Chap.8 Autocorrelated Disturbances

8.1 Fundamental Concepts in Time-series Analysis

A sequence of random variables (vectors) is called a stochastic process. If the index for the random variables is interpreted as representing time, the stochastic process is called time series, i.e. $\{Y(t), t = 0, \pm 1, \pm 2, ...\}$, simply notes as $\{Y_t\}$.

8.1.1 Characteristic Indices of a Time Series

Mean Function : $E(Y_t) = \mu_t$

Variance Function $Var(Y_t) = E(Y_t - u_t)^2 = \gamma_{tt}$

Auto covariance Function $Cov(Y_t, Y_{t-j}) = E\left[(Y_t - u_t)(Y_{t-j} - u_{t-j})\right] = \gamma_{tj}$

Autocorrelation Function $\dfrac{Cov(Y_t, Y_{t-j})}{Var(Y_t)} = \dfrac{\gamma_{tj}}{\gamma_{tt}} = \rho_{t,t-j} \quad j = 0, \pm 1, \pm 2, \cdots$

All these indices are functions of time “t”.

8.1.2 Stationarity

1) Strictly stationary process

A n-dimension stochastic process $\{Y_t\}$ is strictly stationary if, for any given finite integer k and for any set of subscripts $t_1, \cdots, t_n$, the joint distribution equality for $\{Y_t\}$ holds:

$$F(Y_1, \cdots, Y_n; t_1, \cdots, t_n) = F(Y_1, \cdots, Y_n; t_{1-k}, \cdots, t_{n-k}) \tag{8.1}$$

The definition implies that the distribution of $\{Y_t\}$ depends only on k, the relative position in the sequence, but not on $t_1, \cdots, t_n$, the absolute position of $\{Y_t\}$. So the mean variance and other higher moments, if they exist, remain the same across $t_1, \cdots, t_n$. The definition also implies that if $\{Y_t\}$ is strictly stationary, then $\{f(Y_t)\}$ is, where the function $f(\cdot)$ needs to be “measurable”.

2) Weakly (Covariance) stationary process

A stochastic process $\{Y_t\}$ is weakly (covariance) stationary if

i) $E(Y_t) = \mu$

ii) $Cov(Y_t, Y_{t-j}) = E\left[(Y_t - \mu)(Y_{t-j} - \mu)\right] = \gamma_j \quad j = 0, \pm 1, \cdots$

The definition implies that the mean function of $\{Y_t\}$ is not the function of time "t", and the $Cov(Y_t, Y_{t-j})$ exists, is finite, and depends only on j but not on the "start point" of time "t".

Note:

i) If a sequence if strictly stationary and if the variance and covariance are finite, then the sequence is weakly stationary.

ii) For a scalar covariance stationary process $\{Y_t\}$, the j-th order autocovariance γ_j satisfies

$$\gamma_j = \gamma_{-j} \tag{8.2}$$

By covariance stationary, the autocovariance matrix of the process is a band spectrum matrix:

$$Cov(Y_t, Y_{t+1}, \cdots, Y_{t+n-1}) = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{n-1} & \gamma_{n-2} & \cdots & \gamma_0 \end{bmatrix} \tag{8.3}$$

The j-th order autocorrelation coefficient ρ_j, is defined as

$$\rho_j = \frac{Cov(Y_t, Y_{t-j})}{Var(Y_t)} = \frac{\gamma_j}{\gamma_0} \quad j = 0, \pm 1, \pm 2, \cdots \tag{8.3'}$$

For $j = 0, |\rho_j| = 1; \quad j \neq 0, |\rho_j| < 1$. The plot of $\{\rho_j\}$ against $j = 0, 1, \cdots$ is called the correlogram.

For a vector covariance stationary process $\{Y_t\}$, the j-th order autocovariance, denoted Π_j, is defined as

$$\Pi_j = Cov(Y_t, Y_{t-j}) \quad j = 0, \pm 1, \pm 2, \cdots$$

Also by covariance stationary, Π_j satisfies

$$\Pi_j = \Pi'_{-j}$$

3) White noise processes

A covariance stationary process $\{Y_t\}$ is white noise if

$$E(Y_t) = 0 \text{ and } Cov(Y_t, Y_{t-j}) = 0 \text{ for } j \neq 0.$$

Clearly, an i.i.d. sequence with mean zero and finite variance is a special case of a white noise process, and it is called an independent white noise process.

8.1.3 Ergodicity

1) Ergodic stationary

A stationary process $\{Y_t\}$ is said to be ergodic, if for any two bound functions

$$f : R^{a+1} \to R, \quad g : R^{b+1} \to R$$

$$\begin{aligned} &\lim_{n\to\infty} \left| E\left[f\left(Y_t, Y_{t+1}, \cdots Y_{t+a}\right) g\left(Y_{t+n}, Y_{t+n+1}, \cdots, Y_{t+n+b}\right) \right] \right| \\ &= \left| E\left[f\left(Y_t, Y_{t+1}, \cdots Y_{t+a}\right) \right] \right| \left| E\left[g\left(Y_t, Y_{t+1}, \cdots, Y_{t+b}\right) \right] \right| \end{aligned} \tag{8.4}$$

Note:

i) A stationary process is ergodic if it is asymptotically independent, that is, if any two random variables positioned far apart in the sequence are almost independently distributed.

ii) A stationary process that is ergodic will be called ergodic stationary.

iii) For any (measurable) function $f(\cdot)$, $\{f(Y_t)\}$ is ergodic stationary whenever $\{Y_t\}$ is.

2) Ergodic Theorem

Let $\{Y_t\}$ be a stationary and ergodic process with $E(Y_t) = \mu$, then

$$\overline{Y}_n = \frac{1}{n}\sum_{t=1}^{n} Y_t \xrightarrow{a.s.} \mu \tag{8.5}$$

Note:

i) The Ergodic Theorem is a substantial generalization of Kolmogorov's LLN. Serial dependence, which is ruled out by the i.i.d. assumption in Kolmogorov's LLN, is allowed in the Ergodic Theorem, provided that it disappears in the long run.

ii) This theorem implies that any moment of a stationary and ergodic process (if it exists and is finite) is consistently estimated by the sample moment.

8.1.4 Martingales and Martingale Difference Sequence

1) Martingales

Let X_t be an element of Y_t, the scale process $\{X_t\}$ is called a Martingale with respect to $\{Y_t\}$, if

$$E\left(X_t \middle| Y_{t-1},\cdots,Y_1\right) = X_{t-1} \quad \text{for} \quad t \geq 2$$

The conditioning set $I_{t-1} = \left(Y_{t-1}, Y_{t-2}, \cdots, Y_1\right)$ is often called the information set at date t-1. $\{X_t\}$ is called simply a martingale if the information set is its own past values $\left(X_{t-1}, X_{t-2}, \cdots, X_1\right)$. If Y_t includes X_t, then $\{X_t\}$ is a martingale, because

$$\begin{aligned} & E\left(X_t \middle| X_{t-1},\cdots,X_1\right) \\ & = E\left[E\left(X_t \middle| Y_{t-1},\cdots,Y_1\right) \middle| X_{t-1},\cdots,X_1\right] \\ & = E\left(X_{t-1} \middle| X_{t-1},\cdots,X_1\right) = X_{t-1} \end{aligned} \qquad (8.6)$$

Note:

i) If the process started in the infinite past so that t runs from $-\infty$ to $+\infty$, the definition of a martingale with respect to $\{Y_t\}$ is $E\left(X_t \middle| Y_{t-1}, Y_{t-2}, \cdots\right) = X_{t-1}$, and the qualifier "$t \geq 2$" is not need.

ii) A vector process $\{Y_t\}$ is called a martingale if $E\left(Y_t \middle| Y_{t-1},\cdots,Y_1\right) = Y_{t-1}$ for $t \geq 2$.

2) Random Walks

$\{Y_t\}$ is a Random Walk If it can be written as

$$Y_1 = \varepsilon_1, Y_2 = \varepsilon_1 + \varepsilon_2, \cdots$$

$$Y_t = Y_{t-1} + \varepsilon_t \qquad (8.7)$$

Where $\{\varepsilon_t\}$ is i.i.d. with mean zero and finite variance.

3) Martingale Difference Sequence (MDS)

A vector process $\{\varepsilon_t\}$ with $E\left(\varepsilon_t\right) = 0$ is called a martingale difference sequence (MDS) or martingale difference, if the expectation conditional on its past values is zero:

$$E\left(\varepsilon_t \middle| \varepsilon_{t-1},\cdots,\varepsilon_1\right) = 0 \quad \text{for} \quad t \geq 2.$$

The process is called martingale difference because

$$Y_t = Y_{t-1} + \varepsilon_t = \varepsilon_t + \cdots + \varepsilon_1$$

Where $\{Y_t\}$ is a martingale. The proof is as follows.

Proof:

$$\begin{aligned}
&E\left(Y_t \middle| Y_{t-1}, \cdots, Y_1\right) \\
&= E\left(Y_t \middle| \varepsilon_{t-1}, \cdots, \varepsilon_1\right) \\
&= E\left(\varepsilon_t + \cdots + \varepsilon_1 \middle| \varepsilon_{t-1}, \cdots, \varepsilon_1\right) \\
&= E\left(\varepsilon_t \middle| \varepsilon_{t-1}, \cdots, \varepsilon_1\right) + \varepsilon_{t-1} + \cdots + \varepsilon_1 \\
&= \varepsilon_{t-1} + \cdots + \varepsilon_1 \\
&= Y_{t-1}
\end{aligned} \tag{8.8}$$

The first equality comes from which $\left(Y_{t-1}, \cdots, Y_1\right)$ and $\left(\varepsilon_{t-1}, \cdots, \varepsilon_1\right)$ have the same information, the fourth equality comes from $E\left(\varepsilon_t \middle| \varepsilon_{t-1}, \cdots, \varepsilon_1\right) = 0$.

