Addendum to NOTE 4

Sampling Distributions of OLS Estimators of β_0 and β_1

Monte Carlo Simulations

The True Model: is given by the **population regression equation (PRE)**

$$Y_i = \beta_0 + \beta_1 X_i + u_i = 70.0 + 0.90 X_i + u_i \tag{1}$$

where

$\boldsymbol{\beta_0 = 70.0}$ and $\boldsymbol{\beta_1 = 0.90}$;

Y_i = weekly consumption expenditures of the i-th household;

X_i = weekly disposable income of the i-th household;

u_i = an iid random error term that is assumed to be $N(0, \sigma^2)$.

Four Alternative Models: specify different values for $\sigma^2 = \text{Var}(u_i \mid X_i)$.

Model 1: sets $\sigma^2 = \text{Var}(u_i \mid X_i) = 1{,}600$, $\sigma = \sqrt{\text{Var}(u_i \mid X_i)} = \text{se}(u_i \mid X_i) = 40$.

Model 2: sets $\sigma^2 = \text{Var}(u_i \mid X_i) = 6{,}400$, $\sigma = \sqrt{\text{Var}(u_i \mid X_i)} = \text{se}(u_i \mid X_i) = 80$.

Model 3: sets $\sigma^2 = \text{Var}(u_i \mid X_i) = 25{,}600$, $\sigma = \sqrt{\text{Var}(u_i \mid X_i)} = \text{se}(u_i \mid X_i) = 160$.

Model 4: sets $\sigma^2 = \text{Var}(u_i \mid X_i) = 160{,}000$, $\sigma = \sqrt{\text{Var}(u_i \mid X_i)} = \text{se}(u_i \mid X_i) = 400$.

The Monte Carlo Simulations

- Four different sample sizes: N = 30, N = 60, N = 120, N = 300.

- Set population values of X, β_0 and β_1, and $\sigma^2 = \text{Var}(u_i \mid X_i)$.

- Generate 1,000 independent random samples of Y_i and u_i values.

- For each of these 1,000 independent random samples, compute the values of the OLS coefficient estimators:

$$\hat{\beta}_1 = \frac{\sum_i x_i y_i}{\sum_i x_i^2} \tag{2}$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} \tag{3}$$

where $x_i \equiv X_i - \overline{X}$, $y_i \equiv Y_i - \overline{Y}$, $\overline{X} = \sum_i X_i / N$, and $\overline{Y} = \sum_i Y_i / N$.

- Tabulate and plot the 1,000 estimates of β_1 and the 1,000 estimates of β_0, i.e., the 1,000 values of $\hat{\beta}_1$ and the 1,000 values of $\hat{\beta}_0$.

OLS estimates $\hat{\beta}_1$ of β_1 are denoted as **`b1`**.

OLS estimates $\hat{\beta}_0$ of β_0 are denoted as **`b0`**.

Simulation Results for Sample Size N = 30 Observations (1,000 Replications)

N = 30: Simulation of **Model 1** for which σ^2 = 1,600, σ = 40:

. summarize

Variable	Obs	**Mean**	Std. Dev.	Min	Max
b0	1000	**70.7223**	14.89798	24.52773	112.19
b1	1000	**.8997546**	.0083921	**.873687**	**.9290485**

N = 30: Simulation of **Model 2** for which σ^2 = 6,400, σ = 80:

. summarize

Variable	Obs	**Mean**	Std. Dev.	Min	Max
b0	1000	**71.44459**	29.79595	-20.94459	154.3801
b1	1000	**.8995092**	.0167843	**.847374**	**.9580969**

N = 30: Simulation of **Model 3** for which σ^2 = 25,600, σ = 160:

. summarize

Variable	Obs	**Mean**	Std. Dev.	Min	Max
b0	1000	**72.88918**	59.59191	-111.8891	238.7602
b1	1000	**.8990184**	.0335685	**.7947479**	**1.016194**

N = 30: Simulation of **Model 4** for which σ^2 = 160,000, σ = 400:

. summarize

Variable	Obs	**Mean**	Std. Dev.	Min	Max
b0	1000	**77.22296**	148.9798	-384.7229	491.9006
b1	1000	**.8975461**	.0839213	**.6368698**	**1.190485**

Simulation Results for Sample Size N = 60 Observations (1,000 Replications)

N = 60: Simulation of **Model 1** for which σ^2 = 1,600, σ = 40:

. summarize

Variable	Obs	**Mean**	Std. Dev.	Min	Max
b0	1000	**70.4091**	10.66658	34.24415	109.2541
b1	1000	**.8998245**	.0060534	**.8822982**	**.9191378**

N = 60: Simulation of **Model 2** for which σ^2 = 6,400, σ = 80:

. summarize

Variable	Obs	**Mean**	Std. Dev.	Min	Max
b0	1000	**70.81821**	21.33315	-1.511704	148.5082
b1	1000	**.899649**	.0121069	**.8645963**	**.9382755**

N = 60: Simulation of **Model 3** for which σ^2 = 25,600, σ = 160:

. summarize

Variable	Obs	**Mean**	Std. Dev.	Min	Max
b0	1000	**71.63642**	42.66631	-73.02339	227.0164
b1	1000	**.8992979**	.0242138	**.8291926**	**.976551**

N = 60: Simulation of **Model 4** for which σ^2 = 160,000, σ = 400:

. summarize

Variable	Obs	**Mean**	Std. Dev.	Min	Max
b0	1000	**74.09104**	106.6658	-287.5585	462.541
b1	1000	**.8982448**	.0605344	**.7229815**	**1.091377**

Simulation Results for Sample Size N = 120 Observations (1,000 Replications)

N = 120: Simulation of **Model 1** for which σ^2 = 1,600, σ = 40:

```
. summarize

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+-----------------------------------------------------------
          b0 |      1000    70.01442    7.527844    47.7453    97.8483
          b1 |      1000    .9000406    .0042317   .8859292   .9143825
```

N = 120: Simulation of **Model 2** for which σ^2 = 6,400, σ = 80:

```
. summarize

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+-----------------------------------------------------------
          b0 |      1000    70.02884    15.05569    25.4906   125.6966
          b1 |      1000    .9000811    .0084634   .8718585    .928765
```

N = 120: Simulation of **Model 3** for which σ^2 = 25,600, σ = 160:

```
. summarize

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+-----------------------------------------------------------
          b0 |      1000    70.05768    30.11138  -19.01881   181.3933
          b1 |      1000    .9001622    .0169268    .843717     .95753
```

N = 120: Simulation of **Model 4** for which σ^2 = 160,000, σ = 400:

```
. summarize

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+-----------------------------------------------------------
          b0 |      1000    70.14419    75.27844   -152.547   348.4832
          b1 |      1000    .9004055    .0423171   .7592924   1.043825
```

Simulation Results for Sample Size N = 300 Observations (1,000 Replications)

N = 300: Simulation of **Model 1** for which σ^2 = 1,600, σ = 40:

```
. summarize

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+-----------------------------------------------------------
          b0 |      1000    70.05802    4.697781   53.87795   84.48841
          b1 |      1000    .8999697    .0025922   .8908832   .9090104
```

N = 300: Simulation of **Model 2** for which σ^2 = 6,400, σ = 80:

```
. summarize

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+-----------------------------------------------------------
          b0 |      1000    70.11603    9.395562   37.75589   98.97681
          b1 |      1000    .8999394    .0051844   .8817664   .9180208
```

N = 300: Simulation of **Model 3** for which σ^2 = 25,600, σ = 160:

```
. summarize

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+-----------------------------------------------------------
          b0 |      1000    70.23207    18.79112   5.511793   127.9536
          b1 |      1000    .8998787    .0103688   .8635329   .9360417
```

N = 300: Simulation of **Model 4** for which σ^2 = 160,000, σ = 400:

```
. summarize

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+-----------------------------------------------------------
          b0 |      1000    70.58016    46.97781  -91.22052   214.8841
          b1 |      1000    .8996968     .025922   .8088323   .9901041
```

Effect on Sampling Distributions of $\hat{\beta}_0$ and $\hat{\beta}_1$ of Increasing the Error Variance σ^2

<u>Question</u>: What is the effect on the sampling distributions of $\hat{\beta}_0$ and $\hat{\beta}_1$ of ***increasing*** **the error variance** $\sigma^2 = \text{Var}(u_i \mid X_i)$ while holding constant both sample size N and sample variation of X = $TSS_X \equiv \sum_{i=1}^{N} x_i^2 \equiv \sum_{i=1}^{N}(X_i - \bar{X})^2$?

