Chapter 13

Asymptotic Theory and Stochastic Regressors

The nature of the explanatory variable is assumed to be non-stochastic or fixed in repeated samples in any regression analysis. Such an assumption is appropriate for experiments conducted inside the laboratories where the experimenter can control the values of explanatory variables. Then the repeated observations on the study variable can be obtained for fixed values of explanatory variables. In practice, such an assumption may only sometimes be satisfied. Sometimes, the explanatory variables in a given model are the study variable in another model. Thus the study variable depends on the explanatory variables that are stochastic in nature. Under such situations, the statistical inferences drawn from the linear regression model based on the assumption of fixed explanatory variables may not remain valid.

We assume now that the explanatory variables are stochastic but uncorrelated with the disturbance term. If they are correlated, the issue is addressed through instrumental variable estimation. Such a situation arises in the case of measurement error models.

Stochastic regressors model

Consider the linear regression model

$$y = X\beta + \varepsilon$$

where X is a $(n \times k)$ matrix of observations on k explanatory variables, which are **stochastic in nature,** y is a $(n \times 1)$ vector of observations on study variable, β is a $(k \times 1)$ vector of regression coefficients, and ε is the $(n \times 1)$ vector of disturbances. Under the assumption $E(\varepsilon) = 0, V(\varepsilon) = \sigma^2 I,$ the distribution of ε_i, conditional on satisfies these properties for all values of X where x_i' denotes the i^{th} row of X. This is demonstrated as follows:

Let $p(\varepsilon_i \mid x_i')$ be the conditional probability density function of ε_i given x_i' and $p(\varepsilon_i)$ is the unconditional probability density function of ε_i. Then

$$\begin{aligned} E(\varepsilon_i \mid x_i') &= \int \varepsilon_i p(\varepsilon_i \mid x_i') d\varepsilon_i \\ &= \int \varepsilon_i p(\varepsilon_i) d\varepsilon_i \\ &= E(\varepsilon_i) \\ &= 0 \end{aligned}$$

$$
\begin{aligned}
E\left(\varepsilon_i^2 \mid x_i^{'}\right) &= \int \varepsilon_i^2 p\left(\varepsilon_i \mid x_i^{'}\right) d\varepsilon_i \\
&= \int \varepsilon_i^2 p\left(\varepsilon_i\right) d\varepsilon_i \\
&= E\left(\varepsilon_i^2\right) \\
&= \sigma^2.
\end{aligned}
$$

In case, they are independent, then $p\left(\varepsilon_i \mid x_i^{'}\right) = p\left(\varepsilon_i\right)$.

Least squares estimation of parameters

The additional assumption that the explanatory variables are stochastic poses no problem in the ordinary least squares estimation of β and σ^2. The OLSE of β is obtained by minimizing $(y - X\beta)'(y - X\beta)$ with respect β as

$$
b = \left(X'X\right)^{-1} X'y
$$

and estimator of σ^2 is obtained as

$$
s^2 = \frac{1}{n-k}(y - Xb)'(y - Xb).
$$

Maximum likelihood estimation of parameters:

Assuming $\varepsilon \sim N\left(0, \sigma^2 I\right)$ in the model $y = X\beta + \varepsilon$ along with X is stochastic and independent of ε, the joint probability density function ε and X can be derived from the joint probability density function of y and X as follows:

$$
\begin{aligned}
f(\varepsilon, X) &= f\left(\varepsilon_1, \varepsilon_2, ..., \varepsilon_n, x_1^{'}, x_2^{'}, ..., x_n^{'}\right) \\
&= \left(\prod_{i=1}^{n} f\left(\varepsilon_i\right)\right)\left(\prod_{i=1}^{n} f\left(x_i^{'}\right)\right) \\
&= \left(\prod_{i=1}^{n} f\left(y_i \mid x_i^{'}\right)\right)\left(\prod_{i=1}^{n} f\left(x_i^{'}\right)\right) \\
&= \prod_{i=1}^{n}\left(f\left(y_i \mid x_i^{'}\right) f\left(x_i^{'}\right)\right) \\
&= \prod_{i=1}^{n} f\left(y_i, x_i^{'}\right) \\
&= f\left(y_1, y_2, ..., y_n, x_1^{'}, x_2^{'}, ..., x_n^{'}\right) \\
&= f(y, X).
\end{aligned}
$$