Conversely, if $\{Y_t\}$ is a martingale, $\{\varepsilon_t\}$ can be backed out by taking first differences:

$$\varepsilon_1 = Y_1, \quad \varepsilon_2 = Y_2 - Y_1, \quad \cdots, \quad \varepsilon_t = Y_t - Y_{t-1}, \quad \cdots \tag{8.9}$$

And $\{\varepsilon_t\}$ is a MDS.

Note: A MDS has no correlation.
Proof:

Suppose $\{\varepsilon_t\}$ is a MDS, so for any j>0, we have

$$E\left(\varepsilon_t \middle| \varepsilon_{t-1}, \cdots, \varepsilon_1\right) = E\left(\varepsilon_t\right) = 0$$

$$\begin{aligned}
&Cov\left(\varepsilon_t, \varepsilon_{t-j}\right) \\
&= E\left(\varepsilon_t \varepsilon_{t-j}\right) - E\left(\varepsilon_t\right) E\left(\varepsilon_{t-j}\right) \\
&= E\left[E\left(\varepsilon_t \varepsilon_{t-j} \middle| \varepsilon_{t-1}, \varepsilon_{t-2}, \cdots, \varepsilon_1\right)\right] \\
&= E\left[E\left(\varepsilon_t \middle| \varepsilon_{t-1}, \cdots, \varepsilon_{t-j}, \cdots, \varepsilon_1\right) \varepsilon_{t-j}\right] \\
&= 0
\end{aligned}$$

The third equality comes from the linearity of conditional expectation.

8.1.5 ARCH Processes

A process $\{\varepsilon_t\}$ is said to be an ARCH(1) if it can be written as

$$\begin{cases} \varepsilon_t = \sigma_t u_t \\ \sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \\ u_t \sim i.i.d.(0,1) \end{cases} \tag{8.10}$$

If ε_1 is the initial value of the process, it is then easy to show that $\varepsilon_t\,(t \geq 2)$ is a function of ε_1 and $(u_2, \cdots, u_t)$, therefore u_t is independent of $(\varepsilon_1, \cdots, \varepsilon_{t-1})$.

Now, we proof that $\{\varepsilon_t\}$ is a MDS.

$$\begin{aligned} & E\left[\varepsilon_t \middle| \varepsilon_{t-1}, \varepsilon_{t-2}, \cdots, \varepsilon_1\right] \\ &= E\left[\sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2} \cdot u_t \middle| \varepsilon_{t-1}, \varepsilon_{t-2}, \cdots, \varepsilon_1\right] \\ &= \sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2}\, E\left[u_t \middle| \varepsilon_{t-1}, \varepsilon_{t-2}, \cdots, \varepsilon_1\right] \\ &= \sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2}\, E\left[u_t\right] \\ &= 0 \end{aligned}$$

By a similar argument, it follows that

$$E\left(\varepsilon_t^2 \middle| \varepsilon_{t-1}, \cdots, \varepsilon_1\right) = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \tag{8.11}$$

Since $E\left[\varepsilon_t \middle| \varepsilon_{t-1}, \varepsilon_{t-2}, \cdots, \varepsilon_1\right] = 0$, the conditional second moment is just the conditional variance, it is a function of its own history of the process. In this sense, the process exhibits own conditional heteroscedasticity. It can be shown (Engle, 1982) that the process is strictly stationary and ergodic if $|\alpha_1| < 1$, provide that ε_1 is a draw from an appropriate distribution or provide that the process started in the infinite past.

4) Conclusions

Now we have three formulations for covariance stationary processes. They are, in the order of strength,

$\{\varepsilon_t\}$ is i.i.d. with zero-mean and finite variance.

$\Rightarrow$ $\{\varepsilon_t\}$ is stationary MDS with finite variance.

$\Rightarrow$ $\{\varepsilon_t\}$ is white noise.

8.2 Autocorrelated Disturbances

8.2.1 General Formulation

In the usual time-series setting the disturbances are assumed to be:

i) Homoscedastic but correlated across observations.

ii) Stationary, i.e. the covariance between observations t and s is a function only of $|t-s|$.

So that $Y = X\beta + \varepsilon$

$$E\left(\varepsilon\varepsilon' \middle| X\right) = \sigma^2\Omega \tag{8.12}$$

Where $\sigma^2\Omega$ is a full positive matrix by above two assumptions.

$$\sigma^2\Omega = \left(\gamma_{|t-s|}\right) \tag{8.13}$$

It is just the matrix form of equation (8.3).

8.2.2 AR(1) Process

Suppose $Y_t = X_t'\beta + \varepsilon_t \quad t = 1, \cdots, T$

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t \tag{8.14}$$

$|\rho| < 1$, $\{u_t\}$: a white noise series, $Cov\left(\varepsilon_{t-1}, u_t\right) = 0$.

$$\gamma_0 = E\left(\varepsilon_t^2\right) = E\left[\rho\varepsilon_{t-1} + u_t\right]^2 \Rightarrow \sigma_\varepsilon^2 = \rho^2\sigma_\varepsilon^2 + \sigma_u^2 \Rightarrow \sigma_\varepsilon^2 = \frac{\sigma_u^2}{1-\rho^2}$$

So $\gamma_1 = E\left(\varepsilon_t\varepsilon_{t-1}\right) = \rho\sigma_\varepsilon^2$

$$\vdots$$

$$\gamma_s = E\left(\varepsilon_t\varepsilon_{t-s}\right) = E\left[\left(\rho\varepsilon_{t-1} + u_t\right)\varepsilon_{t-s}\right] = \rho^s\sigma_\varepsilon^2$$

$$\therefore \quad \sigma^2\Omega = \frac{\sigma_u^2}{1-\rho^2}\begin{pmatrix} 1 & \rho & \cdots & \rho^{T-1} \\ \rho & 1 & \cdots & \rho^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \cdots & 1 \end{pmatrix} \tag{8.15}$$

8.2.3 MA(1) Process

$\varepsilon_t = u_t - \lambda u_{t-1}$..(8.16)

The assumptions of $\{u_t\}$ are the same as AR(1) process.

$$
\begin{aligned}
\gamma_0 &= E\left(u_t - \lambda u_{t-1}\right)^2 = \left(1+\lambda^2\right)\sigma_u^2 \\
\gamma_1 &= E\left[\left(u_t - \lambda u_{t-1}\right)\left(u_{t-1} - \lambda u_{t-2}\right)\right] = -\lambda\sigma_u^2 \\
\gamma_2 &= E\left[\left(u_t - \lambda u_{t-1}\right)\left(u_{t-2} - \lambda u_{t-3}\right)\right] = 0 \\
&\vdots \\
\gamma_j &= 0 \quad j > 1
\end{aligned}
$$

$$
\therefore \quad \sigma^2\Omega = \sigma_u^2 \begin{pmatrix} 1+\lambda^2 & -\lambda & 0 & 0 & \cdots & 0 \\ -\lambda & 1+\lambda^2 & -\lambda & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1+\lambda^2 \end{pmatrix} \tag{8.17}
$$

8.3 Reasons for Autocorrelated Disturbances (Leave Out)

8.4 Testing for Autocorrelation

8.4.1 The Durbin-Watson Test

1) DW statistic

Suppose that in the model (8.14), one suspects that the disturbance follows an AR(1), then, we have

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

The Durbin-Watson test statistic is computed from the vector of OLS residuals $e = Y - X\hat{\beta}$ and is defined as

$$d = \frac{\sum_{t=2}^{T}\left(e_t - e_{t-1}\right)^2}{\sum_{t=1}^{T} e_t^2} = 2\left(1-r\right) - \frac{e_1^2 + e_T^2}{\sum_{t=1}^{T} e_t^2}, \text{ where } r = \hat{\rho} = \frac{\sum_{t=2}^{T} e_t e_{t-1}}{\sum_{t=1}^{T} e_t^2}$$

For a large T, then the last term will be negligible, leaving

$$d \simeq 2\left(1-r\right) \tag{8.18}$$

Because any computed d value depends on the associated X matrix, exact critical values of d that will cover all empirical applications cannot be tabulated. Dubin and Watson established upper (dU) and lower (dL) bounds for the critical values. The testing procedure is as follows:

i) If d<dL, reject the null hypothesis in favor of the hypothesis of positive first-order autocorrelation.

ii) If d>4-dL, reject the null hypothesis in favor of the hypothesis of negative first-order autocorrelation.

iii) dU<d<4-dU, do not reject the null hypothesis.

iv) dL<d<dU, 4-dU<d<4-dL, the test is inconclusive.

