The Journal of Finance and Data Science

Machine Learning Private Equity Returns

--Manuscript Draft--

Manuscript Number:	JFDS-D-23-00185R1
Article Type:	Research Article
Section/Category:	Financial Economics
Keywords:	Return factor model; Private Equity; Public factor exposure; Model combination; machine learning; Ensemble Learning
Corresponding Author:	Christian Tausch AssetMetrix Munich, Germany
First Author:	Christian Tausch
Order of Authors:	Christian Tausch
	Marcus Pietz
Abstract:	In this paper, we use two machine learning techniques to learn the aggregated return time series of complete private capital fund segments. First, we propose Stochastic Discount Factor (SDF) model combination to determine the public factor exposure of private equity. Here, we describe our theoretical motivation to favor model combination over model selection. This entails that we apply simple coefficient averaging to obtain multivariate SDF models that mimic the factor exposure of all major private capital fund types. As a second step, we suggest component- wise L2 boosting to estimate the error term time series associated with our factor models. The simple addition of the public factor model returns and the error terms then yields the total return time series. These return time series can be applied for proper integrated public and private risk management or benchmarking.
Suggested Reviewers:	Arthur Korteweg, Professor University of Southern California KORTEWEG@MARSHALL.USC.EDU Expert in advanced methods for Private Equity
	Andrew Ang Columbia University aa610@columbia.edu Authored a very similar paper.
Response to Reviewers:	

Machine Learning Private Equity Returns

Christian Tausch
AssetMetrix GmbH
Theresienhoehe 13, D-80339 Munich
christian.tausch@quant-unit.com

Marcus Pietz
AssetMetrix GmbH
Theresienhoehe 13, D-80339 Munich
marcus.pietz@asset-metrix.com

October 7, 2024

Keywords:

Return factor model, Private equity, Public factor exposure, Model combination, Machine learning, Ensemble learning

Acknowledgements

We thank Nicolas Duetsch, Philipp Abel, and Kai Urban for valuable feedback and helpful comments that greatly improved the paper. Further, we want to thank Jingzhi Huang, the editor, and two anonymous referees for their inspiring suggestions that helped to significantly improve the depth and robustness of our paper.

Declaration of interest

The authors report no conflict of interest. The authors alone are responsible for the content and writing of the paper.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The author is an Editorial Board Member/Editor-in-Chief/Associate Editor/Guest Editor for [Journal name] and was not involved in the editorial review or the decision to publish this article.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Response to Referee # 2

Dear Referee,

Thank you sincerely for taking the time to review our paper. We greatly appreciate your thoughtful feedback, which will help us significantly improve our work. In this letter, we have carefully addressed each of your comments and provided our responses to your suggestions. All revisions made in response to referee feedback are highlighted in green in the updated submission.

Major Comments:

1. Comment: Comparison with Existing Models: While the paper presents a new model for generating returns from private equity cash flows, it lacks a direct comparison with existing methods. The mention of calculating the Internal Rate of Return (IRR) directly in Section 4.2 suggests an opportunity for comparison. Even if the proposed model does not always outperform existing methods, its unique advantages can still be highlighted. This comparison would strengthen the paper's contribution by demonstrating its relative performance.
    a. Response: Agree, we newly added Appendix A “Comparison to competing approaches” on pages 22 and 23 where we discuss the intricacies associated with existing methods for (1) return measurement, (2) factor-model estimation, and (3) error-term derivation. Here we also establish that the fundamental problem is stale pricing of NAVs which makes the measurement of “true” PE returns infeasible.
2. Comment: Multi-dimensional Validation: The paper currently validates the model's effectiveness primarily through visual comparisons of cumulative returns and autocorrelation functions (ACFs) between the model-generated returns and NAV returns from different data sources. While these visual comparisons are intuitive, they lack statistical rigor. I recommend incorporating additional statistical metrics like R-squared, Mean Squared Error (MSE), or other relevant measures to quantitatively assess the model's accuracy and performance. This would provide a more comprehensive evaluation of the model's effectiveness across different private equity fund types and time periods.
    a. Response: Agree, that we can add quantitative measures of model performance. However, as “true” PE returns are unobservable, we need to measure the distance between NAV returns (which are only proxy returns) and our model returns. We added Table 3 with quantitative measures like MSE, MAE, … on page 17, which is referenced on page 13.