This implies that the maximum likelihood estimators of β and σ^2 will be based on

$$\prod_{i=1}^{n} f\left(y_i \mid x_i^{'}\right) = \prod_{i=1}^{n} f\left(\varepsilon_i\right)$$

so they will be the same as based on the assumption that ε_i 's, $i = 1, 2, ..., n$ are distributed as $N\left(0, \sigma^2\right)$. So the maximum likelihood estimators of β and σ^2 when the explanatory variables are stochastic are obtained as

$$\tilde{\beta} = \left(X'X\right)^{-1} X'y$$
$$\tilde{\sigma}^2 = \frac{1}{n}\left(y - X\tilde{\beta}\right)'\left(y - X\tilde{\beta}\right).$$

Alternative approach for deriving the maximum likelihood estimates

Alternatively, the maximum likelihood estimators of β and σ^2 can also be derived using the joint probability density function of y and X.

Note: Note that the vector $\underline{x}'$ is represented by an underscore <u>in this section</u> to denote that its order is $\left[1 \times (k-1)\right]$ which excludes the intercept term.

Let $\underline{x}_i^{'}, i = 1, 2, ..., n$ are from a multivariate normal distribution with mean vector $\underline{\mu}_x$ and covariance matrix Σ_{xx}, i.e., $\underline{x}_i^{'} \sim N\left(\underline{\mu}_x, \Sigma_{xx}\right)$ and the joint distribution of y and $\underline{x}_i^{'}$ is

$$\begin{pmatrix} y \\ \underline{x}_i^{'} \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_y \\ \underline{\mu}_x \end{pmatrix}, \begin{pmatrix} \sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{pmatrix}\right].$$

Let the linear regression model is

$$y_i = \beta_0 + \underline{x}_i^{'}\underline{\beta}_1 + \varepsilon_i, \quad i = 1, 2, ..., n$$

where $\underline{x}_i^{'}$ is a $\left[1 \times (k-1)\right]$ vector of observation of random vector x, β_0 is the intercept term and $\underline{\beta}_1$ is the $\left[(k-1) \times 1\right]$ vector of regression coefficients. Further ε_i is disturbance term with $\varepsilon_i \sim N\left(0, \sigma^2\right)$ and is independent of $\underline{x}'$.

Suppose

$$\begin{pmatrix} y \\ \underline{x} \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_y \\ \underline{\mu}_x \end{pmatrix}, \begin{pmatrix} \sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{pmatrix}\right].$$

The joint probability density function of $\left(y, \underline{x}_i^{'}\right)$ based on a random sample of size n is

$$f\left(y, \underline{x}'\right) = \frac{1}{(2\pi)^{\frac{k}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\begin{pmatrix} y-\mu_y \\ \underline{x}-\underline{\mu}_x \end{pmatrix}^{'} \Sigma^{-1} \begin{pmatrix} y-\mu_y \\ \underline{x}-\underline{\mu}_x \end{pmatrix}\right].$$

Now using the following result, we find Σ^{-1}:

Result: Let A be a nonsingular matrix which is partitioned suitably as

$$A = \begin{pmatrix} B & C \\ D & E \end{pmatrix},$$

where E and $F = B - CE^{-1}D$ are nonsingular matrices, then

$$A^{-1} = \begin{pmatrix} F^{-1} & -F^{-1}CE^{-1} \\ -E^{-1}DF^{-1} & E^{-1} + E^{-1}DF^{-1}CE^{-1} \end{pmatrix}.$$

