Ch.10 Autocorrelated Disturbances

(June 15, 2016)

In a time-series linear regression model setting,

$$Y_t = \mathbf{x}_t'\boldsymbol{\beta} + u_t,\ t = 1, 2, ..., T, \tag{10-1}$$

a common problem is autocorrelation, or serial correlation of the disturbance u_t across periods.

Example.

See the plot of the residuals at Figure 12.1 on p. 251 of Greene 5th edition. ■

1 Time-Series Data

A particularly important aspect of real observable phenomena, which the random variables concept cannot accommodate, is their *time dimension*; the concept of random variable is essential static. A number of economic phenomena for which we need to formulate probability models come in the form of dynamic processes for which we have discrete sequence of observations in time. The problem we have to face is to extend the simple probability model,

$$\Phi = \{f(x; \boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta}\},$$

to one which enables us to model *dynamic phenomena*. We have already moved in this direction by proposing the random vector probability model

$$\Phi = \{f(x_1, x_2, ..., x_T; \boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta}\}.$$

The way we viewed this model so far has been as representing *different characteristics* of the phenomenon (at the same time) in question in the form of the jointly distributed r.v.'s $X_1, X_2, ..., X_T$. If we reinterpret this model as representing the *same characteristic but at successive points in time* then this can be viewed as a dynamic probability model. With this as a starting point let us consider the dynamic probability model in the context of $(\mathcal{S}, \mathcal{F}, \mathcal{P})$.

1.1 The Concept of a Stochastic Process

The natural way to make the concept of a random variable dynamic is to extend its domain by attaching a *date* to the elements of the sample space $\mathcal{S}$.

Definition.
Let $(\mathcal{S}, \mathcal{F}, \mathcal{P})$ be a probability space and $\mathcal{T}$ an index set of real numbers and define the function $X(\cdot, \cdot)$ by $X(\cdot, \cdot) : \mathcal{S} \times \mathcal{T} \to \mathbb{R}$. The ordered sequence of random variables $\{X(\cdot, t), t \in \mathcal{T}\}$ is called a stochastic process. ■

This definition suggests that for a stochastic process $\{X(\cdot, t), t \in \mathcal{T}\}$, for each $t \in \mathcal{T}$, $X(\cdot, t)$ represents a random variable on $\mathcal{S}$. On the other hand, for each s in $\mathcal{S}$, $X(s, \cdot)$ represents a function of t which we call a *realization* of the process. $X(s, t)$ for given s and t is just a real number.

Three main elements of a stochastic process $\{X(\cdot, t), t \in \mathcal{T}\}$ are:

(a). its range space (sometimes called the state space), usually $\mathcal{R}$;

(b). the index $\mathcal{T}$, usually one of $\mathcal{R}, \mathcal{R}_+ = [0, \infty)$, and

(c). the dependence structure of the r.v.'s $\{X(\cdot, t), t \in \mathcal{T}\}$. □

In what follows a stochastic process will be denoted by $\{X_t, t \in \mathcal{T}\}$ (s is dropped and $X(t)$ is customary used as continuous stochastic process) and we are concerning exclusively on *discrete stochastic process*.

The dependence structure of $\{X_t, t \in \mathcal{T}\}$, in direct analogy with the case of a random vector, should be determined by the joint distribution of the process. The question arises, however, since $\mathcal{T}$ is commonly an infinite set, do we need an infinite dimensional distribution to define the structure of the process ?

This question was tackled by Kolmogorov (1933) who showed that when the stochastic process satisfies certain regularity conditions the answer is definitely "no". In particular, if we define the "tentative" joint distribution of the process for the subset $(t_1 < t_2 < ... < t_T)$ of $\mathcal{T}$ by $F(X_{t_1}, X_{t_2}, ..., X_{t_T}) = Pr(X_{t_1} \leq x_1, X_{t_2} \leq x_2, ..., X_{t_T} \leq x_T)$, then if the stochastic process $\{X_t, t \in \mathcal{T}\}$ satisfies the conditions:

(a). symmetry:
$F(X_{t_1}, X_{t_2}, ..., X_{t_T}) = F(X_{t_{j1}}, X_{t_{j2}}, ..., X_{t_{jT}})$ where $j1, j2, ..., jT$ is any permutation of the indices $1, 2, ..., T$ (i.e. reshuffling the ordering of the index does not change the distribution).

