Contents

9 Consistent Covariance Matrix Estimation **3**
9.1 Heteroskedasticity 5
9.2 Serial Correlation of Finite Order q 6
9.3 Heteroskedasticity and Infinite Serial Correlation 7
9.4 Newey-West Automatic Lag Selection 8
9.5 The Relation with Spectral Densities 10
9.5.1 The Spectrum 10

Chapter 9

Consistent Covariance Matrix Estimation

The intention of this section is to provide a means by which valid statistical inference from estimation can be made when the regression errors are heteroskedastic and/or serially correlated. The problem could be setup more generally with endogenous regressors in a nonlinear environment. However, since we wish to focus on the consistent covariance estimation we keep the estimation environment very simple.

- In this chapter, we study cases where the ordinary least squares (OLS) yields consistent estimates of parameters.
- The presence of heteroskedasticity and/or serial correlation will not affect the consistency of the parameter estimates, although it will render the estimated asymptotic covariance matrix of the parameter estimates inconsistent.
- The objective then is to obtain a **consistent** estimate of the asymptotic covariance matrix of the parameter estimates which can be used with the consistent estimate of the parameters to conduct valid statistical inference.
- Consider the linear regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon} \tag{9.1}$$

such that

1. $E[\boldsymbol{\varepsilon}] = \mathbf{0}$
2. $\mathbf{X}^T\mathbf{X}/T$ converges to a finite, constant matrix
3. $\mathbf{X}^T\boldsymbol{\varepsilon}/T^{1/2}$ converges to zero
4. $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \Omega_0)$ (Notice that $\Omega_0 \neq \sigma_\varepsilon^2\mathbf{I}_T$)

- This is almost a classical linear regression model except for the non-spherical errors.

- The least-squares estimate of $\boldsymbol{\beta}$ is

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \\ &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{X}\boldsymbol{\beta}_0+\boldsymbol{\varepsilon}) \qquad (9.2)\\ &= \boldsymbol{\beta}_0+(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\varepsilon} \qquad (9.3)\end{aligned}$$

  Hence

$$\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0=(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\varepsilon} \qquad (9.4)$$

- The asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}$ is given by

$$\begin{aligned}\Lambda_0 &= var[T^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)] \\ &= var[T^{1/2}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\varepsilon}] \qquad (9.5)\\ &= (\mathbf{X}^T\mathbf{X}/T)^{-1}var[T^{-1/2}\mathbf{X}^T\boldsymbol{\varepsilon}](\mathbf{X}^T\mathbf{X}/T)^{-1} \qquad (9.6)\\ &= \mathbf{D}_0^{-1}\Sigma_0\mathbf{D}_0^{-1} \qquad (9.7)\end{aligned}$$

  where

$$\Sigma_0=E[\mathbf{X}^T\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\mathbf{X}/T]=E[\mathbf{X}^T\Omega_0\mathbf{X}/T] \qquad (9.8)$$

  since $E[T^{-1/2}\mathbf{X}^T\boldsymbol{\varepsilon}]=\mathbf{0}$ by assumption.

- If we are to conduct valid statistical inference then we need an estimate of Λ_0 which is consistent and positive semi-definite.

- Since the matrix $\mathbf{D}_0$ in (9.7) is observable, we need only to find a consistent estimate of Σ_0 in order to obtain a consistent estimate of Λ_0.

- It is important to be clear that we are dealing with consistency of the matrix Σ_0 and not Ω_0 because the latter has dimensions which are increasing faster than T (the dimension of this is $\frac{T(T+1)}{2}$).

- Before proceeding to discuss various structures of the matrix Σ_0, it is useful to write (9.8) in terms of individual observations for all T as

$$\begin{aligned}\Sigma_0 &= E[T^{-1}\sum_{t=1}^{T}\sum_{\tau=1}^{T}\mathbf{X}_t^T\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_\tau\mathbf{X}_\tau] \\ &= T^{-1}\sum_{t=1}^{T}E[\mathbf{X}_t^T\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t\mathbf{X}_t]+ \\ &\quad T^{-1}\sum_{\tau=1}^{T-1}\sum_{t=\tau+1}^{T}E[\mathbf{X}_t^T\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_{t-\tau}^T\mathbf{X}_{t-\tau}+\mathbf{X}_{t-\tau}^T\boldsymbol{\varepsilon}_{t-\tau}\boldsymbol{\varepsilon}_t^T\mathbf{X}_t] \qquad (9.9)\\ &= T^{-1}\sum_{t=1}^{T}var[\mathbf{X}_t^T\boldsymbol{\varepsilon}_t]+ \\ &\quad T^{-1}\sum_{\tau=1}^{T-1}\sum_{t=\tau+1}^{T}cov[\mathbf{X}_t^T\boldsymbol{\varepsilon}_t,\mathbf{X}_{t-\tau}^T\boldsymbol{\varepsilon}_{t-\tau}]+cov[\mathbf{X}_{t-\tau}^T\boldsymbol{\varepsilon}_{t-\tau},\mathbf{X}_t^T\boldsymbol{\varepsilon}_t] \qquad (9.10)\end{aligned}$$

