

Chapter 17: Econometrics

Prerequisites: Chapter 6, Sections 3.5 - 3.8

17.1 The Problems with Nonrecursive Systems

This chapter contains a mixture of ideas extended from the Chapters on Regression, in particular Chapter 5 and 6, and the chapters on covariance structure, in particular 9 and 10. To anticipate a theme of this chapter, econometricians have come up with a variety of ways to use the basic least squares philosophy to look at models with latent variables and complex causal structures. In this section we are concerned with nonrecursive systems, with equations of the form $\mathbf{y} = \mathbf{B}\mathbf{y} + \mathbf{\Gamma}\mathbf{x} + \boldsymbol{\zeta}$, where $\mathbf{V}(\boldsymbol{\zeta})$ is not diagonal, or it is impossible to arrange the sequence of \mathbf{y} variables such that \mathbf{B} is lower triangular. To illustrate the problems caused by nonrecursion, we start with a deceptively simple two equation system:

$$y_1 = \beta_{12}y_2 + \zeta_1 \quad (17.1)$$

$$y_2 = x_1 + y_1,$$

where y_1 represents the expenditures on our product category, y_2 is income, and x_1 is all other expenditures, including savings. As we did in Chapter 10, we are dropping any subscript that references the individual observation in this section. However, the reader should keep in mind that ζ_1 is a random input to the model, and varies from one observation to the next. The second equation is known as an identity, since there is no error term. If we were to assume that $\mathbf{V}(\zeta_1) = \sigma^2 \mathbf{I}$, can we use the OLS approach of Chapter 5? Unfortunately not, since problems arise due to the covariance between y_2 and ζ_1 . This becomes clear when we substitute the y_1 equation into the y_2 identity:

$$\begin{aligned} y_2 &= (\beta_{12}y_2 + \zeta_1) + x_1 \\ y_2 - \beta_{12}y_2 &= \zeta_1 + x_1 \\ y_2 &= \frac{\zeta_1}{1 - \beta_{12}} + \frac{x_1}{1 - \beta_{12}}. \end{aligned} \quad (17.2)$$

If we assume that $E(\zeta_1) = 0$ then we can say

$$E(y_2) = \frac{x_1}{1 - \beta_{12}} \quad (17.3)$$

which means, by the definition of variance [Equation (4.7)], we get:

$$\begin{aligned} \text{Cov}(\zeta_1, y_2) &= E\{[\zeta_1 - E(\zeta_1)][y_2 - E(y_2)]\} \\ &= E\{\zeta_1[y_2 - E(y_2)]\} \end{aligned}$$

where we get to the second line above since $E(\zeta_1) = 0$. Now, substituting the results of Equation (17.2) and Equation (17.3) into the line above, we get

$$\begin{aligned} \text{Cov}(\zeta_1, y_2) &= E\left[\zeta_1 \left(\frac{\zeta_1}{1-\beta_{12}}\right)\right] \\ &= \frac{1}{1-\beta_{12}} E(\zeta_1 \zeta_1) = \frac{\sigma^2}{1-\beta_{12}}. \end{aligned}$$

Thus, y_2 , which functions as an independent variable in the equation for y_1 , is correlated with the error for that equation, ζ_1 . This is a no-no. In this situation the usual least squares estimator $\hat{\beta}$ is not consistent [consistency is defined in Equation (5.11), but see Johnson p 281-2 for a proof].

There are three solutions to this problem. First, there is what econometricians call *Full Information Maximum Likelihood* which is basically the covariance structure model covered in Chapter 10. Estimating a nonrecursive system using covariance structural models can be tricky however. Second, there is what is known as *Indirect Least Squares* which takes advantage of reduced form, covered elsewhere [Equation (10.6)]:

$$\begin{aligned} \mathbf{y} &= \mathbf{B}\mathbf{y} + \mathbf{\Gamma}\mathbf{x} + \boldsymbol{\zeta} \\ \mathbf{y} - \mathbf{B}\mathbf{y} &= \mathbf{\Gamma}\mathbf{x} + \boldsymbol{\zeta} \\ (\mathbf{I} - \mathbf{B})\mathbf{y} &= \mathbf{\Gamma}\mathbf{x} + \boldsymbol{\zeta} \\ \mathbf{y} &= (\mathbf{I} - \mathbf{B})^{-1}\mathbf{\Gamma}\mathbf{x} + (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\zeta} \\ \mathbf{y} &= \mathbf{G}\mathbf{x} + \mathbf{e} \end{aligned}$$

We can use OLS to estimate the elements in $\hat{\mathbf{G}}$. The major problem here is that unless the model is just identified, with exactly the right number of unknowns, you cannot recover the structural parameters of theoretical importance in \mathbf{B} and $\mathbf{\Gamma}$.

Third, there is a technique called Two Stage Least Squares and we will now cover that.

