\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\Gamma}(L)\mathbf{y}_t + \mathbf{v}_t,$$

or

$$[\mathbf{I} - \boldsymbol{\Gamma}(L)]\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{v}_t.$$

With the stability condition, we have

$$\begin{aligned}\mathbf{y}_t &= [\mathbf{I} - \boldsymbol{\Gamma}(L)]^{-1}(\boldsymbol{\mu} + \mathbf{v}_t)\\ &= (\mathbf{I} - \boldsymbol{\Gamma})^{-1}\boldsymbol{\mu} + \sum_{i=0}^{\infty}\boldsymbol{\Gamma}^i\mathbf{v}_{t-i}\\ &= \bar{\mathbf{y}} + \sum_{i=0}^{\infty}\boldsymbol{\Gamma}^i\mathbf{v}_{t-i}\\ &= \bar{\mathbf{y}} + \mathbf{v}_t + \boldsymbol{\Gamma}\mathbf{v}_{t-1} + \boldsymbol{\Gamma}^2\mathbf{v}_{t-2} + \cdots.\end{aligned} \qquad (\text{~~20~~21-29})$$

The coefficients in the powers of $\boldsymbol{\Gamma}$ are the multipliers in the system. ~~In fact, by renaming things slightly, this set of results is precisely the one we examined in Section 13.9 in our discussion of dynamic simultaneous equations models. We will change the interpretation slightly here, however. As we did in Section 13.9, we~~ We consider the conceptual experiment of disturbing a system in equilibrium. Suppose that $\mathbf{v}$ has equaled $\mathbf{0}$ for long enough that $\mathbf{y}$ has reached equilibrium, $\bar{\mathbf{y}}$. Now we consider injecting a shock to the system by changing one of the v's, for one period, and then returning it to zero thereafter.

21-33

As we saw earlier, y_{mt} will move away from, then return to, its equilibrium. The path whereby the variables return to the equilibrium is called the **impulse response** of the VAR.[10]

In the autoregressive form of the model, we can identify each **innovation,** v_{mt}, with a particular variable in $\mathbf{y}_t$, say, y_{mt}. Consider then the effect of a one-time shock to the system, dv_{mt}. As compared with the equilibrium, we will have, in the current period,

$$y_{mt} - \bar{y}_m = dv_{mt} = \phi_{mm}(0)dv_t.$$

One period later, we will have

$$y_{m,t+1} - \bar{y}_m = (\Gamma)_{mm}dv_{mt} = \phi_{mm}(1)dv_t.$$

Two periods later,

$$y_{m,t+2} - \bar{y}_m = (\Gamma^2)_{mm}dv_{mt} = \phi_{mm}(2)dv_t,$$

and so on. The function, $\phi_{mm}(i)$ gives the impulse response characteristics of variable y_m to innovations in v_m. A useful way to characterize the system is to plot the impulse response functions. The preceding traces through the effect on variable m of a one-time innovation in v_m. We could also examine the effect of a one-time innovation of v_l on variable m. The impulse response function would be

$$\phi_{ml}(i) = \text{element } (m, l) \text{ in } \Gamma^i.$$

Point estimation of $\phi_{ml}(i)$ using the estimated model parameters is straightforward. Confidence intervals present a more difficult problem because the estimated functions $\hat{\phi}_{ml}(i, \hat{\beta})$ are so highly nonlinear in the original parameter estimates. The delta method has thus proved unsatisfactory. Killian (1998) presents results that suggest that bootstrapping may be the more productive approach to statistical inference regarding impulse response functions.

21 ~~20~~.6.7 STRUCTURAL VARs

The VAR approach to modeling dynamic behavior of economic variables has provided some interesting insights and appears [see Litterman (1986)] to bring some real benefits for forecasting. The method has received some strident criticism for its atheoretical approach, however. The "unrestricted" nature of the lag structure in (~~20~~-27) could be synonymous with "unstructured." With no theoretical input to the model, it is difficult to claim that its output provides much of a theoretically justified result. For example, how are we to interpret the impulse response functions derived in the previous section? What lies behind much of this discussion is the idea that there is, in fact, a structure underlying the model, and the VAR that we have specified is a mere hodgepodge of all its components. Of course, that is exactly what reduced forms are. As such, to respond to this sort of criticism, analysts have begun to cast VARs formally as reduced forms and thereby attempt to deduce the structure that they had in mind all along. 21

A VAR model $\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\Gamma}\mathbf{y}_{t-1} + \mathbf{v}_t$ could, in principle, be viewed as the reduced form of the dynamic **structural model**

$$\boldsymbol{\Theta}\mathbf{y}_t = \boldsymbol{\alpha} + \boldsymbol{\Phi}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t,$$

[10] See Hamilton (1994, pp. 318–323 and 336–350) for discussion and a number of related results.

21-34

where we have embedded any exogenous variables x_t in the vector of constants $\boldsymbol{\alpha}$. Thus, $\boldsymbol{\Gamma} = \boldsymbol{\Theta}^{-1}\boldsymbol{\Phi}$, $\boldsymbol{\mu} = \boldsymbol{\Theta}^{-1}\boldsymbol{\alpha}$, $\mathbf{v} = \boldsymbol{\Theta}^{-1}\boldsymbol{\varepsilon}$, and $\boldsymbol{\Omega} = \boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}(\boldsymbol{\Theta}^{-1})'$. Perhaps it is the structure, specified by an underlying theory, that is of interest. For example, we can discuss the impulse response characteristics of this system. For particular configurations of $\boldsymbol{\Theta}$, such as a triangular matrix, we can meaningfully interpret innovations, $\boldsymbol{\varepsilon}$. As we explored at great length in the previous chapter, however, as this model stands, there is not sufficient information contained in the reduced form as just stated to deduce the structural parameters. A possibly large number of restrictions must be imposed on $\boldsymbol{\Theta}$, $\boldsymbol{\Phi}$, and $\boldsymbol{\Sigma}$ to enable us to deduce structural forms from reduced-form estimates, which are always obtainable. The recent work on **structural VARs** centers on the types of restrictions and forms of the theory that can be brought to bear to allow this analysis to proceed. See, for example, the survey in Hamilton (1994, Chapter 11). At this point, the literature on this subject has come full circle because the contemporary development of "unstructured VARs" becomes very much the analysis of quite conventional dynamic structural simultaneous equations models. Indeed, current research [e.g., Diebold (1998)] brings the literature back into line with the structural modeling tradition by demonstrating how VARs can be derived formally as the reduced forms of dynamic structural models. That is, the most recent applications have begun with structures and derived the reduced forms as VARs, rather than departing from the VAR as a reduced form and attempting to deduce a structure from it by layering on restrictions.