- That is, Σ_0 is the **average of the variance** of $\mathbf{X}_t^T\boldsymbol{\varepsilon}_t$ plus a term that takes account of the average of the autocovariances between $\mathbf{X}_t^T\boldsymbol{\varepsilon}_t$ and $\mathbf{X}_{t-\tau}^T\boldsymbol{\varepsilon}_{t-\tau}$ for all t and τ

$$\Sigma_0 = \Pi + \Lambda + \Lambda^T$$

- The matrix Σ_0 is sometimes called the **long-run covariance matrix** (for reasons which will be clear in a moment).

- We now consider various structures for the matrix Ω_0.

9.1 Heteroskedasticity

$$cov[\mathbf{X}_t^T\boldsymbol{\varepsilon}_t, \mathbf{X}_{t-\tau}^T\boldsymbol{\varepsilon}_{t-\tau}] = cov[\mathbf{X}_{t-\tau}^T\boldsymbol{\varepsilon}_{t-\tau}, \mathbf{X}_t^T\boldsymbol{\varepsilon}_t] = 0$$

for all $t \neq \tau$, so that the matrix Ω_0 is diagonal.

- This is the case when the sequence $\{\mathbf{X}_t, \varepsilon_t\}$ is independently but not identically distributedand was considered by White.

- That is, we allow for heteroskedasticity in the regression errors. In this instance, we have

$$\Sigma_0 = T^{-1}\sum_{t=1}^{T} E[\mathbf{X}_t^T\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^T\mathbf{X}_t] = T^{-1}\sum_{t=1}^{T} E[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^T\mathbf{X}_t^T\mathbf{X}_t] \qquad (9.11)$$

- If we assume that the matrix $\mathbf{X}$ is **non-stochastic**, expression in (9.11) can be further simplified to

$$\Sigma_0 = T^{-1}\sum_{t=1}^{T} E[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^T]\mathbf{X}_t^T\mathbf{X}_t$$

- Its estimate is given by

$$\hat{\Sigma} = T^{-1}\sum_{t=1}^{T} \hat{\boldsymbol{\varepsilon}}_t\hat{\boldsymbol{\varepsilon}}_t^T\mathbf{X}_t^T\mathbf{X}_t$$

  where $\hat{\boldsymbol{\varepsilon}}_t = \mathbf{y}_t - \mathbf{X}_t\hat{\boldsymbol{\beta}}$ is the least-squares residuals.

- It is also consistent in the sense that

$$\hat{\Sigma} \xrightarrow{p} \Sigma_0.$$

  since $\hat{\boldsymbol{\beta}}$ is a consistent estimate of $\boldsymbol{\beta}_0$

9.2 Serial Correlation of Finite Order q

$$cov[\mathbf{X}_t^T \boldsymbol{\varepsilon}_t, \mathbf{X}_{t-\tau}^T \boldsymbol{\varepsilon}_{t-\tau}] = cov[\mathbf{X}_{t-\tau}^T \boldsymbol{\varepsilon}_{t-\tau}, \mathbf{X}_t^T \boldsymbol{\varepsilon}_t] = 0$$

for all $\tau \geq q$ (and $1 < q < \infty$), so that the matrix Ω_0 is band-diagonal.

