CHAPTER 4 ASYMPTOTIC THEORY FOR LINEAR REGRESSION MODELS WITH I.I.D. OBSERVATIONS

Key words: Asymptotic analysis, Almost sure convergence, Central limit theorem, Convergence in distribution, Convergence in quadratic mean, Convergence in probability, I.I.D., Law of large numbers, the Slutsky theorem, White's heteroskedasticity-consistent variance estimator.

Abstract: When the conditional normality assumption on the regression error does not hold, the OLS estimator no longer has the finite sample normal distribution, and the t-test statistics and F-test statistics no longer follow a Student t-distribution and a F-distribution in finite samples respectively. In this chapter, we show that under the assumption of i.i.d. observations with conditional homoskedasticity, the classical t-test and F-test are approximately applicable in large samples. However, under conditional heteroskedasticity, the t-test statistics and F-test statistics are not applicable even when the sample size goes to infinity. Instead, White's (1980) heteroskedasticity-consistent variance estimator should be used, which yields asymptotically valid hypothesis test procedures. A direct test for conditional heteroskedasticity due to White (1980) is presented. To facilitate asymptotic analysis in this and subsequent chapters, we first introduce some basic tools in asymptotic analysis.

Motivation

The assumptions of classical linear regression models are rather strong and one may have a hard time finding practical applications where all these assumptions hold exactly. For example, it has been documented that most economic and financial data have heavy tails, and so are not normally distributed. An interesting question now is whether the estimators and tests which are based on the same principles as before still make sense in this more general setting. In particular, what happens to the OLS estimator, the t- and F-tests if any of the following assumptions fails:

- strict Exogeneity $E(\varepsilon_t|\mathbf{X}) = 0$?
- normality $(\varepsilon|\mathbf{X} \sim N(0, \sigma^2 I))$?
- conditional homoskedasticity $(\mathrm{var}(\varepsilon_t|\mathbf{X}) = \sigma^2)$?
- serial uncorrelatedness $(\mathrm{cov}(\varepsilon_t, \varepsilon_s|\mathbf{X}) = 0$ for $t \neq s)$?

When classical assumptions are violated, we do not know the finite sample statistical properties of the estimators and test statistics anymore. A useful tool to obtain the understanding of the properties of estimators and tests in this more general setting is to pretend that we can obtain a limitless number of observations. We can then pose the question how estimators and test statistics would behave when the number of observations increases without limit. This is called *asymptotic analysis*. In practice, the sample size is always finite. However, the asymptotic properties translate into results that hold true approximately in finite samples, provided that the sample size is large enough. We now need to introduce some basic analytic tools for asymptotic theory. For more systematic introduction of asymptotic theory, see, for example, White (1994, 1999).

Introduction to Asymptotic Theory

In this section, we introduce some important convergence concepts and limit theorems. First, we introduce the concept of convergence in mean squares, which is a distance measure of a sequence of random variables from a random variable.

Definition [Convergence in mean squares (or in quadratic mean)] *A sequence of random variables/vectors/matrices $Z_n, n = 1, 2, ...,$ is said to converge to Z in mean squares as $n \to \infty$ if*

$$E||Z_n - Z||^2 \to 0 \text{ as } n \to \infty,$$

where $||\cdot||$ is the sum of the absolute value of each component in $Z_n - Z$.

When Z_n is a vector or matrix, convergence can be understood as convergence in each element of Z_n.When $Z_n - Z$ is a $l \times m$ matrix, where l and m are fixed positive integers, then we can also define the squared norm as

$$||Z_n - Z||^2 = \sum_{t=1}^{l}\sum_{s=1}^{m}[Z_n - Z]^2_{(t,s)}.$$

Note that Z_n converges to Z in mean squares if and only if each component of Z_n converges to the corresponding conponent of Z in mean squares.

Example 1: Suppose $\{Z_t\}$ is i.i.d.(μ, σ^2), and $\bar{Z}_n = n^{-1}\sum_{t=1}^{n} Z_t$. Then

$$\bar{Z}_n \overset{q.m.}{\to} \mu.$$

Solution: Because $E(\bar{Z}_n) = \mu$, we have

$$\begin{aligned}
E(\bar{Z}_n - \mu)^2 &= \text{var}(\bar{Z}_n) \\
&= \text{var}\left(n^{-1}\sum_{t=1}^{n} Z_t\right) \\
&= \frac{1}{n^2}\text{var}\left(\sum_{t=1}^{n} Z_t\right) \\
&= \frac{1}{n^2}\sum_{t=1}^{n}\text{var}(Z_t) \\
&= \frac{\sigma^2}{n} \\
&\to 0 \text{ as } n \to \infty.
\end{aligned}$$

It follows that

$$E(\bar{Z}_n - \mu)^2 = \frac{\sigma^2}{n} \to 0 \text{ as } n \to \infty.$$

Next, we introduce the concept of convergence in probability that is another popular distance measure between a sequence of random variables and random variable.

Definition [Convergence in probability] *Z_n converges to Z in probability if for any given constant $\epsilon > 0$,*

$$\begin{aligned} \Pr[||Z_n - Z|| &> \epsilon] \to 0 \text{ as } n \to \infty \text{ or} \\ \Pr[||Z_n - Z|| &\le \epsilon] \to 1 \text{ as } n \to \infty. \end{aligned}$$

For convergence in probability, we can also write

$$Z_n - Z \xrightarrow{p} 0 \text{ or } Z_n - Z = o_P(1),$$

The notation $o_P(1)$ means that $Z_n - Z$ vanishes to zero in probability. When $Z = b$ is a constant, we can write $Z_n \xrightarrow{p} b$ and $b = p\lim Z_n$ is called the probability limit of Z_n.

Convergence in probability is also called weak convergence or convergence with probability approaching one. When $Z_n \to^p Z$, the probability that the difference $||Z_n - Z||$ exceeds any given small constant ϵ is rather small for all n sufficiently large. In other words, Z_n will be very close to Z with very high probability, when the sample size n is sufficiently large.

To gain more intuition of the convergence in probability, we define the event

$$A_n(\epsilon) = \{\omega \in \Omega : |Z_n(\omega) - Z(\omega)| > \epsilon\},$$

where ω is a basic outcome in sample space Ω. Then convergence in probability says that the probability of event $A_n(\epsilon)$ may be nonzero for any finite n, but such a probability will eventually vanish to zero as $n \to \infty$. In other words, it becomes less and less likely that the difference $|Z_n - Z|$ is larger than a prespecified constant $\epsilon > 0$. Or, we have more and more confidence that the difference $|Z_n - Z|$ will be smaller than ϵ as $n \to \infty$. The constant ϵ can be viewed as a prespecified tolerance level.

Lemma [Weak Law of Large Numbers (WLLN) for I.I.D. Sample] *Suppose $\{Z_t\}$ is i.i.d.(μ, σ^2), and define $\bar{Z}_n = n^{-1}\sum_{t=1}^n Z_t, n = 1, 2, \ldots$. Then*

$$\bar{Z}_n \xrightarrow{p} \mu \text{ as } n \to \infty.$$

Proof: For any given constant $\epsilon > 0$, we have by Chebyshev's inequality

$$\begin{aligned} \Pr(|\bar{Z}_n - \mu| &> \epsilon) \le \frac{E(\bar{Z}_n - \mu)^2}{\epsilon^2} \\ &= \frac{\sigma^2}{n\epsilon^2} \to 0 \text{ as } n \to \infty. \end{aligned}$$

Hence,

$$\bar{Z}_n \xrightarrow{p} \mu \text{ as } n \to \infty.$$

This is the so-called weak law of large numbers (WLLN). In fact, we can weaken the moment condition.

We now provide an economic interpretation of the WLLN using an example. In finance, there is a popular trading strategy called buy-and-hold trading startegy. An investor buys a stock at some day and then hold it for a long time period before he sells it out. This is called a buy-and-hold trading strategy. How is the average return of this trading strategy?

Suppose Z_t is the return of the stock on period t, and the returns over different time periods are i.i.d.(μ, σ^2). Also assume the investor holds the stock for a total of n period. Then the average return over each time period is the sample mean

$$\bar{Z} = \frac{1}{n}\sum_{t=1}^{n} Z_t.$$

When the number n of holding periods is large, we have

$$\bar{Z} \to^p \mu = E(Z_t)$$

as $n \to \infty$. Thus, the average return of the buy-and-hold trading startegy is approximately equal to μ when n is sufficiently large.

Lemma [WLLN for I.I.D. Random Sample] *Suppose $\{Z_t\}$ is i.i.d. with $E(Z_t) = \mu$ and $E|Z_t| < \infty$. Define $\bar{Z}_n = n^{-1}\sum_{t=1}^{n} Z_t$. Then*

$$\bar{Z}_n \xrightarrow{p} \mu \text{ as } n \to \infty.$$

Question: Why do we need the moment condition $E|Z_t| < \infty$?

Yes. We can consider a counter example: Suppose $\{Z_t\}$ is a sequence of i.i.d. Cauchy$(0,1)$ random variables whose moments do not exist. Then $\bar{Z}_n \sim$ Cauchy$(0,1)$ for all $n \geq 1$, and so it does not converge in probability to some constant as $n \to \infty$.

We now introduce a related useful concept:

Definition [Boundedness in Probability] *A sequence of random variables/vectors/matrices $\{Z_n\}$ is bounded in probability if for any small constant $\delta > 0$, there exists a constant $C < \infty$ such that*

$$P(||Z_n|| > C) \leq \delta$$

as $n \to \infty$. We denote

$$Z_n = O_P(1).$$

Intuitively, when $Z_n = O_P(1)$, the probability that $||Z_n||$ exceeds a very large constant is small as $n \to \infty$. Or, equivalently, $||Z_n||$ is smaller than C with a very high probability as $n \to \infty$.

Example: Suppose $Z_n \sim N(\mu, \sigma^2)$ for all $n \geq 1$. Then

$$Z_n = O_P(1).$$

Solution: For any $\delta > 0$, we always have a sufficiently large constant $C = C(\delta) > 0$ such that

$$\begin{aligned}
P(|Z_n| &> C) = 1 - P(-C \leq Z_n \leq C) \\
&= 1 - P\left[\frac{-C-\mu}{\sigma} \leq \frac{Z_n - \mu}{\sigma} \leq \frac{C-\mu}{\sigma}\right] \\
&= 1 - \Phi\left(\frac{C-\mu}{\sigma}\right) + \Phi\left(-\frac{C+\mu}{\sigma}\right) \\
&\leq \delta,
\end{aligned}$$

where $\Phi(z) = P(Z \leq z)$ is the CDF of $N(0,1)$. [We can choose C such that $\Phi[(C-\mu)/\sigma] \geq 1 - \frac{1}{2}\delta$ and $\Phi[-(C+\mu)/\sigma] \leq \frac{1}{2}\delta$.]