Note that $AA^{-1} = A^{-1}A = I.$

Thus

$$\Sigma^{-1} = \frac{1}{\sigma^2}\begin{pmatrix} 1 & -\Sigma_{yx}\Sigma_{xx}^{-1} \\ -\Sigma_{xx}^{-1}\Sigma_{xy} & \sigma^2\Sigma_{xx}^{-1} + \Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yx}\Sigma_{xx}^{-1} \end{pmatrix},$$

where

$$\sigma^2 = \sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}.$$

Then

$$f\left(y, \underline{x}'\right) = \frac{1}{(2\pi)^{\frac{k}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{1}{2\sigma^2}\left\{\left[y - \mu_y - \left(\underline{x} - \underline{\mu}_x\right)'\Sigma_{xx}^{-1}\Sigma_{xy}\right]^2 + \sigma^2\left(\underline{x} - \underline{\mu}_x\right)'\Sigma_{xx}^{-1}\left(\underline{x} - \underline{\mu}_x\right)\right\}\right].$$

The marginal distribution of $\underline{x}'$ is obtained by integrating $f\left(y, \underline{x}'\right)$ over y and the resulting distribution is $(k-1)$ variate multivariate normal distribution as

$$g\left(\underline{x}'\right) = \frac{1}{(2\pi)^{\frac{k-1}{2}} |\Sigma_{xx}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\left(\underline{x} - \underline{\mu}_x\right)'\Sigma_{xx}^{-1}\left(\underline{x} - \underline{\mu}_x\right)\right].$$

The conditional probability density function of the given $\underline{x}'$ is

$$f\left(y \mid \underline{x}'\right)=\frac{f\left(y, \underline{x}'\right)}{g\left(\underline{x}'\right)}$$
$$=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2\sigma^2}\left\{\left(y-\mu_y\right)-\left(\underline{x}-\underline{\mu}_x\right)'\Sigma_{xx}^{-1}\Sigma_{xy}\right\}^2\right]$$

which is the probability density function of normal distribution with

- conditional mean

  $E(y \mid \underline{x}')=\mu_y+\left(\underline{x}-\underline{\mu}_x\right)'\Sigma_{xx}^{-1}\Sigma_{xy}$ and

- conditional variance

  $Var\left(y \mid \underline{x}'\right)=\sigma_{yy}\left(1-\rho^2\right)$

where

$$\rho^2=\frac{\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}}{\sigma_{yy}}$$

is the population multiple correlation coefficient.

In the model

$$y=\beta_0+\underline{x}'\underline{\beta}_1+\varepsilon,$$

the conditional mean is

$$E\left(y_i \mid \underline{x}_i'\right)=\beta_0+\underline{x}'\beta_1+E\left(\varepsilon \mid \underline{x}\right)$$
$$=\beta_0+\underline{x}'\underline{\beta}_1.$$

Comparing this conditional mean with the conditional mean of normal distribution, we obtain the relationship with β_0 and $\underline{\beta}_1$ as follows:

$$\underline{\beta}_1=\Sigma_{xx}^{-1}\Sigma_{xy}$$
$$\beta_0=\mu_y-\underline{\mu}_x'\underline{\beta}_1.$$

The likelihood function of $\left(y, \underline{x}'\right)$ based on a sample of size n is

$$L=\frac{1}{\left(2\pi\right)^{\frac{nk}{2}}\left|\Sigma\right|^{\frac{n}{2}}}\exp\left[\sum_{i=1}^{n}\left\{-\frac{1}{2}\begin{pmatrix}y_i-\mu_y\\ \underline{x}_i-\underline{\mu}_x\end{pmatrix}'\Sigma^{-1}\begin{pmatrix}y_i-\mu_y\\ \underline{x}_i-\underline{\mu}_x\end{pmatrix}\right\}\right].$$