- This is the case when the sequence has finite-order (i.e. q) MA representation.
- The matrix Σ_0 becomes

$$\begin{aligned} \Sigma_0 &= T^{-1} \sum_{t=1}^{T} E[\mathbf{X}_t^T \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^T \mathbf{X}_t] + \\ & \quad T^{-1} \sum_{\tau=1}^{q-1} \sum_{t=\tau+1}^{T} E[\mathbf{X}_t^T \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_{t-\tau}^T \mathbf{X}_{t-\tau}] + E[\mathbf{X}_{t-\tau}^T \boldsymbol{\varepsilon}_{t-\tau} \boldsymbol{\varepsilon}_t^T \mathbf{X}_t] \end{aligned} \tag{9.12}$$

- As an example suppose that $\mathbf{X}_t$ is non-stochastic and ε_t is an $MA(1)$ process such that

$$\varepsilon_t = \theta \varepsilon_{t-1} + u_t,$$

where $u_t \sim$ NID$(0, \sigma_u^2)$.

- Then for $\tau \geq 2$ we have

$$E[\mathbf{X}_t^T \varepsilon_{t-\tau}] = \mathbf{X}_t^T E[\varepsilon_{t-\tau}] = 0$$

and hence

$$\begin{aligned} \Sigma_0 &= T^{-1} \sum_{t=1}^{T} E[\mathbf{X}_t^T \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^T \mathbf{X}_t] + \\ & \quad T^{-1} \sum_{t=2}^{T} E[\mathbf{X}_t^T \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_{t-1}^T \mathbf{X}_{t-1}] + E[\mathbf{X}_{t-1}^T \boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}_t^T \mathbf{X}_t] \end{aligned}$$

-

$$\Sigma_0 = \begin{array}{ccccc} \sigma_{11}^2 & \sigma_{12}^2 & 0 & 0 & 0 \\ \sigma_{12}^2 & \ddots & \sigma_{12}^2 & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \sigma_{12}^2 & \sigma_{11}^2 & 0 \\ 0 & 0 & 0 & \sigma_{12}^2 & \sigma_{11}^2 \end{array}$$

- For the case of q order autocorrelation, a consistent estimate of the matrix Σ_0 in (9.12) is given by

$$\hat{\Sigma} = T^{-1} \sum_{t=1}^{T} \mathbf{X}_t^T \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_t^T \mathbf{X}_t + \tag{9.13}$$

$$T^{-1} \sum_{\tau=1}^{q-1} \sum_{t=\tau+1}^{T} (\mathbf{X}_t^T \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_{t-\tau}^T \mathbf{X}_{t-\tau} + \mathbf{X}_{t-\tau}^T \hat{\boldsymbol{\varepsilon}}_{t-\tau} \hat{\boldsymbol{\varepsilon}}_t^T \mathbf{X}_t) \tag{9.14}$$

where $\hat{\boldsymbol{\varepsilon}}_t = \mathbf{y}_t - \mathbf{X}_t \hat{\boldsymbol{\beta}}$ is the least-squares residuals.

- Since the covariance terms in the second component in (9.13) may be larger (in absolute term) than the variance term in the first component (9.13), the estimated matrix $\hat{\Sigma}$ need not be positive semi-definite.

- This poses a serious problem because when the matrix $\hat{\Sigma}$ is not positive semi-definite so that standard errors for some test statistics for some linear combination of the estimated parameter $\hat{\boldsymbol{\beta}}$ would be negative !

- A simple estimator which is always positive semi-definite and consistent is given by the Newey-West estimator:

$$\begin{aligned}\hat{\Sigma} &= T^{-1}\sum_{t=1}^{T}\mathbf{X}_t^T\hat{\boldsymbol{\varepsilon}}_t\hat{\boldsymbol{\varepsilon}}_t^T\mathbf{X}_t + \\ & T^{-1}\sum_{\tau=1}^{q-1}\sum_{t=\tau+1}^{T}\omega(\tau,q)[\mathbf{X}_t^T\hat{\boldsymbol{\varepsilon}}_t\hat{\boldsymbol{\varepsilon}}_{t-\tau}^T\mathbf{X}_{t-\tau} + \mathbf{X}_{t-\tau}^T\hat{\boldsymbol{\varepsilon}}_{t-\tau}\hat{\boldsymbol{\varepsilon}}_t^T X_t]\end{aligned}$$

  where $\omega(\tau,q) = 1-(\tau/(q+1))$ is a weighting scheme that declines as τ increases.

- Intuitively the positive semi-definiteness of the estimator is ensured by down weighting the higher-order variance of $T^{-1/2}\mathbf{X}^T\boldsymbol{\varepsilon}$ and the consistency ensured by reducing the down weighting as the sample size grows.

