1

# Regression Analysis II: Properties of OLS Estimators


Hüseyin Taştan[1]

[1]Department of Economics
Ph.D. Program - Advanced Econometrics
Yildiz Technical University


September 24, 2013

2

## Properties of OLS Estimators

- What are the statistical properties of OLS estimators?
- Finite sample properties
  - Unbiasedness
  - Efficiency (minimum variance)
  - Gauss-Markov Theorem
- Asymptotic properties
  - Consistency
  - Asymptotic Efficiency
  - Asymptotic Normality

3

## Unbiasedness

Under the first three CLRM assumptions OLS estimator $\hat{\boldsymbol{\beta}}$ is unbiased. This is relatively easy to prove:

- OLS estimator is

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{y}$$

- Substituting the true model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top (\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u})$$

- Rearranging

$$\hat{\boldsymbol{\beta}} = \underbrace{(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{X}}_{\boldsymbol{I}_k}\boldsymbol{\beta} + (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{u}$$

- Thus, the OLS estimator can be written as

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{u}$$

- OLS estimator is a linear function of unobserved error terms.

4

## Unbiasedness

How should we treat the matrix $\boldsymbol{X}$?

1. $\boldsymbol{X}$ is nonstochastic or fixed: reasonable assumption for experiments. The experimenter chooses the values in $\boldsymbol{X}$. This assumption is not reasonable for econometric models.
2. Columns of $\boldsymbol{X}$ are exogenous: randomness in the DGP that generated $\boldsymbol{X}$ is independent of the error terms $\boldsymbol{u}$ in the DGP for $\boldsymbol{y}$. This implies $\mathsf{E}(\boldsymbol{u}|\boldsymbol{X}) = \boldsymbol{0}$. This assumption usually applies to cross-sectional data.
3. $\boldsymbol{X}_t$ is exogenous: $\mathsf{E}(u_t|\boldsymbol{X}_t) = 0$ in the context of time series data. This assumption says the error term $u_t$ cannot depend on the explanatory variables at time $t$.

# Unbiasedness

- We have shown that
$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{u}$$
- Take the conditional expectation w.r.t. $\boldsymbol{X}$
$$\mathsf{E}\left[\hat{\boldsymbol{\beta}}|\boldsymbol{X}\right] = \boldsymbol{\beta} + \mathsf{E}\left[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{u}|\boldsymbol{X}\right]$$
- Since we condition on $\boldsymbol{X}$ we can write
$$\mathsf{E}\left[\hat{\boldsymbol{\beta}}|\boldsymbol{X}\right] = \boldsymbol{\beta} + \boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \mathsf{E}\left[\boldsymbol{u}|\boldsymbol{X}\right]$$
- Using exogeneity assumption:
$$\mathsf{E}\left[\hat{\boldsymbol{\beta}}|\boldsymbol{X}\right] = \boldsymbol{\beta} + \boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \underbrace{\mathsf{E}\left[\boldsymbol{u}|\boldsymbol{X}\right]}_{=\mathbf{0}_n}$$
- Thus, we get $\mathsf{E}\left[\hat{\boldsymbol{\beta}}|\boldsymbol{X}\right] = \boldsymbol{\beta}$
- Applying the law of iterated expectations we obtain $\mathsf{E}\left(\hat{\boldsymbol{\beta}}\right) = \boldsymbol{\beta}$



# Monte Carlo Simulation

- Data Generating Process (DGP):
$$y_i = \beta_1 + \beta_2 x_i + u_i, \quad i = 1, 2, \dots, n$$
$\beta_1 = 1,\ \beta_2 = 2,\ x \sim\ Uniform(0, 10),\ u \sim\ N(0, 1)$
- Number of Monte Carlo Replications = 10000
- Results:

Summary statistics

| True Value | Mean | Std | Min | Max |
|---|---|---|---|---|
| 1.0000 | 1.0037 | 0.2818 | -0.0750 | 2.0792 |
| 2.0000 | 1.9996 | 0.0515 | 1.8082 | 2.1970 |

- Sampling distributions are shown in Figures 1 and 2.

see `OLSMC1.m`



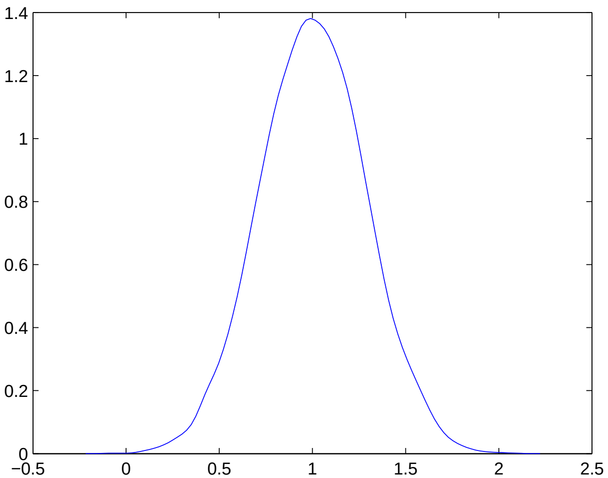


Figure 1 : Sampling Distribution of $\hat{\beta}_1$



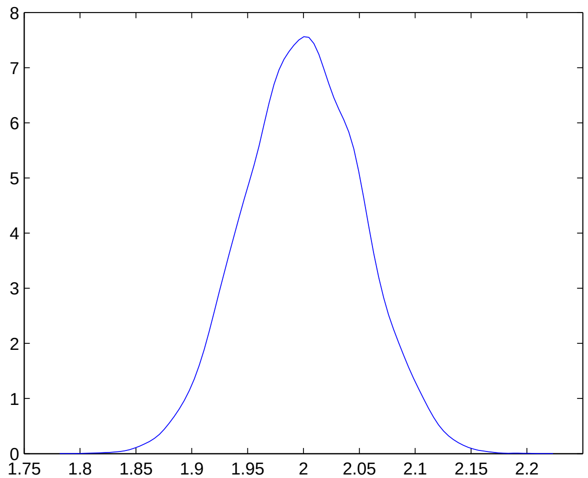


Figure 2 : Sampling Distribution of $\hat{\beta}_2$

## OLS Estimator Can Be Biased

- We have shown that under certain assumptions OLS is unbiased. This is good news.
- The bad news is that the crucial assumption of exogeneity $\mathsf{E}\left[\boldsymbol{u}|\boldsymbol{X}\right] = \mathbf{0}_n$ may not hold in practice.
- There are several cases where this assumption may fail:
- Functional form misspecification
- Omitting relevant variables
- Measurement error in $\boldsymbol{X}$
- Simultaneity (feedback from $y$ to $x$)



