Machine Learning Private Equity Returns

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

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

September 20, 2024

Keywords

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

Acknowledgements

We thank Nicolas Dütsch, Philipp Abel, and Kai Urban for valuable feedback and helpful comments that greatly improved the paper.

Declaration of interest

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

Machine Learning Private Equity Returns

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 L_2 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.

1 Introduction

Factor investing has gained widespread popularity in public markets due to its ability to discern the underlying drivers of returns. Articulating a portfolio's exposure to public factors entails a comprehensive understanding of both its realized returns and the expectations for future risk and return. For private market investors, meaningful insights can be derived from realistic factor decompositions, presuming the existence of shared components influencing both public and private returns. In light of this perspective, our factor analysis dissects private equity returns into a traded return component and its corresponding error terms. The primary applications of our return decomposition extend to public market equivalent benchmarking, comprehensive risk management, and strategic asset allocation.

Generally, more sophisticated methods are required to estimate factor models for private (in comparison to public) asset classes as private returns are not directly observable on liquid secondary markets. However, the challenge of identifying the 'best' factor model is the same as in public markets. There are first attempts to select or create public indices that shall replicate private equity (PE) returns. Here, we can distinguish between (i) factor- and (ii) holding-based approaches. Phalippou (2014) benchmarks buyout funds against several factor indices. More subtle factor-based methods are mainly established in the hedge fund replication literature (Tancar and Viebig, 2008; Weisang, 2014). To avoid overfitting, O'Doherty et al. (2017) propose to combine multiple factor models for passive hedge fund replication. Holding-based approaches to mimic the PE investment style are proposed by L'Her et al. (2016), Stafford (2017), Madhavan and Sobczyk (2019), and Porter and Porter (2019). This paper pursues a factor-based solution that employs publicly traded indices and does not require deal-level information on PE funds. In contrast to the factor-based ansatz of Phalippou (2014), we base our approach on Stochastic Discount Factor (SDF) models that are estimated on PE fund-level data rather than comparing several ad-hoc choices for potentially suitable indices.

This article aims at a data-science-driven solution to describe PE returns.[1] Concretely, we propose SDF model combination (of factor models estimated by the Driessen et al. (2012) method) as a straightforward method to obtain 'strong' public factor models that explain private equity returns. Here, we provisionally define the term 'strong' versus 'weak' model in analogy to the boosting literature (Schapire, 1990). A 'strong' model shall exhibit small (error term) bias and variance, whereas a 'weak' model displays high bias and variance. Only the 'true' or 'best' or 'optimal' model obtains an error term of precisely zero.

As we cannot expect a perfect model with zero error terms in any realistic setting, we try to explicitly model these error terms by means of machine learning. Usually, estimation procedures for linear return factor models like $R_t = \alpha + \beta^\top X_t + e_t$ are mainly interested in estimating the factor loadings β and the constant outperformance term α. For private equity fund returns only Ang et al. (2018) propose a Markov Chain Monte Carlo (MCMC) approach to estimate also the error terms e. As an alternative to their method, our paper suggests a statistical learning technique called componentwise L_2 boosting to estimate the error term time series belonging to a given public factor model. Moreover, in stark contrast to Ang et al. (2018), our model does not assume a parametric distribution for the error term e. Figure 1 illustrates our general machine learning approach to derive the public and private part of private equity returns.

The paper is structured as follows. Section 2 introduces the method's underlying semiparametric setting. Section 3 explains why model selection is problematic. Section 4 introduces two estimation procedures: model combination for the public factor model and componentwise L_2 boosting for the error terms. Section 5 successfully applies the estimation procedures to a real-world dataset. Section 6 concludes. Finally, the `R` code used for model estimation can be found in the following online repository `https://github.com/quant-unit/Fundwise_SDF`.

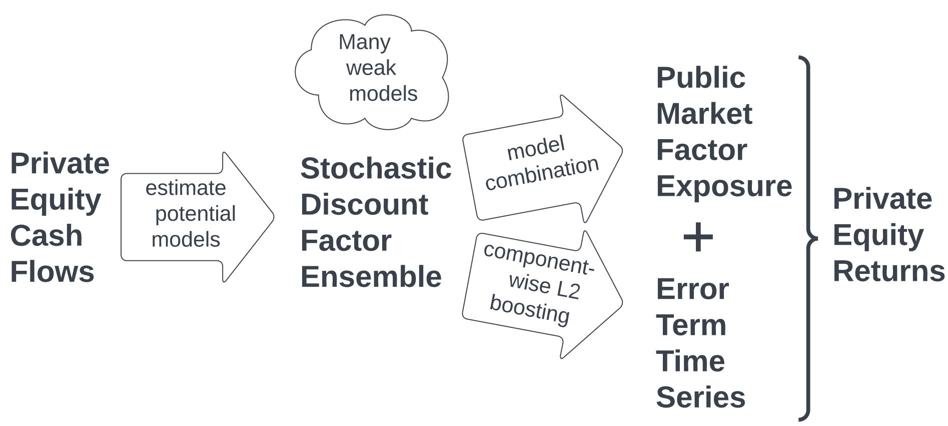


Figure 1: How to translate private equity cash flows to private equity returns?

