CHAPTER 6 LINEAR REGRESSION UNDER CONDITIONAL HETEROSKEDASTICITY AND AUTOCORRELATION

Key words: Heteroskedasticity and Autocorrelation (HAC) consistent variance-covariance matrix, Kernel function, Long-run variance-covariance matrix, Newey-West estimator, Nonparametric estimation, Spectral density matrix.

Abstract: When the regression disturbance $\{\varepsilon_t\}$ displays serial correlation, the asymptotic results in Chapter 5 are no longer applicable, because the asymptotic variance of the OLS estimator will depend on serial correlation in $\{X_t\varepsilon_t\}$. In this chapter, we introduce a method to estimate the asymptotic variance of the OLS estimator in the presence of heteroskedasticity and autocorrelation, and then develop test procedures based on it. Some empirical applications are considered.

6.1 Motivation

In Chapter 5, we assumed that $\{X_t\varepsilon_t\}$ is an MDS. In many economic applications, there may exist serial correlation in the regression error $\{\varepsilon_t\}$. As a consequence, $\{X_t\varepsilon_t\}$ is generally no longer an MDS. We now provide a few examples where $\{\varepsilon_t\}$ is serially correlated.

Example 1 [Testing a zero population mean]: Suppose the daily stock return $\{Y_t\}$ is a stationary ergodic process with $E(Y_t) = \mu$. We are interested in testing the null hypothesis

$$\mathbf{H}_0 : \mu = 0$$

versus the alternative hypothesis

$$\mathbb{H}_A : \mu \neq 0.$$

A test for $\mathbf{H}_0$ can be based on the sample mean

$$\bar{Y}_n = n^{-1}\sum_{t=1}^{n} Y_t.$$

By a suitable CLT (White (1999)), the sampling distribution of the sample mean $\bar{Y}_n$ scaled by $\sqrt{n}$

$$\sqrt{n}\bar{Y}_n \to^d N(0, V),$$

where the asymptotic variance of the sample mean

$$V \equiv avar\left(\sqrt{n}\bar{Y}_n\right).$$

Because

$$\begin{aligned}\text{var}(\sqrt{n}\bar{Y}_n) &= n^{-1}\sum_{t=1}^{n}\text{var}\left(Y_t\right) \\ &\quad +2n^{-1}\sum_{t=2}^{n-1}\sum_{j=1}^{t-1}\text{cov}(Y_t, Y_{t-j}),\end{aligned}$$

serial correlation in $\{Y_t\}$ is expected to affect the asymptotic variance of $\sqrt{n}\bar{Y}_n$. Thus, unlike in Chapter 5, avar$(\sqrt{n}\bar{Y}_n)$ is no longer equal to var(Y_t).

Suppose there exists a variance-covariance estimator $\hat{V}$ such that $\hat{V} \to^p V$. Then, by the Slutsky theorem, we can construct a test statistic which is asymptotically N(0,1) under $\mathbf{H}_0$:

$$\frac{\sqrt{n}\bar{Y}_n}{\sqrt{\hat{V}}} \to^d N(0,1).$$

Example 2 [Unbiasedness Hypothesis]: Consider the following linear regression model

$$S_{t+\tau} = \alpha + \beta F_t(\tau) + \varepsilon_{t+\tau},$$

where $S_{t+\tau}$ is the spot foreign exchange rate at time $t+\tau$, $F_t(\tau)$ is the forward exchange rate (with maturity $\tau > 0$) at time t, and the disturbance $\varepsilon_{t+\tau}$ is not observable. Forward currency contracts are agreements to exchange, in the future, fixed amounts of two currencies at prices set today. No money changes hand over until the contract expires or is offset.

It has been a longstanding controversy on whether the current forward rate $F_t(\tau)$, as opposed to the current spot rate S_t, is a better predictor of the future spot rate $S_{t+\tau}$. The unbiasedness hypothesis states that the forward exchange rate (with maturity τ) at time t is the optimal predictor for the spot exchange rate at time $t+\tau$, namely,

$$E(S_{t+\tau}|I_t) = F_t(\tau) \text{ a.s.},$$

where I_t is the information set available at time t. This implies

$$\mathbf{H}_0 : \alpha = 0, \beta = 1,$$

and

$$E(\varepsilon_{t+\tau}|I_t) = 0 \text{ a.s., } t = 1, 2, ...$$

However, with $\tau > 1$, we generally do not have $E(\varepsilon_{t+j}|I_t) = 0$ a.s. for $1 \le j \le \tau - 1$. Consequently, there exists serial correlation in $\{\varepsilon_t\}$ up to $\tau - 1$ lags under H_0.

Example 3 [Long Horizon Return Predictability]: There has been much interest in regressions of asset returns, measured over various horizons, on various forecasting variables. The latter include ratios of price to dividends or earnings various interest rate measures such as the yield spread between long and short term rates, and the quality yield spread between low and high-grade corporate bonds, ad the short term interest rate.

Consider a regression

$$Y_{t+h,h} = \beta_0 + \beta_1 r_t + \beta_2 (d_t - p_t) + \varepsilon_{t+h,h}$$

where $Y_{t+h,h}$ is the cumulative return over the holding period from time t to time $t+h$, namely,

$$Y_{t+h,h} = \sum_{j=1}^{h} R_{t+j},$$

where R_{t+j} is an asset return in period $t+j$, r_t is the short term interest rate in time t, and $d_t - p_t$ is the log dividend-price ratio, which is expected to be a good proxy for market expectations of future stock return, because $d_t - p_t$ is equal to the expectation of the sum of all discounted future returns and dividend growth rates. In the empirical finance, there has been an interest in investigating how the predictability of asset returns by various forecasting variables depends on time horizon h. For example, it is expected that $d_t - p_t$ is a better proxy for expectations of long horizon returns than for expectations of short horizon returns. When monthly data is used and $h > 1$, there exists an overlapping for observations on $Y_{t+h,h}$. As a result, the regression disturbance $\varepsilon_{t+h,h}$ is expected to display serial correlation up to lag order $h - 1$.

Example 4 [Relationship between GDP and Money Supply]: Consider the linear macroeconomic regression model

$$Y_t = \alpha + \beta M_t + \varepsilon_t,$$

where Y_t is GDP at time t, M_t is the money supply at time t, and ε_t is an unobservable disturbance such that $E(\varepsilon_t|M_t) = 0$ but there may exist strong serial correlation of unknown form in $\{\varepsilon_t\}$.