17.2 Two Stage Least Squares

The basic strategy of Two Stage Least Squares, sometimes called *2SLS*, is to replace y_2 with \hat{y}_2 in Equation (17.1) above. To discuss the technique further, we need to revert to the notational convention of Chapters 5, 6 and 8 which explicitly makes reference to individual observations. Rather than refer to a particular endogenous variable as y_2 , lets say, it is now a particular column of the \mathbf{Y} matrix which has n rows, one row for each observation. To get the discussion started, we introduce some key vectors and matrices:

Array	Order	Description
$\mathbf{y}_{\cdot 1}$	$n \cdot 1$	Endogeneous variable of interest
\mathbf{Y}_2	$n \cdot (p-1)$	Other endogenous variables in the equation for $\mathbf{y}_{\cdot 1}$

β_2	$(p-1) \cdot 1$	Structural parameters for Y_2
X_1	$n \cdot k_1$	Exogenous variables in equation for $y_{\cdot 1}$
γ	$K_1 \cdot 1$	Structural parameters for X_1
$\zeta_{\cdot 1}$	$n \cdot 1$	Error in the equation for $y_{\cdot 1}$

The model looks like

$$y_{\cdot 1} = Y_2 \beta_2 + X_1 \beta_1 + \zeta_{\cdot 1}$$

Now we define the full set of exogenous variables as $X = [X_1 \mid X_2]$. In stage 1 we regress Y_2 on X to produce:

$$\hat{Y}_2 = X(X'X)^{-1}X'Y_2.$$

In stage 2 we regress $y_{\cdot 1}$ on \hat{Y}_2 and X_1 . This produces a formula for the unknowns as below:

$$\begin{bmatrix} \hat{\beta}_2 \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} \hat{Y}_2' \hat{Y}_2 & \hat{Y}_2' X_1 \\ X_1' \hat{Y}_2 & X_1' X_1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{Y}_2' y_{\cdot 1} \\ X_1' y_{\cdot 1} \end{bmatrix}.$$

While Y_2 may be correlated with $\zeta_{\cdot 1}$ we expect that \hat{Y}_2 is not. It is not literally necessary to execute two stage least squares in two stages. Instead you can use

$$\begin{bmatrix} \hat{\beta}_2 \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} Y_2' X(X'X)^{-1} X' Y_2 & Y_2' X_1 \\ X_1' Y_2 & X_1' X_1 \end{bmatrix}^{-1} \begin{bmatrix} Y_2' X(X'X)^{-1} X' y_{\cdot 1} \\ X_1' y_{\cdot 1} \end{bmatrix}$$

or define $Y_2 = \hat{Y}_2 + E_2$ so that

$$\hat{Y}_2' \hat{Y}_2 = \hat{Y}_2' (Y_2 - E_2) = \hat{Y}_2' Y_2 = Y_2' X(X'X)^{-1} X' Y_1$$

Now rewriting,

$$\begin{bmatrix} \hat{\beta}_2 \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} Y_2' Y_2 - k E_2' E_2 & Y_2' X_1 \\ X_1' Y_2 & X_1' X_1 \end{bmatrix}^{-1} \begin{bmatrix} (Y_2' - k E_2') y_{\cdot 1} \\ X_1' y_{\cdot 1} \end{bmatrix}.$$

For $k = 0$ we have OLS and for $k = 1$ we have 2SLS. There is a technique called Limited Information Maximum Likelihood in which k is itself estimated.

17.3 Econometric Approaches to Measurement Error

We begin by noting that measurement error in the y vector is not a problem for regression. Assume the real model is

$$\tilde{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

where $\tilde{\mathbf{y}}$ is the true value of the dependent variable vector. Instead, unfortunately, we observe

$$\mathbf{y} = \tilde{\mathbf{y}} + \boldsymbol{\delta}$$

where $\boldsymbol{\delta}$, in general, is not a null vector. We can write the true model

$$\mathbf{y} - \boldsymbol{\delta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} + \boldsymbol{\delta},$$

so that we just get a slightly different error term. Unless $\text{Cov}(\boldsymbol{\delta}, \mathbf{X}) \neq \mathbf{0}$ we will be OK. Now, however, let's contemplate what happens when there is measurement error on the x side. Imagine that we have the true model

$$\mathbf{y} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \mathbf{e}$$

but we observe

$$\mathbf{X} = \tilde{\mathbf{X}} + \mathbf{F} \tag{17.4}$$

instead. Rewriting the true model, we get

$$\begin{aligned} \mathbf{y} &= \tilde{\mathbf{X}}\boldsymbol{\beta} + \mathbf{e} \\ &= (\mathbf{X} - \mathbf{F})\boldsymbol{\beta} + \mathbf{e} \\ &= \mathbf{X}\boldsymbol{\beta} - \mathbf{F}\boldsymbol{\beta} + \mathbf{e} \\ &= \mathbf{X}\boldsymbol{\beta} + (\mathbf{e} - \mathbf{F}\boldsymbol{\beta}). \end{aligned}$$

In this case we find out that the $\text{Cov}(\mathbf{X}, \mathbf{F}\boldsymbol{\beta})$ is not going to vanish since \mathbf{F} is a component of \mathbf{X} . Thus the error and the independent variables are correlated and the OLS estimator is not consistent. We can get around this problem using a technique called *Instrumental Variables*. We need to find a set of instruments, $\mathbf{X}_{(i)}$, that are independent of both the error vector \mathbf{e} and the errors in the \mathbf{X} -variables, \mathbf{F} . We then estimate $\boldsymbol{\beta}$ below such that

$$\hat{\boldsymbol{\beta}}_{(i)} = (\mathbf{X}'_{(i)}\mathbf{X})^{-1}\mathbf{X}'_{(i)}\mathbf{y}$$

and $\hat{\boldsymbol{\beta}}_{(i)}$ will then consistently estimate $\boldsymbol{\beta}$. From time to time we might use \mathbf{Z} with 1's and -1's from a median split of the x variables.

