Exercise – Understanding the asymptotic normality of the OLS estimator

1. Aim of the exercise

The aim of this exercise is to examine how the sampling distribution of the OLS estimator becomes approximately normal as the sample size grows, despite the regression errors being non-normal.

2. Theory

The asymptotic normality of the OLS estimator describes how the estimator behaves when the sample size becomes large. While consistency ensures that the estimator converges to the true parameter value, asymptotic normality characterizes the distribution of the estimation error in large samples. In particular, after appropriate scaling by $\sqrt{n}$, the OLS estimator behaves like a normally distributed random variable, even when the regression errors themselves are not normal. This result follows from the Central Limit Theorem and provides the theoretical foundation for large-sample inference. Below we derive the asymptotic distribution of the OLS estimator.

Consider the simple linear regression model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

and assume that $\{(x_i, \varepsilon_i)\}$ is i.i.d. with

$$\mathbb{E}[\varepsilon_i \mid x_i] = 0,$$

and

$$\mathbb{E}[x_i^2] > 0.$$

Also assume that

$$\mathbb{E}\left[\varepsilon_i^2 \mid x_{i1}\right] = \sigma^2$$

due to homoskedasticity. The OLS estimator for β_1 can be written as

$$\hat{\beta}_1 - \beta_1 = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} x_i \varepsilon_i\right).$$

This decomposition is the key to the asymptotic distribution. It also continues to hold in a correctly specified multiple-regression model. In that case, the same formula appears with x_i replaced by the part of x_{1i} that is orthogonal to the other regressors. This follows from the Frisch–Waugh–Lovell theorem, which states that the coefficient on x_{1i} in a multiple regression is the same as the coefficient obtained by regressing the residuals of y_i (after removing the other regressors) on the residuals of x_{1i} (after removing the other regressors). Therefore, focusing on the one-regressor case involves no loss of generality for the asymptotic argument.

In this derivation we do not include a constant. Omitting it does not affect the asymptotic distribution of the slope estimator. With a constant, the exact OLS formula uses deviations from sample means, such as $(x_i - \bar{x})$, but these deviation terms vanish asymptotically at rate $\frac{1}{\sqrt{n}}$. Therefore, the asymptotic distribution of the slope estimator is identical whether or not a constant is included in the regression.

To study asymptotic normality, multiply both sides by $\sqrt{n}$:

$$\sqrt{n}\left(\hat{\beta}_1 - \beta_1\right) = \left(\frac{1}{n}\sum_{i=1}^{n} x_{i1}^2\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_{i1}\varepsilon_i\right).$$

We analyze the two factors on the right-hand side separately. Assume $\{x_{i1}\}$ is i.i.d. with

$$\mathbb{E}[x_{i1}^2] < \infty$$

and

$$\mathbb{E}[x_{i1}^2] > 0.$$

By the Weak Law of Large Numbers,

$$\frac{1}{n}\sum_{i=1}^{n} x_{i1}^2 \xrightarrow{p} \mathbb{E}[x_{i1}^2].$$

Since $\mathbb{E}[x_{i1}^2] > 0$, the inverse exists, and by the continuous mapping theorem,

$$\left(\frac{1}{n}\sum_{i=1}^{n} x_{i1}^2\right)^{-1} \xrightarrow{p} \left(\mathbb{E}[x_{i1}^2]\right)^{-1}.$$

Assume that

$$\mathbb{E}[x_{i1}^2\varepsilon_i^2] < \infty.$$

By the law of iterated expectations,

$$\mathbb{E}\left[x_{i1}\varepsilon_i\right] = \mathbb{E}\left[\mathbb{E}\left[x_{i1}\varepsilon_i \mid x_{i1}\right]\right] = \mathbb{E}\left[x_{i1}\mathbb{E}\left[\varepsilon_i \mid x_{i1}\right]\right] = 0,$$

and

$$\mathbb{E}\left[x_{i1}^2\varepsilon_i^2\right] = \mathbb{E}\left[x_{i1}^2\mathbb{E}\left[\varepsilon_i^2 \mid x_{i1}\right]\right] = \sigma^2\mathbb{E}\left[x_{i1}^2\right].$$