Question: What happens to the OLS estimator $\hat{\beta}$ if the disturbance $\{\varepsilon_t\}$ displays conditional heteroskedasticity (i.e., $E(\varepsilon_t^2|X_t) = \sigma^2$ a.s. fails) *and/or* autocorrelation (i.e., $\text{cov}(\varepsilon_t, \varepsilon_{t-j}) \neq 0$ for some $j > 0$)? In particular,

- Is the OLS estimator $\hat{\beta}$ consistent for β^o?
- Is $\hat{\beta}$ asymptotically most efficient?
- Is $\hat{\beta}$, after properly scaled, asymptotically normal?
- Are the t-test and F-test statistics are applicable for large sample inference?

6.2 Assumptions

We now state the set of assumptions which allow for serial correlation and conditional heteroskedasticity of unknown form.

Assumption 6.1 [Ergodic Stationarity]: $\{Y_t, X_t'\}'$ is a stationary ergodic process.

Assumption 6.2 [Linearity]

$$Y_t = X_t'\beta^o + \varepsilon_t,$$

where β^o is a $K \times 1$ unknown parameter and ε_t is the unobservable disturbance.

Assumption 6.3 [Correct Model Specification]: $E(\varepsilon_t|X_t) = 0$ a.s.

Assumption 6.4 [Nonsingularity]: The $K \times K$ matrix

$$Q = E(X_t X_t')$$

is finite and nonsingular.

Assumption 6.5 [Long-run Variance]: (i) For $j = 0, \pm 1, ...$, put the $K \times K$ matrix

$$\begin{aligned} \Gamma(j) &= \text{cov}(X_t\varepsilon_t, X_{t-j}\varepsilon_{t-j}) \\ &= E[X_t\varepsilon_t\varepsilon_{t-j}X_{t-j}']. \end{aligned}$$

Then $\Sigma_{j=-\infty}^{\infty}||\Gamma(j)|| < \infty$, where $||A|| = \Sigma_{i=1}^{K}\Sigma_{j=1}^{K}|A_{(i,j)}|$ for any $K \times K$ matrix, and the long-run variance-covariance matrix

$$V = \sum_{j=-\infty}^{\infty} \Gamma(j)$$

is p.d.

(ii) The conditional expectation

$$E(X_t\varepsilon_t|X_{t-j}\varepsilon_{t-j}, X_{t-j-1}\varepsilon_{t-j-1}, ...) \to^{q.m.} 0 \text{ as } j \to \infty;$$

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

$$\begin{aligned} r_j &= E(X_t\varepsilon_t|X_{t-j}\varepsilon_{t-j}, X_{t-j-1}\varepsilon_{t-j-1}, ...) \\ &\quad -E(X_t\varepsilon_t|X_{t-j-1}\varepsilon_{t-j-1}, X_{t-j-2}\varepsilon_{t-j-2}, ...). \end{aligned}$$

Remarks:

Assumptions 6.1–6.4 have been assumed in Chapter 5 but Assumption 6.5 is new. Assumption 6.5(i) allows for both conditional heteroskedasticity and autocorrelation of unknown form in $\{\varepsilon_t\}$, and no normality assumption is imposed on $\{\varepsilon_t\}$.

We do not assume that $\{X_t\varepsilon_t\}$ is an MDS, although $E(X_t\varepsilon_t) = 0$ as implied by $E(\varepsilon_t|X_t) = 0$ a.s. Note that $E(\varepsilon_t|X_t) = 0$ a.s. does not necessarily imply that $\{X_t\varepsilon_t\}$ is MDS in a time series context. See the aforementioned examples for which $\{X_t\varepsilon_t\}$ is not MDS.

Assumptions 6.5(ii, iii) imply that the serial dependence of $X_t\varepsilon_t$ on its past history in term of mean and variance respectively vanishes to zero as the lag order $j \to \infty$. Intuitively, Assumption 6.5(iii) may be viewed as the net effect of $X_{t-j}\varepsilon_{t-i}$ on the conditional mean of $X_t\varepsilon_t$. It assumes that $E(r_j'r_j) \to 0$ as $j \to \infty$.

6.3 Long-run Variance Estimation

Question: Why are we interested in V?

Recall that for the OLS estimator $\hat{\beta}$, we have

$$\sqrt{n}(\hat{\beta} - \beta^o) = \hat{Q}^{-1}n^{-1/2}\sum_{t=1}^{n} X_t\varepsilon_t.$$

Suppose the CLT holds for $\{X_t\varepsilon_t\}$. That is, suppose

$$n^{-1/2}\sum_{t=1}^{n} X_t\varepsilon_t \xrightarrow{d} N(0, V),$$

where V is an asymptotic variance, namely

$$\begin{aligned} V &\equiv \text{avar}\left(n^{-1/2}\sum_{t=1}^{n} X_t\varepsilon_t\right) \\ &= \lim_{n\to\infty} \text{var}\left(n^{-1/2}\sum_{t=1}^{n} X_t\varepsilon_t\right). \end{aligned}$$

Then, by the Slutsky theorem, we have

$$\sqrt{n}(\hat{\beta} - \beta^o) \xrightarrow{d} N(0, Q^{-1}VQ^{-1})$$

under suitable regularity conditions.

Put

$$g_t = X_t \varepsilon_t.$$

Note that $E(g_t) = 0$ given $E(\varepsilon_t|X_t) = 0$ and the law of iterated expectations. Because $\{g_t\}$ is not an MDS, it may be serially correlated. Thus, the autocovariance function $\Gamma(j) = \text{cov}(g_t, g_{t-j})$ may not be zero at least for some lag order $j > 0$.

Now we consider the variance

$$\begin{aligned}
\text{var}\left(n^{-1/2}\sum_{t=1}^{n} X_t\varepsilon_t\right) &= \text{var}\left(n^{-1/2}\sum_{t=1}^{n} g_t\right) \\
&= E\left[\left(n^{-1/2}\sum_{t=1}^{n} g_t\right)\left(n^{-1/2}\sum_{s=1}^{n} g_s\right)'\right] \\
&= n^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n} E(g_t g_s') \\
&= n^{-1}\sum_{t=1}^{n} E(g_t g_t') \\
&\quad + n^{-1}\sum_{t=2}^{n}\sum_{s=1}^{t-1} E(g_t g_s') \\
&\quad + n^{-1}\sum_{t=1}^{n-1}\sum_{s=t+1}^{n} E(g_t g_s') \\
&= n^{-1}\sum_{t=1}^{n} E(g_t g_t') \\
&\quad + \sum_{j=1}^{n-1} n^{-1}\sum_{t=j+1}^{n} E(g_t g_{t-j}') \\
&\quad + \sum_{j=-(n-1)}^{-1} n^{-1}\sum_{t=1}^{n+j} E(g_t g_{t-j}') \\
&= \sum_{j=-(n-1)}^{n-1} (1 - |j|/n)\Gamma(j) \\
&\to \sum_{j=-\infty}^{\infty} \Gamma(j) \text{ as } n \to \infty
\end{aligned}$$

by dominated convergence. Therefore, we have $V = \sum_{j=-\infty}^{\infty} \Gamma(j)$.

