Please provide your assessment of each of the following four categories, summary and suggested questions for the discussion. The minimum length of the report is 300 words.

The thesis focuses on finding the optimal portfolio for a risk-averse investor and applies the Modern Portfolio Theory by Markowitz on the Czech peer-to-peer lending market using a dataset from the platform Zonky. The author concludes that the optimal portfolio consists mainly of low risk loans with low interest rates. Further, the study implies that the recovery rate is a significant determinant of performance (in terms of expected return and Sharpe ratio) of high risk categories and that the correlation between individual risk categories should not be ignored in a portfolio analysis.

Contribution

The author uses the historical data provided by the platform Zonky to analyse expected return, standard deviation and Sharpe ratio of individual rating categories and he then uses this information for a construction of the optimal portfolio. The analysis is based on the average return on investment calculated separately on repaid loans and defaulted loans and three assumptions of recovery rate, as the information is not available from the dataset.

The study has several important implications, it shows that the expected return of the risk averse investor’s optimal portfolios for all recovery rate scenarios is lower than 4% while Zonky promises average rate of return of 6%, this is most probably driven by low return on high risk loans which results in very low representation of the medium and high risk loans in the optimal portfolios reaching less than 3%, while the representation of these loans in the examined sample is more than 25%. This is an interesting imbalance rising a question about its driver. Even though the study contains all of these results, it does not summarize them in this way in the text.

Further, the author shows that the diversification leads to a lower standard deviation than a standard deviation of any of the assets in the portfolios.

Methods

The used methodology is appropriate for a bachelor thesis, it builds on the Modern Portfolio Theory by Markowitz and uses return on investment as a measure of investment performance, the concepts are well described and the analysis is done properly. However, I would appreciate more detailed reasoning regarding the chosen methodology and discussion of limitations. There are multiple typos in the formulas (e.g. 4.3 should sum across t and no i; difference in the variance formula 4.5 is not squared and the explanation of the formula mentions that $\tilde{i}$s are different assets, the same term is used in the definition of the portfolio variance a couple of lines later which is confusing; 5.1 states that weights have to be larger than 1; 5.3 includes months but text refers to days; notation of matrices $X$ vs $T$ on page 59), the author skips some crucial steps in formula derivation (4.3.2 to 4.4.2), some calculations are not clearly described and explained (penalty calculation, description of the calculation on page 59).

Further, I think that the study should contain more descriptive statistics mainly distribution of loans and their size across credit categories, number and size of defaults by credit category, length of the loans. Also distribution of return on investment for each credit category would be a useful output.

Finally, I am not convinced that the presented way of calculation of variance-covariance matrix is the optimal solution. The author divides loan into groups based on issuance month and credit category and calculates the average return on investment, where the month represent the time series
dimension. The time series are then used for calculation of the variance-covariance matrix. The only
difference within a rating group is driven by defaults and experienced losses. Within such a short time-
frame, what is the reasoning of correlation in default rates and default losses of loans issued in
different months? Moreover, the time series are 36 months long focusing only on repaid or defaulted
loans, this could lead to a selection bias when only short term loans with potentially similar behavioural
patterns across rating groups are included in the analysis. Could this cause a bias in the return on
investment analysis? Some of the samples (month-rating groups) are rather small with less than 10
loans; this together with potential fat tails driven by defaults may lead to extreme observations, which
are overweighed in the average calculation. None of these potential issues is mentioned in the thesis.

Literature
The thesis provides a comprehensive list on foreign P2P studies listing different topics covered by the
papers. I would appreciate a more detailed overview of optimal portfolio literature, a short summary of
what other theses have found about Zonky data (there is several theses based on their dataset), and
an explanation how this study fits to the presented overview.

The section would benefit from a better structuring or introductory paragraph listing the topics covered
by the review. There is at least one missing reference in the List of references, Poláč (2017). There are
some parts of the study that are short on citations, for example the following would benefit from citing
the source:
- Introduction
  - This theory is usually used for analyzing stocks but proved to be useful also for
    analyzing P2P loans. (page 6)
- Chapter 1
  - some studies show that the growth will continue (page 6)

Manuscript form
The format of the thesis could be significantly improved. There are inconsistencies in fonts and
alignment (Table of content – alignment, List of figures – font inconsistencies).

The language is rather simplistic with short sentences, the text does not flow well in some parts (Introduction
- The data of this company will be analyzed. First, the P2P lending will be described
including the main characteristics, advantages as well as disadvantages, market and the Czech
company itself (page 6); Chapter 1 - The reason for that is the fact that the platforms do not hold any
of the loans themselves but they lend. (page 7); Chapter 2 - Until September 2017 the fee was 1 % for
every risk category. However, changed its fee policy which should enable it further expansion. (page
28)).