21 ~~20~~.6.8 APPLICATION: POLICY ANALYSIS WITH A VAR

Cecchetti and Rich (2001) used a structural VAR to analyze the effect of recent disinflationary policies of the Fed on aggregate output in the U.S. economy. The Fed's policy of the last two decades has leaned more toward controlling inflation and less toward stimulation of the economy. The authors argue that the long-run benefits of this policy include economic stability and increased long-term trend output growth. But, there is a short-term cost in lost output. Their study seeks to estimate the "sacrifice ratio," which is a measure of the cumulative cost of this policy. The specific indicator they study measures the cumulative output loss after τ periods of a policy shock at time t, where the (persistent) shock is measured as the change in the level of inflation.

21 ~~20~~.6.8.a A VAR Model for the Macroeconomic Variables

The model proposed for estimating the ratio is a structural VAR,

$$\Delta y_t = \sum_{i=1}^{p} b_{11}^i \Delta y_{t-i} + b_{12}^0 \Delta\pi_t + \sum_{i=1}^{p} b_{12}^i \Delta\pi_{t-i} + \varepsilon_t^y,$$

$$\Delta\pi_t = b_{21}^0 \Delta y_t + \sum_{i=1}^{p} b_{21}^i \Delta y_{t-i} + \sum_{i=1}^{p} b_{22}^i \Delta\pi_{t-i} + \varepsilon_t^\pi,$$

where y_t is aggregate real output in period t and π_t is the rate of inflation from period $t-1$ to t and the model is cast in terms of rates of changes of these two variables. (Note, therefore, that sums of $\Delta\pi_t$ measure accumulated changes in the rate of inflation, not changes in the CPI.) The vector of innovations, $\boldsymbol{\varepsilon}_t = (\varepsilon_t^y, \varepsilon_t^\pi)'$ is assumed to have mean $\mathbf{0}$, contemporaneous covariance matrix $E[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'] = \boldsymbol{\Omega}$ and to be strictly nonautocorrelated. (We have retained Cecchetti and Rich's notation for most of this discussion, save for the number of lags, which is denoted n in their paper and p here, and some other minor changes which will be noted in passing where necessary.)[11] The equation system may also be written

FN 11

$$\mathbf{B}(L)\begin{bmatrix}\Delta y_t\\ \Delta\pi_t\end{bmatrix} = \begin{bmatrix}\varepsilon_t^y\\ \varepsilon_t^\pi\end{bmatrix},$$

where $\mathbf{B}(L)$ is a 2×2 matrix of polynomials in the lag operator. The components of the disturbance (innovation) vector $\boldsymbol{\varepsilon}_t$ are identified as shocks to aggregate supply and aggregate demand, respectively.

21 ~~20~~.6.8.b The Sacrifice Ratio

Interest in the study centers on the impact over time of structural shocks to output and the rate of inflation. To calculate these, the authors use the **vector moving average** (VMA) form of the model, which would be KT

$$\begin{bmatrix}\Delta y_t\\ \Delta\pi_t\end{bmatrix} = [\mathbf{B}(L)]^{-1}\begin{bmatrix}\varepsilon_t^y\\ \varepsilon_t^\pi\end{bmatrix} = \mathbf{A}(L)\begin{bmatrix}\varepsilon_t^y\\ \varepsilon_t^\pi\end{bmatrix} = \begin{bmatrix}A_{11}(L) & A_{12}(L)\\ A_{21}(L) & A_{22}(L)\end{bmatrix}\begin{bmatrix}\varepsilon_t^y\\ \varepsilon_t^\pi\end{bmatrix}$$

$$= \begin{bmatrix}\sum_{i=0}^{\infty} a_{11}^i \varepsilon_{t-i}^y & \sum_{i=0}^{\infty} a_{12}^i \varepsilon_{t-i}^\pi\\ \sum_{i=0}^{\infty} a_{21}^i \varepsilon_{t-i}^y & \sum_{i=0}^{\infty} a_{22}^i \varepsilon_{t-i}^\pi\end{bmatrix}.$$

(Note that the superscript "i" in the last form of the preceding model is not an exponent; it is the index of the sequence of coefficients.) The impulse response functions for the model corresponding to (21 ~~20~~-27) are precisely the coefficients in $\mathbf{A}(L)$. In particular, the effect on the change in inflation τ periods later of a change in ε_t^π in period t is a_{22}^τ. The total effect from time $t+0$ to time $t+\tau$ would be the sum of these, $\sum_{i=0}^{\tau} a_{22}^i$. The counterparts for the rate of output would be $\sum_{i=0}^{\tau} a_{12}^i$. However, what is needed is not the effect only on period τ's output, but the cumulative effect on output from the time of the shock up to period τ. That would be obtained by summing these period-specific effects, to obtain $\sum_{i=0}^{\tau}\sum_{j=0}^{i} a_{12}^j$. Combining terms, the sacrifice ratio is

$$S_{\varepsilon^\pi}(\tau) = \frac{\sum_{j=0}^{\tau}\frac{\partial y_{t+j}}{\partial \varepsilon_t^\pi}}{\frac{\partial \pi_{t+\tau}}{\partial \varepsilon_t^\pi}} = \frac{\sum_{i=0}^{0} a_{12}^i + \sum_{i=0}^{1} a_{12}^i + \cdots + \sum_{i=0}^{\tau} a_{12}^i}{\sum_{i=0}^{\tau} a_{22}^i} = \frac{\sum_{i=0}^{\tau}\sum_{j=0}^{i} a_{12}^j}{\sum_{i=0}^{\tau} a_{22}^i}.$$

The function $S(\tau)$ is then examined over long periods to study the long-term effects of monetary policy.

