

LECTURE 5

Serially Correlated Regression Disturbances

Autoregressive Disturbance Processes

The interpretation which is given to the disturbance term of a regression model depends upon the context in which the analysis is conducted. Imagine that we are fitting the regression equation

$$(206) \quad y_t = \beta_0 + x_{t1}\beta_1 + \cdots + x_{tk}\beta_k + \varepsilon_t$$

to a set of economic variables observed through time. Then it is usual to assume that the disturbance ε_t represents the net effect upon the dependent variable y_t of a large number of subsidiary variables which, individually, are of insufficient importance to be included in the systematic part of the model.

It may be reasonable to imagine that, instead of accumulating over time, the effects of the subsidiary variables will tend to cancel each other in any period. Then their overall effect might have a small constant expected value. The inclusion of the intercept term in the regression equation allows us to assume that $E(\varepsilon_t) = 0$; for any nonzero net effect of the subsidiary variables will be absorbed by β_0 .

Economic variables are liable to follow slowly-evolving trends and they are also liable to be strongly correlated with each other. If the disturbance term is indeed compounded from such variables, then we should expect that it too will follow a slowly-evolving trend.

The assumptions of the classical regression model regarding the disturbance term are at variance with these expectations. In the classical model, it is assumed that the disturbances constitute a sequence $\varepsilon(t) = \{\varepsilon_t; t = 0, \pm 1, \pm 2, \dots\}$ of independently and identically distributed random variables such that

$$(207) \quad E(\varepsilon_t \varepsilon_s) = \begin{cases} \sigma^2, & \text{if } t = s; \\ 0, & \text{if } t \neq s. \end{cases}$$

The process which generates such disturbances is often called a white-noise process. A sequence of 50 observations generated by a white-noise process is plotted in figure 1. The sequence is of a highly volatile nature; and its past values are of no use in predicting its future values.

Our task is to find models for the disturbance process which are more in accordance with economic circumstances. Given the paucity of econometric data, we are unlikely to be able to estimate the parameters of complicated models with any degree of precision; and, in econometrics the traditional means of representing the inertial properties of the disturbance process has been to adopt a simple first-order autoregressive model, or AR(1) model, whose equation takes the form of

$$(208) \quad \eta_t = \phi\eta_{t-1} + \varepsilon_t, \quad \text{where} \quad \phi \in (-1, 1).$$

Here it continues to be assumed that ε_t is generated by a white-noise process with $E(\varepsilon_t) = 0$. In many econometric applications, the value of ϕ falls in the more restricted interval $[0, 1)$.

According to this model, the conditional expectation of η_t given η_{t-1} is $E(\eta_t|\eta_{t-1}) = \phi\eta_{t-1}$. That is to say, the expectation of the current disturbance is ϕ times the value of the previous disturbance. This implies that, for a value of ϕ which is closer to unity than to zero, there will be a high degree of correlation amongst successive elements of the sequence $\eta(t) = \{\eta_t; t = 0, \pm 1, \pm 2, \dots\}$. This result is illustrated in figure 2 which gives a sequence of 50 observation on an AR(1) process with $\phi = 0.9$

We can show that the covariance of two elements of the sequence $\eta(t)$ which are separated by τ time periods is given by

$$(209) \quad C(\eta_{t-\tau}, \eta_t) = \gamma_\tau = \sigma^2 \frac{\phi^\tau}{1 - \phi^2}.$$

It follows that variance of the process, which is formally the autocovariance of lag $\tau = 0$, is given by

$$(210) \quad V(\eta_t) = \gamma_0 = \frac{\sigma^2}{1 - \phi^2}.$$

As ϕ tends to unity, the variance increases without bound. In fact, the sequences in figures 1 and 2 share the same underlying white noise-process which has a unit variance; and it is evident that the autocorrelated sequence of figure 2 has the wider dispersion.

