CHAPTER 5 LINEAR REGRESSION MODELS WITH DEPENDENT OBSERVATIONS

Key words: Dynamic regression model, Ergodicity, Martingale difference sequence, Random walk, Serial correlation, Static regression model, Stationarity, Time series, Unit root, White noise.

Abstract: In this chapter, we will show that the asymptotic theory for linear regression models with i.i.d. observations carries over to linear time series regression models with martingale difference sequence disturbances. Some basic concepts in time series analysis are introduced, and some tests for serial correlation are described.

Motivation

The asymptotic theory developed in Chapter 4 is applicable for cross-sectional data (due to the i.i.d. random sample assumption). What happens if we have time series data? Could the asymptotic theory for linear regression models with i.i.d. observations be applicable to linear regression models with time series observations?

Consider a simple regression model

$$\begin{aligned} Y_t &= X_t'\beta^o + \varepsilon_t \\ &= \beta_0 + \beta_1 Y_{t-1} + \varepsilon_t, \\ \{\varepsilon_t\} &\sim \text{i.i.d.} N(0, \sigma^2). \end{aligned}$$

Here, $X_t = (1, Y_{t-1})'$. This is called an autoregression model, which violates the i.i.d. assumption for $\{Y_t, X_t'\}'$ in Chapter 4. Here, we have

$$E(\varepsilon_t | X_t) = 0 \text{ a.s.}$$

but we no longer have

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

because X_{t+j} contains ε_t when $j > 0$. Hence, Assumption 3.2 (strict exogeneity) fails.

In general, the i.i.d. assumption for $\{Y_t, X_t'\}'$ in Chapter 4 rules out time series data. Most economic and financial data are time series observations.

Question: Under what conditions will the asymptotic theory developed in Chapter 4 carry over to linear regression models with dependent observations?

Introduction to Time Series Analysis

To establish asymptotic theory for linear regression models with time series observations, we need to first introduce some basic concepts in time series.

Question: What is a time series process?

A time series process can be stochastic or deterministic. In this book, we only consider stochastic time series processes, which is consistent with the fundamental axiom of modern econometrics discussed in Chapter 1.

Definition [Stochastic Time Series Process]: *A stochastic time series* $\{Z_t\}$ *is a sequence of random variables or random vectors indexed by time* $t \in \{..., 0, 1, 2, ...\}$ *and governed by some probability law* (Ω, F, P), *where* Ω *is the sample space,* F *is a* σ*-field, and* P *is a probability measure, with* $P : F \to [0, 1]$.

Remarks:

More precisely, we can write $Z_t = Z(t, \cdot)$, and its realization $z_t = Z(t, \omega)$, where $\omega \in \Omega$ is a basic outcome in sample space Ω.

For each ω, we can obtain a sample path $z_t = Z(t, \omega)$ of the process $\{Z_t\}$ as a deterministic function of time t. Different ω's will give different sample paths.

The dynamics of $\{Z_t\}$ is completely determined by the *transition probability* of Z_t; that is, the *conditional probability* of Z_t given its past history $I_{t-1} = \{Z_{t-1}, Z_{t-2}, ...\}$.

Time Series Random sample: Consider a subset (or a segment) of a time series process $\{Z_t\}$ for $t = 1, \cdots, n$. This is called a time series random sample of size n, denoted as

$$Z^n = \{Z_1, \cdots, Z_n\}'.$$

Any realization of this random sample is called a data set, denoted as

$$z^n = \{z_1, \cdots, z_n\}'.$$

This corresponds to the occurrence of some specific outcome $\omega \in \Omega$. In theory, a random sample Z^n can generate many data sets, each corresponding to a specific $\omega \in \Omega$. In reality, however, one only observes a data set for any random sample of the economic process, due to the nonexperimental nature of the economic system.

Question: Why can the dynamics of $\{Z_t\}$ be completely captured by its conditional probability distribution?

Consider the random sample Z^n. It is well-known from basic statistics courses that the joint probability distribution of the random sample Z^n,

$$f_{Z^n}(z^n) = f_{Z_1,Z_2,...,Z_n}(z_1, z_2, ..., z_n), \qquad z^n \in \mathbb{R}^n,$$

completely captures all the sample information contained in Z^n. With $f_{Z^n}(z^n)$, we can, in theory, obtain the sampling distribution of any statistic (e.g., sample mean estimator, sample variance estimator, confidence interval estimator) that is a function of Z^n.

Now, by sequential partition (repeating the multiplication rule $P(A \cap B) = P(A|B)P(B)$ for any event A and B), we can write

$$f_{Z^n}(z^n) = \prod_{t=1}^{n} f_{Z_t|I_{t-1}}(z_t|I_{t-1}),$$

where by convention, for $t = 1$, $f(z_1|I_0) = f(z_1)$, the marginal density of Z_1. Thus, the conditional density function $f_{Z_t|I_{t-1}}(z|I_{t-1})$ completely describes the joint probability of the random sample Z^n.

Example 1: Let Z_t be the US Gross Domestic Product (GDP) in quarter t. Then the quarterly records of U.S. GDP from the first quarter of 1961 to the last quarter of 2001 constitute a time series data set, denoted as $z^n = (z_1, \cdots, z_n)'$, with $n = 164$.

Example 2: Let Z_t be the S&P 500 closing price index at day t. Then the daily records of S & P 500 index from July 2, 1962 to December 31, 2001 constitute a time series data set, denoted as $z^n = (z_1, \cdots, z_n)'$, with $n = 9987$.

Here is a fundamental feature of economic time series: Each random variable Z_t only has one observed realization z_t in practice. It is impossible to obtain more realizations for each economic variable Z_t, due to the nonexperimental nature of an economic system. In order to "aggregate" realizations from different random variables $\{Z_t\}_{t=1}^n$, we need to impose stationarity—a concept of stability for certain aspects of the probability law $f_{Z_t|I_{t-1}}(z|I_{t-1})$. For example, we may need to assume:
(i) The marginal probability of each Z_t shares some common features (e.g., the same mean, the same variance).
(ii) The relationship (joint distribution) between Z_t and I_{t-1} is time-invariant in certain aspects (e.g., $\text{cov}(Z_t, Z_{t-j}) = \gamma(j)$ does not depend on time t; it only depends on the time distance j).

With these assumptions, observations from different random variables $\{Z_t\}$ can be viewed to contain some common features of the data generating process, so that one can conduct statistical inference by pooling them together.

Stationarity

A stochastic time series $\{Z_t\}$ can be stationary or nonstationary. There are at least two notions for stationarity. The first is strict stationarity.

Definition [Strict Stationarity]: *A stochastic time series process $\{Z_t\}$ is strictly stationary if for any admissible $t_1, t_2, \cdots, t_m$, the joint probability distribution of $\{Z_{t_1}, Z_{t_2}, \cdots, Z_{t_m}\}$ is the same as the joint distribution of $\{Z_{t_1+k}, Z_{t_2+k}, \cdots, Z_{t_m+k}\}$ for all integers k. That is,*

$$f_{Z_{t_1}Z_{t_2}\cdots Z_{t_m}}(z_1, \cdots, z_m) = f_{Z_{t_1+k}Z_{t_2+k}\cdots Z_{t_m+k}}(z_1, \cdots, z_m).$$

Remarks:

If Z_t is strictly stationary, the conditional probability of Z_t given I_{t-1} will have a time-invariant functional form. In other words, the probabilistic structure of a completely stationary process is invariant under a shift of the time origin.

Strict stationarity is also called "complete stationarity", because it characterizes the time-invariance property of the entire joint probability distribution of the process $\{Z_t\}$.

No moment condition on $\{Z_t\}$ is needed when defining strict stationarity. Thus, a strictly stationary process may not have finite moments (e.g., $\text{var}(Z_t) = \infty$). However, if moments (e.g., $E(Z_t)$) and cross-moments ($e.g., E(Z_t Z_{t-j})$) of $\{Z_t\}$ exist, then they are time-invariant when $\{Z_t\}$ is strictly stationary.

Any measurable transformation of a strictly stationary process is still strictly stationary.

Strict stationarity implies identical distribution for each of the Z_t. Thus, although a strictly stationary time series data are realizations from different random variables, they can be viewed as realizations from the same (marginal) population distribution.

Example 1: Suppose $\{Z_t\}$ is an i.i.d. Cauchy $(0, 1)$ sequence with marginal pdf

$$f(z) = \frac{1}{\pi(1+z^2)}, \qquad -\infty < z < \infty.$$

Note that Z_t has no moment. Consider $\{Z_{t_1}, \cdots, Z_{t_m}\}$. Because the joint distribution

$$f_{Z_{t_1}Z_{t_2}\cdots Z_{t_m}}(z_1, \cdots, z_m) = \prod_{j=1}^{m} f(z_j)$$

is time-invariant, $\{Z_t\}$ is strictly stationary.

We now introduce another concept of stationarity based on the time-invariance property of the joint moments of $\{Z_{t_1}, Z_{t_2}, ..., Z_{t_m}\}$.

Definition [N-th order stationarity]: *The time series process $\{Z_t\}$ is said to be stationary up to order N if, for any admissible $t_1, t_2, \cdots, t_m$, and any k, all the joint moments up to order N of $\{Z_{t_1}, Z_{t_2}, \cdots, Z_{t_m}\}$ exist and equal to the corresponding joint moments up to order N of $\{Z_{t_1+k}, \cdots, Z_{t_m+k}\}$. That is,*

$$E\left[(Z_{t_1})^{n_1} \cdots (Z_{t_m})^{n_m}\right] = E\left[(Z_{t_1+k})^{n_1} \cdots (Z_{t_m+k})^{n_m}\right],$$

for any k and all nonnegative integers $n_1, \cdots, n_m$ satisfying $\sum_{j=1}^m n_j \leq N$.

Remarks:

Setting $n_2 = n_3 = \cdots = n_m = 0$, we have

$$E\left[(Z_t)^{n_1}\right] = E\left[(Z_0)^{n_1}\right] \text{ for all } t.$$

On the other hand, for $n_1 + n_2 \leq N$, we have the pairwise joint product moment

$$\begin{aligned} E\left[(Z_t)^{n_1}(Z_{t-j})^{n_2}\right] &= E\left[(Z_0)^{n_1}(Z_{-j})^{n_2}\right] \\ &= \text{function of } j, \end{aligned}$$

where j is called a lag order.

We now consider a special case: $N = 2$. This yields a concept called weak stationarity.

Definition [Weak Stationarity] *A stochastic time series process $\{Z_t\}$ is weakly stationary if*

(i) $E(Z_t) = \mu$ for all t;
(ii) $var(Z_t) = \sigma^2 < \infty$ for all t;
(iii) $cov(Z_t, Z_{t-j}) = \gamma(j)$ is only a function of lag order j for all t.

Remarks:

Strict stationarity is defined in terms of the "time invariance" property of the entire distribution of $\{Z_t\}$, while weak-stationarity is defined in terms of the "time-invariance" property in the first two moments (means, variances and covariances) of $\{Z_t\}$. Suppose all moments of $\{Z_t\}$ exist. Then it is possible that the first two moments are time-invariant but the higher order moments are time-varying. In other words, a process $\{Z_t\}$ can be weakly stationary but not strictly stationary. However, Example 1 shows that a process can be strictly stationary but not weakly stationary, because the first two moments simply do not exist.

Weak stationarity is also called "covariance-stationarity", or "2nd order stationarity" because it is based on the time-invariance property of the first two moments. It does not require identical distribution for each of the Z_t. The higher order moments of Z_t can be different for different t's.

Question: Which, strict or weak stationarity, is more restrictive?

We consider two cases:

(i) If $E(Z_t^2) < \infty$, then strict stationarity implies weak stationarity.

(ii) However, if $E(Z_t^2) = \infty$, strict stationarity does not imply weak stationarity. In other words, a time series process can be strictly stationary but not weakly stationary.

Example: An i.i.d. Cauchy$(0, 1)$ process is strictly stationary but not weakly stationary.

A special but important weakly stationary time series is a process with zero autocorrelations.

Definition [White Noise]: *A time series process $\{Z_t\}$ is a white noise (or serially uncorrelated) process if*

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

(ii) $var(Z_t) = \sigma^2,$

(iii) $cov(Z_t, Z_{t-j}) = \gamma(j) = 0$ *for all* $j > 0.$

Remarks:

Later we will explain why such a process is called a white noise (WN) process. WN is a basic building block for linear time series modeling.

When $\{Z_t\}$ is a white noise and $\{Z_t\}$ is a Gaussian process (i.e., any finite set $(Z_{t_1}, Z_{t_2}, ..., Z_{t_m})$ of $\{Z_t\}$ has a joint normal distribution), we call $\{Z_t\}$ is a Gaussian white noise. For a Gaussian white noise process, $\{Z_t\}$ is an i.i.d. sequence.

Example 2: A first order autoregressive (AR(1)) process

$$\begin{aligned} Z_t &= \alpha Z_{t-1} + \varepsilon_t, \\ \varepsilon_t &\sim \text{white noise } (0, \sigma^2) \end{aligned}$$

is weakly stationary if $|\alpha| < 1$ (Z_t is a unit root process if $\alpha = 1$) because $Z_t = \sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j}$, and

$$\begin{aligned} E(Z_t) &= 0, \\ \mathrm{var}(Z_t) &= \frac{\sigma^2}{1-\alpha^2}, \\ \gamma(j) &= \frac{\sigma^2}{1-\alpha^2}\alpha^{|j|}, \qquad j = 0, \pm 1, \pm 2, \ldots . \end{aligned}$$

Here, ε_t may be interpreted as a random shock or an innovation that derives the movement of the process $\{Z_t\}$ over time.