## Monte Carlo Simulation: OLS estimation of AR(1) model

- Data Generating Process (DGP):

$$y_t = \phi_0 + \phi_1 y_{t-1} + u_t, \quad t = 1, 2, \ldots, n$$

$\phi_0 = 1,\ \phi_1 = 0.8,\ y_0 = \mathsf{E}(y),\ u \sim \ iidN(0,1)$

- Initial 200 simulated observations are discarded to lessen the impact of initial conditions.
- Number of Monte Carlo Replications = 10000
- Results:

Summary statistics

| True Value | Mean | Std | Min | Max |
|---|---|---|---|---|
| 1.0000 | 1.6237 | 0.8122 | -0.5283 | 5.9482 |
| 0.8000 | 0.6751 | 0.1498 | -0.1722 | 1.0595 |

- Sampling distributions are shown in Figures 3 and 4.

see `OLSMC1.m`



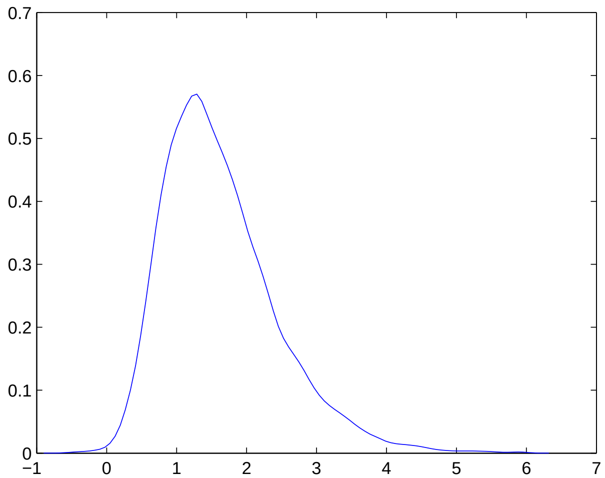


Figure 3 : Sampling Distribution of $\hat{\phi}_0$



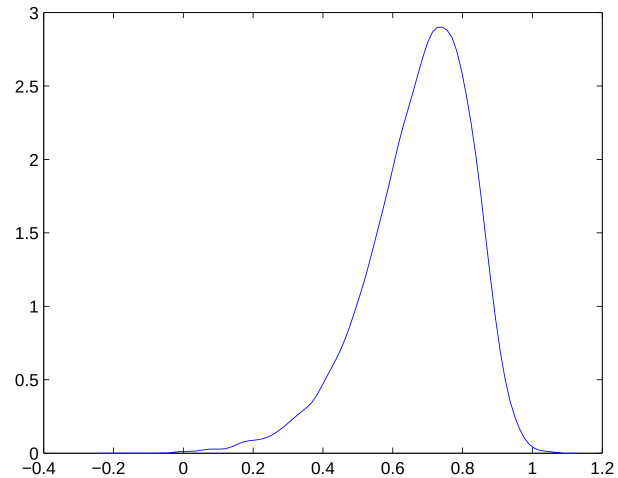


Figure 4 : Sampling Distribution of $\hat{\phi}_1$

## Variance-covariance Matrix of OLS estimator

Recall the fourth CLRM assumption:

$$\mathsf{V}(\boldsymbol{u}|\boldsymbol{X}) = \mathsf{E}(\boldsymbol{u}\boldsymbol{u}^\top|\boldsymbol{X}) = \sigma^2 \boldsymbol{I}_n$$

which simply says that there is **no heteroskedasticity** and **no serial correlation** in the errors. Conditional on $\boldsymbol{X}$:

$$\begin{aligned}\mathsf{E}(\boldsymbol{u}\boldsymbol{u}^\top) &= \mathsf{E}\left\{\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}\begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix}\right\} \\ &= \mathsf{E}\left\{\begin{bmatrix} u_1^2 & u_1u_2 & \dots & u_1u_n \\ u_2u_1 & u_2^2 & \dots & u_2u_n \\ \vdots & \vdots & \ddots & \vdots \\ u_nu_1 & u_nu_2 & \dots & u_n^2 \end{bmatrix}\right\}\end{aligned}$$



## Variance-covariance Matrix of OLS estimator

$$\begin{aligned}\mathsf{E}(\boldsymbol{u}\boldsymbol{u}^\top) &= \begin{bmatrix} \mathsf{E}(u_1^2) & \mathsf{E}(u_1u_2) & \dots & \mathsf{E}(u_1u_n) \\ \mathsf{E}(u_2u_1) & \mathsf{E}(u_2^2) & \dots & \mathsf{E}(u_2u_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{E}(u_nu_1) & \mathsf{E}(u_nu_2) & \dots & \mathsf{E}(u_n^2) \end{bmatrix} \\ &= \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix} \\ &= \sigma^2 \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \\ &= \sigma^2 \boldsymbol{I}_n\end{aligned}$$



## Variance-covariance Matrix of OLS estimator

Now let us derive the covariance matrix of $\hat{\boldsymbol{\beta}}$.

