Assignment 3

Cherryl Chico, Dmitrii Kuptsov, Olta Recica

23 October 2025

1

1.a

(i) The data are generated according to

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \varepsilon_i,$$

where $\varepsilon_i \sim N(0, \sigma^2)$ and $n = 35$.

The true parameter values in the DGP are $\beta_1 = 10$ and $\beta_2 = \beta_3 = \beta_4 = 0.5$. The model is estimated by Ordinary Least Squares (OLS):

$$\hat{\beta} = (X'X)^{-1}X'y,$$

where X is the $n \times K$ matrix of observations on the constant and regressors, and y is the $n \times 1$ vector of dependent-variable observations. Under the Gauss–Markov assumptions and normal disturbances, the statistic

$$t_k = \frac{\hat{\beta}_k - \beta_k}{se(\hat{\beta}_k)} \sim t(n - K),$$

follows a Student-t distribution in finite samples, where

$$se(\hat{\beta}_k) = \sqrt{\hat{\sigma}^2[(X'X)^{-1}]_{kk}}, \qquad \hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n - K}.$$

Hence, the 95% confidence interval for each parameter is

$$\hat{\beta}_k \pm t_{0.975}(n - K)\, se(\hat{\beta}_k).$$

The estimation was carried out in `R` using:

```
model <- lm(y ~ x2 + x3 + x4, data = data33)
summary(model)
confint(model, level = 0.95)
```

Table 1: OLS estimates and 95% confidence intervals.

Variable	Estimate	Std. Error	t value	Pr(>|t|)	95% CI
(Intercept)	7.9365	1.6427	4.832	0.00003	[4.586 , 11.287]
x_2	0.5953	0.0747	7.970	0.00000	[0.443 , 0.748]
x_3	0.2996	0.5264	0.569	0.573	[-0.774 , 1.373]
x_4	0.7819	0.5278	1.481	0.149	[-0.294 , 1.858]

Residual standard error: 3.508 (df = 31)

$R^2 = 0.8856$, Adjusted $R^2 = 0.8745$, F-statistic = 79.95, p-value = 1.10×10^{-14}

The estimate of $\hat{\beta}_2 = 0.595$ is positive and statistically significant, consistent with the true value $\beta_2 = 0.5$. Coefficients $\hat{\beta}_3$ and $\hat{\beta}_4$ have the expected positive signs but are imprecisely estimated, with wide confidence intervals that include zero. This is due to the high correlation between x_3 and x_4 ($\text{corr}(x_3, x_4) \approx 0.9$), which increases the sampling variance of their estimators:

$$\text{var}(\hat{\beta}_k \mid X) = \sigma^2[(X'X)^{-1}]_{kk}.$$

The model exhibits a high $R^2 = 0.8856$, indicating a good overall fit, although inference on x_3 and x_4 is unreliable because of collinearity.

(ii) Under the classical linear model assumptions, the OLS estimator satisfies

$$\hat{\beta} \mid X \sim N\big(\beta,\, \sigma^2 (X'X)^{-1}\big).$$

For the parameters β_3 and β_4, the joint 95% confidence region is defined as the set of all points (β_3, β_4) that satisfy

$$(\hat{\beta} - \beta)' \Big[\hat{\sigma}^2 (X'X)^{-1}\Big]^{-1} (\hat{\beta} - \beta) \leq \chi^2_2(0.95),$$

where $\chi^2_2(0.95)$ denotes the 95th percentile of the chi-squared distribution with two degrees of freedom. This inequality describes an ellipse centered at $(\hat{\beta}_3, \hat{\beta}_4)$, since $(\hat{\beta}_3, \hat{\beta}_4)$ are jointly normally distributed.

The 95% confidence region was obtained in `R` using the `confidenceEllipse()` function from the `car` package:

```
library(car)
confidenceEllipse(model, c("x3", "x4"), levels = 0.95, lwd = 2)
abline(v = confint(model)["x3", ], lty = 2)
abline(h = confint(model)["x4", ], lty = 2)
points(coef(model)["x3"], coef(model)["x4"], pch = 19)
title("95% Confidence Ellipse for (β3, β4)")
```