More generally, $\{Z_t\}$ is an AR(p) process if

$$\begin{aligned} Z_t &= \alpha_0 + \sum_{j=1}^{p} \alpha_j Z_{t-j} + \varepsilon_t, \\ \varepsilon_t &\sim \text{White noise } (0, \sigma^2). \end{aligned}$$

Example 3: $\{Z_t\}$ is a q-th order moving-average process (MA(q)) if

$$\begin{aligned} Z_t &= \alpha_0 + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j} + \varepsilon_t, \\ \{\varepsilon_t\} &\sim \text{White noise } (0, \sigma^2). \end{aligned}$$

This is a weakly stationary process. For an MA(q) process, we have $\gamma(j) = 0$ for all $|j| > q$.

Example 4: $\{Z_t\}$ is an autoregressive-moving average (ARMA) process of orders (p, q) if

$$\begin{aligned} Z_t &= \alpha_0 + \sum_{j=1}^{p} \alpha_j Z_{t-j} + \sum_{j=1}^{q} \beta_j \varepsilon_{t-j} + \varepsilon_t, \\ \{\varepsilon_t\} &\sim \text{white noise } (0, \sigma^2). \end{aligned}$$

ARMA models include AR models and MA models as special cases. An estimation method for ARMA models can be found in Chapter 9.

Under rather mild regularity conditions, a zero-mean weakly stationary process can be represented by an MA(∞) process

$$\begin{aligned} Z_t &= \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j}, \\ \varepsilon_t &\sim \mathrm{WN}(0, \sigma^2), \end{aligned}$$

where $\Sigma_{j=1}^{\infty}\alpha_j^2 < \infty$. This is called Wold's decomposition. The partial derivative

$$\frac{\partial Z_{t+j}}{\partial \varepsilon_t} = \alpha_j, j = 0, 1, ...$$

is called the impulse response function of the time series process $\{Z_t\}$ with respect to a random shock ε_t. This function characterizes the impact of a random shock ε_t on the immediate and subsequent observations $\{Z_{t+j}, j \geq 0\}$. For a weakly stationary process, the impact of any shock on a future Z_{t+j} will always diminish to zero as the lag order $j \to \infty$, because $\alpha_j \to 0$. The ultimate cumulative impact of ε_t on the process $\{Z_t\}$ is the sum $\Sigma_{j=0}^{\infty}\alpha_j$.

The function $\gamma(j) = \mathrm{cov}(Z_t, Z_{t-j})$ is called the autocovariance function of the weakly stationary process $\{Z_t\}$, where j is a lag order. It characterizes the (linear) serial dependence of Z_t on its own lagged variable Z_{t-j}. Note that $\gamma(j) = \gamma(-j)$ for all integers j.

The normalized function $\rho(j) = \gamma(j)/\gamma(0)$ is called the autocorrelation function of $\{Z_t\}$. It has the property that $|\rho(j)| \leq 1$. The plot of $\rho(j)$ as a function of j is called the autocorrelogram of the time series process $\{Z_t\}$. It can be used to judge which linear time series model (e.g., AR, MA, or ARMA) should be used to fit a particular time series data set.

We now consider the Fourier transform of the autocovariance function $\gamma(j)$.

Definition [Spectral Density Function] *The Fourier transform of* $\gamma(j)$

$$h(\omega) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\gamma(j)e^{-ij\omega}, \qquad \omega \in [-\pi, \pi],$$

where $i = \sqrt{-1}$, *is called the power spectral density of process* $\{Z_t\}$.

The normalized version

$$f(\omega) = \frac{h(\omega)}{\gamma(0)} = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\rho(j)e^{-ij\omega}, \qquad \omega \in [-\pi, \pi],$$

is called the standardized spectral density of $\{Z_t\}$.

Question: What are the properties of $f(\omega)$?

It can be shown that (i) $f(\omega)$ is real-valued, and $f(\omega) \geq 0$; (ii) $\int_{-\pi}^{\pi} f(\omega)d\omega = 1$; (iii) $f(-\omega) = f(\omega)$.

The spectral density $h(\omega)$ is widely used in economic analysis. For example, it can be used to search for business cycles. Specifically, a frequency ω_0 corresponding to a special peak is closely associated with a business cycle with periodicity $T_0 = 2\pi/\omega_0$. Intuitively, time series can be decomposed as the sum of many cyclical components with different frequencies ω, and $h(\omega)$ is the strength or magnitude of the component with frequency ω. When $h(\omega)$ has a peak at ω_0, it means that the cyclical component with frequency ω_0 or periodicity $T_0 = 2\pi/\omega_0$ dominates all other frequencies. Consequently, the whole time series behaves as mainly having a cycle with periodicity T_0.

The functions $h(\omega)$ and $\gamma(j)$ are Fourier transforms of each other. Thus, they contain the same information on serial dependence in $\{Z_t\}$. In time series analysis, the use of $\gamma(j)$ is called the time domain analysis, and the use of $h(\omega)$ is called the frequency domain analysis. Which tool to use depends on the convenience of the user. In some applications, the use of $\gamma(j)$ is simpler and more intuitive, while in other applications, the use of $h(\omega)$ is more enlightening. This is exactly the same as the case that it is more convenient to use Chinese in China, while it is more convenient to use English in U.S.

Example 1: Hamilton, James (1994, *Time Series Analysis*): Business cycles of U.S. industrial production

Example 2: Steven Durlauf (1990, *Journal of Monetary Economics*): Income tax rate changes

Reference: Sargent, T. (1987): *Macroeconomic Theory*, 2nd Edition. Academic Press: Orlando, U.S.A.

For a serially uncorrelated sequence, the spectral density $h(\omega)$ is flat as a function of frequency ω :

$$\begin{aligned} h(\omega) &= \frac{1}{2\pi}\gamma(0) \\ &= \frac{1}{2\pi}\sigma^2 \text{ for all } \omega \in [-\pi, \pi]. \end{aligned}$$

This is analogous to the power (or energy) spectral density of a physical white color light. It is for this reason that we call a serially uncorrelated time series a white noise process.

Intuitively, a white color light can be decomposed via a lens as the sum of equal magnitude components of different frequencies. That is, a white color light has a flat physical spectral density function.

It is important to point out that a white noise may not be i.i.d., as is illustrated by the example below:

Example 3: Consider an autoregressive conditional heteroskedastic (ARCH) process

$$\begin{aligned} Z_t &= \varepsilon_t h_t^{1/2}, \\ h_t &= \alpha_0 + \alpha_1 Z_{t-1}^2, \\ \varepsilon_t &\sim \text{ i.i.d.(0,1)}. \end{aligned}$$

This is first proposed by Engle (1982) and it has been widely used to model volatility in economics and finance. We have $E(Z_t|I_{t-1}) = 0$ and $\text{var}(Z_t|I_{t-1}) = h_t$, where $I_{t-1} = \{Z_{t-1}, Z_{t-2}, ...\}$ is the information set containing all past history of Z_t.

It can be shown that

$$\begin{aligned} E(Z_t) &= 0, \\ \text{cov}(Z_t, Z_{t-j}) &= 0 \text{ for } j > 0, \\ \text{var}(Z_t) &= \frac{\alpha_0}{1-\alpha_1}. \end{aligned}$$

When $\alpha_1 < 1$, $\{Z_t\}$ is a stationary white noise. But it is not weakly stationary if $\alpha_1 = 1$, because $\text{var}(Z_t) = \infty$. In both cases, $\{Z_t\}$ is strictly stationary (e.g., Nelson 1990, *Journal of Econometrics*).

Although $\{Z_t\}$ is a white noise, it is not an i.i.d. sequence because the correlation in $\{Z_t^2\}$ is $\text{corr}(Z_t^2, Z_{t-j}^2) = \alpha_1^{|j|}$ for $j = 0, 1, 2, ...$ In other words, an ARCH process is uncorrelated in level but is autocorrelated in squares.

Nonstationarity

Usually, we call $\{Z_t\}$ a nonstationary time series when it is not covariance-stationary. In time series econometrics, there have been two types of nonstationary processes that display similar sample paths when the sample size is not large but have quite different implications. We first discuss a nonstationary process called trend-stationary process.

Example 3: $\{Z_t\}$ is called a trend-stationary process if

$$Z_t = \alpha_0 + \alpha_1 t + \varepsilon_t,$$

where ε_t is a weakly stationary process with mean 0 and variance σ^2. To see why $\{Z_t\}$ is not weakly stationary, we consider a simplest case where $\{\varepsilon_t\}$ is i.i.d. $(0, \sigma^2)$. Then

$$\begin{aligned} E(Z_t) &= \alpha_0 + \alpha_1 t, \\ var(Z_t) &= \sigma^2, \\ \text{cov}(Z_t, Z_{t-j}) &= 0. \end{aligned}$$

Question: What happens if $\Delta Z_t = Z_t - Z_{t-1}$?

More generally, a trend-stationary time series process can be defined as follows:

$$Z_t = \alpha_0 + \sum_{j=1}^{p} \alpha_j t^j + \varepsilon_t,$$

where $\{\varepsilon_t\}$ is a weakly stationary process. The reason that $\{Z_t\}$ is called trend-stationary is that it will become weakly stationary after the deterministic trend is removed.

Next, we discuss the second type of nonstationary process called difference-stationary process. Again, we start with a special case:

Example 4: $\{Z_t\}$ is a random walk with a drift if

$$Z_t = \alpha_0 + Z_{t-1} + \varepsilon_t,$$

where $\{\varepsilon_t\}$ is i.i.d. $(0, \sigma^2)$. For simplicity, we assume $Z_0 = 0$. Then

$$\begin{aligned} E(Z_t) &= \alpha_0 t, \\ \mathrm{var}(Z_t) &= \sigma^2 t, \\ \mathrm{cov}(Z_t, Z_{t-j}) &= \sigma^2 (t-j). \end{aligned}$$

Note that for any given j,

$$\mathrm{corr}(Z_t, Z_{t-j}) = \sqrt{\frac{t-j}{t}} \to 1 \text{ as } t \to \infty,$$

which implies that the impact of an infinite past event on today's behavior never dies out. Indeed, this can be seen more clearly if we write

$$Z_t = Z_0 + \alpha_0 t + \sum_{j=0}^{t-1} \varepsilon_{t-j}.$$

Note that $\{Z_t\}$ has a deterministic linear time trend but with an increasing variance over time. The impulse response function $\partial Z_{t+j}/\partial \varepsilon_t = 1$ for all $j \geq 0$, which never dies off to zero as $j \to \infty$.

There is another nonstationary process called martingale process which is closely related to a random walk.

Definition [Martingale] *A time series process $\{Z_t\}$ is a martingale with drift if*

$$Z_t = \alpha + Z_{t-1} + \varepsilon_t,$$

and $\{\varepsilon_t\}$ *satisfies*

$$E(\varepsilon_t|I_{t-1}) = 0 \text{ a.s.},$$

where I_{t-1} *is the* σ*-field generated by* $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots\}$*. We call that* $\{\varepsilon_t\}$ *is a martingale difference sequence (MDS).*

Question: Why is ε_t called an MDS?

Because ε_t is the difference of a martingale process. That is, $\varepsilon_t = Z_t - Z_{t-1}$.

Example [Martingale and Efficient Market Hypothesis]: Suppose a stock log-price $\ln P_t$ follows a martingale process, i.e.,

$$\ln P_t = \ln P_{t-1} + \varepsilon_t,$$

where $E(\varepsilon_t|I_{t-1}) = 0$. Then $\varepsilon_t = \ln P_t - \ln P_{t-1} \approx \frac{P_t - P_{t-1}}{P_{t-1}}$ is the stock relative price change or stock return (if no dividend) from time $t-1$ to time t, which can be viewed as the proxy for the new information arrival from time $t-1$ to time t that derives the stock price change in the same period. For this reason, ε_t is also called an innovation sequence. The MDS property of ε_t implies that the price change ε_t is unpredictable using the past information available at time $t-1$, and the market is called informationally efficient. Thus, the best predictor for the stock price at time t using the information available at time $t-1$ is P_{t-1}, that is, $E(P_t|I_{t-1}) = P_{t-1}$.

Question: What is the relationship between a random walk and a martingale?

A random walk is a martingale because IID with zero mean implies $E(\varepsilon_t|I_{t-1}) = E(\varepsilon_t) = 0$. However, the converse is not true.

Example: Reconsider an ARCH(1) process

$$\begin{aligned} \varepsilon_t &= h_t^{1/2} z_t, \\ h_t &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2, \\ \{z_t\} &\sim \text{i.i.d.(0,1)}. \end{aligned}$$

where $\alpha_0, \alpha_1 > 0$. It follows that

$$\begin{aligned} E(\varepsilon_t|I_{t-1}) &= 0, \\ \text{var}(\varepsilon_t|I_{t-1}) &= h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2, \end{aligned}$$

where I_{t-1} denotes the information available at time $t-1$. Clearly $\{\varepsilon_t\}$ is MDS but not IID, because its conditional variance h_t is time-varying (depending on the past information set I_{t-1}).