- Using the definition of covariance matrix for random vectors (all expectations are conditional on $\boldsymbol{X}$):

$$\mathsf{Var}(\hat{\boldsymbol{\beta}}) = \mathsf{E}\left[\left(\hat{\boldsymbol{\beta}} - \mathsf{E}(\hat{\boldsymbol{\beta}})\right)\left(\hat{\boldsymbol{\beta}} - \mathsf{E}(\hat{\boldsymbol{\beta}})\right)^\top\right]$$

- We have shown that the OLS estimator is unbiased: $\mathsf{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$. Substituting into expression above we get

$$\mathsf{Var}(\hat{\boldsymbol{\beta}}) = \mathsf{E}\left[\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)^\top\right]$$

- Recall that we have shown $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{u}$, substituting this

$$\mathsf{Var}(\hat{\boldsymbol{\beta}}) = \mathsf{E}\left[\left((\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{u}\right)\left((\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{u}\right)^\top\right]$$



## Variance-covariance Matrix of OLS estimator

- Now the expression becomes

$$\mathsf{Var}(\hat{\boldsymbol{\beta}}) = \mathsf{E}\left[(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{u}\boldsymbol{u}^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\right]$$

- Taking expectations conditional on $\boldsymbol{X}$

$$\mathsf{Var}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top \underbrace{\mathsf{E}(\boldsymbol{u}\boldsymbol{u}^\top)}_{\sigma^2\boldsymbol{I}_n} \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}$$

- Rearranging we get:

$$\begin{aligned}\mathsf{Var}(\hat{\boldsymbol{\beta}}) &= \sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\underbrace{\boldsymbol{X}^\top\boldsymbol{I}_n\boldsymbol{X}}_{\boldsymbol{X}^\top\boldsymbol{X}}(\boldsymbol{X}^\top\boldsymbol{X})^{-1} \\ &= \sigma^2\underbrace{(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X}}_{\boldsymbol{I}_k}(\boldsymbol{X}^\top\boldsymbol{X})^{-1} \\ &= \sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\end{aligned}$$

## Covariance Matrix of $\hat{\beta}$

- We have shown that the covariance matrix is a function of the error variance and information content of the regressor matrix:
$$\mathsf{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1}$$
- Note that error variance $\sigma^2$ is unknown. We can find an unbiased estimator of the error variance:
$$\begin{aligned} s^2 &= \frac{1}{n-k}SSR \\ &= \frac{1}{n-k}\hat{\boldsymbol{u}}^\top\hat{\boldsymbol{u}} \\ &= \frac{1}{n-k}(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})^\top(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}) \end{aligned}$$
It can easily be shown that $\mathsf{E}(s^2)=\sigma^2$.
- Now **the covariance matrix estimator** of the OLS:
$$\widehat{\mathsf{Var}(\hat{\boldsymbol{\beta}})} = s^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1}$$
This matrix is $k\times k$ square, symmetric and positive definite.



## Residuals and Error Terms

- Residual vector and OLS estimator are defined as
$$\hat{\boldsymbol{u}} = \boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}, \quad \hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}$$
- Combining these two expressions and $\boldsymbol{y}=\boldsymbol{X}\beta+\boldsymbol{u}$ we obtain
$$\begin{aligned} \hat{\boldsymbol{u}} &= \boldsymbol{X}\beta+\boldsymbol{u}-\boldsymbol{X}\left((\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\boldsymbol{X}\beta+\boldsymbol{u})\right) \\ &= \boldsymbol{u}-\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{u} \\ &= \boldsymbol{M}\boldsymbol{u} \end{aligned}$$
where $\boldsymbol{M}=\boldsymbol{I}_n-\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$.
- This simply says that the estimated residual vector is a linear combination of unobserved error terms.
- It can also be shown that $\hat{\boldsymbol{u}}=\boldsymbol{M}\boldsymbol{y}$.
- Using this we can calculate residual sum of squares (SSR).



## Expected Value of SSR

- Now we know that $\hat{\boldsymbol{u}}=\boldsymbol{M}\boldsymbol{u}$. Using this SSR can be written as
$$\begin{aligned} \hat{\boldsymbol{u}}^\top\hat{\boldsymbol{u}} &= \boldsymbol{u}^\top\boldsymbol{M}^\top\boldsymbol{M}\boldsymbol{u} \\ &= \boldsymbol{u}^\top\boldsymbol{M}\boldsymbol{u} \end{aligned}$$
- Since trace is a linear operator we can write
$$\mathsf{E}(\hat{\boldsymbol{u}}^\top\hat{\boldsymbol{u}}) = \mathsf{E}\left(\mathsf{tr}(\boldsymbol{u}^\top\boldsymbol{M}\boldsymbol{u})\right)$$
- Recall the invariance of trace operator under cyclical permutations: $\mathsf{tr}(ABC)=\mathsf{tr}(BCA)=\mathsf{tr}(CAB)$. Using this we can write
$$\mathsf{E}(\hat{\boldsymbol{u}}^\top\hat{\boldsymbol{u}}) = \mathsf{E}\left(\mathsf{tr}(\boldsymbol{u}\boldsymbol{u}^\top\boldsymbol{M})\right) = \mathsf{tr}\left(\boldsymbol{M}\mathsf{E}(\boldsymbol{u}\boldsymbol{u}^\top)\right)$$
the last expression follows from the linearity of tr and E.



## Expected Value of SSR

- Now, using the assumption $\mathsf{E}(\boldsymbol{u}\boldsymbol{u}^\top)=\sigma^2\boldsymbol{I}_n$ and properties of $\boldsymbol{M}$
$$\begin{aligned} \mathsf{E}(\hat{\boldsymbol{u}}^\top\hat{\boldsymbol{u}}) &= \mathsf{tr}(\boldsymbol{M}\sigma^2\boldsymbol{I}_n) = \sigma^2\mathsf{tr}(\boldsymbol{M}) \\ &= \sigma^2\mathsf{tr}\left(\boldsymbol{I}_n-\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\right) \\ &= \sigma^2\left(\mathsf{tr}(\boldsymbol{I}_n)-\mathsf{tr}(\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top)\right) \\ &= \sigma^2\left(\mathsf{tr}(\boldsymbol{I}_n)-\mathsf{tr}((\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X})\right) \\ &= \sigma^2\left(\mathsf{tr}(\boldsymbol{I}_n)-\mathsf{tr}(\boldsymbol{I}_k)\right) \\ &= \sigma^2(n-k) \end{aligned}$$
- Using this an unbiased estimator of $\sigma^2$ is easily found
$$s^2 = \frac{\hat{\boldsymbol{u}}^\top\hat{\boldsymbol{u}}}{n-k} = \frac{1}{n-k}\sum_{i=1}^{n}\hat{u}_i^2$$

## Expected Value of SSR

- Note that
$$\mathsf{E}(s^2) = \frac{1}{n-k}\mathsf{E}(\hat{\boldsymbol{u}}^\top \hat{\boldsymbol{u}}) = \frac{1}{n-k}\sigma^2(n-k) = \sigma^2$$
- The (positive) square root of $s^2$
$$s = \sqrt{s^2}$$
is called the **standard error of the regression**.