In contrast, when $\{g_t\}$ is MDS, we have

$$\begin{aligned} V &\equiv \text{avar}\left(n^{-1/2}\sum_{t=1}^{n} X_t\varepsilon_t\right) \\ &= E(g_t g_t') \\ &= E(X_t X_t' \varepsilon_t^2) \\ &= \Gamma(0) \end{aligned}$$

when $\{g_t\}$ is MDS.

When $\text{cov}(g_t, g_{t-j})$ is p.s.d. for all $j > 0$, the difference $\Sigma_{j=-\infty}^{\infty}\Gamma(j) - \Gamma(0)$ is a p.s.d matrix. Intuitively, when $\Gamma(j)$ is p.s.d., a large deviation of g_t from its mean will tend to be followed by another large deviation. As a result, $V - \Gamma(0)$ is p.s.d.

To explore the link between the long-run variance V and the spectral density matrix of $\{X_t\varepsilon_t\}$, which is crucial for consistent estimation of V, we now extend the concept of the spectral density of a univariate time series to a multivariate time series context.

Definition [Spectral Density Matrix] *Suppose $\{g_t = X_t\varepsilon_t\}$ is a $K\times 1$ weakly stationary process with $E(g_t) = 0$ and autocovariance function $\Gamma(j) \equiv cov(g_t, g_{t-j}) = E(g_t g_{t-j}')$, which is a $K \times K$ matrix. Suppose*

$$\sum_{j=-\infty}^{\infty} ||\Gamma(j)|| < \infty.$$

Then the Fourier transform of the autocovariance function $\Gamma(j)$ exists and is given by

$$H(\omega) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty} \Gamma(j)\exp(-ij\omega), \qquad \omega \in [-\pi, \pi],$$

where $i = \sqrt{-1}$. The $K \times K$ matrix-valued function $H(\omega)$ is called the spectral density matrix of the weakly stationary time series vector-valued process $\{g_t\}$.

Remark:

The inverse Fourier transform of the spectral density matrix is

$$\Gamma(j) = \int_{-\pi}^{\pi} H(\omega)e^{ij\omega}d\omega.$$

Both $H(\omega)$ and $\Gamma(j)$ are Fourier transforms of each other. They contain the same amount of information on serial dependence of the process $\{g_t = X_t\varepsilon_t\}$. The spectral distribution function $H(\omega)$ is useful to identify business cycles (see Sargent 1987, *Dynamic Marcoeconomics*, 2nd Edition). For example, if g_t is the GDP growth rate at time t, then $H(\omega)$ can be used to identify business cycle of the economy.

When $\omega = 0$, then the long-run variance-covariance matrix

$$V = 2\pi H(0) = \sum_{j=-\infty}^{\infty} \Gamma(j).$$

That is, the long-run variance V is 2π times the spectral density matrix of the time series process $\{g_t\}$ at frequency zero. As will be seen below, this link provides a basis for consistent nonparametric estimation of V.

Question: What are the elements of the $K \times K$ matrix $\Gamma(j)$?

Recall that $g_t = (g_{0t,}g_{1t}, ..., g_{kt})'$, where $g_{lt} = X_{lt}\varepsilon_t$ for $0 \le l \le k$. Then the $(l+1, m+1)$-th element of $\Gamma(j)$ is

$$\begin{aligned} [\Gamma(j)]_{(l+1,m+1)} &= \Gamma_{lm}(j) \\ &= \mathrm{cov}[g_{lt}, g_{m(t-j)}] \\ &= \mathrm{cov}[X_{lt}\varepsilon_t, X_{m(t-j)}\varepsilon_{(t-j)}], \end{aligned}$$

which is the cross-covariance between $X_{lt}\varepsilon_t$ and $X_{m(t-j)}\varepsilon_{(t-j)}$. We note that

$$\Gamma_{lm}(j) \neq \Gamma_{lm}(-j),$$

because g_t is a vector, not a scalar. Instead, we have

$$\Gamma(j) = \Gamma(-j)',$$

which implies $\Gamma_{lm}(j) = \Gamma_{ml}(-j)$.

Question: What is the $(l+1, m+1)$-th element of $H(\omega)$ when $l \neq m$? The function

$$H_{lm}(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \Gamma_{lm}(j)e^{-ij\omega}$$

is called the cross-spectral density between $\{g_{lt}\}$ and $\{g_{mt}\}$. The cross-spectrum is very useful in investigating the comovements between different economic time series. The popular concept of Granger causality was first defined using the cross-spectrum (see Granger 1969, *Econometrica*). In general, $H_{lm}(\omega)$ is complex-valued.

Question: How to estimate V?

Recall the important identity:

$$V = 2\pi H(0) = \sum_{j=-\infty}^{\infty} \Gamma(j),$$

where $\Gamma(j) = \text{cov}(g_t, g_{t-j})$. The long-run variance V is 2π times $H(0)$, the spectral density matrix at frequency zero. This provides the basis to use a nonparametric approach to estimating V.

A possible naive estimation method:

Given a random sample $\{Y_t, X_t'\}_{t=1}^n$, we can obtain the estimated OLS residual e_t from the linear regression model $Y_t = X_t'\beta^o + \varepsilon_t$. Because

$$V = \sum_{j=-\infty}^{\infty} \Gamma(j),$$

we first consider a naive estimator

$$\hat{V} = \sum_{j=-(n-1)}^{n-1} \hat{\Gamma}(j),$$

where the sample autocovariance function

$$\hat{\Gamma}(j) = \begin{cases} n^{-1}\sum_{t=j+1}^{n} X_t e_t X_{t-j}' e_{t-j}, & j = 0, 1, ..., n-1, \\ n^{-1}\sum_{t=1-j}^{n} X_t e_t X_{t-j}' e_{t-j}, & j = -1, -2, ..., -(n-1). \end{cases}$$

There is no need to subtract the same mean from $X_t e_t$ and $X_{t-j} e_{t-j}$ because $\mathbf{X}'e = \Sigma_{t=1}^n X_t e_t = 0$. Also, note that the summation over lag orders in $\hat{V}$ extends to the maximum lag order $n-1$ for the sample autocovariance function $\hat{\Gamma}(j)$. Unfortunately, although $\hat{\Gamma}(j)$ is consistent for $\Gamma(j)$ for each given j as $n \to \infty$, the estimator $\hat{V}$ is not consistent for V.