To find the correlation of two elements from the autoregressive sequence, we note that

$$(211) \quad \text{Corr}(\eta_{t-\tau}, \eta_t) = \frac{C(\eta_{t-\tau}, \eta_t)}{\sqrt{V(\eta_{t-\tau})V(\eta_t)}} = \frac{C(\eta_{t-\tau}, \eta_t)}{V(\eta_t)} = \frac{\gamma_\tau}{\gamma_0}.$$

5: CORRELATED DISTURBANCES

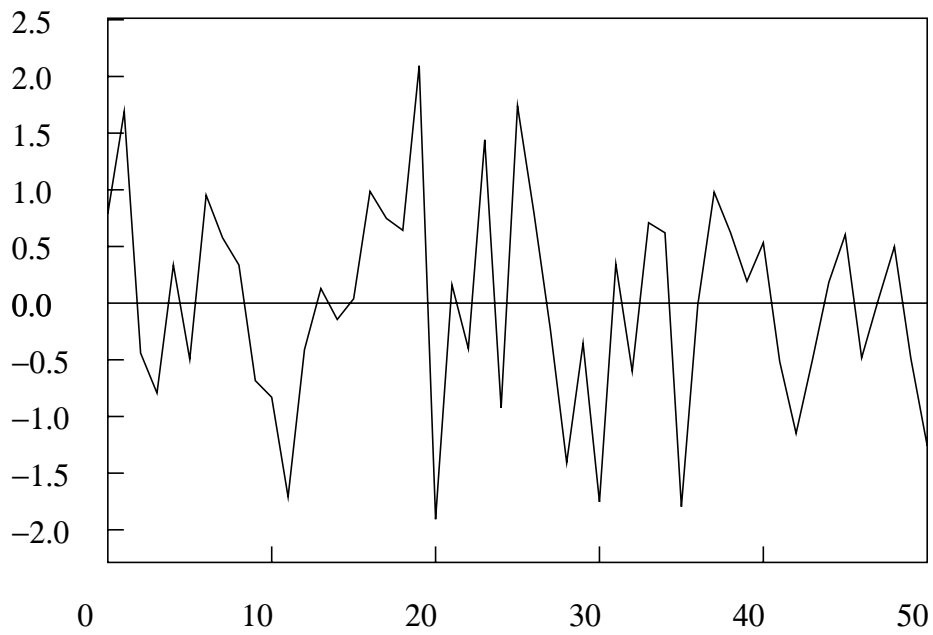


Figure 1. 50 observations on a white-noise process $\varepsilon(t)$ of unit variance.

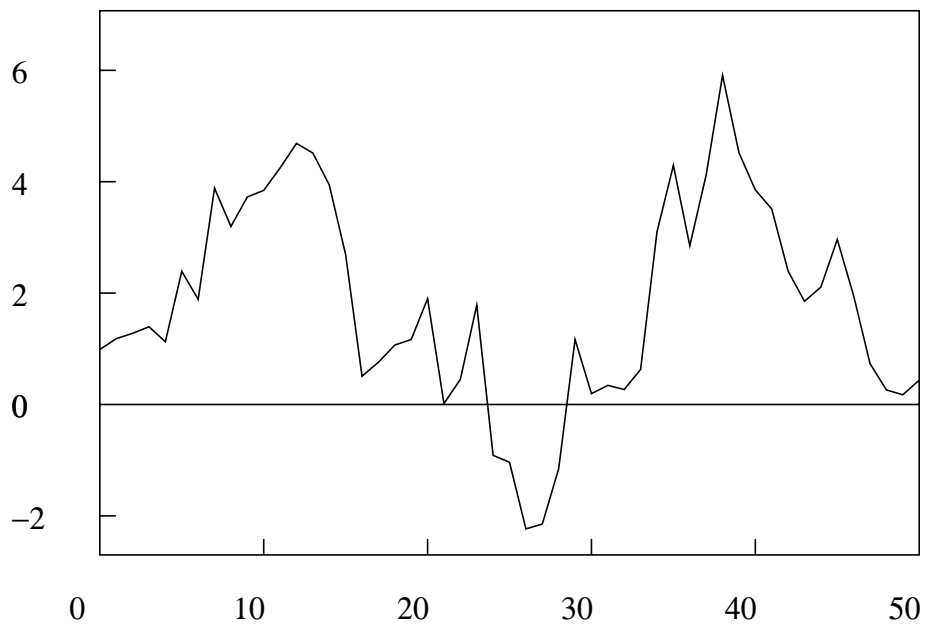


Figure 2. 50 observations on an AR(1) process $\eta(t) = 0.9\eta(t - 1) + \varepsilon(t)$.