Since the only condition for MDS is $E(\varepsilon_t|I_{t-1}) = 0$ a.s., an MDS need not be strictly stationary or weakly stationary. However, if it is assumed that $\text{var}(\varepsilon_t) = \sigma^2$ exists, then an MDS is weakly stationary.

When the variance $E(\varepsilon_t^2)$ exists, we have the following directional relationships:

$$\text{IID (with } \mu = 0) \Longrightarrow \text{ MDS} \Longrightarrow \text{ WHITE NOISE.}$$

Lemma: *If $\{\varepsilon_t\}$ is an MDS with $E(\varepsilon_t^2) = \sigma^2 < \infty$, then $\{\varepsilon_t\}$ is a white noise.*

Proof: By the law of iterated expectations, we have

$$E(\varepsilon_t) = E[E(\varepsilon_t|I_{t-1})] = 0,$$

and for any $j > 0$,

$$\begin{aligned}
\text{cov}(\varepsilon_t, \varepsilon_{t-j}) &= E(\varepsilon_t \varepsilon_{t-j}) - E(\varepsilon_t)E(\varepsilon_{t-j}) \\
&= E[E(\varepsilon_t \varepsilon_{t-j}|I_{t-1})] \\
&= E[E(\varepsilon_t|I_{t-1})\varepsilon_{t-j}] \\
&= E(0 \cdot \varepsilon_{t-j}) \\
&= 0.
\end{aligned}$$

This implies that MDS, together with $\text{var}(\varepsilon_t) = \sigma^2$, is a white noise.

However, a white noise does not imply a MDS.

Example: A nonlinear MA process

$$\begin{aligned}
\varepsilon_t &= \alpha z_{t-1} z_{t-2} + z_t, \\
\{z_t\} &\sim i.i.d.(0,1).
\end{aligned}$$

Then it can be shown that $\{\varepsilon_t\}$ is a white noise but not MDS, because $\text{cov}(\varepsilon_t, \varepsilon_{t-j}) = 0$ for all $j > 0$ but

$$E(\varepsilon_t|I_{t-1}) = \alpha z_{t-1} z_{t-2} \neq 0.$$

Question: When will the concepts of IID, MDS and White noise coincide?

When $\{\varepsilon_t\}$ is a stationary Gaussian process. A time series is a stationary Gaussian process if $\{\varepsilon_{t_1}, \varepsilon_{t_2}, ..., \varepsilon_{t_m}\}$ is multivariate normally distributed for any admissible sets of integers $\{t_1, t_2, ..., t_m\}$. Unfortunately, an important stylized fact for economic and financial time series is that they are typically non-Gaussian. Therefore, it is important to emphasize the difference among the concepts of IID, MDS and White Noises in time series econometrics.

When $\mathrm{var}(\varepsilon_t)$ exists, both random walk and martingale processes are special cases of the so-called unit root process which is defined below.

Definition [Unit root or difference stationary process]: $\{Z_t\}$ *is a unit root process if*

$$\begin{aligned} Z_t &= \alpha_0 + Z_{t-1} + \varepsilon_t, \\ & \quad \{\varepsilon_t\} \text{ is covariance-stationarty } (0, \sigma^2). \end{aligned}$$

The process $\{Z_t\}$ is called a unit root process because its autoregressive coefficient is unity. It is also called a difference-stationary process because its first difference,

$$\Delta Z_t = Z_t - Z_{t-1} = \alpha_0 + \varepsilon_t,$$

becomes weakly stationary. In fact, the first difference of a linear trend-stationary process $Z_t = \alpha_0 + \alpha_1 t + \varepsilon_t$ is also weakly stationary:

$$\Delta Z_t = \alpha_1 + \varepsilon_t - \varepsilon_{t-1}.$$

The inverse of differencing is "integrating". For the difference-stationary process $\{Z_t\}$, we can write it as the integral of the weakly stationary process $\{\varepsilon_t\}$ in the sense that

$$Z_t = \alpha_0 t + Z_0 + \sum_{j=0}^{t-1} \varepsilon_{t-j},$$

where Z_0 is the starting value of the process $\{Z_t\}$. This is analogous to differentiation and integration in calculus which are inverses of each other. For this reason, $\{Z_t\}$ is also called an Integrated process of order 1, denoted as $I(1)$.Obviously, a random walk and a martingale process are $I(1)$ processes if the variance of the innovation ε_t is finite.

We will assume strict stationarity in most cases in the present and subsequent chapters. This implies that some economic variables have to be transformed before used in $Y_t = X_t'\beta^o + \varepsilon_t$. Otherwise, the asymptotic theory developed here cannot be applied. Indeed, a different asymptotic theory should be developed for unit root processes (see, e.g., Hamilton (1994), *Time Series Analysis*).

In macroeconomics, it is important to check whether a nonstationary macroeconomic time series is trend-stationary or difference-stationary. If it is a unit root process, then a shock to the economy will never die out to zero as time evolves. In contrast, a random shock to a trend-stationary process will die out to zero eventually.

Question: Why has the unit root econometrics been so popular in econometrics?

It was found in empirical studies (e.g., Nelson and Plosser (1982, *Journal of Monetary Economics*)) that most macroeconomic time series display unit root properties.

Ergodicity

Next, we introduce a concept of asymptotic independence.

Question: Consider the following time series

$$\begin{aligned} Z^n &= (Z_1, Z_2, \cdots, Z_n)' \\ &= (W, W, \cdots, W)', \end{aligned}$$

where W is a random variable that does not depend on time index t. Obviously, the stationarity condition holds. However, any realization of this random sample Z^n will be

$$z^n = (w, w, \cdots, w)',$$

i.e., it will contain the same realization w for all n observations (no new information as n increases). In order to avoid this, we need to impose a condition called ergodicity that assumes that $(Z_t, \cdots, Z_{t+k})$ and $(Z_{m+t}, \cdots, Z_{m+t+l})$ are asymptotically independent when their time distance $m \to \infty$.

Statistically speaking, independence or little correlation generates new or more information as the sample size n increases. Recall that X and Y are independent if and only if

$$E[f(X)g(Y)] = E[f(X)]E[g(Y)]$$

for any measurable functions $f(\cdot)$ and $g(\cdot)$. We now extend this definition to define ergodicity.

Definition [Ergodicity]: *A strictly stationary process $\{Z_t\}$ is said to be ergodic if for any two bounded functions $f : R^{k+1} \to R$ and $g : \mathbb{R}^{l+1} \to \mathbb{R}$,*

$$\begin{aligned} & \lim_{m\to\infty} |E[f(Z_t, \cdots, Z_{t+k})g(Z_{m+t}, \cdots, Z_{m+t+l})]| \\ = & \ |E[f(Z_t, \cdots, Z_{t+k})]| \cdot |E[g(Z_{m+t}, \cdots, Z_{m+t+l})]|. \end{aligned}$$

Remarks:

Clearly, ergodicity is a concept of asymptotic independence. A strictly stationary process that is ergodic is called ergodic stationary. If $\{Z_t\}$ is ergodic stationary, then $\{f(Z_t)\}$ is also ergodic stationary for any measurable function $f(\cdot)$.

Ergodicity Theorem [WLLN for Ergodic Stationary Random Samples]: *Let $\{Z_t\}$ be an ergodic stationary process with $E(Z_t) = \mu$ and $E|Z_t| < \infty$. Then the sample mean*

$$\bar{Z}_n = n^{-1} \sum_{t=1}^{n} Z_t \xrightarrow{p} \mu \text{ as } n \to \infty.$$

Question: Why do we need to assume ergodicity?

Consider a counter example which does not satisfy the ergodicity condition: $Z_t = W$ for all t. Then $\bar{Z}_n = W$, a random variable which will not converge to μ as $n \to \infty$.

Next, we state a CLT for ergodic stationary MDS random samples.

Central Limit Theorem for Ergodic Stationary MDS: *Suppose $\{Z_t\}$ is a stationary ergodic MDS process, with var$(Z_t) \equiv E(Z_t Z_t') = V$ finite, symmetric and positive definite. Then as $n \to \infty$,*

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

or equivalently,

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

Question: Is avar$(\sqrt{n}\bar{Z}_n) = V =var(Z_t)$? That is, is the asymptotic variance of $\sqrt{n}\bar{Z}_n$

coincides with the individual variance var(Z_t).
To check this, we have

$$\begin{aligned}
\text{var}(\sqrt{n}\bar{Z}_n) &= E[\sqrt{n}\bar{Z}_n\sqrt{n}\bar{Z}_n'] \\
&= E\left[\left(n^{-1/2}\sum_{t=1}^{n} Z_t\right)\left(n^{-1/2}\sum_{s=1}^{n} Z_s\right)'\right] \\
&= n^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n} E(Z_t Z_s') \\
[E(Z_t Z_s) &= 0 \text{ for } t \neq s, \text{ by the LIE}] \\
&= n^{-1}\sum_{t=1}^{n} E(Z_t Z_t') \\
&= E(Z_t Z_t') \\
&= V.
\end{aligned}$$

Here, the MDS property plays a crucial rule in simplifying the asymptotic variance of $\sqrt{n}\bar{Z}_n$ because it implies $\text{cov}(Z_t, Z_s) = 0$ for all $t \neq s$.MDS is one of the most important concepts in modern economics, particularly in macroeconomics, finance, and econometrics. For example, rational expectations theory can be characterized by an expectational error being an MDS.

Large Sample Theory for Linear Regression Models with Dependent Observations

5.1 Assumptions

With the basic time series concepts and analytic tools introduced above, we can now develop an asymptotic theory for linear regression models with time series observations. We first state the assumptions that allow for time series observations.

Assumption 5.1 [Ergodic stationarity]: The stochastic process $\{Y_t, X_t'\}'$ is jointly stationary and ergodic.

Assumption 5.2 [Linearity]:

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

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

Assumption 5.3 [Correct Model Specification]: $E(\varepsilon_t|X_t) = 0$ a.s. with $E(\varepsilon_t^2) = \sigma^2 < \infty$.

Assumption 5.4 [Nonsingularity]: The $K \times K$ matrix

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

is finite and nonsingular.

Assumption 5.5 [MDS]: $\{X_t\varepsilon_t\}$ is an MDS process with respect to the σ-field generated by $\{X_s\varepsilon_s, s < t\}$ and 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.

Remarks:

In Assumption 5.1, the ergodic stationary process $Z_t = \{Y_t, X_t'\}'$ can be independent or serially dependent across different time periods. we thus allow for time series observations from a stationary stochastic process.

It is important to emphasize that the asymptotic theory to be developed below and in subsequent chapters is not applicable to nonstationary time series. A problem associated with nonstationary time series is the so-called spurious regression or spurious correlation problem. If the dependent variable Y_t and the regressors X_t display similar trending behaviors over time, one is likely to obtain seemly highly "significant" regression coefficients and high values for R^2, even if they do not have any causal relationship. Such results are completely spurious. In fact, the OLS estimator for nonstationary time series regression model does not follow the asymptotic theory to be developed below. A different asymptotic theory for nonstationary time series regression models has to be used (see, e.g., Hamiltion 1994). Using the correct asymptotic theory, the seemingly highly "significant" regression coefficient estimators would become insignificant in the spurious regression models.

Unlike the i.i.d. case, where $E(\varepsilon_t|X_t) = 0$ is equivalent to the strict exogeneity condition that

$$E(\varepsilon_t|X) = E(\varepsilon_t|X_1, ..., X_t, ..., X_n) = 0,$$

the condition $E(\varepsilon_t|X_t) = 0$ is weaker than $E(\varepsilon_t|X) = 0$ in a time series context. In other words, it is possible that $E(\varepsilon_t|X_t) = 0$ but $E(\varepsilon_t|X) \neq 0$. Assumption 5.3 allows for the inclusion of predetermined variables in X_t, the lagged dependent variables Y_{t-1}, Y_{t-2}, etc.

For example, suppose $X_t = (1, Y_{t-1})'$. Then we obtain an AR(1) model

$$\begin{aligned} Y_t &= X_t'\beta^o + \varepsilon_t \\ &= \beta_0 + \beta_1 Y_{t-1} + \varepsilon_t, \qquad t = 2, \cdots, n. \\ \{\varepsilon_t\} &\sim \text{MDS}(0, \sigma^2). \end{aligned}$$

Then $E(\varepsilon_t|X_t) = 0$ holds if $E(\varepsilon_t|I_{t-1}) = 0$, namely if $\{\varepsilon_t\}$ is an MDS, where I_{t-1} is the sigma-field generated by $\{\varepsilon_{t-1}, \varepsilon_{t-2}, ...\}$. However, we generally have $E(\varepsilon_t|X) \neq 0$ because $E(\varepsilon_t X_{t+1}) \neq 0$.

When X_t contains an intercept the MDS condition for $\{X_t\varepsilon_t\}$ in Assumption 5.5 implies that $E(\varepsilon_t|I_{t-1}) = 0$; that is, $\{\varepsilon_t\}$ is an MDS, where $I_{t-1} = \{\varepsilon_{t-1}, \varepsilon_{t-2}, ...\}$.

Question: When can an MDS disturbance ε_t arise in economics and finance?