AQ: Term "structural model" was a KT on msp 21-33

21 ~~20~~.6.8.c Identification and Estimation of a Structural VAR Model

Estimation of this model requires some manipulation. The **structural model** is a conventional linear simultaneous equations model of the form,

$$\mathbf{B}_0\mathbf{y}_t = \mathbf{B}\mathbf{x}_t + \boldsymbol{\varepsilon}_t,$$

[11] The authors examine two other VAR models, a three-equation model of Shapiro and Watson (1988), which adds an equation in real interest rates $(i_t - \pi_t)$ and a four-equation model by Gali (1992), which models Δy_t, Δi_t, $(i_t - \pi_t)$, and the real money stock, $(\Delta m_t - \pi_t)$. Among the foci of Cecchetti and Rich's paper was the surprisingly large variation in estimates of the sacrifice ratio produced by the three models. In the interest of brevity, we will restrict our analysis to Cecchetti's (1994) two-equation model.

where $\mathbf{y}_t$ is $(\Delta y_t, \Delta \pi_t)'$ and $\mathbf{x}_t$ is the lagged values on the right-hand side. As we saw in ~~Section 13.3~~ Chapter 10, without further restrictions, a model such as this is not identified (estimable). A total of M^2 restrictions—M is the number of equations, here two—are needed to identify the model. In the familiar cases of simultaneous equations models that we examined in Chapter ~~13~~ 10, identification is usually secured through exclusion restrictions (i.e., zero restrictions), either in $\mathbf{B}_0$ or $\mathbf{B}$. This type of exclusion restriction would be unnatural in a model such as this one—there would be no basis for poking specific holes in the coefficient matrices. The authors take a different approach, which requires us to look more closely at the different forms the time-series model can take.

Write the structural form as

$$\mathbf{B}_0\mathbf{y}_t = \mathbf{B}_1\mathbf{y}_{t-1} + \mathbf{B}_2\mathbf{y}_{t-2} + \cdots + \mathbf{B}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t,$$

where

$$\mathbf{B}_0 = \begin{bmatrix} 1 & -b_{12}^0 \\ -b_{21}^0 & 1 \end{bmatrix}.$$

As noted, this is in the form of a conventional simultaneous equations model. Assuming that $\mathbf{B}_0$ is nonsingular, which for this two-equation system requires only that $1 - b_{12}^0 b_{21}^0$ not equal zero, we can obtain the reduced form of the model as

$$\begin{aligned}\mathbf{y}_t &= \mathbf{B}_0^{-1}\mathbf{B}_1\mathbf{y}_{t-1} + \mathbf{B}_0^{-1}\mathbf{B}_2\mathbf{y}_{t-2} + \cdots + \mathbf{B}_0^{-1}\mathbf{B}_p\mathbf{y}_{t-p} + \mathbf{B}_0^{-1}\boldsymbol{\varepsilon}_t \\ &= \mathbf{D}_1\mathbf{y}_{t-1} + \mathbf{D}_2\mathbf{y}_{t-2} + \cdots + \mathbf{D}_p\mathbf{y}_{t-p} + \boldsymbol{\mu}_t,\end{aligned} \quad (\cancel{20}\text{-}30)\ 21$$

where $\boldsymbol{\mu}_t$ is the vector of reduced form innovations. Now, collect the terms in the equivalent form

$$[\mathbf{I} - \mathbf{D}_1 L - \mathbf{D}_2 L^2 - \cdots]\mathbf{y}_t = \boldsymbol{\mu}_t.$$

The moving-average form that we obtained earlier is

$$\mathbf{y}_t = [\mathbf{I} - \mathbf{D}_1 L - \mathbf{D}_2 L^2 - \cdots]^{-1}\boldsymbol{\mu}_t.$$

Assuming stability of the system, we can also write this as

$$\begin{aligned}\mathbf{y}_t &= [\mathbf{I} - \mathbf{D}_1 L - \mathbf{D}_2 L^2 - \cdots]^{-1}\boldsymbol{\mu}_t \\ &= [\mathbf{I} - \mathbf{D}_1 L - \mathbf{D}_2 L^2 - \cdots]^{-1}\mathbf{B}_0^{-1}\boldsymbol{\varepsilon}_t \\ &= [\mathbf{I} + \mathbf{C}_1 L + \mathbf{C}_2 L^2 + \cdots]\boldsymbol{\mu}_t \\ &= \boldsymbol{\mu}_t + \mathbf{C}_1\boldsymbol{\mu}_{t-1} + \mathbf{C}_2\boldsymbol{\mu}_{t-2} \cdots \\ &= \mathbf{B}_0^{-1}\boldsymbol{\varepsilon}_t + \mathbf{C}_1\boldsymbol{\mu}_{t-1} + \mathbf{C}_2\boldsymbol{\mu}_{t-2} \cdots.\end{aligned}$$

So, the $\mathbf{C}_j$ matrices correspond to our $\mathbf{A}_j$ matrices in the original formulation. But this manipulation has added something. We can see that $\mathbf{A}_0 = \mathbf{B}_0^{-1}$. Looking ahead, the reduced form equations can be estimated by least squares. Whether the structural parameters, and thereafter, the VMA parameters can as well depends entirely on whether $\mathbf{B}_0$ can be estimated. From (~~20~~-30) we can see that if $\mathbf{B}_0$ can be estimated, then $\mathbf{B}_1 \ldots \mathbf{B}_p$ can also just by premultiplying the reduced form coefficient matrices by this estimated

21

21-37

$\mathbf{B}_0$. So, we must now consider this issue. ~~(This is precisely the conclusion we drew at the beginning of Section 13.3.)~~

Recall the initial assumption that $E[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'] = \boldsymbol{\Omega}$. In the reduced form, we assume $E[\boldsymbol{\mu}_t\boldsymbol{\mu}_t'] = \boldsymbol{\Sigma}$. As we know, reduced forms are always estimable (indeed, by least squares if the assumptions of the model are correct). That means that $\boldsymbol{\Sigma}$ is estimable by the least squares residual variances and covariance. From the earlier derivation, we have that $\boldsymbol{\Sigma} = \mathbf{B}_0^{-1}\boldsymbol{\Omega}(\mathbf{B}_0^{-1})' = \mathbf{A}_0\boldsymbol{\Omega}\mathbf{A}_0'$. (Again, see the beginning of Section 13.3.) The authors have secured identification of the model through this relationship. In particular, they assume first that $\boldsymbol{\Omega} = \mathbf{I}$. Assuming that $\boldsymbol{\Omega} = \mathbf{I}$, we now have that $\mathbf{A}_0\mathbf{A}_0' = \boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ is an estimable matrix with three free parameters. Because $\mathbf{A}_0$ is 2×2, one more restriction is needed to secure identification. At this point, the authors, invoking Blanchard and Quah (1989), assume that "demand shocks have no permanent effect on the level of output. This is equivalent to $A_{12}(1) = \sum_{i=0}^{\infty} a_{12}^i = 0$." This might seem like a cumbersome restriction to impose. But, the matrix $\mathbf{A}(1)$ is $[\mathbf{I} - \mathbf{D}_1 - \mathbf{D}_2 - \cdots - \mathbf{D}_p]^{-1}\mathbf{A}_0 = \mathbf{F}\mathbf{A}_0$ and the components, $\mathbf{D}_j$, have been estimated as the reduced form coefficient matrices, so $\mathbf{A}_{12}(1) = 0$ assumes only that the upper right element of this matrix is zero. We now obtain the equations needed to solve for $\mathbf{A}_0$. First,