17.4 Generalized Least Squares

GLS estimation has been discussed in Sections 6.8, 12.4 and 13.3. Here we review and further develop the concept of GLS with an eye to applying it to data that are collected across time and so cannot be considered independent. In the basic linear model,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e},$$

in this section we will assume that $\mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{V})$ where in general, $\mathbf{V} \neq \mathbf{I}$. Regardless as to the distribution of \mathbf{e} , if we estimate

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y},$$

we find that $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$, but this estimator no longer produces the best, or smallest, variance, $V(\hat{\boldsymbol{\beta}})$. Assuming that \mathbf{V} is of full rank (see Section 3.7), \mathbf{V}^{-1} exists and we can decompose it in the manner of Equation 3.38) such that

$$\mathbf{V} = \mathbf{P}'\mathbf{P}.$$

Using \mathbf{P} to premultiply the linear model, we get

$$\mathbf{P}\mathbf{y} = \mathbf{P}\mathbf{X}\boldsymbol{\beta} + \mathbf{P}\mathbf{e} \text{ or}$$

$$\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta} + \mathbf{e}^*.$$

What are the properties of the new error term, \mathbf{e}^* ? According to Theorem (4.9) we have

$$\begin{aligned} V(\mathbf{e}^*) &= \mathbf{P}[\sigma^2 \mathbf{V}] \mathbf{P}' \\ &= \sigma^2 \mathbf{P}[\mathbf{P}'\mathbf{P}]^{-1} \mathbf{P}' \end{aligned}$$

and since \mathbf{V} is of full rank, \mathbf{P} is square and also of full rank so we can say that

$$V(\mathbf{e}^*) = \sigma^2 \mathbf{P} \mathbf{P}^{-1} (\mathbf{P}')^{-1} \mathbf{P}' = \sigma^2 \mathbf{I}.$$

While we cannot believe in the Gauss-Markov assumption with \mathbf{e} , we can with \mathbf{e}^* ! Rather than minimizing $\mathbf{e}'\mathbf{e}$ as in OLS, we should minimize

$$\mathbf{e}^{*'}\mathbf{e}^* = \mathbf{e}'\mathbf{P}'\mathbf{P}\mathbf{e} = \mathbf{e}'\mathbf{V}^{-1}\mathbf{e}$$

instead. Doing so, we pick our objective function as

$$\begin{aligned} f &= \mathbf{e}^{*'}\mathbf{e}^* = (\mathbf{y}^* - \mathbf{X}^*\boldsymbol{\beta})'(\mathbf{y}^* - \mathbf{X}^*\boldsymbol{\beta}) \\ &= \mathbf{y}^{*'}\mathbf{y}^* - 2\mathbf{y}^{*'}\mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}^{*'}\mathbf{X}^*\boldsymbol{\beta}. \end{aligned}$$

In order to minimize f , we should set $\partial f / \partial \boldsymbol{\beta} = \mathbf{0}$ and solve for $\boldsymbol{\beta}$, as we will now do:

$$\frac{\partial f}{\partial \boldsymbol{\beta}} = 2\mathbf{X}^{*'} \boldsymbol{\beta} - 2\mathbf{X}^{*'} \mathbf{y}^* = \mathbf{0}$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{*'} \mathbf{X}^*)^{-1} \mathbf{X}^{*'} \mathbf{y}^*$$

and of course we end up with the usual formula, but using the transformed data matrices \mathbf{X}^* and \mathbf{y}^* . Substituting back $\mathbf{P}\mathbf{X} = \mathbf{X}^*$ and $\mathbf{P}\mathbf{y} = \mathbf{y}^*$, we have

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{X}^{*'} \mathbf{X}^*)^{-1} \mathbf{X}^{*'} \mathbf{y}^* \\ &= (\mathbf{X}' \mathbf{P}' \mathbf{P} \mathbf{X})^{-1} \mathbf{X}' \mathbf{P}' \mathbf{P} \mathbf{y} \\ &= (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{y}. \end{aligned}$$

The variance of this estimator is

$$\begin{aligned} \mathbf{V}(\hat{\boldsymbol{\beta}}) &= (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} (\sigma^2 \mathbf{V}) \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1}. \end{aligned}$$

This is all fine and dandy, but since \mathbf{V} contains $\frac{n(n+1)}{2}$ unique elements, it is necessary that most of them be known a priori. But there is another identification issue. Since $\mathbf{V}(\mathbf{e}) = \sigma^2 \mathbf{V}$, we cannot uniquely identify both σ^2 and the elements of \mathbf{V} . That this is so can be seen by simply multiplying σ^2 by some value a and then dividing all of the elements of \mathbf{V} by a and the model is unchanged. What we do is to set $\text{Tr}(\mathbf{V}) = \text{Tr}(\mathbf{I}) = n$.