9.3 Heteroskedasticity and Infinite Serial Correlation

-
$$\Sigma_0 = T^{-1}\sum_{t=1}^{T}E[\mathbf{X}_t^T\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^T\mathbf{X}_t] + \tag{9.15}$$
$$T^{-1}\sum_{\tau=1}^{T-1}\sum_{t=\tau+1}^{T}E[\mathbf{X}_t^T\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_{t-\tau}^T\mathbf{X}_{t-\tau}] + E[\mathbf{X}_{t-\tau}^T\boldsymbol{\varepsilon}_{t-\tau}\boldsymbol{\varepsilon}_t^T\mathbf{X}_t] \tag{9.16}$$

- This is the most general case.

- We need to impose the restriction that as $\tau \to \infty$, the covariance between $\mathbf{X}_t^T\boldsymbol{\varepsilon}_t$ and $\mathbf{X}_{t-\tau}^T\boldsymbol{\varepsilon}_{t-\tau}$ goes to zero.

- This amounts to imposing the restriction that the sequence $\{\mathbf{X}_t, \varepsilon_t\}$ is a *mixing* sequence (For a definition of mixing sequence, see, e.g., White (1984)).

- For such a sequence the matrix Σ_0 can be approximated by using a finite-lag m which increases with T

$$\begin{aligned}\Sigma_0^* &= T^{-1}\sum_{t=1}^{T}E[\mathbf{X}_t^T\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^T\mathbf{X}_t] + \\ & T^{-1}\sum_{\tau=1}^{m}\sum_{t=\tau+1}^{T}E[\mathbf{X}_t^T\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_{t-\tau}^T\mathbf{X}_{t-\tau}] + E[\mathbf{X}_{t-\tau}^T\boldsymbol{\varepsilon}_{t-\tau}\boldsymbol{\varepsilon}_t^T\mathbf{X}_t]\end{aligned}$$

- Note that if m is kept fixed, as T grows, the number of omitted things would grow such that their sum may no longer be negligible.

- This suggests that m will have to grow with T for the term in Σ_0 which are omitted by Σ_0^* to remain negligible.

- Nevertheless, if the above approximation is to yield a consistent estimator, m must grow at a slower rate than T.

- One estimator of Σ_0 proposed by Newey and West (1987) is

$$\hat{\Sigma}^* = T^{-1}\sum_{t=1}^{T}\mathbf{X}_t^T\hat{\boldsymbol{\varepsilon}}_t\hat{\boldsymbol{\varepsilon}}_t^T\mathbf{X}_t + \tag{9.17}$$

$$T^{-1}\sum_{j=1}^{m}\sum_{t=j+1}^{T}\omega(j,m)[\mathbf{X}_t^T\hat{\boldsymbol{\varepsilon}}_t\hat{\boldsymbol{\varepsilon}}_{t-\tau}^T\mathbf{X}_{t-\tau} + \mathbf{X}_{t-\tau}^T\hat{\boldsymbol{\varepsilon}}_{t-\tau}\hat{\boldsymbol{\varepsilon}}_t^T\mathbf{X}_t] \tag{9.18}$$

  where $\omega(j,m) = 1 - \frac{j}{m+1}$ is the weighting scheme (called the *kernel, in this case the Bartlett kernel*). The positive semi-definiteness of the estimator is ensured using an argument similar to case (2).

- Consistency requires:

1. $m \to \infty$ as $T \to \infty$

2. $\frac{m}{T^{\frac{1}{2}}} \to 0$

3. For the Bartlett kernel, the rate is $m = o(T^{1/3})$.

9.4 Newey-West Automatic Lag Selection

- Newey and West (1994, **Review of Economic Studies, 61,** 631-653) suggested a way to choose m. This nonparametric method is, for a given kernel for weighting the autocovariances, asymptotically equivalent to one that is optimal under mean-square error loss function.

- The simulations suggest that it does as well as anything else available, although there is some size distortion.

- Let $f_t = \mathbf{X}_t^T\hat{\boldsymbol{\varepsilon}}_t$ be the $(k\times 1)$ vector we wish to estimate the variance covariance matrix Σ.

- We will assume there is a constant in the model: $X_t = (1, \tilde{x}_t^T)$ and that we wish to estimate the long-run covariance matrix (Σ) of $\bar{f}_t$ using the Newey-West estimator:

$$\Sigma = \hat{\Lambda}_0 + \sum_{j=1}^{m}(1 - \frac{j}{m+1})(\hat{\Lambda}_j + \hat{\Lambda}_j^T)$$

where

$$\hat{\Lambda}_j = \frac{1}{T-1} \sum_{t=j+2}^{T} f f_{t-j}^T \quad j = 0, 1, \ldots, m \tag{9.19}$$

- Now the trick is to choose m optimally in a data dependent way.
- The form of the decision is:

$$m = \gamma \times T^{\frac{1}{3}} \tag{9.20}$$

where γ is estimated from the data.