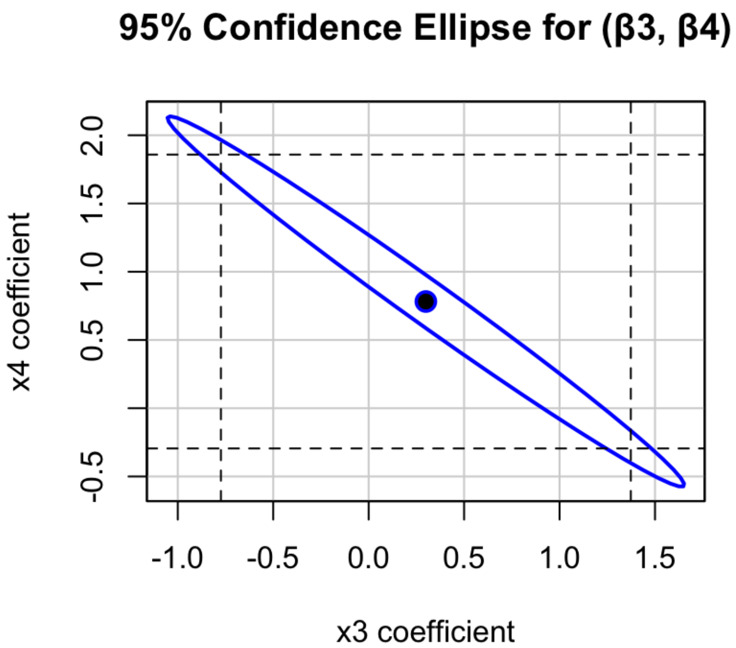


Figure 1: 95% joint confidence region for (β_3 and β_4) (blue ellipse) with marginal confidence intervals (dashed lines).

(iii) The ellipse in Figure 1 represents the joint 95% confidence region for (β_3, β_4). The dashed vertical and horizontal lines correspond to the marginal 95% confidence intervals for each parameter, and the black point marks the OLS estimates $(\hat{\beta}_3, \hat{\beta}_4)$. The ellipse is elongated and negatively sloped, reflecting the negative correlation between $\hat{\beta}_3$ and $\hat{\beta}_4$, caused by the strong collinearity between x_3 and x_4.

Since the origin $(0, 0)$ does not lie within the ellipse, the joint null hypothesis

$$H_0 : \beta_3 = \beta_4 = 0$$

can be rejected at the 5% significance level. In comparison, we cannot reject the hypothesis that $\beta_3 = 0$ or $\beta_4 = 0$ separately.

(iv) The figure illustrates the difference between testing the significance of regressors separately and jointly.

Separate tests: Testing each coefficient individually (e.g., $H_0 : \beta_3 = 0$ and $H_0 : \beta_4 = 0$) relies on the marginal confidence intervals for β_3 and β_4, represented by the vertical and horizontal dashed lines in the figure. Each test ignores the covariance between the two estimates. Since both intervals include zero, we cannot reject either of the null hypothesis.

Joint test: The ellipse represents the set of (β_3, β_4) values that are jointly consistent with the data at the 5% significance level. A joint test takes into account the correlation between the estimates. If the origin $(0, 0)$ lies inside the ellipse, the joint null hypothesis

$$H_0 : \beta_3 = \beta_4 = 0$$

cannot be rejected at the 5% level. If the origin were outside the ellipse, the null would be rejected, indicating that at least one coefficient differs from zero when considered jointly.

The difference between separate and joint tests arises because $\hat{\beta}_3$ and $\hat{\beta}_4$ are correlated when x_3 and x_4 are collinear. Collinearity inflates their standard errors, making individual t-tests unreliable. Each regressor may appear insignificant separately, even though they are jointly statistically relevant. In comparison, a joined test will produce lower CIs (if there are no other strong collinearities in the model), and therefore more likely to result to statistical significant (joint) regressors. The joint confidence region therefore provides a more accurate assessment of significance under collinearity by accounting for the covariance between regressors.

1.b

The regression was re-estimated using a sample of $n = 3500$ observations generated from the same DGP:

$$y_i = 10 + 0.5x_{i2} + 0.5x_{i3} + 0.5x_{i4} + \varepsilon_i, \qquad \varepsilon_i \sim N(0, 4^2).$$

The estimated coefficients are $\hat{\beta}_2 = 0.490$, $\hat{\beta}_3 = 0.551$, and $\hat{\beta}_4 = 0.448$, which remain very close to the true values $\beta_2 = \beta_3 = \beta_4 = 0.5$.

