3 Asymptotic Theory

The following are the assumptions of the classical linear model.

Theorem 3.1

1. The model is known to be $Y = X\beta + \epsilon$, where $\beta < \infty$, Y and ϵ are $n \times 1$ vectors, and X is a $n \times k$ matrix.
2. X is a nonstochastic and finite $n \times k$ matrix.
3. $X'X$ is nonsingular for all $n \geq k$.
4. $E(\epsilon) = 0$.
5. $\epsilon \sim N(0, \sigma^2 I)$, $\sigma^2 < \infty$.

   (a) (Existence) Given Assumptions 1–3, $\hat{\beta}$ exists for all $n \geq k$ and is unique.

   (b) (Unbiasedness) Given Assumptions 1–4, $E(\hat{\beta}) = \beta$.

   (c) (Normality) Given Assumptions 1–5, $\hat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1})$.

   (d) (Efficiency) Given Assumptions 1–5, $\hat{\beta}$ is the maximum likelihood estimator and is the best unbiased estimator in the sense that the variance-covariance matrix of any other unbiased estimator exceeds that of $\hat{\beta}$ by a positive semidefinite matrix, regardless of the value of β.

The properties of existence, unbiasedness, normality, and efficiency of an estimator are the small sample analogs of the properties that will be the focus of interest here. Unbiasedness tells us that the distribution of $\hat{\beta}$ is centered around the unknown true value β, whereas the normality property allows us to construct confidence intervals and test hypotheses using the $t-$ and $F-$distributions. The efficiency property guarantees that our estimator has the greatest possible precision within a given class of estimators and also helps ensure that tests of hypotheses have high power.

Of course, the classical assumptions are rather stringent and can easily fail in situations faced by the analyst. Since failures of assumptions 3 and 4 are easily remedied (exclude linearly dependent regressors if 3 fails, include a constant in the model if 4 fails, we will concern ourselves primarily with the failure of assumptions 2 and 5. The possible failure of assumption 1 is a subject that requires a course in itself and will not be fully considered here. Nevertheless, the tools developed here will be essential to understanding and treating the consequences of the failure of assumptions 1.

We shall briefly examine the consequences of various failures of assumptions 2 and 5. First, suppose that ϵ exhibits heteroskedasticity or serial correlation,k so that $E(\epsilon\epsilon') = \Omega \neq \sigma^2 I$. We have the following result for the OLS estimator.

Theorem 3.2

Suppose the classical assumptions 1–4 hold but replace 5 with 5′:

$$\epsilon \sim N(0, \Omega),\ \Omega < \infty.$$

Then the existence (a) and unbiasedness (b) results hold as before, but the normality result (c) is replaced by (c′)

$$\hat{\beta} \sim N(\beta, (X'X)^{-1}X'\Omega X(X'X)^{-1}),$$

and the efficiency result (d) does not hold, that is, $\hat{\beta}$ is no longer the best unbiased estimator.

PROOF:

By definition, $\hat{\beta} = (X'X)^{-1}X'Y$. Given assumption 1,

$$\hat{\beta} = \beta + (X'X)^{-1}X'\epsilon, \tag{70}$$

where $(X'X)^{-1}X'\epsilon$ is a linear combination of jointly normal random variables and is therefore jointly normal with

$$E[(X'X)^{-1}X'\epsilon] = (X'X)^{-1}X'E[\epsilon] \tag{71}$$

$$= 0, \tag{72}$$

given assumptions 2 (nonstochastic X) and 4 ($E(\epsilon) = 0$ and

$$var[(X'X)^{-1}X'\epsilon] = E[(X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1}] \tag{73}$$

$$= (X'X)^{-1}X'E[\epsilon\epsilon']X(X'X)^{-1} \tag{74}$$

$$= (X'X)^{-1}X'\Omega X(X'X)^{-1}, \tag{75}$$

$$\tag{76}$$

given assumptions 2 (X is nonstochastic) and $5'$. Hence,

$$\hat{\beta} \sim N(\beta, (X'X)^{-1}X'\Omega X(X'X)^{-1}). \tag{77}$$

That efficiency (d) does not hold follows because there exists an unbiased estimator with smaller covariance matrix than $\hat{\beta}$, namely, $\beta^* = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$.