The thesis contains numerous grammar errors, typos and wrong punctuation:
- Introduction
  - was primarily concerned with the credit risk with this kind of business (page 5)
  - as disadvantages. market (page 6)
  - data in THE best possible way (page 6)
  - an investor who wish to reach (page 6)
- Chapter 1
  - studies shows (page 6)
  - United states (page 8)
  - due to the stringer regulatory policy (page 9)

The text is not gender neutral (Introduction - an investor lending his money). The preferred alignment
is “justify”. Footnote number should be placed outside of punctuation (institutions was only 48 %3.
(page 8)).
Report on Bachelor / Master Thesis

Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague

<table>
<thead>
<tr>
<th>Student:</th>
<th>Filip Jonáš</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advisor:</td>
<td>Mgr. Petr Polák, MSc.</td>
</tr>
<tr>
<td>Title of the thesis:</td>
<td>Portfolio optimization for an P2P investor on Zonky</td>
</tr>
</tbody>
</table>

Summary and suggested questions for the discussion during the defense

The thesis analyses an interesting question of optimal portfolio selection in the peer-to-peer lending market using Zonky data. The methodology is appropriate for a bachelor thesis and the results are interesting. However, the thesis would significantly benefit from a proof-reading. Some of the methodology considerations and conclusions are not mentioned in the text, the text often does not flow well, the formatting could be significantly improved and there are many typos both in the text and the formulas.

I recommend the thesis for defence and for reasons stated above I suggest grade D.

- What is the share of P2P lending in the lending market? What are the differences between countries (Czech Republic vs UK, US, EU, Asia)?
- How is the penalty in equation 5.5 calculated?
- How many loan are excluded from the analysis because they have not matured yet? Can this cause a selection bias?
- What is the reasoning of correlation in default rates and default losses of loans issued in different months across the rating categories within such a short time-period?
- What might drive the difference between optimal portfolio selection and the distribution of loans across credit categories in the Zonky market (there are almost no medium and high-risk loans in the optimal portfolio but almost 25% of the loans in the Zonky dataset are in these categories)?
- How might credit risk estimation in Zonky compare to traditional banks? Are the datasets that they use for the estimation comparable?
- What advice would you give to an investor interested in peer-to-peer lending based on the analysis and based on the phase of the business cycle?

**SUMMARY OF POINTS AWARDED** (for details, see below):

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>POINTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contribution</td>
<td>(max. 30 points) 20</td>
</tr>
<tr>
<td>Methods</td>
<td>(max. 30 points) 18</td>
</tr>
<tr>
<td>Literature</td>
<td>(max. 20 points) 15</td>
</tr>
<tr>
<td>Manuscript Form</td>
<td>(max. 20 points) 8</td>
</tr>
<tr>
<td>TOTAL POINTS</td>
<td>(max. 100 points) 61</td>
</tr>
</tbody>
</table>

GRADE (A – B – C – D – E – F) D

**NAME OF THE REFEREE:** Ing. Mgr. Barbora Máková M.A.

**DATE OF EVALUATION:** 28/05/2019

Referee Signature
EXPLANATION OF CATEGORIES AND SCALE:

CONTRIBUTION: The author presents original ideas on the topic demonstrating critical thinking and ability to draw conclusions based on the knowledge of relevant theory and empirics. There is a distinct value added of the thesis.

<table>
<thead>
<tr>
<th>Strong</th>
<th>Average</th>
<th>Weak</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>15</td>
<td>0</td>
</tr>
</tbody>
</table>

METHODS: The tools used are relevant to the research question being investigated, and adequate to the author’s level of studies. The thesis topic is comprehensively analyzed.

<table>
<thead>
<tr>
<th>Strong</th>
<th>Average</th>
<th>Weak</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>15</td>
<td>0</td>
</tr>
</tbody>
</table>

LITERATURE REVIEW: The thesis demonstrates author’s full understanding and command of recent literature. The author quotes relevant literature in a proper way.

<table>
<thead>
<tr>
<th>Strong</th>
<th>Average</th>
<th>Weak</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>10</td>
<td>0</td>
</tr>
</tbody>
</table>

MANUSCRIPT FORM: The thesis is well structured. The student uses appropriate language and style, including academic format for graphs and tables. The text effectively refers to graphs and tables and disposes with a complete bibliography.

<table>
<thead>
<tr>
<th>Strong</th>
<th>Average</th>
<th>Weak</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>10</td>
<td>0</td>
</tr>
</tbody>
</table>

Overall grading:

<table>
<thead>
<tr>
<th>TOTAL</th>
<th>GRADE</th>
</tr>
</thead>
<tbody>
<tr>
<td>91 – 100</td>
<td>A</td>
</tr>
<tr>
<td>81 - 90</td>
<td>B</td>
</tr>
<tr>
<td>71 - 80</td>
<td>C</td>
</tr>
<tr>
<td>61 – 70</td>
<td>D</td>
</tr>
<tr>
<td>51 – 60</td>
<td>E</td>
</tr>
<tr>
<td>0 – 50</td>
<td>F</td>
</tr>
</tbody>
</table>