2) The expansions of DW statistic

i) Savin and White's extensive tables: $6 \le n \le 200, \quad k \le 10$.

ii) Farebrother's table: for no constant model, dL to be replaced by dM.

iii) Wallis Test: for fourth-order autocorrelation

$$\varepsilon_t = \rho_4 \varepsilon_{t-4} + u_t$$
$$H_0 : \rho_4 = 0 \quad H_1 : \rho_4 \neq 0$$
$$d_4 = \sum_{t=5}^{T} \left(e_t - e_{t-4}\right)^2 \Big/ \sum_{t=1}^{T} e_t^2$$

8.4.2 Testing in the Presence of a Lagged Dependent Variable

There are two important qualifications to the use of DW test. First, it is necessary to include a constant term in the model. Second, it is strictly valid only for a non-stochastic matrix X. Thus, the DW Test is not likely to be valid when there is a lagged dependent variable in the equation. Consider the relation

$$Y_t = \beta_1 Y_{t-1} + \cdots + \beta_s Y_{t-s} + \beta_{s+1} X_{t1} + \cdots + \beta_{s+k} X_{tk} + \varepsilon_t \tag{8.19}$$

with $\varepsilon_t = \rho \varepsilon_{t-1} + u_t \quad |\rho| < 1 \quad u \sim N\left(0, \sigma^2 I\right)$.

$$H_0 : \rho = 0$$

The statistic is:

$$h = \hat{\rho} \sqrt{\frac{T}{1 - T \cdot \operatorname{var}\left(\hat{\beta}_1\right)}} \overset{a}{\sim} N(0,1) \tag{8.20}$$

Where $\hat{\rho}$ can be computed from the residuals e_t or from (8.18), use the approximation $\hat{\rho} = 1 - \frac{d}{2}$, when the DW statistic has been computed.The test breaks down if $\operatorname{var}\left(\hat{\beta}_1\right) \ge \frac{1}{T}$, an alternative is to regress e_t on $X_{t1}, \cdots, X_{tk}, Y_{t-1}, \cdots, Y_{t-s}, e_{t-1}$; if the coefficient of e_{t-1} in this regress is significantly different from zero by t test, reject the null hypothesis $H_0 : \rho = 0$.

Durbin indicates that this last procedUre can be extended to test for AR(p):

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \cdots + \rho_p \varepsilon_{t-p} + u_t \tag{8.21}$$
$$H_0 : \rho_1 = \cdots = \rho_p = 0$$

Let $X_{0t} = \left(Y_{t-1}, \cdots, Y_{t-s}, X_{t1}, \cdots, X_{tk}\right)$, and $e = Y - X_0 \left(X_0' X_0\right)^{-1} X_0' Y$.

Define $E_p = \left(e_1, e_2, \cdots, e_p\right) = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ e_1 & 0 & \cdots & 0 \\ e_2 & e_1 & \cdots & 0 \\ & & & \vdots \\ \vdots & \vdots & \ddots & e_1 \\ & & & \vdots \\ e_{T-1} & e_{T-2} & \cdots & e_{T-p} \end{bmatrix}$

In the second-stage regression, the relevant restricted regression is:

Regress e on X_0

The unrestricted regression is:

Regress e on $E_p \quad X_0$

Therefore, if SSE_U and SSE_R are Error Sum of Squares from the above two regressions respectively, we have

$$F = \frac{\left(SSE_R - SSE_U\right)/p}{SSE_U / \left[T - \left(s+K\right) - p\right]} \qquad (8.22)$$

Since e is the residual vector from the original regression (8.19), the regressors X_0 in the restricted regression have no explanatory power, so

$SSR_R = 0$ and $SSE_R = e'e$,

$$F = \frac{SSR_U / p}{SSE_U / \left[T - \left(s+K\right) - p\right]} \qquad (8.23)$$

The SSR_U from the unrestricted regression is

$$\begin{aligned} SSR_U &= e'\begin{pmatrix} E_p & X_0 \end{pmatrix} \begin{pmatrix} E_p'E_p & E_p'X_0 \\ X_0'E_p & X_0'X_0 \end{pmatrix}^{-1} \begin{pmatrix} E_p' \\ X_0' \end{pmatrix} e \\ &= \begin{pmatrix} e'E_p & 0 \end{pmatrix} \begin{pmatrix} E_p'E_p & E_p'X_0 \\ X_0'E_p & X_0'X_0 \end{pmatrix}^{-1} \begin{pmatrix} E_p'e \\ 0 \end{pmatrix} \\ &= e'E_p \left[E_p'E_p - E_p'X_0 \left(X_0'X_0\right)^{-1} X_0'E_p \right]^{-1} E_p'e \end{aligned} \qquad (8.24)$$

So the F statistic to test the joint significance of the coefficients on the lagged residuals is then

$$F = \frac{e'E_p \left[E_p'E_p - E_p'X_0 \left(X_0'X_0\right)^{-1} X_0'E_p \right]^{-1} E_p'e / p}{SSE_U / \left[T - \left(s+K\right) - p\right]} \qquad (8.25)$$

This statistic does not have exact, finite sample validity since the regressor matrix in the both restricted and unrestricted regressions is stochastic. As $T \to \infty$, $p \cdot F$ tends in distribution to $\chi^2(p)$.

8.4.3 The Breusch-Goldfrey Test (LM Test)

We first discuss the Breusch-Goldfrey Test for AR(1)

Suppose the specified equation is

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t \tag{8.26}$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t \tag{8.27}$$

Where $|\rho| < 1, \quad u_t \sim IIN\left(0, \sigma_u^2\right)$.

Substituting (8.27) in (8.26) gives

$$y = \beta_1 (1-\rho) i + \beta_2 x + \rho y_{-1} - \beta_2 \rho x_{-1} + u = F\left(\beta; x, x_{-1}, y_{-1}\right) + u \tag{8.28}$$

Where $\beta = (\beta_1, \beta_2, \rho)$ is the true parameter vector.

We wish to test the hypothesis: $H_0 : \rho = 0$ (no autocorrelation)

The log-likelihood function for (8.28) is

$$\ln L = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \sigma_u^2 - \frac{\left[y - F\left(\beta; x, x_{-1}, y_{-1}\right)\right]'\left[y - F\left(\beta; x, x_{-1}, y_{-1}\right)\right]}{2\sigma_u^2}$$

The information matrix for this type of regression model is block diagonal, so the parameter vector β can be treated separately from σ_u^2. The score vector is then

$$\frac{\partial \ln L}{\partial \beta} = -\frac{\partial \left\{ \frac{\left[y - F\left(\beta; x, x_{-1}, y_{-1}\right)\right]'\left[y - F\left(\beta; x, x_{-1}, y_{-1}\right)\right]}{2\sigma_u^2} \right\}}{\partial \beta}$$

$$= -\frac{\partial \left[\sum_{t=1}^{n} u_t^2 \Big/ 2\sigma_u^2 \right]}{\partial \beta} = -\frac{1}{\sigma_u^2} \sum_{t=1}^{n} u_t \frac{\partial u_t}{\partial \beta} = \frac{1}{\sigma_u^2} \sum_{t=1}^{n} u_t w_t$$

Where $w_t = -\partial u_t / \partial \beta$. The information matrix is

$$I(\beta) = E\left[\left(\frac{\partial \ln L}{\partial \beta}\right)\left(\frac{\partial \ln L}{\partial \beta'}\right)\right] = E\left[\frac{1}{\sigma_u^4}\left(\sum u_t w_t\right)\left(\sum u_t w_t\right)'\right]$$

$$= E\left[\frac{1}{\sigma_u^4}\left(\sum u_t^2 w_t w_t' + \sum_{t \neq s} u_t u_s w_t w_s'\right)\right] = \frac{1}{\sigma_u^2} E\left(\sum w_t w_t'\right)$$

Where the last line follows the assumptions about u's. Asymptotically it makes no difference if $E\left(\sum w_t w_t'\right)$ replaced by $\sum w_t w_t'$.

Under H_0, $y_t = \beta_1 + \beta_2 x_t + u_t$, thus the LM statistic is

$$LM = \frac{1}{\tilde{\sigma}_u^2}\left(\sum \tilde{u}_t \tilde{w}_t\right)'\left(\sum \tilde{w}_t \tilde{w}_t'\right)^{-1}\left(\sum \tilde{u}_t \tilde{w}_t\right) \qquad (8.29)$$
$$= \frac{1}{\tilde{\sigma}_u^2}\tilde{u}'\tilde{w}\left(\tilde{w}'\tilde{w}\right)^{-1}\tilde{w}'\tilde{u}$$

Where $\tilde{w} = \begin{pmatrix} \cdots & \tilde{w}_{1i} & \cdots \\ \cdots & \tilde{w}_{2i} & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & \tilde{w}_{ni} & \cdots \end{pmatrix}$ $\left(i = 1,2,3\right)$, and $\tilde{u} = \begin{pmatrix} \tilde{u}_1 \\ \tilde{u}_2 \\ \vdots \\ \tilde{u}_n \end{pmatrix}$.

The "$\sim$" indicate that all elements in (8.29) are evaluated at the restricted estimates $\tilde{\beta}$, $\tilde{\sigma}_u^2 = \tilde{u}'\tilde{u}/n$, so $LM = n\tilde{R}^2$ .Where $\tilde{R}^2$ is the coefficient of determination from the regression of $\tilde{u}$ on $\tilde{w}$.

Under H_0, $\tilde{u}_t = y_t - \tilde{\beta}_1 - \tilde{\beta}_2 x_t$,

$$w_t = \begin{pmatrix} -\partial u_t / \partial \beta_1 \\ -\partial u_t / \partial \beta_2 \\ -\partial u_t / \partial \rho \end{pmatrix} = \begin{pmatrix} 1-\rho \\ x_t - \rho x_{t-1} \\ y_{t-1} - \beta_1 - \beta_2 x_{t-1} \end{pmatrix}$$

$$\tilde{w}_t = \begin{pmatrix} 1 \\ x_t \\ \tilde{u}_{t-1} \end{pmatrix}$$

The test of $\rho = 0$ is therefore obtained in two steps:

Step 1: Apply OLS to (8.29) to obtain the residual $\tilde{u}_t$ (labeled e_t).