This implies that the correlation of the two elements separated by τ periods is just ϕ^τ ; and thus, as the temporal separation increases, the correlation tends to zero in the manner of a convergent geometric progression.

To demonstrate these results, let us consider substituting for $\eta_{t-1} = \phi\eta_{t-2} + \varepsilon_{t-1}$ in the equation under (208) and then substituting for $\eta_{t-2} = \phi\eta_{t-3} + \varepsilon_{t-2}$, and so on indefinitely. By this process we find that

$$\begin{aligned}
 \eta_t &= \phi\eta_{t-1} + \varepsilon_t \\
 &= \phi^2\eta_{t-2} + \varepsilon_t + \phi\varepsilon_{t-1} \\
 &\vdots \\
 (212) \quad &= \{\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \dots\} \\
 &= \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i}.
 \end{aligned}$$

Here the final expression is justified by the fact that $\phi^n \rightarrow 0$ as $n \rightarrow \infty$ in consequence of the restriction that $|\phi| < 1$. Thus we see that η_t is formed as a geometrically declining weighted average of all past values of the sequence $\varepsilon(t)$.

Using this result, we can now write

$$\begin{aligned}
 \gamma_\tau &= C(\eta_{t-\tau}, \eta_t) = E(\eta_{t-\tau}\eta_t) \\
 (213) \quad &= E\left(\left\{\sum_{i=0}^{\infty} \phi^i \varepsilon_{t-\tau-i}\right\}\left\{\sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}\right\}\right) \\
 &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \phi^i \phi^j E(\varepsilon_{t-\tau-i}\varepsilon_{t-j}).
 \end{aligned}$$

But the assumption that $\varepsilon(t)$ is a white-noise process with zero-valued autocovariances at all nonzero lags implies that

$$(214) \quad E(\varepsilon_{t-\tau-i}\varepsilon_{t-j}) = \begin{cases} \sigma^2, & \text{if } j = \tau + i; \\ 0, & \text{if } j \neq \tau + i. \end{cases}$$

Therefore, on using the above conditions in (213) and on setting $j = \tau + i$, we find that

$$\begin{aligned}
 \gamma_\tau &= \sigma^2 \sum_i \phi^i \phi^{i+\tau} = \sigma^2 \phi^\tau \sum_i \phi^{2i} \\
 (215) \quad &= \sigma^2 \phi^\tau \{1 + \phi^2 + \phi^4 + \phi^6 + \dots\} \\
 &= \sigma^2 \frac{\phi^\tau}{1 - \phi^2}.
 \end{aligned}$$

5: CORRELATED DISTURBANCES

This establishes the result under (209).

The Algebra of the Lag Operator

The analysis of stochastic dynamic processes of a more complicated sort than the first-order autoregressive process can pose problems which are almost insurmountable without the aid of more powerful algebraic techniques. One of the techniques which is widely used is the calculus of operators.

The essential operator in the context of discrete-time stochastic systems is the so-called lag operator L . When this is applied to the sequence $x(t) = \{x_t, t = 0, \pm 1, \pm 2, \dots\}$, the effect is to replace each element by its predecessor. That is to say,

$$(216) \quad \begin{aligned} x(t) &= \{\dots, x_{-2}, x_{-1}, x_0, x_1, x_2, \dots\} \quad \text{becomes} \\ x(t-1) &= \{\dots, x_{-3}, x_{-2}, x_{-1}, x_0, x_1, \dots\}. \end{aligned}$$

Using the summary notation of the lag operator, this transformation is expressed by writing

$$(217) \quad Lx(t) = x(t-1).$$

Now, $L\{Lx(t)\} = Lx(t-1) = x(t-2)$; so it makes sense to define L^2 by $L^2x(t) = x(t-2)$. More generally, $L^kx(t) = x(t-k)$ and, likewise, $L^{-k}x(t) = x(t+k)$. For the sake of completeness, we also define the identity operator $I = L^0$ whereby $Ix(t) = x(t)$.