We shall now examine the properties of β^*.

As long as Ω is known, the presence of serial correlation or heteroskedasticity does not render us incapable of testing hypotheses or constructing confidence intervals. This can still be done using (c$'$), although the failure of the efficiency (d) property indicates that the OLS estimator my not be the best for these purposes. However, if Ω is unknown (apart from a factor of proportionality), testing hypotheses and constructing confidence intervals is no longer a simple matter. One might be able to construct tests based on estimates of Ω, but the resulting statistics may have very complicated distributions. As we shall see later, his difficulty is lessened in large samples by the availability of convenient approximations based on the central limit theorem and laws of large numbers.

If Ω is known, efficiency can be regained by applying OLS to a linear transformation of the original model, i.e.,

$$C^{-1}Y = C^{-1}X\beta + C^{-1}\epsilon \tag{78}$$

or

$$Y^* = X^*\beta + \epsilon^*, \tag{79}$$

where $Y^* = C^{-1}Y$, $X^* = C^{-1}X$, $\epsilon^* = C^{-1}\epsilon$ and C is a nonsingular factorization of Ω such that $CC' = \Omega$ and $C^{-1}\Omega C^{-1\prime} = I$. This transformation ensures that $E(\epsilon^*\epsilon^{*\prime}) = E(C^{-1}\epsilon\epsilon' C^{-1\prime}) = C^{-1}E(\epsilon\epsilon')C^{-1\prime} = C^{-1}\Omega C^{-1\prime} = I$, so that the efficiency result (d) once again holds.

The least squares estimator for the transformed model is:

$$\beta^* = (X^{*\prime}X^*)^{-1}X^{*\prime}Y^* \tag{80}$$

$$= (X'C^{-1\prime}C^{-1}X)^{-1}X'C^{-1\prime}C^{-1}Y \tag{81}$$

$$= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y. \tag{82}$$

The estimators β^* is called the **generalized least squares** (GLS) estimators and its properties are given by the following result.

The following are the "generalized" classical assumptions.

Theorem 3.3

1 The model is known to be $Y = X\beta + \epsilon$, where $\beta < \infty$, Y and ϵ are $n \times 1$ vectors, and X is a $n \times k$ matrix.

2 X is a nonstochastic and finite $n \times k$ matrix.

3* $X'\Omega^{-1}X$ is nonsingular for all $n \geq k$ and Ω.

4 $E(\epsilon) = 0$.

5* $\epsilon \sim N(0, \Omega)$, is finite and nonsingular.

(a) (Existence) Given Assumptions 1–3*, β^* exists for all $n \geq k$ and is unique.

(b) (Unbiasedness) Given Assumptions 1–4, $E(\beta^*) = \beta$.

(c) (Normality) Given Assumptions 1–5*, $\beta^* \sim N(\beta, X'\Omega^{-1}X)$.

(d) (Efficiency) Given Assumptions 1–5*, β^* is the maximum likelihood estimator and is the best unbiased estimator in the sense that the variance-covariance matrix of any other unbiased estimator exceeds that of $\hat{\beta}$ by a positive semidefinite matrix, regardless of the value of β.

The proof for Theorem 3.3 may be obtained by apply Theorem 3.1 to the model $Y^* = X^*\beta + \epsilon^*$.

If Ω is known, we obtain efficiency by transforming the model "back" to a form in which OLS gives the efficient estimator. However, if Ω is unknown, this transformation is not immediately available. It might be possible to estimate Ω, say be $\hat{\Omega}$, but $\hat{\Omega}$ is then random and so is the factorization $\hat{C}$. 3.1 no longer applies.

Hypothesis testing in the classical linear model relies heavily on being able to make use of the $t-$ and $F-$distributions. Nevertheless, the central limit theorem can can be applied when n is large to guarantee that $\hat{\beta}$ or β^* is distributed approximately as normal, as we shall see.