(i) This similarity is expected, as the OLS estimator is unbiased and consistent:

$$E[\hat{\beta}] = \beta, \qquad \hat{\beta} \xrightarrow{p} \beta \text{ as } n \to \infty.$$

The variance of the estimator,

$$\mathrm{Var}(\hat{\beta} \mid X) = \sigma^2 (X'X)^{-1},$$

decreases with n since $X'X$ increases proportionally to the sample size. Therefore, larger n reduces the sampling variability of $\hat{\beta}$. The considerably smaller standard errors ($se(\hat{\beta}_2) = 0.0078$, $se(\hat{\beta}_3) = 0.0685$, and $se(\hat{\beta}_4) = 0.0680$) shows that the collinearity problem for testing decreased, but have not been totally solved.

R output:

```
Call:
lm(formula = y ~ x2 + x3 + x4, data = data33)

Residuals:
     Min       1Q   Median       3Q      Max
-13.2997  -2.5950  -0.0316   2.5728  14.3972

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.229438   0.181098  56.486  < 2e-16 ***
x2           0.490229   0.007763  63.151  < 2e-16 ***
```

```
x3           0.551196    0.068484    8.048 1.14e-15 ***
x4           0.447534    0.067977    6.584 5.28e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.952 on 3496 degrees of freedom
Multiple R-squared: 0.8519,   Adjusted R-squared: 0.8517
F-statistic: 6702 on 3 and 3496 DF,  p-value: < 2.2e-16
```

(ii) The 95% confidence intervals were recalculated for the larger sample. They are substantially narrower compared with those obtained when $n = 35$. For example, the interval for β_2 changes from [0.443, 0.748] to approximately [0.486, 0.514].

This occurs because the variance of $\hat{\beta}_k$ depends inversely on sample size:

$$\text{Var}(\hat{\beta}_k \mid X) = \sigma^2[(X'X)^{-1}]_{kk} \propto \frac{1}{n}.$$

Hence, the standard errors decrease at rate $1/\sqrt{n}$, and the corresponding confidence intervals become tighter around the true parameter values.

This illustrates the **consistency** of OLS: as n increases, $\hat{\beta} \xrightarrow{p} \beta$, and the confidence intervals converge to the true β's. The narrowing of the intervals reflects reduced sampling uncertainty and greater estimator precision.

(iii) The 95% confidence region for (β_3, β_4) becomes smaller when the sample size increases to $n = 3500$. The shape of the region (an ellipse) remains the same, but its area decreases as the estimators become more precise.

The covariance matrix of the OLS estimator is

$$\widehat{\text{Var}}(\hat{\beta}) = \hat{\sigma}^2(X'X)^{-1}.$$

Since $X'X$ grows proportionally with n, $(X'X)^{-1}$ decreases in magnitude, leading to smaller variances for $\hat{\beta}_3$ and $\hat{\beta}_4$. Consequently, the confidence ellipse contracts around the true parameter values.

Formally, the 95% joint confidence region satisfies

$$(\hat{\beta} - \beta)' \big[\widehat{\text{Var}}(\hat{\beta})\big]^{-1} (\hat{\beta} - \beta) \leq c_{0.95}.$$

As $\widehat{\text{Var}}(\hat{\beta})$ decreases with n, the ellipse becomes smaller. This visual reduction in the region reflects higher precision and the asymptotic properties of the OLS estimator.

(iv) The reduction in the variance of $\hat{\beta}_3$ (and $\hat{\beta}_4$) when n increases can be explained using the variance expression of the OLS estimator:

$$\text{Var}(\hat{\beta}_j \mid X) = \sigma^2 \big[(X'X)^{-1}\big]_{jj}.$$

Because $X'X = \sum_{i=1}^n x_i x_i'$ scales with n, its inverse $(X'X)^{-1}$ decreases approximately at rate $1/n$. Therefore, for a given σ^2,

$$\text{Var}(\hat{\beta}_j \mid X) \propto \frac{\sigma^2}{n}.$$

This implies that as n increases, the sampling variability of $\hat{\beta}_3$ and $\hat{\beta}_4$ decreases, producing smaller standard errors and tighter confidence regions.

Intuitively, a larger sample provides more information about the relationship between y and the regressors, which improves the precision of the estimated slope coefficients. Hence, the observed reduction in the size of the confidence region directly follows from the $1/n$ relationship in the variance expression.

2

Given the following regression model:

$$RPmsoft_t = \beta_1 + \beta_2 RPsandp_t + \beta_3 Dprod_t + \beta_4 Dinflation_t + \beta_5 Dterm_t + \beta_6 m1_t + \epsilon_t$$

where

- $RPmsoft_t$ is the excess return of the Microsoft stock,
- $RPsandp_t$ is the risk premium of the S&P 500 index,
- $Dprod_t$ is the change in production,
- $Dinflation_t$ is the change in inflation,
- $Dterm_t$ is the change in term structure,
- $m1_t$ is the money supply growth,
- ϵ_t is the error term.