Maximizing the log-likelihood function with respect to $\mu_y, \underline{\mu}_x, \Sigma_{xx}$ and Σ_{xy}, the maximum likelihood estimates of respective parameters are obtained as

$$\tilde{\mu}_y = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$\underline{\tilde{\mu}}_x = \underline{\bar{x}} = \frac{1}{n}\sum_{i=1}^{n} \underline{x}_i = (\bar{x}_2, \bar{x}_3, ..., \bar{x}_k)$$

$$\tilde{\Sigma}_{xx} = S_{xx} = \frac{1}{n}\left(\sum_{i=1}^{n} \underline{x}_i \underline{x}_i^{'} - n\underline{\bar{x}}\,\underline{\bar{x}}'\right)$$

$$\tilde{\Sigma}_{xy} = S_{xy} = \frac{1}{n}\left(\sum_{i=1}^{n} \underline{x}_i y_i - n\underline{\bar{x}}\,\bar{y}\right)$$

where $\underline{x}_i^{'} = (x_{i2}, x_{i3}, ..., x_{ik}),\ S_{xx}$ is $[(k\text{-}1)\times(k\text{-}1)]$ matrix with elements $\frac{1}{n}\sum_t (x_{ti} - \bar{x}_i)(x_{tj} - \bar{x}_j)$ and S_{xy} is $[(k\text{-}1)\times 1]$ vector with elements $\frac{1}{n}\sum_t (x_{ti} - \bar{x}_i)(y_i - \bar{y})$.

Based on these estimates, the maximum likelihood estimators are obtained as

$$\underline{\tilde{\beta}}_1 = S_{xx}^{-1} S_{xy}$$

$$\tilde{\beta}_0 = \bar{y} - \bar{x}'\underline{\tilde{\beta}}_1$$

$$\tilde{\beta} = \begin{pmatrix} \tilde{\beta}_0 \\ \underline{\tilde{\beta}}_1 \end{pmatrix} = \left(X'X\right)^{-1} X'y.$$

Properties of the estimators of least squares estimator:

The estimation error of OLSE is

$$\begin{aligned} b - \beta &= \left(X'X\right)^{-1} X'y - \beta \\ &= \left(X'X\right)^{-1} X'\left(X\beta + \varepsilon\right) - \beta \\ &= \left(X'X\right)^{-1} X'\varepsilon. \end{aligned}$$

Then assuming that $E\left[\left(X'X\right)^{-1} X'\right]$ exists, we have

$$\begin{aligned} E(b - \beta) &= E\left[\left(X'X\right)^{-1} X'\varepsilon\right] \\ &= E\left[E\left\{\left(X'X\right)^{-1} X'\varepsilon \middle| X\right\}\right] \\ &= E\left[\left(X'X\right)^{-1} X'\right] E\left(\varepsilon\right) \\ &= 0 \end{aligned}$$

because $\left(X'X\right)^{-1} X'$ and ε are independent. So b is an unbiased estimator of β.

The covariance matrix of b is obtained as

$$
\begin{aligned}
V(b) &= E(b-\beta)(b-\beta)' \\
&= E\left[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}\right] \\
&= E\left[E\left\{(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}|X\right\}\right] \\
&= E\left[(X'X)^{-1}X'E(\varepsilon\varepsilon')X(X'X)^{-1}|X\right] \\
&= E\left[(X'X)^{-1}X'\sigma^2X(X'X)^{-1}\right] \\
&= \sigma^2 E\left[(X'X)^{-1}\right].
\end{aligned}
$$

Thus the covariance matrix involves a mathematical expectation. The unknown σ^2 can be estimated by

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{e'e}{n-k} \\
&= \frac{(y-Xb)'(y-Xb)}{n-k}
\end{aligned}
$$

where $e = y - Xb$ is the residual and

$$
\begin{aligned}
E(\hat{\sigma}^2) &= E\left[E(\hat{\sigma}^2|X)\right] \\
&= E\left[E\left(\frac{e'e}{n-k}\right)|X\right] \\
&= E(\sigma^2) \\
&= \sigma^2.
\end{aligned}
$$

Note that the OLSE $b = (X'X)^{-1}X'y$ involves the stochastic matrix X and stochastic vector y, so b is not a linear estimator. It is also no more the best linear unbiased estimator of β as in the case when X is non-stochastic. The estimator of σ^2 as being conditional on given X is an efficient estimator.