A Special Case: What happens to C if $Z_n \sim N(0,1)$?

In this case,

$$\begin{aligned}
P(|Z_n| &> C) = 1 - \Phi(C) + \Phi(-C) \\
&= 2[1 - \Phi(C)].
\end{aligned}$$

Suppose we set

$$2[1 - \Phi(C)] = \delta,$$

that is, we set

$$C = \Phi^{-1}\left(1 - \frac{\delta}{2}\right),$$

where $\Phi^{-1}(\cdot)$ is the inverse function of $\Phi(\cdot)$. Then we have

$$P(|Z_n| > C) = \delta.$$

The following lemma provides a convenient way to verify convergence in probability.

Lemma: If $Z_n - Z \xrightarrow{q.m.} 0$, *then* $Z_n - Z \xrightarrow{p} 0$.

Proof: By Chebyshev's inequality, we have

$$P(|Z_n - Z| > \epsilon) \leq \frac{E[Z_n - Z]^2}{\epsilon^2} \to 0$$

for any given $\epsilon > 0$ as $n \to \infty$. This completes the proof.

Example: Suppose Assumptions 3.1–3.4 hold. Does the OLS estimator $\hat{\beta}$ converges in probability to β^o?

Solution: From Theorem, we have

$$\begin{aligned} \tau' E[(\hat{\beta} - \beta^o)(\hat{\beta} - \beta^o)'|\mathbf{X}]\tau &= \sigma^2 \tau'(X'X)^{-1}\tau \\ &\to 0 \end{aligned}$$

for any $\tau \in R^K, \tau'\tau = 1$ as $n \to \infty$ with probability one. It follows that $E||\hat{\beta} - \beta^o||^2 = E\{E[||\hat{\beta} - \beta^o||^2|X]\} \to 0$ as $n \to \infty$. Therefore, by Lemma, we have $\hat{\beta} \to^p \beta^o$.

Example: Suppose Assumptions 3.1, 3.3 and 3.5 hold. Does s^2 converge in probability to σ^2?

Solution: Under the given assumptions,

$$(n - K)\frac{s^2}{\sigma^2} \sim \chi^2_{n-K},$$

and therefore we have $E(s^2) = \sigma^2$ and $\text{var}(s^2) = \frac{2\sigma^4}{n-K}$. It follows that $E(s^2 - \sigma^2)^2 = 2\sigma^4/(n - K) \to 0$, $s^2 \to^{q.m.} \sigma^2$ and so $s^2 \xrightarrow{p} \sigma^2$ because convergence in quadratic mean implies convergence in probability.

While convergence in mean squares implies convergence in probability, the converse is not true. We now give an example.

Example: Suppose

$$Z_n = \begin{cases} 0 & \text{with prob } 1 - \frac{1}{n} \\ n & \text{with prob } \frac{1}{n}. \end{cases}$$

Then $Z_n \xrightarrow{p} 0$ as $n \to \infty$ but $E(Z_n - 0)^2 = n \to \infty$. Please verify it.

Solution:

(i) For any given $\varepsilon > 0$, we have

$$P(|Z_n - 0| > \varepsilon) = P(Z_n = n) = \frac{1}{n} \to 0.$$

(ii)

$$\begin{aligned} E(Z_n - 0)^2 &= \sum_{z_n \in \{0,n\}} (z_n - 0)^2 f(z_n) \\ &= (0-0)^2 \cdot \left(1 - n^{-1}\right) + (n-0)^2 \cdot n^{-1} \\ &= n \to \infty. \end{aligned}$$

Next, we provide another convergence concept called almost sure convergence.

Definition [Almost Sure Convergence] $\{Z_n\}$ *converges to* Z *almost surely if*

$$\Pr\left[\lim_{n\to\infty} ||Z_n - Z|| = 0\right] = 1.$$

We denote $Z_n - Z \xrightarrow{a.s.} 0$.

To gain intuition for the concept of almost sure convergence, recall the definition of a random variable: any random variable is a mapping from the sample space Ω to the real line, namely $Z : \Omega \to \mathbb{R}$. Let ω be a basic outcome in the sample space Ω. Define a subset in Ω :

$$A^c = \{\omega \in \Omega : \lim_{n\to\infty} Z_n(\omega) = Z(\omega)\}.$$

That is, A^c is the set of basic outcomes on which the sequence of $\{Z_n(\cdot)\}$ converges to $Z(\cdot)$ as $n \to \infty$. Then almost sure convergence can be stated as

$$P(A^c) = 1.$$

In other words, the convergent set A^c has probability one to occur.

Example: Let ω be uniformly distributed on $[0, 1]$, and define

$$Z(\omega) = \omega \text{ for all } \omega \in [0, 1].$$

and

$$Z_n(\omega) = \omega + \omega^n \text{ for } \omega \in [0, 1].$$

Is $Z_n - Z \xrightarrow{a.s.} 0$?

Solution: Consider

$$A^c = \{\omega \in \Omega : \lim_{n\to\infty} |Z_n(\omega) - Z(\omega)| = 0\}.$$

Because for any given $\omega \in [0, 1)$, we always have

$$\begin{aligned} \lim_{n\to\infty} |Z_n(\omega) - Z(\omega)| &= \lim_{n\to\infty} |(\omega + \omega^n) - \omega| \\ &= \lim_{n\to\infty} \omega^n = 0. \end{aligned}$$

In contrast, for $\omega = 1$, we have

$$\lim_{n\to\infty} |Z_n(1) - Z(1)| = 1^n = 1 \neq 0.$$

Thus, $A^c = [0, 1)$ and $P(A^c) = 1$. We also have $P(A) = P(\omega = 1) = 0$.

In probability theory, almost sure convergence is closely related to pointwise convergence (almost everywhere). It is also called strong convergence.

Lemma [Strong Law of Large Numbers (SLLN) for I.I.D. Random Samples] *Suppose $\{Z_t\}$ be i.i.d. with $E(Z_t) = \mu$ and $E|Z_t| < \infty$. Then*

$$\bar{Z}_n \xrightarrow{a.s.} \mu \text{ as } n \to \infty.$$

Almost sure convergence implies convergence in probability but not vice versa.

Question: If $s^2 \to^p \sigma^2$, do we have $s \to^p \sigma$?

Answer: Yes. It follows from the following continuity lemma with the choice of $g(s^2) = \sqrt{s^2} = s$.

Lemma [Continuity]: *(i) Suppose $a_n \xrightarrow{p} a$ and $b_n \xrightarrow{p} b$, and $g(\cdot)$ and $h(\cdot)$ are continuous functions. Then*

$$\begin{aligned} &g(a_n) + h(b_n) \xrightarrow{p} g(a) + h(b), \text{ and} \\ &g(a_n)h(b_n) \xrightarrow{p} g(a)h(b). \end{aligned}$$

(ii) Similar results hold for almost sure convergence.

The last convergence concept we will introduce is called convergenve in distribution.

It may be emphasized that convergence in mean squres, convergence in probability and almost sure convergence all measure the closeness between the random variable Z_n

and the random variable Z. This differs from the concept of convergence in distribution introduced in Chapter 3. There, convergence in distribution is defined in terms of the closeness of the CDF $F_n(z)$ of Z_t to the CDF $F(z)$ of Z, not between the closeness of the random variable Z_n to the random variable Z. As a result, for convergence in mean squares, convergence in probability and almost sure convergence, Z_n converges to Z if and only if convergence of Z_n to Z occurs element by element (that is, each element of Z_n converges to the corresponding element of Z). For the convergence in distribution of Z_n to Z, however, element by element convergence does not imply convergence in distribution of Z_n to Z,because element-wise convergence in distribution ignores the relationships among the components of Z_n. Nevertheless, $Z_n \xrightarrow{d} Z$ does imply element by element convergence in distribution. That is, convergence in joint distribution implies convergence in marginal distribution.

The main purpose of asymptotic analysis is to derive the large sample distribution of the estimator or statistic of interest and use it as an approximation in statistical inference. For this purpose, we need to make use of an important limit theorem, namely Central Limit Theorem (CLT). We now state and prove the CLT for i.i.d. random samples, a fundamental limit theorem in probability theory.

Lemma [Central Limit Theorem (CLT) for I.I.D. Random Samples]: *Suppose* $\{Z_t\}$ *is i.i.d.*(μ, σ^2), *and* $\bar{Z}_n = n^{-1}\sum_{t=1}^n Z_t$. *Then as* $n \to \infty$,

$$
\begin{aligned}
\frac{\bar{Z}_n - E(\bar{Z}_n)}{\sqrt{\mathrm{var}(\bar{Z}_n)}} &= \frac{\bar{Z}_n - \mu}{\sqrt{\sigma^2/n}} \\
&= \frac{\sqrt{n}(\bar{Z}_n - \mu)}{\sigma} \\
&\xrightarrow{d} N(0,1).
\end{aligned}
$$

Proof: Put

$$Y_t = \frac{Z_t - \mu}{\sigma},$$

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

$$\frac{\sqrt{n}(\bar{Z}_n - \mu)}{\sigma} = \sqrt{n}\bar{Y}_n.$$

The characteristic function of $\sqrt{n}\bar{Y}_n$

$$\begin{aligned}
\phi_n(u) &= E[\exp(iu\sqrt{n}\bar{Y}_n)], \qquad i=\sqrt{-1} \\
&= E\left[\exp\left(\frac{iu}{\sqrt{n}}\sum_{t=1}^{n} Y_t\right)\right] \\
&= \prod_{t=1}^{n} E\left[\exp\left(\frac{iu}{\sqrt{n}} Y_t\right)\right] \text{ by independence} \\
&= \left[\phi_Y\left(\frac{u}{\sqrt{n}}\right)\right]^n \text{ by identical distribution.} \\
&= \left[\phi_Y(0) + \phi'(0)\frac{u}{\sqrt{n}} + \frac{1}{2}\phi''(0)\frac{u^2}{n} + \cdots\right]^n \\
&= \left(1 - \frac{u^2}{2n}\right)^n + o(1) \\
&\to \exp\left(-\frac{u^2}{2}\right) \text{ as } n \to \infty,
\end{aligned}$$

where the third equality follows by independence, the fourth equality follows by identical distribution, the fifth equality follows by the Taylor series expansion, and $\phi(0) = 1, \phi'(0) = 0, \phi''(0) = -1$. Note that $o(1)$ means a reminder term that vanishes to zero as $n \to \infty$, and we have also made use of the fact that $\left(1 + \frac{a}{n}\right)^n \to e^a$.