Now consider what happens when assumption 2 fails, so that the explanatory variables X are stochastic. In some cases, this causes no real problems because we can examine the properties of our estimators "conditional" on X. for example, consider the unbiasedness property. To demonstrate unbiasedness we use 1 (correct specification) to write

$$\hat{\beta} = \beta(X'X)^{-1}X'\epsilon. \tag{83}$$

If X is random, we can no longer write $E[(X'X)^{-1}X'\epsilon] = (X'X)^{-1}X'E[\epsilon]$. However, by taking conditional expectations, we can treat X as "fixed", so we have

$$E(\hat{\beta}|X) = \beta + E[(X'X)^{-1}X'\epsilon|X] \tag{84}$$

$$= \beta + (X'X)^{-1}X'E[\epsilon|X]. \tag{85}$$

If we are willing to assume $E[\epsilon|X] = 0$, then conditional unbiasedness follows, i.e.,

$$E[\hat{\beta}|X] = \beta. \tag{86}$$

Unconditional unbiasedness also holds as a consequence of the law of iterated expectations which we will come across in more detail later.

$$E(\hat{\beta}) = E[E(\hat{\beta}|X)] \tag{87}$$

$$= E(\beta) \tag{88}$$

$$= \beta. \tag{89}$$

The other properties can be similarly considered. However, the assumption that $E(\epsilon|X) = 0$ is crucial. If $E(\epsilon|X) \neq 0$, $\hat{\beta}$ need not be unbiased, either conditionally or unconditionally.

Situations in which $E(\epsilon|X) \neq 0$ can arise easily. For example, X_t may **contain errors**

of measurement. Suppose the model is

$$Y_t = W_t\beta + V_t, \quad E(W_t'V_t) = 0, \tag{90}$$

but we measure W_t subject to errors η_t as $X_t = W_t + \eta_t$, $E(W_t'\eta_t) = 0$, $E(\eta_t'\eta_t) \neq 0$, $E(\eta_t'V_t) = 0$. Then

$$Y_t = X_t\beta + V_t - \eta_t\beta \tag{91}$$

$$= X_t\beta + \epsilon_t. \tag{92}$$

With $\epsilon_t = V_t - \eta_t\beta$, we have $E(X_t'\epsilon_t) = E[(W_t' + \eta_t')(V_t - \eta_t\beta)] = E(\eta_t'\eta_t)\beta \neq 0$. Now $E(\epsilon|X) = 0$ implies that for all t, $E(X_t'\epsilon_t) = 0$, since $E(X_t'\epsilon_t) = E[E(X_t'\epsilon_t|X)] = E[X_t'E(\epsilon_t|X)] = 0$. Hence, $E(X_t'\epsilon_t) \neq 0$ implies $E(\epsilon|X) \neq 0$. The OLS estimator will not be unbiased in the presence of measurement error.

For another example, let us consider the case of **serially correlated errors in the presence of a lagged dependent variable** Y_{t-1}:

$$Y_t = Y_{t-1}\alpha + W_t\delta + \epsilon_t, \quad E(W_t'\epsilon_t) = 0 \tag{93}$$

$$\epsilon_t = \rho\epsilon_{t-1} + V_t, \quad E(\epsilon_{t-1}V_t) = 0. \tag{94}$$

Let $X_t = (Y_{t-1}, W_t)$ and $\beta' = (\alpha, \delta')$. Again the model is,

$$Y_t = X_t\beta + \epsilon_t,$$

but we have $E(X_t'\epsilon_t) = E[(Y_{t-1}, W_t)'\epsilon_t] = [E(Y_{t-1}\epsilon_t), 0]'$. If we also assume $E(Y_{t-1}V_t) = 0$, $E(Y_{t-1}\epsilon_{t-1}) = E(Y_t\epsilon_t)$, and $E(\epsilon_t^2) = \sigma^2$, it can be shown that

$$E(Y_{t-1}\epsilon_t) = \sigma^2\rho/(1 - \rho\alpha).$$

Thus $E(X_t'\epsilon_t) \neq 0$ so that $E(\epsilon|X') \neq 0$ and OLS is not generally unbiased.