Asymptotic theory:

The asymptotic properties of an estimator concern the properties of the estimator when sample size n grows large.

For the need and understanding of asymptotic theory, we consider an example. Consider the simple linear regression model with one explanatory variable and n observations as

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,\ E(\varepsilon_i) = 0,\ Var(\varepsilon_i) = \sigma^2,\ i = 1, 2, ..., n.$$

The OLSE of β_1 is

$$b_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

and its variance is

$$Var(b_1) = \frac{\sigma^2}{n}.$$

If the sample size grows large, then the variance of b_1 gets smaller. The shrinkage in variance implies that as sample size n increases, the probability density of OLSE b collapses around its mean because $Var(b)$ becomes zero.

Let there be three OLSEs based on sample sizes n_1, n_2 and n_3 respectively such that $n_1 < n_2 < n_3$, say. If c and δ are some arbitrarily chosen positive constants, then the probability that the value of b lies within the interval $\beta \pm c$ can be made to be greater than $(1-\delta)$ for a large value of n. This property is the consistency of b which ensures that even if the sample is very large, then we can be confident with a high probability that b will yield an estimate that is close to β.

Probability in limit

Let $\hat{\beta}_n$ be an estimator of β based on a sample of size n. Let γ be any small positive constant. Then for large n, the requirement that b_n takes values with probability almost one in an arbitrarily small neighborhood of the true parameter value β is

$$\lim_{n\to\infty} P\left[\left|\hat{\beta}_n - \beta\right| < \gamma\right] = 1$$

which is denoted as

$$\text{plim } \hat{\beta}_n = \beta$$

and it is said that $\hat{\beta}_n$ converges to β in probability. The estimator $\hat{\beta}_n$ is said to be a consistent estimator of β.

A sufficient but not necessary condition for $\hat{\beta}_n$ to be a consistent estimator of β is that

$$\lim_{n\to\infty} E\left[\hat{\beta}_n\right] = \beta$$

and $\lim_{n\to\infty} Var\left[\hat{\beta}_n\right] = 0.$

Consistency of estimators

Now we look at the consistency of the estimators of β and σ^2.

(i) Consistency of *b*

Under the assumption that $\lim_{n\to\infty}\left(\frac{X'X}{n}\right)=\Delta$ exists as a nonstochastic and nonsingular matrix (with finite elements), we have

$$\begin{aligned}\lim_{n\to\infty} V(b) &= \sigma^2 \lim_{n\to\infty} \frac{1}{n}\left(\frac{X'X}{n}\right)^{-1} \\ &= \sigma^2 \lim_{n\to\infty} \frac{1}{n}\Delta^{-1} \\ &= 0.\end{aligned}$$

This implies that OLSE converges to β in quadratic mean. Thus OLSE is a consistent estimator of β. This also holds true for maximum likelihood estimators also.

Same conclusion can also be proved using the concept of convergence in probability.

The consistency of OLSE can be obtained under the weaker assumption that

$$\text{plim}\left(\frac{X'X}{n}\right)=\Delta_*.$$

exists and is a nonsingular and nonstochastic matrix and

$$\text{plim}\left(\frac{X'\varepsilon}{n}\right)=0.$$

Since

$$\begin{aligned}b-\beta &= (X'X)^{-1}X'\varepsilon \\ &= \left(\frac{X'X}{n}\right)^{-1}\frac{X'\varepsilon}{n}.\end{aligned}$$

So

$$\begin{aligned}\text{plim}(b-\beta) &= \text{plim}\left(\frac{X'X}{n}\right)^{-1}\text{plim}\left(\frac{X'\varepsilon}{n}\right) \\ &= \Delta_*^{-1}.0 \\ &= 0.\end{aligned}$$

Thus b is a consistent estimator of β. The same is true for maximum likelihood estimators also.

