FACULTY
OF SOCIAL SCIENCES
Charles University

Thesis Evaluation Report

Author:	Jan Picálek
Advisor:	Martin Hronec
Title:	Mispricing in leveraged value small-capitalization stocks
Opponent	Jiří Novák

Summary

The author uses machine learning to identify future winners and losers in a sample of highly leveraged, high book-to-market (i.e. value), small market capitalization stocks. The thesis builds on a recent working paper by Chingono and Rasmussen (2016) who identify that high financial leverage increases the small size premium and the value premium in realized stock returns. Using the machine learning algorithm applied to this set of firms, the author documents a substantial risk-adjusted returns earned on a long-short strategy out of the estimation sample.

Contribution

I find the author's motivation of the aim and the contribution of his thesis somewhat unfocused and disorganized, which makes it more difficult for me to evaluate how valuable the empirical analysis is. It seems to me that the thesis could be presented as an analysis that aims at discovering or discriminating between the underlying drivers of a recently documented empirical pattern on the impact of high financial leverage on the magnitude of the small size premium and the value premium in realized stock returns (Chingono and Rasmussen, 2016). If presented from this perspective, the author could aim to use machine learning to help identify potential economic justification for the recently discovered empirical pattern. This would certainly constitute an interesting research question. But to achieve this the introduction, the conclusion, and perhaps some parts of the empirical analysis would have to be presented differently.

The way the thesis is formulated it seems that the author simply finds that the machine learning algorithm produces substantial out of sample returns. While this could be considered important it is not easy to see what we can learn from such an analysis. Is leverage relevant because it is related to an additional risk dimension? Or is it related to imperfections in market infrastructure? Or is it related to mispricing due to some behavioral biases? I do not think the thesis helps us discriminate between these alternative explanations. What new insights does the algorithmic prediction of debt repayment offer? To a great extent the algorithm remains a black box giving little scope for an intuitive economic interpretation. Furthermore, it is not clear whether the results would be similar in different samples and how generalizable they are.

Methods

I appreciate that the author uses modern machine learning techniques. I also appreciate that he uses several risk models including the relatively recent five-factor model.

I do not understand why the author uses data from several stock markets that may differ in the form of regulation, liquidity, etc. Would not be a more homogenous sample from a single stock market be preferable?

If I interpret Tables 3.1 and 3.2 correctly only about 12% of the total observations meet the three qualification criteria. This also raises questions about generalizability of the findings.

The author claims he uses the criteria Rasmussen & Chingono (2015). I suggest he motivates the deviations (i.e. using 85th instead of 75th percentile, using book-to-market rather than *EBITDA*-to-enterprise value.

> *"This thesis builds upon the findings of Rasmussen & Chingono (2015), so our definition of the universe is equivalent to theirs."*

In computing accounting measures, it is typically preferable to use longer lags to make sure that all market participants have access to the information and that it is well-processed and incorporated to prices.

A more detailed discussion and motivation of the variables would benefit the thesis.

I think it would be helpful to discuss the implications of the findings that most of the effect comes from the short leg (Table 5.3). Small stocks may not be easy to short. Furthermore, does the author offer any explanation why the short leg has strongly negative loading on the market factor (column (5) in Table 5.3)?

Can the author comment on why Figure 5.3 suggests (in contrast to Table 5.3) that the strongest effect comes from the long leg (rather than short leg)?

Excluding the delisted firms without considering the delisting return introduces a look-ahead bias.

> *"We treat the delisted stocks by excluding them from the data one month prior its delisting date."*

I believe that the criteria for setting some accounting items to missing should be more explicit.

> *"Individual accounting figures that appear unreliable due to a conflict with accounting standards are set to missing. E.g., positive cost items or negative cash dividends paid. On top of that, firm-year observations with apparently illogical or highly anomalous values in essential accounting variables are dropped from the sample. This covers long-term debt greater than assets, profit margins greater than 100%, or deeply negative etc."*

Literature

I appreciate that the author reviews a large number of relevant studies.

It is not always clear to me why some sections of literature review are included. For example, the section on the portfolio theory and the discussion of stock returns following equity IPOs, SEOs and debt issues.

I believe the literature review ends rather abruptly. It lacks a clear conclusion that would lead up the author to formulating his hypotheses and motivate his contribution to the reviewed literature.

I find some statements highly questionable and not sufficiently supported by references. For example, is it actually well-established that capital structure is a risk-based factor? Furthermore, is it actually clear that returns following capital issues result from mispricing?

> *"While capital structure is a risk-based factor, the events of capital raising and their implications on future returns are recognized as rather behavioral anomalies."*

Form

I think the thesis could be better organized and focused on the main message. I am not convinced that discussing at length the portfolio theory and the construction of pricing models is essential. A sharper emphasis on the small size premium and the value premium and especially on the discussion of financial leverage as a potential moderating variable would benefit the thesis. I would also expect a more explicit discussion of why using machine learning can yield new insights. It is not quite clear to me whether the main intention of machine learning is to predict returns or to predict reductions of financial leverage (that could be correlated with returns).

It would be helpful to include page numbers.

The thesis is written in good English.

Conclusion

The results of the Urkund analysis do not indicate significant text similarity with other sources.

I believe that the Thesis fulfills the requirements stipulated by the Faculty of Social Sciences, Charles University. I recommend the Thesis for defense. I suggest a grade C.

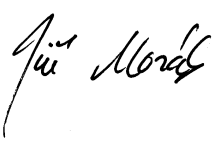
Questions

I recommend the examination committee to ask the author about the following:

- What is the main research benefit in using machine-learning algorithms in this setting? What insights does it give?
- What can we learn from the thesis about the nature of the Rasmussen & Chingono's (2015) findings?
- To what extent are the results driven by the long leg and the short leg (compare Figure 5.3 and Table 5.3)?

Awarded Points and Grade

Contribution (max 30)	20
Methods (max 30)	23
Literature (max 20)	16
Form (max 20)	18
Total (max 100)	**77**
Grade (A – B – C – D – E – F)	**C**



Referee's Signature

9 June, 2022

Evaluation Date

Jiří Novák

Referee's Name

Grading Scale

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

Strong	Average	Weak
20	10	0

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.

Strong	Average	Weak
30	15	0

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.

Strong	Average	Weak
30	15	0

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.

Strong	Average	Weak
20	10	0

OVERALL GRADING:

Total Points	Grade
91 – 100	A
81 – 90	B
71 – 80	C
61 – 70	D
51 – 60	E
0 – 50	F