More rigorously, we can show

$$\begin{aligned}
\ln \phi_n(u) &= n \ln \phi_Y\left(\frac{u}{\sqrt{n}}\right) \\
&= \frac{\ln \phi_Y\left(\frac{u}{\sqrt{n}}\right)}{n^{-1}} \\
&\to \frac{u}{2}\lim_{n\to\infty} \frac{\frac{\phi_Y'(u/\sqrt{n})}{\phi_Y(u/\sqrt{n})}}{n^{-1/2}} \\
&= \frac{u^2}{2}\lim_{n\to\infty} \frac{\phi_Y''(u/\sqrt{n})\phi_Y(u/\sqrt{n}) - [\phi_Y'(u/\sqrt{n})]^2}{\phi_Y^2(u/\sqrt{n})} \\
&= -\frac{u^2}{2}.
\end{aligned}$$

It follows that

$$\lim_{n\to\infty} \phi_n(u) = e^{-\frac{1}{2}u^2}.$$

This is the characteristic function of $N(0,1)$. By the uniqueness of the characteristic function, the asymptotic distribution of

$$\frac{\sqrt{n}(\bar{Z}_n - \mu)}{\sigma}$$

is $N(0,1)$. This completes the proof.

Lemma [Cramer-Wold Device] *A* $p \times 1$ *random vector* $Z_n \xrightarrow{d} Z$ *if and only if for any nonzero* $\lambda \in R^p$ *such that* $\lambda'\lambda = \Sigma_{j=1}^p \lambda_j^2 = 1$, *we have*

$$\lambda' Z_n \xrightarrow{d} \lambda' Z.$$

This lemma is useful for obtaining asymptotic multivariate distributions.

Slutsky Theorem: *Let* $Z_n \xrightarrow{d} Z$, $a_n \xrightarrow{p} a$ *and* $b_n \xrightarrow{p} b$, *where* a *and* b *are constants. Then*

$$a_n + b_n Z_n \xrightarrow{d} a + bZ \text{ as } n \to \infty.$$

Question: If $X_n \to^d X$ and $Y_n \to^d Y$. Is $X_n + Y_n \to^d X + Y$?

Answer: No. We consider two examples:

Example 1: X_n and Y_n are independent N(0,1). Then

$$X_n + Y_n \to^d N(0,2).$$

Example 2: $X_n = Y_n \sim N(0,1)$ for all $n \geq 1$. Then

$$X_n + Y_n = 2X_n \sim N(0,4).$$

Example 3: Suppose Assumptions 3.1, 3.3(a) and 3.5, and the hypothesis $\mathbf{H}_0 : R\beta^o = r$ hold, where R is a $J \times K$ nonstochastic matrix with rank J, r is a $J \times 1$ nonstochastic vector, and $J \leq K$. Then the quadratic form

$$\frac{(R\hat{\beta} - r)'[R(\mathbf{X}'\mathbf{X})^{-1}R']^{-1}(R\hat{\beta} - r)}{\sigma^2} \sim \chi^2_J.$$

Suppose now we replace σ^2 by s^2. What is the asymptotic distribution of the quadratic form

$$\frac{(R\hat{\beta} - r)'[R(\mathbf{X}'\mathbf{X})^{-1}R']^{-1}(R\hat{\beta} - r)}{s^2}?$$

Finally, we introduce a lemma which is very useful in deriving the asymptotic distributions of nonlinear statistics (i.e., nonlinear functions of the random sample).

Lemma [Delta Method] *Suppose* $\sqrt{n}(\bar{Z}_n - \mu)/\sigma \to^d N(0,1)$, *and* $g(\cdot)$ *is continuously differentiable with* $g'(\mu) \neq 0$. *Then as* $n \to \infty$,

$$\sqrt{n}[g(\bar{Z}_n) - g(\mu)] \to^d N(0, [g'(\mu)]^2\sigma^2).$$

Proof: First, because $\sqrt{n}(\bar{Z}_n - \mu)/\sigma \to^d N(0,1)$ implies $\sqrt{n}(\bar{Z}_n - \mu)/\sigma = O_P(1)$, we have $\bar{Z}_n - \mu = O_P(n^{-1/2}) = o_P(1)$.

Next, by a first order Taylor series expansion, we have

$$Y_n = g(\bar{Z}_n) = g(\mu) + g'(\bar{\mu}_n)(\bar{Z}_n - \mu),$$

where $\bar{\mu}_n = \lambda\mu + (1-\lambda)\bar{Z}_n$ for $\lambda \in [0,1]$. It follows by the Slutsky theorem that

$$\begin{aligned} \sqrt{n}\frac{g(\bar{Z}_n) - g(\mu)}{\sigma} &= g'(\bar{\mu}_n)\sqrt{n}\frac{\bar{Z}_n - \mu}{\sigma} \\ &\to {}^d N(0, [g'(\mu)]^2), \end{aligned}$$

where $g'(\bar{\mu}_n) \to^p g'(\mu)$ given $\bar{\mu}_n \to^p \mu$.

By the Slutsky theorem again, we have

$$\sqrt{n}[Y_n - g(\mu)] \to^d N(0, \sigma^2[g'(\mu)]^2).$$

This completes the proof.

The Delta method is a Taylor series approximation in a statistical context. It linearizes a smooth (i.e., differentiable) nonlinear statistic so that the CLT can be applied to the linearized statistic. Therefore, it can be viewed as a generalization of the CLT from a sample average to a nonlinear statistic. This method is very useful when more than one parameter makes up the function to be estimated and more than one random variable is used in the estimator.

Example: Suppose $\sqrt{n}(\bar{Z}_n - \mu)/\sigma \to^d N(0,1)$ and $\mu \neq 0$ and $0 < \sigma < \infty$. Find the limiting distribution of $\sqrt{n}(\bar{Z}_n^{-1} - \mu^{-1})$.

Solution: Put $g(\bar{Z}_n) = \bar{Z}_n^{-1}$. Because $\mu \neq 0$, $g(\cdot)$ is continuous at μ. By a first order Taylor series expansion, we have

$$\begin{aligned} g(\bar{Z}_n) &= g(\mu) + g'(\bar{\mu}_n)(\bar{Z}_n - \mu), \text{ or} \\ \bar{Z}_n^{-1} - \mu^{-1} &= (-\bar{\mu}_n^{-2})(\bar{Z}_n - \mu) \end{aligned}$$

where $\bar{\mu}_n = \lambda\mu + (1-\lambda)\bar{Z}_n \to^p \mu$ given $\bar{Z}_n \to^p \mu$ and $\lambda \in [0,1]$. It follows that

$$\begin{aligned} \sqrt{n}(\bar{Z}_n^{-1} - \mu^{-1}) &= -\frac{\sigma}{\bar{\mu}_n^2}\frac{\sqrt{n}(\bar{Z}_n - \mu)}{\sigma} \\ &\to {}^d N(0, \sigma^2/\mu^4). \end{aligned}$$

Taylor series expansions, various convergence concepts, laws of large numbers, and central limit theorems, and slutsky theorem constitute a tool kit of asymptotic analysis. We now use these asymptotic tools to investigate the large sample behavior of the OLS estimator and related statistics in subsequent chapters.

Large Sample Theory for Linear Regression Models with I.I.D. Observations

4.1 Assumptions

We first state the assumptions under which we will estalish the asymptotic theory for linear regression models.

Assumption 4.1 [I.I.D.]: $\{Y_t, X_t'\}_{t=1}^n$ is an i.i.d. random sample.

Assumption 4.2 [Linearity]:

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

for some unknown $K \times 1$ parameter β^o and some unobservable random variable ε_t.

Assumption 4.3 [Correct Model Specification]: $E(\varepsilon_t|X_t) = 0$ a.s. with $E(\varepsilon_t^2) = \sigma^2 < \infty$.

Assumption 4.4 [Nonsingularity]: The $K \times K$ matrix

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

is nonsingular and finite.

Assumption 4.5: The $K \times K$ matrix $V \equiv \text{var}(X_t\varepsilon_t) = E(X_t X_t'\varepsilon_t^2)$ is finite and positive definite (p.d.).

Remarks:

The i.i.d. observations assumption in Assumption 4.1 implies that the asymptotic theory developed in this chapter will be applicable to cross-sectional data, but not time series data. The observations of the later are usually correlated and will be considered in Chapter 5. Put $Z_t = (Y_t, X_t')'$. Then I.I.D. implies that Z_t and Z_s are independent when $t \neq s$, and the Z_t have the same distribution for all t. The identical distribution means that the observations are generated from the same data generating process, and independence means that different observations contain new information about the data generating process.

Assumptions 4.1 and 4.3 imply the strict exogeneity condition (Assumption 3.2) holds, because we have

$$
\begin{aligned}
E(\varepsilon_t|\mathbf{X}) &= E(\varepsilon_t|X_1, X_2, ...X_t, ...X_n) \\
&= E(\varepsilon_t|X_t) \\
&= 0 \ a.s.
\end{aligned}
$$

As a most important feature of Assumptions 4.1–4.5 together, we allow for conditional heteroskedasticity (i.e., $\text{var}(\varepsilon_t|X_t) \neq \sigma^2$ a.s.), and do not assume normality for the conditional distribution of $\varepsilon_t|X_t$. It is possible that $\text{var}(\varepsilon_t|X_t)$ may be correlated with X_t. For example, the variation of the output of a firm may depend on the size of the firm, and the variation of a household may depend on its income level. In economics and finance, conditional heteroskedasticity is more likely to occur in cross-sectional observations than in time series observations, and for time series observations, conditional heteroskedasticity is more likely to occur for high-frequency data than low-frequency data. In this chapter, we will consider the effect of conditional heteroskedasticity in cross-section observations. The effect of conditional heteroskedasticity in time series observations will be considered in Chapter 5.

On the other hand, relaxation of the normality assumption is more realistic for economic and financial data. For example, it has been well documented (Mandelbrot 1963, Fama 1965, Kon 1984) that returns on financial assets are not normally distributed. However, the I.I.D. assumption implies that $\text{cov}(\varepsilon_t, \varepsilon_s) = 0$ for all $t \neq s$. That is, there exists no serial correlation in the regression disturbance.

Among other things, Assumption 4.4 implies $E(X_{jt}^2) < \infty$ for $0 \leq j \leq k$. By the SLLN for i.i.d. random samples, we have

$$
\begin{aligned}
\frac{\mathbf{X}'\mathbf{X}}{n} &= \frac{1}{n}\sum_{t=1}^{n} X_t X_t' \\
&\to^{a.s.} E(X_t X_t') = Q
\end{aligned}
$$

as $n \to \infty$. Hence, when n is large, the matrix $\mathbf{X}'\mathbf{X}$ behaves approximately like nQ, whose minimum eigenvalue $\lambda_{\min}(nQ) = n\lambda_{\min}(Q) \to \infty$ at the rate of n. Thus, Assumption 4.4 implies Assumption 3.3.