## Efficiency of OLS

**Theorem (Gauss-Markov Theorem)**

*Under the first four CLRM assumptions OLS estimator has the minimum variance among all unbiased linear estimators (Best Linear Unbiased Estimator - BLUE). In other words, the OLS estimator $\hat{\boldsymbol{\beta}}$ is more efficient than any other linear unbiased estimator, $\tilde{\boldsymbol{\beta}}$, in the sense that*
$$Var(\tilde{\boldsymbol{\beta}}) - Var(\hat{\boldsymbol{\beta}})$$
*is a positive semidefinite matrix. Note that normality assumption is not required.*



## Proof of Gauss-Markov Theorem

- Define another linear unbiased estimator s.t.
$$\tilde{\boldsymbol{\beta}} = \left[(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top + \boldsymbol{C}\right]\boldsymbol{y}$$
where $\boldsymbol{C}$ is $k \times n$ matrix that depend on $\boldsymbol{X}$.
- For unbiasedness we must have
$$\boldsymbol{C}\boldsymbol{X} = \mathbf{0}_{k\times k}$$
so that
$$\mathsf{E}(\tilde{\boldsymbol{\beta}}) = \mathsf{E}\left\{\left[(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top + \boldsymbol{C}\right](\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u})\right\} = \boldsymbol{\beta}$$
- Note that $\tilde{\boldsymbol{\beta}}$ can be written as
$$\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta} + \boldsymbol{D}\boldsymbol{u}$$
where $\boldsymbol{D} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top + \boldsymbol{C}$.



## Proof of Gauss-Markov Theorem

- Now let us derive the covariance matrix of $\tilde{\boldsymbol{\beta}}$:
$$\begin{aligned} \mathsf{Var}(\tilde{\boldsymbol{\beta}}) &= \mathsf{E}\left[(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top\right] \\ &= \boldsymbol{D}\mathsf{E}(\boldsymbol{u}\boldsymbol{u}^\top)\boldsymbol{D}^\top \\ &= \sigma^2\boldsymbol{D}\boldsymbol{D}^\top \end{aligned}$$
where
$$\boldsymbol{D}\boldsymbol{D}^\top = (\boldsymbol{X}^\top\boldsymbol{X})^{-1} + \boldsymbol{C}\boldsymbol{C}^\top$$
- Now the difference between the two covariance matrices is
$$\begin{aligned} \mathsf{Var}(\tilde{\boldsymbol{\beta}}) - \mathsf{Var}(\hat{\boldsymbol{\beta}}) &= \sigma^2\left[(\boldsymbol{X}^\top\boldsymbol{X})^{-1} + \boldsymbol{C}\boldsymbol{C}^\top\right] - \sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1} \\ &= \sigma^2\boldsymbol{C}\boldsymbol{C}^\top \end{aligned}$$
which is always positive semi-definite because for any nonzero $k \times 1$ vector $\boldsymbol{q}$
$$\boldsymbol{q}^\top\boldsymbol{C}\boldsymbol{C}^\top\boldsymbol{q} = \boldsymbol{f}^\top\boldsymbol{f} \geq 0$$
where $\boldsymbol{f} = \boldsymbol{C}^\top\boldsymbol{q}$.

## Consistency of OLS Estimator

- OLS estimator $\hat{\boldsymbol{\beta}}_n$, indexed by the sample size, is said to be consistent for $\boldsymbol{\beta}$ if, as the sample size $n$ increases
$$\hat{\boldsymbol{\beta}}_n \xrightarrow{p} \boldsymbol{\beta}$$
in the limit.
- In terms of probability limit: $\operatorname{plim}(\hat{\boldsymbol{\beta}}_n) = \boldsymbol{\beta}$.
- Recall that OLS estimator was found as
$$\begin{aligned} \hat{\boldsymbol{\beta}}_n &= \left(\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^\top\right)^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i y_i \\ &\equiv (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y} \end{aligned}$$
- We assume that $\boldsymbol{X}^\top \boldsymbol{X}$ converges to a **positive definite matrix of constants**, $\boldsymbol{Q}$, in the limit:
$$\lim_{n\to\infty} \frac{1}{n} \boldsymbol{X}^\top \boldsymbol{X} = \boldsymbol{Q}$$



## Consistency of OLS Estimator

- OLS estimator can be written as
$$\begin{aligned} \hat{\boldsymbol{\beta}}_n &= \boldsymbol{\beta} + (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{u} \\ &= \boldsymbol{\beta} + \left(\frac{1}{n} \boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \left(\frac{1}{n} \boldsymbol{X}^\top \boldsymbol{u}\right) \end{aligned}$$
- If the assumption made above holds then $\boldsymbol{Q}^{-1}$ exists. By Slutsky's theorem
$$\operatorname{plim} \hat{\boldsymbol{\beta}}_n = \boldsymbol{\beta} + \boldsymbol{Q}^{-1} \operatorname{plim}\left(\frac{1}{n} \boldsymbol{X}^\top \boldsymbol{u}\right)$$
- Now we need to find the plim of the last term in this expression.
- If $\boldsymbol{X}$ is fixed (nonstochastic) in repeated samples
$$\mathsf{E}\left(\frac{1}{n} \boldsymbol{X}^\top \boldsymbol{u}\right) = \frac{1}{n} \boldsymbol{X}^\top \mathsf{E}(\boldsymbol{u}) = \mathbf{0} \quad (k \times 1)$$