Other operators are the difference operator $\nabla = I - L$ which has the effect that

$$(218) \quad \nabla x(t) = x(t) - x(t-1),$$

the forward-difference operator $\Delta = L^{-1} - I$, and the summation operator $S = (I - L)^{-1} = \{I + L + L^2 + \dots\}$ which has the effect that

$$(219) \quad Sx(t) = \sum_{i=0}^{\infty} x(t-i).$$

In general, we can define polynomials of the lag operator of the form $\alpha(L) = \alpha_0 + \alpha_1L + \dots + \alpha_pL^p$ having the effect that

$$(220) \quad \alpha(L)x(t) = \alpha_0x(t) + \alpha_1x(t-1) + \dots + \alpha_px(t-p).$$

There is a considerable advantage in using the lag operator when a summary notation is needed. However, the real importance of the device stems

from the fact that we may employ the ordinary algebra of polynomials and rational functions for the purpose of manipulating lag-operator polynomials and their ratios.

Consider, for example, the following expansion:

$$(221) \quad \frac{1}{1 - \phi z} = \{1 + \phi z + \phi^2 z^2 + \dots\}.$$

The expression on the RHS, which is a geometric series, may be obtained by a variety of means including long division and the binomial expansion. Conversely, the sum of the geometric progression is obtained via the following subtraction:

$$(222) \quad \begin{array}{r} S = 1 + \phi z + \phi^2 z^2 + \dots \\ \phi z S = \quad \phi z + \phi^2 z^2 + \dots \\ \hline S - \phi z S = 1. \end{array}$$

This gives $S(1 - \phi z) = 1$, from which $S = (1 - \phi z)^{-1}$.

The symbol z , which is described as an algebraic indeterminate, usually stands for a real or a complex number. However, it is perfectly in order to replace z by the lag operator L . In that case, equation (221) becomes

$$(223) \quad \frac{1}{1 - \phi L} = \{1 + \phi L + \phi^2 L^2 + \dots\}.$$

To see the use of this, consider the equation (212). We may replace the generic elements η_t , η_{t-1} and ε_t by the corresponding sequences so as to obtain the following equation for the AR(1) process:

$$(224) \quad \eta(t) = \phi \eta(t - 1) + \varepsilon(t).$$

With the use of the lag operator, this may be rendered as

$$(225) \quad (I - \phi L)\eta(t) = \varepsilon(t).$$

On multiplying both sides by $(I - \phi L)^{-1}$ and using the expansion of (223), the latter becomes

$$(226) \quad \begin{aligned} \eta(t) &= \frac{1}{1 - \phi L} \varepsilon(t) \\ &= \{1 + \phi L + \phi^2 L^2 + \dots\} \varepsilon(t) \\ &= \{\varepsilon(t) + \phi \varepsilon(t - 1) + \phi^2 \varepsilon(t - 2) + \dots\}. \end{aligned}$$

5: CORRELATED DISTURBANCES

Thus the result which was given under (212) may be obtained without recourse to a process of substitution.

Serial Correlation in the Regression Disturbances

We shall now consider the problems which arise from the serial correlation of the disturbances of a regression model, as well as the means of treating these problems.

Recall that, amongst the features of the Classical Linear Regression Model, there is the assumption that the elements of the disturbance sequence $\varepsilon(t) = \{\varepsilon_t; t = 0, \pm 1, \pm 2 \dots\}$ are independently and identically distributed (i.i.d) with zero mean, such that

$$(227) \quad E(\varepsilon_t) = 0, \quad \text{for all } t \quad \text{and} \quad C(\varepsilon_t, \varepsilon_s) = \begin{cases} \sigma^2, & \text{if } t = s; \\ 0, & \text{if } t \neq s. \end{cases}$$

This may be described as the null hypothesis concerning the disturbances; and it specifically excludes any serial correlation of the disturbances.