We can estimate σ^2 using

$$s^2 = \frac{\text{SS}_{\text{Error}}}{n - k}$$

where

$$\text{SS}_{\text{Error}} = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}).$$

We can construct t -statistics that allow us to test hypotheses of the form

$$H_0: \beta_i = 0$$

using the i th diagonal element of $s^2 (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1}$ in the denominator to create a t . One can also test one degree of freedom hypotheses such as

$$\mathbf{a}' \boldsymbol{\beta} = c$$

using

$$\hat{t} = \frac{\mathbf{a}'\hat{\boldsymbol{\beta}} - c}{s^2 \mathbf{a}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{a}}$$

and for more complex hypotheses of the form

$$H_0: \mathbf{A}\boldsymbol{\beta} - \mathbf{c} = \mathbf{0}$$

we use

$$SS_H = (\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{c})'[\mathbf{A}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{A}]^{-1}(\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{c})$$

to construct an F ratio numerator (with degrees of freedom equal to the number of rows in \mathbf{A}), with s^2 in the denominator (with $n - k$ degrees of freedom).

One area that we can apply GLS to occurs when the error in a regression model is not independent because the data are collected over time, leading to autocorrelated error. This may happen if we are analyzing the behavior of a particular firm, a particular store, category sales, purchases in a particular geographic region, and in many other cases in marketing where we look at data not collected across independent subjects. The next section speaks to that application of GLS.

17.5 Autocorrelated Error

When we collect data over time, rather than across a set of independent individuals, we run the risk that the error from observations that are closer together in time will be more closely related than a pair of errors that are farther apart from time. For example, looking at industry-wide sales of motor homes, we may fail to include every possible exogenous factor that there could be in a model for such sales. In fact, unless our model fits without error, it must be the case that we have omitted some important independent variables. Now, if any of those independent variables that did not find their way into our regression equation vary in a systematic way over time, for example, the weather, or consumer confidence, then the errors in our regression equation will also vary systematically over time. Of course, that would violate the Gauss-Markov assumption and necessitate some counter measure. Such as GLS. To begin to sketch this out, consider observation t on the dependent variable and the model for it,

$$\mathbf{y}_t = \mathbf{x}'_t \boldsymbol{\beta} + \mathbf{e}_t \quad (17.5)$$

where, needless to say, \mathbf{x}'_t represents the t -th row of the matrix of independent variables, \mathbf{X} . Given the argument in the preceding paragraph, we note that values of \mathbf{e}_t are not independently distributed, but rather, adjacent observations follow the model

$$\mathbf{e}_t = \rho \mathbf{e}_{t-1} + \boldsymbol{\varepsilon}_t \quad (17.6)$$

In this context, the values $\boldsymbol{\varepsilon}_t$ represent an error for the error, if you will. We would also be remiss if we did not point out that a requirement of the model is that $|\rho| < 1$. The distribution of the $\boldsymbol{\varepsilon}_t$ is characterized as

$$\varepsilon_t \sim N(\mathbf{0}, \sigma^2 \mathbf{I}), \quad (17.7)$$

which is to say that the the ε_t , unlike the e_t , are independently distributed. They behave like a white noise process, in summary. Repeating our model for the error,

$$e_t = \rho e_{t-1} + \varepsilon_t \quad (17.8)$$

we see that, since e_{t-1} appears in the right hand side, the model for e_{t-1} would contain e_{t-2} in it. Making that obvious substitution, we get

$$\begin{aligned} e_t &= \rho(\rho e_{t-2} + \varepsilon_{t-2}) + \varepsilon_t \\ &= \rho[\rho(\rho e_{t-3} + \varepsilon_{t-2}) + \varepsilon_{t-1}] + \varepsilon_t \\ &= \dots \end{aligned}$$

At this point the pattern should be obvious. Continuing the process of substitution, we end up with

$$\begin{aligned} e_t &= \varepsilon_t + \rho \varepsilon_{t-1} + \rho^2 \varepsilon_{t-2} + \rho^3 \varepsilon_{t-3} + \dots \\ &= \sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i}. \end{aligned} \quad (17.9)$$

This last equation will look quite familiar if you have looked at Equation (15.17) or (18.15), being an infinite series. Now we wish to find out the expectation of the error. To determine the expectation of e_t from Equation (17.9), we keep in mind that the expectation of a sum is equal to the sum of the expectations [Equation (4.4)], and that therefore

$$\begin{aligned} E(e_t) &= E\left[\sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i} \right] \\ &= \sum_{i=0}^{\infty} \rho^i E(\varepsilon_{t-i}) = 0, \end{aligned} \quad (17.10)$$

since ρ is a constant parameter that describes the population and by assumption $E(e_t) = 0$ for all t . Now we wish to figure out the variance of e_t , that is $V(e_t) = E[e_t - E(e_t)]^2$ according to Equation (4.7). Given that $E(e_t) = 0$, which we have just shown in Equation (17.10), we will only need to figure out $E(e_t^2)$. That will be made easier by recalling that all cross terms of the form $E(\varepsilon_{t_i}, \varepsilon_{t_j})$ will vanish as the ε_t are presumed independent, and that $a^0 = 1$ for any value a . So, squaring the second line of Equation (17.10) we have

$$\begin{aligned}
E(e_t^2) &= E(\varepsilon_t^2) + \rho^2 E(\varepsilon_{t-1}^2) + \rho^4 E(\varepsilon_{t-2}^2) + \dots \\
&= \sigma^2 + \rho^2 \sigma^2 + \rho^4 \sigma^2 + \dots \\
&= (1 + \rho^2 + \rho^4 + \dots) \sigma^2.
\end{aligned}$$