## Consistency of OLS Estimator

- Since $\mathsf{E}(\boldsymbol{u}\boldsymbol{u}^\top) = \sigma^2 \boldsymbol{I}_n$ we can write
$$\begin{aligned} \mathsf{Var}\left(\frac{1}{n} \boldsymbol{X}^\top \boldsymbol{u}\right) &= \frac{1}{n} \boldsymbol{X}^\top \mathsf{E}(\boldsymbol{u}\boldsymbol{u}^\top) \boldsymbol{X} \frac{1}{n} \\ &= \frac{\sigma^2}{n} \frac{\boldsymbol{X}^\top \boldsymbol{X}}{n} \end{aligned}$$
- Obviously then
$$\lim_{n\to\infty} \mathsf{Var}\left(\frac{1}{n} \boldsymbol{X}^\top \boldsymbol{u}\right) = \mathbf{0} \cdot \boldsymbol{Q} = \mathbf{0}$$
- In other words $\boldsymbol{X}^\top \boldsymbol{u}$ converges in mean square to zero. This in turn implies that
$$\operatorname{plim}\left(\frac{1}{n} \boldsymbol{X}^\top \boldsymbol{u}\right) = 0$$
and
$$\operatorname{plim} \hat{\boldsymbol{\beta}}_n = \boldsymbol{\beta} + \boldsymbol{Q}^{-1} \mathbf{0} = \boldsymbol{\beta}$$



## Consistency of OLS Estimator

- OLS estimator is consistent
- There are two crucial assumptions: the first one is that each regressor is uncorrelated with the error term: $\mathsf{Cov}(\boldsymbol{X}_j, \boldsymbol{u}) = 0$, $j = 1, 2, \ldots, k$
- And the other crucial assumption is that the matrix $\boldsymbol{X}^\top \boldsymbol{X}/n$ converges to a positive definite matrix of constants.
- In fact, this assumption is stronger than required. For milder conditions on $\boldsymbol{X}$ matrix see Grenander conditions (Greene, p.65).
- As an example, consider a model with a constant and a trend term, and compute the limit.

29

## Consistency of OLS Estimator

- OLS estimator can be **biased but consistent**: consider the AR(1) model

$$y_t = \phi_0 + \phi_1 y_{t-1} + u_t, \quad t = 2, 3, \ldots, n$$

- In finite samples OLS estimator of $\phi_1$ will be biased because strict exogeneity does not hold: $y_{t-1}$ depends on $u_{t-1}$, $u_{t-2}$, ...
- But the OLS estimator is consistent because $y_{t-1}$ and $u_t$ are uncorrelated. $y_{t-1}$ is predetermined but its realized value has no impact on the expectation of $u_t$.
- OLS estimator will be inconsistent if there are important omitted variables, or measurement errors in the regressors.

30

## Monte Carlo Simulation: OLS estimation of AR(1) model

- Data Generating Process (DGP):

$$y_t = \phi_0 + \phi_1 y_{t-1} + u_t, \quad t = 1, 2, \ldots, n$$

$\phi_0 = 1$, $\phi_1 = 0.8$, $y_0 = \mathrm{E}(y)$, $u \sim \; iidN(0, 1)$

- Consider three sample sizes: `n = [50 200 500]`
- Initial 200 simulated observations are discarded to lessen the impact of initial conditions.
- Number of Monte Carlo Replications $=$ 10000
- For each $n$ we simulate AR(1) process and estimate the parameters by OLS 10000 times.
- Kernel density estimates of OLS sampling distributions: Figures 5 and 6.

see `OLSMC1.m`

31

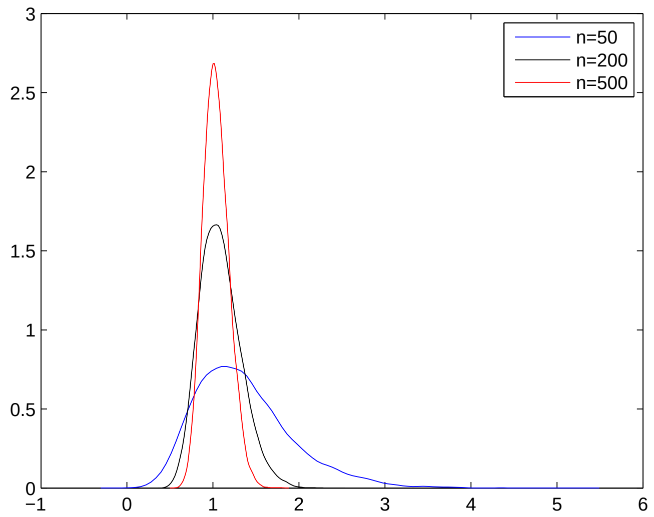


Figure 5 : Sampling Distributions of $\hat{\phi}_0$

32

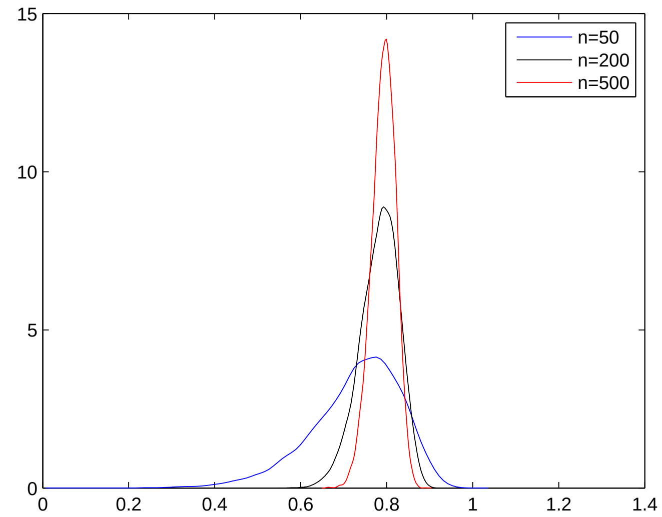


Figure 6 : Sampling Distributions of $\hat{\phi}_1$

33

## Asymptotic Normality of OLS Estimator

- Now we are interested in the distribution of the random vector
$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{d} ?$$
- In other words, what is the asymptotic distribution of $\hat{\boldsymbol{\beta}}_n$:
$$\hat{\boldsymbol{\beta}}_n \overset{a}{\sim} ?$$
- We know that
$$\hat{\boldsymbol{\beta}}_n = \boldsymbol{\beta} + (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{u}$$
- Rewrite this as
$$\begin{aligned}\hat{\boldsymbol{\beta}}_n &= \boldsymbol{\beta} + (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{u}\\ \hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta} &= (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{u}\\ \sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) &= \left(\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1}\left(\frac{1}{\sqrt{n}}\boldsymbol{X}^\top \boldsymbol{u}\right)\end{aligned}$$