If all of the other assumptions of the Classical Model are granted, then the presence of serial correlation amongst the disturbances will affect neither the unbiasedness of the ordinary least-squares estimates nor their consistency. However, the standard errors and confidence intervals which are calculated under the i.i.d assumption will be incorrect, which will affect any inferences which we might make.

To overcome the problems of serial correlation, we should have to build a model for the disturbance process as an integral part of our regression strategy. First, however, we have to detect the presence of serial correlation.

The residuals of an ordinary least-squares regression represent estimates of the disturbances from which they may differ significantly; and the misrepresentation of the disturbances can affect our ability to detect serial correlation.

To demonstrate the differences between the disturbances and the residuals, we may resort to matrix algebra. Recall that the fitted regression equation is represented by

$$(228) \quad y = X\hat{\beta} + e, \quad \text{where} \quad \hat{\beta} = (X'X)^{-1}X'y.$$

The residual vector is

$$(229) \quad e = y - X\hat{\beta} = y - X(X'X)^{-1}X'y;$$

and, on defining $P = X(X'X)^{-1}X'$, we can write this as

$$(230) \quad e = (I - P)y = (I - P)(X\beta + \varepsilon).$$

Now observe that $(I - P)X = 0$. It follows that

$$(231) \quad e = (I - P)\varepsilon.$$

This shows that the residuals represent a transformation of the disturbances; and, since the matrix $I - P$ of the transformation is non-invertible, it is impossible to recover the disturbances. Next observe that $PP' = P$ and that $(I - P)(I - P)' = (I - P)^2 = (I - P)$. It follows that the variance-covariance or dispersion matrix of the residuals is given by

$$(232) \quad \begin{aligned} D(e) &= E\{(I - P)\varepsilon\varepsilon'(I - P)'\} \\ &= (I - P)E(\varepsilon\varepsilon')(I - P)' \\ &= (I - P)\{\sigma^2 I\}(I - P)' = \sigma^2(I - P). \end{aligned}$$

This is to be compared with the variance-covariance matrix of the disturbances which is $D(\varepsilon) = \sigma^2 I$. Thus we see that the residuals will be serially correlated even when the disturbances are independently and identically distributed. Nevertheless, if the remainder of regression model, which is to say its systematic part, is correctly specified, then we shall find that $e \rightarrow \varepsilon$ as $T \rightarrow \infty$. Likewise, we shall find that, as the sample size increases, $D(e) = \sigma^2(I - P)$ will tend to $D(\varepsilon) = \sigma^2 I$.

Now let us imagine that the sequence $\eta(t) = \{\eta_t; t = 0, \pm 1, \pm 2, \dots\}$ of the disturbances follows an AR(1) process such that

$$(233) \quad \eta(t) = \rho\eta(t - 1) + \varepsilon(t), \quad \text{with } \rho \in [0, 1).$$

Then, if we could observe the sequence directly, the best way of detecting the serial correlation would be to estimate the value of ρ and to test the significance of the estimate. The appropriate estimate would be

$$(234) \quad r = \frac{\sum_{t=2}^T \eta_{t-1}\eta_t}{\sum_{t=1}^T \eta_t^2}.$$

When residuals are used in the place of disturbances, we face a variety of options.

The first of these options, which we may be constrained to follow in some circumstances, is to treat the residuals as if they were equal to the disturbances. This approach is acceptable when the sample contains sufficient information to determine estimates of the systematic parameters which are close to the true values. Thus it may be said that the procedure has asymptotic validity in the sense that it becomes increasingly acceptable as the sample size increases.

A second approach is to derive a sequence of revised residuals which have the same statistical properties as the disturbances, such that they are independently and identically distributed whenever the disturbances are. Then we

5: CORRELATED DISTURBANCES

should be able to make exact inferences based on the usual statistical tables. The problem with this approach is that the test statistics which are based upon the revised residuals tend to lack power in small samples.