2.a

Using the data microsoft.csv, the resulting regression model is as follows:

$$\begin{aligned}\widehat{RPmsoft_t} = &-0.9291 + 1.3232 RPsandp_t - 1.5216 Dprod_t \\ &+0.4716 Dinflation_t + 4.1588 Dterm_t + 5.4352 m1_t\end{aligned}$$

```
. reg rpmsoft rpsandp dprod dinflation dterm m1

      Source |       SS           df       MS      Number of obs   =       324
-------------+----------------------------------   F(5, 318)       =     17.26
       Model |  13628.1699         5  2725.63398   Prob > F        =    0.0000
    Residual |  50211.9204       318  157.899121   R-squared       =    0.2135
-------------+----------------------------------   Adj R-squared   =    0.2011
       Total |  63840.0903       323  197.647338   Root MSE        =    12.566

------------------------------------------------------------------------------
     rpmsoft | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     rpsandp |   1.323168   .1524728     8.68   0.000     1.023185    1.623151
       dprod |  -1.521569   1.283074    -1.19   0.237    -4.045955    1.002817
  dinflation |   .4715841   2.351165     0.20   0.841     -4.15422    5.097388
       dterm |   4.158737   2.487168     1.67   0.095    -.7346462     9.05212
          m1 |   5.435169   2.869448     1.89   0.059    -.2103327    11.08067
       _cons |  -.9290695   .7598421    -1.22   0.222    -2.424022    .5658834
------------------------------------------------------------------------------
```

2.b

To test the *January effect* which is that on avereage, every else equal, the returns (or excess returns) are larger in the month of January than the rest of the months, we set-up the following hypothesis test:

$$H_0 : \beta_6 = 0 \quad \text{(No January effect)} \quad vs \quad H_a : \beta_6 > 0 \quad \text{(January effect exists and is positive)}$$

where β_6 is the estimated coefficient for the regressor $m1$ which is a 1 for January and 0 otherwise.

From there OLS regression results, we have $t - value = 1.89$. At $\alpha = 1\%$ and degrees of freedom $n - K = 324 - 6 = 298$, the $t - critical_{0.01}(298) = 2.339$. Because $t - value < t - critical$, we fail to reject the null hypothesis. That is, the data does not provide evidence to reject the claim that there is no January effect in the excess returns of Microsoft stock.

2.c

The starting point for use of the $t-test$ statistic is the (conditional) sampling distribution of the $\hat{\beta}_k$ which is derived from the classifical assumptions plus normality. Thus, the assumptions other than normality are:

(A1) Linearity: The regression model has been correctly specified such that $y = X\beta + \epsilon$

(A2) Strict exogeneity: The error term has an expected value of zero given any values of the regressors in all time periods, i.e. $E(\epsilon_i|X) = 0$ for all i.

(A3) Homoskedasticity: The variance of the error term is constant across all levels of the regressors, i.e. $Var(\epsilon_i|X) = \sigma^2$ for all i.

(A4) Disturbances are uncorrelated: The error terms are uncorrelated across observations, i.e. $Cov(\epsilon_i, \epsilon_j|X) = 0$ for all $i \neq j$.

In addition, to derive the distribution of the $t - test$ statistic, we used

$$\frac{\hat{\beta}_k - r}{\sqrt{\sigma^2 (X'X)^{-1}_{kk}}} \underset{\text{under } H_0}{\sim} N(0,1)$$

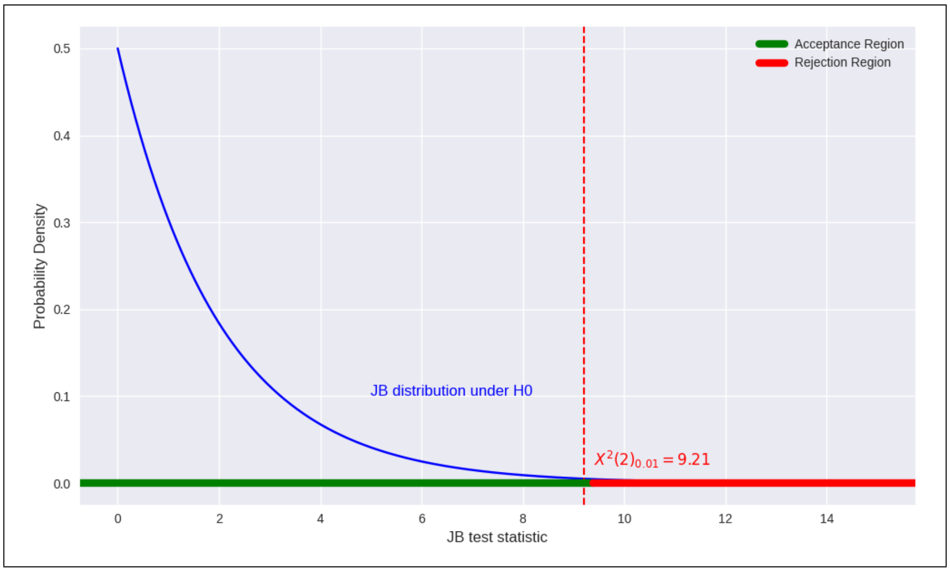
where r is the value of β_k under the null hypothesis. Thus, the distribution of the $t - test$ statistic is derived under the assumption that H_0 is true.