<u>Answer</u>: ***Increasing*** **the error variance $\sigma^2 = \text{Var}(u_i \mid X_i)$** ***increases*** **the variances of the sampling distributions of $\hat{\beta}_0$ and $\hat{\beta}_1$**, i.e., increases $\text{Var}(\hat{\beta}_0)$ and $\text{Var}(\hat{\beta}_1)$.

Illustration for Sample Size N = 60 Observations

- **Model 1** for which σ^2 = 1,600, σ = 40:

```
. summarize

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+-----------------------------------------------------------
          b0 |      1000     70.4091    10.66658   34.24415   109.2541
          b1 |      1000    .8998245    .0060534   .8822982   .9191378
```

- **Model 2** for which σ^2 = 6,400, σ = 80:

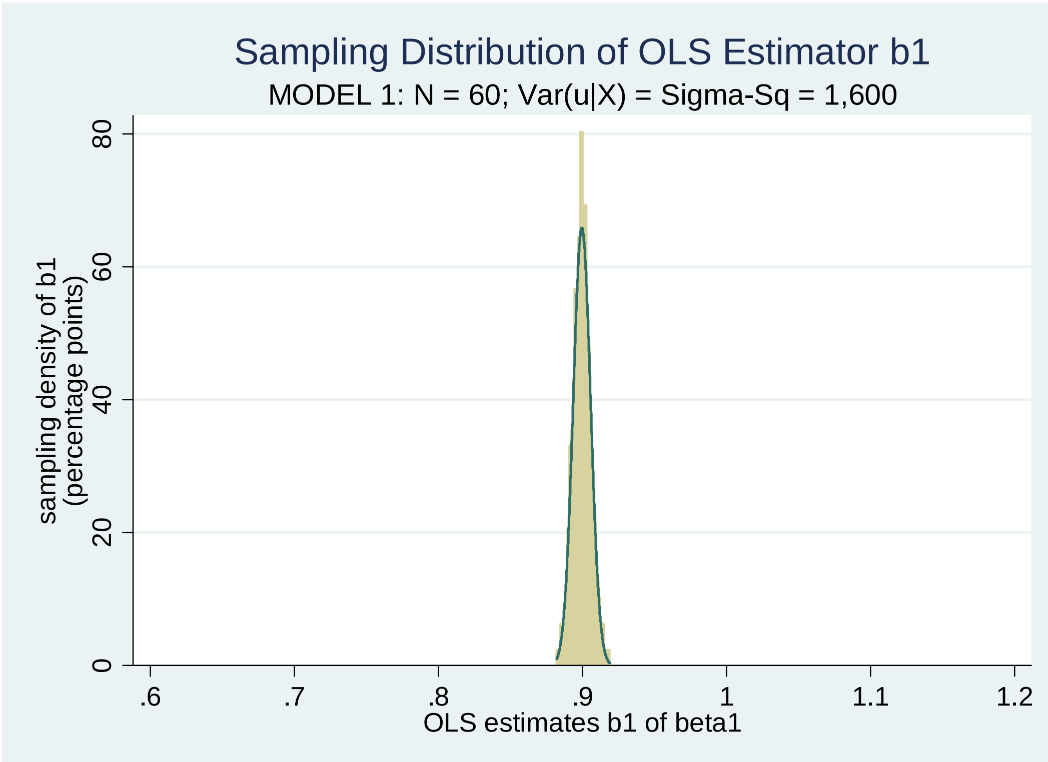
```
. summarize

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+-----------------------------------------------------------
          b0 |      1000    70.81821    21.33315  -1.511704   148.5082
          b1 |      1000     .899649    .0121069   .8645963   .9382755
```