(b). Compatibility:
$\lim_{x_T \to \infty} F(X_{t_1}, X_{t_2}, ..., X_{t_T}) = F(X_{t_1}, X_{t_2}, ..., X_{t_{T-1}})$ (i.e. the dimensionality of the joint distribution can be reduced by marginalisation); □

then there exists a probability space $(\mathcal{S}, \mathcal{F}, \mathcal{P})$ and a stochastic process $\{X_t, t \in \mathcal{T}\}$ defined on it whose finite dimensional distribution is the distribution $F(X_{t_1}, X_{t_2}, ..., X_{t_T})$ as defined above. That is, the probability structure of the stochastic process $\{X_t, t \in \mathcal{T}\}$ is completely specified by the joint distribution of $F(X_{t_1}, X_{t_2}, ..., X_{t_T})$ for all values of $\mathcal{T}$ (a positive integer) and any subset $(t_1, t_2, ..., t_T)$ of $\mathcal{T}$.

1.1.1 The First Two Moment of a Stochastic Process

Given that, for a specific t, X_t is a random variable, we can denote its distribution and density function by $F(X_t)$ and $f(X_t)$ respectively. Moreover the mean, variance and higher moments of X_t (as a r.v.) can be defined as standard form as:

$$\begin{aligned} E(X_t) &= \int_{x_t} x_t f(x_t) dx_t = \mu_t \\ E(X_t - \mu_t)^2 &= \int_{x_t} (x_t - \mu_t)^2 f(x_t) dx_t = v^2(t) \\ E(X_t)^r &= \mu_{rt}, \quad r \geq 1, \quad t \in \mathcal{T}. \end{aligned}$$

The linear dependence measures between X_{t_i} and X_{t_j}

$$v(t_i, t_j) = E[(X_{t_i} - \mu_{t_i})(X_{t_j} - \mu_{t_j})], \quad t_i, t_j \in \mathcal{T},$$

is now called the **autocovariance function**. In standardized form

$$r(t_i, t_j) = \frac{v(t_i, t_j)}{v(t_i)v(t_j)}, \quad t_i, t_j \in \mathcal{T},$$

is called is **autocorrelation function**. These numerical characteristics of the stochastic process $\{X_t, t \in \mathcal{T}\}$ play an important role in the analysis of the process and its application to modeling real observable phenomena. We say that $\{X_t, t \in \mathcal{T}\}$ is an uncorrelated process if $r(t_i, t_j) = 0$ for any $t_i, t_j \in \mathcal{T}, t_i \neq t_j$.

Example.
One of the most important example of a stochastic process is the normal process. The stochastic process $\{X_t, t \in \mathcal{T}\}$ is said to be normal (or Gaussian) if any finite subset of $\mathcal{T}$, say $t_1, t_2, ..., t_T$, $(X_{t_1}, X_{t_2}, ..., X_{t_T}) \equiv \mathbf{X}_T'$ has a multivariate normal distribution, i.e.

$$f(X_{t_1}, X_{t_2}, ..., X_{t_T}) = (2\pi)^{-T/2}|\mathbf{V}_T|^{-1/2} \exp[-\frac{1}{2}(\mathbf{X}_T - \boldsymbol{\mu}_T)'\mathbf{V}_T^{-1}(\mathbf{X}_T - \boldsymbol{\mu}_T)],$$

where

$$\boldsymbol{\mu}_T = E(\mathbf{X}_T) = \begin{bmatrix} \mu_1 \\ \mu_2 \\ . \\ . \\ . \\ \mu_T \end{bmatrix} \mathbf{V}_T = \begin{bmatrix} v^2(t_1) & v(t_1, t_2) & . & . & . & v(t_1, t_T) \\ v(t_2, t_1) & v^2(t_2) & . & . & . & v(t_2, t_T) \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ v(t_T, t_1) & . & . & . & . & v^2(t_T) \end{bmatrix}. \quad \blacksquare$$

As in the case of a normal random variable, the distribution of a normal stochastic process is characterized by the first two moment but now they are function of t.

One problem so far in the definition of a stochastic process given above is much too general to enable us to obtain an operational probability model. In the analysis of stochastic process we only have a single realization of the process and we will have to deduce the value of μ_t and $v(t)$ with the help of a single observation. (which is impossible !)

The main purpose of the next three sections is to consider various special forms of stochastic process where we can construct probability models which are manageable in the context of statistical inference. Such manageability is achieved by imposing certain

restrictions which enable us to reduce the number of unknown parameters involved in order to be able to deduce their value from a single realization. These restrictions come in two forms:

(a). restriction on the **time-heterogeneity** of the process; and

(b). restriction on the **memory** of the process.

1.2 Restricting the Time-Heterogeneity of a Stochastic Process

For an arbitrary stochastic process $\{X_t, t \in \mathcal{T}\}$ the distribution function $F(X_t; \theta_t)$ depends on t with the parameter θ_t characterizing it being function of t as well. That is, a stochastic process is time-heterogeneous in general. This, however, raises very difficult issues in modeling real phenomena because usually we only have one observation for each t. Hence in practice we will have to estimate θ_t on the basis of a single observation, which is impossible. For this reason we are going to consider an important class of *stationary* process which exhibit considerable time-homogeneity and can be used to model phenomena approaching their $equilibrium\ steady-state$, but continuously undergoing "random" functions. This is the class of stationary stochastic processes.