By the Central Limit Theorem,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_{i1}\varepsilon_i \xrightarrow{d} N\left[0, \sigma^2\mathbb{E}\left[x_{i1}^2\right]\right].$$

By Slutsky's theorem,

$$\sqrt{n}\left(\hat{\beta}_1 - \beta_1\right) \xrightarrow{d} N\left[0,\ \left(\mathbb{E}\left[x_{i1}^2\right]\right)^{-2}\cdot\sigma^2\mathbb{E}\left[x_{i1}^2\right]\right] = N\left[0,\ \frac{\sigma^2}{\mathbb{E}\left[x_{i1}^2\right]}\right].$$

That is, the OLS estimator for β_1 is asymptotically normal:

$$\sqrt{n}\left(\hat{\beta}_1 - \beta_1\right) \xrightarrow{d} N\left(0,\ \frac{\sigma^2}{\mathbb{E}\left[x_{i1}^2\right]}\right),$$

or equivalently,

$$\hat{\beta}_1 \approx N\left[\beta_1,\ \frac{1}{n}\cdot\frac{\sigma^2}{\mathbb{E}\left[x_{i1}^2\right]}\right]$$

for large n. This shows that, even when the regression errors are non-normal, the sampling distribution of the OLS estimator becomes approximately normal as the sample size grows, by the Central Limit Theorem.

3. Clear the workspace

Clear the workspace.

```
%% 3. Clear the workspace

% Clear previous results
clear;
```

4. Simulation setup

Here we set up the simulation environment that we will use to study the asymptotic sampling distribution of the OLS estimator. We begin by choosing a sequence of sample sizes, since asymptotic normality concerns how the estimator behaves as the sample size becomes large. We then specify how many Monte Carlo replications to run for each sample size. To highlight that asymptotic normality does not depend on normally distributed errors, we generate the regression errors from a t-distribution with a small number of degrees of freedom. A small degrees-of-freedom value produces heavy-tailed errors, meaning the error distribution has much more probability mass far from zero than a normal distribution. This makes the finite-sample behavior of the OLS estimator more irregular, so any convergence toward a normal distribution becomes easier to see in the simulation. Finally, we define true parameter values for the regression model.

```
%% 4. Simulation setup

% 4.1. Define a sequence of sample sizes
N_obs_grid = [10 30 1000];

% 4.2. Define the number of simulations
N_sim = 5000;

% 4.3. Define the degrees of freedom for t distribution
t_df = 4;

% 4.4. Define true values for the coefficients
B_true = [0.2 3.5]';
```