When $X_{0t} = 1$, Assumption 4.5 implies $E(\varepsilon_t^2) < \infty$. If $E(\varepsilon_t^2|X_t) = \sigma^2 < \infty$ a.s., i.e., there exists conditional homoskedasticity, then Assumption 4.5 can be ensured by Assumption 4.4. More generally, there exists conditional heteroskedasticity, the moment condition in Assumption 4.5 can be ensured by the moment conditions that $E(\varepsilon_t^4) < \infty$ and $E(X_{jt}^4) < \infty$ for $0 \leq j \leq k$, because by repeatedly using the Cauchy-Schwarz inequality twice, we have

$$\begin{aligned}|E(\varepsilon_t^2 X_{jt} X_{lt})| &\leq [E(\varepsilon_t^4)]^{1/2}[E(X_{jt}^2 X_{lt}^2)]^{1/2} \\ &\leq [E(\varepsilon_t^4)]^{1/2}[E(X_{jt}^4)E(X_{lt}^4)]^{1/4}\end{aligned}$$

where $0 \leq j, l \leq k$ and $1 \leq t \leq n$.

We now address the following questions:

- Consistency of OLS?
- Asymptotic normality?
- Asymptotic efficiency?
- Confidence interval estimation?
- Hypothesis testing?

In particular, we are interested in knowing whether the statistical properties of OLS $\hat{\beta}$ and related test statistics derived under the classical linear regression setup are still valid under the current setup, at least when n is large.

4.2 Consistency of OLS

Suppose we have a random sample $\{Y_t, X_t'\}_{t=1}^n$. Recall that the OLS estimator:

$$\begin{aligned}\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y \\ &= \left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}\frac{\mathbf{X}'Y}{n} \\ &= \hat{Q}^{-1}n^{-1}\sum_{t=1}^n X_t Y_t,\end{aligned}$$

where

$$\hat{Q} = n^{-1}\sum_{t=1}^n X_t X_t'.$$

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

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

We will consider the consistency of $\hat{\beta}$ directly.

Theorem [Consistency of OLS] *Under Assumptions 4.1-4.4, as* $n \to \infty$,

$$\hat{\beta} \xrightarrow{p} \beta^o \text{ or } \hat{\beta} - \beta^o = o_P(1).$$

Proof: Let $C > 0$ be some bounded constant. Also, recall $X_t = (X_{0t}, X_{1t}, ..., X_{kt})'$. First, the moment condition holds: for all $0 \leq j \leq k$,

$$\begin{aligned} E|X_{jt}\varepsilon_t| &\leq (EX_{jt}^2)^{\frac{1}{2}}(E\varepsilon_t^2)^{\frac{1}{2}} \text{ by the Cauchy-Schwarz inequality} \\ &\leq C^{\frac{1}{2}}C^{\frac{1}{2}} \\ &\leq C \end{aligned}$$

where $E(X_{jt}^2) \leq C$ by Assumption 4.4, and $E(\varepsilon_t^2) \leq C$ by Assumption 4.3. It follows by WLLN (with $Z_t = X_t\varepsilon_t$) that

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

where

$$\begin{aligned} E(X_t\varepsilon_t) &= E[E(X_t\varepsilon_t|X_t)] \text{ by the law of iterated expectations} \\ &= E[X_tE(\varepsilon_t|X_t)] \\ &= E(X_t \cdot 0) \\ &= 0. \end{aligned}$$

Applying WLLN again (with $Z_t = X_tX_t'$) and noting that

$$E|X_{jt}X_{lt}| \leq [E(X_{jt}^2)E(X_{lt}^2)]^{\frac{1}{2}} \leq C$$

by the Cauchy-Schwarz inequality for all pairs (j, l), where $0 \leq j, l \leq k$, we have

$$\hat{Q} \xrightarrow{p} E(X_tX_t') = Q.$$

Hence, we have $\hat{Q}^{-1} \to^p Q^{-1}$ by continuity. It follows that

$$\begin{aligned} \hat{\beta} - \beta^o &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon \\ &= \hat{Q}^{-1}n^{-1}\sum_{t=1}^{n} X_t\varepsilon_t \\ \xrightarrow{p} Q^{-1} \cdot 0 &= 0. \end{aligned}$$

This completes the proof. ■

4.3 Asymptotic Normality of OLS

Next, we derive the asymptotic distribution of $\hat{\beta}$. We first provide a multivariate CLT for I.I.D. random samples.

Lemma [Multivariate Central Limit Theorem (CLT) for I.I.D. Random Samples]: *Suppose $\{Z_t\}$ is a sequence of i.i.d. random vectors with $E(Z_t) = 0$ and $var(Z_t) = E(Z_t Z_t') = V$ is finite and positive definite. Define*

$$\bar{Z}_n = n^{-1} \sum_{t=1}^{n} Z_t.$$

Then as $n \to \infty$,

$$\sqrt{n}\bar{Z}_n \xrightarrow{d} N(0, V)$$

or

$$V^{-\frac{1}{2}} \sqrt{n}\bar{Z}_n \xrightarrow{d} N(0, I).$$

Question: What is the variance-covariance matrix of $\sqrt{n}\bar{Z}_n$?
Answer: Noting that $E(Z_t) = 0$, we have

$$\begin{aligned}
\text{var}(\sqrt{n}\bar{Z}_n) &= \text{var}\left(n^{-\frac{1}{2}} \sum_{t=1}^{n} Z_t\right) \\
&= E\left[\left(n^{-\frac{1}{2}} \sum_{t=1}^{n} Z_t\right)\left(n^{-\frac{1}{2}} \sum_{s=1}^{n} Z_s\right)'\right] \\
&= n^{-1} \sum_{t=1}^{n} \sum_{s=1}^{n} E(Z_t Z_s') \\
&= n^{-1} \sum_{t=1}^{n} E(Z_t Z_t') \quad \text{(because } Z_t \text{ and } Z_s \text{ are independent for } t \neq s) \\
&= E(Z_t Z_t') \\
&= V.
\end{aligned}$$

In other words, the variance of $\sqrt{n}\bar{Z}_n$ is identical to the variance of each individual random vector Z_t.

Theorem [Asymptotic Normality of OLS] *Under Assumptions 4.1-4.5, we have*

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

as $n \to \infty$, *where* $V \equiv \text{var}(\varepsilon_t|X_t) = E(X_t X_t' \varepsilon_t^2)$.

Proof: Recall that

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

First, we consider the second term

$$n^{-\frac{1}{2}} \sum_{t=1}^{n} X_t \varepsilon_t.$$

Noting that $E(X_t \varepsilon_t) = 0$ by Assumption 4.3, and $\text{var}(X_t \varepsilon_t) = E(X_t X_t' \varepsilon_t^2) = V$, which is finite and p.d. by Assumption 4.5. Then, by the CLT for i.i.d. random sequences $\{Z_t = X_t \varepsilon_t\}$, we have

$$\begin{aligned} n^{-\frac{1}{2}} \sum_{t=1}^{n} X_t \varepsilon_t &= \sqrt{n} \left(n^{-1} \sum_{t=1}^{n} X_t \varepsilon_t \right) \\ &= \sqrt{n} \bar{Z}_n \\ &\xrightarrow{d} Z \sim N(0, V). \end{aligned}$$

On the other hand, as shown earlier, we have

$$\hat{Q} \xrightarrow{p} Q,$$

and so

$$\hat{Q}^{-1} \xrightarrow{p} Q^{-1}$$

given that Q is nonsingular so that the inverse function is continuous and well defined. It follows by the Slutsky Theorem that

$$\begin{aligned} \sqrt{n}(\hat{\beta} - \beta^o) &= \hat{Q}^{-1} n^{-\frac{1}{2}} \sum_{t=1}^{n} X_t \varepsilon_t \\ &\xrightarrow{d} Q^{-1} Z \sim N(0, Q^{-1} V Q^{-1}). \end{aligned}$$

This completes the proof. ■

Remarks:

The theorem implies that the asymptotic mean of $\sqrt{n}(\hat{\beta} - \beta^o)$ is equal to 0. That is, the mean of $\sqrt{n}(\hat{\beta} - \beta^o)$ is approximately 0 when n is large.

It also implies that the asymptotic variance of $\sqrt{n}(\hat{\beta} - \beta^o)$ is $Q^{-1} V Q^{-1}$. That is, the variance of $\sqrt{n}(\hat{\beta} - \beta^o)$ is approximately $Q^{-1} V Q^{-1}$. Because the asymptotic variance is a different concept from the variance of $\sqrt{n}(\hat{\beta}-\beta^o)$, we denote the asymptotic variance of $\sqrt{n}(\hat{\beta}-\beta^o)$ as follows: $\text{avar}(\sqrt{n}\hat{\beta}) = Q^{-1}VQ^{-1}$.

We now consider a special case under which we can simplfy the expression of $\text{avar}(\sqrt{n}\hat{\beta})$.

Special Case : Conditional Homoskedasticity

Assumption 4.6: $E(\varepsilon_t^2|X_t) = \sigma^2$ a.s.

Theorem: *Suppose Assumptions 4.1–4.6 hold. Then as* $n \to \infty$,

$$\sqrt{n}(\hat{\beta}-\beta^o) \xrightarrow{d} N(0, \sigma^2 Q^{-1}).$$

Proof: Under Assumption 4.6, we can simplify

$$\begin{aligned} V &= E(X_t X_t' \varepsilon_t^2) \\ &= E[E(X_t X_t' \varepsilon_t^2 | X_t)] \text{ by the law of iterated expectations} \\ &= E[X_t X_t' E(\varepsilon_t^2 | X_t)] \\ &= \sigma^2 E(X_t X_t') \\ &= \sigma^2 Q. \end{aligned}$$

The results follow immediately because

$$Q^{-1}VQ^{-1} = Q^{-1}\sigma^2 Q Q^{-1} = \sigma^2 Q^{-1}.$$

Remarks:

Under conditional homoskedasticity, the asymptotic variance of $\sqrt{n}(\hat{\beta}-\beta^o)$ is

$$\text{avar}(\sqrt{n}\hat{\beta}) = \sigma^2 Q^{-1}.$$

Question: Is the OLS estimator $\hat{\beta}$ the BLUE estimator asymptotically (i.e., when $n \to \infty$)?

