1) Standard error consistency and test statistic asymptotic normality in linear models

Consider the model for the observable data \( \{y_t, x_t^T\}_{t=1}^n \):

\[
Y = X \beta + U,
\]

where \( \beta \) is a \( k \times 1 \) parameter vector. Let \( X \) be the \( n \times k \) matrix of covariates and let \( U \) be the \( n \times 1 \) vector of latent errors, where \( \mathbb{E}(UU^T|X) = \sigma^2 I_n \). Let \( \hat{\beta}_n \) denote the OLS estimator of \( \beta \). For this model, consider the following conditions.

C1:

\[
n^{-1} (X^T X) \xrightarrow{p} M,
\]

where \( M \) is a nonsingular matrix.

C2:

\[
n^{-1} (X^T U) \xrightarrow{p} 0,
\]

where \( 0 \) is the zero vector of dimension \( k \).

C3:

\[
n^{-\frac{1}{2}} (X^T U) \rightsquigarrow N (0, S),
\]

where \( \rightsquigarrow \) denotes convergence in distribution and \( S \) is a \( k \times k \) matrix.

C4:

\[
n^{-1} (U^T U) \xrightarrow{p} \sigma^2,
\]

where \( \sigma^2 > 0 \).

a) Establish consistency for \( \hat{\beta}_n \) under C1-C3.

b) Derive \( S \).

c) Establish \( n^{\frac{1}{2}} \left( \hat{\beta}_n - \beta \right) \rightsquigarrow N (0, V) \) under C1-C3. Derive \( V \).

d) Let \( \tilde{U} = Y - X \hat{\beta}_n \) and let \( \hat{\sigma}^2 = n^{-1} \tilde{U}^T \tilde{U} \). Establish consistency of the variance estimator

\[
\tilde{V}_n = (n^{-1}X^T X)^{-1} [n^{-1}X^T (\hat{\sigma}^2 I_n) X] (n^{-1}X^T X)^{-1},
\]

under C1-C4.
Proposed Answer to Question 1

a) First
\[ \hat{\beta}_n = (X^T X)^{-1} (X^T Y) = \beta + (X^T X)^{-1} (X^T U). \]
For consistency,
\[ \hat{\beta}_n - \beta = [n^{-1} (X^T X)]^{-1} [n^{-1} (X^T U)]. \]
Therefore, by C1, C2 and the continuous mapping theorem
\[ \left( \hat{\beta}_n - \beta \right) \xrightarrow{p} 0, \]
which implies \( \hat{\beta}_n \xrightarrow{p} \beta. \)

b) We have
\[ S := \lim_{n \to \infty} n^{-1} \text{Var}(X^T U) = \lim_{n \to \infty} n^{-1} \mathbb{E}(X^T U U^T X). \]
If \( \mathbb{E}(U U^T | X) = \sigma^2 I_n \), then
\[ \lim_{n \to \infty} n^{-1} \mathbb{E}(X^T U U^T X) = \lim_{n \to \infty} n^{-1} \mathbb{E}(X^T \mathbb{E}(U U^T | X) X) = \sigma^2 \lim_{n \to \infty} n^{-1} \mathbb{E}(X^T X) = \sigma^2 M. \]

c) For asymptotic normality,
\[ n^{\frac{1}{2}} \left( \hat{\beta}_n - \beta \right) = [n^{-1} (X^T X)]^{-1} \left[ n^{-\frac{1}{2}} (X^T U) \right]. \]
Therefore, by C1, C3 and Slutsky’s Theorem
\[ n^{\frac{1}{2}} \left( \hat{\beta}_n - \beta \right) \sim N \left( 0, M^{-1} S M^{-1} \right). \]
Precisely, the limit distribution has variance \( M^{-1} S (M^T)^{-1} \), where \( M \) is a symmetric matrix. Because \( S = \sigma^2 M, V = \sigma^2 M^{-1}. \)

d) First, observe that the estimator is simplified as
\[ \hat{V}_n = \sigma^2 \left( n^{-1} X^T X \right)^{-1} \left[ n^{-1} X^T X \right] \left( n^{-1} X^T X \right)^{-1} = \sigma^2 \left( n^{-1} X^T X \right)^{-1}. \]
We then have
\[ \hat{U} \hat{U} = \left[ (Y - X \beta) - X \left( \hat{\beta}_n - \beta \right) \right] \left[ (Y - X \beta) - X \left( \hat{\beta}_n - \beta \right) \right]^T \]
\[ = U^T U - U^T X \left( \hat{\beta}_n - \beta \right) \left( \hat{\beta}_n - \beta \right)^T X^T U + \left( \hat{\beta}_n - \beta \right)^T X^T X \left( \hat{\beta}_n - \beta \right). \]