Au: Confirm x-ref to Section 13.3 is OK

$$\mathbf{A}_0\mathbf{A}_0' = \boldsymbol{\Sigma} \Rightarrow \begin{bmatrix} (a_{11}^0)^2 + (a_{12}^0)^2 & a_{11}^0 a_{21}^0 + a_{12}^0 a_{22}^0 \\ a_{11}^0 a_{21}^0 + a_{12}^0 a_{22}^0 & (a_{21}^0)^2 + (a_{22}^0)^2 \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{11} \end{bmatrix}, \qquad (\cancel{20}\,21\text{-}31)$$

which provides three equations. Second, the theoretical restriction is

$$\mathbf{F}\mathbf{A}_0 = \begin{bmatrix} * & f_{11}a_{12}^0 + f_{12}a_{22}^0 \\ * & * \end{bmatrix} = \begin{bmatrix} * & 0 \\ * & * \end{bmatrix}.$$

FN 12

This provides the four equations needed to identify the four elements in $\mathbf{A}_0$.[12]

Collecting results, the estimation strategy is first to estimate $\mathbf{D}_1, \ldots, \mathbf{D}_p$ and $\boldsymbol{\Sigma}$ in the reduced form, by least squares. (They set $p = 8$.) Then use the restrictions and (~~20~~21-31) to obtain the elements of $\mathbf{A}_0 = \mathbf{B}_0^{-1}$ and, finally, $\mathbf{B}_j = \mathbf{A}_0^{-1}\mathbf{D}_j$.

The last step is estimation of the matrices of impulse responses, which can be done as follows: We return to the reduced form which, using our augmentation trick, we

[12] At this point, an intriguing loose end arises. We have carried this discussion in the form of the original papers by Blanchard and Quah (1989) and Cecchetti and Rich (2001). Returning to the original structure, we see that because $\mathbf{A}_0 = \mathbf{B}_0^{-1}$, if $\mathbf{B}_0$ has ones on the diagonal, then $\mathbf{A}_0$ actually does not have four unrestricted and unknown elements, it has two. The model is thus overidentified. We could have predicted this at the outset. In our conventional simultaneous equations model, the normalizations in $\mathbf{B}_0$ (ones on the diagonal) provide two restrictions of the $M^2 = 4$ required for identification. Assuming that $\boldsymbol{\Omega} = \mathbf{I}$ provides three more, and the theoretical restriction provides a sixth. Therefore, the four unknown elements in an unrestricted $\mathbf{B}_0$ are overidentified. It might seem convenient at this point to forego the theoretical restriction on long-term impacts, but it seems more natural to omit the restrictions on the scaling of $\boldsymbol{\Omega}$. With the two normalizations already in place, assuming that the innovations are uncorrelated ($\boldsymbol{\Omega}$ is diagonal) and "demand shocks have no permanent effect on the level of output" together suffice to identify the model. Blanchard and Quah appear to reach the same conclusion (page 656), but then they also assume the unit variances [page 657, equation (1)]. They argue that the assumption of unit variances is just a convenient normalization, which for their model is actually the case, because they do not assume that $\mathbf{B}_0$ is diagonal. Cecchetti and Rich, however, *do* appear to normalize $\mathbf{B}_0$ in their equation (1). They then (evidently) drop the assumption after (10), however, "[B]ecause $\mathbf{A}_0$ has $(n \times n)$ unique elements ...." This would imply that the normalization they impose on their (1) has not, in fact, been carried through the later manipulations, so, once again, the model is exactly identified.

21-38

write as

$$\begin{bmatrix} \mathbf{y}_t \\ \mathbf{y}_{t-1} \\ \cdots \\ \mathbf{y}_{t-p+1} \end{bmatrix} = \begin{bmatrix} \mathbf{D}_1 & \mathbf{D}_2 & \cdots & \mathbf{D}_p \\ \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{y}_{t-1} \\ \mathbf{y}_{t-2} \\ \cdots \\ \mathbf{y}_{t-p} \end{bmatrix} + \begin{bmatrix} \mathbf{A}_0\boldsymbol{\varepsilon}_t \\ \mathbf{0} \\ \cdots \\ \mathbf{0} \end{bmatrix}. \qquad (\cancel{20}\ 21\text{-}32)$$

For convenience, arrange this result as

$$\mathbf{Y}_t = (\mathbf{D}L)\mathbf{Y}_t + \mathbf{w}_t.$$

Now, solve this for $\mathbf{Y}_t$ to obtain the final form

$$\mathbf{Y}_t = [\mathbf{I} - \mathbf{D}L]^{-1}\mathbf{w}_t.$$

Write this in the spectral form and expand as we did earlier, to obtain

$$\mathbf{Y}_t = \sum_{i=0}^{\infty} \mathbf{P}\boldsymbol{\Lambda}^i\mathbf{Q}\mathbf{w}_{t-i}. \qquad (\cancel{20}\ 21\text{-}33)$$

We will be interested in the uppermost subvector of $\mathbf{Y}_t$, so we expand (~~20~~ 21-33) to yield

$$\begin{bmatrix} \mathbf{y}_t \\ \mathbf{y}_{t-1} \\ \cdots \\ \mathbf{y}_{t-p+1} \end{bmatrix} = \left[\sum_{i=0}^{\infty} \mathbf{P}\boldsymbol{\Lambda}^i\mathbf{Q} \begin{bmatrix} \mathbf{A}_0\boldsymbol{\varepsilon}_{t-i} \\ \mathbf{0} \\ \cdots \\ \mathbf{0} \end{bmatrix} \right].$$

The matrix in the summation is $Mp \times Mp$. The impact matrices we seek are the $M \times M$ matrices in the upper left corner of the spectral form, multiplied by $\mathbf{A}_0$.