4.4 Asymptotic Variance Estimator

To construct confidence interval estimators or hypothesis tests, we need to estimate the asymptotic variance of $\sqrt{n}(\hat{\beta}-\beta^o)$, $\text{avar}(\sqrt{n}\hat{\beta})$. Because the expression of $\text{avar}(\sqrt{n}\hat{\beta})$ differs under conditional homoskedasticity and conditional heteroskedasticity respectively, we consider the estimator for $\text{avar}(\sqrt{n}\hat{\beta})$ under these two cases separately.

Case I: Conditional Homoskedasticity

In this case, the asymptotic variance of $\sqrt{n}(\hat{\beta} - \beta^o)$ is

$$\text{avar}(\sqrt{n}\hat{\beta}) = Q^{-1}VQ^{-1} = \sigma^2 Q^{-1}.$$

Question: How to estimate Q?

Lemma: *Suppose Assumptions 4.1, 4.2 and 4.4 hold. Then*

$$\hat{Q} = n^{-1}\sum_{t=1}^{n} X_t X_t' \xrightarrow{p} Q.$$

Question: How to estimate σ^2?

Recalling that $\sigma^2 = E(\varepsilon_t^2)$, we use the sample residual variance estimator

$$\begin{aligned} s^2 &= e'e/(n-K) \\ &= \frac{1}{n-K}\sum_{t=1}^{n} e_t^2 \\ &= \frac{1}{n-K}\sum_{t=1}^{n}(Y_t - X_t'\hat{\beta})^2. \end{aligned}$$

Theorem [Consistent Estimator for σ^2]: *Under Assumptions 4.1-4.4,*

$$s^2 \xrightarrow{p} \sigma^2.$$

Proof: Given that $s^2 = e'e/(n-K)$ and

$$\begin{aligned} e_t &= Y_t - X_t'\hat{\beta} \\ &= \varepsilon_t + X_t'\beta^o - X_t'\hat{\beta} \\ &= \varepsilon_t - X_t'(\hat{\beta} - \beta^o), \end{aligned}$$

we have

$$\begin{aligned} s^2 &= \frac{1}{n-K}\sum_{t=1}^{n}[\varepsilon_t - X_t'(\hat{\beta} - \beta^o)]^2 \\ &= \frac{n}{n-K}\left(n^{-1}\sum_{t=1}^{n}\varepsilon_t^2\right) \\ &\quad +(\hat{\beta} - \beta^o)'\left[(n-K)^{-1}\sum_{t=1}^{n}X_t X_t'\right](\hat{\beta} - \beta^o) \\ &\quad -2(\hat{\beta} - \beta^o)'(n-K)^{-1}\sum_{t=1}^{n}X_t\varepsilon_t \\ &\quad \xrightarrow{p} 1 \cdot \sigma^2 + 0 \cdot Q \cdot 0 - 2 \cdot 0 \cdot 0 \\ &= \sigma^2 \end{aligned}$$

given that K is a fixed number (i.e., K does not grow with the sample size n), where we have made use of the WLLN in three places respectively.

We can then consistently estimate $\sigma^2 Q^{-1}$ by $s^2\hat{Q}^{-1}$.

Theorem [Asymptotic Variance Estimator of $\sqrt{n}(\hat{\beta}-\beta^o)$] *Under Assumptions 4.1-4.4, we have*

$$s^2\hat{Q}^{-1} \xrightarrow{p} \sigma^2 Q^{-1}.$$

Remark:

The asymptotic variance estimator of $\sqrt{n}(\hat{\beta}-\beta^o)$ is

$$s^2\hat{Q}^{-1} = s^2(\mathbf{X}'\mathbf{X}/n)^{-1}.$$

This is equivalent to saying that the variance estimator of $\hat{\beta}-\beta^o$ is approximately equal to

$$s^2\hat{Q}^{-1}/n = s^2(\mathbf{X}'\mathbf{X})^{-1}$$

when for a large n. Thus, when $n \to \infty$ and there exists conditional homoskedasticity, the variance estimator of $\hat{\beta}-\beta^o$ coincides with the form of the variance estimator for $\hat{\beta}-\beta^o$ in the classical regression case. Because of this, as will be seen below, the conventional t-test and F-test are still valid for large samples under conditional homoskedasticity.

CASE II: Conditional Heteroskedesticity

In this case,

$$\text{avar}(\sqrt{n}\hat{\beta}) = Q^{-1}VQ^{-1},$$

which cannot be simplified.

Question: We can still estimate $\hat{Q}$ to estimate Q. How to estimate $V = E(X_t X_t' \varepsilon_t^2)$?

We can use its sample analog

$$\hat{V} = n^{-1}\sum_{t=1}^{n} X_t X_t' e_t^2 = \frac{\mathbf{X}'\mathrm{D}(e)\mathrm{D}(e)'\mathbf{X}}{n},$$

where

$$\mathrm{D}(e) = diag(e_1, e_2, ..., e_n)$$

is an $n \times n$ diagonal matrix with diagonal elements equal to e_t for $t = 1, ..., n$. To ensure consistency of $\hat{V}$ to V, we impose the following additional moment conditions.

Assumption 4.7: (i) $E(X_{jt}^4) < \infty$ for all $0 \leq j \leq k$; and (ii) $E(\varepsilon_t^4) < \infty$.

Lemma: *Suppose Assumptions 4.1–4.5 and 4.7 hold. Then*

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

Proof: Because $e_t = \varepsilon_t - (\hat{\beta} - \beta^o)' X_t$, we have

$$\begin{aligned}
\hat{V} &= n^{-1} \sum_{t=1}^{n} X_t X_t' \varepsilon_t^2 \\
&\quad + n^{-1} \sum_{t=1}^{n} X_t X_t' [(\hat{\beta} - \beta^o)' X_t X_t' (\hat{\beta} - \beta^o)] \\
&\quad - 2n^{-1} \sum_{t=1}^{n} X_t X_t' [\varepsilon_t X_t' (\hat{\beta} - \beta^o)] \\
&\to^p \quad V + 0 - 2 \cdot 0,
\end{aligned}$$

where for the first term, we have

$$n^{-1} \sum_{t=1}^{n} X_t X_t' \varepsilon_t^2 \xrightarrow{p} E(X_t X_t' \varepsilon_t^2) = V$$

by the WLLN and Assumption 4.7, which implies

$$E|X_{it} X_{jt} \varepsilon_t^2| \leq [E(X_{it}^2 X_{jt}^2) E(\varepsilon_t^4)]^{\frac{1}{2}}.$$

For the second term, we have

$$\begin{aligned}
& n^{-1} \sum_{t=1}^{n} X_{it} X_{jt} (\hat{\beta} - \beta^o)' X_t X_t' (\hat{\beta} - \beta^o) \\
=& \sum_{l=0}^{k} \sum_{m=0}^{k} (\hat{\beta}_l - \beta_l^o)(\hat{\beta}_m - \beta_m^o) \left(n^{-1} \sum_{t=1}^{n} X_{it} X_{jt} X_{lt} X_{mt} \right) \\
\to& \; ^p 0
\end{aligned}$$

given $\hat{\beta} - \beta^o \to^p 0$, and

$$n^{-1} \sum_{t=1}^{n} X_{it} X_{jt} X_{lt} X_{mt} \to^p E\left(X_{it} X_{jt} X_{lt} X_{mt}\right) = O(1)$$

by the WLLN and Assumption 4.7.

Similarly, for the last term, we have

$$
\begin{aligned}
& n^{-1}\sum_{t=1}^{n} X_{it}X_{jt}\varepsilon_t X_t'(\hat\beta-\beta^o) \\
= & \sum_{l=0}^{k}(\hat\beta_l-\beta_l^o)\left(n^{-1}\sum_{t=1}^{n} X_{it}X_{jt}X_{lt}\varepsilon_t\right) \\
\to^p & \; 0
\end{aligned}
$$

given $\hat\beta-\beta^o \to^p 0$, and

$$
n^{-1}\sum_{t=1}^{n} X_{it}X_{jt}X_{lt}\varepsilon_t \to^p E\left(X_{it}X_{jt}X_{lt}\varepsilon_t\right)=0
$$

by the WLLN and Assumption 4.7. This completes the proof.

We now construct a consistent estimator for avar$(\sqrt{n}\hat\beta)$ under conditional heteroskedasticity.

Theorem [Asymptotic variance estimator for $\sqrt{n}(\hat\beta-\beta^o)$]: *Under Assumptions 4.1–4.5 and 4.7, we have*

$$
\hat Q^{-1}\hat V\hat Q^{-1} \xrightarrow{p} Q^{-1}VQ^{-1}.
$$

Remark:

This is the so-called White's (1980) heteroskedasticity-consistent variance-covariance matrix of the estimator $\sqrt{n}(\hat\beta-\beta^o)$. It follows that when there exists conditional heteroskedasticity, the estimator for the variance of $\hat\beta-\beta^o$ is

$$
\begin{aligned}
& (\mathbf{X}'\mathbf{X}/n)^{-1}\hat V(\mathbf{X}'\mathbf{X}/n)^{-1}/n \\
= & (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathrm{D}(e)\mathrm{D}(e)'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1},
\end{aligned}
$$

which differs from the estimator $s^2(\mathbf{X}'\mathbf{X})^{-1}$ in the case of conditional homoskedasticity.

Question: What happens if we use $s^2\hat Q^{-1}$ as an estimator for the avar$[\sqrt{n}(\hat\beta-\beta^o)]$ while there exists conditional heteroskedasticity?

Observe that

$$
\begin{aligned}
V & \equiv E(X_tX_t'\varepsilon_t^2) \\
& = \sigma^2 Q+\mathrm{cov}(X_tX_t',\varepsilon_t^2) \\
& = \sigma^2 Q+\ \mathrm{cov}[X_tX_t',\sigma^2(X_t)],
\end{aligned}
$$

where $\sigma^2 = E(\varepsilon_t^2)$, $\sigma^2(X_t) = E(\varepsilon_t^2|X_t)$, and the last equality follows by the LIE. Thus, if $\sigma^2(X_t)$ is positively correlated with X_tX_t', $\sigma^2 Q$ will underestimate the true variance-covariance $E(X_tX_t'\varepsilon_t^2)$ in the sense that $V - \sigma^2 Q$ is a positive definite matrix. Consequently, the standard t-test and F-test will overreject the correct null hypothesis at any given significance level. There will exist substantial Type I errors.

Question: What happens if one use the asymptotic variance estimator $\hat{Q}^{-1}\hat{V}\hat{Q}^{-1}$ but there exists conditional homoskedasticity?