- We would like to optimally weight the f_t.
- Define a weight vector $(k \times 1)$, w.
- The idea is to choose w optimally to minimize the MSE of the estimated long-run covariance. The optimal choice of w will depend on the data, implying a different w and hence a different γ from problem to problem.
- Since there is really no way of guiding this choice without actually knowing the data generating process, Newey and West suggest w as:

$$w = (0, 1, \ldots 1)^T \tag{9.21}$$

where the 0 entry corresponds to the constant.

- This choice makes sense within the context of IV estimation where the first instrument is a constant. In more general cases, and without further information equal weighting seems appropriate.
- We need to do the following calculations. We can remove much of the serial correlation in f_t by filtering and recoloring at the end of the procedure (Andrews and Monahan, 1993 **Econometrica** found this to have excellent finite sample properties in terms of coverage probabilities).
- Pre-whiten f_t using a $VAR(1)$ to obtain the $(k \times k)$ matrix

$$\hat{A} = \sum f_t f_{t-1}^T \left(\sum_{t=2}^{T} f_{t-1} f_{t-1}^T \right)^{-1} \tag{9.22}$$

Obtain the "residuals"

$$f_t^+ = f_t - \hat{A} f_{t-1}. \tag{9.23}$$

- Denote n as the *lag selection parameter.*

- In order to gauge its effect on the lag length parameter m, Newey-West suggest one should do some sensitivity analysis. The initial setting for n that is given is:

$$n = 4\left(\frac{T}{100}\right)^{\frac{2}{9}} \tag{9.24}$$

- Using this n, calculate the autocovariances:

$$\hat{\sigma}_j = \frac{1}{T-1}\sum_{t=j+2}^{T}\left\{\left(w^T f_t^+\right)\times\left(w^T f_{t-j}^+\right)\right\} \quad j = 0, \ldots, n \tag{9.25}$$

From these estimated autocovariances, we need

$$\hat{s}^{(1)} = 2\sum_{j=1}^{n} j\hat{\sigma}_j \qquad \hat{s}^{(0)} = \hat{\sigma}_0 + 2\sum_{j=1}^{n}\hat{\sigma}_j \tag{9.26}$$

Finally, the estimated γ is obtained as:

$$\hat{\gamma} = 1.1447\left(\left[\frac{\hat{s}^{(1)}}{\hat{s}^{(0)}}\right]^2\right)^{\frac{1}{3}} \tag{9.27}$$

9.5 The Relation with Spectral Densities

There are many good references to spectral analysis available. For example Jenkins and Watts, Fuller, the third chapter of Harvey and the sixth chapter of Chow. Two particularly good introductory references are Chatfield and Granger and Newbold. Our intention in this section is to briefly introduce the subject and show some advantages are from moving from the time domain to the frequency domain. In the last 5 years the methods of Fourier analysis have reappeared in the study of economic time series. This is especially true in calculating so-called **long-run covariances matrices used in** HAC estimation.

Spectral methods allow one to get at the **cyclical** (and irregular and aperiodic) features in the data that the time domain representations fail to bring out.

9.5.1 The Spectrum

The fundamental relation in the frequency domain is the (power) spectrum. The spectrum for an indeterministic stationary process y_t is a continuous **even** (**real**) function:

$$f(\omega) \;=\; (2\pi)^{-1}\left[\gamma(0) + 2\sum_{\tau=1}^{\infty}\gamma(\tau)\cos\omega\tau\right] \tag{9.28}$$

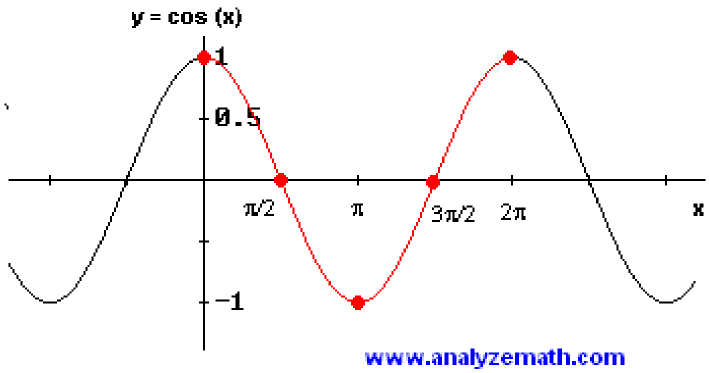

Figure 9.1:

where ω is the frequency in radians (a radian is equal to $180/\pi$ degrees; about 59 degrees) which takes on values over the range $[-\pi, \pi]$ and $\gamma(\tau)$ are the autocovariances of y_t.