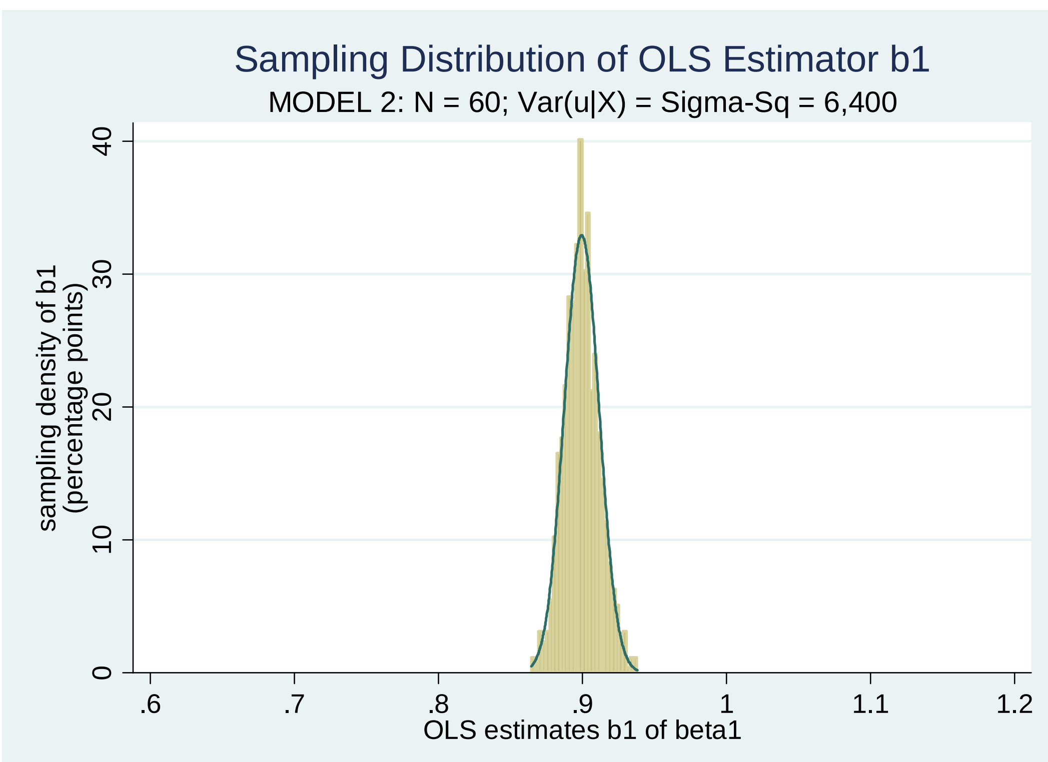
- **Model 3** for which σ^2 = 25,600, σ = 160:

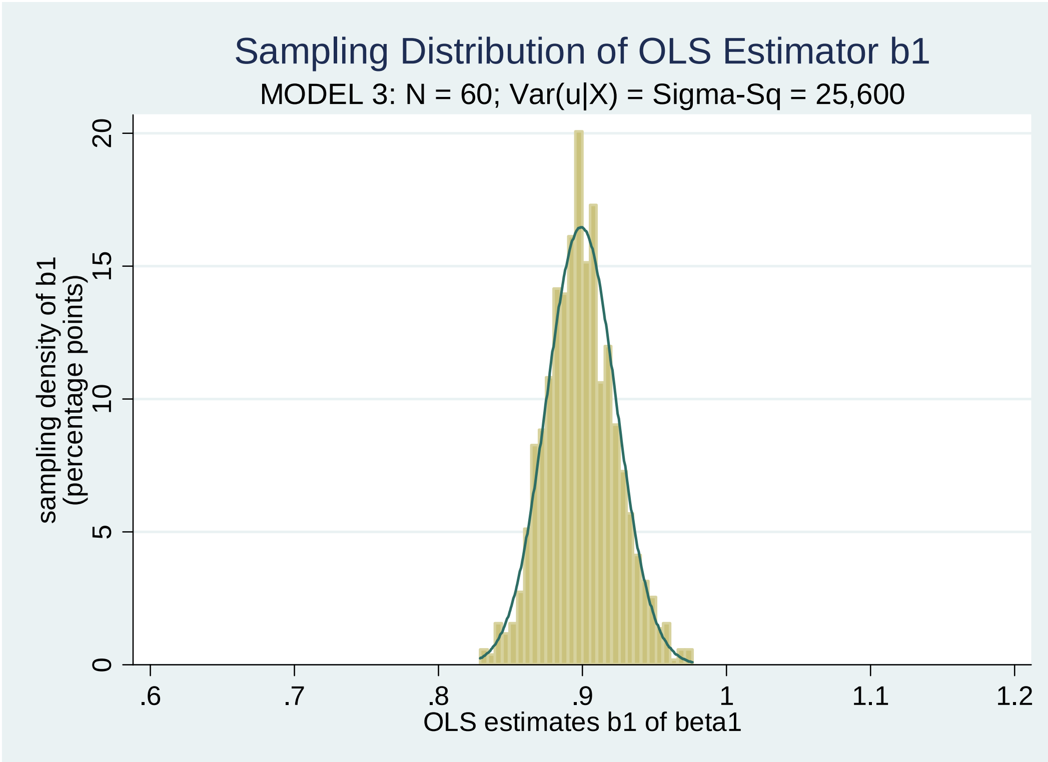
```
. summarize

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+-----------------------------------------------------------
          b0 |      1000    71.63642    42.66631  -73.02339   227.0164
          b1 |      1000    .8992979    .0242138   .8291926    .976551
```

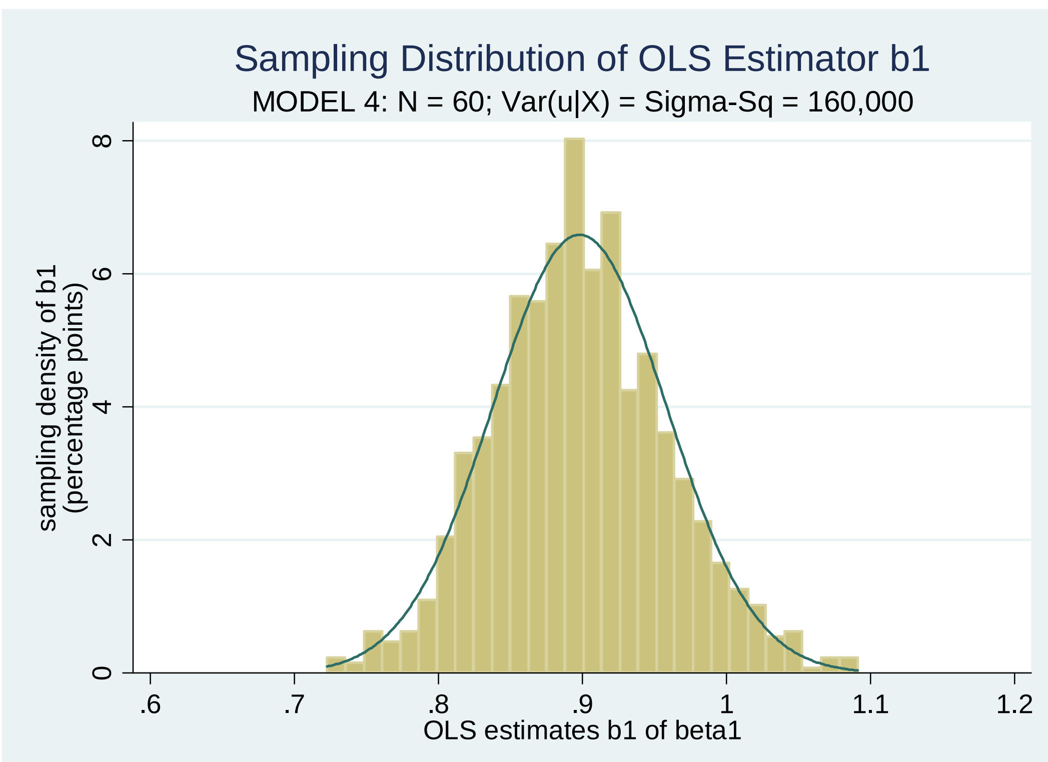
- **Model 4** for which σ^2 = 160,000, σ = 400:

```
. summarize

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+-----------------------------------------------------------
          b0 |      1000    74.09104    106.6658  -287.5585    462.541
          b1 |      1000    .8982448    .0605344   .7229815   1.091377
```