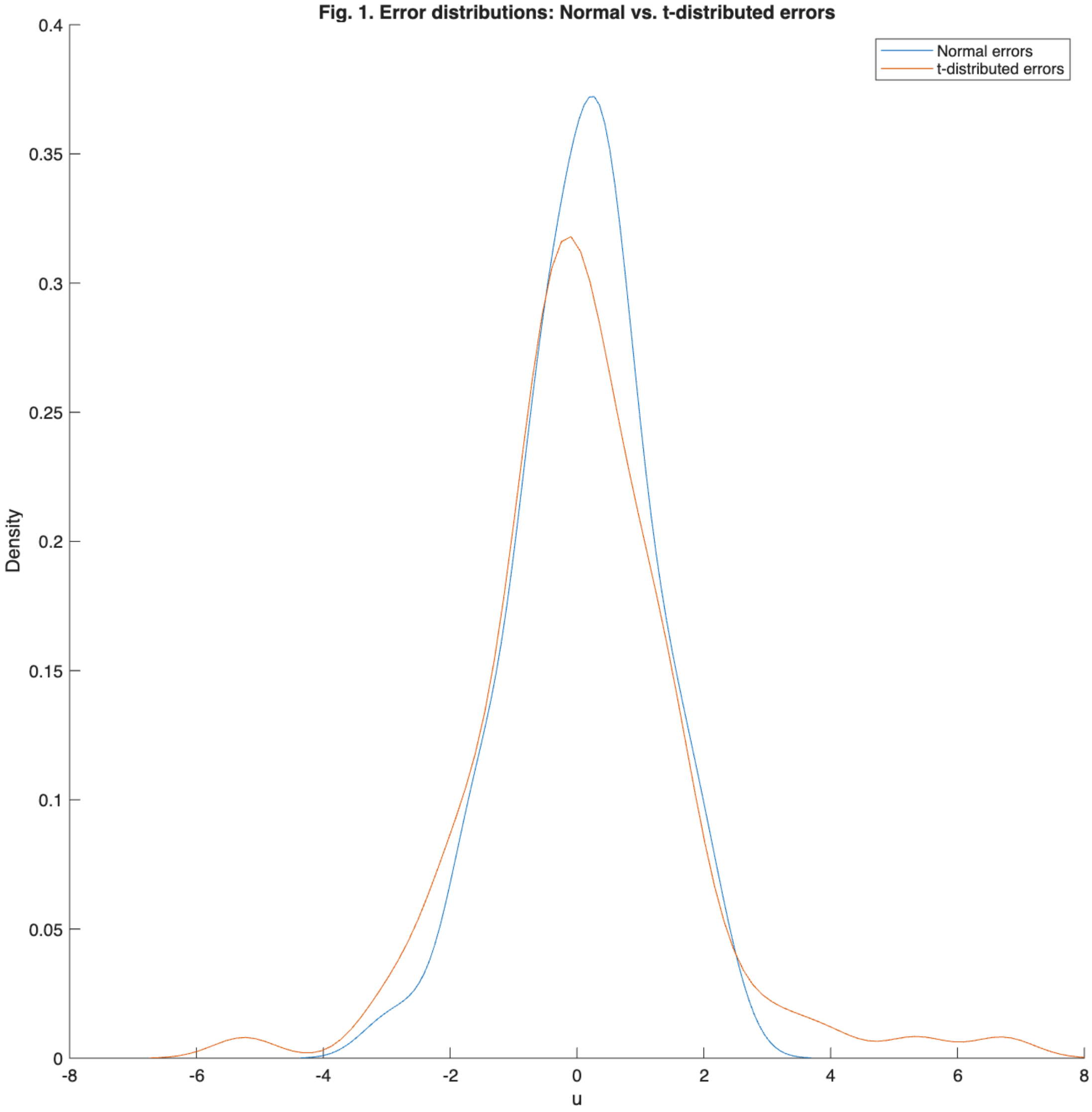
5. Plot the error distributions: Normal vs. t

Here we visualize the difference between normally distributed errors and the heavy-tailed t-distributed errors used in the simulation. We draw a small sample from each distribution and plot their kernel density estimates. This illustrates how the t-distribution places more probability mass in the tails, which motivates examining how such heavy-tailed errors affect the sampling distribution of the OLS estimator in later steps.

```
%% 5. Plot the error distributions: Normal vs. t

% 5.1. Draw errors from the normal distribution
u_normal = random('Normal',0,1,[100 1]);

% 5.2. Draw errors from the t distribution
u_t = random('t',t_df,[100 1]);
```

```
% 5.3. Plot the error distributions
figure
hold on
ksdensity(u_normal)
ksdensity(u_t)
% Theoretical SD of N(0,1)
sd_normal = 1;
% Choose how many standard deviations to show
k = 8;
% Symmetric x-limits based on the theoretical SD
xlim([-k*sd_normal,k*sd_normal])
xlabel('u')
ylabel('Density')
legend('Normal errors','t-distributed errors')
title('Fig. 1. Error distributions: Normal vs. t-distributed errors')
hold off
```

Fig. 1. Error distributions: Normal vs. t-distributed errors



6. Simulate the asymptotic sampling distribution of the OLS estimator

Here we generate the sampling distribution of the OLS estimator for different sample sizes. For each value of the sample size in the grid, we repeatedly draw a sample, obtain an OLS estimate, and record the scaled estimation error.