Step 2: Regress e_t on $\left[1, x_t, e_{t-1}\right]$ to find $\tilde{R}^2$.

Under H_0, $n\tilde{R}^2 \sim \chi^2(1)$.

This procedure easily extends to testing for higher orders of autocorrelation, including AR(p) and MA(q). It may be seen that $\tilde{w}$ in (8.29) is the $\left(E_p \quad X_0\right)$ matrix in the unrestricted regression in Durbin-h test, and $\tilde{u}$ is e. Thus, using the unrestricted regression ,the LM statistic in (8.29) is

$$LM = \frac{e'E_p\left[E_p'E_p - E_p'X_0\left(X_0'X_0\right)^{-1}X_0'E_p\right]^{-1}E_p'e}{e'e/T} = TR_0^2 \qquad (8.29')$$

Where R_0^2 is the coefficient determinant from the regression of e on $\left(E_p \quad X_0\right)$. The only difference between the Durbin Statistic in (8.25) and the LM test in (8.29') is in the variance term in the denominator. Breusch shows that these terms have the same probability limit and so the two procedures are asymptotically equivalent.

In the concrete, it can be shown that the $p \cdot F$ Statistic in (8.25) is asymptotically equivalent to TR_0^2 in (8.29').

8.4.4 Box-Pierce-Ljung Statistic

1) Special case of Theorem 6.7 of Hall and Heyde (1980)

Suppose $\{Z_t\}$ can be written as $\mu + \varepsilon_t$, where ε_t is MDS with "own" conditional homoscedasticity:

$$E\left(\varepsilon_t^2 \middle| \varepsilon_{t-1}, \varepsilon_{t-2}, \cdots\right) = \sigma^2 \quad \sigma^2 > 0 \tag{8.30}$$

The sample j-th order autocorrelation coefficient $\hat{\rho}_j$ is defined as

$$\hat{\rho}_j = \frac{\hat{\gamma}_j}{\hat{\gamma}_0} \quad j = 1, 2, \cdots \tag{8.31}$$

Where

$$\hat{\gamma}_j = \frac{1}{T}\sum_{t=j+1}^{T}\left(Z_t - \bar{Z}_T\right)\left(Z_{t-j} - \bar{Z}_T\right) = \frac{1}{T}\sum_{t=j+1}^{T} e_t e_{t-j} \tag{8.32}$$

$$\bar{Z}_T = \frac{1}{T}\sum_{t=1}^{T} Z_t$$

Then $\sqrt{T}\hat{\gamma} \xrightarrow{d} N\left(0, \sigma^4 I_p\right)$ and $\sqrt{T}\hat{\rho} \xrightarrow{d} N\left(0, I_p\right)$.

Where $\hat{\gamma} = \left(\hat{\gamma}_1, \hat{\gamma}_2, \cdots, \hat{\gamma}_p\right)'$, $\quad \hat{\rho} = \left(\hat{\rho}_1, \hat{\rho}_2, \cdots, \hat{\rho}_p\right)'$.

Now, we can test whether $\hat{\rho} = \left(\hat{\rho}_1, \hat{\rho}_2, \cdots, \hat{\rho}_p\right)'$ are simultaneously zero. Since the elements of $\sqrt{T}\hat{\rho}$ are asymptotically independent and individually distributed as standard normal, the Box-Pierce Q (Box and Pierce, 1970), is asymptotically Chi-squared:

$$Q = T\sum_{j=1}^{p}\hat{\rho}_j^2 = \sum_{j=1}^{p}\left(\sqrt{T}\hat{\rho}_j\right)^2 \xrightarrow{d} \chi^2\left(p\right) \tag{8.33}$$

And the modification of Box-Pierce Q, called the Ljung-Box Q', is asymptotically equivalent in that its difference from the Box-Pierce Q vanishes in large sample. This statistic is:

$$Q' = T\left(T+2\right)\sum_{j=1}^{p}\frac{1}{T-j}\hat{\rho}_j^2 = \sum_{j=1}^{p}\frac{T+2}{T-j}\left(\sqrt{T}\hat{\rho}_j\right)^2 \xrightarrow{d} \chi^2\left(p\right) \qquad (8.34)$$

The Ljung-Box Q' often provides a better approximation to the χ^2 distribution for moderate samples.

2) Sample autocorrelations calculated from residuals

Suppose model

$$Y_t = X_t'\beta + \varepsilon_t \qquad (8.35)$$

Satisfying the following set of assumptions:

A1. The (K+1)-dimensional vector stochastic process $\left\{Y_t \quad X_t\right\}$ is jointly stationary and ergodic.

A2. All the regressors are orthogonal to the contemporaneous disturbance:

$E\left(X_{tk}\varepsilon_t\right) = 0$ for all t and k (k=1,⋯,K).

A3. The matrix $E\left(X_t X_t'\right) = \Sigma_{XX}$ is nonsingular and finite.

A4. $\left\{X_t\varepsilon_t\right\}$ is a MDS with finite second moments.

If the error term ε_t were observable, we would calculate the sample autocorrelations as

$$\tilde{\rho}_j = \frac{\tilde{\gamma}_j}{\tilde{\gamma}_0} \quad j = 1, 2, \cdots \qquad (8.31')$$

Where $\tilde{\gamma}_j = \frac{1}{T}\sum_{t=j+1}^{T}\varepsilon_t\varepsilon_{t-j}$ $\qquad (8.32')$

Now consider the realistic case where we replace ε_t in the formula $\left(8.32'\right)$ by the OLS estimate e_t and calculate the sample autocorrelations defined as (8.31). Because the regressors include a constant, we have $\bar{e} = 0$. So there is no need to subtract the sample mean in the calculating of $\hat{\gamma}_j$. Under this situation, only if the regressors are strictly exogenous, the residual-based Q statistic derived from $\left\{\hat{\rho}_j\right\}$ is all right for testing serial correlation. Using the following relation: $e_t = \varepsilon_t - X_t'\left(\hat{\beta}-\beta\right)$, we have

$$\begin{aligned}\hat{\gamma}_j &= \frac{1}{T}\sum_{t=j+1}^{T} e_t e_{t-j} \\ &= \frac{1}{T}\sum_{t=j+1}^{T}\left[\varepsilon_t - X_t'\left(\hat{\beta}-\beta\right)\right]\left[\varepsilon_{t-j} - X_{t-j}'\left(\hat{\beta}-\beta\right)\right] \\ &= \tilde{\gamma}_j - \frac{1}{T}\sum_{t=j+1}^{T}\left(X_{t-j}\varepsilon_t + X_t\varepsilon_{t-j}\right)'\left(\hat{\beta}-\beta\right) \\ &\qquad + \left(\hat{\beta}-\beta\right)'\left(\frac{1}{T}\sum_{t=j+1}^{T} X_t X_{t-j}'\right)\left(\hat{\beta}-\beta\right)\end{aligned} \quad (8.36)$$

Where $\tilde{\gamma}_j = \frac{1}{T}\sum_{t=j+1}^{T}\varepsilon_t\varepsilon_{t-j}$.

If $E\left(X_t\varepsilon_{t-j}\right)$, $E\left(X_{t-j}\varepsilon_t\right)$ and $E\left(X_t X_{t-j}'\right)$ are all finite, then the second and the third terms converge to zero in probability because $\hat{\beta}\xrightarrow{p}\beta$. Therefore

$$\hat{\gamma}_j - \tilde{\gamma}_j \xrightarrow{P} 0 \quad j = 0,1,2,\cdots \quad (8.37)$$

However $\sqrt{T}$ times the difference does not. Now, by multiplying both sides of (8.36) by $\sqrt{T}$, we obtain

$$\begin{aligned}\sqrt{T}\hat{\gamma}_j &= \sqrt{T}\left(\frac{1}{T}\sum_{t=j+1}^{T}\varepsilon_t\varepsilon_{t-j}\right) - \frac{1}{T}\sum_{t=j+1}^{T}\left(X_{t-j}\varepsilon_t + X_t\varepsilon_{t-j}\right)'\sqrt{T}\left(\hat{\beta}-\beta\right) \\ &\qquad + \sqrt{T}\left(\hat{\beta}-\beta\right)'\left(\frac{1}{T}\sum_{t=j+1}^{T} X_t X_{t-j}'\right)\left(\hat{\beta}-\beta\right)\end{aligned} \quad (8.38)$$

Where $\sqrt{T}\left(\hat{\beta}-\beta\right)$ converges to a normal random variable, by the Lemma "$X_n \xrightarrow{p} 0, Y_n \xrightarrow{d} Y \Rightarrow X_n Y_n \xrightarrow{p} 0$",the third term on the right hand side vanishes. Regarding the second term, we have

$$\frac{1}{T}\sum_{t=j+1}^{T}\left(X_{t-j}\varepsilon_t + X_t\varepsilon_{t-j}\right) \xrightarrow{P} E\left(X_{t-j}\varepsilon_t\right) + E\left(X_t\varepsilon_{t-j}\right) \quad (8.39)$$

If the regressors are strictly exogenous in the sense that $E\left(X_t\varepsilon_s\right) = 0$ for all t and s, then

$$E\left(X_{t-j}\varepsilon_t\right) + E\left(X_t\varepsilon_{t-j}\right) = 0 \quad (8.40)$$

The second term converges to zero in probability, and thus

$$\sqrt{T}\hat{\gamma}_j - \sqrt{T}\tilde{\gamma}_j \xrightarrow{P} 0$$

Since $\tilde{\gamma}_0 = \frac{1}{T}\sum \varepsilon_t^2 \xrightarrow{P} \sigma^2, \quad \hat{\gamma}_0 = \frac{1}{T}\sum e_t^2 \xrightarrow{P} \sigma^2$,

$$\sqrt{T}\hat{\rho}_j - \sqrt{T}\tilde{\rho}_j = \sqrt{T}\frac{\hat{\gamma}_j}{\hat{\gamma}_0} - \sqrt{T}\frac{\tilde{\gamma}_j}{\tilde{\gamma}_0} \xrightarrow{P} 0 \qquad (8.41)$$

(8.41) means that the Q statistic calculated from the regression residuals $\{e_t\}$, too, is asymptotically χ^2 distributed, and we can use this residual-based Q to test for serial correlation.