Sampling Distribution of OLS Estimator b1
MODEL 1: N = 60; Var(u|X) = Sigma-Sq = 1,600
sampling density of b1
(percentage points)
0
20
40
60
80
.6
.7
.8
.9
1
1.1
1.2
OLS estimates b1 of beta1


Sampling Distribution of OLS Estimator b1
MODEL 2: N = 60; Var(u|X) = Sigma-Sq = 6,400
sampling density of b1
(percentage points)
0
10
20
30
40
.6
.7
.8
.9
1
1.1
1.2
OLS estimates b1 of beta1


Sampling Distribution of OLS Estimator b1
MODEL 3: N = 60; Var(u|X) = Sigma-Sq = 25,600
sampling density of b1
(percentage points)
0
5
10
15
20
.6
.7
.8
.9
1
1.1
1.2
OLS estimates b1 of beta1


Sampling Distribution of OLS Estimator b1
MODEL 4: N = 60; Var(u|X) = Sigma-Sq = 160,000
sampling density of b1
(percentage points)
0
2
4
6
8
.6
.7
.8
.9
1
1.1
1.2
OLS estimates b1 of beta1

Effect on Sampling Distributions of $\hat{\beta}_0$ and $\hat{\beta}_1$ of Increasing Sample Size N

<u>Question</u>: What is the effect on the sampling distributions of $\hat{\beta}_0$ and $\hat{\beta}_1$ of increasing the sample size N while holding constant both the error variance $\sigma^2 = \mathrm{Var}(u_i \mid X_i)$ and the sample variation of X = $\mathrm{TSS}_X \equiv \sum_{i=1}^{N} x_i^2 \equiv \sum_{i=1}^{N} (X_i - \overline{X})^2$?

<u>Answer</u>: ***Increasing* the sample size N *decreases*** the ***variances* of the sampling distributions of $\hat{\beta}_0$ and $\hat{\beta}_1$**, i.e., decreases $\mathrm{Var}\left(\hat{\beta}_0\right)$ and $\mathrm{Var}\left(\hat{\beta}_1\right)$.

Illustration for Model 3 for which σ^2 = 25,600, σ = 160:

- **For N = 30:**

. summarize

Variable	Obs	Mean	**Std. Dev.**	**Min**	**Max**
b0	1000	**72.88918**	**59.59191**	**-111.8891**	**238.7602**
b1	1000	**.8990184**	**.0335685**	**.7947479**	**1.016194**

- **For N = 60:**

. summarize

Variable	Obs	Mean	**Std. Dev.**	**Min**	**Max**
b0	1000	**71.63642**	**42.66631**	**-73.02339**	**227.0164**
b1	1000	**.8992979**	**.0242138**	**.8291926**	**.976551**