34

## Asymptotic Normality of OLS Estimator

- We know $\operatorname{plim}(\boldsymbol{X}^\top \boldsymbol{X}/n)^{-1} = \boldsymbol{Q}^{-1}$.
- But what about the limit distribution of the last term?
- As we saw before $\boldsymbol{X}^\top \boldsymbol{u}$ is just a linear combination of the error term. Under certain conditions we can invoke the Central Limit Theorem to find the limit distribution of this expression.
- If $u_i$ is iid with zero mean and finite variance, $\boldsymbol{X}$ is asymptotically cooperative (Grenander conditions) and $\operatorname{plim}(\boldsymbol{X}^\top \boldsymbol{X}/n) = \boldsymbol{Q}$ then
$$\frac{1}{\sqrt{n}}\boldsymbol{X}^\top \boldsymbol{u} \xrightarrow{d} N(\boldsymbol{0}, \sigma^2 \boldsymbol{Q})$$

35

## Asymptotic Normality of OLS Estimator

- Thus
$$\boldsymbol{Q}^{-1}\frac{1}{\sqrt{n}}\boldsymbol{X}^\top \boldsymbol{u} \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{Q}^{-1}\sigma^2\boldsymbol{Q}^{-1}\boldsymbol{Q})$$
or
$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{d} N(\boldsymbol{0}, \sigma^2\boldsymbol{Q}^{-1})$$
- Therefore
$$\hat{\boldsymbol{\beta}}_n \overset{a}{\sim} N\left(\boldsymbol{\beta},\ \frac{\sigma^2}{n}\boldsymbol{Q}^{-1}\right)$$
- Note that in practice, we approximate $\frac{1}{n}\boldsymbol{Q}^{-1}$ by $(\boldsymbol{X}^\top \boldsymbol{X})^{-1}$.
- **Importance of asymptotic normality of OLS estimator**: If the explanatory variable matrix is asymptotically cooperative then the OLS estimator is asymptotically normal. We do not need the assumption of the normality for the error term. All inference procedures are valid asymptotically.

36

## Numerical Example

- See MATLAB file `OLSnumex1.m`, data file is `OLSnumex1.txt`
- Using the data file the following quantities are easily computed:
$$\boldsymbol{X}^\top \boldsymbol{X} = \begin{bmatrix} 20 & 80 \\ 80 & 468 \end{bmatrix}, \quad (\boldsymbol{X}^\top \boldsymbol{X})^{-1} = \begin{bmatrix} 0.1581 & -0.0270 \\ -0.0270 & 0.0068 \end{bmatrix}$$
$$\boldsymbol{X}^\top \boldsymbol{y} = \begin{bmatrix} 186 \\ 1042 \end{bmatrix}$$
- OLS estimates are
$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{y}\\ &= \begin{bmatrix} 0.1581 & -0.0270 \\ -0.0270 & 0.0068 \end{bmatrix}\begin{bmatrix} 186 \\ 1042 \end{bmatrix}\\ \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} &= \begin{bmatrix} 1.2459 \\ 2.0135 \end{bmatrix}\end{aligned}$$

37

## Numerical Example

- Estimated OLS regression in equation form:

$$\hat{y} = 1.2459 + 2.0135x$$

$$SSR = \hat{\boldsymbol{u}}^\top \hat{\boldsymbol{u}} = 20.173, \quad s^2 = \frac{1}{18} SSR = 1.1207$$

- Covariance matrix:

$$\mathsf{V}(\hat{\boldsymbol{\beta}}) = s^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1} = 1.1207 \begin{bmatrix} 0.1581 & -0.0270 \\ -0.0270 & 0.0068 \end{bmatrix}$$

$$= \begin{bmatrix} 0.1772 & -0.0303 \\ -0.0303 & 0.0076 \end{bmatrix}$$

- Variances:

$$\mathsf{V}(\hat{\beta}_1) = 0.1772, \quad \mathsf{V}(\hat{\beta}_2) = 0.0076, \quad \mathsf{Cov}(\hat{\beta}_1, \hat{\beta}_2) = -0.0303$$

38

## Numerical Example

- Standard errors:

$$se(\hat{\beta}_1) = \sqrt{0.1772} = 0.4209$$

$$se(\hat{\beta}_2) = \sqrt{0.0076} = 0.087$$

- $t$-ratios:

$$t_{\hat{\beta}_1} = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = \frac{1.2459}{0.4209} = 2.9599$$

$$t_{\hat{\beta}_2} = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)} = \frac{2.0135}{0.087} = 23.1386$$

39

## Overspecification: Including Irrelevant Variables in a Regression Model

- What happens if we add an irrelevant variable in the model? (overspecifying the model)
- Irrelevance of the variable means that its coefficient in the population model is zero.
- E.g., suppose that in the regression below the partial effect of $x_3$ is zero, $\beta_3 = 0$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

- Taking the conditional expectation we have:

$$\mathsf{E}(y|x_1, x_2, x_3) = \mathsf{E}(y|x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

- Even though the true model is given above $x_3$ is added to the model by mistake.

40

## Including Irrelevant Variables in a Regression Model

- In this case sample regression function (SRF) is given by

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$$

- OLS estimators are still unbiased:

$$\mathsf{E}(\hat{\beta}_0) = \beta_0, \quad \mathsf{E}(\hat{\beta}_1) = \beta_1, \quad \mathsf{E}(\hat{\beta}_2) = \beta_2, \quad \mathsf{E}(\hat{\beta}_3) = 0,$$

- True parameter value for the irrelevant variable is 0. Since this variable would have no explanatory power the expected value of the OLS estimator will also be zero - i.e. unbiased.
- However, even if they are unbiased, **the variance of the regression will be larger** if the model is overspecified.