Taking each term separately,
\[ n^{-1} \left( U^T U \right) \xrightarrow{p} \sigma^2, \]
by C4;
\[ n^{-1} U^T X \left( \hat{\beta}_n - \beta \right) \xrightarrow{p} \sigma^2, \]
by C2, result 1 that \( \left( \hat{\beta}_n - \beta \right) \xrightarrow{p} 0 \), and the continuous mapping theorem;
\[ \left( \hat{\beta}_n - \beta \right)^T n^{-1} X^T U, \]
by C2, result 1 that \( \left( \hat{\beta}_n - \beta \right) \xrightarrow{p} 0 \), and the continuous mapping theorem; and
\[ \left( \hat{\beta}_n - \beta \right)^T n^{-1} X^T X \left( \hat{\beta}_n - \beta \right), \]
by C1, result 1 that \( \left( \hat{\beta}_n - \beta \right) \xrightarrow{p} 0 \), and the continuous mapping theorem. Hence \( \sigma^2 \xrightarrow{p} \sigma^2 \) and
\[ \hat{V}_n = \sigma^2 \left( n^{-1} X^T X \right)^{-1} \xrightarrow{p} \sigma^2 M^{-1}. \]
2) Identification, Consistency, Variance Estimation, Consistency in Intercept Model

You are an analyst who needs to estimate how many iPhones are being produced. The structure: In each month, \( J \) iPhones are produced and each phone carries a unique serial number indexed by \( j \), with \( j \in \{1, 2, \ldots, J\} \).

The data: For each month \( t \), where \( t = 1, \ldots, n \), you observe one phone, selected at random from the \( J \) phones produced in that month. Thus your data set consists of \( T \) independent observations \( \{y_t\}_{t=1}^n \).

The model:

\[
Y_t = \beta + U_t.
\]

where \( \beta \) is a scalar parameter. Let \( \hat{\beta}_n \) denote the OLS estimator of \( \beta \). Let \( U \) be the \( n \times 1 \) vector of latent errors.

a) Interpret \( \beta \). What assumption is needed to identify \( \beta \)?

b) What is the distribution of \( U_t \)? Verify that the distribution you derive satisfies the identification assumption from part a.

c) Derive \( \hat{\beta}_n \). Explain why the OLS estimator is also a method-of-moments estimator. Establish that \( \hat{\beta}_n \) is a consistent estimator of \( \beta \).

d) Establish that \( \hat{\beta}_n \) is asymptotically normal and derive the variance of the asymptotic distribution. Construct an estimator of this variance and establish that your variance estimator is consistent.

Extra Credit) How could \( \hat{\beta}_n \) be used to form an estimator of \( J \)? What is the variance of this estimator?
Proposed Answer to Question 2

a) The coefficient $\beta$ captures the location of the distribution of serial numbers. As we typically model the mean of the dependent variable, in the typical setting $\beta$ is the mean of the distribution of serial numbers. For the case at hand, $Y_t$ is a discrete random variable that takes the values $1, \ldots, J$, so

$$
\mathbb{E}(Y) = J^{-1} \sum_{j=1}^{J} j = J^{-1} \frac{J(J+1)}{2} = \frac{(J + 1)}{2}.
$$

For identification,

$$
\mathbb{E}(Y) = \beta + \mathbb{E}(U),
$$

so the identification condition is $\mathbb{E}(U) = 0$.