Definition.
A stochastic process $\{X_t, t \in \mathcal{T}\}$ is said to be **(strictly) stationary** if any subset $(t_1, t_2, ..., t_T)$ of $\mathcal{T}$ and any τ,

$$F(X_{t_1}, ..., X_{t_T}) = F(X_{t_1+\tau}, ..., X_{t_T+\tau}). \qquad \blacksquare$$

That is, the distribution of the process remains unchanged when shifted in time by an arbitrary value τ. In terms of the marginal distributions, (strictly) stationarity implies that

$$F(X_t) = F(X_{t+\tau}), \ \ t \in \mathcal{T},$$

and hence $F(X_{t_1}) = F(X_{t_2}) = ... = F(X_{t_T})$. That is stationarity implies that $X_{t_1}, X_{t_2}, ..., X_{t_T}$ are (individually) identically distributed.

The concept of stationarity, although very useful in the context of probability theory, is very difficult to verify in practice because it is defined in terms of distribution function. For this reason the concept of the *second* order stationarity, defined in terms of the first two moments, is commonly preferred.

Definition.
A stochastic process $\{X_t, t \in \mathcal{T}\}$ is said to be (**weakly) stationary** if

$$\begin{aligned} E(X_t) &= \mu \ \ for\ all\ t; \\ v(t_i, t_j) &= E[(X_{t_i} - \mu)(X_{t_j} - \mu)] = \gamma_{|t_j - t_i|}, \ \ t_i, t_j \in \mathcal{T}. \end{aligned}$$

These suggest that weakly stationarity for $\{X_t, t \in \mathcal{T}\}$ implies that its mean and variance $v^2(t_i) = \gamma_0$ are constant and free of t and its autocovariance depends on the interval $|t_j - t_i|$; not t_i and t_j. ■

Example.
Consider the normal stochastic process in the above example. With the weakly stationarity assumption, now

$$\boldsymbol{\mu}_T = E(\mathbf{X}_T) = \begin{bmatrix} \mu \\ \mu \\ . \\ . \\ . \\ \mu \end{bmatrix}, \ \ \mathbf{V}_T = \begin{bmatrix} \gamma_0 & \gamma_1 & . & . & . & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & . & . & . & \gamma_{T-2} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \gamma_{T-1} & . & . & . & . & \gamma_0 \end{bmatrix},$$

a sizeable reduction in the number of unknown parameters from $T + [T(T+1)/2]$ to $(T+1)$. ■

It is important, however, to note that even in the case of stationarity the number of parameters increases with the size of the subset $(t_1, ..., t_T)$ although the parameters do not depend on $t \in \mathcal{T}$. This is because time-homogeneity does not restrict the 'memory' of the process. In the next section we are going to consider 'memory' restrictions in an obvious attempt to 'solve' the problem of the parameters increasing with the size of the subset $(t_1, t_2, ..., t_T)$ of $\mathcal{T}$.

1.3 Restricting the Memory of a Stochastic Process

In the case of a typical economic times series, viewed as a particular realization of a stochastic process $\{X_t,\ t \in \mathcal{T}\}$ one would expect that the dependence between X_{t_i} and X_{t_j} would tend to weaken as the distance $(t_j - t_i)$ increases. Formally, this dependence can be described in terms of the joint distribution $F(X_{t_1}) = F(X_{t_2}) = ... = F(X_{t_T})$ as follows.

Definition. (Asymptotically Uncorrelated):
A time series $\{X_t,\ t \in \mathcal{T}\}$ is said to be *asymptotically uncorrelated* if there exists a sequence of constants $\{\rho(\tau), \tau \geq 1\}$ defined by

$$r(t, t+\tau) = \left| \frac{\gamma_{t,t+\tau}}{\gamma_t \gamma_{t+\tau}} \right| \leq \rho(\tau), \quad for\ all\ t \in \mathcal{T},$$

such that

$$0 \leq \rho(\tau) \leq 1 \quad and \quad \sum_{\tau=0}^{\infty} \rho(\tau) < \infty. \qquad \blacksquare$$

1.4 Some Special Stochastic Processes

We will consider briefly several special stochastic processes which play an important role in econometric modeling. These stochastic processes will be divided into *parametric* and *non-parametric* process. The non-parametric processes are defined in terms of their joint distribution function or the first few joint moments. On the other hand, parametric processes are defined in terms of a generating mechanism which is commonly a functional form based on a non-parametric process.