Example: Rational Expectations Economics

Recall the dynamic asset pricing model under a rational expectations framework in Chapter 1. The behavior of the economic agent is characterized by the Euler equation:

$$\begin{aligned} E\left[\beta \frac{u'(C_t)}{u'(C_{t-1})} R_t \middle| I_{t-1}\right] &= 1 \text{ or} \\ E[M_t R_t | I_{t-1}] &= 1, \end{aligned}$$

where β is the time discount factor of the representative economic agent, C_t is the consumption, R_t is the asset gross return, and M_t is the stochastic discount factor defined as follows:

$$\begin{aligned} M_t &= \beta \frac{u'(C_t)}{u'(C_{t-1})} \\ &= \beta + \frac{u''(C_{t-1})}{u'(C_{t-1})} \Delta C_t + \text{ higher order} \\ &\sim \text{ risk adjustment factor.} \end{aligned}$$

Using the formula that $\text{cov}(X_t, Y_t | I_{t-1}) = E(X_t Y_t | I_{t-1}) - E(X_t | I_{t-1}) E(Y_t | I_{t-1})$ and rearranging, we can write the Euler equation as

$$E(M_t | I_{t-1}) E(R_t | I_{t-1}) + \text{cov}(M_t, R_t | I_{t-1}) = 1.$$

It follows that

$$\begin{aligned} E(R_t | I_{t-1}) &= \frac{1}{E(M_t | I_{t-1})} + \frac{\text{cov}(M_t, R_t | I_{t-1})}{\text{var}(M_t | I_{t-1})} \frac{-\text{var}(M_t | I_{t-1})}{E(M_t | I_{t-1})} \\ &= \alpha_t + \beta_t \lambda_t, \end{aligned}$$

where $\alpha_t = \alpha(I_{t-1})$ is the riskfree interest rate, $\lambda_t = \lambda(I_{t-1})$ is the market risk, and $\beta_t = \beta(I_{t-1})$ is the price of market risk, or the so-called investment beta factor.

Equivalently, we can write a regression equation for the asset return

$$\begin{aligned} R_t &= \alpha_t + \beta_t \lambda_t + \varepsilon_t, \text{ where} \\ E(\varepsilon_t | I_{t-1}) &= 0. \end{aligned}$$

A conventional CAPM usually assumes $\alpha_{t-1} = \alpha, \beta_t = \beta$ and use some proxies for λ_t.

Like in Chapter 4, no normality on $\{\varepsilon_t\}$ is imposed. Furthermore, no conditional homoskedasticity is imposed. We now allow that $\text{var}(\varepsilon_t | X_t)$ is a function of X_t. Because X_t may contain lagged $Y_{t-1}, Y_{t-2}, \cdots$, $\text{var}(\varepsilon_t | X_t)$ may change over time (e.g., volatility clustering). Volatility clustering is a well-known financial phenomenon where a large volatility today tends to be followed by another large volatility tomorrow, and a small volatility today tends to be followed by another small volatility tomorrow.

Although Assumptions 5.1–5.5 allow for temporal dependences between observations, we will still obtain the same asymptotic properties for the OLS estimator and related test procedures as in the i.i.d. case. Put it differently, all the large sample properties for the OLS and related tests established under the i.i.d. assumption in Chapter 4 remain applicable to time series observations with the stationary MDS assumptions for $\{X_t\varepsilon_t\}$. We now show that this is indeed the case in subsequent sections.

5.2 Consistency of OLS

We first investigate the consistency of OLS $\hat{\beta}$. Recall the OLS estimator

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

where, as before,

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

Substituting $Y_t = X_t'\beta^o + \varepsilon_t$ from Assumption 5.2, we have

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

Theorem: *Suppose Assumptions 5.1–5.5 hold. Then*

$$\hat{\beta} - \beta^o \xrightarrow{p} 0 \text{ as } n \to \infty.$$

Proof: Because $\{X_t\}$ is ergodic stationary, $\{X_t X_t'\}$ is also ergodic stationary. Thus, given Assumption 5.4, which implies $E|X_{it}X_{jt}| \leq C < \infty$ for $0 \leq i, j \leq k$ and for some constant C, we have

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

by the WLLN for ergodic stationary processes. Because Q^{-1} exists, by continuity we have

$$\hat{Q}^{-1} \xrightarrow{p} Q^{-1} \text{ as } n \to \infty.$$

Next, we consider $n^{-1}\sum_{t=1}^{n} X_t\varepsilon_t$. Because $\{Y_t, X_t'\}'$ is ergodic stationary, $\varepsilon_t = Y_t - X_t'\beta^o$ is ergodic stationary, and so is $X_t\varepsilon_t$. In addition,

$$E|X_{jt}\varepsilon_t| \leq \left[E(X_{jt}^2)E(\varepsilon_t^2)\right]^{1/2} \leq C < \infty \text{ for } 0 \leq j \leq k$$

by the Cauchy-Schwarz inequality and Assumptions 5.3 and 5.4. It follows that

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

by the WLLN for ergodic stationary processes, where

$$\begin{aligned} E(X_t\varepsilon_t) &= E[E(X_t\varepsilon_t|X_t)] \\ &= E[X_t E(\varepsilon_t|X_t)] \\ &= E(X_t \cdot 0) \\ &= 0 \end{aligned}$$

by the law of iterated expectations and Assumption 5.3. Therefore, we have

$$\hat{\beta} - \beta^o = \hat{Q}^{-1} n^{-1}\sum_{t=1}^{n} X_t\varepsilon_t \xrightarrow{p} Q^{-1}\cdot 0 = 0.$$

This completes the proof.

5.3 Asymptotic Normality of OLS

Next, we derive the asymptotic distribution of $\hat{\beta}$.

Theorem: *Suppose Assumptions 5.1–5.5 hold. Then*

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

Proof: Recall

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

First, we consider the second term

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

Because $\{Y_t, X_t'\}'$ is stationary ergodic, $X_t\varepsilon_t$ is also stationary ergodic. Also, $\{X_t\varepsilon_t\}$ is a MDS with $\text{var}(X_t\varepsilon_t) = E(X_tX_t'\varepsilon_t^2) = V$ being finite and positive definite (Assumption 5.5). By the CLT of stationary ergodic MDS processes, we have

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

Moreover, $\hat{Q}^{-1} \xrightarrow{p} Q^{-1}$, as shown earlier. It follows by the Slutsky theorem that

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

This completes the proof.

The asymptotic distribution of $\hat{\beta}$ under Assumptions 5.1–5.5 is exactly the same as that of $\hat{\beta}$ in Chapter 4. In particular, the asymptotic mean of $\sqrt{n}(\hat{\beta} - \beta^o)$ is 0, and the asymptotic variance of $\sqrt{n}(\hat{\beta} - \beta^o)$ is $Q^{-1} V Q^{-1}$; we denote

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

Special Case: Conditional Homoskedasticity

The asymptotic variance of $\sqrt{n}\hat{\beta}$ can be simplified if there exists conditional homoskedasticity.

Assumption 5.6: $E(\varepsilon_t^2 | X_t) = \sigma^2$ a.s.

This assumption rules out the possibility that the conditional variance of ε_t changes with X_t. For low-frequency macroeconomic time series, this might be a reasonable assumption. For high-frequency financial time series, however, this assumption will be rather restrictive.

Theorem: *Suppose Assumptions 5.1–5.6 hold. Then*

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

Proof: Under Assumption 5.6, we can simplify

$$\begin{aligned} V &= E(X_t X_t' \varepsilon_t^2) \\ &= E[E(X_t X_t' \varepsilon_t^2 | X_t)] \\ &= E[X_t X_t' E(\varepsilon_t^2 | X_t)] \\ &= \sigma^2 E(X_t X_t') \\ &= \sigma^2 Q. \end{aligned}$$

The desired results follow immediately from the previous theorem. This completes the proof.

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

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

This is rather convenient to estimate.

5.4 Asymptotic Variance Estimator

To construct confidence interval estimators or hypothesis test statistics, we need to estimate the asymptotic variance of $\sqrt{n}(\hat{\beta} - \beta^o)$, namely $\text{avar}(\sqrt{n}\hat{\beta})$. We consider consistent estimation for $\text{avar}(\sqrt{n}\hat{\beta})$ under conditional homoskedasticity and conditional heteroskedastcity respectively.

Case I: Conditional Homoskedasticity

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

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

It suffices to have consistent estimators for σ^2 and Q respectively.

Question: How to estimate Q?

Lemma: *Suppose Assumptions 5.1 and 5.3 hold. Then*

$$\hat{Q} \xrightarrow{p} Q \text{ as } n \to \infty.$$

Question: How to estimate σ^2?

To estimate the residual sample variance estimator

$$s^2 = \frac{e'e}{n-K}.$$

Theorem [Consistent Estimator for σ^2]: *Under Assumptions 5.1-5.5, as $n \to \infty$,*

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

Proof: The proof is analogous to the proof of Theorem in Chapter 4. We have

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

given that K is a fixed number, where we have made use of the WLLN for ergodic stationary processes in several places. This completes the proof.

We can then estimate $\text{avar}(\sqrt{n}\hat{\beta}) = \sigma^2 Q^{-1}$ by $s^2\hat{Q}^{-1}$.

Theorem: [Asymptotic Variance Estimator of $\hat{\beta}$]: *Under Assumptions 5.1-5.4, we can consistently estimate the asymptotic variance avar$(\sqrt{n}\hat{\beta})$ by*

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

This implies that the variance estimator of $\hat{\beta}$ is calculated as

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

which is the same as in the classical linear regression case.

CASE II: Conditional Heteroskedasticity

In this case,

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

cannot be further simplified.

Question: How to estimate $Q^{-1}VQ^{-1}$?

Question: It is straightforward to estimate Q by $\hat{Q}$. How to estimate $V = E(X_tX_t'\varepsilon_t^2)$?

We can use its sample analog

$$\hat{V} = n^{-1}\sum_{t=1}^{n} X_t X_t' e_t^2.$$

To ensure consistency of $\hat{V}$ for V, we impose the following moment condition:

Assumption 5.7: $E(X_{jt}^4) < \infty$ for $0 \le j \le k$ and $E(\varepsilon_t^4) < \infty$.

Lemma: *Suppose Assumptions 5.1–5.5 and 5.7 hold. Then*

$$\hat{V} \xrightarrow{p} V \text{ as } n \to \infty.$$

Proof: The proof is analogous to the proof of Theorem in Chapter 4. Because $e_t = \varepsilon_t - (\hat{\beta} - \beta^o)' X_t$, we have

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

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 for ergodic stationary processes and Assumption 5.5. For the second term, it suffices to show that for any combination (i, j, l, m), where $0 \le i, j, l, m \le k$,

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

which follows by $\hat{\beta} - \beta^o \xrightarrow{p} 0$ and $n^{-1}\sum_{t=1}^{n} X_{it} X_{jt} X_{lt} X_{mt} \to^p E(X_{it} X_{jt} X_{lt} X_{mt}) = O(1)$ by the WLLN and Assumption 5.7.

For the last term, it suffices to show

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

which follows from $\hat{\beta}-\beta^o \rightarrow^p 0$, $n^{-1}\sum_{t=1}^{n} X_{it}X_{jt}X_{lt}\varepsilon_t \xrightarrow{p} E(X_{it}X_{jt}X_{lt}\varepsilon_t) = 0$ by the WLLN for ergodic stationary processes, the law of iterated expectations, and $E(\varepsilon_t|X_t) = 0$ a.s.

We have proved the following result.

Theorem [Asymptotic variance estimator for $\sqrt{n}(\hat{\beta}-\beta^o)$]: *Under Assumptions 5.1–5.5 and 5.7, we can estimate avar*$(\sqrt{n}\hat{\beta})$ *by*

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

The variance estimator $\hat{Q}^{-1}\hat{V}\hat{Q}^{-1}$ is the so-called White's heteroskedasticity-consistent variance-covariance matrix of estimator $\sqrt{n}(\hat{\beta}-\beta^o)$ in a linear time series regression model with MDS disturbances.

5.5 Hypothesis Testing

Question: How to construct a test 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?

Because

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

we have under $\mathbf{H}_0$,

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

When $E(\varepsilon_t^2|X_t) = \sigma^2$ a.s., we have $V = \sigma^2 Q$, and so

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

The test statistics differ in two cases. We first construct a test under conditional homoskedasticity.

Case I: Conditional Homoskedasticity

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

Theorem [t-test]: *Suppose Assumptions 5.1-5.6 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

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

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 can consider an asymptotic χ^2 test that is based on the conventional F-statistic.

Theorem [Asymptotic χ^2 Test]: *Suppose Assumptions 5.1-5.6 hold. Then under* $\mathbf{H}_0$,

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

as $n \to \infty$.

Proof: We write

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

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

$$\begin{aligned} \sqrt{n}(R\hat{\beta} - r) &= R\sqrt{n}(\hat{\beta} - \beta^o) \\ &\xrightarrow{d} N(0, \sigma^2 RQ^{-1}R'). \end{aligned}$$

It follows that the quadratic form

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

Also, because $s^2\hat{Q}^{-1} \xrightarrow{p} \sigma^2 Q^{-1}$, we have the Wald test statistic

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

by the Slutsky theorem. This can be written equivalently as follows:

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

namely

$$W = J \cdot F \xrightarrow{d} \chi^2_J,$$

where F is the conventional F-test statistic derived in Chapter 3.

Remarks:

We cannot use the F distribution for a finite sample size n, 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 as $n \to \infty$. That is,

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

Put it differently, the classical F-test is still approximately applicable under Assumptions 5.1–5.5 for a large n.

We now give two examples that are not covered under the assumptions of classical linear regression models.