We work with the estimation error $\hat{\beta}_1 - \beta_1$ because this difference shows how far the estimator is from the true parameter in each sample. Looking at $\hat{\beta}_1$ alone would not isolate this deviation.

We then multiply the estimation error by $\sqrt{n}$ because the unscaled error gets smaller as the sample size increases. Without scaling, its distribution would collapse toward zero. The $\sqrt{n}$ factor keeps the deviations at a comparable scale across different sample sizes, and this scaled error has a well-defined asymptotic normal distribution.

For each sample size, a new set of regressors is generated. The errors follow a heavy-tailed t-distribution. Repeating this procedure many times produces an empirical sampling distribution for each sample size, which we will later compare to the asymptotic normal distribution.

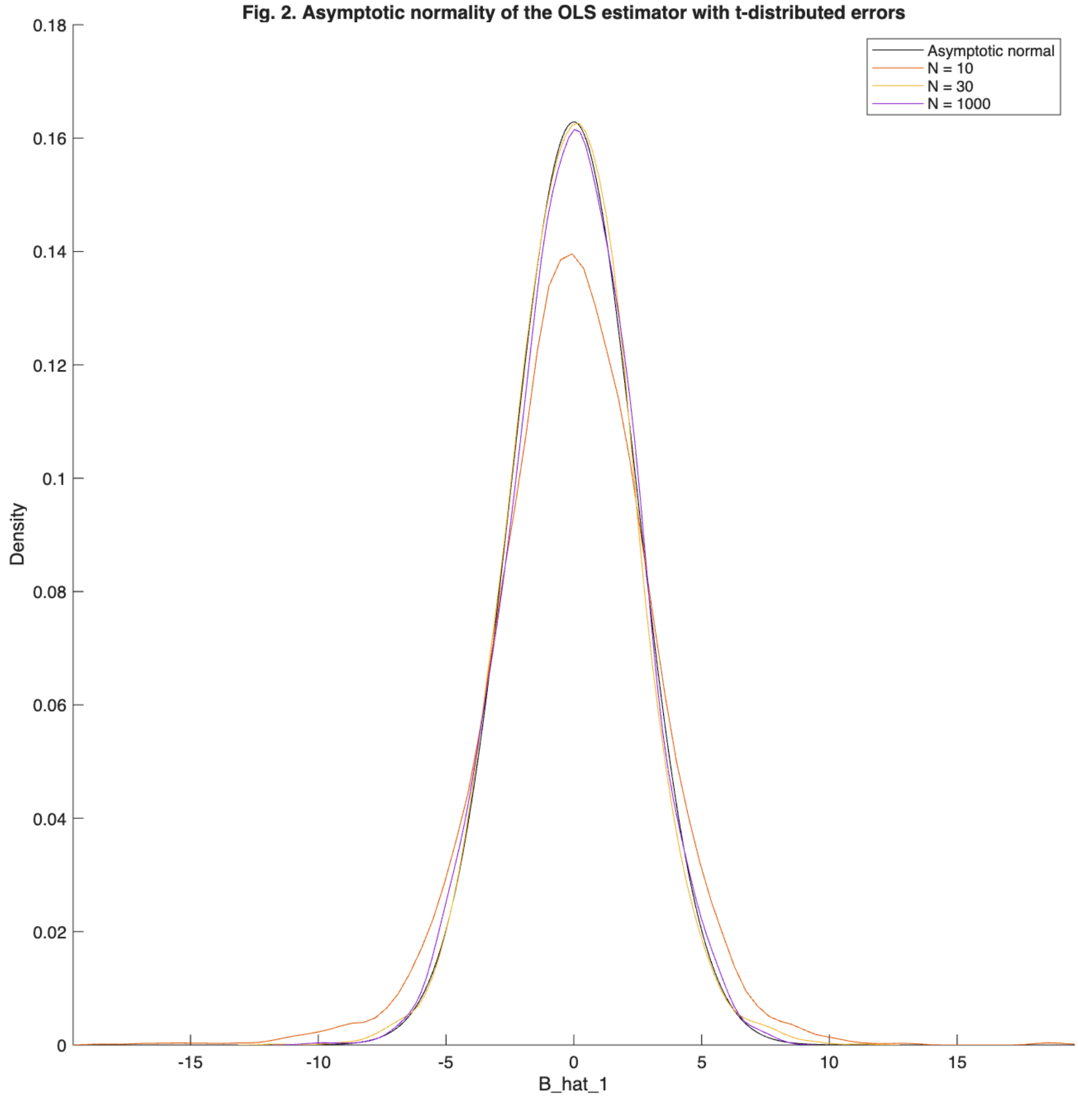
```
%% 6. Simulate the asymptotic sampling distribution of the OLS estimator

% 6.1. Preallocate matrix to store scaled estimation errors
B_hat_by_N = NaN(N_sim,length(N_obs_grid));

% 6.2. Loop over each sample size and simulate a sampling distribution
for j = 1:length(N_obs_grid)
    N_obs_j = N_obs_grid(j);
    x_0 = ones(N_obs_j,1);
    x_1 = random('Uniform',-1,1,[N_obs_j 1]);
    X_j = [x_0 x_1];
    B_hat_temp = NaN(N_sim,1);
    for i = 1:N_sim
        u = random('t',t_df,[N_obs_j 1]);
        y = X_j*B_true+u;
        LSS = exercisefunctionlss(y,X_j);
        % Store the asymptotically scaled coefficient estimation error
        B_hat_temp(i) = sqrt(N_obs_j)*(LSS.B_hat(2,1)-B_true(2));
    end
    % Save results for this sample size
    B_hat_by_N(:,j) = B_hat_temp;
end
```