21 ~~20~~.6.8.d Inference

As noted at the end of Section ~~20~~ 21.6.6, obtaining usable standard errors for estimates of impulse responses is a difficult (as yet unresolved) problem. Killian (1998) has suggested that bootstrapping is a preferable approach to using the delta method. Cecchetti and Rich reach the same conclusion and likewise resort to a bootstrapping procedure. Their bootstrap procedure is carried out as follows: Let $\hat{\delta}$ and $\hat{\Sigma}$ denote the full set of estimated coefficients and estimated reduced form covariance matrix based on direct estimation. As suggested by Doan (2007), they construct a sequence of N draws for the reduced form parameters, then recompute the entire set of impulse responses. The narrowest interval, which contains 90 percent of these draws, is taken to be a confidence interval for an estimated impulse function.

21 ~~20~~.6.8.e Empirical Results

Cecchetti and Rich used quarterly observations on real aggregate output and the consumer price index. Their data set spanned 1959.1 to 1997.4. This is a subset of the data described in the Appendix Table F5.~~1~~ 2. Before beginning their analysis, they subjected the data to the standard tests for stationarity. Figures ~~20~~ 21.5–~~20~~ 21.7 show the log of real output, the rate of inflation, and the changes in these two variables. The

FIGS 21.5 21.6 21.7

21-39

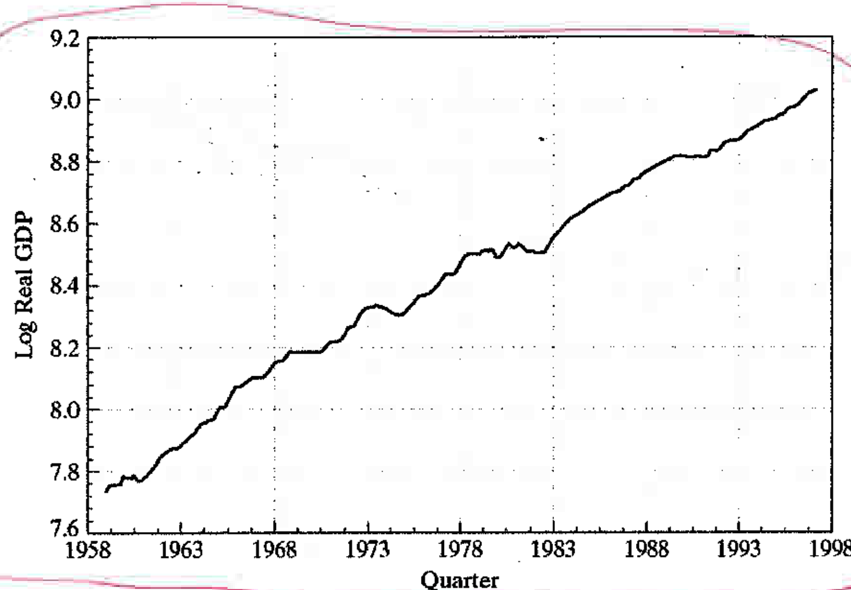

FIGURE ~~20~~.5 21 Log GDP.

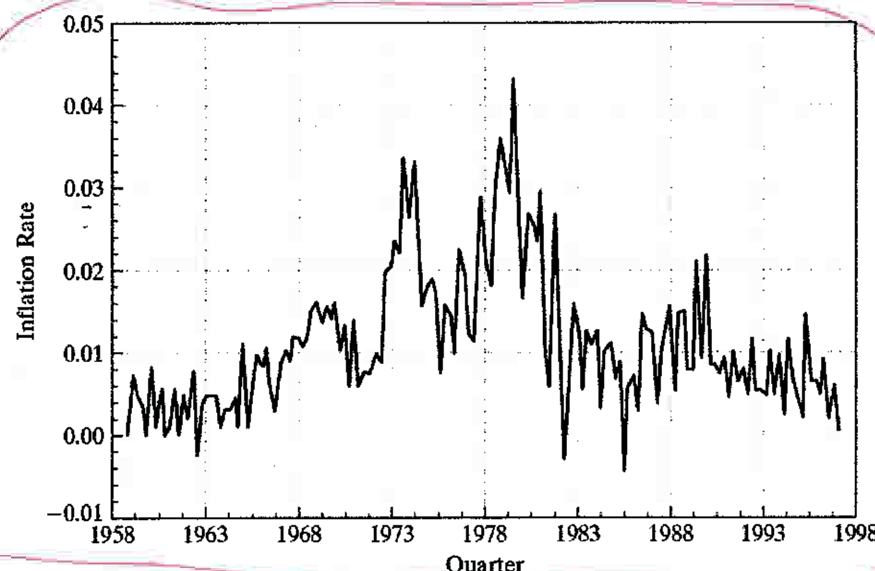

FIGURE ~~20~~.6 21 The Quarterly Rate of Inflation.

first two figures do suggest that neither variable is stationary. On the basis of the Dickey–Fuller (1981) test (see Section ~~22~~ 23.2.4), they found (as might be expected) that the y_t and π_t series both contain unit roots. They conclude that because output has a unit root, the identification restriction that the long-run effect of aggregate demand shocks on output is well defined and meaningful. The unit root in inflation allows for

21-40

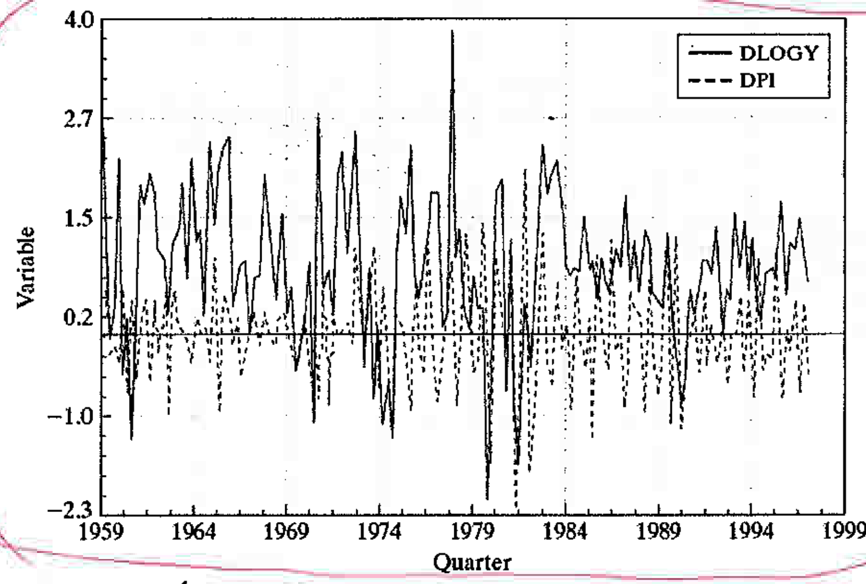

FIGURE ~~20~~.7 21 Rates of Change, Log Real GDP, and the Rate of Inflation.