Question: Why?

There are too many estimated terms in the summation over lag orders. In fact, there are n estimated parameters $\{\hat{\Gamma}(j)\}_{j=0}^{n-1}$ in $\hat{V}$. The asymptotic variance of the estimator $\hat{V}$ defined above is proportional to the ratio of the number of estimated autocovariance matrices $\{\hat{\Gamma}(j)\}$ to the sample size n, which will not vanish to zero if the number of estimated covariances is the same as or close to the sample size n.

Nonparametric Kernel Estimation

The above explanation motivates us to consider the following truncated sum

$$\hat{V} = \sum_{j=-p}^{p} \hat{\Gamma}(j),$$

where p is a positive integer. If p is fixed (i.e., p does not grow when the sample size n increases), however, we expect

$$\hat{V} \xrightarrow{p} \sum_{j=-p}^{p} \Gamma(j) \neq 2\pi H(0) = V,$$

because the resulting bias

$$2\pi H(0) - \sum_{j=-p}^{p} \Gamma(j) = \sum_{|j|>p} \Gamma(j)$$

will never vanish to zero as $n \to \infty$ when p is fixed. Hence, we should let p grows to infinity as $n \to \infty$; that is, let $p = p(n) \to \infty$ as $n \to \infty$. The bias will then vanish to zero as $n \to \infty$. However, we cannot let p grow as fast as the sample size n. Otherwise, the variance of $\hat{V}$ will never vanish to zero. Therefore, to ensure consistency of $\hat{V}$ to V, we should balance the bias and the variance of $\hat{V}$ properly. This requires using a truncated variance estimator

$$\hat{V} = \sum_{j=-p_n}^{p_n} \hat{\Gamma}(j),$$

where $p_n \to \infty, p_n/n \to 0$. An example $p_n = n^{1/3}$.

Although this estimator is consistent for V, it may not be positive semi-definite for all n. To ensure that it is always positive semi-definite, we can use a weighted average estimator

$$\hat{V} = \sum_{j=-p_n}^{p_n} k(j/p_n)\hat{\Gamma}(j)$$

where the weighting function $k(\cdot)$ is called a kernel function. An example of such kernels is the Bartlett kernel

$$k(z) = (1 - |z|)\mathbf{1}(|z| \leq 1),$$

where $\mathbf{1}(\cdot)$ is the indicator function, which takes value 1 if the condition inside holds, and takes value 0 if the condition inside does not hold. Newey and West (1987, *Econometrica*; 1994, *Review of Economic Studies*) first used this kernel function to estimate V in econometrics. The truncated variance estimator $\hat{V}$ can be viewed as a kernel-based estimator with the use of the truncated kernel $k(z) = \mathbf{1}(|z| \leq 1)$, which assigns equal weighting to each of the first p_n lags.

Most kernels are downward-weighting in the sense that $k(z) \to 0$ as $|z| \to \infty$. The use of a downward weighting kernel may enhance estimation efficiency of V because when $\Sigma_{j=-\infty}^{\infty}||\Gamma(j)|| < \infty$, we have $\Gamma(j) \to 0$ as $j \to \infty$, and so it is more efficient to assign a larger weight to a lower order j and a smaller weight to a higher order j.

In fact, we can consider a more general form of estimator for V :

$$\hat{V} = \sum_{j=1-n}^{n-1} k(j/p_n)\hat{\Gamma}(j),$$

where $k(\cdot)$ may have unbounded support. Although the lag order j sums up from $1-n$ to $n-1$, the variance of the estimator $\hat{V}$ still vanishes to zero, provided $p_n \to \infty, p_n/n \to 0$, and $k(\cdot)$ discounts higher order lags as $j \to \infty$. An example of $k(\cdot)$ that has unbounded support is the Quadratic-Spectral kernel:

$$k(z) = \frac{3}{(\pi z)^2}\left\{\frac{\sin(\pi z)}{\pi z} - \cos(\pi z)\right\}, \quad -\infty < z < \infty.$$

Andrews (1991, *Econometrica*) uses it to estimate for V. This kernel also delivers a p.s.d. matrix. Moreover, it minimizes the asymptotic MSE of the estimator $\hat{V}$ over a class of kernel functions.

Under certain regularity conditions on random sample $\{Y_t, X_t'\}'$, kernel function $k(\cdot)$, and lag order p_n (Newey and West 1987, Andrews 1991), we have

$$\hat{V} \xrightarrow{p} V$$

provided $p_n \to \infty, p_n/n \to 0$. Intuitively, although the summation over lag orders in $\hat{V}$ extends to the maximum lag order $n-1$, the lag orders that are much larger than p_n are expected to have negligible contributions to $\hat{V}$, given that $k(\cdot)$ discounts higher order lags. As a consequence, we have $\hat{V} \to^p V$. There are many rules to satisfy $p_n \to \infty, p_n/n \to 0$. Andrews (1991) and Newey and West (1994) discuss data-driven methods to choose p_n.

Question: What are the regularity conditions on $k(\cdot)$?

Assumption on the kernel function: $k : \mathbb{R} \to [-1, 1]$ is symmetric about 0, is continuous at all points except a finite number of points on $\mathbb{R}$, with $k(0) = 1$ and $\int_{-\infty}^{\infty} k^2(z)dz < \infty$.

At point 0, $k(\cdot)$ attains the maximal value, and the fact that $k(\cdot)$ is square-integrable implies $k(z) \to 0$ as $|z| \to \infty$.

For derivations of asymptotic variance and asymptotic bias of the long-run variance estimator $\hat{V}$, see Newey and West (1987) and Andrews (1991).

6.4 Consistency of OLS

When there exists conditional heteroskedasticity and autocorrelation of unknown form in $\{\varepsilon_t\}$, it is very difficult, if not impossible, to use the GLS estimation. Instead, the OLS estimator $\hat{\beta}$ is convenient to use in practice. We now investigate the asymptotic properties of the OLS $\hat{\beta}$ when there exist conditional heteroskedasticity and autocorrelation of unknown form.