So in the above equation we have an infinite series of the form $1 + \rho^2 + \rho^4 + \dots$, call it s such that

$$\begin{aligned}
s &= 1 + \rho^2 + \rho^4 + \rho^8 + \dots \\
\rho^2 s &= \rho^2 + \rho^4 + \rho^8 + \dots \\
s - \rho^2 s &= 1
\end{aligned}$$

so that

$$s = \frac{1}{1 - \rho^2}. \quad (17.11)$$

Putting all of this together, we conclude that

$$E(e_t^2) = \sigma_e^2 = \frac{\sigma^2}{1 - \rho^2}. \quad (17.12)$$

To explore the covariances between e_t and e_{t-j} , we begin with $j = 1$. By definition, the covariance between e_t and e_{t-1} is given by

$$E(e_t e_{t-1}) = E[(\varepsilon_t + \rho \varepsilon_{t-1} + \rho^2 \varepsilon_{t-2} + \dots)(\varepsilon_{t-1} + \rho \varepsilon_{t-2} + \rho^2 \varepsilon_{t-3} + \dots)].$$

Looking at the right hand side of that equation, we will factor the ρ that appears in the left parentheses to give us

$$E(e_t e_{t-1}) = E\left\{[\varepsilon_t + \rho(\varepsilon_{t-1} + \rho \varepsilon_{t-2} + \dots)](\varepsilon_{t-1} + \rho \varepsilon_{t-2} + \rho^2 \varepsilon_{t-3} + \dots)\right\}$$

Now, the two terms in the two parentheses on the right hand side are identical. We can write them as a single term squared. What's more, you will notice an ε_t all alone on the left of the right hand side. Its expectation is zero, and since there are no other values ε_t on the right hand side, the covariance of it and every other term will be zero. It thus vanishes without a trace. Rewriting, that gives us

$$E(e_t e_{t-1}) = \rho E[(\varepsilon_{t-1} + \rho \varepsilon_{t-2} + \rho^2 \varepsilon_{t-3} + \dots)^2]. \quad (17.13)$$

You will note that since ρ is a constant it can pass through the expectation operator [for a review, take a peek at Equation (4.5)]. Again, we remind you that $E(\varepsilon_t, \varepsilon_{t-1})$, that is the covariance between two different values of the ε_t are zero by the assumption of Equation (17.7). However, just

because the ε_t are independent does not mean that the e_t are. In fact, looking at Equation (17.13), we are almost ready to make a conclusion about the autocovariance of the e_t . The part in parentheses is just the model for the e_t i.e. Equation (17.8). Its expectation squared must then be the variance of e_t , so that

$$E(e_t e_{t-1}) = \rho \sigma_\varepsilon^2. \quad (17.14)$$

Following the same reasoning we find that the

$$\text{Cov}(e_t, e_{t-j}) = \rho^j \sigma_\varepsilon^2. \quad (17.15)$$

Summarizing, we can say that the variance matrix of the e_t is $\sigma_\varepsilon^2 \mathbf{V} = \frac{\sigma_\varepsilon^2}{1-\rho^2} \mathbf{V}$ with

$$\mathbf{V} = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{bmatrix}.$$

Thus the GLS approach only needs to estimate two error related parameters, ρ and σ_ε^2 . In the *Cochrane-Orcutt Iterative Procedure* we pick a starting value for ρ , calculate $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$, then pick ρ in such a way as to minimize $\mathbf{e}'\mathbf{e}$ while holding $\hat{\boldsymbol{\beta}}$ fixed, and then re-estimate $\hat{\boldsymbol{\beta}}$ holding ρ fixed. One alternates between those two least squares steps until there is convergence. More general specifications of the nature of the error are possible. While in this section we have discussed a single autoregressive parameter, in much the same way that we talk about an AR(1) model in Section 18.4, just like with ARIMA models, you can have AR(2) or other processes.

17.6 Testing for Autocorrelated Error

Durbin and Watson (1950) proposed using

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \quad (17.16)$$

as a test statistic for autocorrelated residuals. Here, the hypothesis being tested is $H_0: \rho = 0$. For positive autocorrelation the numerator will be small, while for negative autocorrelation the numerator will tend to be large. There is an upper limit (d_u) and a lower limit (d_l) for this statistic such that

$$\text{if } d < d_l, \text{ reject } H_0,$$

if $d > d_u$, fail to reject H_0 , and if

if $d_l < d < d_u$

the test is inconclusive.