[1]See Coqueret and Guida (2023) for a textbook treatment of machine learning methods for factor investing.

2 Semiparametric setting

Stochastic Discount Factors (SDFs) are econometric models to price a given cash flow stream (Hansen and Richard, 1987). They are usually estimated by semiparametric approaches that go without a parametric model for the error term (Ferson, 2019). Following Driessen et al. (2012), the cash flows stream under consideration is generated by $i = 1, 2, \ldots, n$ private equity funds (or portfolios). SDFs can be applied in net present value calculations for realized cash flow paths

$$P_{\tau,i} = \sum_{t=1}^{T} \Psi_{\tau,t} \; CF_{t,i} \tag{1}$$

with price P, SDF functional Ψ, and cash flow stream CF. Time is discrete with $t = 1, 2, \ldots, T$. If the 'true' SDF model is applied, the expected price $\mathbb{E}[P] = 0$.[2] Thus the functional form and explanatory variables used in the 'true' SDF explain or describe the risk and return properties of the cash flows. Further, if we have an appropriate SDF for a given private equity fund type, we can apply it in a simple net present value calculation (as in the formula above) to assess the outperformance of a specific private equity fund. A positive/negative net present value indicates an out/under-performance compared to other PE funds. To discount a time t cash flow to time τ, we use the simple linear multi-period SDF model

$$\Psi_{\tau,t} = \frac{\prod_{h=0}^{\tau} \left(1 + \alpha + r_h + \sum_j \; \beta_j \; F_{j,h} + e_h\right)}{\prod_{h=0}^{t} \left(1 + \alpha + r_h + \sum_j \; \beta_j \; F_{j,h} + e_h\right)} \tag{2}$$

with usually $\alpha = 0$, $e_h = 0 \; \forall \; h$, risk-free rate r, zero-net-investment factor return F, and factor coefficient β that has to be estimated from data. Consequently, the expected multi-period return for a given asset is modeled by

$$E\left[R_{\tau,t}\right] = E\left[\frac{1}{\Psi_{\tau,t}} + \epsilon_{\tau,t}\right]$$

where the period-specific error term $\epsilon_{\tau,t}$ has zero expectation $E[\epsilon_{\tau,t}] = 0$.

As commonly seen in the asset pricing literature, we estimate SDF model coefficients by semiparametric approaches that require no distributional assumption for $\epsilon_{\tau,t}$. This means we select α and β values that yield average pricing errors close to zero. Here the typical loss function choices are quadratic forms like in Generalized Method of Moments (GMM) or a quadratic or least absolute deviance loss function $L()$

$$\hat{\theta} = \; \arg\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} L\left(P; \theta, \gamma\right) \tag{3}$$

where $\theta = (\alpha, \beta)$ and γ denotes the vector of potential hyper-parameters like data cutoff dates or weighting/averaging methods. More details on the semiparametric estimation framework can be found in the original paper of Driessen et al. (2012) and

[2] We also refer to P as pricing error since we want to avoid any average price deviation from zero, i.e., minimize the expected pricing error.

its credit market factor application by Hüther et al. (2023). As a little modification of their native method, we average over multiple compounding dates $\tau \in \mathcal{T}$ in our quadratic loss function

$$L\left(P;\theta,\gamma\right) = \left(\frac{1}{|\mathcal{T}|}\sum_{\tau\in\mathcal{T}} P_{\tau,i}\right)^2 \tag{4}$$

to decrease the small-sample bias and variance of the estimator.

3 So many weak SDF model candidates

In the semiparametric setting described in section 2, we likely obtain many weak model candidates with no clear winner among them. This scenario is comparable to the extensive "factor model zoo" observed in public equity markets, as coined by Cochrane (2011); Feng et al. (2020). Furthermore, the extensive body of public market literature fails to definitively determine which factors provide the most comprehensive explanation for their assets' returns. In this section, we explain in greater detail why we also expect to be confronted with (i) many but (ii) weak SDF model candidates in the private equity case. Moreover, we discuss why it is challenging to select the single 'best' model from an extensive collection of homogeneous competitors (that all use the same SDF model as specified in equation 2).