Example 1 [Testing for Granger Causality]: Consider a bivariate time series $\{Y_t, X_t\}$, where t is the time index, $I_{t-1}^{(Y)} = \{Y_{t-1}, ..., Y_1\}$ and $I_{t-1}^{(X)} = \{X_{t-1}, ..., X_1\}$. For example, Y_t is the GDP growth, and X_t is the money supply growth. We say that X_t does not Granger-cause Y_t in conditional mean with respect to $I_{t-1} = \{I_{t-1}^{(Y)}, I_{t-1}^{(X)}\}$ if

$$E(Y_t|I_{t-1}^{(Y)}, I_{t-1}^{(X)}) = E(Y_t|I_{t-1}^{(Y)}).$$

In other words, the lagged variables of X_t have no impact on the level of Y_t.

Granger causality is defined in terms of incremental predictability rather than the real cause-effect relationship. From an econometric point of view, it is a test of omitted variables in a time series context. It is first introduced by Granger (1969).

Question: How to test Granger causality?

We consider two approaches to testing Granger causality. The first test is proposed by Granger (1969). Consider now a linear regression model

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 Y_{t-1} + \cdots + \beta_p Y_{t-p} \\ &\quad + \beta_{p+1} X_{t-1} + \cdots + \beta_{p+q} X_{t-q} + \varepsilon_t. \end{aligned}$$

Under non-Granger causality, we have

$$\mathbf{H}_0 : \beta_{p+1} = \cdots = \beta_{p+q} = 0.$$

The F-test statistic

$$F \sim F_{q,n-(p+q+1)}.$$

The classical regression theory of Chapter 3 (Assumption 3.2: $E(\varepsilon_t|\mathbf{X}) = 0$) rules out this application, because it is a dynamic regression model. However, we have justified in this chapter that under H_0,

$$q \cdot F \xrightarrow{d} \chi^2_q$$

as $n \to \infty$ under conditional homoskedasticity even for a linear dynamic regression model.

There is another well-known test for Granger causality proposed by Sims (1980), which is based on the fact that the future cannot cause the present in any notion of causality. To test whether $\{X_t\}$ Granger-causes $\{Y_t\}$, we consider the following linear regression model

$$X_t = \alpha_0 + \sum_{j=1}^{p} \alpha_j X_{t-j} + \sum_{j=1}^{J} \beta_j Y_{t+j} + \sum_{j=1}^{q} \gamma_j Y_{t-j} + \varepsilon_t.$$

Here, the dependent variable is X_t rather than Y_t. If $\{X_t\}$ Granger-causes $\{Y_t\}$, we expect some relationship between the current X_t and the future values of Y_t. Note that nonzero values for any of $\{\beta_j\}_{j=1}^J$ cannot be interpreted as causality from the future values of Y_t to the current X_t, simply because the future cannot cause the present. Nonzero values of any β_j must imply that there exists causality from current X_t to future values of Y_t. Therefore, we test the null hypothesis

$$\mathbf{H}_0 : \beta_j = 0 \text{ for } 1 \leq j \leq J.$$

Let F be the associated F-test statistic. Then under H_0,

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

as $n \to \infty$ under conditional homoskedasticity.

Example 2 [Wage Determination]: Consider the wage function

$$\begin{aligned} W_t &= \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + \beta_3 U_t \\ &\quad + \beta_4 V_t + \beta_5 W_{t-1} + \varepsilon_t, \end{aligned}$$

where W_t = wage, P_t = price, U_t = unemployment, and V_t = unfilled vacancies. We will test the null hypothesis

$$\mathbf{H}_0 : \beta_1 + \beta_2 = 0, \beta_3 + \beta_4 = 0, \text{ and } \beta_5 = 1.$$

Question: What is the economic interpretation of the null hypothesis $\mathbf{H}_0$?

Under $\mathbf{H}_0$, we have the restricted wage equation:

$$\Delta W_t = \beta_0 + \beta_1 \Delta P_t + \beta_4 D_t + \varepsilon_t,$$

where $\Delta W_t = W_t - W_{t-1}$ is the wage growth rate, $\Delta P_t = P_t - P_{t-1}$ is the inflation rate, and $D_t = V_t - U_t$ is an index for job market situation (excess job supply). This implies that the wage increase depends on the inflation rate and the excess labor supply.

Under $\mathbf{H}_0$, we have

$$3F \to^d \chi^2_3.$$

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

Theorem $[(n-K)R^2$ **Test]:** *Suppose Assumption 5.1-5.6 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_j^o, $1 \leq j \leq k$, *are the slope coefficients in the linear regression model* $Y_t = X_t'\beta^o + \varepsilon_t$.

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

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

Then under $\mathbf{H}_0$,

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

Proof: First, note that as shown earlier, we have in this case,

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

Here, we have $J = k$, and under $\mathbf{H}_0$,

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

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 fixed (i.e., does not grow with the sample size n), we have

$$R^2/(1-R^2) \xrightarrow{p} 0$$

or equivalently,

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

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

$$\begin{aligned} (n-K)R^2 &= \frac{(n-K)R^2}{1-R^2} \cdot (1-R^2) \\ &\to \chi^2_k. \end{aligned}$$

This completes the proof. ■

Example 3 [Efficient Market Hypothesis]: Suppose Y_t is the exchange rate return in period t, and I_{t-1} is the information available at time $t-1$. Then a classical version of the efficient market hypothesis (EMH) can be stated as follows:

$$E(Y_t|I_{t-1}) = E(Y_t)$$

To check whether exchange rate changes are unpredictable using the past history of exchange rate changes, we specify a linear regression model:

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

where

$$X_t = (1, Y_{t-1}, ..., Y_{t-k})'.$$

Under EMH, we have

$$\mathbf{H}_0 : \beta_j^o = 0 \text{ for all } j = 1, ..., k.$$

If the alternative

$$\mathbb{H}_A : \beta_j^o \neq 0 \text{ at least for some } j \in \{1, ..., k\}$$

holds, then exchange rate changes are predictable using the past information.

Remarks:

What is the appropriate interpretation if $\mathbf{H}_0$ is not rejected? Note that there exists a gap between the efficiency hypothesis and $\mathbf{H}_0$, because the linear regression model is just one of many ways to check EMH. Thus, $\mathbf{H}_0$ is not rejected, at most we can only say that no evidence against the efficiency hypothesis is found. We should not conclude that EMH holds.

In using $k \cdot F$ or $(n-K)R^2$ statistic to test EMH, although the normality assumption is not needed for this result, we still require conditional homoskedasticity, which rules out autoregressive conditional heteroskedasticity (ARCH) in the dynamic time series regression framework. ARCH effects arise in high-frequency financial time series processes.

Case II: Conditional Heteroskedasticity

Next, we construct hypothesis tests for $\mathbf{H}_0$ under conditional heteroskedasticity. Recall that under $\mathbf{H}_0$,

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

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

For $J = 1$, we have

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

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

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

Theorem [Robust t-Test Under Conditional Heteroskedasticity] *Suppose Assumptions 5.1–5.5 and 5.7 hold. Then under* $\mathbf{H}_0$ with $J = 1$, *as* $n \to \infty$, the robust t-test statistic

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

For $J > 1$, the quadratic form

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

under $\mathbf{H}_0$. Given $\hat{Q} \xrightarrow{p} Q$ and $\hat{V} \xrightarrow{p} V$, where $\hat{V} = \mathbf{X}'\mathrm{D}(e)\mathrm{D}(e)'\mathbf{X}/n$, we have a robust Wald test statistic

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

by the Slutsky theorem. We can equivalently write

$$W = (R\hat{\beta} - r)'[R(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathrm{D}(e)\mathrm{D}(e)'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}R']^{-1}(R\hat{\beta} - r) \to^d \chi^2_J.$$

Theorem [Robust Wald Test Under Conditional Heteroskedasticity] *Suppose Assumptions 5.1–5.5 and 5.7 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, $J \cdot F$ and $(n - K)R^2$ cannot be used even when $n \to \infty$.

On the other hand, although the general form of the test statistic W developed here can be used no matter whether there exists conditional homoskedasticity, W may perform poorly in small samples (i.e., the asymptotic χ^2_J approximation may be poor in small samples, or Type I errors are large). 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 in a time series context.

5.6 Testing for Conditional Heteroskedasticity and Autoregressive Conditional Heteroskedasticity

Question: How to test conditional heteroskedasticity in a time series regression context?

Question: Can we still use White's (1980) test for conditional heteroskedasticity?

Yes. Although White's (1980) test is developed under the independence assumption, it is still applicable to a time series linear regression model when $\{X_t\varepsilon_t\}$ is an MDS process. Thus, the test procedure to implement White's (1980) test as is discussed in Chapter 4 can be used here.

In the time series econometrics, there is an alternative approach to testing conditional heteroskedasticity in an autoregressive time series context. This is Engle's (1982, *Econometrica*) Lagrange Multiplier test for autoregressive conditional heterosekedasticity (ARCH) in $\{\varepsilon_t\}$.

Consider the regression model

$$\begin{aligned} Y_t &= X_t'\beta^o + \varepsilon_t, \\ \varepsilon_t &= \sigma_t z_t, \\ \{z_t\} &\sim i.i.d.(0,1). \end{aligned}$$

The null hypothesis

$$\mathbf{H}_0 : \sigma_t^2 = \sigma^2 \text{ for some } \sigma^2 > 0.$$

where $I_{t-1} = \{\varepsilon_{t-1}, \varepsilon_{t-2}, ...\}$.

Here, to allow for a possibly time-varying conditional variance of the rergression disturbance ε_t given I_{t-1}, ε_t is formulated as the product between a random shock z_t and $\sigma_t = \sigma(I_{t-1})$. When the random shock series $\{z_t\}$ is i.i.d.$(0,1)$, we have

$$\begin{aligned} \text{var}(\varepsilon_t|I_{t-1}) &= E(z_t^2\sigma_t^2|I_{t-1}) \\ &= \sigma_t^2 E(z_t^2|I_{t-1}) \\ &= \sigma_t^2. \end{aligned}$$

That is, σ_t^2 is the conditional variance of ε_t given I_{t-1}. The null hypothesis $\mathbf{H}_0$ says that the conditional variance of ε_t given I_{t-1} does not change over time.

The alternative hypothesis to $\mathbf{H}_0$ is that σ_t^2 is a function of I_{t-1}, so it changes over time. In particular, we consider the following auxiliary regression for ε_t^2 :

$$\varepsilon_t^2 = \alpha_0 + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 + v_t,$$

where $E(v_t|I_{t-1}) = 0$ a.s. This is called an ARCH(q) process in Engle (1982). ARCH models can capture a well-known empirical styles fact called volatility clustering in financial markets, that is, a high volatility today tends to be followed by another large volatility tomorrow, and a small volatility today tends to be followed by another small volatility tomorrow, and such patterns alternate over time. To see this more clearly, we consider an ARCH(1) model where

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2,$$

where, to ensure nonnegativity of σ_t^2, both α_0 and α_1 are required to be nonnegative parameters. Suppose $\alpha_1 > 0$. Then if ε_{t-1} is an unusually large deviation from its expectation of 0 so that ε_{t-1}^2 is large, then the conditional variance of ε_t is larger than usual. Therefore, ε_t is expected to have an unusually large deviation from its mean of 0, with either direction. Similarly, if ε_{t-1}^2 is usually small. then σ_t^2 is small, and ε_t^2 is expected to be small as well. Because of this behavior, volatility clustering arises.

In addition to volatility clustering, the ARCH(1) model can also generate heavy tails for ε_t even when the random shock z_t is i.i.d.$N(0,1)$. This can be seen from its kurtosis

$$\begin{aligned} K &= \frac{E(\varepsilon_t^4)}{[E(\varepsilon_t^2)]^2} \\ &= \frac{E(z_t^4)(1-\alpha_1^2)}{(1-3\alpha_1^2)} \\ &> 3 \end{aligned}$$

given $\alpha_1 > 0$.

With an ARCH modelling framework, all autoregressive coefficients $\alpha_j, 1 \leq j \leq q$, are identically zero when $\mathbf{H}_0$ holds. Thus, we can test $\mathbf{H}_0$ by checking whether all $\alpha_j, 1 \leq j \leq q$, are jointly zero. If $\alpha_j \neq 0$ for some $1 \leq j \leq q$, then there exists autocorrelation in $\{\varepsilon_t^2\}$ and $\mathbf{H}_0$ is false.

Observe that with $\varepsilon_t = \sigma_t z_t$ and $\{z_t\}$ is i.i.d.(0,1), the disturbance v_t in the auxiliary autoregression model is an i.i.d. sequence under $\mathbf{H}_0$, which implies that $E(v_t^2|I_{t-1}) = \sigma_v^2$, that is, $\{v_t\}$ is conditionally homoskedastic. Thus, when $\mathbf{H}_0$ holds, we have

$$(n-q-1)\tilde{R}^2 \to^d \chi_q^2,$$

where $\tilde{R}^2$ is the centered R^2 from the auxiliary regression.

The auxiliary regression for ε_t^2, unfortunately, is infeasible because ε_t is not observable. However, we can replace ε_t by the estimated residual e_t and consider the regression

$$e_t^2 = \alpha_0 + \sum_{j=1}^{q} \alpha_j e_{t-j}^2 + \tilde{v}_t.$$

Then we have

$$(n - q - 1)R^2 \to^d \chi_q^2.$$

Note that the replacement of ε_t by e_t has no impact on the asymptotic distribution of the test statistic, for the same reason as in White's (1980) direct test for conditional heteroskedasticity. See Chapter 4 for more discussions.