3) Testing with predetermined, but not strictly exogenous, regressors

When the regressors are not strictly exogenous, there is no guarantee that (8.40) holds. Consequently, we need to modify the Q statistic to restore its asymptotical χ^2 distribution. For this purpose, consider two restrictions

i) $E\left(\varepsilon_t \middle| \varepsilon_{t-1}, \varepsilon_{t-2}, \cdots, X_t, X_{t-1}, \cdots\right) = 0 \qquad (8.42)$

ii) $E\left(\varepsilon_t^2 \middle| \varepsilon_{t-1}, \varepsilon_{t-2}, \cdots, X_t, X_{t-1}, \cdots\right) = \sigma^2 \qquad (8.43)$

where restriction (i) implies that $\{X_t\varepsilon_t\}$ is an MDS and restriction (ii) is stronger than the own conditional homoscedasticity assumption because the conditioning set includes current and past X as well as past ε.

Suppose that A.1, A.3, (8.42) and (8.43) are satisfied, the sample autocorrelation of the OLS residuals be defined as in (8.31).

Then

$$\sqrt{T}\hat{\gamma} \xrightarrow{d} N\left[0, \sigma^4\left(I_p - \Phi\right)\right]$$

and

$$\sqrt{T}\hat{\rho} \xrightarrow{d} N\left[0, \left(I_p - \Phi\right)\right] \qquad (8.44)$$

Where $\hat{\gamma} = \left(\hat{\gamma}_1, \hat{\gamma}_2, \cdots, \hat{\gamma}_p\right)'$, $\hat{\rho} = \left(\hat{\rho}_1, \hat{\rho}_2, \cdots, \hat{\rho}_p\right)'$, and $\Phi = \left(\phi_{jk}\right)$ is given by

$$\phi_{jk} = E\left(X_t\varepsilon_{t-j}\right)' E\left(X_t X_t'\right)^{-1} E\left(X_t\varepsilon_{t-k}\right) \Big/ \sigma^2 \qquad (8.45)$$

By the Ergodic Theorem, matrix Φ is consistently estimated by its sample counterpart:

$$\hat{\Phi} = \left(\hat{\phi}_{jk}\right), \quad \hat{\phi}_{jk} = \bar{\mu}_j' S_{XX}^{-1} \bar{\mu}_k \Big/ S^2, \quad j,k = 1,2,\cdots.p \qquad (8.46)$$

Where $S^2 = \frac{1}{T-K}\sum_{t=1}^{T} e_t^2, \quad \bar{\mu}_j = \frac{1}{T}\sum_{t=j+1}^{T} X_t e_{t-j}$.

It follows from this and (8.44) that:

Modified $Q = T\hat{\rho}'\left(I_p - \hat{\Phi}\right)^{-1}\hat{\rho} \xrightarrow{d} \chi^2(p)$ (8.47)

8.5 Estimation for Autocorrelation Model

8.5.1 GLS Estimation

We now assume that the specification $Y_t = X_t'\beta + \varepsilon_t$ is associated autocorrelation structure, by far the most common assumption is an AR(1) process. So the covariance matrix for ε is (8.14), where

$$\Omega = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & \rho & \cdots & \rho^{T-1} \\ \rho & 1 & \cdots & \rho^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \cdots & 1 \end{pmatrix} \qquad (8.48)$$

$$\Omega^{-1} = \begin{pmatrix} 1 & -\rho & 0 & \cdots & 0 & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & 0 & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\rho & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \cdots & 0 & -\rho & 1 \end{pmatrix} \qquad (8.49)$$

It can be seen that the matrix

$$P = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 & 0 \\ -\rho & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{pmatrix} \qquad (8.50)$$

satisfies the condition $\Omega^{-1} = P'P$. If ρ be known, there are two equivalent ways of deriving GLS estimates of β. One is to substitute ρ in (8.49) and compute $\hat{\beta}_{GLS} = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}Y$ directly. The alternative is to transform the data by pre-multiplication by the P matrix and then estimate the OLS regression of $Y_*(=PY)$ on $X_*(=PX)$. For the AR(1) case, data for the transformed model are:

$$Y_* = \begin{pmatrix} \sqrt{1-\rho^2} Y_1 \\ Y_2 - \rho Y_1 \\ \vdots \\ Y_T - \rho Y_{T-1} \end{pmatrix}, \quad X_* = \begin{pmatrix} \sqrt{1-\rho^2} X_1' \\ X_2' - \rho X_1' \\ \vdots \\ X_T' - \rho X_{T-1}' \end{pmatrix} \tag{8.51}$$

Where $X_t = \left(X_{t1}, \cdots, X_{tK} \right)'$. The data transformed model also can be written as:

$$\begin{cases} \sqrt{1-\rho^2} Y_1 = \sqrt{1-\rho^2} \beta_1 + \sqrt{1-\rho^2} \beta_2 X_{12} + \cdots + \sqrt{1-\rho^2} \beta_K X_{1K} + \sqrt{1-\rho^2} \varepsilon_1 \\ Y_t - \rho Y_{t-1} = \beta_1 (1-\rho) + \beta_2 \left(X_{t2} - \rho X_{t-1,2} \right) + \cdots \\ \qquad + \beta_K \left(X_{tK} - \rho X_{t-1,K} \right) + \left(\varepsilon_t - \rho \varepsilon_{t-1} \right) \qquad (t = 2, \cdots, T) \end{cases} \tag{8.52}$$

It is easy to be shown that the disturbances in (8.52):

$$u_t = \begin{cases} \sqrt{1-\rho^2} \varepsilon_1, & t = 1 \\ \varepsilon_t - \rho \varepsilon_{t-1}, & t = 2, \cdots, T \end{cases}$$

are spherical disturbances.

If, however, the first row in P is dropped out, the regression between variables would simply be the second formula in (8.52). In small samples dropping the first observation can have a marked effect on the coefficient estimate, although asymptotically it is of little importance.

Corresponding results have been derived for higher-order autoregressive process, see Green, 5th, P272.

8.5.2 FGLS Estimation

In practice, Ω usually is unknown, we must estimate the $\Omega(\rho)$ along with the other parameters of the model; and all that is needed for efficient estimation of β is a consistent estimator of $\Omega(\rho)$.

For an AR(1) process, the most common procedure is to begin FGLS with a natural estimator of ρ.

(i) $r = \hat{\rho} = \dfrac{\sum_{t=2}^{T} e_t e_{t-1}}{\sum_{t=1}^{T} e_t^2}$ (e_t: OLS residual)

(ii) $\hat{\rho}^* = \dfrac{T-K}{T-1} \hat{\rho}$ (Theil Estimator)

(iii) $\hat{\rho}^{**} = 1 - \dfrac{d}{2}$ (Durbin-Watson Estimator)

(iv) The slope on Y_{t-1} in a regression of Y_t on Y_{t-1}, X_t and X_{t-1}.

After ρ be estimated, there are two procedures of calculating FGLS estimates of β.

(i) The Cochrane-Orcutt (C-O) procedure

This procedure is to substitute the estimated ρ in the second formula in (8.52), which is equivalent to dropping the first row of the P matrix in (8.50).

(ii) The Prais and Winsten procedure

This procedure is to use the full P matrix, so that the first observation receives exploit treatment.

It is possible to iterate any of the estimators to convergence.

(i) The C-O iterated procedure

The second formula in (8.52) can be rearranged in two equivalent forms as

$$\left(Y_t - \rho Y_{t-1}\right) = \beta_1\left(1-\rho\right) + \beta_2\left(X_{t2} - \rho X_{t-1,2}\right) + \cdots + \beta_K\left(X_{tK} - \rho X_{t-1,K}\right) + u_t \quad (8.53a)$$

$$\left(Y_t - \beta_1 - \beta_2 X_{t2} - \cdots - \beta_K X_{tK}\right) = \rho\left(Y_{t-1} - \beta_1 - \beta_2 X_{t-1,2} - \cdots - \beta_K X_{t-1,K}\right) + u_t \quad (8.53b)$$

We start the C-O iterative procedure with an estimate of $\hat{\rho}^{(1)}$ and then substitute it in (8.53a) and compute OLS regression, yielding estimated coefficients $\hat{\beta}_j^{(1)}$, $\left(j = 1,\cdots,K\right)$. These in turn are used to compute the variables in (8.53b); and an OLS regression yields a new estimate $\hat{\rho}^{(2)}$. The iteration continues until a satisfactory degree of convergence is reached.