17.7 Lagged Variables

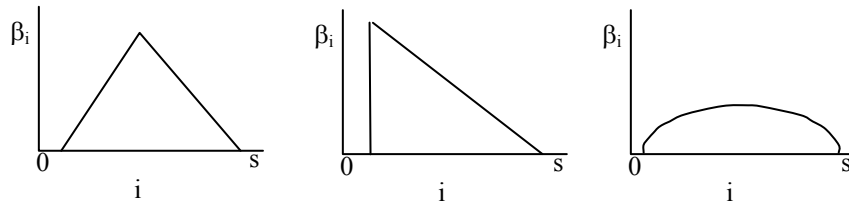
Suppose it is the case that consumers do not immediately react to a change in a marketing variable. In that case we would expect to see a relationship like the one below,

$$y_t = \beta_0 + x_{t-1}\beta_1 + e_t$$

or perhaps their reaction begins immediately but is distributed across several time periods, as in

$$y_t = \beta_0 + x_{t-1}\beta_1 + x_{t-2}\beta_2 + \dots + x_{t-s}\beta_s + e_t.$$

This is reasonable under many real life marketing situations. For example, the consumer may not immediately learn about a change in the market. Or perhaps, they are encumbered in their actions by inventory already on hand. However realistic this may be, there are unfortunately some problems with this approach. For one thing, what should "s" be? For another, we will be losing a degree of freedom for each lag, which is to say that the model is not very parsimonious. Finally, successive values of x might well be highly correlated, so that multicollinearity rears its head. What we can do is impose some sort of a priori structure on the values of the β_i . A graph of some possible structural assumptions is below:



Of course, any function can be represented by a polynomial of sufficiently high degree, fact exploited in ANOVA in Section 7.6. We can approximate, for example, a system with $s = 7$ lags with a polynomial of the third degree:

$$\beta_0 = a_0$$

$$\beta_1 = a_0 + a_1 + a_2 + a_3$$

$$\beta_2 = a_0 + 2a_1 + 4a_2 + 8a_3$$

$$\beta_3 = a_0 + 3a_1 + 9a_2 + 27a_3$$

... = ...

$$\beta_7 = a_0 + 7a_1 + 49a_2 + 343a_3$$

The reader will perhaps recognize that the coefficients for the a values are constant in the first column, linear in the second, quadratic in the third and cubic in the fourth. If we substitute these equations back into the model for $s = 7$, i. e.

$$y_t = \beta_0 + x_{t-1}\beta_1 + x_{t-2}\beta_2 + \dots + x_{t-7}\beta_7 + e_t$$

we get after collecting the a_i terms

$$y_t = \beta_0 + (x_t + x_{t-1} + x_{t-2} + \dots + x_{t-7})a_0 + (x_t + 2x_{t-1} + 3x_{t-2} + \dots + 7x_{t-7})a_1 + (x_t + 4x_{t-1} + 9x_{t-2} + \dots + 49x_{t-7})a_2 + (x_t + 8x_{t-1} + 27x_{t-2} + \dots + 343x_{t-7})a_3 + e_t \quad (17.17)$$

which is equivalent to a model with

$$y_t = \beta_0 + w_0a_0 + w_1a_1 + w_2a_2 + w_3a_3 + e_t \quad (17.18)$$

where $w_0 = x_t + x_{t-1} + x_{t-2} + \dots + x_{t-7}$, and the other w values are defined as above in Equation (17.17). This is known as *Almon's Scheme*. If we define

$$\mathbf{K} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \\ \dots & \dots & \dots & \dots \\ 1 & 7 & 27 & 343 \end{bmatrix}$$

using the coefficients for the x 's, then

$$V(\hat{\boldsymbol{\beta}}) = \sigma^2 \mathbf{K}(\mathbf{W}'\mathbf{W})^{-1} \mathbf{K}' \quad (17.19)$$

lets you test the value s of the maximum lag, while

$$V(\hat{\mathbf{a}}) = \sigma^2 (\mathbf{W}'\mathbf{W})^{-1} \quad (17.20)$$

lets you test the degree of the polynomial required to represent the lag structure.

While Almon's Scheme is quite compelling, another approach was proposed by Koyck, who used a geometric sequence. Koyck started with the infinite sequence

$$y_t = x_t\beta_0 + x_{t-1}\beta_1 + x_{t-2}\beta_2 + \dots + e_t \quad (17.21)$$

Now, assume that the β values are all of the same sign, and that

$$\sum_{i=0}^{\infty} \beta_i = c < \infty. \quad (17.22)$$

We now introduce the *backshift operator*, B , which is also prominently featured in Chapter 18. We define

$$Bx_t = x_{t-1}. \quad (17.23)$$

Of course, one can also say

$$BBx_t = B^2x_t = x_{t-2} \quad (17.24)$$

and so forth with $B^j x_t = x_{t-j}$. Given our two assumptions of Equations (17.21) and (17.22), we can rewrite the model as

$$\begin{aligned} y_t &= x_t\beta_0 + x_{t-1}\beta_1 + x_{t-2}\beta_2 + \cdots + e_t \\ &= \beta[w_0x_t + w_1x_{t-1} + w_2x_{t-2} + \cdots] + e_t \end{aligned}$$

where $w_i \geq 0$ for $i = 0, 1, 2, \dots, \infty$ and $\sum_{i=0}^{\infty} w_i = 1$. Given that, we can now rewrite the above equation as

$$= \beta[w_0 + w_1B + w_2B^2 + \cdots]x_t + e_t.$$

Now we introduce the major assumption of the Koyck scheme. The w 's have a geometric relationship to each other as in

$$w_i = (1 - \lambda)\lambda^i \quad (17.25)$$

where $0 < \lambda < 1$. In that case

$$\begin{aligned} w_0 + w_1B + w_2B^2 + \cdots &= (1 - \lambda)(1 + \lambda B + \lambda^2 B^2 + \cdots) \\ &= (1 - \lambda) \frac{1}{1 - \lambda B}. \end{aligned}$$