Remarks:

The existence of ARCH effect for $\{\varepsilon_t\}$ does not automatically imply that we have to use White's heteroskedasticity-consistent variance-covariance matrix $Q^{-1}VQ^{-1}$ for the OLS estimator $\hat{\beta}$. Suppose $Y_t = X_t'\beta^o + \varepsilon_t$ is a static time series model such that the two time series $\{X_t\}$ and $\{\varepsilon_t\}$ are independent of each other, and $\{\varepsilon_t\}$ displays ARCH effect, i.e.,

$$var(\varepsilon_t|I_{t-1}) = \alpha_0 + \sum_{j=1}^{p} \alpha_j \varepsilon_{t-j}^2$$

with at least some $\alpha_j \neq 0$. Then Assumption 5.6 still holds because $\mathrm{var}(\varepsilon_t|X_t) = \mathrm{var}(\varepsilon_t) = \sigma^2$ given the assumption that $\{X_t\}$ and $\{\varepsilon_t\}$ are independent. In this case, we have $\mathrm{avar}(\sqrt{n}\hat{\beta}) = \sigma^2 Q^{-1}$.

Next, suppose $Y_t = X_t'\beta^o + \varepsilon_t$ is a dynamic time series regression model such that X_t contains some lagged dependent variables (say Y_{t-1}). Then if $\{\varepsilon_t\}$ displays ARCH effect, Assumption 5.6 may fail because we may have $E(\varepsilon_t^2|X_t) \neq \sigma^2$, which generally occurs when X_t and $\{\varepsilon_{t-j}^2, j = 1, ..., p\}$ are not independent. In this case, we have to use $\mathrm{avar}(\sqrt{n}\hat{\beta}) = Q^{-1}VQ^{-1}$.

5.7 Testing Serial Correlation

Question: Why is it important to test serial correlation for $\{\varepsilon_t\}$?

We first provide some motivation for doing so. Recall that under Assumptions 5.1–5.5,

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

where V =var$(X_t\varepsilon_t)$. Among other things, this implies that the asymptotic variance of $n^{-1/2}\sum_{t=1}^{n} X_t\varepsilon_t$ is the same as the variance of $X_t\varepsilon_t$. This follows from the MDS assumption for $\{X_t\varepsilon_t\}$:

$$\begin{aligned}
& \text{var}\left(n^{-1/2}\sum_{t=1}^{n} X_t\varepsilon_t\right) \\
= \; & n^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n} E(X_t\varepsilon_t X_s'\varepsilon_s) \\
= \; & n^{-1}\sum_{t=1}^{n} E(X_t X_t'\varepsilon_t^2) \\
= \; & E(X_t X_t'\varepsilon_t^2) \\
= \; & V.
\end{aligned}$$

This result will not generally hold if the MDS property for $\{X_t\varepsilon_t\}$ is violated.

Question: How to check $E(X_t\varepsilon_t|I_{t-1}) = 0$, where I_{t-1} is the σ-field generated by $\{X_s\varepsilon_s, s < t\}$?

When X_t contains the intercept, we have that $\{\varepsilon_t\}$ is MDS with respect to the σ-field generated by $\{\varepsilon_s, s < t\}$, which implies that $\{\varepsilon_t\}$ is serially uncorrelated (or is a white noise).

If $\{\varepsilon_t\}$ is serially correlated, then $\{X_t\varepsilon_t\}$ will not be MDS, and consequently we will generally have var$(n^{-1/2}\sum_{t=1}^{n} X_t\varepsilon_t) \neq V$. Therefore, serial uncorrelatedness is an important necessary condition for the validity of avar$(\sqrt{n}\hat{\beta}) = Q^{-1}VQ^{-1}$ with $V = E(X_t X_t'\varepsilon_t^2)$.

On the other hand, let us revisit the correct model specification condition that

$$E(\varepsilon_t|X_t) = 0 \text{ a.s.}$$

in a time series context. Note that this condition does not necessarily imply that $\{\varepsilon_t\}$ or $\{X_t\varepsilon_t\}$ is MDS in a time series context.

To see this, consider the case when $Y_t = X_t'\beta^o + \varepsilon_t$ is a static regression model (i.e., when $\{X_t\}$ and $\{\varepsilon_t\}$ are mutually independent, or at least when cov$(X_t, \varepsilon_s) = 0$ for all t, s), it is possible that $E(\varepsilon_t|X_t) = 0$ but $\{\varepsilon_t\}$ is serially correlated. An example is that $\{\varepsilon_t\}$ is an AR(1) process but $\{\varepsilon_t\}$ and $\{X_t\}$ are mutually independent. In this case, serial dependence in $\{\varepsilon_t\}$ does not cause inconsistency of OLS $\hat{\beta}$ to β^o, but we no longer have $\mathrm{var}(n^{-1/2}\Sigma_{t=1}^{n}X_t\varepsilon_t) = V = E(X_tX_t'\varepsilon_t^2)$. In other words, the MDS property for $\{\varepsilon_t\}$ is crucial for $\mathrm{var}(n^{-1/2}\Sigma_{t=1}^{n}X_t\varepsilon_t) = V$ in a static regression model, although it is not needed to ensure $E(\varepsilon_t|X_t) = 0$. For a static regression model, the regressors X_t are usually called exogenous variables. In particular, if $\{X_t\}$ and $\{\varepsilon_t\}$ are mutually independent, then X_t is called strictly exogenous.

On the other hand, when $Y_t = X_t'\beta^o + \varepsilon_t$ is a dynamic model (i.e., when X_t includes lagged dependent variables such as $\{Y_{t-1}, \cdots, Y_{t-k}\}$ so that X_t and ε_{t-j} are generally not independent for $j > 0$), the correct model specification condition

$$E(\varepsilon_t|X_t) = 0 \text{ a.s.}$$

holds when $\{\varepsilon_t\}$ is MDS. If $\{\varepsilon_t\}$ is not an MDS, the condition that $E(\varepsilon_t|X_t) = 0$ a.s. generally does not hold. To see this, we consider, for example, an AR(1) model

$$\begin{aligned} Y_t &= \beta_0^o + \beta_1^o Y_{t-1} + \varepsilon_t \\ &= X_t'\beta^o + \varepsilon_t. \end{aligned}$$

Suppose $\{\varepsilon_t\}$ is an MA(1) process. Then $E(X_t\varepsilon_t) \neq 0$, and so $E(\varepsilon_t|X_t) \neq 0$. Thus, to ensure correct specification $(E(Y_t|X_t) = X_t'\beta^o \text{ a.s.})$ of a dynamic regression model in a time series context, it is important to check MDS for $\{\varepsilon_t\}$. In this case, tests for MDS can be viewed as specification tests for dynamic regression models.

In time series econometrics such as rational expectations econometrics, correct model specification usually requires that ε_t be MDS:

$$E(\varepsilon_t|I_{t-1}) = 0 \text{ a.s.}$$

where I_{t-1} is the information set available to the economic agent at time $t-1$. In this content, X_t is usually a subset of I_{t-1}, namely $X_t \in I_{t-1}$. Thus both Assumptions 5.3 and 5.5 hold simultaneously:

$$\begin{aligned} E(\varepsilon_t|X_t) &= E[E(\varepsilon_t|I_{t-1})|X_t] = 0 \text{ a.s.} \\ &\text{and} \\ E(X_t\varepsilon_t|I_{t-1}) &= X_tE(\varepsilon_t|I_{t-1}) = 0 \text{ a.s.} \end{aligned}$$

because X_t belongs to I_{t-1}.

To check the MDS property of $\{\varepsilon_t\}$, one may check whether there exists serial correlation in $\{\varepsilon_t\}$. Evidence of serial correlation in $\{\varepsilon_t\}$ will indicate that $\{\varepsilon_t\}$ is not

MDS. The existence of serial correlation may be due to various sources of model misspecfication. For example, it may be that in the linear regression model, an important explanatory variable is missing (omitted variables), or that the functional relationship is nonlinear (functional form misspecification), or that lagged dependent variables or lagged explanatory variables should be included as regressors (neglected dynamics or dynamic misspecification). Therefore, tests for serial correlation can also be viewed as a model specification check in a dynamic time series regression context.

Question: How to check serial dependence in $\{\varepsilon_t\}$?

We now introduce a number of tests for serial correlation of the disturbance $\{\varepsilon_t\}$ in a linear regression model.

Breusch and Godfrey's Lagrange Multiplier Test for Serial Correlation

The null hypothesis

$$\mathbf{H}_0 : E(\varepsilon_t|I_{t-1}) = 0,$$

where ε_t is the regression error in the linear regression model

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

$I_{t-1} = \{\varepsilon_{t-1}, \varepsilon_{t-2}, ...\}$, and $E(\varepsilon_t^2|X_t) = \sigma^2$ a.s.

Below, following the vast literature, we will first assume conditional homoskedasticity in testing serial correlation for $\{\varepsilon_t\}$. Thus, this method is not suitable for high-frequency financial time series, where volatility clustering has been well-documented. Extensions to conditional heteroskedasticity will be discussed later.

First, suppose ε_t is observed, and we consider the auxiliary regression model (an AR(p))

$$\begin{aligned} \varepsilon_t &= \sum_{j=1}^{p} \alpha_j \varepsilon_{t-j} + u_t, \\ t &= p+1, \cdots, n, \end{aligned}$$

where $\{u_t\}$ is MDS. Under $\mathbf{H}_0$, we have $\alpha_j = 0$ for $1 \le j \le p$. Thus, we can test $\mathbf{H}_0$ by checking whether the α_j are jointly equal to 0. Assuming $E(\varepsilon_t^2|X_t) = \sigma^2$ (which implies $E(u_t^2|X_t) = \sigma^2$ under $\mathbf{H}_0$), then we can run an OLS regression and obtain

$$(n-2p)\tilde{R}^2_{uc} \xrightarrow{d} \chi^2_p,$$

where $\tilde{R}^2_{uc}$ is the uncentered R^2 in the auxiliary regression (note that there is no intercept), and p is the number of the regressors.
Unfortunately, ε_t is not observable. However, we can replace ε_t with the estimated residual $e_t = Y_t - X_t'\hat{\beta}$. Unlike White's (1980) test for heteroskedasticity of unknown form, this replacement will generally change the asymptotic χ^2_p distribution for $(n-2p)R^2_{uc}$ here. To remove the impact of the estimation error $X_t'(\hat{\beta}-\beta^o)$, we have to modify the auxiliary regression as follows:

$$\begin{aligned}
e_t &= \sum_{j=1}^{K} \gamma_j X_{jt} + \sum_{j=1}^{p} \alpha_j e_{t-j} + u_t \\
&= \gamma' X_t + \sum_{j=1}^{p} \alpha_j e_{t-j} + u_t, \\
t &= p+1, \cdots, n,
\end{aligned}$$

where X_t contains the intercept. The inclusion of the regressors X_t in the auxiliary regression will purge the impact of the estimation error $X_t'(\hat{\beta}-\beta^o)$ of the test statistic, because X_t and $X_t'(\hat{\beta}-\beta^o)$ are perfectly correlated. Therefore, the resulting statistic

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

under $\mathbf{H}_0$, where R^2 is the centered squared multi-correlation coefficient in the feasible auxiliary regression model.

Question: Why should X_t be generally included in the auxiliary regression?

When we replace ε_t by $e_t = \varepsilon_t - X_t'(\hat{\beta}-\beta^o)$, the estimation error $X_t'(\hat{\beta}-\beta^o)$ will have nontrivial impact on the asymptotic distribution of a test statistic for $\mathbf{H}_0$, because X_t may be correlated with ε_{t-j} at least for some lag order $j > 0$ (this occurs when the regression model is dynamic). To remove the impact of $X_t'(\hat{\beta}-\beta^o)$, we have to add the regressor X_t in the auxiliary regression, which is perfectly correlated with the estimation error $X_t'(\hat{\beta}-\beta^o)$, and thus can extract its impact. This can be proven rigorously but we do not attempt to do so here, because it would be very tedious and offer no much new insight than the above intuition. Below, we provide a heuristic explanation.

First, we consider the infeasible auxiliary autoregression. Under the null hypothesis of no serial correlation, the OLS estimator

$$\sqrt{n}(\tilde{\alpha}-\alpha^0) = \sqrt{n}\tilde{\alpha}$$

converges to an asymptotic normal distribution, which implies $\tilde{\alpha} = O_P(n^{-1/2})$ vanishes in probability at a rate of $n^{-1/2}$. The test statistic $n\tilde{R}^2_{uc}$ is asymptotically equivalent to a quadratic form in $\sqrt{n}\tilde{\alpha}$ which follows an asymptotic χ^2_p distribution. In other words, the asymptotic distribution of $n\tilde{R}^2_{uc}$ is determined by the asymptotic distribution of $\sqrt{n}\tilde{\alpha}$.