permanent shifts in its level. The lag length for the model is set at $p = 8$. Long-run impulse response functions are truncated at 20 years (80 quarters). Analysis is based on the rate of change data shown in Figure ~~20~~.7. 21

As a final check on the model, the authors examined the data for the possibility of a structural shift using the tests described in Andrews (1993) and Andrews and Ploberger (1994). None of the Andrews/Quandt supremum LM test, Andrews/Ploberger exponential LM test, or the Andrews/Ploberger average LM test suggested that the underlying structure had changed (in spite of what seems likely to have been a major shift in Fed policy in the 1970s). On this basis, they concluded that the VAR is stable over the sample period.

FIG 21.8 Figure ~~20~~.8 21 (Figures 3A and 3B taken from the article) shows their two separate estimated impulse response functions. The dotted lines in the figures show the bootstrap-generated confidence bounds. Estimates of the sacrifice ratio for Cecchetti's model are 1.3219 for $\tau = 4$, 1.3204 for $\tau = 8$, 1.5700 for $\tau = 12$, 1.5219 for $\tau = 16$, and 1.3763 for $\tau = 20$.

The authors also examined the forecasting performance of their model compared to Shapiro and Watson's and Gali's. The device used was to produce one step ahead, period $T + 1 \mid T$ forecasts for the model estimated using periods $1, \ldots, T$. The first reduced form of the model is fit using 1959.1 to 1975.1 and used to forecast 1975.2. Then, it is reestimated using 1959.1 to 1975.2 and used to forecast 1975.3, and so on. Finally, the root mean squared error of these out of sample forecasts is compared for three models. In each case, the level, rather than the rate of change of the inflation rate is forecasted. Overall, the results suggest that the smaller model does a better job of estimating the impulse responses (has smaller confidence bounds and conforms more

21-41

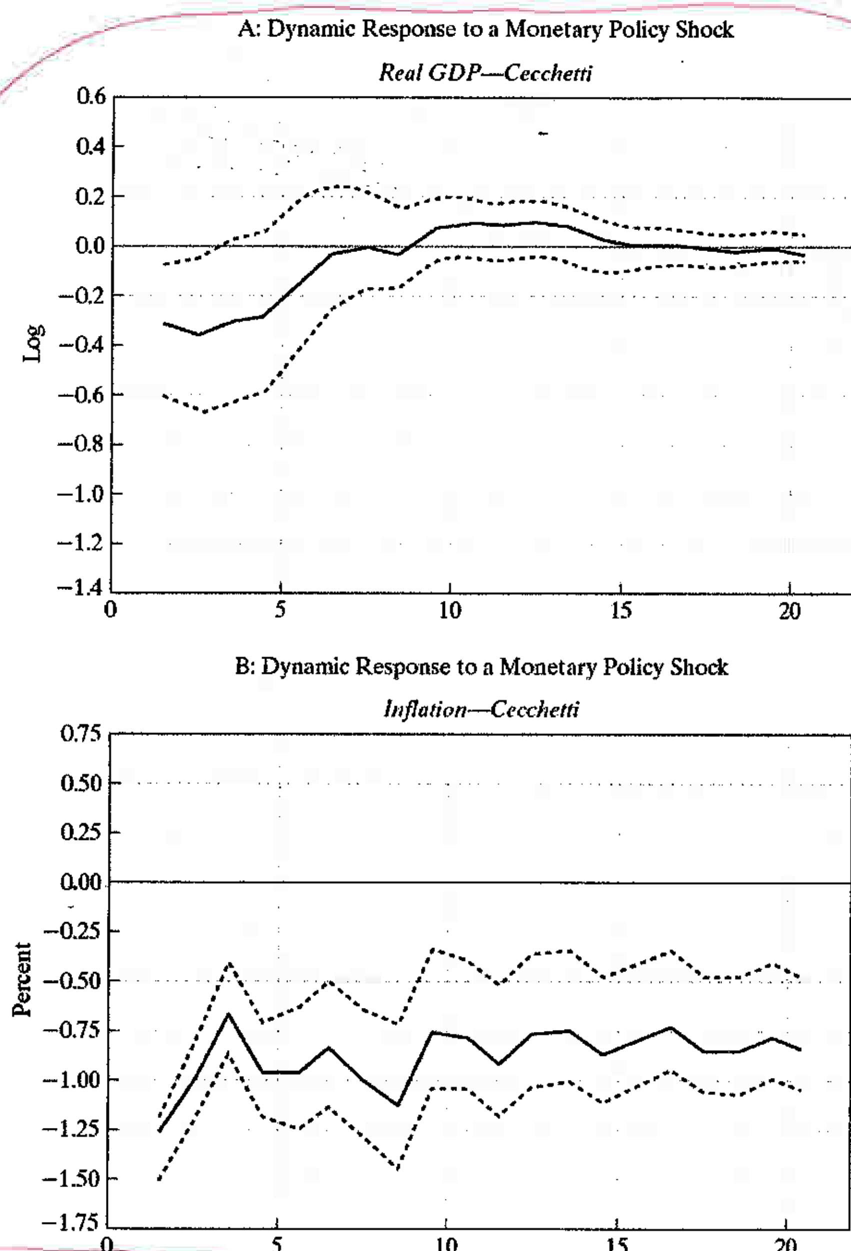

FIGURE ~~20~~.8 Estimated Impulse Response Functions.
21

nearly with theoretical predictions) but performs worst of the three (slightly) in terms of the mean squared error of the out-of-sample forecasts. Because the unrestricted reduced form model is being used for the latter, this comes as no surprise. The end result follows essentially from the result that adding variables to a regression model improves its fit.