41

## Underspecification: Omitting a Relevant Variable

- What happens if we exclude an important variable?
- If a relevant variable is omitted this implies that its parameter is **not** 0 in the PRF. This is called underspecification of the model.
- In this case OLS estimators will, in general, be **biased**.
- E.g. suppose that the true model includes 2 independent variables:
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$
- Suppose that we omitted $x_2$ because, say, it is unobservable. Now the SRF is
$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$$
- Is the OLS estimator on $x_1$, $\tilde{\beta}_1$, still unbiased?

42

## Omitting a Relevant Variable

- The impact of the omitted variable will be included in the error term:
$$y = \beta_0 + \beta_1 x_1 + \nu$$
- True PRF includes $x_2$'yi. Thus the error term $\nu$ can be written as:
$$\nu = \beta_2 x_2 + u$$
- OLS estimator of $\beta_1$ in the model above is:
$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)y_i}{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2}$$
- To determine the magnitude and sign of the bias we will substitute $y$ in the formula for $\tilde{\beta}_1$, re-arrange and take expectation.

43

## Omitting a Relevant Variable

$$\begin{aligned}\tilde{\beta}_1 &= \frac{\sum(x_{i1} - \bar{x}_1)(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i)}{\sum(x_{i1} - \bar{x}_1)^2} \\ &= \beta_1 + \beta_2 \frac{\sum(x_{i1} - \bar{x}_1)x_{i2}}{\sum(x_{i1} - \bar{x}_1)^2} + \frac{\sum(x_{i1} - \bar{x}_1)u_i}{\sum(x_{i1} - \bar{x}_1)^2}\end{aligned}$$

Taking (conditional) expectation we obtain

$$\begin{aligned}E(\tilde{\beta}_1) &= \beta_1 + \beta_2 \frac{\sum(x_{i1} - \bar{x}_1)x_{i2}}{\sum(x_{i1} - \bar{x}_1)^2} + \frac{\sum(x_{i1} - \bar{x}_1)\overbrace{\mathrm{E}(u_i)}^{=0, MLR.3}}{\sum(x_{i1} - \bar{x}_1)^2} \\ &= \beta_1 + \beta_2 \left(\frac{\sum(x_{i1} - \bar{x}_1)x_{i2}}{\sum(x_{i1} - \bar{x}_1)^2}\right)\end{aligned}$$

44

## Omitting a Relevant Variable

$$E(\tilde{\beta}_1) = \beta_1 + \beta_2 \left(\frac{\sum(x_{i1} - \bar{x}_1)x_{i2}}{\sum(x_{i1} - \bar{x}_1)^2}\right)$$

The expression in the parenthesis to the right of $\beta_2$ is just the regression of $x_2$ on $x_1$:

$$\tilde{x}_2 = \tilde{\delta}_0 + \tilde{\delta}_1 x_1$$

Thus

$$E(\tilde{\beta}_1) = \beta_1 + \beta_2 \tilde{\delta}_1$$
$$bias = E(\tilde{\beta}_1) - \beta_1 = \beta_2 \tilde{\delta}_1$$

This is called **omitted variable bias**.

## Omitted Variable Bias

$$bias = E(\tilde{\beta}_1) - \beta_1 = \beta_2\tilde{\delta}_1$$

- Iff $\tilde{\delta}_1 = 0$ or $\beta_2 = 0$ then bias is 0.
- The sign of bias depends on both $\beta_2$ and the correlation between omitted variable ($x_2$) and included variable ($x_1$).
- It is not possible to calculate this correlation if omitted variable cannot be observed.
- The following table summarizes possible cases:

Direction of Bias

| | $Corr(x_1, x_2) > 0$ | $Corr(x_1, x_2) < 0$ |
|---|---|---|
| $\beta_2 > 0$ | positive bias | negative bias |
| $\beta_2 < 0$ | negative bias | positive bias |



## Omitted Variable Bias

$$bias = E(\tilde{\beta}_1) - \beta_1 = \beta_2\tilde{\delta}_1$$

- The size of the bias is also important. It depends on both $\tilde{\delta}_1$ and $\beta_2$.
- A small bias relative to $\beta_1$ may not be a problem in practice.
- But in most cases we are not able to calculate the size of the bias.
- In some cases we may have an idea about the direction of bias. For example, suppose that in the wage equation true PRF contains both education and ability.
- Suppose also that ability is omitted because it cannot be observed, leading to omitted variable bias.
- In this case we can say that sign of the bias is + because it is reasonable to think that people with more ability tend to have higher levels of education and ability is positively related to wage.



## Omitted Variable Bias

- The effect of omitted variable will be in $u$. Thus, exogeneity assumption fails.

$$wage = \beta_0 + \beta_1 educ + \beta_2 ability + u$$

- Instead of this model we estimate

$$wage = \beta_0 + \beta_1 educ + \nu$$

$$\nu = \beta_2 ability + u$$

- Education will be correlated with the error term ($\nu$).

$$\mathrm{E}(\nu|educ) \neq 0$$

- Education is endogenous. If we omit ability the effect of education on wages will be overestimated. Some part of the effect of education on wage comes from ability.



## Omitted Variable Bias

- In models with more than two variables, omitting a relevant variable generally causes OLS estimators to be biased.
- True PRF is given by:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

- $x_3$ is omitted and the following model is estimated

$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + \tilde{\beta}_2 x_2$$

- Also suppose that $x_3$ is correlated with $x_1$ but uncorrelated with $x_2$.
- In this case both $\tilde{\beta}_1$ and $\tilde{\beta}_2$ will be biased. Iff $x_1$ and $x_2$ are correlated then $\tilde{\beta}_2$ will be unbiased.