Theorem: *Suppose Assumptions 6.1–6.5(i) hold. Then*

$$\hat{\beta} \xrightarrow{p} \beta^o \text{ as } n \to \infty.$$

Proof: Recall that we have

$$\hat{\beta} - \beta^o = \hat{Q}^{-1} n^{-1} \sum_{t=1}^{n} X_t \varepsilon_t.$$

By Assumptions 6.1, 6.2 and 6.4 and the WLLN for stationary ergodic processes, we have

$$\hat{Q} \xrightarrow{p} Q \text{ and } \hat{Q}^{-1} \to^p Q^{-1}.$$

Similarly, by Assumptions 6.1–6.3 and 6.5(i), we have

$$n^{-1} \sum_{t=1}^{n} X_t \varepsilon_t \xrightarrow{p} E(X_t \varepsilon_t) = 0$$

using the WLLN for ergodic stationary processes, where $E(X_t \varepsilon_t) = 0$ given Assumption 6.2 ($E(\varepsilon_t | X_t) = 0$ a.s.) and LIE.

6.5 Asymptotic normality of OLS

Next, we derive the asymptotic distribution of $\sqrt{n}(\hat{\beta} - \beta^o)$.

Theorem: *Suppose Assumptions 6.1–6.5 hold. Then*

$$\sqrt{n}(\hat{\beta} - \beta^o) \xrightarrow{d} N(0, Q^{-1} V Q^{-1}),$$

where $V = \sum_{j=-\infty}^{\infty} \Gamma(j)$ *is as in Assumption 6.5.*

The proof of this theorem calls for the use of a new CLT.

Lemma [CLT for Zero Mean Ergodic Stationary Processes (White 1984, Theorem 5.15)]: *Suppose* $\{Z_t\}$ *is a stationary ergodic process with*

(i) $E(Z_t) = 0$;

(ii) $V = \sum_{j=-\infty}^{\infty} \Gamma(j)$ *is finite and nonsingular, where* $\Gamma(j) = E(Z_t Z'_{t-j})$;

(iii) $E(Z_t|Z_{t-j}, Z_{t-j-1}, ...) \to^{q.m.} 0$;

(iv) $\sum_{j=0}^{\infty} [E(r'_j r_j)]^{1/2} < \infty$, *where*

$$r_j = E(Z_t|Z_{t-j}, Z_{t-j-1}, ...) - E(Z_t|Z_{t-j-1}, Z_{t-j-2}, ...).$$

Then as $n \to \infty$,

$$n^{1/2}\bar{Z}_n = n^{-1/2} \sum_{t=1}^{n} Z_t \xrightarrow{d} N(0, V).$$

We now use this CLT to derive the asymptotic distribution of $\sqrt{n}(\hat{\beta} - \beta^o)$.

Proof: Recall that

$$\sqrt{n}(\hat{\beta} - \beta^o) = \hat{Q}^{-1} n^{-1/2} \sum_{t=1}^{n} X_t \varepsilon_t.$$

By Assumptions 6.1–6.3 and 6.5 and the CLT for stationary ergodic processes, we have

$$n^{-1/2} \sum_{t=1}^{n} X_t \varepsilon_t \xrightarrow{d} N(0, V),$$

where $V = \sum_{j=-\infty}^{\infty} \Gamma(j)$ is as in Assumption 6.5. Also, $\hat{Q} \to^p Q$ and $\hat{Q}^{-1} \to^p Q^{-1}$ by Assumption 6.4 and the WLLN for ergodic stationary processes. We then have by the Slutsky theorem

$$\sqrt{n}(\hat{\beta} - \beta^o) \xrightarrow{d} N(0, Q^{-1} V Q^{-1}).$$

6.6 Hypothesis Testing

We now consider testing the null hypothesis

$$\mathbf{H}_0 : R\beta^o = r,$$

where R is a nonstochastic $J \times K$ matrix, and r is a $J \times 1$ nonstochastic vector.

When there exists autocorrelation in $\{X_t \varepsilon_t\}$, there is no need (and in fact there is no way) to consider the cases of conditional homoskedasticity and conditional heteroskedasticity separately (why?).

Corollary: *Suppose Assumptions 6.1–6.5 hold. Then under* $\mathbf{H}_0$, *as* $n \to \infty$,

$$\sqrt{n}(R\hat{\beta} - r) \to^d N(0, RQ^{-1} V Q^{-1} R').$$

We directly assume a consistent estimator $\hat{V}$ for V.

Assumption 6.6: $\hat{V} \to^p V$.

When there exists serial correlation of unknown form, we can estimate V using the nonparametric kernel estimator $\hat{V}$, as described in Section 6.3. In some special scenarios, we may have $\Gamma(j) = 0$ for all $j > p_0$, where p_0 is a fixed lag order. An example of this case is Example 2 in Section 6.1. In this case, we can use the following estimator

$$\hat{V} = \sum_{j=-p_0}^{p_0} \hat{\Gamma}(j).$$

It can be shown that $\hat{V} \to^p V$ in this case.

For the case where $J = 1$, a robust t-type test statistic

$$\frac{\sqrt{n}(R\hat{\beta} - r)}{\sqrt{R\hat{Q}^{-1}\hat{V}\hat{Q}^{-1}R'}} \to^d N(0,1),$$

where the convergence to $N(0,1)$ in distribution holds under $\mathbf{H}_0$.

Question: Why is it called a "robust" t-type test?

This statistic has used the asymptotic variance estimator that is robust to conditional heteroskedasticity and autocorrelation of unknown form.

For the case where $J > 1$, we consider a "robust" Wald test.

Theorem: *Under Assumptions 6.1–6.6, we have the Wald test statistic*

$$\hat{W} = n^{-1}(R\hat{\beta} - r)'[R(\mathbf{X}'\mathbf{X})^{-1}\hat{V}(\mathbf{X}'\mathbf{X})^{-1}R']^{-1}(R\hat{\beta} - r) \xrightarrow{d} \chi^2_J$$

as $n \to \infty$ *under* $\mathbf{H}_0 : R\beta^o = r$.

