ELE 3911 Introduction to Quantitative Finance

– Group Assignment –

Autumn 2023

Exam component weight: 40%

To be answered in groups of 1-2 students

Upload your solution paper to WISEflow. Check the date and time of the deadline in WISEflow. Your solution paper may have up to 15 pages plus an appendix, which has to be uploaded as a separate attachment and does not count towards the 15 pages.

For your solution paper, please use the standard BI template paper and stick to the default font style, font size, line spacing, margins, etc. To save your Word document as PDF file, go to "File," "Save As," click on "Word document (.docx)" to open the drop-down menu, select "PDF (*.PDF)," and hit "Save." Upload your PDF file as the solution paper in WISEflow.*

A separate attachment (as a PDF file) has to be uploaded in WISEflow. This appendix will contain your complete R code with clarifying comments. You can either produce this by using RStudio or copy your code into Word and export it as a PDF file. You do not need to show the output of the code. Upload the resulting PDF file as an attachment to your solution paper in WISEflow.

Read each question carefully and give precise answers. Report numerical answers using ***at least four digits*** *after the decimal place.*

Please note that I cannot assist you with the exam and will only answer basic clarifying questions.

The student code of conduct applies: the solution paper must be written and prepared by the corresponding group members only. ***Collaboration*** *with classmates or other individuals outside the group is* ***not permitted*** *and is considered cheating.*

All papers are automatically subject to ***plagiarism control****.*

All literature used to answer the assignment must be listed as references at the end of the solution paper.

Good luck!

Assignment introduction

In this assignment, you will play the role of an analyst for a fictional peer-to-peer (P2P) lending platform. P2P lending platforms allow investors to fund retail loans directly, rather than indirectly through a financial intermediary. When a borrower applies for a loan, the platform collects information about their credit characteristics and then posts it on the platform. Investors can then select individual loans on the platform and invest in increments as small as $1. While P2P lending gives investors more control, it also requires them to evaluate the creditworthiness of borrowers and to construct their own portfolios of loans, which can be costly.

In an attachment to this assignment, you will find the CSV file `p2ploans.csv`, which contains characteristics of loans posted to the platform.

- `id`: the unique id number for each loan.
- `dti_ratio`: the borrower's debt as a percentage of their income.
- `interest_rate`: interest rate for a loan.
- `internal_rating`: a borrower credit rating produced by the platform.
- `maturity`: loan maturity in years.
- `yearly_payment`: yearly payment to the platform in U.S. dollars.
- `risk_free`: the risk free rate in percent.

  - General guidelines:
    - The appendix should have been submitted as a separate document with clarifying comments.
    - If an answer misinterprets the question or provides a different, but equally valid answer, full or partial credit is given.
    - For any results that depend on random number generation, the numerical answers do not need to match the example answers provided exactly.

Total assignment points **(100 points)**

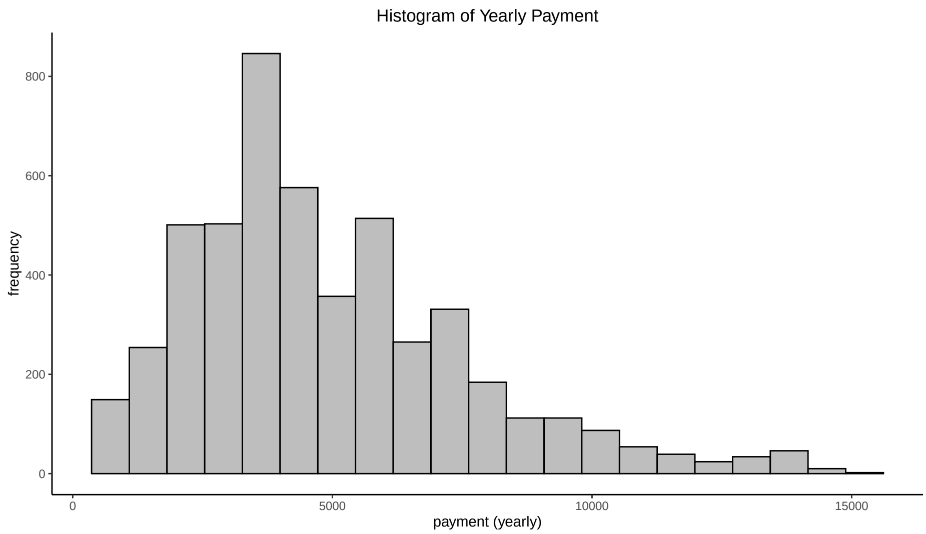
1. **Task 1** **(30 points)**

   Your first task is to explore the properties of loans posted on the platform, focusing on yearly loan payments. These payments are fixed in size over the duration of the loan and are sent directly to the platform, rather than to individual investors.

   (a) Compute and report the mean and median of `yearly_payment`. What can you infer about the distribution of yearly payments from these values?

   - Max: 2 points.
   - Mean $\approx \$4918.38$.
   - Median $\approx \$4322.94$.
   - The mean is larger than the median, which indicates a right-skew.