Now, suppose we replace ε_t by $e_t = \varepsilon_t - (\hat{\beta} - \beta^o)'X_t$, and consider the feasible autoregression

$$e_t = \sum_{j=1}^{p} \alpha_j e_{t-j} + v_t.$$

Suppose the OLS estimator is $\hat{\alpha}$. We can then decompose

$$\hat{\alpha} = \tilde{\alpha} + \hat{\delta} + \text{ reminder term},$$

where $\tilde{\alpha}$, as discussed above, is the OLS estimator of regressing ε_t on $\varepsilon_{t-1}, ..., \varepsilon_{t-p}$, and $\hat{\delta}$ is the OLS estimator of regressing $(\hat{\beta} - \beta^o)'X_t$ on $\varepsilon_{t-1}, ..., \varepsilon_{t-p}$. For a dynamic regression model, the regressor X_t contains lagged dependent variables and so $E(X_t\varepsilon_{t-j})$ is likely nonzero for some $j \in \{1, ..., p\}$. It follows that $\hat{\delta}$ will converge to zero at the same rate as $\tilde{\alpha} - \alpha^0$, which is $n^{-1/2}$. Because $\hat{\delta} \to^p 0$ at the same rate as $\tilde{\alpha}$, $\hat{\delta}$ will have impact on the asymptotic distribution of nR^2_{uc}, where R^2_{uc} is the uncentered R^2 in the auxiliary autoregression. To remove the impact of $\hat{\delta}$, we need to include X_t as additional regressors in the auxiliary regression.

Question: When do we need not include X_t in the auxiliary regression?

Answer: When we have a static regression model, $\text{cov}(X_t, \varepsilon_s) = 0$ for all t, s (so $E(X_t\varepsilon_{t-j}) = 0$ for all $j = 1, ..., p$), the estimation error $X_t'(\hat{\beta} - \beta^o)$ has no impact on the asymptotic distribution of nR^2_{uc}. It follows that we do not need to include X_t in the auxiliary autoregression. In other words, we can test serial correlation for $\{\varepsilon_t\}$ by running the following auxiliary regression model

$$e_t = \sum_{j=1}^{p} \alpha_j e_{t-j} + u_t.$$

The resulting nR^2_{uc} is asymptotically χ^2_p under the null hypothesis of no serial correlation.

Question: Suppose we have a static regression model, and we include X_t in the auxiliary regression in testing serial correlation of $\{\varepsilon_t\}$. What will happen?

For a static regression model, whether X_t is included in the auxiliary regression has no impact on the asymptotic χ^2_p distribution of $(n-2p)R^2_{uc}$ or $(n-2p)R^2$ under the null hypothesis of no serial correlation in $\{\varepsilon_t\}$. Thus, we will still obtain an asymptotic valid test statistic $(n-2p)R^2$ under $\mathbf{H}_0$. In fact, the size performance of the test can be better in finite samples. However, the test may be less powerful than the test without including X_t, because X_t may take away some serial correlation in $\{\varepsilon_t\}$ under the alternative to $\mathbf{H}_0$.

Question: What happens if we include an intercept in the auxiliary regression

$$e_t = \alpha_0 + \sum_{j=1}^{p} \alpha_j e_{t-j} + u_t,$$

where e_t is the OLS residual from a static regression model.

With the inclusion of the intercept here, we can then use $(n-2p)R^2$ to test serial correlation in $\{\varepsilon_t\}$, which is more convenient to compute than $(n-2p)R^2_{uc}$. (Most statistical software report R^2 but not R^2_{uc}.) Under $\mathbf{H}_0$, $(n-2p)R^2 \to^d \chi^2_p$. However, the inclusion of the intercept α_0 may have some adverse impact on the power of the test in small samples, because there is an additional parameter to estimate.

As discussed at the beginning of this section, a test for serial correlation can be viewed as a specification test for dynamic regression models in a time series context, because existence of serial correlation in the estimated model residual $\{e_t\}$ will generally indicate misspecification of a dynamic regression model.

On the other hand, for static regression models with time series observations, it is possible that a static regression model $Y_t = X_t'\beta^o + \varepsilon_t$ is correctly specified in the sense that $E(\varepsilon_t|X_t) = 0$ but $\{\varepsilon_t\}$ displays serial correlation. In this case, existence of serial correlation in $\{\varepsilon_t\}$ does not affect the consistency of the OLS estimator $\hat{\beta}$ but affects the asymptotic variance and therefore the efficiency of the OLS estimator $\hat{\beta}$. However, since ε_t is unobservable, one has to use the estimated residual e_t in testing for serial correlation in a static regression model in the same way as in a dynamic regression model. Because the estimated residual

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

it contains the true disturbance $\varepsilon_t = Y_t - E(Y_t|X_t)$ and model approximation error $E(Y_t|X_t) - X_t'\beta^*$, where $\beta^* = [E(X_tX_t')]^{-1}E(X_tY_t)$ is the best linear least squares approximation coefficient which the OLS $\hat{\beta}$ always converges to as $n \to \infty$. If the linear regression model is misspecified for $E(Y_t|X_t)$, then the approximation error $E(Y_t|X_t) - X_t'\beta^*$

will never vanish to zero and this term can cause serial correlation in e_t if X_t is a time series process. Thus, when one finds that there exists serial correlation in the estimated residuals $\{e_t\}$ of a static linear regression model, it is also likely due to the misspecification of the static regression model. In this case, the OLS estimator $\hat{\beta}$ is generally not consistent. Therefore, one has to first check correct specification of a static regression model in order to give correct interpretation of any documented serial correlation in the estimated residuals.

In the development of tests for serial correlation in regression disturbances, there have been two very popular tests that have historical importance. One is the Durbin-Watson test and the other is Durbin's h test. The Durbin-Watson test is the first formal procedure developed for testing first order serial correlation

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t, \qquad \{u_t\} \sim i.i.d.\left(0, \sigma^2\right),$$

using the OLS residuals $\{e_t\}_{t=1}^n$ in a static linear regression model $Y_t = X_t'\beta^o + \varepsilon_t$. Durbin and Watson (1950,1951) propose a test statistic

$$d = \frac{\sum_{t=2}^n \left(e_t - e_{t-1}\right)^2}{\sum_{t=1}^n e_t^2}.$$

Durbin and Watson present tables of bounds at the 0.05, 0.025 and 0.01 significance levels of the d statistic for static regressions with an intercept. Against the one-sided alternative that $\rho > 0$, if d is less than the lower bound d_L, the null hypothesis that $\rho = 0$ is rejected; if d is greater than the upper bound d_U, the null hypothesis is accepted. Otherwise, the test is equivocal. Against the one-sided alternative that $\rho < 0$, $4 - d$ can be used to replace d in the above procedure.

The Durbin-Watson test has been extended to test for lag 4 autocorrelation by Wallis (1972) and for autocorrelation at any lag by Vinod (1973).

The Durbin-Watson d test is not applicable to dynamic linear regression models, because parameter estimation uncertainty in the OLS estimator $\hat{\beta}$ will have nontrivial impact on the distribution of d. Durbin (1970) developed the so-called h test for first-order autocorrelation in $\{\varepsilon_t\}$ that takes into account parameter estimation uncertainty in $\hat{\beta}$. Consider a simple dynamic linear regression model

$$Y_t = \beta_0^o + \beta_1^o Y_{t-1} + \beta_2^o X_t + \varepsilon_t,$$

where X_t is strictly exogenous. Durbin's h statistic is defined as:

$$h = \hat{\rho}\sqrt{\frac{n}{1 - n \cdot \hat{\mathrm{var}}(\hat{\beta}_1)}},$$

where $\hat{\text{var}}(\hat{\beta}_1)$ is an estimator for the asymptotic variance of $\hat{\beta}_1$, $\hat{\rho}$ is the OLS estimator from regressing e_t on e_{t-1} (in fact, $\hat{\rho} \approx 1 - d/2$). Durbin (1970) shows that $h \xrightarrow{d} N(0,1)$ as $n \to \infty$ under null hypothesis that $\rho = 0$. In fact, Durbin's h test is asymptotically equivalent to the Lagrange multiplier test introduced above.

The Box-Pierce Portmanteau Test

Define the sample autocovariance function

$$\hat{\gamma}(j) = n^{-1} \sum_{t=j+1}^{n} (e_t - \bar{e})(e_{t-j} - \bar{e}),$$

where $\bar{e} = n^{-1} \sum_{t=1}^{n} e_t$ (this is zero when X_t contains an intercept). The Box-Pierce portmanteau test statistic is defined as

$$Q(p) = n \sum_{j=1}^{p} \hat{\rho}^2(j),$$

where the sample aucorrelation function

$$\hat{\rho}(j) = \hat{\gamma}(j)/\hat{\gamma}(0).$$

When $\{e_t\}$ is a directly observed data or is the estimated residual from a static regression model, we can show

$$Q(p) \to^d \chi^2_p$$

under the null hypothesis of no serial correlation.

On the other hand, when e_t is an estimated residual from an ARMA(r, s) model

$$Y_t = \alpha_0 + \sum_{j=1}^{r} \alpha_j Y_{t-j} + \sum_{j=1}^{s} \beta_j \varepsilon_{t-j} + \varepsilon_t,$$

then

$$Q(p) \to^d \chi^2_{p-(r+s)}$$

where $p > r + s$. See Box and Pierce (1970).

To improve small sample performance of the $Q(p)$ test, Ljung and Box (1978) propose a modified $Q(p)$ test statistic:

$$Q^*(p) \equiv n(n+2) \sum_{j=1}^{p} (n-j)^{-1} \hat{\rho}^2(j) \xrightarrow{d} \chi^2_{p-(r+q)}.$$

The modification matches the first two moments of $Q^*(p)$ with those of the χ^2 distribution. This improves the size in small samples, although not the power of the test.

When $\{e_t\}$ is an estimated residual from a dynamic regression model with regressors including both lagged dependent variables and exogenous variables, then the asymptotic distribution of $Q(p)$ is generally unknown (Breusch and Pagan 1980). One solution is to modify the $Q(p)$ test statistic as follows:

$$\hat{Q}(p) \equiv n\hat{\rho}'(I-\hat{\Phi})^{-1}\hat{\rho} \xrightarrow{d} \chi^2_p \text{ as } n \to \infty,$$

where $\hat{\rho} = [\hat{\rho}(1), \cdots, \hat{\rho}(p)]'$, and $\hat{\Phi}$ captures the impact caused by nonzero correlation between $\{X_t\}$ and $\{\varepsilon_{t-j}, 1 \leq j \leq p\}$. See Hayashi (2000, Section 2.10) for more discussion and the expression of $\hat{\Phi}$.

Like the $(n-p)R^2$ test, the $Q(p)$ test also assumes conditional homoskedasticity. In fact, it can be shown to be asymptotically equivalent to the $(n-p)R^2$ test statistic when e_t is the estimated residual of a static regression model.

The Kernel-Based Test for Serial Correlation

Hong (1996, *Econometrica*)

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

Examples of $k(\cdot)$:

(i) The truncated kernel

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

(ii) The Bartlett kernel

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

(iii) The Daniell kernel

$$k(z) = \frac{\sin(\pi z)}{\pi z}, \qquad z \in \mathbf{R},$$

Here, $\mathbf{1}(|z| \leq 1)$ is the indicator function that takes value 1 if $|z| \leq 1$ and 0 otherwise.

Define a test statistic

$$M(p) = \left[n\sum_{j=1}^{n-1} k^2(j/p)\hat{\rho}^2(j) - C(p)\right] / \sqrt{D(p)},$$

where $\hat{\rho}(j)$ is the sample autocorrelation function,

$$\begin{aligned} C(p) &= \sum_{j=1}^{n-1} k^2(j/p), \\ D(p) &= 2\sum_{j=1}^{n-2} k^4(j/p). \end{aligned}$$

Under the null hypothesis of no serial correlation with conditional homoskedasticity, it can be shown that

$$M(p) \to^p N(0,1)$$

as $p = p(n) \to \infty, p/n \to 0$. This holds no matter whether e_t is the estimated residual from a static regression model or a dynamic regression model.

To appreciate why $M(p) \to^d N(0,1)$, we consider a special case of using the truncated kernel $k(z) = \mathbf{1}(|z| \leq 1)$, which assigns an equal weight to each of the first p lags. In this case, $M(p)$ becomes

$$M_T(p) = \frac{n \sum_{j=1}^{p} \hat{\rho}^2(j) - p}{\sqrt{2p}}.$$

This can be viewed as a generalized version of the Box-Pierce test. In other words, the Box-Pierce test can be viewed as a kernel-based test with the choice of the truncated kernel.

For a static regression model, we have $n\Sigma_{j=1}^{p}\hat{\rho}^2(j) \to^d \chi_p^2$ under the null hypothesis of no serial correlation. When p is large, we can obtain a normal approximation for χ_p^2 by substracting its mean p and dividing by its standard deviation $\sqrt{2p}$:

$$\frac{\chi_p^2 - p}{\sqrt{2p}} \to^d N(0,1) \text{ as } p \to \infty.$$

In fact, when $p \to \infty$ as $n \to \infty$, we have the same asymptotic result even when the regression model is dynamic.

Question: Why is it not needed to correct for the impact of the estimation error contained in e_t even when the regression model is dynamic?

Answer: The estimation error indeed does have some impact but such impact becomes asymptotically negligible when p grows to infinity as $n \to \infty$. In contrast, the Box-Pierce portmanteau test has some problem because it uses a fixed lag order p (i.e., p is fixed when $n \to \infty$.)

Question: What is the advantage of using a kernel function?

For a weakly stationary process $\{\varepsilon_t\}$, the autocorrelation function $\rho(j)$ typically decays to zero as j increases. Consequently, it is more powerful if one can discount higher order lags rather than treat all lags equally. This can be achieved by using a downward weighting kernel function such as the Bartlett kernel and the Daniell kernel. Hong (1996) shows that the Daniell kernel gives a most powerful test among a class of kernel functions.