The $cos(\omega) = \cos(-\omega)$ is an even function

Since f is even (and nonnegative), we need only evaluate it over the range $[0, \pi]$. If y_t is a white noise process, then we know $\gamma(\tau) = 0$ for all $\tau \neq 0$ and $\gamma(0) \;=\; \sigma^2$ so:

$$f(\omega) \;=\; \sigma^2/2\pi. \tag{9.29}$$

where the variance of y_t is σ^2 for all t. Thus, the spectrum for white noise is flat and trivially periodic.The spectrum for white noise is a constant multiple of the variance and analogous to white light. We might say that all frequencies contribute equally to the variance.

Integrating (9.28) over $[-\pi, \pi]$ we see that **the area under the spectrum is equal to the variance**:

$$\int_{-\pi}^{\pi} f(\omega) d\omega \;=\; \gamma(0) \tag{9.30}$$

What is important about (9.30) is that the variance and thhe autocovariances can be decomposed as a process in terms of frequency. We may also write (9.28) in terms of the autocorrelations $\rho(\tau)$ by dividing by $\gamma(0)$. Sometimes this standardized equation is referred to as the spectral density. (However Fuller simply calls (9.28) the spectral density associated with the covariance). We follow Harvey by referring to (9.28) as the spectrum or spectral density.

Now it is often easier to work with the complex representations of spectrum leading from Fourier transforms is the identity:

$$e^{i\omega} = \cos\omega + i\sin\omega \tag{9.31}$$

where $i = \sqrt{-1}$. The exponential function is an extremely attractive representation and we can rewrite (9.28) as

$$f(\omega) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\gamma(j)e^{-ih\omega}. \tag{9.32}$$

- To relate the estimation of the matrix Σ_0 with the **spectral density** estimator, we rewrite the matrix Σ_0 in (9.15) as

$$\Sigma_0 = T^{-1}\sum_{\tau=1}^{T}\sum_{t=1}^{T}E[\mathbf{X}_\tau^T\boldsymbol{\varepsilon}_\tau\boldsymbol{\varepsilon}_t^T\mathbf{X}_t]$$

- Using the change of variable technique, the matrix Σ_0 can be rewritten as

$$\Sigma_0 = \sum_{j=-T+1}^{T-1}\boldsymbol{\gamma}(j)$$

where

$$\boldsymbol{\gamma}(j) = \begin{cases} T^{-1}\sum_{t=j+1}^{T}E[\mathbf{X}_t^T\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_{t-j}^T\mathbf{X}_{t-j}] & for\ j \geq 0 \\ T^{-1}\sum_{t=-j+1}^{T}E[\mathbf{X}_{t+j}^T\boldsymbol{\varepsilon}_{t+j}\boldsymbol{\varepsilon}_t^T\mathbf{X}_t] & for\ j < 0 \end{cases}$$

- When the random variable $T^{-1/2}\mathbf{X}^T\boldsymbol{\varepsilon}$ is covariance stationary, it has a spectral density matrix given by (9.32) where

$$\boldsymbol{\gamma}(j) = E[\mathbf{X}_t^T\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_{t-j}^T\mathbf{X}_{t-j}]$$

- It is easy to see that the following is true

$$\lim_{T\to\infty}\Sigma_0 = 2\pi\mathbf{f}(0)$$

- This fact motivates the use of spectral density estimators to estimate the matrix Σ_0.

- Moreover, given covariance stationarity and known $\boldsymbol{\beta}_0$, the estimate proposed by Newey and West (1987) corresponds to the spectral density estimator evaluated at $\omega = 0$ and using the tent window (or usually called the Bartlett kernel).

- Since we associate the low frequencies (i.e. $\omega = 0$) with the long run covariance matrix, it is now obvious why the matrix Σ_0 is sometimes called the **long-run covariance matrix**.

- See the *hac.do* program for illustration