1.4.1 Non-Parametric Process

A common used non-parametric process which forms a basis of most parametric processes is the white noise process.

Definition.
A stochastic process $\{\varepsilon_t,\ t \in \mathcal{T}\}$ is said to be a **white-noise** process if

(a). $E(\varepsilon_t) = 0$;

(b). $E(\varepsilon_t \varepsilon_\tau) = \begin{cases} \sigma^2, & if\ t = \tau \\ 0, & if\ t \neq \tau. \end{cases}$ ■

Hence, a white-noise process is both time-homogeneous, in view of the fact that it is a weakly-stationary process, and has no memory. In the case where $\{\varepsilon_t,\ t \in \mathcal{T}\}$ is also assumed to be normal the process is also strictly stationary.

1.4.2 Parametric Stochastic Processes

The main difference between the type of time series considered, white noise and the one to be considered in this section is that the latter are defined in terms of a generating mechanism; they are "derived" stochastic processes from a white noise process.

Definition.
A stochastic process $\{u_t,\ t \in \mathcal{T}\}$ is said to be a **autoregressive of order one** (AR(1)) if it satisfies the stochastic difference equation,

$$u_t = \phi u_{t-1} + \varepsilon_t, \tag{10-2}$$

where ϕ is a constant and ε_t is a white-noise process. ■

We first consider the case that $|\phi| < 1$. Let the index set $\mathcal{T}^* = \{0, \pm 1, \pm 2, ...\}$ and assume that $X_{-T} \to 0$ as $T \to \infty$. Define a *lag-operator* "L" by

$$LX_t \equiv X_{t-1},$$

then the AR(1) process in (10-1) can be rewritten as

$$(1-\phi L)u_t = \varepsilon_t,$$

or

$$\begin{aligned} u_t &= (1-\phi L)^{-1}\varepsilon_t \\ &= (1+\phi L+\phi^2 L^2+....)\varepsilon_t \\ &= \varepsilon_t+\phi\varepsilon_{t-1}+\phi^2\varepsilon_{t-2}+..... \\ &= \sum_{i=0}^{\infty}\phi^i\varepsilon_{t-i}, \end{aligned}$$

from which we can deduce that

$$\begin{aligned} E(u_t) &= 0, \\ v(\tau) &= E(u_t u_{t+\tau}) = E\left\{\left(\sum_{i=0}^{\infty}\phi^i\varepsilon_{t-i}\right)\left(\sum_{j=0}^{\infty}\phi^i\varepsilon_{t+\tau-j}\right)\right\} \\ &= \sigma_u^2\left(\sum_{i=0}^{\infty}\phi^i\phi^{i+\tau}\right) \\ &= \sigma_\varepsilon^2\phi^\tau\left(\sum_{i=0}^{\infty}\phi^{2i}\right) \\ &= \frac{\sigma_\varepsilon^2}{(1-\phi^2)}\cdot\phi^\tau, \ \ \tau\geq 0. \end{aligned}$$

Hence, for $|\phi|<1$, the stochastic process $\{X_t,\ t\in\mathcal{T}^*\}$ is both weakly-stationary and asymptotically uncorrelated since the autocovariance function

$$v(\tau)=\frac{\sigma_\varepsilon^2}{(1-\phi^2)}\phi^\tau\to 0,\ as\ \tau\to\infty.$$

Therefore, if any finite subset of $\mathcal{T}^*$, say $t_1,t_2,...,t_T$ of a AR(1) process, $(u_{t_1},u_{t_2},...,u_{t_T})\equiv\mathbf{u}_T'$ has covariance matrix

$$E(\mathbf{u}_T\mathbf{u}_T')=\sigma_\varepsilon^2\frac{1}{(1-\phi^2)}\begin{bmatrix} 1 & \phi & . & . & . & \phi^{T-1} \\ \phi & 1 & \phi & . & . & \phi^{T-2} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \phi^{T-1} & . & . & . & . & 1 \end{bmatrix}=\sigma_\varepsilon^2\cdot\mathbf{\Sigma}, \tag{10-3}$$

where

$$\mathbf{\Sigma} = \frac{1}{(1-\phi^2)} \begin{bmatrix} 1 & \phi & . & . & . & \phi^{T-1} \\ \phi & 1 & \phi & . & . & \phi^{T-2} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \phi^{T-1} & . & . & . & . & 1 \end{bmatrix}.$$

It is straightforward to show by direct multiplication that

$$\mathbf{P}'\mathbf{P} = \mathbf{\Sigma}^{-1},$$

for

$$\mathbf{P} = \begin{bmatrix} \sqrt{1-\phi^2} & 0 & . & . & . & 0 \\ -\phi & 1 & 0 & . & . & 0 \\ 0 & -\phi & 1 & 0 & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & -\phi & 1 \end{bmatrix}. \tag{10-4}$$

2 OLS Estimation

We showed in Section 8.1 that in the presence of autocorrelation in the disturbance, the OLS estimator $\hat{\boldsymbol{\beta}}$ for (10-1) is unbiased and consistent. However it is inefficient relative to the GLS estimator.