(b) Use `ggplot` to construct a histogram of the yearly payment with 20 bins. Include an appropriate title, x axis label, and y axis label. Paste the histogram into your solution paper.

- Max: 3 points.
- Full credit for correct histogram with valid formatting and labels.
- Example:
- 

(c) Based on your findings in 1(a) and 1(b), does the distribution of yearly payments appear to be symmetric? Explain your answer.

- Max: 6 points.
- The distribution is right-skewed and not symmetric.
- The mean is greater than the median.
- Discuss the histogram's features.

(d) Compute the skewness and kurtosis of the yearly payment. How do these values compare to the skewness and kurtosis of a normal distribution?

- Max: 4 points.
- Skewness ≈ 1.019249.
- kurtosis ≈ 4.073969.
- Normal distribution has a skewness of 0.
- Normal distribution has a kurtosis of 3.

(e) Based on your findings in 1(a)-(d), would you use a normal distribution, a triangular distribution, or a uniform distribution to model yearly payments? Justify your answer by discussing each distribution's parameters. Compare this to the histogram you plotted in 1(b) and the statistics you calculated.

A correct answer should have discussed the properties of the triangular distribution, the normal distribution, and uniform distributions. The answer should either be the triangular distribution or compelling reasoning for another distribution. See the points below for an example description of the distributions:

- Max: 7 points.

- The triangular distribution is defined by its mean, median, and mode. It has finite bounds, can be skewed, and is unimodal.
- The normal distribution is defined by its mean and variance. It has infinite bounds, zero skew, and is unimodal.
- The uniform distribution is defined by its minimum and maximum. It has finite bounds, zero skew, and no mode.

(f) Identify the loan with the largest yearly payment and report its `id`. Assume that payments are made at the end of each year and that you discount payments at the risk free rate of 1.72%. What is the present value of the first yearly payment made on this loan?

- Max: 4 points.
- Loan ID = 137.
- PV $\approx \$14,766.18$.

(g) Recall that the yearly payment is fixed over the duration of the loan. Compute the present value of yearly payments to the platform over the duration of the loan. Use the risk free rate of 1.72% to discount payments.

- Max: 4 points.
- PV $\approx \$71,375.93$.

2. **Task 2** **(25 points)**

Your manager is concerned with the impact that borrower default could have on aggregate payment flows to the platform. In this task, you will assume that borrowers have the option to default. In such an event, the platform will receive \$0 from the borrower in all remaining periods.

(a) Use the data for the loan with `id` = 5. Assume the probability of default is 0.05 for loans that have not entered default, irrespective of the year of payment. What is the expected value of the first yearly payment?

- Max: 3 points.
- EV $\approx \$1967.18$

(b) Under the same assumptions as task 2(a), what is the expected value of the final yearly payment?

- Max: 3 points.
- EV $\approx \$1775.38$

(c) Is there a difference between the expected values you reported in 2(a) and 2(b)? Explain why or why not.

- Max: 4 points.
- Yes, there is a difference.
- The probability of not having defaulted after three years is lower than the probability of not having defaulted after one year.

(d) What distribution can you use to model the number of defaults in the first period of repayment, given the probability of default? Report that distribution and its parameter values.

- Max: 3 points.
- Binomial distribution
- n = 5000
- p = 0.05
- q = 0.95
  - Partial credit for different distribution with compelling reasons.

(e) Based on your assumption in 2(d), what is the expected number of defaults in the first year? What is the variance of the number of defaults? What is the skewness of the number of defaults?

- Max: 6 points.
- Expectation $= np = 250$
- Variance $= npq = 237.5$
- Skewness $= \frac{(q-p)}{\sqrt{npq}} = 0.0584$

(f) Consider the aggregate yearly payment flows to the platform from the loans with `id` values of 1-10. If the default probability for each loan is 0.05, what is the sum of expected payments at the end of the first year?

- Max: 3 points.
- $EV \approx \$41669.51$

(g) Now assume that default is perfectly correlated across borrowers. With a 0.05 probability, all borrowers default; otherwise, they all repay. From a mean-variance investor perspective, is this situation preferable to the case where defaults are independent? Or would you be indifferent between the two options? Explain your reasoning.

- Max: 3 points.
- The expected payment is the same, but the variance is higher when defaults are correlated. If nothing else changes, independent defaults are preferable from a mean-variance perspective.

3. **Task 3** **(25 points)**

You are next asked to evaluate the platform's system of internal ratings, which assign a risk category to each loan based on borrower and loan characteristics. Your manager would like to determine the extent to which interest rates, which are set in a bidding process among investors, vary with internal ratings and other observable risk characteristics.

(a) Compute the mean of the interest rate for `AA` and `HR` groups. What is the difference between the two in percentage points?

- Max: 5 points.

- $\approx 24.7002 ppt$

(b) What could explain the difference in rates between `AA` and `HR`-rated loans? Discuss variables that are both in and outside of the dataset.