Proof: Because

$$\sqrt{n}(R\hat{\beta} - r) \xrightarrow{d} N(0, RQ^{-1}VQ^{-1}R'),$$

we have the quadratic form

$$\sqrt{n}(R\hat{\beta} - r)'\left(RQ^{-1}VQ^{-1}R'\right)^{-1}\sqrt{n}(R\hat{\beta} - r) \xrightarrow{d} \chi^2_J.$$

By the Slutsky theorem, we have the Wald test statistic

$$\hat{W} = n(R\hat{\beta} - r)'\left(R\hat{Q}^{-1}\hat{V}\hat{Q}^{-1}R'\right)^{-1}(R\hat{\beta} - r) \xrightarrow{d} \chi^2_J.$$

Using the expression of $\hat{Q} = \mathbf{X}'\mathbf{X}/n$, we have an equivalent expression for $\hat{W}$:

$$\hat{W} = n^{-1}(R\hat{\beta} - r)'[R(\mathbf{X}'\mathbf{X})^{-1}\hat{V}(\mathbf{X}'\mathbf{X})^{-1}R']^{-1}(R\hat{\beta} - r) \xrightarrow{d} \chi^2_J.$$

Remark:

The standard t-statistic and F-statistic cannot be used when there exists autocorrelation and conditional heteroskedasticity in $\{X_t\varepsilon_t\}$.

Question: Can we use this Wald test when $\Gamma(j) = 0$ for all nonzero j?

Yes. But this is not a good test statistic because it may perform poorly in finite samples. In particular, it usually overrejects the correct null hypothesis $\mathbf{H}_0$ in finite samples even if $\Gamma(j) = 0$ for all $j \neq 0$. In the case where $\Gamma(j) = 0$ for all $j \neq 0$, a better estimator to use is

$$\begin{aligned}
\hat{V} &= \hat{\Gamma}(0) \\
&= n^{-1}\sum_{t=1}^{n} X_t e_t e_t X_t' \\
&= \mathbf{X}'\mathrm{D}(e)\mathrm{D}(e)'\mathbf{X}/n.
\end{aligned}$$

This is essentially White's heteroskedasticity consistent variance estimator (also see Chapter 5).

Question: Why do the robust t- and Wald tests tend to overreject $\mathbf{H}_0$ in the presence of HAC?

We use the robust t-test as an example. Recall $\hat{V}$ is an estimator for $H(0)$ up to a factor of 2π. When there exists strong positive serial correlation in $\{\varepsilon_t\}$, as is the case of economic time series, $H(\omega)$ will display a peak or mode at frequency zero. The kernel estimator, which is a local averaging estimator, always tends to underestimate $H(0)$, because it has an asymptotic negative bias. Consequently, the robust t-statistic tends to be a larger statistic value, because it is the ratio of $R\hat{\beta} - r$ to the square root of a variance estimator which tends to be smaller than the true variance.

Simulation Evidence

6.7 Testing Whether Long-run Variance-Covariance Estimation is Needed

Because of the notorious poor performance of the robust t- and W tests even when $\Gamma(j) = 0$ for all $j \neq 0$, it is very important to test whether we really have to use a long-run variance estimator.

Question: How to test whether we need to use the long-run variance-covariance matrix estimator? That is, how to test whether the null hypothesis that

$$\mathbf{H}_0 : 2\pi H(0) \equiv \sum_{j=-\infty}^{\infty} \Gamma(j) = \Gamma(0)?$$

The null hypothesis $\mathbf{H}_0$ can be equivalently written as follows:

$$\mathbf{H}_0 : \sum_{j=1}^{\infty} \Gamma(j) = 0.$$

It can arise from two cases:

(i) $\Gamma(j) = 0$ for all $j \neq 0$.

(ii) $\Gamma(j) \neq 0$ for some $j \neq 0$, but $\sum_{j=1}^{\infty} \Gamma(j) = 0$. For simplicity, we will consider the first case only. Case (ii) is pathological, although it could occur in practice.

We now provide a test for $\mathbf{H}_0$ under case (i). See Hong (1997) in a related univariate context.

To test the null hypothesis that $\Sigma_{j=1}^{\infty}\Gamma(j) = 0$, we can use a consistent estimator $\hat{A}$ (say) for $\Sigma_{j=1}^{\infty}\Gamma(j)$ and then check whether $\hat{A}$ is close to a zero matrix. Any significant difference of $\hat{A}$ from zero will indicate the violation of the null hypothesis, and thus a long-run variance estimator is needed.

To estimate $\Sigma_{j=1}^{\infty}\Gamma(j)$ consistently, we can use a nonparametric kernel estimator

$$\hat{A} = \sum_{j=1}^{n-1} k(j/p_n)\text{vech}[\hat{\Gamma}(j)],$$

where $p_n = p(n) \to \infty$ at a suitable rate as $n \to \infty$. We shall derive the asymptotic distribution of $\hat{A}$ (with suitable scaling) under the assumption that $\{g_t = X_t\varepsilon_t\}$ is MDS, which implies the null hypothesis H_o that $\Sigma_{j=1}^{\infty}\Gamma(j) = 0$. First, we consider the case when $\{g_t = X_t\varepsilon_t\}$ is autoregressively conditionally homoskedastic, namely, $\text{var}(g_t|I_{t-1}) = \text{var}(g_t)$, where $I_{t-1} = \{g_{t-1}, g_{t-2}, ...\}$. In this case, we can show

$$\left[p\int_0^{\infty} k^2(z)dz\right]^{-1/2} \text{vech}^{-1}\left[\Gamma(0)\right]\sqrt{n}\hat{A} \to^d N(0, I_{K(K+1)/2}).$$

We can then construct a test statistic

$$\begin{aligned}\hat{M} &= \left[p\int_0^{\infty} k^2(z)dz\right]^{-1} n\hat{A}'\text{vech}^{-2}\left[\hat{\Gamma}(0)\right]\hat{A} \\ &\to {}^d\chi^2_{K(K+1)/2}.\end{aligned}$$

Next, we consider the case when $\{g_t = X_t\varepsilon_t\}$ is autoregresively conditionally heterskedastic, namely $\text{var}(g_t|I_{t-1}) \neq \text{var}(g_t)$. In this case, the test statistic is

$$\hat{M} = \hat{A}'\hat{B}^{-1}\hat{A},$$

where

$$\begin{aligned}
\hat{B} &= \sum_{j=1}^{n-1}\sum_{l=1}^{n-1} k(j/p)k(l/p)\hat{C}(j,l), \\
\hat{C}(j,l) &= \frac{1}{n}\sum_{t=1+\max(j,l)}^{n-1} \text{vech}(\hat{g}_t\hat{g}'_{t-j})\text{vech}'(\hat{g}_t\hat{g}'_{t-l}),
\end{aligned}$$

with $\hat{g}_t = X_t e_t$. Under the assumption that $\{g_t = X_t\varepsilon_t\}$ is an MDS, we have

$$\hat{M} \to^d \chi^2_{K(K+1)/2}.$$

This test is robust to autoregressive conditional heteroskedasticity of unknown form for $\{g_t = X_t\varepsilon_t\}$.