2.1 Estimating the Appropriate Covariance Matrix for OLS Estimators

Following White's suggestion for heteroscedasticity, Newey and West's (1987) robust, consistent covariance of $\hat{\boldsymbol{\beta}}$ estimator for heteroscedastic and autocorrelated disturbance (HAC) with an unspecified structure is at (8-9), repeated here,

$$\begin{aligned}&\widehat{Var(\hat{\boldsymbol{\beta}})}_{HAC}\\&= (\mathbf{X}'\mathbf{X})^{-1}\left(\sum_{t=1}^{T} e_t^2\mathbf{x}_t\mathbf{x}_t' + \sum_{l=1}^{L}\sum_{t=l+1}^{T} w_l e_t e_{t-l}(\mathbf{x}_t\mathbf{x}_{t-l}' + \mathbf{x}_{t-l}\mathbf{x}_t')\right)(\mathbf{X}'\mathbf{X})^{-1}\end{aligned} \tag{10-5}$$

Here, e_t is the tth OLS residual and

$$w_l = 1 - \frac{l}{L+1}.$$

The maximum lag L must be determined in advance to be large enough that autocorrelations at lags larger than L are small enough to ignore.

Exercise 1.
Reproduce the results at Table 12.1 on p. 267 of Greene 5th Edition.

3 Testing for Autocorrelation

Most of the available tests for autocorrelation are based on the principle that if the true disturbances u_t are autocorrelated, this fact will be revealed through the autocorrelation of the OLS residuals e_t, because

$$\begin{aligned} Y_t &= \mathbf{x}_t'\boldsymbol{\beta} + u_t \\ &= \mathbf{x}_t'\hat{\boldsymbol{\beta}} + e_t. \end{aligned}$$

The simplest indicator is the slope estimator in the artificial regression

$$e_t = re_{t-1} + v_t,$$

i.e.

$$\hat{r} = \frac{\sum_{t=2}^{T} e_t e_{t-1}}{\sum_{t=1}^{T} e_t^2} \tag{10-6}$$

If there is autocorrelation, then the slope in this regression will be an estimator of $\rho = Corr[u_t, u_{t-1}]$. The complication in the analysis lies in determining a formal means of evaluating when the estimator is "large", that is, on what statistical basis to reject the null hypothesis that ρ equals zero.

3.1 The Durbin-Watson Test

The most extensively used test for AR(1) disturbance is the Durbin-Watson test developed by Durbin and Watson (1950,1951). The DurbinWatson statistic was the first formal procedure developed for testing for autocorrelation using the least squares residuals. Before making formal statement about the test, we have the following lemma which is used to prove this test.

Lemma.
Let $\mathbf{z}$ and $\mathbf{v}$ be $T \times 1$ random vector such that $\mathbf{z} = \mathbf{M}\mathbf{v}$, where $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and $\mathbf{X}$ is a $T \times k$ nonstochastic matrix of rank k. Furthermore, let $r = \mathbf{z}'\mathbf{A}\mathbf{z}/\mathbf{z}'\mathbf{z}$, where $\mathbf{A}$ is a real symmetric matrix. Then

(a). There exists an orthogonal transformation $\mathbf{v} = \mathbf{H}\boldsymbol{\delta}$ such that

$$r = \frac{\sum_{i=1}^{T-k} u_i \delta_i^2}{\sum_{i=1}^{T-k} \delta_i^2}, \tag{10-7}$$

where $u_1, u_2, ..., u_{T-k}$ are the $T-k$ nonzero (ordered) eigenvalues of $\mathbf{MA}$, the rest being zero and $\delta_i \sim N(0,1)$. (That is, u_i is function of $\mathbf{X}$. Therefore the distribution of r is unknown.)

(b). If s of the columns of $\mathbf{X}$ are linear combinations of s of the eigenvectors of $\mathbf{A}$ and if the eigenvalues of $\mathbf{A}$ associated with the remaining $T-s$ eigenvalues of $\mathbf{A}$ are renumbered so that

$$\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_{T-s},$$

then

$$\lambda_i \leq u_i \leq \lambda_{i+k-s} \quad (i = 1, 2, ..., T-k). \qquad \square$$

From the above lemma the following corollary can be deduced.