- **For N = 120:**

. summarize

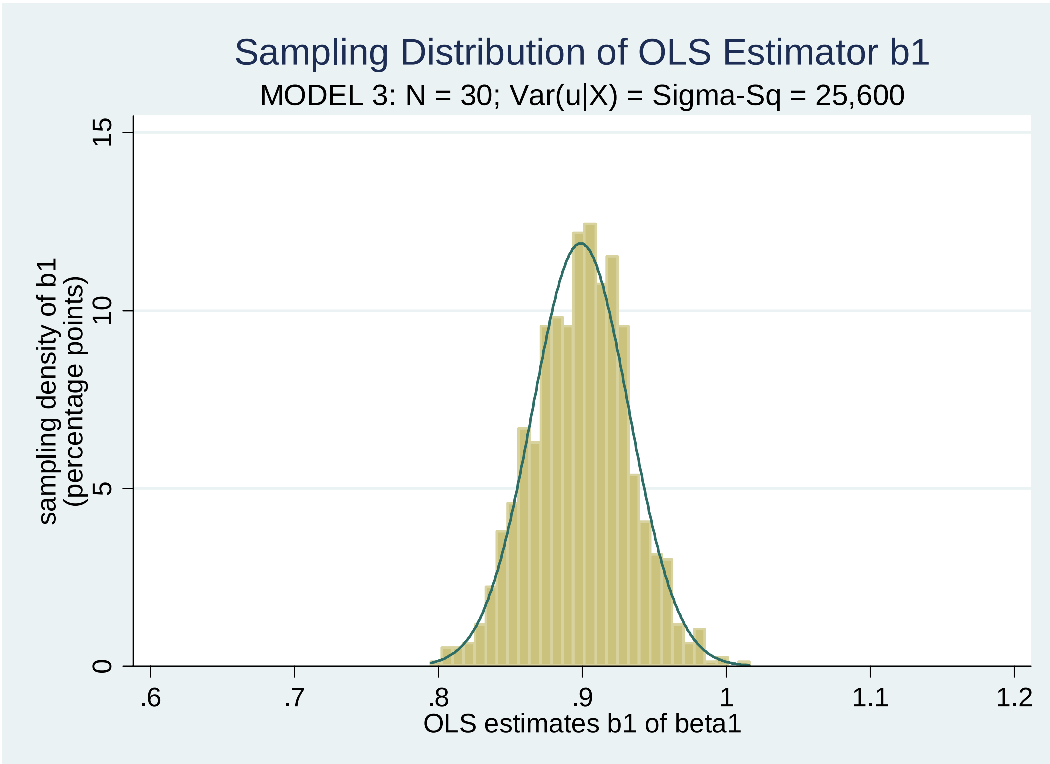
Variable	Obs	Mean	**Std. Dev.**	**Min**	**Max**
b0	1000	**70.05768**	**30.11138**	**-19.01881**	**181.3933**
b1	1000	**.9001622**	**.0169268**	**.843717**	**.95753**