(ii) The Prais and Winsten iterated procedure

The problem of this iterative procedure is that they may converge to a local minimum and not necessarily to the global minimum. A precaution is to fit equations like (8.53a) for a grid of ρ values in steps of 0.1 from, say, -0.9 to 0.9 and then iterate from the regression with the smallest SSE.

Note:

(i) The iterative procedure can be used directly to the following nonlinear model:

$$Y_t = \rho Y_{t-1} + \beta_1\left(1-\rho\right) + \beta_2\left(X_{t2} - \rho X_{t-1,2}\right) + \cdots + \beta_K\left(X_{tK} - \rho X_{t-1,K}\right) + u_t \quad (8.54)$$

The nonlinear least squares (NLS) is required for this model. In order to avoid the problem exits with Prais-Winsten iterated procedure, it is advisable to start the NLS process with several different coefficient vectors to see if convergence takes place at the same vector.

(ii) The iterated FGLS procedure does not yield ML estimates, even with special treatment of the first observation.

The log-likelihood in (8.14), namely

$$\ln L = -\frac{n}{2}\ln\left(2\pi\right) - \frac{1}{2}\ln\left|\sigma^2\Omega\right| - \frac{1}{2}\varepsilon'\left(\sigma^2\Omega\right)^{-1}\varepsilon$$

From the relations already defined in (8.14) and (8.48), it follows that

$$\ln L = -\frac{n}{2}\ln\left(2\pi\right) - \frac{n}{2}\ln\sigma_u^2 + \frac{1}{2}\ln\left(1-\rho^2\right) - \frac{1}{2\sigma_u^2}u'u \qquad (8.55)$$

Maximizing the log-likelihood takes account of term in $\ln\left(1-\rho^2\right)$, which is ignored in the GLS procedure. Beach and Mackinnon drew attention to this point and have devised an iterative procedure for maximizing (8.55), see Green,5th, P273.

8.6 Newey-West Autocorrelation Consistent Covariance Estimator

8.6.1 The CLT for serial dependence process

1) Ergodic Stationary Martingale Differences CLT

Let $\{Z_t\}$ be a vector MDS that is stationary and ergodic with $E(Z_t Z_t') = \Sigma$ and $\bar{Z}_T = \frac{1}{T}\sum_{t=1}^{T} Z_t$, then

$$\sqrt{T}\bar{Z}_T = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t \xrightarrow{d} N(0,\Sigma) \qquad (8.56)$$

This CLT, being applicable not just to i.i.d. sequences but also to stationary martingale differences such as ARCH(1) processes, is more general than Linderg-Levy.

2) Gordin's Central Limit Theorem

We now consider a CLT that is broad enough to include the case that interested us at the outset, stochastically dependent observations on X_t and autocorrelation in ε_t.

Gordin's conditions for Ergodic Stationary process:

i) Summability of autocovariances

With dependent observations,

$$\lim_{T\to\infty} Var\left(\sqrt{T}\bar{Z}_T\right) = \sum_{t=0}^{\infty}\sum_{s=0}^{\infty} Cov\left(Z_t Z_s'\right) = \sum_{j=-\infty}^{+\infty}\Gamma_j = \Gamma^* \qquad (8.57)$$

Where $Z_t = X_t\varepsilon_t$, Γ^* is a finite matrix. If the sum is to be finite, then $j=0$ term must be finite, which gives us a necessary condition $E(Z_t Z_t') = \Gamma_0$, a finite matrix.

ii) Asymptotic uncorrelated: $E\left[Z_t \middle| Z_{t-j}, Z_{t-j-1},\cdots\right] \xrightarrow{m.s.} 0$ as $j\to\infty$.

iii) Asymptotic negligibility of innovations

$$\sum_{j=0}^{+\infty}\left[E\left(r_{tj}'r_{tj}\right)\right]^{1/2} < \infty \quad\text{.........}(8.58)$$

Where

$$r_{tj} = E\left[Z_t \middle| Z_{t-j}, Z_{t-j-1},\cdots\right] - E\left[Z_t \middle| Z_{t-j-1}, Z_{t-j-2},\cdots\right] \quad\text{.........}(8.59)$$

It can be shown that $Z_t = \sum_{s=0}^{\infty} r_{ts}$:

$$Z_t = \left\{ E\left[Z_t \middle| Z_{t,} Z_{t-1}, \cdots \right] - E\left[Z_t \middle| Z_{t-1}, Z_{t-2}, \cdots \right] \right\} + \left\{ E\left[Z_t \middle| Z_{t-1}, Z_{t-2}, \cdots \right] - E\left[Z_t \middle| Z_{t-3}, Z_{t-4}, \cdots \right] \right\}$$
$$+ \left\{ E\left[Z_t \middle| Z_{t-3}, Z_{t-4}, \cdots \right] - E\left[Z_t \middle| Z_{t-5}, Z_{t-6}, \cdots \right] \right\} + \cdots + E\left[Z_t \middle| Z_{t-j}, Z_{t-j-1}, \cdots \right]$$

$$= r_{t0} + r_{t1} + r_{t2} + \cdots + r_{t,j-1} + E\left[Z_t \middle| Z_{t-j}, Z_{t-j-1}, \cdots \right]$$

$$E\left[Z_t \middle| Z_{t-j}, Z_{t-j-1}, \cdots \right] \xrightarrow{m.s.} 0$$

$$\therefore Z_t - \left(r_{t0} + r_{t1} + r_{t2} + \cdots + r_{t,j-1} \right) \xrightarrow{m.s.} 0$$

$$Z_t = \sum_{j=0}^{\infty} r_{tj}$$

The vector r_{tj} can be viewed as the information in this accumulated sum that entered the process at time t－j. This condition states that information eventually becomes negligible as it fades far back in time from the current observation.

Gordin's Central Limit Theorem:

If Gordin's condition holds for vector ergodic stationary process $\{Z_t\}$, then $E(Z_t) = 0$, $\{\Gamma_j\}$ is absolutely summable ,and $\sqrt{T}\bar{Z}_T \xrightarrow{d} N(0, \Gamma^*)$.

8.6.2 Large Sample Theory for Linear Regression under Conditional Heterocedasticity and Autocorrelation

1) Assumptions

A 8.6.1 $\{Y_t, X_t\}$ is a stationary ergodic process with $Y_t = X_t'\beta + \varepsilon_t$.

A 8.6.2 $E\left(\varepsilon_t \middle| X_t\right) = 0$

A 8.6.3 The $K \times K$ matrix $Q = E\left(X_t X_t'\right)$ is finite and nonsingular.

A 8.6.4 i) Put $\Gamma_j = Cov\left(Z_t, Z_{t-j}\right) = Cov\left(X_t\varepsilon_t, X_{t-j}\varepsilon_{t-j}\right) = E\left(X_t\varepsilon_t\varepsilon_{t-j}X_{t-j}'\right)$,

$$\Gamma^* = \sum_{j=-\infty}^{+\infty} \Gamma_j \qquad (8.60)$$

is finite and positive definite.

ii) $E\left[Z_t \middle| Z_{t-j}, Z_{t-j-1}, \cdots \right] = E\left[X_t\varepsilon_t \middle| X_{t-j}\varepsilon_{t-j}, X_{t-j-1}\varepsilon_{t-j-1}, \cdots \right] \xrightarrow{m.s.} 0$ as $j \to \infty$.

iii) $\sum_{j=0}^{\infty}\left[E\left(\gamma_{ij}'\gamma_{ij}\right)\right]^{1/2}<\infty$, where

$$\gamma_{ij}=E\left[X_t\varepsilon_t\middle|X_{t-j}\varepsilon_{t-j},X_{t-j-1}\varepsilon_{t-j-1},\cdots\right]-E\left[X_t\varepsilon_t\middle|X_{t-j-1}\varepsilon_{t-j-1},X_{t-j-2}\varepsilon_{t-j-2},\cdots\right]$$

Assumption 8.6.4 on $\{\varepsilon_t\}$ allows for both conditional heteroscedasticity and autocorrelation of unknown form.

2) Long-run Variance Estimation

i) Recall OLS Estimator, we have

$$\sqrt{T}\left(\hat{\beta}-\beta\right)=\hat{Q}^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}X_t\varepsilon_t$$

Suppose $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}X_t\varepsilon_t\xrightarrow{d}N\left(0,V\right)$.

Where V is the asymptotic variance

$$V=A\text{var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}X_t\varepsilon_t\right)$$

Then we have $\sqrt{T}\left(\hat{\beta}-\beta\right)\xrightarrow{d}N\left(0,Q^{-1}VQ^{-1}\right)$.