Corollary.
The r in (10-6) is bounded by

$$r_L \leq r \leq r_U,$$

where

$$r_L = \frac{\sum_{i=1}^{T-k} \lambda_i \delta_i^2}{\sum_{i=1}^{T-k} \delta_i^2},$$

and

$$r_U = \frac{\sum_{i=1}^{T-k} \lambda_{i+k-s} \delta_i^2}{\sum_{i=1}^{T-k} \delta_i^2}. \qquad \blacksquare$$

The importance of this results is that it set bounds on r which are independent of $\mathbf{X}$. We now turn to the test $H_0 : \rho = 0$, of the AR(1) disturbances in the linear model:

$$\begin{aligned} y_t &= \mathbf{x}_t'\boldsymbol{\beta} + u_t; \\ u_t &= \rho u_{t-1} + \varepsilon_t, \quad t = 1, 2, ..., T, \end{aligned}$$

where ε_t is a white noise process and $-1 < \rho < 1$.

The Durbin-Watson d-statistics is written as

$$d = \frac{\sum_{t=1}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} = \frac{\mathbf{e}'\mathbf{A}\mathbf{e}}{\mathbf{e}'\mathbf{e}} = \frac{\sum_{t=2}(e_t^2 - 2e_t e_{t-1} + e_{t-1}^2)}{\sum_{t=1} e_t^2} \simeq 2(1 - corr(e_t, e_{t-1})),$$

where

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 & . & . & 0 \\ -1 & 2 & -1 & 0 & . & 0 \\ 0 & -1 & 2 & -1 & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & 0 \\ . & . & . & -1 & 2 & -1 \\ 0 & & . & 0 & -1 & 1 \end{bmatrix}.$$

Therefore a small value of d would like to reject H_0 for testing a positive ρ.

The eigenvalues of $\mathbf{A}$ are

$$\lambda_i = 2\left\{1 - \cos\frac{\pi(i-1)}{T}\right\}, \quad i = 1, 2, ..., T.$$

The eigenvector of $\mathbf{A}$ corresponding to the zero eigenvalues λ_1 is $(1, 1, ..., 1)'$, which is the regression vector corresponding to a constant term in the regression model. (Notice that, in this discussion, as well as in the statistical table, the existence of a constant is implicitly assumed.) From the Corollary above and using the fact that $\mathbf{e} = \mathbf{M}\boldsymbol{\varepsilon}$ we have

$$d_L \leq d \leq d_U,$$

where

$$d_L = \frac{\sum_{i=1}^{T-k} \lambda_i \delta_i^2}{\sum_{i=1}^{T-k} \delta_i^2},$$

and

$$d_U = \frac{\sum_{i=1}^{T-k} \lambda_{i+k-1} \delta_i^2}{\sum_{i=1}^{T-k} \delta_i^2}.$$

Since the λ_i are the same in any regression models with T observations and k regressors including the constant term, the distribution of d_L and d_U have been computed critical values by DW.

Three hypotheses of interest with respect to AR(1) disturbance processes are

DW(a). $H_0 : \rho = 0$ versus $H_1 : \rho > 0$;

DW(b). $H_0 : \rho = 0$ versus $H_1 : \rho < 0$; and

DW(c). $H_0 : \rho = 0$ versus $H_1 : \rho \neq 0$. □

For testing DW(a)., the null hypothesis is rejected if $d < F(d)_{5\%}$ which is guaranteed if $d < F(d_L)_{5\%}$ for the un-availability of $F(d)_{5\%}$. Equivalently speak, the null hypothesis is accepted if $d > F(d)_{5\%}$ which is guaranteed if $d > F(d_U)_{5\%}$.

It is important to emphasize that:

(a). the statistical tables of Durbin and Watson assume the existence of a constant,

(b). no allowance is made for missing observations, and

(c). the DW test was derived under the assumption that $\mathbf{X}$ is nonstochastic and thus is not applicable, for example, when lagged values of the dependent variable appear among the regressors. □

3.2 The Box Q test

Box and Pierce (1970)'s Q test is carried out by referring

$$Q = T \sum_{j=1}^{k} \hat{r}_j^2(e),$$

where $\hat{r}_j$ is the estimation of the jth covariance in u_t from e_t,.i.e.

$$\hat{r}_j = \frac{\sum_{t=j+1}^{T} e_t e_{t-j}}{\sum_{t=1}^{T} e_t^2}. \tag{10-8}$$

When the null hypothesis, $\rho = 0$, Q is approximately distributed as χ^2_k. On the other hand, if the null hypothesis is inappropriate, the average value of Q will be inflated. A refinement that appears to have better finite-sample properties is the Ljung-Box (1979) statistics:

$$Q' = T(T+2)\sum_{j=1}^{k}\frac{\hat{r}_j^2(e)}{T-k}.$$

The limiting distribution of Q' is the same as that of Q.