Testing Serial Correlation Under Conditional Heteroskedasticity

We have been testing serial correlation under conditional homoskedasticity. All aforementioned tests assume conditional homoskedasticity or even *i.i.d.* on $\{\varepsilon_t\}$ under the null hypothesis of no serial correlation, which rules out high frequency financial time series, which has been documented to have persistent volatility clustering. To test serial correlation under conditional hetersokedasticity, we need to use different procedures because the F-test and $(n-p)R^2$ are no longer valid.

Question: Under what conditions will conditional homoskedasticity be a reasonable assumption? And under what conditions will it not be a reasonable assumption?

Answer: It is a reasonable assumption for low-frequency macroeconomic time series. It is not a reasonable assumption for high-frequency financial time series.

Question: How to construct a test for serial correlation under conditional heteroskedasticity?

Wooldridge's (1991) Robust Test

Some effort has been devoted to robustifying tests for serial correlation. Wooldridge (1990,1991) proposes some regression-bas new procedures to test serial correlation that are robust to conditional heteroskedasticity. Specifically, Wooldridge (1990,1991) proposes a two-stage procedure to robustify the nR^2 test for serial correlation in estimated residuals $\{e_t\}$ of a linear regression model (2.1):

- Step 1: Regress $(e_{t-1}, \cdots, e_{t-p})$ on X_t and save the estimated $p \times 1$ residual vector $\hat{v}_t$;
- Step 2: Regress 1 on $\hat{v}_t e_t$ and obtain SSR, the sum of squared residuals;
- Step 3: Compare the $n - SSR$ statistic with the asymptotic χ^2_p distribution.

The first auxiliary regression purges the impact of parameter estimation uncertainty in the OLS estimator $\hat{\beta}$ and the second auxiliary regression delivers a test statistic robust to conditional heteroskedasticity of unknown form.

The Robust Kernel-based Test

Hong and Lee (2006) have recently robustified Hong's (1996) spectral density-based consistent test for serial correlation of unknown form:

$$\hat{M} \equiv \left[n^{-1} \sum_{j=1}^{n-1} k^2(j/p) \hat{\gamma}^2(j) - \hat{C}(p) \right] / \sqrt{\hat{D}(p)},$$

where the centering and scaling factors

$$\begin{aligned} \hat{C}(p) &\equiv \hat{\gamma}^2(0) \sum_{j=1}^{n-1} k^2(j/p) + \sum_{j=1}^{n-1} k^2(j/p) \hat{\gamma}_{22}(j), \\ \hat{D}(p) &\equiv 2\hat{\gamma}^4(0) \sum_{j=1}^{n-2} k^4(j/p) + 4\hat{\gamma}^2(0) \sum_{j=1}^{n-2} k^4(j/p) \hat{\gamma}_{22}(j) \\ &\quad +2 \sum_{j=1}^{n-2} \sum_{l=1}^{n-2} k^2(j/p) k^2(l/p) \hat{C}(0,j,l)^2, \end{aligned}$$

with

$$\hat{\gamma}_{22}(j) \equiv n^{-1} \sum_{t=j+1}^{n-1} [e_t^2 - \hat{\gamma}(0)][e_{t-j}^2 - \hat{\gamma}(0)]$$

and

$$\hat{C}(0,j,l) \equiv n^{-1} \sum_{t=\max(j,l)+1}^{n} [e_t^2 - \hat{\gamma}(0)] e_{t-j} e_{t-l}.$$

Intuitively, the centering and scaling factors have taken into account possible volatility clustering and asymmetric features of volatility dynamics, so the $\hat{M}$ test is robust to these effects. It allows for various volatility processes, including GARCH models, Nelson's (1991) EGARCH, and Glosten *et al.*'s (1993) Threshold GARCH models.

5.7 Summary and Conclusion

In this chapter, after introducing some basic concepts in time series analysis, we show that the asymptotic theory established under the i.i.d. assumption in Chapter 4 carries over to linear ergodic stationary time series regression models with MDS disturbances. The MDS assumption for the regression disturbances plays a key role here. For a static linear regression model, the MDS assumption is crucial for the validity of White's (1980) variance-covariance matrix estimator. For a dynamic linear regression model, the MDS assumption is crucial for correct model specification for the conditional mean $E(Y_t|I_{t-1})$.

To check the validity of the MDS assumption, one can test serial correlation in the regression disturbances. We introduce a number of tests for serial correlation and discuss the difference in testing serial correlation between a static regression model and a dynamic regression model.

References

Box and Pierce (1970), Journal of American Statistical Association.

Engle, R.F. (1982), Econometrica.

Hong, Y. (1996), Econometrica.

Wooldrige, J. (1990), Journal of Econometrics.

Wooldridge, J. (1991), Journal of Econometrics.

EXERCISES

5.1. (a) Suppose that using the Lagrange Multiplier test, one finds that there exists serial correlation in $\{\varepsilon_t\}$. Can we conclude that $\{\varepsilon_t\}$ is not a martingale difference sequence $(m.d.s)$? Give your reasoning.

(b) Suppose one finds that there exists no serial correlation in $\{\varepsilon_t\}$. Can we conclude that $\{\varepsilon_t\}$ is a $m.d.s.$? Give your reasoning. [Hint: Consider a process $\varepsilon_t = z_{t-1}z_{t-2} + z_t$, where $z_t \sim i.i.d.(0, \sigma^2)$.]

5.2. Suppose $\{Z_t\}$ is a zero-mean weakly stationary process with spectral density function $h(\omega)$ and normalized spectral density function $f(\omega)$. Show that:

(a) $f(\omega)$ is real-valued for all $\omega \in [-\pi, \pi]$;

(b) $f(\omega)$ is a symmetric function, i.e., $f(-\omega) = f(\omega)$;

(c) $\int_{-\pi}^{\pi} f(\omega)d\omega = 1$;

(d) $f(\omega) \geq 0$ for all $\omega \in [-\pi, \pi]$. [Hint: Consider the limit of $E|n^{-1/2}\sum_{t=1}^{n} Z_t e^{it\omega}|^2$, the variance of the complex-valued random variable $n^{-1/2}\sum_{t=1}^{n} Z_t e^{it\omega}$.

5.3. Suppose a time series linear regression model

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

where the disturbance ε_t is directly observable, satisfies Assumptions 5.1–5.3. This class of models contains both static regression models and dynamic regression models.

(a) Does the condition $E(\varepsilon_t|X_t) = 0$ imply that $\{\varepsilon_t\}$ is a white noise? Explain.

(b) If $\{\varepsilon_t\}$ is MDS, does it imply $E(\varepsilon_t|X_t) = 0$? Explain.

(c) If $\{\varepsilon_t\}$ is serially correlated, does it necessarily imply $E(\varepsilon_t|X_t) \neq 0$, i.e., the linear regression model is misspecified for $E(Y_t|X_t)$? Explain.

5.4. Suppose that in a linear regression model

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

the disturbance ε_t is directly observable. We are interested in testing the null hypothesis $\mathbf{H}_0$ that $\{\varepsilon_t\}$ is serially uncorrelated. Suppose Assumptions 5.1–5.6 hold.

(a) Consider the auxiliary regression

$$\varepsilon_t = \sum_{j=1}^{p} \alpha_j \varepsilon_{t-j} + u_t, \ t = p+1, ..., n.$$

Let $\tilde{R}_{uc}^2$ is the uncentered R^2 from the OLS estimation of this auxiliary regression. Show that $(n-2p)\tilde{R}_{uc} \to^d \chi_p^2$ as $n \to \infty$ under $\mathbf{H}_0$.

(b) Now consider another auxiliary regression

$$\varepsilon_t = \alpha_0 + \sum_{j=1}^{p} \alpha_j \varepsilon_{t-j} + u_t, t = p+1, ..., n.$$

Let $\tilde{R}^2$ be the centered R^2 from this auxiliary regression model. Show that $(n-2p)\tilde{R}^2 \to \chi_p^2$ as $n \to \infty$ under $\mathbf{H}_0$.

(c) Which test statistic, $(n-2p)\tilde{R}^2_{uc}$ or $(n-2p)\tilde{R}^2$, performs better in finite samples? Give your heuristic reasoning.

5.5. Suppose that in a linear regression model

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

the disturbance ε_t is directly observable. We are interested in testing the null hypothesis $\mathbf{H}_0$ that $\{\varepsilon_t\}$ is serially uncorrelated. Suppose Assumptions 5.1–5.5 hold, and $E(\varepsilon_t^2|X_t) \neq \sigma^2$.

(a) Consider the auxiliary regression

$$\varepsilon_t = \sum_{j=1}^{p} \alpha_j \varepsilon_{t-j} + u_t, \;\; t = p+1, ..., n.$$

Construct an asymptotically valid test statistic for the null hypothesis that there exists no serial correlation in $\{\varepsilon_t\}$.

5.6. Suppose ε_t follows an ARCH(1) process

$$\begin{aligned} \varepsilon_t &= z_t \sigma_t, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2, \\ \{z_t\} &\sim i.i.d.N(0,1) \end{aligned}$$

(a) Show $E(\varepsilon_t|I_{t-1}) = 0$ and $\text{cov}(\varepsilon_t, \varepsilon_{t-j}) = 0$ for all $j > 0$, where $I_{t-1} = \{\varepsilon_{t-1}, \varepsilon_{t-2}, ...\}$.

(b) Show $\text{cov}(\varepsilon_t^2, \varepsilon_{t-1}^2) = \alpha_1$.

(c) Show the kurtosis of ε_t is given by

$$\begin{aligned} K &= \frac{E(\varepsilon_t^4)}{[E(\varepsilon_t^2)]^2} = \frac{3(1-\alpha_1^2)}{1-3\alpha_1^2} \\ &> 3 \text{ if } \alpha_1 > 0. \end{aligned}$$

5.7. Suppose a time series linear regression model

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

where the disturbance ε_t is directly observable, satisfies Assumptions 5.1–5.5. Both static and dynamic regression models are covered.

Suppose there exists autoregressive conditional heteroskedasticity (ARCH) for $\{\varepsilon_t\}$, namely,

$$E(\varepsilon_t^2|I_{t-1}) = \alpha_0 + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2,$$

where I_{t-1} is the sigma-field generated by $\{\varepsilon_{t-1}, \varepsilon_{t-2}, ...\}$. Does this imply that one has to use the asymptotic variance formula $Q^{-1}VQ^{-1}$ for avar$(\sqrt{n}\hat{\beta})$? Explain.

5.8. Suppose a time series linear regression model

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

where the disturbance ε_t is directly observable, satisfies Assumptions 5.1–5.5, and the two time series $\{X_t\}$ and $\{\varepsilon_t\}$ are independent of each other.

Suppose there exists autoregressive conditional heteroskedasticity for $\{\varepsilon_t\}$, namely,

$$E(\varepsilon_t^2|I_{t-1}) = \alpha_0 + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2,$$

where I_{t-1} is the sigma-field generated by $\{\varepsilon_{t-1}, \varepsilon_{t-2}, ...\}$.

What is the form of avar$(\sqrt{n}\hat{\beta})$, where $\hat{\beta}$ is the OLS estimator?

5.9. Suppose a dynamic time series linear regression model

$$\begin{aligned} Y_t &= \beta_0^o + \beta_1^o Y_{t-1} + \varepsilon_t \\ &= X_t'\beta^o + \varepsilon_t \end{aligned}$$

satisfies Assumptions 5.1–5.5. Suppose further there exists autoregressive conditional heteroskedasticity for $\{\varepsilon_t\}$ in form of the following:

$$E(\varepsilon_t^2|I_{t-1}) = \alpha_0 + \alpha_1 Y_{t-1}^2.$$

What is the form of avar$(\sqrt{n}\hat{\beta})$, where $\hat{\beta}$ is the OLS estimator?

5.10. Suppose a time series linear regression model

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

satisfies Assumptions 5.1, 5.2 and 5.4, the two time series $\{X_t\}$ and $\{\varepsilon_t\}$ are independent of each other, and $E(\varepsilon_t) = 0$. Suppose further that there exist serial correlation in $\{\varepsilon_t\}$.

(a) Does the presence of serial correlation in $\{\varepsilon_t\}$ affect the consistency of $\hat{\beta}$ for β^o? Explain.

(b) Does the presence of serial correlation in $\{\varepsilon_t\}$ affect the form of asymptotic variance $\text{avar}(\sqrt{n}\hat{\beta}) = Q^{-1}VQ^{-1}$, where $V = \lim_{n\to\infty}\text{var}(n^{-1/2}\Sigma_{t=1}^n X_t\varepsilon_t)$? In particular, do we still have $V = E(X_tX_t'\varepsilon_t^2)$? Explain.

5.11. Suppose a dynamic time series linear regression model

$$\begin{aligned} Y_t &= \beta_0^o + \beta_1^o Y_{t-1} + \varepsilon_t \\ &= X_t'\beta^o + \varepsilon_t, \end{aligned}$$

where $X_t = (1, Y_{t-1})'$, satisfies Assumptions 5.1, 5.2 and 5.4. Suppose further $\{\varepsilon_t\}$ follows an MA(1) process:

$$\varepsilon_t = \rho v_{t-1} + v_t,$$

where $\{v_t\}$ is i.i.d.$(0, \sigma_v^2)$. Thus, there exists first order serial correlation in $\{\varepsilon_t\}$.

Is the OLS estimator $\hat{\beta}$ consistent for β^o? Explain.