The fraction on the right hand side of the line immediately above is a consequence of the logic worked out in Equation (17.11) where we previously worked out the solution to an infinite series just like the one above. The upshot is that we can now write the model

$$\begin{aligned} y_t &= \frac{\beta(1 - \lambda)}{1 - \lambda B} x_t + e_t \\ (1 - \lambda B)y_t &= \beta(1 - \lambda)x_t + (1 - \lambda B)e_t \\ y_t - \lambda B y_t &= \beta(1 - \lambda)x_t + e_t - \lambda B e_t \\ y_t &= \beta(1 - \lambda)x_t + \lambda y_{t-1} + (e_t - \lambda e_{t-1}) \end{aligned} \quad (17.26)$$

As you can see, Koyck's scheme is characterized by autocorrelated error and lagged endogenous variables on the right hand side. Why would that be? Is there any marketing theory in which that would make sense? We will be finding out shortly.

17.8 Partial Adjustment by Consumers

The partial adjustment model posits that the optimal value of the y variable, y^* , might depend on x . For example, y could be an amount spent on our brand and x is income. As the consumer wants to make an optimal choice, and if the relationship is linear, we would have

$$\tilde{y}_t = \beta_0 + x_t \beta_1, \quad (17.27)$$

but due to less than perfect information about the market, inventory considerations, inertia, or the cognitive costs of change, the consumer can only adjust a certain proportion of the way from his or her previous value, y_{t-1} , to the optimal value at \tilde{y}_t . In mathematical terms,

$$y_t - y_{t-1} = \gamma(\tilde{y}_t - y_{t-1}) + e_t \quad (17.28)$$

with $0 < \gamma < 1$. Substituting Equation (17.27) into Equation (17.28), we see that

$$y_t = \beta_0 \gamma + \beta_1 \gamma x_t + (1 - \gamma)y_{t-1} + e_t$$

which bears a resemblance to Koyck's scheme, only here we have an intercept, and the error is not autocorrelated.

17.9 Adaptive Adjustment by Consumers

Another way that a similar equation may come about is through consumers adapting their expectations. Define \tilde{x}_t as the expected level of x , and assume that some key consumer behavior depends on \tilde{x}_t . The value \tilde{x}_t could be the best guess of the price of a good, something to do with its availability in the market, and so forth. The consumer's behavior should then appear as below

$$y_t = \beta_0 + \tilde{x}_t \beta_1 + e_t. \quad (17.29)$$

Now if we assume that the expectations are updated by a fraction of the discrepancy between the current observation and the previous expectation, we get

$$\tilde{x}_t - \tilde{x}_{t-1} = \delta(x_t - \tilde{x}_{t-1})$$

$$\tilde{x}_t = \delta x_t - \delta \tilde{x}_{t-1} + \tilde{x}_{t-1}$$

$$= \delta x_t + (1 - \delta)\tilde{x}_{t-1}$$

$$\tilde{x}_t - (1 - \delta)\tilde{x}_{t-1} = \delta x_t.$$

Define $\lambda = 1 - \delta$. Then starting with the last line of the above equation,

$$\tilde{x}_t - \lambda \tilde{x}_{t-1} = \delta x_t$$

$$(1 - \lambda B)\tilde{x}_t = \delta x_t$$

$$\tilde{x}_t = \frac{\delta}{1 - \lambda B} x_t.$$

Finally, substituting this result into the model of Equation (17.29), we find out that

$$\begin{aligned} y_t &= \beta_0 + \tilde{x}_t \beta_1 + e_t \\ &= \beta_0 + \frac{\delta \beta_1}{1 - \lambda B} x_t + e_t \end{aligned}$$

which is the same as the Koyck Scheme of Equation (17.26). As far as estimating these models, OLS is not consistent [see Equation (5.11) for a definition of consistency] if there is autocorrelated error. You can use a two-stage estimator substituting \hat{y}_{t-1} for y_{t-1} . You can also use x_{t-1} as an instrument for y_{t-1} .

17.10 Pooling Time Series and Cross Section Data

Suppose we had for a particular sales region, call it region 1, the model

$$\mathbf{y}^{(1)} = \mathbf{X}^{(1)}\boldsymbol{\beta}^{(1)} + \mathbf{e}^{(1)}$$

where $\mathbf{y}^{(1)}$ is the $T \cdot 1$ vector of observations on the response of the market in that region and $\mathbf{X}^{(1)}$ is a matrix of marketing instruments including such factors as advertising effort, and so forth. In region 1 we might have k_1 such instruments. Data have been collected from time period 1 through time period T and analogously, in region 2, we have done the same thing but with k_2 different independent variables:

$$\mathbf{y}^{(2)} = \mathbf{X}^{(2)}\boldsymbol{\beta}^{(2)} + \mathbf{e}^{(2)}.$$