- **For N = 300:**

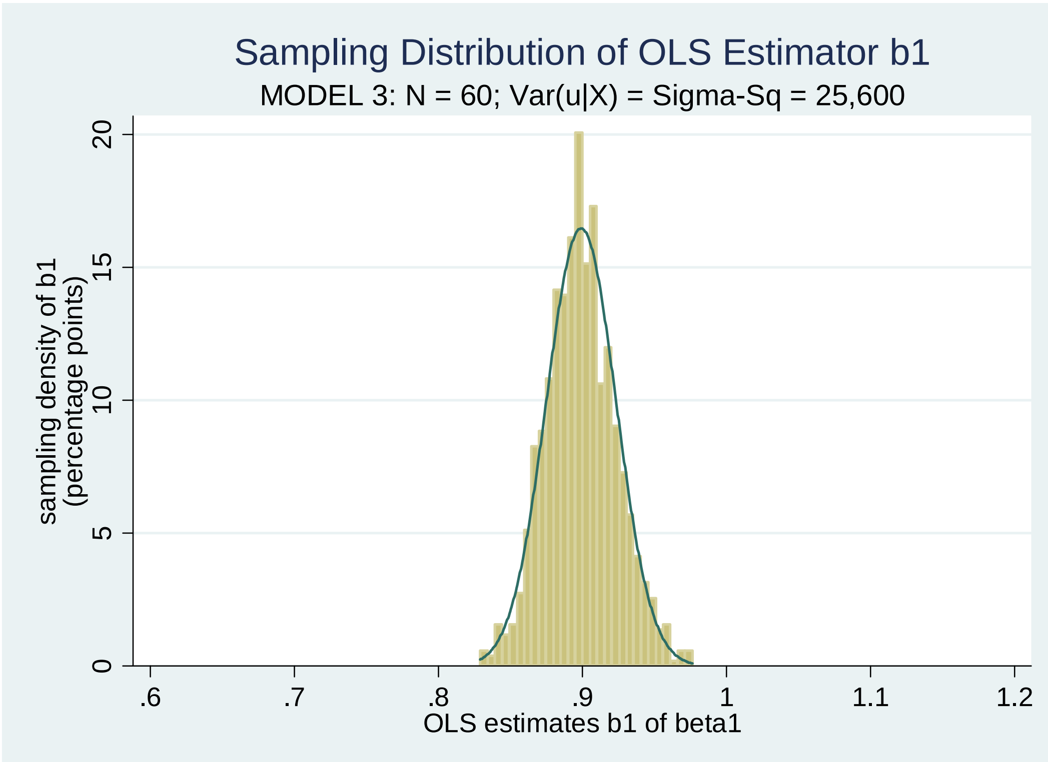
. summarize

Variable	Obs	Mean	Std. Dev.	Min	Max
b0	1000	**70.23207**	**18.79112**	**5.511793**	**127.9536**
b1	1000	**.8998787**	**.0103688**	**.8635329**	**.9360417**

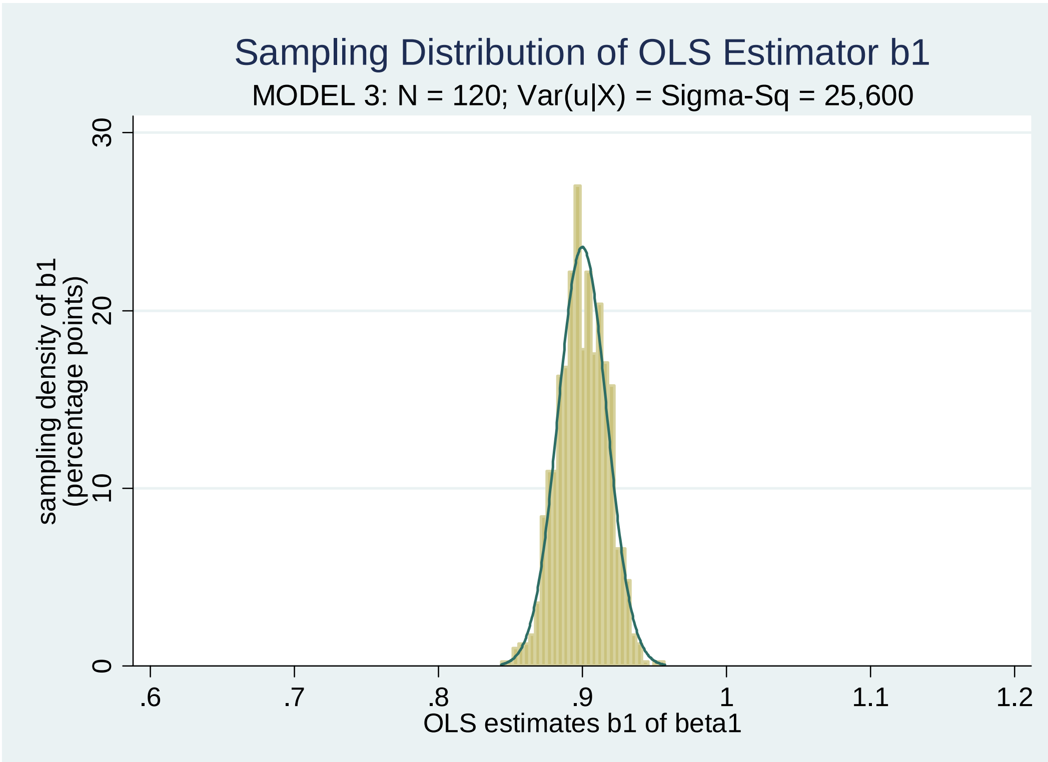
- **Model 3 for N = 30:**



- **Model 3 for N = 60:**



- **Model 3 for N = 120:**



- **Model 3 for N = 300:**