b) Because

$$
U_t = Y_t - \beta,
$$

the potential values of $U_t$ are

$$
1 - \mathbb{E}(Y), \ldots, J - \mathbb{E}(Y),
$$

each of which is equally likely under the fact that the number is chosen randomly, and so occurs with probability $J^{-1}$. By direct calculation, the mean of the distribution of $U_t$ is

$$
\mathbb{E}(U_t) = J^{-1} \sum_{j=1}^{J} (j - \mathbb{E}(Y)) = J^{-1} \sum_{j=1}^{J} j - \mathbb{E}(Y) = 0,
$$

which satisfies the identification assumption.

c) The OLS estimator is defined as

$$
\hat{\beta}_n = \arg \min_b \sum_{t=1}^{n} (y_t - b)^2,
$$

for which the first-order condition is

$$
-2 \sum_{t=1}^{n} (y_t - \hat{\beta}_n) = 0 \Rightarrow \hat{\beta}_n = n^{-1} \sum_{t=1}^{n} y_t.
$$
The OLSE is the sample mean, which is in turn the sample analog of the population mean, and so the OLSE is a method-of-moments estimator. To establish consistency

\( \left( \hat{\beta}_n - \beta \right) = n^{-1} \sum_{t=1}^{n} y_t - \mathbb{E}(Y_t) . \)

Because the data are i.i.d., a law of large numbers implies

\[ n^{-1} \sum_{t=1}^{n} y_t \xrightarrow{p} \mathbb{E}(Y_t) \]

which implies \( \hat{\beta}_n \xrightarrow{p} \beta . \)

d) To establish asymptotic normality

\[ \sqrt{n} \left( \hat{\beta}_n - \beta \right) = \sqrt{n} \left( n^{-1} \sum_{t=1}^{n} (y_t - \beta) \right) = n^{-\frac{1}{2}} \sum_{t=1}^{n} u_t . \]

Because the data are i.i.d., a central limit theorem implies

\[ n^{-\frac{1}{2}} \sum_{t=1}^{n} u_t \sim N \left( 0, \sigma^2 \right) , \]

where \( \sigma^2 := Var \left( U_t \right) . \) Hence \( \sqrt{n} \left( \hat{\beta}_n - \beta \right) \sim N \left( 0, \sigma^2 \right) \) and the variance of the asymptotic distribution is \( \sigma^2 . \) The estimator is

\[ \hat{\sigma}_n^2 \text{ with } \hat{\sigma}_n^2 := n^{-1} \sum_{t=1}^{n} \left( y_t - \hat{\beta}_n \right)^2 . \]

For consistency of the estimator of the asymptotic variance,

\[ y_t - \hat{\beta}_n = u_t - \left( \hat{\beta}_n - \beta \right) , \]

so

\[ n^{-1} \sum_{t=1}^{n} \left( y_t - \hat{\beta}_n \right)^2 = n^{-1} \sum_{t=1}^{n} u_t^2 - 2 \left( \hat{\beta}_n - \beta \right) n^{-1} \sum_{t=1}^{n} u_t + \left( \hat{\beta}_n - \beta \right)^2 . \]
Because $U_t$ is an i.i.d. random variable,

$$n^{-1} \sum_{t=1}^{n} u_t^2 \xrightarrow{P} \sigma^2 \text{ and } n^{-1} \sum_{t=1}^{n} u_t \xrightarrow{P} 0,$$

and by the consistency of $\hat{\beta}_n$ for $\beta$,

$$\left( \hat{\beta}_n - \beta \right) \xrightarrow{P} 0.$$

Hence the continuous mapping theorem implies

$$2 \left( \hat{\beta}_n - \beta \right) n^{-1} \sum_{t=1}^{n} u_t \xrightarrow{P} 0 \text{ and } \left( \hat{\beta}_n - \beta \right)^2 \xrightarrow{P} 0,$$

so $\hat{\sigma}^2_n \xrightarrow{P} \sigma^2$.

extra credit) Because

$$\mathbb{E} (Y_t) = \frac{J + 1}{2},$$

an estimator of $J$ is

$$\hat{J} = 2\hat{\beta}_n - 1.$$

The estimator is consistent, by consistency of $\hat{\beta}_n$, asymptotically normal, by the asymptotic normality of $\hat{\beta}_n$, and the asymptotic variance of the estimator is $4\sigma^2/n$ (which can be consistently estimated with use of $\hat{\sigma}^2_n$).