Now we consider the $\{\varepsilon_t\}$ under A8.6.1~A8.6.4, it is clear that Gordin's condition is satisfied for $\{Z_t\}=\{X_t\varepsilon_t\}$. So we have

$$\sqrt{T}\bar{Z}_T\xrightarrow{d}N\left(0,\Gamma^*\right)\qquad(8.61)$$

ii) How to Estimate Γ^*?

$$T\cdot Var\left(\bar{Z}_T\right)=Var\left(\sqrt{T}\bar{Z}_T\right)$$

$$=\frac{1}{T}\left[Cov\left(Z_1,Z_1+\cdots+Z_T\right)+\cdots+Cov\left(Z_T,Z_1+\cdots+Z_T\right)\right]$$

$$=\frac{1}{T}\left[\Gamma_0+\Gamma_1+\cdots+\Gamma_{T-1}\right]+\left[\Gamma_1+\Gamma_0+\cdots+\Gamma_{T-2}\right]+\cdots+\left[\Gamma_{T-1}+\cdots+\Gamma_1+\Gamma_0\right]$$

$$=\frac{1}{T}\left\{T\Gamma_0+2\left(T-1\right)\Gamma_1+\cdots+2\left(T-j\right)\Gamma_j+\cdots+2\Gamma_{T-1}\right\}$$

$$=\Gamma_0+2\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\Gamma_j$$

$$\approx\sum_{j=-(T-1)}^{T-1}\Gamma_j$$

$$\rightarrow \sum_{j=-\infty}^{+\infty} \Gamma_j = \Gamma^* \qquad (8.62)$$

where the last equality follows by Assumption A8.6.4. For (8.62), we can consider the estimator.

$$\hat{\Gamma} = \sum_{j=-(T-1)}^{T-1} \hat{\Gamma}_j \qquad (8.63)$$

Where

$$\hat{\Gamma}_j = \begin{cases} \frac{1}{T} \sum_{t=j+1}^{T} X_t e_t X'_{t-j} e_{t-j} & j = 0,1,\cdots,T-1 \\ \frac{1}{T} \sum_{t=1-j}^{T} X_{t+j} X'_t e_t e_{t+j} & j = -1,-2,\cdots,-(T-1) \end{cases} \qquad (8.64)$$

Unfortunately, although $\hat{\Gamma}_j$ is consistent for Γ_j for each given j, the estimator $\hat{\Gamma}$ is not consistent for Γ.

iii) Nonparametric Kernel Estimation

Definition (Spetral Density Matrix): Suppose $Z_t = X_t \varepsilon_t$ is a $K \times 1$ weakly stationary process with $E(Z_t) = 0$ and autocovariance function $\Gamma_j = Cov(Z_t, Z_{t-j})$ which is a $K \times K$ matrix. Suppose

$$\sum_{j=-\infty}^{+\infty} \left\| \Gamma_j \right\| < \infty$$

Then the Fourier transform of the autocovariance function Γ_j exists and is given by

$$h(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{+\infty} \Gamma_j \exp(-ij\omega), \quad \omega \in [-\pi, +\pi]$$

where $i = \sqrt{-1}$. The $K \times K$ matrix-valued function $h(\omega)$ is called the spectral density matrix of $\{Z_t\}$.

The inverse Fourier transform

$$\Gamma_j = \int_{-\pi}^{\pi} h(\omega) e^{ij\omega} d\omega$$

Both $h(\omega)$ and Γ_j are Fourier transform of each other, they contain the same amount of information on serial dependence of $\{X_t \varepsilon_t\}$. When $\omega = 0$,

$$\Gamma^* = 2\pi h(0) = \sum_{j=-\infty}^{+\infty} \hat{\Gamma}_j \qquad (8.65)$$

This is an important identity, the long-run variance is 2π times the spetral density matrix at frequency zero. This identity provides an approach to estimating Γ^*.
First, we consider

$$\hat{\Gamma} = \sum_{j=-q}^{q} \Gamma_j \qquad (8.66)$$

If q is fixed, then

$$\hat{\Gamma} \xrightarrow{P} \sum_{j=-q}^{q} \Gamma_j \neq 2\pi h(0)$$

We should let q grows to infinity as $T \to \infty$; that is $q = q(T) \to \infty$. The largest q we can use is $q = T - 1$:

$$\hat{\Gamma} = \sum_{j=-(T-1)}^{T-1} \hat{\Gamma}_j \qquad (8.67)$$

But this will not be consistent for Γ^*, because we essentially have T-1 unknown parameters using T data points. To ensure consistency of $\hat{\Gamma}$ to Γ^*, we should use

$$\hat{\Gamma} = \sum_{j=-q_T}^{q_T} \hat{\Gamma}_j \qquad (8.68)$$

where $q_T \to \infty, q_T/T \to 0$.

Although this estimator is consistent for Γ^*, it may not be positive-definite for all T. To ensure it is positive, we use

$$\hat{\Gamma} = \sum_{j=-q_T}^{q_T} k(j/q_T)\hat{\Gamma}_j \qquad (8.69)$$

Where $k(\bullet)$ is called a kernel function. When the Bartlett kernel is used:

$$k(x) = (1-|x|)1(|x|<1)$$

We obtain the so-called Newey-West (1987, 1994) Estimator for Γ^*. When the Quadratic-spectral kernel is used:

$$k(x) = \frac{3}{(\pi x)^2}\left\{\frac{\sin(\pi x)}{\pi x} - \cos(\pi x)\right\}$$

We obtain Andrews' (1991) Quadratic-spectral estimator for Γ^*. Not all kernel functions give positive semi-definite matrix $\hat{\Gamma}^*$, but many of them do.

3.Newey-West Heteroscedasticity Autocorrelation Consistent Covariance Estimator

This is am extension of white estimator. For the OLS estimator $\hat{\beta}$.

$$Avar\left(\hat{\beta}\middle|X\right)=\frac{1}{T}\left[\left(\frac{X'X}{T}\right)^{-1}\left[\frac{1}{T}X'\left(\sigma^2\Omega\right)X\right]\left(\frac{X'X}{T}\right)^{-1}\right]$$

$$Q_*=\frac{1}{T}X'\left(\sigma^2\Omega\right)X$$
$$=\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\sigma_{ts}X_tX_s'$$

Under Assumption A8.6.1~A8.6.4, $\sigma_{ts}\neq 0\quad(t\neq s)$. The natural counterpart for estimating Q_* would be $\hat{Q}_*=\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}e_te_sX_tX_s'$, in order that $\hat{Q}_*$ would be consistent and positive definite , we use Newey-West HAC Estimator:

$$S_*=\hat{Q}_*=S_0+\frac{1}{T}\sum_{j=1}^{L}\sum_{t=j+1}^{T}w(j)e_te_{t-j}'\left(X_tX_{t-j}'+X_{t-j}X_t'\right)$$

$w(j)=1-\frac{j}{L+1}$ is Bartlett Kernel

$$k\left(j/q_T\right)=\begin{cases}1-j/L+1, & 0\le j\le L+1\\ 0, & otherwise\end{cases}$$

8.7 Another Three Topics about Autocorrelation

8.7.1 Estimation with a Lagged Dependent Variable

1) The consequence of OLS estimation

When the model contains both autocorrelation and lagged dependent variables, the OLS estimator $\hat{\beta}$ is inconsistent. Suppose

$$Y_t=\beta_1+\beta_2X_t+\beta_3Y_{t-1}+\varepsilon_t$$

Where X_t is non-stochastic, $\varepsilon_t=\rho\varepsilon_{t-1}+u_t$, $|\rho|<1$, $u_t\sim iid\left(0,\sigma_u^2\right)$. We have

$\hat{\beta}_{LS}=\beta+\left(X'X\right)^{-1}X'\varepsilon$, where $X=\left(1\quad X_t\quad Y_{t-1}\right)$.

$$p\lim \hat{\beta}_{LS} = \beta + p\lim\left(\frac{X'X}{T}\right)^{-1} p\lim\left(\frac{X'\varepsilon}{T}\right)$$
$$= \beta + Q^{-1} p\lim\left(\frac{X'\varepsilon}{T}\right)$$

Where $Q = p\lim\left(\frac{X'X}{T}\right)$, $p\lim\left(\frac{X'\varepsilon}{T}\right) = p\lim\frac{1}{T}\begin{pmatrix}\sum \varepsilon_t \\ \sum X_t\varepsilon_t \\ \sum Y_{t-1}\varepsilon_t\end{pmatrix}$.

Using the theorem “consistency of sample mean”, we obtain,

$$p\lim\frac{1}{T}\sum \varepsilon_t = E(\varepsilon_t) = 0$$
$$p\lim\frac{1}{T}\sum X_t\varepsilon_t = E(X_t\varepsilon_t) = X_t E(\varepsilon_t) = 0$$
$$p\lim\frac{1}{T}\sum Y_{t-1}\varepsilon_t = E(Y_{t-1}\varepsilon_t)$$

and

$$E(Y_{t-1}\varepsilon_t)$$
$$= E\left[Y_{t-1}(\rho\varepsilon_{t-1} + u_t)\right]$$
$$= \rho E(Y_{t-1}\varepsilon_{t-1}) + E(Y_{t-1}u_t)$$
$$= \rho E\left[(\beta_1 + \beta_2 X_{t-1} + \beta_3 Y_{t-2} + \varepsilon_{t-1})\varepsilon_{t-1}\right] \ldots\ldots\ldots\ldots\ldots \left(E(Y_{t-1}u_t) = 0\right)$$
$$= \rho E(\beta_3 Y_{t-2}\varepsilon_{t-1}) + \rho E(\varepsilon_{t-1}^2)$$

Since $E(Y_{t-2}\varepsilon_{t-1}) = E(Y_{t-1}\varepsilon_t)$, $E(\varepsilon_{t-1}^2) = \frac{\sigma_u^2}{1-\rho^2}$.

$$E(Y_{t-1}\varepsilon_t) = \frac{\rho\sigma_u^2}{(1-\rho\beta_3)(1-\rho^2)} \neq 0$$

$$p\lim \hat{\beta} = \begin{pmatrix}\beta_1 \\ \beta_2 \\ \beta_3\end{pmatrix} + Q^{-1}\begin{pmatrix} 0 \\ 0 \\ \frac{\rho\sigma_u^2}{(1-\rho\beta_3)(1-\rho^2)}\end{pmatrix} \neq \begin{pmatrix}\beta_1 \\ \beta_2 \\ \beta_3\end{pmatrix} \qquad (8.70)$$

Therefore, Least Squares is inconsistent unless ρ equals zero. Since $\hat{\beta}_{LS}$ is inconsistent, the residuals on which an estimator of ρ would be based are likewise inconsistent. The results is, both $\hat{\rho}$ and d statistic are inconsistent, and the FGLS cannot proceed.