(ii) Consistency of s^2

Now we look at the consistency of s^2 as an estimate of σ^2. We have

$$\begin{aligned} s^2 &= \frac{1}{n-k} e'e \\ &= \frac{1}{n-k} \varepsilon' \bar{H} \varepsilon \\ &= \frac{1}{n}\left(1-\frac{k}{n}\right)^{-1} \left[\varepsilon'\varepsilon - \varepsilon' X (X'X)^{-1} X'\varepsilon\right] \\ &= \left(1-\frac{k}{n}\right)^{-1} \left[\frac{\varepsilon'\varepsilon}{n} - \frac{\varepsilon' X}{n}\left(\frac{X'X}{n}\right)^{-1} \frac{X'\varepsilon}{n}\right]. \end{aligned}$$

Note that $\frac{\varepsilon'\varepsilon}{n}$ consists of $\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2$ and $\{\varepsilon_i^2, i=1,2,...,n\}$ is a sequence of independently and identically distributed random variables with mean σ^2. Using the law of large numbers

$$\text{plim}\left(\frac{\varepsilon'\varepsilon}{n}\right) = \sigma^2$$

$$\begin{aligned} \text{plim}\left[\frac{\varepsilon' X}{n}\left(\frac{X'X}{n}\right)^{-1} \frac{X'\varepsilon}{n}\right] &= \left(\text{plim}\frac{\varepsilon' X}{n}\right)\left[\text{plim}\left(\frac{X'X}{n}\right)^{-1}\right]\left(\text{plim}\frac{X'\varepsilon}{n}\right) \\ &= 0.\Delta_*^{-1}.0 \\ &= 0 \end{aligned}$$

$$\begin{aligned} \Rightarrow \text{plim}(s^2) &= (1-0)^{-1}\left[\sigma^2 - 0\right] \\ &= \sigma^2. \end{aligned}$$

Thus s^2 is a consistent estimator of σ^2. The same holds true for maximum likelihood estimates also.

Asymptotic distributions:

Suppose we have a sequence of random variables $\{\alpha_n\}$ with a corresponding sequence of cumulative density functions $\{F_n\}$ for a random variable α with cumulative density function F. Then α_n converges in distribution to α if F_n converges to F point wise. In this case, F is called the asymptotic distribution of α_n.

Note that since convergence in probability implies the convergence in distribution, so $\text{plim}\ \alpha_n = \alpha \Rightarrow \alpha_n \xrightarrow{D} \alpha$ (α_n tend to α in distribution), i.e., the asymptotic distribution of α_n is F which is the distribution of α.

Note that

$E(\alpha)$: Mean of asymptotic distribution

$Var(\alpha)$: Variance of asymptotic distribution

$\lim_{n\to\infty} E(\alpha_n)$: Asymptotic mean

$\lim_{n\to\infty} E\left[\alpha_n - \lim_{n\to\infty} E(\alpha_n)\right]^2$: Asymptotic variance.

Asymptotic distribution of sample mean and least squares estimation

Let $\alpha_n = \bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$ be the sample mean based on a sample of size n. Since the sample mean is a consistent estimator of the population mean $\bar{Y}$, so

$$\text{plim}\, \bar{Y}_n = \bar{Y}$$

which is constant. Thus the asymptotic distribution of $\bar{Y}_n$ is the distribution of a constant. This is not a regular distribution as all the probability mass is concentrated at one point. Thus as sample size increases, the distribution of $\bar{Y}_n$ collapses.

Suppose consider only the one-third observations in the sample and find the sample mean as

$$\bar{Y}_n^* = \frac{3}{n}\sum_{i=1}^{\frac{n}{3}} Y_i.$$

Then $E\left(\bar{Y}_n^*\right) = \bar{Y}$

and
$$\begin{aligned} Var\left(\bar{Y}_n^*\right) &= \frac{9}{n^2}\sum_{i=1}^{\frac{n}{3}} Var\left(Y_i\right) \\ &= \frac{9}{n^2}\frac{n}{3}\sigma^2 \\ &= \frac{3}{n}\sigma^2 \\ &\to 0 \text{ as } n \to \infty. \end{aligned}$$

Thus $\text{plim}\, \bar{Y}_n^* = \bar{Y}$ and Y_n^* has the same degenerate distribution as $\bar{Y}_n$. Since $Var\left(\bar{Y}_n^*\right) > Var\left(\bar{Y}_n\right)$, so $\bar{Y}_n$ is preferred over $\bar{Y}_n^*$.