21-42

21 ~~20~~.6.9 VARs IN MICROECONOMICS

VARs have appeared in the microeconometrics literature as well. Chamberlain (1980) suggested that a useful approach to the analysis of panel data would be to treat each period's observation as a separate equation. For the case of $T = 2$, we would have

$$y_{i1} = \alpha_i + \mathbf{x}_{i1}'\beta + \varepsilon_{i1},$$
$$y_{i2} = \alpha_i + \mathbf{x}_{i2}'\beta + \varepsilon_{i2},$$

where i indexes individuals and α_i are unobserved individual effects. This specification produces a multivariate regression, to which Chamberlain added restrictions related to the individual effects. Holtz-Eakin, Newey, and Rosen's (1988) approach is to specify the equation as

$$y_{it} = \alpha_{0t} + \sum_{l=1}^{m} \alpha_{lt} y_{i,t-l} + \sum_{l=1}^{m} \delta_{lt} x_{i,t-l} + \Psi_t f_i + \mu_{it}.$$

In their study, y_{it} is hours worked by individual i in period t and x_{it} is the individual's wage in that period. A second equation for earnings is specified with lagged values of hours and earnings on the right-hand side. The individual, unobserved effects are f_i. This model is similar to the VAR in (21 ~~20~~-27), but it differs in several ways as well. The number of periods is quite small (14 yearly observations for each individual), but there are nearly 1,000 individuals. The dynamic equation is specified for a specific period, however, so the relevant sample size in each case is n, not T. Also, the number of lags in the model used is relatively small; the authors fixed it at three. They thus have a two-equation VAR containing 12 unknown parameters, six in each equation. The authors used the model to analyze causality, measurement error, and parameter stability—that is, constancy of α_{lt} and δ_{lt} across time.

Example 21 ~~20~~.7 VAR for Municipal Expenditures

In Example 13 ~~15~~.10, we examined a model of municipal expenditures proposed by Dahlberg and Johansson (2000): Their equation of interest is

$$\Delta S_{i,t} = \mu_t + \sum_{j=1}^{m} \beta_j \Delta S_{i,t-j} + \sum_{j=1}^{m} \gamma_j \Delta R_{i,t-j} + \sum_{j=1}^{m} \delta_j \Delta G_{i,t-j} + u_{i,t}^S$$

for $i = 1, \ldots, N = 265$ and $t = m+1, \ldots, 9$. $S_{i,t}$, $R_{i,t}$, and $G_{i,t}$ are municipal spending, receipts (taxes and fees), and central government grants, respectively. Analogous equations are specified for the current values of $R_{i,t}$ and $G_{i,t}$. This produces a vector autoregression for each municipality,

$$\begin{bmatrix} \Delta S_{i,t} \\ \Delta R_{i,t} \\ \Delta G_{i,t} \end{bmatrix} = \begin{pmatrix} \mu_{S,t} \\ \mu_{R,t} \\ \mu_{G,t} \end{pmatrix} + \begin{pmatrix} \beta_{S,1} & \gamma_{S,1} & \delta_{S,1} \\ \beta_{R,1} & \gamma_{R,1} & \delta_{R,1} \\ \beta_{G,1} & \gamma_{G,1} & \delta_{G,1} \end{pmatrix} \begin{bmatrix} \Delta S_{i,t-1} \\ \Delta R_{i,t-1} \\ \Delta G_{i,t-1} \end{bmatrix} + \cdots$$
$$+ \begin{pmatrix} \beta_{S,m} & \gamma_{S,m} & \delta_{S,m} \\ \beta_{R,m} & \gamma_{R,m} & \delta_{R,m} \\ \beta_{G,m} & \gamma_{G,m} & \delta_{G,m} \end{pmatrix} \begin{bmatrix} \Delta S_{i,t-m} \\ \Delta R_{i,t-m} \\ \Delta G_{i,t-m} \end{bmatrix} + \begin{bmatrix} u_{i,t}^S \\ u_{i,t}^R \\ u_{i,t}^G \end{bmatrix}.$$

The model was estimated by GMM, so the discussion at the end of the preceding section applies here. We will be interested in testing whether changes in municipal spending, $\Delta S_{i,t}$, are Granger-caused by changes in revenues, $\Delta R_{i,t}$, and grants, $\Delta G_{i,t}$. The hypothesis to be tested

21-43

13

is $\gamma_{S,j} = \delta_{S,j} = 0$ for all j. This hypothesis can be tested in the context of only the first equation. Parameter estimates and diagnostic statistics are given in Example ~~18~~.10. We can carry out the test in two ways. In the unrestricted equation with all three lagged values of all three variables, the minimized GMM criterion is $q = 22.8287$. If the lagged values of ΔR and ΔG are omitted from the ΔS equation, the criterion rises to 42.9182.[13] There are six restrictions. The difference is 20.090 so the F statistic is $20.09/6 = 3.348$. We have more than 1,000 degrees of freedom for the denominator, with 265 municipalities and 5 years, so we can use the limiting value for the critical value. This is 2.10, so we may reject the hypothesis of noncausality and conclude that changes in revenues and grants do Granger cause changes in spending. (This hardly seems surprising.) The alternative approach is to use a Wald statistic to test the six restrictions. Using the full GMM results for the ΔS equation with 14 coefficients we obtain a Wald statistic of 15.3030. The critical chi-squared would be $6 \times 2.1 = 12.6$, so once again, the hypothesis is rejected.

FN 13

five

Dahlberg and Johansson approach the causality test somewhat differently by using a sequential testing procedure. (See their page 413 for discussion.) They suggest that the intervening variables be dropped in turn. By dropping first G, then R and G, and then first R then G and R, they conclude that grants do not Granger-cause changes in spending ($\Delta q =$ only 0.07) but in the absence of grants, revenues do (Δq|grants excluded) = 24.6. The reverse order produces test statistics of 12.2 and 12.4, respectively. Our own calculations of the four values of q yields 22.829 for the full model, 23.1302 with only grants excluded, 23.0894 with only R excluded, and 42.9182 with both excluded, which disagrees with their results but is consistent with our earlier ones.

AU: Is this part of Example 21.7?