The asymptotic variance estimator is asymptotically valid, but it will not perform as well as the estimator $s^2\hat{Q}^{-1}$ in finite samples, because the latter exploits the information of conditonal homoskedasticity.

4.5 Hypothesis Testing

Question: How to construct a test statistic for the null hypothesis

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

where R is a $J \times K$ constant matrix, and r is a $J \times 1$ constant vector?

We first consider

$$R\hat{\beta} - r = R(\hat{\beta} - \beta^o) + R\beta^o - r.$$

It follows that under $\mathbf{H}_0 : R\beta^o = r$, we have

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

The test procedures will differ depending on whether there exists conditional heteroskedasticity. We first consider the case of conditional homoskedasticity.

Case I: Conditional Homoskedasticity

Under conditional homoskedasticity, we have $V = \sigma^2 Q$ and so

$$\sqrt{n}(R\hat{\beta} - r) \xrightarrow{d} N(0, \sigma^2 RQ^{-1}R')$$

when $\mathbf{H}_0$ holds.

When $J = 1$, we can use the conventional t-test statistic for large sample inference.

Theorem [t-test]: *Suppose Assumptions 4.1-4.4 and 4.6 hold. Then under* $\mathbf{H}_0$ with $J = 1$,

$$T = \frac{R\hat{\beta} - r}{\sqrt{s^2 R(\mathbf{X}'\mathbf{X})^{-1}R'}} \xrightarrow{d} N(0, 1)$$

as $n \to \infty$.

Proof: Give $R\sqrt{n}(\hat{\beta} - \beta^o) \xrightarrow{d} N(0, \sigma^2 RQ^{-1}R')$, $R\beta^o = r$ under $\mathbf{H}_0$, and $J = 1$, we have

$$\frac{\sqrt{n}(R\hat{\beta} - r)}{\sqrt{\sigma^2 RQ^{-1}R'}} = \frac{R\sqrt{n}(\hat{\beta} - \beta^o)}{\sqrt{\sigma^2 RQ^{-1}R'}} \to^d N(0,1).$$

By the Slutsky theorem and $\hat{Q} = \mathbf{X}'\mathbf{X}/n$, we obtain

$$\frac{\sqrt{n}(R\hat{\beta} - r)}{\sqrt{s^2 R\hat{Q}^{-1}R'}} \to^d N(0,1).$$

This ratio is the conventional t-test statistic we examined in Chapter 3, namely:

$$\frac{\sqrt{n}(R\hat{\beta} - r)}{\sqrt{s^2 R\hat{Q}^{-1}R'}} = \frac{R\hat{\beta} - r}{\sqrt{s^2 R(\mathbf{X}'\mathbf{X})^{-1}R'}} = T.$$

For $J > 1$, we use a quadratic form test statistic.

Theorem [Asymptotic χ^2 Test] *Suppose Assumptions 4.1–4.4 and 4.6 hold. Then under* $\mathbf{H}_0$,

$$\begin{aligned} J \cdot F &\equiv (R\hat{\beta} - r)' \left[s^2 R(\mathbf{X}'\mathbf{X})^{-1}R'\right]^{-1} (R\hat{\beta} - r) \\ &\xrightarrow{d} \chi^2_J \end{aligned}$$

as $n \to \infty$.

Proof: Under $\mathbf{H}_0$, the quadratic form

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

Also, $s^2\hat{Q}^{-1} \xrightarrow{p} \sigma^2 Q^{-1}$, so we have by the Slutsky theorem

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

or equivalently

$$J \cdot \frac{(R\hat{\beta} - r)'[R(\mathbf{X}'\mathbf{X})^{-1}R']^{-1}(R\hat{\beta} - r)/J}{s^2} = J \cdot F \xrightarrow{d} \chi^2_J,$$

namely

$$J \cdot F \xrightarrow{d} \chi^2_J.$$

Remarks:

When $\{\varepsilon_t\}$ is not i.i.d.$N(0, \sigma^2)$ conditional on X_t, we cannot use the F distribution, but we can still compute the F-statistic and the appropriate test statistic is J times the F-statistic, which is asymptotically χ^2_J. That is,

$$J \cdot F = \frac{(\tilde{e}'\tilde{e} - e'e)}{e'e/(n-K)} \xrightarrow{d} \chi^2_J.$$

Because $J \cdot F_{J,n-K}$ approaches χ^2_J as $n \to \infty$, we may interpret the above theorem in the following way: the classical results for the F-test are still approximately valid under conditional homoskedasiticity when n is large.

When the null hypothesis is that all slope coefficients except the intercept are jointly zero, we can use a test statistic based on R^2.

A Special Case: Testing for Joint Significance of All Economic Variables

Theorem [$(n-K)R^2$ **Test**]: *Suppose Assumption 4.1-4.6 hold, and we are interested in testing the null hypothesis that*

$$\mathbf{H}_0 : \beta_1^o = \beta_2^o = \cdots = \beta_k^o = 0,$$

where the β *are the regression coefficients from*

$$Y_t = \beta_0^o + \beta_1^o X_{1t} + \cdots + \beta_k^o X_{kt} + \varepsilon_t.$$

Let R^2 *be the coefficient of determination from the unrestricted regression model*

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

Then under $\mathbf{H}_0$,

$$(n-K)R^2 \xrightarrow{d} \chi^2_k,$$

where $K = k + 1$.

Proof: First, recall that in this special case we have

$$\begin{aligned} F &= \frac{R^2/k}{(1-R^2)/(n-k-1)} \\ &= \frac{R^2/k}{(1-R^2)/(n-K)}. \end{aligned}$$

By the above theorem and noting $J = k$, we have

$$k \cdot F = \frac{(n-K)R^2}{1-R^2} \xrightarrow{d} \chi^2_k$$

under $\mathbf{H}_0$. This implies that $k \cdot F$ is bounded in probability; that is,

$$\frac{(n-K)R^2}{1-R^2} = O_P(1).$$

Consequently, given that k is a fixed integer,

$$\frac{R^2}{1-R^2} = O_P(n^{-1}) = o_P(1)$$

or

$$R^2 \xrightarrow{p} 0.$$

Therefore, $1 - R^2 \xrightarrow{p} 1$. By the Slutsky theorem, we have

$$\begin{aligned} (n-K)R^2 = &\ k \cdot \frac{(n-K)R^2/k}{1-R^2}(1-R^2) \\ = &\ (k \cdot F)(1-R^2) \\ &\xrightarrow{d} \chi^2_k, \end{aligned}$$

or asymptotically equivalently,

$$(n-K)R^2 \xrightarrow{d} \chi^2_k.$$

This completes the proof. ■

Question: Do we have $nR^2 \xrightarrow{d} \chi^2_k$?

Yes, we have

$$nR^2 = \frac{n}{n-K}(n-K)R^2 \text{ and } \frac{n}{n-K} \to 1.$$

Case II: Conditional Heteroskedasticity

Recall that under $\mathbf{H}_0$,

$$\begin{aligned} \sqrt{n}(R\hat{\beta} - r) &= R\sqrt{n}(\hat{\beta} - \beta^o) + \sqrt{n}(R\beta^o - r) \\ &= R\sqrt{n}(\hat{\beta} - \beta^o) \\ &\xrightarrow{d} N(0, RQ^{-1}VQ^{-1}R'), \end{aligned}$$

where

$$V = E(X_t X_t' \varepsilon_t^2).$$

Therefore, when $J = 1$,we have

$$\frac{\sqrt{n}(R\hat{\beta} - r)}{\sqrt{RQ^{-1}VQ^{-1}R'}} \to^d N(0,1) \text{ as } n \to \infty.$$

Given $\hat{Q} \to^p Q$ and $\hat{V} \xrightarrow{p} V$, where $\hat{V} = \mathbf{X}'\mathrm{D}(e)\mathrm{D}(e)'\mathbf{X}/n$, and the Slutsky theorem, we can define a robust t-test statistic

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

when $\mathbf{H}_0$ holds. By robustness, we mean that W is valid no matter whether there exists conditional heteroskedasticity.

Theorem [Robust t-Test Under Conditional Heteroskedasticity] *Suppose Assumptions 4.1–4.5 and 4.7 hold. Then under* $\mathbf{H}_0$ with $J = 1$, *as* $n \to \infty$, the robust t-test statistic

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

When $J > 1$, we have the quadratic form

$$\begin{aligned} W &= \sqrt{n}(R\hat{\beta} - r)'[RQ^{-1}VQ^{-1}R']^{-1}\sqrt{n}(R\hat{\beta} - r) \\ &\xrightarrow{d} \chi^2_J \end{aligned}$$

under $\mathbf{H}_0$. Given $\hat{Q} \xrightarrow{p} Q$ and $\hat{V} \xrightarrow{p} V$, the robust Wald test statistic

$$\begin{aligned} W &= \sqrt{n}(R\hat{\beta} - r)'[R\hat{Q}^{-1}\hat{V}\hat{Q}^{-1}R']^{-1}\sqrt{n}(R\hat{\beta} - r) \\ &\xrightarrow{d} \chi^2_J \end{aligned}$$

by the Slutsky theorem.

We can write W equivalently as follows:

$$W = (R\hat{\beta} - r)'[R(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathrm{D}(e)\mathrm{D}(e)'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}R']^{-1}(R\hat{\beta} - r),$$

where we have used the fact that

$$\begin{aligned} \hat{V} &= \frac{1}{n}\sum_{t=1}^{n} X_t e_t e_t X_t' \\ &= \frac{\mathbf{X}'\mathrm{D}(e)\mathrm{D}(e)'\mathbf{X}}{n}, \end{aligned}$$

where D(e)= $diag(e_1, e_2, ..., e_n)$.

Theorem [Robust Wald Test Under Conditional Heteroskedasticity] *Suppose Assumptions 4.1–4.5 and 4.7 hold. Then under* $\mathbf{H}_0$*, as* $n \to \infty$*,*

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

Remarks:

Under conditional heteroskedasticity, the test statistics $J \cdot F$ and $(n-K)R^2$ cannot be used.

Question: What happens if there exists conditional heteroskedasticity but $J \cdot F$ or $(n-K)R^2$ is used.

There will exist Type I errors because $J \cdot F$ or $(n-K)R^2$ will be no longer asymptotically χ^2-distributed under $\mathbf{H}_0$.

Although the general form of the Wald test statistic developed here can be used no matter whether there exists conditional homoskedasticity, this general form of test statistic may perform poorly in small samples. Thus, if one has information that the error term is conditionally homoskedastic, one should use the test statistics derived under conditional homoskedasticity, which will perform better in small sample sizes. Because of this reason, it is important to test whether conditional homoskedasticity holds.

4.6 Testing Conditional Homoskedasticity

We now introduce a method to test conditional heteroskedasticity.