## Monte Carlo Experiment

- DGP Setup:
$$y = \beta_1 + \beta_2 x_1 + \beta_3 x_2 + u$$
- Let $\boldsymbol{x} = [x_1 \quad x_2]^\top$ be fixed in repeated samples and
$$\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \ \boldsymbol{\mu} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \ \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$$
so that $Cor(x_1, x_2) = 0.9$
- Sample size $n = 50$, the number of Monte Carlo replications is 10000.
- $u_i \sim N(0, 1)$, $i = 1, 2, \ldots, n$.

see `OLSMC2.m`



## Monte Carlo Experiment

- We consider four cases:
- CASE 1: $x_2$ is irrelevant and not included in the sample regression:
$$\beta_1 = 5, \quad \beta_2 = 2, \quad \beta_3 = 0$$
In this case sample regression is correctly specified.
- CASE 2: $x_2$ is irrelevant and included in the sample regression. The same DGP as the Case 1. In this case, sample regression is overspecified.
- CASE 3: $x_2$ is relevant and excluded from the regression:
$$\beta_1 = 5, \quad \beta_2 = 2, \quad \beta_3 = 1$$
In this case the model is underspecified (omitting a relevant variable).
- CASE 4: $x_2$ is relevant and included in the regression. The same DGP as the Case 3. In this case the model correctly specified.



## MC Experiment: Numerical Results

```
CASE 1: x2 is irrelevant and not included in the OLS regression
True values: beta1=5, beta2=2, beta3=0

    betahat1    STD1       betahat2    STD2
    4.9979      0.28638    2.0013      0.20353

CASE 2: x2 is irrelevant and included in the OLS regression
True values: beta1=5, beta2=2, beta3=0

    betahat1    STD1       betahat2    STD2       betahat3     STD3
    4.9969      0.29933    2.0004      0.36052    0.0024772    0.34291

CASE 3:  X2 is relevant and excluded from the OLS regression
True values: beta1=5, beta2=2, beta3=1

    betahat1    STD1       betahat2    STD2
    5.2396      0.29082    2.8654      0.20764

CASE 4: X2 is relevant and included
True values: beta1=5, beta2=2, beta3=1

    betahat1    STD1       betahat2    STD2       betahat3    STD3
    5.0047      0.29718    1.9929      0.36391    1.0019      0.34351
```



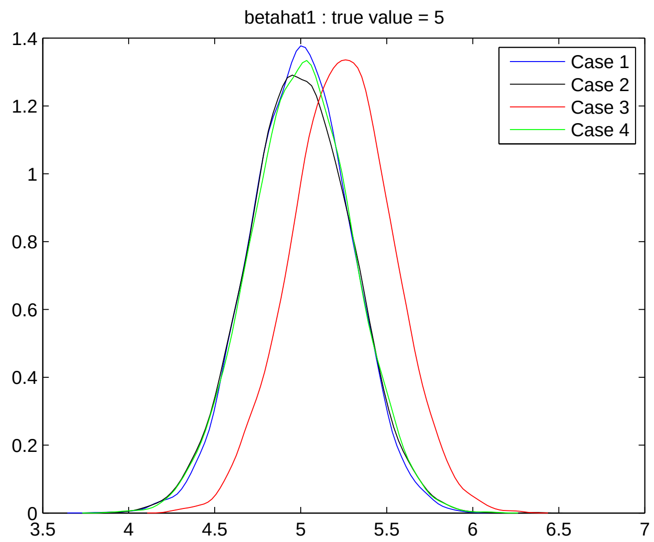


Figure 7 : Sampling Distributions of $\hat{\beta}_1$

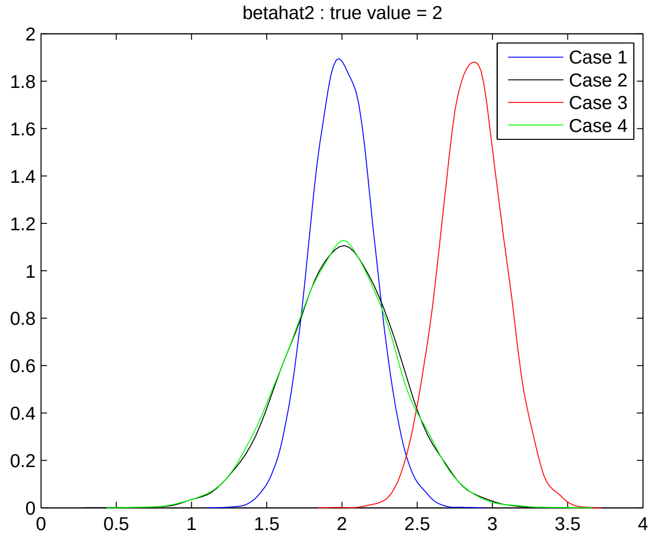


Figure 8 : Sampling Distributions of $\hat{\beta}_2$



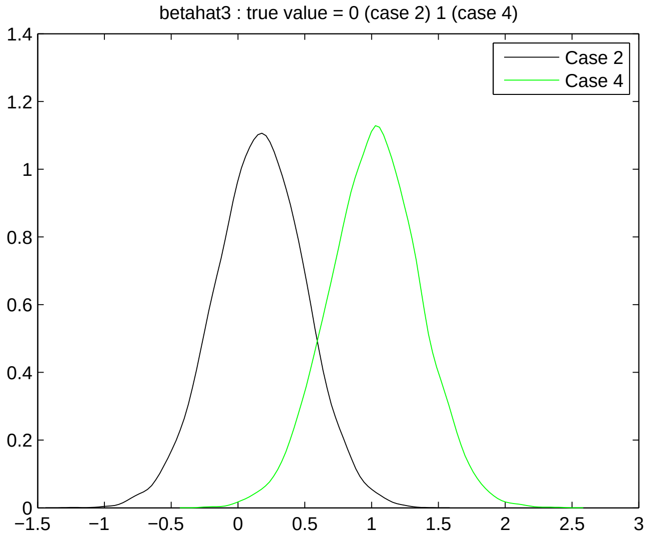


Figure 9 : Sampling Distributions of $\hat{\beta}_3$