3.1 Why are there so many models?

First, there exists a great variety of different public return factor candidates that might be also relevant for private equity. It is not even clear which factors 'best' explain public equity returns, cf. references in Karolyi and Van Nieuwerburgh (2020).

Second, we can choose from many feasible semiparametric estimators, loss functions, or ancillary machine learning methods, cf. Gu et al. (2020) for a public market overview. Additionally, there are discretionary choices for regularization- or other hyper-parameters. Further, bootstrapping and other resampling-based techniques may be helpful but likewise dramatically increase the number of model candidates.

Third, in contrast to common public databases available for public equity, the private equity literature has to draw on various private – often proprietary – data sources. When estimating the same model on different datasets, the coefficient estimates will differ. Unclear data cutoff dates or horizons exacerbate this issue.

3.2 Why are the models so weak?

First, private equity data is notoriously sparse as the first PE funds were established in the 1970s to 80s. Moreover, it is currently (almost) impossible to collect data on all funds from a given vintage. These issues render finite [infinite] sample inference more [less] reliable.

Second, for fund-level data, the dependency issue introduced by overlapping fund cash flows further reduces the number of truly independent observations. Similarly, financial engineering applied on fund level like bridge financing or Net Asset Value (NAV) facilities undermine the explanatory power of observed fund cash flows.

Third, in private equity, limited sample sizes typically allow the inclusion of just 1-2 factors in a SDF model; even many two-factor models appear to be statistically insignificant (weak models); models with more covariates almost surely overfit (invalid models).

3.3 Why is model selection difficult?

First, model uncertainty is particularly high for weak models, as we cannot be sure which single model performs best when we test several competing models/factors.

Second, as there are many model candidates but limited data, data snooping may be a general problem (White, 2000).

Third, correct inference post model selection is essential but known to be challenging (Bachoc et al., 2019). This means classical statistical inference results only apply for testing a single model. When we want to compare several model candidates, the application of these basic statistical methods yields invalid inference results.

4 And the jury selects: model combination

A powerful alternative to model selection is model combination. Generally, Bayesian and frequentist model averaging is a well-studied field in classical statistics (Hansen, 2014; Moral-Benito, 2015). Similarly, ensemble learning, a prominent subfield of machine learning, builds on the notion to form a strong learner by combining many weak learners (Bühlmann, 2012). In more applied settings, predictive forecast combination often helps to increase forecast precision (Huang and Lee, 2010). In a public market setting, Rapach et al. (2010) forecast the public equity premium by combining 15 univariate OLS regressions of different macroeconomic predictors. Probably most closely related to our case, O'Doherty et al. (2017) advertise model combination for passive hedge fund replication.

4.1 Model averaging

As we potentially want to exclude invalid models from combination, the set of valid models is defined as the M^* models with the smallest absolute pricing error $M^* = \{m \in M : P^{(m)} \leq Q(x; P)\}$ where Q is the pricing error quantile function and quantile value $x = \frac{M^*}{M}$. Here, M is the number of all valid and invalid SDF models. Generally, we can also apply more sophisticated – but also more time-consuming – cross-validation procedures (Arlot and Celisse, 2010) or the Model Confidence Set approach of Hansen et al. (2011) to distinguish between valid and invalid models. The simpler alternative is to assume that all SDF models estimated by the researcher are valid, i.e. $M^* = M$ with $x = 1$. Finally, the weighted pricing error obtained by

SDF model averaging is defined as

$$\mathcal{E}_{\tau,i}^{(M^*)} = \sum_{m=1}^{M^*} w_m \sum_{t=1}^{T} \Psi_{\tau,t}^{(m)} \; CF_{t,i} \tag{5}$$

with model weight $w_m \geq 0$ and all weights sum to one $\sum_m^{M^*} w_m = 1$. If a similar predictive performance for all valid models can be assumed, equal weighting $w = (M^*)^{-1}$ is advised.[3]. Otherwise, we shall overweight more predictive models with smaller pricing errors, e.g., the weight may be proportional to the inverse of the absolute pricing error $(|P|)^{-1}$.