7. Overlay convergence to normality

Here we compare the sampling distributions obtained in the simulation with the theoretical asymptotic normal distribution. We first compute the population variances of the t-distributed errors and the uniform regressor, which together determine the asymptotic standard deviation of the scaled OLS estimator. Using this value, we construct the corresponding normal density. We then overlay kernel density estimates of the simulated scaled estimation errors for each sample size. This allows us to see how the sampling distribution approaches its asymptotic normal shape as the sample size increases. Even at a moderate sample size such as 30, the simulated distribution already resembles the asymptotic normal curve quite closely.

```
%% 7. Overlay convergence to normality

% 7.1. Compute population variance of the t-distributed errors
var_u = t_df/(t_df-2);

% 7.2. Compute population variance of the uniform regressor
var_x_1 = 1/3;

% 7.3. Compute asymptotic standard deviation of the scaled OLS estimator
sd_asymptotic = sqrt(var_u/var_x_1);

% 7.4. Create grid of x-values for plotting the theoretical density
x_grid = linspace(min(B_hat_by_N(:)),max(B_hat_by_N(:)),400);

% 7.5. Construct theoretical asymptotic normal density
pdf_normal = pdf('Normal',x_grid,0,sd_asymptotic);

% 7.6. Define bandwidth that is fixed for all N
Bandwidth = 0.5;

% 7.7. Figure
figure
hold on
plot(x_grid,pdf_normal,'Color',[0 0 0]);
for j = 1:length(N_obs_grid)
    [f,xi] = ksdensity(B_hat_by_N(:,j),'Bandwidth',Bandwidth);
    plot(xi,f)
end
% Choose how many standard deviations to show
k = 8;
% Symmetric x-limits based on the theoretical asymptotic SD
xlim([-k*sd_asymptotic,k*sd_asymptotic])
xlabel('B\_hat\_1')
ylabel('Density')
legend(["Asymptotic normal",cellstr("N = " + string(N_obs_grid))])
title(['Fig. 2. Asymptotic normality of the OLS estimator with ' ...
      't-distributed errors'])
hold off
```

Fig. 2. Asymptotic normality of the OLS estimator with t-distributed errors



8. Final notes