2.d

The $Jarque-Bera$ (JB) test can be used to check for normality of the distrubances. The JB-test statistic is given by:

$$JB = \frac{n}{6}\left(sk^2 + \frac{(kur-3)^2}{4}\right) \underset{\text{under } H_0}{\overset{a}{\sim}} \chi^2(2)$$

where sk is the sample coefficient of the skewness of the variable, kur is its sample coefficient of kurtosis, and n is the sample size. Using significance of level of $\alpha = 0.01$ and the critical value $X^2_{0.01}(2) = 9.21$. The acceptance and rejection regions are in the illustration below:



A normal distribution has skewness of 0 and kurtosis of 3. Thus, a deviation from these values can indicate a departure from normality. The corresponding hypothesis test is set-up as follows:

$$H_0 : SK = 0 \text{ and } KUR = 3 \quad vs \quad H_A : \text{ not } H_0$$

where SK and KUR are the population coefficients of skewness and kurtosis for the disturbances, respectively.

Under normality, the mean and variance of skewness is 0 and $\frac{6}{n}$ and 3 and $\frac{24}{n}$ for kurtosis, respectively. Thus, we can see the $JB-test$ statistic is composed of mahalonobis distances for SK and KUR. If the disturbances are normally distributed, then the $JB-test$ statistic should be small, otherwise it will have a large positive value. Therefore, the acceptance region is from 0 up to the critical value determined by the significance level.

2.e

If the assumptions regarding the dgp for consistency of OLS estimator are met, then $\hat{\beta} \xrightarrow{p} \beta$. Note that $\hat{\epsilon} = y - \hat{y} = X\beta + \epsilon - X\hat{\beta} = X(\beta - \hat{\beta}) + \epsilon$, by *Continuous Mapping Theorem*, we have $\hat{\epsilon} \xrightarrow{p} X(\beta - \beta) + \epsilon = \epsilon$. Thus, $\hat{\epsilon} \xrightarrow{p} \epsilon$.

2.f

Using Stata, the $JB-test$ value is 1809. At $\alpha = 0.01$ and degrees of freedom $= 2$, the critical value is $X^2_{0.01}(2) = 9.21$. Since 1809>9.21, we reject the null hypothesis that the disturbances are normally distributed. Thus, there is evidence to suggest that the disturbances are not normally distributed.

2.g

Testing again the $January\ effect$ using asymptotic t-test, we have critical value $= z_{0.01} = 2.326$. Since the $t-value = 1.89 < 2.326$, we fail to reject the null hypothesis. Thus, the conclusion remains the same as in part (b) that there is no evidence to suggest the existence of January effect in the excess returns of Microsoft stock.

2.h

The use of asymptotic tests is justified in this case for two reasons. Firstly, the normality assumption for exact $t-test$ is not satisfied. Secondly, we have 325 observations which seems to be large enough for asymptotic teststs. As an alternative, we may propose bootstrap, but the difference would probably be minor.

2.i

The statement is correct. The exact $t-test$ relies on the t-distribution which posseses heavier tails than the standard normal distribution used in the asymptotic test. This property has two distinct consequences:

1. For a given significance level α, the cirtical value from the t_{n-1} distribution exceeds its standard normal counterpart $(t_{n-1,1-\alpha/2} > z_{1-\alpha/2})$. This results ina wider acceptance region for the exact test.
2. For any computed test statistic, the $p-value$ derived from the t_{n-1} distribution is larger than that from the asymptotic normal approximation.

Both a wider acceptance region and a larger $p-value$ mean that the exact $t-test$ requires stronger evidence against the null hypothesis to reject it, making its inference more conservative.