4.2 Coefficient averaging

The general model combination approach defined by equation 5 allows the aggregation of structurally distinct SDF models. If we always employ the same linear SDF as defined in equation 2, which can be replicated by an investable trading strategy, model combination can be interpreted as a diversification strategy. The idea is to invest in several promising strategies rather than committing entirely to a single "optimal" trading strategy. Using always the same linear SDF model enables the following coefficient averaging formula, which only approximates the diversified pricing error from equation 5 (the better when the factor returns – thus return horizons – are small).

$$\beta_j^{(M^*)} = \sum_{m=1}^{M^*} w_m \; \beta_{j,m} \tag{6}$$

The idea here is that one multivariate model that includes all traded factors can be better perceived and interpreted as an ensemble of M^* heterogeneous one- or two-factor models.

Finally, with the $\beta_j^{(M^*)}$ coefficient estimated, we can focus on the α term in equation 2. To quantify the public market out/underperformance, we can apply the Excess-IRR method described by Phalippou and Gottschalg (2009), which estimates the α term by setting the pricing error to precisely zero for each i

$$\hat{\alpha}_i = \arg\min_{\alpha_i} \left| P_{\tau,i}^{(M^*)} \right| \tag{7}$$

Alternatively, we can directly calculate the Internal Rate of Return (IRR) associated with the SDF discounted cash flows (with $\alpha = 0$ in equation 2), which corresponds to the direct-alpha method described by Gredil et al. (2023). In both cases, we estimate the constant out/underperformance term for the ith portfolio after adjusting for systematic risk factors.

[3] The forecast combination puzzle states that simple equal-weighting practically often performs better than more complex (theoretically optimal) combination schemes (Smith and Wallis, 2009; Claeskens et al., 2016; Qian et al., 2019).

4.3 Estimating idiosyncratic returns

In the spirit of equation 7, we can also try to find an estimate for the series $e = (e_h)_{h=1,2,...,T}$ from equation 2 instead of estimating the constant α. For illustrative purposes, we select here a conceptually simple, but brute-force machine learning algorithm called componentwise L_2 boosting (Bühlmann, 2006). In general, we can choose any high-dimensional statistical method that can be used for variable selection like the canonical lasso (Bühlmann and Van De Geer, 2011). In the probably most closely related paper to our problem, Ang et al. (2018) use a Bayesian Markov Chain Monte Carlo (MCMC) method to estimate the systematic and also idiosyncratic return component of private equity returns.

In our approach, we start with a set of α and β estimates and learn the idiosyncratic error term time series e conditional on these initial estimates. Concretely, we estimate in the first step an ensemble of size $M_{\text{systematic}}$ consisting of reasonable estimates for α and β by iteratively applying traditional non-linear optimization (using a different set of factors F, dataset samples, or hyperparameters γ in each iteration).

In the second step, we use componentwise L_2 boosting to estimate a new series e for each of the $M_{\text{systematic}}$ public factor ensembles. Componentwise L_2 boosting (CLB), also known as gradient boosting with the L_2 loss function, is a machine learning algorithm that belongs to the family of boosting methods. Boosting is an ensemble learning technique where weak learners (usually simple models like decision trees) are sequentially trained, and each new learner focuses on correcting the errors made by the existing ensemble. In the context of boosting, "componentwise" refers to the fact that each weak learner added to the ensemble is trained to fit the residual errors of the existing ensemble. This is done in a way that the new learner addresses a specific component of the overall prediction. The original paper of (Bühlmann, 2006, Section 2) gives the exact mathematical definition of the CLB algorithm. Financial applications of the CLB method can be found in Bai and Ng (2009), Mittnik et al. (2015), and Tausch (2019).

In our case, the CLB algorithm can be described by the following "brute-force" iterative procedure. We apply these CLB steps to each public factor model (of the previously estimated ensemble of size $M_{\text{systematic}}$) separately.

1. **Initialization:** The boosting process starts with the public factor model without error terms ($e_t = 0 \ \forall\ t$) as initial model. The first weak learner is then trained to minimize the L_2 loss on the training data, considering the difference between the true outcomes and the predictions made by the current ensemble (which is the initial factor model).

2. **Sequential Training:** In each subsequent iteration, a new weak learner – in the form of an error vector e update – is added to the ensemble. This learner is trained to fit the negative gradient of the L_2 loss with respect to the current ensemble's predictions. In other words, it focuses on the residuals, the differences between the true outcome (which is a zero pricing error in equation 4) and the pricing error made by the current ensemble. This means that, at each step, the boosting algorithm is effectively optimizing the ensemble by sequentially addressing the weaknesses of the current predictions.