Now we observe the asymptotic behaviour of $\bar{Y}_n$ and $\bar{Y}_n^*$. Consider a sequence of random variables $\{\alpha_n\}$.

Thus for all n, we have

$$\alpha_n = \sqrt{n}\left(\bar{Y}_n - \bar{Y}\right)$$
$$\alpha_n^* = \sqrt{n}\left(\bar{Y}_n^* - \bar{Y}\right)$$
$$E\left(\alpha_n\right) = \sqrt{n}\ E\left(\bar{Y}_n - \bar{Y}\right) = 0$$
$$E\left(\alpha_n^*\right) = \sqrt{n}\ E\left(\bar{Y}_n^* - \bar{Y}\right) = 0$$
$$Var\left(\alpha_n\right) = nE\left(\bar{Y}_n - \bar{Y}\right)^2 = n\ \frac{\sigma^2}{n} = \sigma^2$$
$$Var\left(\alpha_n^*\right) = nE\left(\bar{Y}_n^* - \bar{Y}\right)^2 = n\ \frac{3\sigma^2}{n} = 3\sigma^2.$$

Assuming the population to be normal, the asymptotic distribution of

- α_n is $N\left(0, \sigma^2\right)$
- α_n^* is $N\left(0, 3\sigma^2\right)$.

So now $\bar{Y}_n$ is preferable over $\bar{Y}_n^*$. The central limit theorem can be used to show that α_n will have an asymptotically normal distribution even if the population is not normally distributed.

Also, since

$$\sqrt{n}\left(\bar{Y}_n - \bar{Y}\right) \sim N\left(0, \sigma^2\right)$$
$$\Rightarrow Z = \frac{\sqrt{n}\left(\bar{Y}_n - \bar{Y}\right)}{\sigma} \sim N\left(0,1\right)$$

and this statement holds true in the finite sample as well as asymptotic distributions.

Consider the ordinary least squares estimate $b = \left(X'X\right)^{-1} X'y$ of β in a linear regression model $y = X\beta + \varepsilon$. If X is nonstochastic then the finite covariance matrix of b is

$$V(b) = \sigma^2 (X'X)^{-1}.$$

The asymptotic covariance matrix of b under the assumption that $\lim_{n\to\infty} \frac{X'X}{n} = \Sigma_{xx}$ exists and is nonsingular. It is given by

$$\begin{aligned}\sigma^2 \lim_{n\to\infty} (X'X) &= \sigma^2 \lim_{n\to\infty} \left(\frac{1}{n}\right) \lim_{n\to\infty} \left(\frac{X'X}{n}\right)^{-1} \\ &= \sigma^2 .0.\Sigma_{xx}^{-1} \\ &= 0\end{aligned}$$

which is a null matrix.

Consider the asymptotic distribution of $\sqrt{n}(b-\beta)$. Then even if ε is not necessarily normally distributed, then asymptotically

$$\sqrt{n}(b-\beta) \sim N\left(0, \sigma^2 \Sigma_{xx}^{-1}\right)$$

$$\frac{n(b-\beta)'\Sigma_{xx}(b-\beta)}{\sigma^2} \sim \chi_k^2.$$

If $\frac{X'X}{n}$ is considered as an estimator of Σ_{xx}, then

$$\frac{n(b-\beta)'\frac{X'X}{n}(b-\beta)}{\sigma^2} = \frac{(b-\beta)'X'X(b-\beta)}{\sigma^2}$$

is the usual test statistic as is in the case of finite samples with $b \sim N\left(\beta, \sigma^2 (X'X)^{-1}\right)$.