A third approach, which has become practicable only in recent years, is to devise tests which take full account of whatever statistical properties the ordinary least-squares residuals would possess if the null hypothesis asserting the i.i.d nature of the disturbances were true. This means that we must calculate the exact sampling distribution of the test statistic under the null hypothesis in view of the precise value of the moment matrix $X'X/T$ which characterises the regression. Such an approach requires considerable computing resources.

The traditional approach to the problem of testing for the presence of serial correlation in a regression model is due to Durbin and Watson. They have attempted to make an explicit allowance for the uncertainties which arise from not knowing the precise distribution of the test statistic in any particular instance. This has led them to acknowledge and to identify the circumstances where their test statistic delivers inconclusive results.

The test statistic of Durbin and Watson, which is based upon the sequence $\{e_t; t = 1, \dots, T\}$ of ordinary least-square residuals, is defined by

$$(235) \quad d = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2}.$$

In fact, this statistic may be used quite generally for detecting any problem of misspecification whose symptoms are seeming violations of the i.i.d assumption concerning the disturbances.

In attempting to understand the nature of the statistic, attention should be focussed on its numerator. The numerator is a sum of squares of differences between adjacent residuals. Clearly, if the disturbance sequence is highly correlated, with adjacent values lying close to each other, then we should expect the numerator to have a low value. Conversely, if adjacent values of the disturbance sequence are uncorrelated, then we should expect adjacent residuals, likewise, to show little correlation, and we should expect the numerator to have a high value.

On expanding the numerator of the Durbin–Watson statistic, we find that

$$(236) \quad d = \frac{1}{\sum_{t=1}^T e_t^2} \left\{ \sum_{t=2}^T e_t^2 - 2 \sum_{t=2}^T e_t e_{t-1} + \sum_{t=2}^T e_{t-1}^2 \right\} \simeq 2 - 2r,$$

where

$$(237) \quad r = \frac{\sum_{t=2}^T e_t e_{t-1}}{\sum_{t=1}^T e_t^2}$$

is an estimate of the coefficient ρ of serial correlation which is based on the ordinary least-squares residuals. If ρ is close to 1, and r likewise, then d will be close to zero; and we shall have a strong indication of the presence of serial correlation. If ρ is close to zero, so that the i.i.d assumption is more or less valid, then d will be close to 2.

The response of Durbin and Watson to the fact that their statistic is based on residuals rather than on disturbances was to provide a table of critical values which included a region of indecision. The resulting decision rules can be expressed as follows

- if $d < d_L$, then acknowledge the presence of serial correlation,
- if $d_L \leq d \leq d_U$, then remain undecided,
- if $d_U < d$, then deny the presence of serial correlation.

The values of d_L and d_U to be found in the table depend upon two parameters. The first is the number of degrees of freedom available to the regression and the second is the number of variables included in the regression. The latter is reckoned without including the intercept term. It is evident from the table that, as the number of degrees of freedom increase, the region of indecision lying between d_L and d_U becomes smaller, until a point is reached where it is no longer necessary to make any allowance for it.

Estimating a Regression Model with AR(1) Disturbances

Now let us consider various ways of estimating a regression model with a first-order autoregressive disturbance process. Assume that the regression equation takes the form of

$$(238) \quad y(t) = \alpha + \beta x(t) + \eta(t),$$

where

$$(239) \quad \eta(t) = \rho\eta(t-1) + \varepsilon(t) \quad \text{with} \quad \rho \in [0, 1).$$

With the use of the lag operator, the latter equation can be written as

$$(240) \quad \eta(t) = \frac{1}{1 - \rho L} \varepsilon(t)$$

which is substituted into equation (238) to give

$$(241) \quad y(t) = \alpha + \beta x(t) + \frac{1}{1 - \rho L} \varepsilon(t).$$