Instability of a VAR Model

The coefficients for the three-variable VAR model in Example ~~20~~.7 appear in Table ~~15~~.5. The characteristic roots of the 9×9 coefficient matrix are -0.6025, 0.2529, 0.0840, $(1.4586 \pm 0.6584i)$, $(-0.6992 \pm 0.2019i)$, and $(0.0611 \pm 0.6291i)$. The first pair of complex roots has modulus greater than one, so the estimated VAR is unstable. The data do not appear to be consistent with this result, though with only five usable years of data, that conclusion is a bit fragile. One might suspect that the model is overfit. Because the disturbances are assumed to be uncorrelated across equations, the three equations have been estimated separately. The GMM criterion for the system is then the sum of those for the three equations. For $p = 3, 2$, and 1, respectively, these are $(22.8287 + 30.5398 + 17.5810) = 70.9495$, $(30.4526 + 34.2590 + 20.5416) = 85.2532$, and $(34.4986 + 53.2506 + 27.5927) = 115.6119$. The difference statistic for testing down from three lags to two is 14.3037. The critical chi-squared for nine degrees of freedom is 19.62, so it would appear that $m = 3$ may be too large. The results clearly reject the hypothesis that $m = 1$, however. The coefficients for a model with two lags instead of one appear in Table 15.5. If we construct Γ from these results instead, we obtain a 6×6 matrix whose characteristic roots are 1.5817, -0.2196, $-0.3509 \pm 0.4362i$, and $0.0968 \pm 0.2791i$. The system remains unstable.

AU: Confirm Table 15.5 x-ref is OK

20.7 SUMMARY AND CONCLUSIONS

This chapter has surveyed a particular type of regression model, the dynamic regression. The signature feature of the dynamic model is effects that are delayed or that persist through time. In a static regression setting, effects embodied in coefficients are assumed to take place all at once. In the dynamic model, the response to an innovation is distributed through several periods. The first three sections of this chapter examined several different forms of single-equation models that contained lagged effects. The

[13]Once again, these results differ from those given by Dahlberg and Johansson. As before, the difference results from our use of the same weighting matrix for all GMM computations in contrast to their recomputation of the matrix for each new coefficient vector estimated.

21-44

progression, which mirrors the current literature, is from tightly structured lag "models" (which were sometimes formulated to respond to a shortage of data rather than to correspond to an underlying theory) to unrestricted models with multiple period lag structures. We also examined several hybrids of these two forms, models that allow long lags but build some regular structure into the lag weights. Thus, our model of the formation of expectations of inflation is reasonably flexible, but does assume a specific behavioral mechanism. We then examined several methodological issues. In this context as elsewhere, there is a preference in the methods toward forming broad unrestricted models and using familiar inference tools to reduce them to the final appropriate specification. The second half of the chapter was devoted to a type of seemingly unrelated regressions model. The vector autoregression, or VAR, has been a major tool in recent research. After developing the econometric framework, we examined two applications, one in macroeconomics centered on monetary policy and one from microeconomics.

AU: Terms with blue checks were not bold KTs in text

Key Terms and Concepts

- Autocorrelation
- Autoregression
- Autoregressive distributed lag ?ARDL?
- Autoregressive form
- Autoregressive model
- Characteristic equation
- Distributed lag form
- Dynamic regression model
- Elasticity
- Equilibrium
- Equilibrium error
- Equilibrium multiplier
- Equilibrium relationship
- Error correction
- Exogeneity
- Expectation
- Finite lags
- General-to-simple method
- Granger causality
- Impact multiplier
- Impulse response
- Infinite lag model
- Infinite lags
- Innovation
- Invertible
- Lagged variables
- Lag operator
- Lag weight
- Long-run multiplier
- Mean lag
- Median lag
- Moving-average form
- One-period-ahead forecast
- Partial adjustment
- Phillips curve
- Polynomial in lag operator
- Random walk with drift
- Rational lag
- Simple-to-general approach
- Specification
- Stability
- Stationary
- Strong exogeneity
- Structural model
- Structural VAR
- Superconsistent
- Univariate autoregression
- Vector autoregression (VAR)
- Vector moving average (VMA)

Exercises

1. Obtain the mean lag and the long- and short-run multipliers for the following distributed lag models:
   a. $y_t = 0.55(0.02x_t + 0.15x_{t-1} + 0.43x_{t-2} + 0.23x_{t-3} + 0.17x_{t-4}) + e_t$.
   b. The model in Exercise 3.
   c. The model in Exercise 4. (Do for either x or z.)
2. Expand the rational lag model $y_t = [(0.6 + 2L)/(1 - 0.6L + 0.5L^2)]x_t + e_t$. What are the coefficients on $x_t, x_{t-1}, x_{t-2}, x_{t-3}$, and x_{t-4}?
3. Suppose that the model of Exercise 2 were specified as

$$y_t = \alpha + \frac{\beta + \gamma L}{1 - \delta_1 L - \delta_2 L^2} x_t + e_t.$$

21-45
End 21

Describe a method of estimating the parameters. Is ordinary least squares consistent?

4. Describe how to estimate the parameters of the model

$$y_t = \alpha + \beta \frac{x_t}{1-\gamma L} + \delta \frac{z_t}{1-\phi L} + \varepsilon_t,$$

where ε_t is a serially uncorrelated, homoscedastic, classical disturbance.

Applications

1. We are interested in the long-run multiplier in the model

$$y_t = \beta_0 + \sum_{j=0}^{6} \beta_j x_{t-j} + \varepsilon_t.$$

Assume that x_t is an autoregressive series, $x_t = r x_{t-1} + v_t$ where $|r| < 1$.

a. What is the long run multiplier in this model?
b. How would you estimate the long-run multiplier in this model?
c. Suppose you knew that the preceding is the true model but you linearly regress y_t only on a constant and the first five lags of x_t. How does this affect your estimate of the long run multiplier?

part/

d. Same as c. for four lags instead of five.
e. Using the macroeconomic data in Appendix Table F5.1 2, let y_t be the log of real investment and x_t be the log of real output. Carry out the computations suggested and report your findings. Specifically, how does the omission of a lagged value affect estimates of the short-run and long-run multipliers in the unrestricted lag model?

2. Explain how to estimate the parameters of the following model:

$$y_t = \alpha + \beta x_t + \gamma y_{t-1} + \delta y_{t-2} + e_t,$$
$$e_t = \rho e_{t-1} + u_t.$$

Is there any problem with ordinary least squares? Let y_t be consumption and let x_t be disposable income. Using the method you have described, fit the previous model to the data in Appendix Table F5.1 2. Report your results.