- Max: 5 points.
- Default probabilities, which might be affected by the DTI ratio and loan maturity.
- Partial credit for mostly correct answer.

(c) Using the `lm()` function, regress the interest rate on the internal ratings. You can treat the rating group, which is a categorical variable, as a `factor` in `R`. How well do the internal ratings explain variation in the interest rate?

- Max: 5 points.
- Example regression output:

```
Call:
lm(formula = interest_rate ~ internal_rating, data = df)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.3143 -0.9471  0.0805  0.7605  3.5842 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)        7.97951    0.04755  167.81   <2e-16 ***
internal_ratingAA -2.33243    0.07612  -30.64   <2e-16 ***
internal_ratingB   3.01974    0.06878   43.90   <2e-16 ***
internal_ratingC   7.47501    0.06601  113.25   <2e-16 ***
internal_ratingD  13.42630    0.06842  196.24   <2e-16 ***
internal_ratingE  18.77475    0.07993  234.89   <2e-16 ***
internal_ratingHR 22.36781    0.08610  259.79   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.427 on 4993 degrees of freedom
Multiple R-squared:  0.9675,	Adjusted R-squared:  0.9674 
F-statistic: 2.476e+04 on 6 and 4993 DF,  p-value: < 2.2e-16
```

-
- Other valid regressions specifications were accepted for full credit.
- A correct answer should specify code that regresses `interest_rate` on `internal_rating` in the form of a `factor`. It should also discuss model fit, possibly by referencing adjusted-R^2.

(d) In task 3(b), you proposed variables that might explain variability in interest rates. Include them as regressors in the specification you estimated in task 3(c). Discuss the magnitude and significance of the impact of these variables. Also, discuss the impact they have on the regression's fit.

- Max: 5 points.
- The regression should include the DTI ratio as an additional regressor, possibly along with other variables. There should be clear justification for the inclusion of variables and their transformation. Only DTI is significant, but the magnitude is small.
- Adjusted R^2 may increase slightly, depending on the specification, but the increase is very small.

(e) What do your findings in 3(d) suggest about the internal ratings? Do they already incorporate most of the information contained in the other variables in the dataset?

- Max: 5 points.

- The internal ratings appear to explain most of the variation in interest rates and probably already incorporate information from the other variables we included in the regression.

4. **Task 4** **(20 points)**

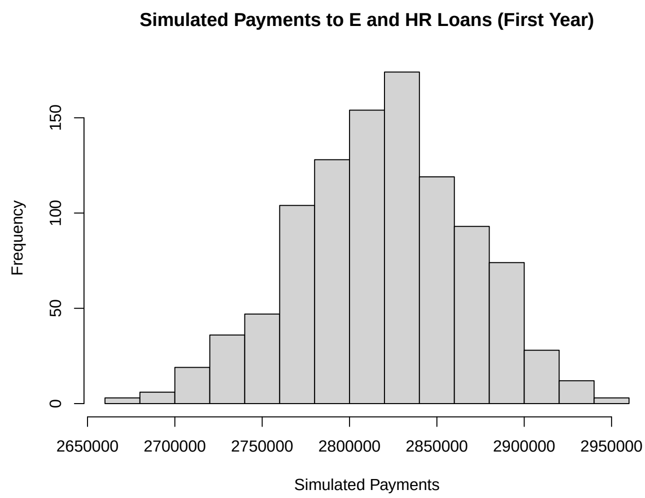
   In earlier tasks, you assumed that the default rate was the same for each loan. Your manager has now provided you with historical default data by internal rating group, which is shown in the table below.

rating	AA	A	B	C	D	E	HR
default probability	0.01	0.02	0.03	0.05	0.08	0.15	0.30

   In this task, you will simulate the sum of the first year of payments to the platform for groups `E` and `HR`. Assume each borrower either makes payment or defaults (pays $0) with the probability given in the table. Also, assume that default is independent across borrowers.

   (a) Simulate the first year of payments 1000 times for all loans in the `E` and `HR` internal rating groups. Construct a histogram with 20 bins and an appropriate title and axis labels. Paste it into your solution paper.

   - Max: 5 points.
   - Full credit for histogram that resembles normal distribution.
   - Example:



   -
   - Full credit if E and HR plotted separately under the interpretation that the payments should not be combined.
   - The numerical values on the axes do not need to align with the example plot, since there will be some random variation.

   (b) What probability distribution does your histogram most closely resemble? Why does it resemble this distribution?

   - Max: 5 points.
   - Normal distribution.
   - Central limit theorem.

(c) Report the mean and standard deviation of the sum of payments over the 1000 simulations.

- Max: 5 points.
- Mean $\approx \$2,820,542$
- Standard deviation ≈ 49009.15
- Since the results are based on a simulation, full credit is given for values that are close to the answers provided.

(d) Compute the 95% VaR for total payments across the 1000 simulations.

- Max: 5 points.
- VaR 95% $\approx \$2,734,096$.
- Since the results are based on a simulation, full credit is given for values that are close to the answers provided.
- Partial credit if VaR interpreted and implemented differently (but correctly).