4 GLS when $\mathbf{\Sigma}$ is known

As a prelude to deriving feasible estimator for $\boldsymbol{\beta}$ in this model, we consider full generalized least squares estimation assuming that $\mathbf{\Sigma}$ is known. In the next section, we will turn to the more realistic case in which $\mathbf{\Sigma}$ must be estimated as well.

4.1 Generalized Least Squares Estimators

If the parameters of $\mathbf{\Sigma}$ are known, then the GLS estimator,

$$\begin{aligned}\tilde{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{y} \\ &= (\mathbf{X}'\mathbf{P}'\mathbf{P}\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}'\mathbf{P}\mathbf{y}\end{aligned}$$

can be computed directly.

For the disturbance AR(1) process, $u_t = \rho u_{t-1} + \varepsilon_t$, as we have shown in (10-3) that for $\mathbf{P}'\mathbf{P} = \mathbf{\Sigma}^{-1}$,

$$\mathbf{P} = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & . & . & . & 0 \\ -\rho & 1 & 0 & . & . & 0 \\ 0 & -\rho & 1 & 0 & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & -\rho & 1 \end{bmatrix}.$$

The data for the transformed model therefore are

$$\mathbf{P}\mathbf{y} = \begin{bmatrix} \sqrt{1-\rho^2}Y_1 \\ Y_2 - \rho Y_1 \\ Y_3 - \rho Y_2 \\ . \\ . \\ . \\ Y_T - \rho Y_{T-1} \end{bmatrix}, \quad \mathbf{P}\mathbf{X} = \begin{bmatrix} \sqrt{1-\rho^2}\mathbf{x}_1' \\ \mathbf{x}_2' - \rho\mathbf{x}_1' \\ \mathbf{x}_3' - \rho\mathbf{x}_2' \\ . \\ . \\ . \\ \mathbf{x}_T' - \rho\mathbf{x}_{T-1}' \end{bmatrix}, \tag{10-9}$$

and

$$\mathbf{Pu} = \begin{bmatrix} \sqrt{1-\rho^2}u_1 \\ u_2 - \rho u_1 \\ u_3 - \rho u_2 \\ . \\ . \\ . \\ u_T - \rho u_{T-1} \end{bmatrix} = \begin{bmatrix} \sqrt{1-\rho^2}u_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ . \\ \varepsilon_T \end{bmatrix}.$$

Since $E(\sqrt{1-\rho^2}u_1)^2 = \sigma_\varepsilon^2$, thus

$$E(\mathbf{Puu'P'}) = \sigma_\varepsilon^2 \mathbf{I}_T$$

as expected.

4.2 Maximum Likelihood Estimators

If the parameters of the disturbance process are known, then GLS and the maximum likelihood estimation of the AR(1) model are equivalent. To obtain the likelihood function for normally distributed disturbances, we use the multivariate normal density.

Result.

If the parameter in the AR(1) disturbances ρ is known, then the Gaussian MLE of $\boldsymbol{\beta}$ is identical to the GLS.

Proof.

From (10-3) that we assume

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad E(\mathbf{uu'}) = \sigma_\varepsilon^2 \boldsymbol{\Sigma},$$

Because $\mathbf{u} \sim N(\mathbf{0}, \sigma_\varepsilon^2 \boldsymbol{\Sigma})$ and $\mathbf{X}$ is nonstochastic, we view the observed sample $\mathbf{y}$ as a single draw from a $N(\mathbf{X}\boldsymbol{\beta}, \sigma_\varepsilon^2 \boldsymbol{\Sigma})$. The sample likelihood could be written down immediately from the formula for the multivariate Gaussian density:

$$\begin{aligned} f_{\mathbf{y}}(\mathbf{y}; \boldsymbol{\beta}) &= (2\pi)^{-T/2} \left|(\sigma_\varepsilon^2 \boldsymbol{\Sigma})^{-1}\right|^{1/2} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\sigma_\varepsilon^2 \boldsymbol{\Sigma})^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right] \\ &= (2\pi)^{-T/2} \left|\sigma_\varepsilon^{-2} \mathbf{P'P}\right|^{1/2} \exp\left[-\frac{1}{2\sigma_\varepsilon^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{P'P}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right], \end{aligned}$$

with log likelihood

$$\mathcal{L}(\boldsymbol{\beta}) = (-T/2)\log(2\pi) + \frac{1}{2}\log\left|\sigma_\varepsilon^{-2}\mathbf{P}'\mathbf{P}\right| - \frac{1}{2\sigma_\varepsilon^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\mathbf{P}'\mathbf{P}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}). \quad (10\text{-}10)$$