These two regression are only *seemingly unrelated* because we would expect $\text{Cov}(e_t^{(1)}, e_t^{(2)}) \neq 0$ as long as the two regions are interconnected economically. In point of fact, there are hardly two regions left on Earth that are not interconnected economically. The covariance identified above is called *contemporaneous* for obvious reasons. Alternating the regions so as to keep contemporaneous observations next to each other as we move from row to row, we could combine the two models as below:

$$\begin{bmatrix} y_1^{(1)} \\ y_1^{(2)} \\ y_2^{(1)} \\ y_2^{(2)} \\ \dots \\ y_T^{(1)} \\ y_T^{(2)} \end{bmatrix} = \begin{bmatrix} 1 & x_{11}^{(1)} & \dots & x_{1k_1}^{(1)} & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & x_{11}^{(2)} & \dots & x_{1k_2}^{(2)} \\ 1 & x_{21}^{(1)} & \dots & x_{2k_1}^{(1)} & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & x_{21}^{(2)} & \dots & x_{2k_2}^{(2)} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & x_{T1}^{(1)} & \dots & x_{Tk_1}^{(1)} & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & x_{T1}^{(2)} & \dots & x_{Tk_2}^{(2)} \end{bmatrix} \begin{bmatrix} \beta_0^{(1)} \\ \beta_1^{(1)} \\ \dots \\ \beta_{k_1}^{(1)} \\ \beta_0^{(2)} \\ \beta_1^{(2)} \\ \dots \\ \beta_{k_2}^{(2)} \end{bmatrix} + \begin{bmatrix} e_1^{(1)} \\ e_1^{(2)} \\ \dots \\ e_2^{(1)} \\ e_2^{(2)} \\ \dots \\ e_T^{(1)} \\ e_T^{(2)} \end{bmatrix}$$

so that our model is now simply $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, with \mathbf{y} having n times T rows, assuming n regions and T observations across time. For now we will continue to assume that $n = 2$. With contemporaneous covariance we can model the error covariance matrix as

$$V(\mathbf{e}) = \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\Sigma} \end{bmatrix} = {}_T\mathbf{I}_T \otimes {}_n\boldsymbol{\Sigma}_n$$

where each null matrix and each copy of $\boldsymbol{\Sigma}$ is $2 \cdot 2$, and the Kronecker product operator \otimes is defined in Section 1.10 (and elaborated on in Section 8.6). In fact, we might note the similarity between this and the error structure for the Multivariate General Linear model in Equation (8.36) and Equation (8.37). The difference is that in the General Linear Model, all dependent variables have the same set of independent variables. In this case, the seemingly unrelated model, we will use GLS, in which the error matrix, usually notated \mathbf{V} , will be $\mathbf{I} \otimes \boldsymbol{\Sigma}$ as you can see now

$$\hat{\boldsymbol{\beta}} = [\mathbf{X}'(\mathbf{I} \otimes \boldsymbol{\Sigma})^{-1} \mathbf{X}]^{-1} \mathbf{X}'(\mathbf{I} \otimes \boldsymbol{\Sigma})^{-1} \mathbf{y} \quad (17.30)$$

and

$$V(\hat{\boldsymbol{\beta}}) = [\mathbf{X}'(\mathbf{I} \otimes \boldsymbol{\Sigma})^{-1} \mathbf{X}]^{-1}.$$

We can apply a two step procedure in which we use OLS and estimate $\mathbf{S} = \hat{\boldsymbol{\Sigma}}$, then we apply GLS using \mathbf{S} as a substitute for $\boldsymbol{\Sigma}$ in Equation (17.30). One could also use Maximum Likelihood.

Another way to deal with the pooling problem is to use dummy variables, as we did in Section 7.3 in Equation (7.5) in the context of the analysis of variance. Of course, one could also use effect or orthogonal coding. If we set up dummies for each time period and each region, this purges the dependent variable of all variance associated with regions and time. This is a somewhat drastic approach, since some of the variance of interest will get thrown out with the bath water. A less drastic approach is to treat time and cross-sections like a random effect in ANOVA. Suppose that our error term is composed of

$$e_{ij} = \alpha_i + \varphi_t + \varepsilon_{it}$$

with $V(\alpha_i) = \sigma_\alpha^2$, $V(\varphi_t) = \sigma_\varphi^2$ and $V(\varepsilon_{it}) = \sigma_\varepsilon^2$. Given that the data from each region are next to each other in the \mathbf{y} and \mathbf{e} vectors,

$$V(\mathbf{e}) = \sigma_{\alpha}^2 (\mathbf{1}_T \mathbf{1}'_T \otimes \mathbf{I}_n) + \sigma_{\phi}^2 (\mathbf{I}_T \otimes \mathbf{1} \mathbf{1}'_n) + \sigma_{\epsilon}^2 \mathbf{I}_{nT}$$

and we can apply a first step ANOVA to estimate the random variances σ_{α}^2 and σ_{ϕ}^2 while in the second step we use those estimates in GLS. Parks has proposed a model like Zellner's seemingly unrelated regressions but with autoregressive error, and De Silva has suggested a model with variance components for the cross sections but autoregression for the time component.

References

Johnston, J. (1972) *Econometric Methods, Second Edition*. New York: McGraw-Hill.

Theil, Henri (1971) *Principles of Econometrics*. New York: Wiley.

Maddala, G. S. (1977) *Econometrics*. New York: McGraw-Hill.