Question: How to test conditional homoskedasticity for $\{\varepsilon_t\}$ in a linear regression model?

There have been many tests for conditional homoskedasticity. Here, we introduce a popular one due to White (1980).

White's (1980) test

The null hypothesis

$$\mathbf{H}_0 : E(\varepsilon_t^2 | X_t) = \sigma^2,$$

where ε_t is the regression error in the linear regression model

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

First, suppose ε_t were observed, and we consider the auxiliary regression

$$\begin{aligned} \varepsilon_t^2 &= \gamma_0 + \sum_{j=1}^{k} \gamma_j X_{jt} + \sum_{1 \le j \le l \le k} \gamma_{jl} X_{jt} X_{lt} + v_t \\ &= \gamma' \text{vech}(X_t X_t') + v_t \\ &= \gamma' U_t + v_t, \end{aligned}$$

where $\text{vech}(X_t X_t')$ is an operator stacks all lower triangular elements of matrix $X_t X_t'$ into a $\frac{K(K+1)}{2} \times 1$ column vector. For example, when $X_t = (1, X_{1t}, X_{2t})'$, we have

$$\text{vech}(X_t X_t') = (1, X_{1t}, X_{2t}, X_{1t}^2, X_{2t}^2, X_{1t}X_{2t})'.$$

For the auxiliary regression, there is a total of $\frac{K(K+1)}{2}$ regressors in U_t. This is essentially regressing ε_t^2 on the intercept, X_t, and the quadratic terms and cross-product terms of X_t. Under $\mathbf{H}_0$, all coefficients except the intercept are jointly zero. Any nonzero coefficients will indicate the existence of conditional heteroskedasticity. Thus, we can test $\mathbf{H}_0$ by checking whether all coefficients except the intercept are jointly zero. Assuming that $E(\varepsilon_t^4|X_t) = \mu_4$ (which implies $E(v_t^2|X_t) = \sigma_v^2$ under $\mathbf{H}_0$), then we can run an OLS regression and construct a R^2-based test statistic. Under $\mathbf{H}_0$, we can obtain

$$(n - J - 1)\tilde{R}^2 \xrightarrow{d} \chi_J^2,$$

where $J = \frac{K(K+1)}{2} - 1$ is the number of the regressors except the intercept.
Unfortunately, ε_t is not observable. However, we can replace ε_t with $e_t = Y_t - X_t'\hat{\beta}$, and run the following feasible auxiliary regression

$$\begin{aligned} e_t^2 &= \gamma_0 + \sum_{j=1}^{k} \gamma_j X_{jt} + \sum_{1 \le j \le l \le k} \gamma_{jl} X_{jt} X_{lt} + \tilde{v}_t \\ &= \gamma' \text{vech}(X_t X_t') + \tilde{v}_t, \end{aligned}$$

the resulting test statistic

$$(n - J - 1)R^2 \xrightarrow{d} \chi_J^2.$$

It can be shown that the replacement of ε_t^2 by e_t^2 has no impact on the asymptotic χ_J^2 distribution of $(n - J - 1)R^2$. The proof, however, is rather tedious. For the details of the proof, see White (1980). Below, we provide some intuition.

Question: Why does the use of e_t^2 in place of ε_t^2 have no impact on the asymptotic distribution of $(n-J-1)R^2$?

To explain this, we put $U_t = \text{vech}(X_t X_t')$. Then the infeasible auxiliary regression is

$$\varepsilon_t^2 = U_t'\gamma + v_t.$$

We have $\sqrt{n}(\tilde{\gamma} - \gamma^0) \to^d N(0, \sigma_v^2 Q_{\text{uu}}^{-1})$, where $Q_{\text{uu}} = E(U_t U_t')$, and under $\mathbf{H}_0 : R\gamma = 0$, where R is a $J \times J$ diagonal matrix with the first diaagonal element being 0 and other diagonal elements being 1, we have

$$\sqrt{n}R\tilde{\gamma} \to^d N(0, \sigma_v^2 R Q_{\text{uu}}^{-1} R'),$$

where $\tilde{\gamma}$ is the OLS estimator and $\sigma_v^2 = E(v_t^2)$. This implies $R\tilde{\gamma} = O_P(n^{-1/2})$, which vanishes to zero in probability at rate $n^{-1/2}$. It is this term that yields the asymptotic χ_J^2 distribution for $(n-J-1)\tilde{R}^2$, which is asymptotically equiavalent to the test statistic

$$\sqrt{n}\tilde{\gamma}' R' [s_v^2 R \hat{Q}_{\text{uu}}^{-1} R']^{-1} \sqrt{n} R\tilde{\gamma}.$$

Now suppose we replace ε_t^2 with e_t^2, and consider the auxliary regression

$$e_t^2 = U_t'\gamma^0 + \tilde{v}_t.$$

Denote the OLS estimator by $\hat{\gamma}$. We decompose

$$\begin{aligned} e_t^2 &= \left[\varepsilon_t - X_t'(\hat{\beta} - \beta^o)\right]^2 \\ &= \varepsilon_t^2 + (\hat{\beta} - \beta^o)' X_t X_t' (\hat{\beta} - \beta^o) - 2(\hat{\beta} - \beta^o)' X_t \varepsilon_t \\ &= \gamma' U_t + \tilde{v}_t. \end{aligned}$$

Thus, $\hat{\gamma}$ can be written as follows:

$$\hat{\gamma} = \tilde{\gamma} + \hat{\delta} + \hat{\eta},$$

where $\tilde{\gamma}$ is the OLS estimator of the infeasible auxiliary regression, $\hat{\delta}$ is the effect of the second term, and $\hat{\eta}$ is the effect of the third term. For the third term, $X_t\varepsilon_t$ is uncorrelated with U_t given $E(\varepsilon_t|X_t) = 0$. Therefore, this term, after scaled by the factor $\hat{\beta} - \beta^o$ that itself vanishes to zero in probability at the rate $n^{-1/2}$, will vanish to zero in probability at a rate n^{-1}, that is, $\hat{\eta} = O_P(n^{-1})$. This is expected to have negligible impact on the asymptotic distribution of the test statistic. For the second term, $X_t X_t'$ is perfectly correlated with U_t. However, it is scaled by a factor of $||\hat{\beta} - \beta^o||^2$ rather than by $||\hat{\beta} - \beta^o||$ only. As a consequence, the regression coefficient of $(\hat{\beta} - \beta^o)' X_t X_t' (\hat{\beta} - \beta^o)$

on U_t will also vanish to zero at rate n^{-1}, that is, $\hat{\delta} = O_P(n^{-1})$. Therefore, it also has negligible impact on the asymptotic distribution of $(n - J - 1)R^2$.

Question: How to test conditional homoskedasticity if $E(\varepsilon_t^4|X_t)$ is not a constant (i.e., $E(\varepsilon_t^4|X_t) \neq \mu_4$ for some μ_4 under $\mathbf{H}_0$)? This corresponds to the case when v_t displays conditional heteroskedasticity.

Question: Suppose White's (1980) test rejects the null hypothesis of conditional homoskedasticity, one can then conclude that there exists evidence of conditional heteroskedasticity. What conclusion can one reach if White's test fails to reject $\mathbf{H}_0$: $E(\varepsilon_t^2|X_t) = \sigma^2$?

Because White (198) consider a quadratic alternative to test $\mathbf{H}_0$, it may have no power against some conditional heteroskedastic alternatives for which $E(\varepsilon_t^2|X_t)$ does not depend on the quadratic form of X_t but depend on cubic or higher order polynomials of X_t. Thus, when White's test fails to reject $\mathbf{H}_0$, one can only say that we find no evidence against $\mathbf{H}_0$.

However, when White's test fails to reject $\mathbf{H}_0$, we have

$$E(\varepsilon_t^2 X_t X_t') = \sigma^2 E(X_t X_t') = \sigma^2 Q$$

even if $\mathbf{H}_0$ is false. Therefore, one can use the conventional variance-covariance estimator $s^2(X'X)^{-1}$ for $\hat{\beta}$. Indeed, the main motivation for White's (1980) test for conditional heteroskedasticity is whether the heteroskedasticity-consistent variance-covariance matrix of $\hat{\beta}$ has to be used, not really whether conditional heteroskedasticity exists. For this purpose, it suffices to regress ε_t^2 or e_t^2 on the quadratic form of X_t. This can be seen from the decomposition

$$V = E(X_t X_t' \varepsilon_t^2) = \sigma^2 Q + \mathrm{cov}(X_t X_t', \varepsilon_t^2),$$

which indicates that $V = \sigma^2 Q$ if and only if ε_t^2 is uncorrelated with $X_t X_t'$.

The validity of White's test procedure and associated interpretations is built upon the assumption that the linear regression model is correctly specified for the conditional mean $E(Y_t|X_t)$.Suppose the linear regression model is not correctly specified, i.e., $E(Y_t|X_t) \neq X_t'\beta$ for all β. Then the OLS $\hat{\beta}$ will converge to $\beta^* = [E(X_t X_t')]^{-1} E(X_t Y_t)$, the best linear least squares approximation coefficient, and $E(Y_t|X_t) \neq X_t'\beta^*$. In this case, the estimated residual

$$\begin{aligned} e_t &= Y_t - X_t'\hat{\beta} \\ &= \varepsilon_t + [E(Y_t|X_t) - X_t'\beta^*] + X_t'(\beta^* - \hat{\beta}), \end{aligned}$$

where $\varepsilon_t = Y_t - E(Y_t|X_t)$ is the true disturbance with $E(\varepsilon_t|X_t) = 0$, the estimation error $X_t'(\beta^* - \hat{\beta})$ vanishes to 0 as $n \to \infty$, but the approximation error $E(Y_t|X_t) - X_t'\beta^*$ never disappears. In other words, when the linear regression model is misspecified for $E(Y_t|X_t)$, the estimated residual e_t will contain not only the true disturbance but also the approximation error which is a function of X_t. This will result in a spurious conditional heteroskedasticity when White's test is used. Therefore, before using White's test or any other tests for conditional heteroskedasticity, it is important to first check whether the linear regression model is correctly specified. For tests of correct specification of a linear regression model, see Hausman's test in Chapter 7 and other specification tests mentioned there.