Because

$$\begin{aligned}
\frac{1}{2}\log\left|\sigma_\varepsilon^{-2}\mathbf{P}'\mathbf{P}\right| &= \frac{1}{2}\log\{\sigma_\varepsilon^{-2T}\cdot|\mathbf{P}'\mathbf{P}|\} \\
&= -\frac{T}{2}\log\sigma_\varepsilon^2 + \frac{1}{2}\log\{|\mathbf{P}'\mathbf{P}|\} \\
&= -\frac{T}{2}\log\sigma_\varepsilon^2 + \frac{1}{2}\log\{|\mathbf{P}'||\mathbf{P}||\}, \quad (\because \mathbf{P} \text{ is symmtric}) \\
&= -\frac{T}{2}\log\sigma_\varepsilon^2 + \log|\mathbf{P}| \\
&= -\frac{T}{2}\log\sigma_\varepsilon^2 + \frac{1}{2}\log(1-\rho^2). \qquad (10\text{-}11)
\end{aligned}$$

Substitute (10-11) and (10-9) into (10-10) we obtain that

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\beta},\sigma_\varepsilon^2) &= (-T/2)[\log(2\pi)+\log\sigma_\varepsilon^2] + \frac{1}{2}\log(1-\rho^2) \\
&\quad -\frac{1-\rho^2}{2\sigma_\varepsilon^2}(Y_1-\mathbf{x}_1'\boldsymbol{\beta}) - \frac{1}{2\sigma_\varepsilon^2}\sum_{t=2}^{T}\left[(Y_t-\rho Y_{t-1}) - (\mathbf{x}_t-\rho\mathbf{x}_{t-1})'\boldsymbol{\beta}\right]^2. \qquad (10\text{-}12)
\end{aligned}$$

The sums of squares terms is the residual sums of squares from the transformed classical regression model in (10-9). If ρ is known, then the MLE solution is GLS for $\boldsymbol{\beta}$. ■

5 FGLS When $\boldsymbol{\Sigma}$ is Unknown

We consider specifically the case the disturbance is a AR(1) process but with an unknown ρ.

5.1 Feasible Generalized Least Squares Estimators

For a FGLS of $\boldsymbol{\beta}$, all that need is a consistent estimator of $\boldsymbol{\Omega}(\rho)$. Since the OLS $\hat{\boldsymbol{\beta}}$ is consistent, we can use the OLS residual e_t in the first step,

$$\hat{\rho} = \frac{\sum_{t=2}^{T} e_t e_{t-1}}{\sum_{t=1}^{T} e_t^2}$$

as a consistent estimator of ρ. With this $\hat{\boldsymbol{\Omega}} = \boldsymbol{\Omega}(\hat{\rho})$, the FGLS is in the second step

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

For the second step of FGLS, there are two possibilities:

(a). Full FGLS. This estimator is the Prais-Winsten (1954) estimator.

(b). FGLS omitting the first observation. This estimator was first suggested by Cochrane and Orcutt (1949).

5.2 Maximum Likelihood Estimator

Full maximum likelihood estimators can be obtained by maximizing the log likelihood in (10-12) with respect to $\boldsymbol{\beta}, \sigma_\varepsilon^2$, and ρ, i.e.

$$\begin{aligned}\mathcal{L}^*(\boldsymbol{\beta}, \sigma_\varepsilon^2, \rho) &= (-T/2)[\log(2\pi) + \log \sigma_\varepsilon^2] + \frac{1}{2}\log(1-\rho^2) \\ &\quad -\frac{1-\rho^2}{2\sigma_\varepsilon^2}(Y_1 - \mathbf{x}_1'\boldsymbol{\beta}) - \frac{1}{2\sigma_\varepsilon^2}\sum_{t=2}^{T}\left[(Y_t - \rho Y_{t-1}) - (\mathbf{x}_t - \rho\mathbf{x}_{t-1})'\boldsymbol{\beta}\right]^2.\end{aligned}$$

The solutions are nonlinear functions and the numerical methods in Ch. 3 are needed. However, for a given ρ, the MLE of $\boldsymbol{\beta}$ and σ_ε^2 are the usual ones, GLS. The problem is to estimate ρ. One possibility is to search the range $-1 < \rho < 1$ for the value that with the implied estimates of the other parameters maximizes $\mathcal{L}^*(\boldsymbol{\beta}, \sigma_\varepsilon^2, \rho)$, which is analogous to the Hildreth-Lu estimator.

Exercise 2.
Reproduce the results at Table 12.2 on p.275 of Greene 5th edition. ■

Kaohsiung Harbor, Taiwan.

End of this Chapter