2) Hatanaka (1974, 1976) method: an efficient two-step estimation

We consider estimation of the model

$$Y_t = X_t'\beta + \gamma Y_{t-1} + \varepsilon_t$$
$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t$$

Step i): Find out the IV for Y_{t-1}

An appropriate instrument variable can be obtained by using the fitted values in the regression of Y_t on X_t and X_{t-1}, the residuals from the IV regression are then used to construct

$$\hat{\rho} = \frac{\sum_{t=3}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_{t-1}}{\sum_{t=3}^{T}\hat{\varepsilon}_{t-1}^2} \qquad (8.71)$$

Where $\hat{\varepsilon}_t = Y_t - X_t'\hat{\beta}_{IV} - \hat{\gamma}_{IV}Y_{t-1}$.

Step ii): Compute FGLS estimates

$$Y_t - \hat{\rho}Y_{t-1} = \left(X_{t-1} - \hat{\rho}X_{t-1}\right)'\hat{\beta}^* + \hat{\gamma}^*\left(Y_{t-1} - \hat{\rho}Y_{t-2}\right) + \hat{d}\hat{\varepsilon}_{t-1} + u_t \qquad (8.72)$$

The efficient estimator of ρ is

$$\hat{\hat{\rho}} = \hat{\rho} + \hat{d} \qquad (8.73)$$

Appropriate asymptotic standard errors for the estimators, including $\hat{\hat{\rho}}$, are obtained from the $S^2\left(X_*'X_*\right)^{-1}$ computed at the second step, where X_* is the expplaining variables data matrix in (8.72). Hatanaka shows that these estimators are asymptotically equivalent to MLE.

8.7.2 Common Factors

See Chap 12.10, Green 5th

8.7.3 Forecasting in the Presence of Autocorrelation

1) The best linear unbiased prediction formula of Y_{T+n}

Suppose in the GR model $Y = XB + \varepsilon$, $\varepsilon \sim \left(0,\sigma^2\Omega\right) = \left(0,V\right)$, the value of regressor at time $T+n$ is $X_{T+n} = (1, \mathrm{X}_{2,\mathrm{T+n}}, \cdots, \mathrm{X}_{\mathrm{K,T+n}})'$, while the actual value of Y at the same time is

$$Y_{T+\mathrm{n}} = \mathrm{X}'_{\mathrm{T+n}}\beta + \varepsilon_{T+n} \qquad (8.74)$$

Since the disturbance $\varepsilon_{T+\mathrm{n}}$ satisfies assumption conditions of the GR model, we have

$$\varepsilon_{T+n} \sim N\,(0, \sigma^2_{T+n}) \text{, and } E(\varepsilon_{T+n}\varepsilon) = \begin{pmatrix} \varepsilon_{T+n}\varepsilon_1 \\ \vdots \\ \varepsilon_{T+n}\varepsilon_T \end{pmatrix} = W \qquad (8.75)$$

Suppose P is the best linear unbiased estimate of Y_{T+n}, then P need to satisfiy

$$P = C'Y \qquad (8.76a)$$

$$EP = EC'Y = C'X\beta = X'_{T+n}\beta \qquad (8.76b)$$

$$E(P - Y_{T+n})^2 = Min \qquad (8.76c)$$

where C is a $T \times 1$ scalar vector. So solving P is equivaentl to solving C, which satisfies equation （8.76a） ~ （8.76c）. That is to say, minimize $E(P - Y_{T+n})^2$ under constrained condition $E(P - Y_{T+n}) = 0$.

From （8.76b）, we have, $(\mathrm{C}'\mathrm{X}\text{-}\mathrm{X}'_{T+n})\beta = 0$.

From （8.76a）, we have

$$\begin{aligned} \mathrm{p}\text{-}\mathrm{Y}_{T+n} &= (\mathrm{C}'\mathrm{X}\text{-}\mathrm{X}'_{T+n})\beta + (\mathrm{C}'\varepsilon\text{-}\varepsilon_{T+n}) \\ &= \mathrm{C}'\varepsilon\text{-}\varepsilon_{T+n} \end{aligned}$$

So

$$\begin{aligned} \mathrm{E}(\mathrm{P}\text{-}\mathrm{Y}_{T+n})^2 &= E\left[(\mathrm{P}\text{-}\mathrm{Y}_{T+n})(\mathrm{P}\text{-}\mathrm{Y}_{T+n})'\right] \\ &= E\left[(\mathrm{C}'\varepsilon\text{-}\varepsilon_{T+n})(\mathrm{C}'\varepsilon\text{-}\varepsilon_{T+n})'\right] \\ &= E(\mathrm{C}'\varepsilon\varepsilon'C + \varepsilon^2_{T+n}\text{-}2\mathrm{C}'\varepsilon\varepsilon_{T+n}) \\ &= C'VC + \sigma^2_{T+n}\text{-}2\mathrm{C}'W \end{aligned}$$

Using Lagrange multiplier vector $\lambda = (\lambda_1, \cdots, \lambda_{\mathrm{k}})'$ to construct object function

$$\Phi = C'VC + \sigma^2_{T+n} - 2C'W - 2(\mathrm{C}'\mathrm{X} - \mathrm{X}'_{\mathrm{T+n}})\lambda$$

Then C is the solution of equations as below:

$$\begin{cases} \dfrac{\partial \Phi}{\partial C}\Big|_{C=\hat{C},\lambda=\hat{\lambda}} = 2V\hat{C} - 2W - 2X\hat{\lambda} = 0 \\ \dfrac{\partial \Phi}{\partial \lambda}\Big|_{C=\hat{C},\lambda=\hat{\lambda}} = 2(\hat{C}'X - X'_{T+n}) = 0 \end{cases}$$

That is $\begin{cases} V\hat{C} - X\hat{\lambda} = W \\ X'\hat{C} = X_{T+n} \end{cases}$.

And we can obtain

$$\hat{\lambda} = (X'V^{-1}X)^{-1}(X_{T+n} - X'V^{-1}W)$$
$$\hat{C} = V^{-1}X\hat{\lambda} + V^{-1}W$$
$$= V^{-1}X[(X'V^{-1}X)^{-1}(X_{T+n} - X'V^{-1}W)] + V^{-1}W$$
$$= V^{-1}X[(X'V^{-1}X)^{-1}X_{T+n} - (X'V^{-1}X)^{-1}X'V^{-1}W] + V^{-1}W$$

So,

$$P = \hat{C}'Y$$
$$= X'_{T+n}(X'V^{-1}X)^{-1}X'V^{-1}Y + W'V^{-1}Y - W'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}Y$$
$$= X'_{T+n}\hat{\beta}_{GLS} + W'V^{-1}Y - W'V^{-1}X\hat{\beta}_{GLS}$$
$$= X'_{T+n}\hat{\beta}_{GLS} + W'V^{-1}(Y - X\hat{\beta}_{GLS}) \quad (8.77)$$
$$= X'_{T+n}\hat{\beta}_{GLS} + W'V^{-1}e$$

When $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$ and u_t is a white noise series, we have

$$\hat{Y}_{T+n} = X'_{T+n}\hat{\beta}_{GLS} + \rho^n e_T \quad (8.78)$$

It is easy to see, the BLUP of Y_{n+T} including a modified factor, which calculate from the information of autocorrelated disturbance structure V.

2) An Alternative Explanation

An alternative explanation of (8.78) will be given at following. Since heteroscedasticity doesn't affect forecasting alone in GR model, only the existence of autocorrelation will affect BLUP, we just consider the simple case of AR(1), and the same idea can be expanded to solute more complicate case. Suppose in GR model: $Y = X\beta + \varepsilon$, $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$, where u_t is a white noise series. From GLS estimation, we can get $Y_t = X'_t\hat{\beta}_{GLS} + e_t, t = 1, \cdots, T$, then the forecasting of Y_{T+1} is:

$$E(Y_{T+1} \mid F_T) = X_{T+1}'\hat{\beta}_{GLS} + E(\varepsilon_{T+1} \mid F_T)$$
$$= X_{T+1}'\hat{\beta}_{GLS} + \rho E(\varepsilon_T \mid F_T)$$
$$= X_{T+1}'\hat{\beta}_{GLS} + \rho e_T$$

Where F_T denotes all information from period 1 to T.

Similarly, the forecasting of Y_{T+2} is:

$$
\begin{aligned}
E\left(Y_{T+2} \mid F_T\right) &= X_{T+2}{}' \hat{\beta}_{GLS} + E\left(\varepsilon_{T+2} \mid F_T\right) \\
&= X_{T+2}{}' \hat{\beta}_{GLS} + \rho E\left(\varepsilon_{T+1} \mid F_T\right) \\
&= X_{T+2}{}' \hat{\beta}_{GLS} + \rho^2 E\left(\varepsilon_{T} \mid F_T\right) \\
&= X_{T+2}{}' \hat{\beta}_{GLS} + \rho^2 e_T
\end{aligned}
$$

It can be easy to show, the forecasting of Y_{T+n} is:

$$
E\left(Y_{T+n} \mid F_T\right) = X_{T+n}{}' \hat{\beta}_{GLS} + \rho^n e_T
$$

Apparently, the conclusion is the same to (8.78).