Minor Comments:

1. Comment: Section 3 discusses the reasons for using ensemble methods for private equity funds, setting the stage for the use of combination and boosting techniques. Some reasons are repetitive and could be consolidated or reduced in length. For example, 3.1 mentions data privacy as a reason for limited data, while 3.2 discusses

the late start of data collection and limited sample sizes, both of which contribute to data scarcity. Additionally, 3.3 also mentions limited data.

   a. Response: Generally, we agree that we could further streamline Section 3. However, given that Section 3 is only a little longer than one page, we prefer to keep our original structure where we briefly highlight the data scarcity issue from several dimensions. We kept the original version unchanged and hope this is acceptable to you.

2. Comment: Section 4.1 proposes methods for selecting valid models (M*), but the empirical section does not specify the value of x used (likely x = 1).
   a. Response: Agree, we confirm now that we do not exclude any model on page 9-10.
3. We newly introduced Section “5.3 Model limitations for other fund types” on page 17 as response to your final three points that are all tackled in this section. Thus, we agree to all three observations and chose to answer them in a combined manner.
   a. Comment: In the conclusion part, the results for Buyout funds could be more thoroughly discussed in the empirical section. We could focus more on the model's strengths and weaknesses in this part. The conclusion also mentions potential unreliability of the model at the beginning and end of the error term time series. It would be better to suggest possible solutions for this issue.
   b. Additional Suggestion: Exploration of Model Limitations: Discuss potential limitations of the proposed model, such as its assumptions or potential biases. This would enhance the paper's transparency and credibility.
   c. Additional Suggestion: Investigation of Model Robustness: Compare model from 3 data sources with 3 NAV returns by metrics to validate the model's robustness. This would further strengthen the paper's conclusions and enhance its generalizability.
   d. Response to (a), (b) and (c): Agree, we newly introduced Section 5.3 on page 16-17 which refers to the new Appendix B that contains return charts for multiple fund types on pages 24-38. So, we discuss the model limitations with reference to other fund types where the model’s shortcomings (outliers) are more obvious. We hope Appendix B is not too long for the Journal.

In total, we have added 17-18 new pages to incorporate your suggestions, with 16 of those pages included in the Appendix and 14 pages being only charts. Thank you very much, for your valuable feedback!

Response to Referee # 1

Dear Referee,

Thank you sincerely for taking the time to review our paper. We greatly appreciate your thoughtful feedback, which will help us significantly improve our work. In this letter, we have carefully addressed each of your comments and provided our responses to your suggestions. All revisions made in response to referee feedback are highlighted in green in the updated submission.

Must be addressed

Comment: Even if you have already mention the abbreviation for SDF in the abstract, please use the whole expression in the main text for the first time that you are using it.

Response: Agree, we added the whole expression on page 2 now.

Could be addressed

Comment: It would be interesting to use the Model Confidence Set method (Hansen et al., 2021) to determine the valid models and use those for the model averaging to see whether the performance improves or not. The Model Confidence Set method is quite efficient and not very time-consuming. Souto (2023) provided easy-to-implement Python code for this test.

Response: The Model Confidence Set method is indeed a very interesting and relevant approach that can support and improve our model selection efforts. We added the original reference for Hansen et al. (2011) on page 6 of the paper. For the empirical experiments in our R code, we used the MCS package since it is written in R instead of Python. We updated the code in our online repository accordingly.

I uploaded the Revised Manuscript (without author info) – Clean as Supplementary Material since it is in PDF format and the system only allows docx files for the revised manuscript. This is a bug that shall be fixed as soon as possible.