A Related Test: Variance Ratio Test

In fact, the above test is closely related to a variance ratio test that is popular in financial econometrics. Extending an idea of Cochrane (1988), Lo and MacKinlay (1988) first rigorously present an asymptotic theory for a variance ratio test for the MDS hypothesis of asset returns $\{Y_t\}$. Recall that $\sum_{j=1}^p Y_{t-j}$ is the cumulative asset return over a total of p periods. Then under the MDS hypothesis, which implies $\gamma(j) \equiv \text{cov}(Y_t, Y_{t-j}) = 0$ for all $j > 0$, one has

$$\frac{\text{var}\left(\sum_{j=1}^p Y_{t-j}\right)}{p \cdot \text{var}(Y_t)} = \frac{p\gamma(0) + 2p\sum_{j=1}^p(1-j/p)\gamma(j)}{p\gamma(0)} = 1.$$

This unity property of the variance ratio can be used to test the MDS hypothesis because any departure from unity is evidence against the MDS hypothesis.

The variance ratio test is essentially based on the statistic

$$\text{VR}_o \equiv \sqrt{n/p}\sum_{j=1}^p(1-j/p)\hat{\rho}(j) = \frac{\pi}{2}\sqrt{n/p}\left[\hat{f}(0) - \frac{1}{2\pi}\right],$$

where

$$\hat{f}(0) = \frac{1}{2\pi}\sum_{j=-p}^{p}\left(1 - \frac{|j|}{p}\right)\hat{\rho}(j)$$

is a kernel-based normalized spectral density estimator at frequency 0, with the Bartlett kernel $K(z) = (1 - |z|)\mathbf{1}(|z| \leq 1)$ and a lag order equal to p. This, the variance ratio test is the same as checking whether the long-run variance is equal to the individual variance $\gamma(0)$. Because VR_o is based on a spectral density estimator of frequency 0, and because of this, it is particularly powerful against long memory processes, whose spectral density at frequency 0 is infinity (see Robinson 1994, for discussion on long memory processes).

Under the MDS hypothesis with conditional homoskedasticity for $\{Y_t\}$, Lo and MacKinlay (1988) show that for any fixed p,

$$\mathrm{VR}_o \xrightarrow{d} N[0, 2(2p-1)(p-1)/3p] \text{ as } n \to \infty.$$

When $\{Y_t\}$ displays conditional heteroskedasticity, Lo and MacKinlay (1988) also consider a heteroskedasticity-consistent variance ratio test:

$$\mathrm{VR} \equiv \sqrt{n/p} \sum_{j=1}^{p} (1 - j/p)\hat{\gamma}(j)/\sqrt{\hat{\gamma}_2(j)},$$

where $\hat{\gamma}_2(j)$ is a consistent estimator for the asymptotic variance of $\hat{\gamma}(j)$ under conditional heteroskedasticity. Lo and MacKinlay (1988) assume a fourth order cumulant condition that

$$E\left[(Y_t - \mu)^2 (Y_{t-j} - \mu)(Y_{t-l} - \mu)\right] = 0, \qquad j, l > 0, j \neq l. \qquad ((2.4))$$

Intuitively, this condition ensures that the sample autocovariances at different lags are asymptotically uncorrelated; that is, $\mathrm{cov}[\sqrt{n}\hat{\gamma}(j), \sqrt{n}\hat{\gamma}(l)] \to 0$ for all $j \neq l$. As a result, the heteroskedasticity-consistent VR has the same asymptotic distribution as VR_o. However, the condition in (2.4) rules out many important volatility processes, such as EGARCH and Threshold GARCH models. Moreover, the variance ratio test only exploits the implication of the MDS hypothesis on the spectral density at frequency 0; it does not check the spectral density at nonzero frequencies. As a result, it is not consistent against serial correlation of unknown form. See Durlauf (1991) for more discussion.

6.8 A Classical Ornut-Cochrane Procedure

Long-run variance estimators are necessary for statistical inference of the OLS estimation in a linear regression model when there exists serial correlation of *unknown form*. If serial correlation in the regression error has a known special pattern, then simpler statistical inference procedures are possible. One example is the classical Ornut-Cochrane procedure. Consider a linear regression model with serially correlated errors:

$$Y_t = X_t' \beta^o + \varepsilon_t,$$

where $E(\varepsilon_t|X_t) = 0$ but $\{\varepsilon_t\}$ follows an AR(p) process

$$\varepsilon_t = \sum_{j=1}^{p} \alpha_j \varepsilon_{t-j} + v_t, \{v_t\} \sim \text{i.i.d.}(0, \sigma^2).$$

The OLS estimator $\hat{\beta}$ is consistent for β^o given $E(X_t\varepsilon_t) = 0$ but its asymptotic variance depends on serial correlation in $\{\varepsilon_t\}$. We can consider the following transformed linear regression model

$$\begin{aligned} Y_t - \sum_{j=1}^{p} \alpha_j Y_{t-j} &= \left(X_t - \sum_{j=1}^{p} \alpha_j X_{t-j} \right)' \beta^o \\ &\quad + \left(\varepsilon_t - \sum_{j=1}^{p} \alpha_j \varepsilon_{t-j} \right) \\ &= \left(X_t - \sum_{j=1}^{p} \alpha_j X_{t-j} \right)' \beta^o + v_t. \end{aligned}$$

We can write it as follows:

$$Y_t^* = X_t^{*\prime} \beta^o + v_t,$$

where

$$\begin{aligned} Y_t^* &= Y_t - \sum_{j=1}^{p} \alpha_j Y_{t-j}, \\ X_t^* &= X_t - \sum_{j=1}^{p} \alpha_j X_{t-j}. \end{aligned}$$

The OLS estimator $\tilde{\beta}$ of this transformed regression will be consistent for β^o and asymptotically normal:

$$\sqrt{n}(\tilde{\beta} - \beta^o) \to^d N(0, \sigma_v^2 Q_{x^*x^*}^{-1}),$$

where $Q_{x^*x^*} = E(X_t^* X_t^{*\prime})$. Moreover it is asymptotically BLUE. However, the OLS estimator $\tilde{\beta}$ is infeasible, because (Y_t^*, X_t^*) is not available due to the unknown parameters $\{\alpha_j\}_{j=1}^{p}$. As a solution, one can use a feasible two-step procedure:

- Step 1: Regress

  $$Y_t = X_t'\beta^o + \varepsilon_t, t = 1, ..., n,$$

  Y_t on X_t, and obtain the estimated OLS residual $e_t = Y_t - X_t'\hat{\beta}$;

- Step 2: Regress an AR(p) model

$$e_t = \sum_{j=1}^{p} \alpha_j e_{t-j} + \tilde{v}_t, t = p+1, ..., n,$$

obtain the OLS estimators $\{\hat{\alpha}_j\}_{j=1}^p$;

- Step 3: Regress the transformed model

$$\hat{Y}_t^* = \hat{X}_t^{*\prime} \beta^o + v_t^*, t = p+1, ..., n,$$

where $\hat{Y}_t^*$ and $\hat{X}_t^*$ are defined in the same way as Y_t and X_t respectively, with $\{\hat{\alpha}_j\}_{j=1}^p$ replacing $\{\alpha_j\}_{j=1}^p$. The resulting OLS estimator is denoted as $\tilde{\beta}_a$.

It can be shown that the adaptive feasible OLS estimator $\tilde{\beta}_a$ has the same asymptotic properties as the infeasible OLS estimator $\tilde{\beta}$. In other words, the sampling error resulting from the first step estimation has no impact on the asymptotic properties of the OLS estimator in the second step. The asymptotic variance estimator of $\tilde{\beta}_a$ is given by

$$\hat{s}_v^2 \hat{Q}_{x^*x^*}^{-1},$$

where

$$\begin{aligned} \hat{s}_v^2 &= \frac{1}{n-K} \sum_{t=1}^{n} \hat{v}_t^{*2}, \\ \hat{Q}_{x^*x^*} &= \frac{1}{n} \sum_{t=1}^{n} \hat{X}_t^* \hat{X}_t^{*\prime}, \end{aligned}$$

with $\hat{v}_t = \hat{Y}_t^* - \hat{X}_t^{*\prime} \tilde{\beta}_a$. The t-test statistic which is asymptotically $N(0,1)$ and the $J \cdot F$-test statistic which is asymptotically χ_J^2 from the last stage regression are applicable when the sample size n is large.

The estimator $\tilde{\beta}_a$ is essentially the adaptive feasible GLS estimator described in Chapter 3, and it is asymptotically BLUE. This estimation method is therefore asymptotically more efficient than the robust test procedures developed in Section 6, but it is based on the assumption that the AR(p) process for the disturbance $\{\varepsilon_t\}$ is known. The robust test procedures in Section 6 are applicable when $\{\varepsilon_t\}$ has conditional heteroscedasticity and serial correlation of unknown form.

6.9 Empirical Applications

6.10 Conclusion

In this chapter, we have first discussed some motivating economic examples where a long-run variance estimator is needed. Then we discussed consistent estimation of a long-run variance-covariance matrix by a nonparametric kernel method. The asymptotic properties of the OLS estimator are investigated, which calls for the use of a new CLT because $\{X_t\varepsilon_t\}$ is not a MDS. Robust t- and Wald test statistics that are valid under conditional heteroskedasticity and autocorrelation of unknown form are then derived. When there exists serial correlation of unknown form, there is no need (and no way) to separate the cases of conditional homoskedasticity and conditional heteroskedasticity. Because robust t- and Wald tests have very poor finite sample performances even if $\{X_t\varepsilon_t\}$ is a MDS, it is desirable to first check whether we really need a long-run variance estimator. We provide such a test. Finally, some empirical applications are considered. We also introduce a classical estimation method called Ornut-Ochrance procedure when it is known that the regression disburbance follows an AR process with a known order.

Long-run variances have been also widely used in nonstationary time series econometrics such as in unit root and cointegration (e.g., Phillips 1987).

References

Andrews, D.W.K. (1991), Econometrica.

Hong, Y. (1997), *Journal of Time Series Analysis.*

Lo, A. and MacKinlay (1988), *Review of Financial Studies.*

Newey, W. and K. West (1987), Econometrica.

Newey, W. and K. West (1994), Review of Economic Studies.

Phillips, P. (1987), Econometrica.

EXERCISES

6.1. Suppose Assumptions 6.1–6.3 and 6.5(i) hold. Show

$$\begin{aligned}\text{avar}\left(n^{-1/2}\sum_{t=1}^{n} X_t\varepsilon_t\right) &\equiv \lim_{n\to\infty}\text{var}\left(n^{-1/2}\sum_{t=1}^{n} X_t\varepsilon_t\right)\\ &= \sum_{j=-\infty}^{\infty}\Gamma(j).\end{aligned}$$

6.2. Suppose $\Gamma(j) = 0$ for all $j > p_0$, where p_0 is a fixed lag order. An example of this case is Example 2 in Section 6.1. In this case, the long-run variance $V = \sum_{j=-p_0}^{p_0}\Gamma(j)$ and we can estimate it by using the following estimator

$$\hat{V} = \sum_{j=-p_0}^{p_0}\hat{\Gamma}(j).$$

where $\hat{\Gamma}(j)$ is defined as in Section 6.1. Show that for each given j, $\hat{\Gamma}(j) \to^p \Gamma(j)$ as $n \to \infty$.

Given that p_0 is a fixed interger, an important implication that $\hat{\Gamma}(j) \to^p \Gamma(j)$ for each given j as $n \to \infty$ is that $\hat{V} \to^p V$ as $n \to \infty$.

6.3. Suppose $\{Y_t\}$ is a stationary time series process with the following spectral density function exists:

$$h(\omega) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\gamma(j)e^{-ij\omega}.$$

Show that

$$\text{var}\left(\sum_{j=1}^{p} Y_{t-j}\right) \to 2\pi h(0) \text{ as } p \to \infty.$$

6.4. Suppose $\{Y_t\}$ is a weakly stationary process with $\gamma(j) = \text{cov}(Y_t, Y_{t-j})$.

(a) Find an example of $\{Y_t\}$ such that $\sum_{j=1}^{\infty}\gamma(j) = 0$ but there exists at least one $j > 0$, such that $\gamma(j) \neq 0$.

(b) Can the variance ration test detect the time series process in part (a) with a high probability.