5: CORRELATED DISTURBANCES

Multiplying this equation throughout by $I - \rho L$ gives

$$(242) \quad \begin{aligned} (1 - \rho L)y(t) &= (1 - \rho L)\alpha + (1 - \rho L)\beta x(t) + \varepsilon(t) \\ &= \mu + (1 - \rho L)\beta x(t) + \varepsilon(t), \end{aligned}$$

where $\mu = (1 - \rho)\alpha$. This can be written as

$$(243) \quad q(t) = \mu + \beta w(t) + \varepsilon(t),$$

where

$$(244) \quad q(t) = (1 - \rho L)y(t) \quad \text{and} \quad w(t) = (1 - \rho L)x(t).$$

If the value of ρ were known, then the sequences $q(t)$ and $w(t)$ could be formed and the parameters μ and β could be estimated by applying ordinary least-squares regression to equation (243). An estimate for $\alpha = \mu/(1 - \rho)$ would then be recoverable from the estimate of μ .

There are various various ways in which we may approach the estimation of the equation (241) when ρ is unknown.

A simple approach to the estimation of the equation, which requires only a single application of ordinary least-squares regression, depends upon rewriting equation (242) as

$$(245) \quad y(t) = \mu + \rho y(t - 1) + \beta_0 x(t) + \beta_1 x(t - 1) + \varepsilon(t),$$

where $\beta_0 = \beta$ and $\beta_1 = -\rho\beta$. The estimation of this equation by ordinary least-squares regression takes no account of the fact that the parameters are bound by the restriction that $\beta_1 = -\rho\beta_0$; and therefore the number of parameters to be estimated becomes four instead of three. We should expect to lose some statistical efficiency by ignoring the parametric restriction.

The second approach to the estimation of equation (238) is based on equation (243), and it involves searching for the optimal value of ρ by running a number of trial regressions. For a given value of ρ , the elements of $q(t) = (1 - \rho L)y(t)$ and $w(t) = (1 - \rho L)x(t)$ for $t = 2, \dots, T$ are constructed according to the following scheme:

$$(246) \quad \begin{array}{ll} q_2 = y_2 - \rho y_1, & w_2 = x_2 - \rho x_1, \\ q_3 = y_3 - \rho y_2, & w_3 = x_3 - \rho x_2, \\ \vdots & \vdots \\ q_T = y_T - \rho y_{T-1}, & w_T = x_T - \rho x_{T-1}. \end{array}$$

Then the corresponding equation

$$(247) \quad q_t = \mu + \beta w_t + u_t$$

is subjected to an ordinary least-squares regression; and the value of the residual sum of squares is recorded. The procedure is repeated for various values of ρ ; and the definitive estimates of ρ , α and β are those which correspond to the least value of the residual sum of squares.

The procedure of searching for the optimal value of ρ may be conducted in a systematic and efficient manner by using a line-search algorithm such as the method of Fibonacci Search or the method of Golden-Section Search. These algorithms are described in textbooks on unconstrained numerical optimisation.

The third method for estimating the parameters of equation (238) is the well-known Cochrane–Orcutt procedure. This is an iterative method in which each stage comprises two ordinary least-squares regressions.

The first of these regressions takes a value for ρ and estimates the parameters of the equation

$$y_t - \rho y_{t-1} = \mu + \beta(x_t - \rho x_{t-1}) + \varepsilon_t$$

(248) or, equivalently,

$$q_t = \mu + \beta w_t + \varepsilon_t$$

in the manner which we have just described. This gives rise to the conditional estimates $\beta = \beta(\rho)$ and $\mu = \mu(\rho)$. From the latter, we obtain the estimate $\alpha = \alpha(\rho)$.

The second of the regressions is applied to the equation

$$(y_t - \alpha - \beta x_t) = \rho(y_{t-1} - \alpha - \beta x_{t-1}) + \varepsilon_t$$

(249) or, equivalently,

$$\eta_t = \rho \eta_{t-1} + \varepsilon_t$$

which incorporates the previously-determined values of α and β . This gives rise to the conditional estimate $\rho = \rho(\alpha, \beta)$. The later is then fed back into the previous equation (248); at which point another stage of the iterative process may ensue. The process is continued until the estimates generated by successive stages are virtually identical. The final estimates correspond to the unconditional minimum of the residual sum of squares.