4.7 Empirical Applications

4.8 Summary and Conclusion

In this chapter, within the context of i.i.d. observations, we have relaxed some key assumptions of the classical linear regression model. In particular, we do not assume conditional normality for ε_t and allow for conditional heteroskedasticity. Because the exact finite sample distribution of the OLS is generally unknown, we have replied on asymptotic analysis. It is found that for large samples, the results of the OLS estimator $\hat{\beta}$ and related test statistics (e.g., t-test statistic and F-test statistic) are still applicable under conditional homoskedasticity. Under conditional heteroskedasticity, however, the statistic properties of $\hat{\beta}$ are different from those of $\hat{\beta}$ under conditional homoskedasticity, and as a consequence, the conventional t-test and F-test are invalid even when the sample size $n \to \infty$. One has to use White's (1980) heteroskedasticity-consistent variance estimator for the OLS estimator $\hat{\beta}$ and use it to construct robust test statistics. A direct test for conditional heteroskedastcity, due to White (1980), is described.

The asymptotic theory provides convenient inference procedures in practice. However, the finite sample distribution of $\hat{\beta}$ may be different from its asymptotic distribution. How well the approximation of the asymptotic distribution for the unknown finite sample distribution depends on the data generating process and the sample size of the data. In econometrics, simulation studies have been used to examine how well asymptotic theory can approximate the finite sample distributions of econometric estimators or related statistics. They are the nearest approach that econometricians can make to the laboratory experiments of the physical sciences and are a very useful way of reinforcing or checking the theoretical results. Alternatively, resampling methods called bootstrap have been proposed in econometrics to approximate the finite sample distributions of econometric estimators or related statistics by simulating data on a computer. In this book, we focus on asymptotic theory.

References

Hong, Y. (2006), Probability Theory and Statistics for Economists.

White, H. (1980), *Econometrica*.

White, H. (1999), *Asymptotic Theory for Econometricians,* 2nd Edition.

Hayashi, F. (2000), *Econometrics,* Ch.2.

EXERCISES

4.1. Suppose Assumptions 3.1, 3.3 and 3.5 hold. Show (a) s^2 converges in probability to σ^2, and (b) s converges in probability to σ.

4.2. Let $Z_1, ..., Z_n$ be a random sample from a population with mean μ and variance σ^2. Show that

$$E\left[\frac{\sqrt{n}(\bar{Z}_n - \mu)}{\sigma}\right] = 0 \text{ and } Var\left[\frac{\sqrt{n}(\bar{Z}_n - \mu)}{\sigma}\right] = 1.$$

4.3. Suppose a sequence of random variables $\{Z_n, n = 1, 2, ...\}$ is defined as

$$\begin{array}{ccc} Z_n & \frac{1}{n} & n \\ P_{Z_n} & 1 - \frac{1}{n} & \frac{1}{n} \end{array}$$

(a) Does Z_n converges in mean square to 0? Give your reasoning clearly.
(b) Does Z_n converges in probability to 0? Give your reasoning clearly.

4.4. Let the sample space S be the closed interval [0,1] with the uniform probability distribution. Define $Z(s) = s$ for all $s \in [0, 1]$. Also, for $n = 1, 2, ...$, define a sequence of random variables

$$Z_n(s) = \begin{cases} s + s^n & \text{if } s \in [0, 1 - n^{-1}] \\ s + 1 & \text{if } s \in (1 - n^{-1}, 1]. \end{cases}$$

(a) Does Z_n converge in quadratic mean to Z?
(a) Does Z_n converge in probability to Z?
(b) Does Z_n converge almost surely to Z?

4.5. Suppose $g(\cdot)$ is a real-valued continuous function, and $\{Z_n, n = 1, 2, ...\}$ is a sequence of real-valued random variables which converges in probability to random variable Z. Show $g(Z_n) \to^p g(Z)$.

4.6. Suppose a stochastic process $\{Y_t, X_t'\}'$ satisfies the following assumptions:

Assumption 1.1 [Linearity] $\{Y_t, X_t'\}'$ is an i.i.d. process with

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

for some unknown parameter β^o and some unobservable disturbance ε_t;

Assumption 1.2 [i.i.d.] The $K \times K$ matrix $E(X_t X_t') = Q$ is nonsingular and finite;

Assumption 1.3 [conditional heteroskedasticity]:
(i) $E(X_t \varepsilon_t) = 0$;
(ii) $E(\varepsilon_t^2 | X_t) \neq \sigma^2$;
(iii) $E(X_{jt}^4) \leq C$ for all $0 \leq j \leq k$, and $E(\varepsilon_t^4) \leq C$ for some $C < \infty$.

(a) Show that $\hat{\beta} \to^p \beta^o$?
(b) Show that $\sqrt{n}(\hat{\beta} - \beta^o) \to^d N(0, \Omega)$, where $\Omega = Q^{-1} V Q^{-1}$, and $V = E(X_t X_t' \varepsilon_t^2)$.
(c) Show that the asymptotic variance estimator

$$\hat{\Omega} = \hat{Q}^{-1} \hat{V} \hat{Q}^{-1} \to^p \Omega,$$

where $\hat{Q} = n^{-1} \sum_{t=1}^n X_t X_t'$ and $\hat{V} = n^{-1} \sum_{t=1}^n X_t X_t' e_t^2$. This is called White's (1980) heteroskedasticity consistent variance-covariance matrix estimator.
(d) Consider a test for hypothesis $\mathbf{H}_0 : R\beta^o = r$. Do we have $J \cdot F \to^d \chi_J^2$, where

$$F = \frac{(R\hat{\beta} - r)'[R(\mathbf{X}'\mathbf{X})^{-1} R']^{-1} (R\hat{\beta} - r)/J}{s^2}$$

is the usual F-test statistic? If it holds, give the reasoning. If it does not, could you provide an alternative test statistic that converges in distribution to χ_J^2.

4.7. Put $Q = E(X_t X_t'), V = E(\varepsilon_t^2 X_t X_t')$ and $\sigma^2 = E(\varepsilon_t^2)$. Suppose there exists conditional heteroskedasticity, and $\text{cov}(\varepsilon_t^2, X_t X_t') = V - \sigma^2 Q$ is positive semi-definite, i.e, $\sigma^2(X_t)$ is positively correlated with $X_t X_t'$. Show that $Q^{-1} V Q^{-1} - \sigma^2 Q^{-1}$ is positive semi-definite.

4.8. Suppose the following assumptions hold:

Assumption 2.1: $\{Y_t, X_t'\}'$ is an i.i.d. random sample with

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

for some unknown parameter β^o and unobservable random disturbance ε_t.

Assumption 2.2: $E(\varepsilon_t | X_t) = 0$ a.s.

Assumption 2.3:
(i) $W_t = W(X_t)$ is a positive function of X_t;
(ii) The $K \times K$ matrix $E(X_t W_t X_t') = Q_w$ is finite and nonsingular.
(iii) $E(W_t^8) \leq C < \infty$, $E(X_{jt}^8) \leq C < \infty$ for all $0 \leq j \leq k$, and $E(\varepsilon_t^4) \leq C$;

Assumption 2.4: $V_w = E(W_t^2 X_t X_t' \varepsilon_t^2)$ is finite and nonsingular.

We consider the so-called weighted least squares (WLS) estimator for β^o :

$$\hat{\beta}_w = \left(n^{-1} \sum_{t=1}^{n} X_t W_t X_t' \right)^{-1} n^{-1} \sum_{t=1}^{n} X_t W_t Y_t.$$

(a) Show that $\hat{\beta}_w$ is the solution to the following problem

$$\min_{\beta} \sum_{t=1}^{n} W_t (Y_t - X_t'\beta)^2.$$

(b) Show that $\hat{\beta}_w$ is consistent for β^o;
(c) Show that$\sqrt{n}(\hat{\beta}_w - \beta^o) \to^d N(0, \Omega_w)$ for some $K \times K$ finite and positive definite matrix Ω_w. Obtain the expression of Ω_w under (i) conditional homoskedasticity $E(\varepsilon_t^2|X_t) = \sigma^2$ a.s. and (ii) conditional heteroskedasticity $E(\varepsilon_t^2|X_t) \neq \sigma^2$.
(d) Propose an estimator $\hat{\Omega}_w$ for Ω_w, and show that $\hat{\Omega}_w$ is consistent for Ω_w under conditional homoskedasticity and conditional heteroskedasticity respectively.
(e) Construct a test statistic for $\mathbf{H}_0 : R\beta^o = r$, where R is a $J \times K$ matrix and r is a $J \times 1$ vector under conditional homoskedasticity and under conditional heteroskedasticity respectively. Derive the asymptotic distribution of the test statistic under $\mathbf{H}_0$ in each case.
(f) Suppose $E(\varepsilon_t^2|X_t) = \sigma^2(X_t)$ is
known, and we set $W_t = \sigma^{-1}(X_t)$. Construct a test statistic for $\mathbf{H}_0 : R\beta^o = r$, where R is a $J \times K$ matrix and r is a $J \times 1$ vector. Derive the asymptotic distribution of the test statistic under $\mathbf{H}_0$.

4.9. Consider the problem of testing conditional homoskedasticity ($\mathbf{H}_0 : E(\varepsilon_t^2|X_t) = \sigma^2$) for a linear regression model

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

where X_t is a $K \times 1$ vector consisting of an intercept and explanatory variables. To test conditional homoskedasticity, we consider the auxiliary regression

$$\begin{aligned} \varepsilon_t^2 &= \text{vech}(X_t X_t')'\gamma + v_t \\ &= U_t'\gamma + v_t, \end{aligned}$$

Show that under $\mathbf{H}_0 : E(\varepsilon_t^2|X_t) = \sigma^2$, (a) $E(v_t|X_t) = 0$, and (b) $E(v_t^2|X_t) = \sigma_v^2$ if and only if $E(\varepsilon_t^4|X_t) = \mu_4$ for some constant μ_4.

4.10. Consider the problem of testing conditional homoskedasticity ($\mathbf{H}_0 : E(\varepsilon_t^2|X_t) = \sigma^2$) for a linear regression model

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

where X_t is a $K \times 1$ vector consisting of an intercept and explanatory variables. To test conditional homoskedasticity, we consider the auxiliary regression

$$\begin{aligned}\varepsilon_t^2 &= \text{vech}(X_t X_t')'\gamma + v_t \\ &= U_t'\gamma + v_t.\end{aligned}$$

Suppose Assumptions 4.1, 4.2, 4.3, 4.4, 4.7 hold, and $E(\varepsilon_t^4|X_t) \neq \mu_4$. That is, $E(\varepsilon_t^4|X_t)$ is a function of X_t.

(a) Show $\text{var}(v_t|X_t) \neq \sigma_v^2$ under $\mathbf{H}_0$. That is, the disturbance v_t in the auxiliary regression model displays conditional heteroskedasticity.

(b) Suppose ε_t is directly observable. Construct an asymptotically valid test for the null hypothesis $\mathbf{H}_0$ of conditional homoskedasticity of ε_t. Justify your reasoning and test statistic.