3. **Componentwise Correction:** The term "componentwise" emphasizes that each weak learner corrects only one specific component (date) of the overall error time series e. Concretely, one boosting step simultaneously optimizes each component of the vector e to reduce the pricing error. However, only the component of the e vector (i.e., a single date t) that yields the minimal pricing error, when compared to all other optimized e vector components, is used to update the factor model.
4. **Shrinkage:** Componentwise L_2 boosting often incorporates a shrinkage parameter between zero and one, which controls the contribution of each weak learner to the ensemble (in our empirical example: 0.33). This helps prevent overfitting.
5. **Early Stopping:** The steepness of the gradient or cross validation techniques can be employed to determine the optimal stopping time of the algorithm. With too many boosting iterations CLB likely overfits. In our empirical example, we stop after 200 boosting steps.
6. **Final Prediction:** The final prediction for the error term time series e is the sum of the predictions made by all the weak learners in the error ensemble.

Finally, in the spirit of model combination, we average over the $M_{\text{systematic}}$ estimates of α, β and e to obtain our final factor model estimates. Since $\alpha = E[e]$, we can, without loss of generality, set $\alpha = 0$ in our estimation procedure. In practice, it could however be beneficial in some cases to start with a "good initial guess value" for α.

5 Empirical application

5.1 Factor exposure by fund segment

In this section, we apply our model combination methodology to replicate the factor exposure of PE funds. The resulting SDF models employ MSCI indices and are estimated by our augmented version of the Driessen et al. (2012) method. Specifically, all SDF ensembles draw on (valid) two-factor models, where the first factor is always the MSCI World excess return over the risk-free rate and the second factor is a long-short combination of MSCI World style indices:

MKT-RF: MSCI Market Return Minus Risk-free Rate

SMB: MSCI Small Cap Minus MSCI Large Cap Return

HML: MSCI Value Minus MSCI Growth Return

HDY-MKT: MSCI High Dividend Yield Minus MSCI Market Return

QLT-MKT: MSCI Quality Minus MSCI Market Return

For each of the 4 factors (SMB, HML, HDY-MKT, QLT-MKT), we estimate an ensemble of $2\times2\times5$ bivariate two-factor models with (i) quadratic and least absolute deviance loss function $L()$, (ii) both equal- and fund-size-weighted cash flows, and (iii) maximum months 120, 150, 180, 210, 240. Next, we use the simple coefficient averaging as described in subsection 4.2 to obtain five-factor models.

Type	MKT-RF	HML	SMB	HDY-MKT	QLT-MKT
BO	1.4 (0.15)	-0.1 (0.02)	-0.09 (0.02)	-0.15 (0.03)	-0.1 (0.04)
DD	1.04 (0.25)	0.25 (0.02)	0.29 (0.02)	0.54 (0.07)	0.01 (0.06)
INF	0.71 (0.45)	0.13 (0.09)	0.13 (0.04)	0.22 (0.08)	0.73 (0.38)
MEZZ	0.75 (0.16)	0.03 (0.04)	0.16 (0.07)	0.14 (0.05)	0.14 (0.15)
NATRES	0.48 (0.29)	0.02 (0.15)	0.06 (0.17)	0.1 (0.23)	-0.37 (0.09)
PD	0.69 (0.25)	0.13 (0.05)	0.19 (0.08)	0.31 (0.11)	0.21 (0.14)
RE	1.05 (0.5)	0.02 (0.1)	-0.33 (0.07)	-0.43 (0.17)	-0.55 (0.11)
VC	1.05 (0.71)	-0.67 (0.08)	-0.56 (0.06)	-0.92 (0.1)	0.77 (0.8)
MKT	1	0	0	0	0

Table 1: PITCHBOOK 2023: Multivariate five-factor models obtained by simple coefficient averaging (with standard deviations in parenthesis).

We use the Pitchbook dataset as of 19th April 2023 to analyze the following private capital fund types: Buyout (BO), Distressed Debt (DD), Infrastructure (INF), Mezzanine (MEZZ), Natural Resources (NATRES), Private Debt (PD), Real Estate (RE), and Venture Capital (VC). For these fund types, we extract all (i) equal-weighted and (ii) fund-size-weighted cash flow series. For non-liquidated funds, we treat the latest net asset value as a final cash flow. We explicitly refrain from including the most recent vintage years to analyze only funds with at least a few years of history. Thus, the minimum vintage year is 1996, and the maximum is 2017. Table 1 provides an overview of the Pitchbook dataset used for SDF model estimation.

Type	Annualized Return				Sharpe Ratio
	mean.R	stdv.R	mean.R-RF	stdv.R-RF	mean/stdv.R-RF
BO	**0.141**	0.223	**0.117**	0.223	0.526
DD	0.123	0.160	0.099	0.160	0.616
INF	0.106	0.093	0.082	0.093	**0.887**
MEZZ	0.096	0.113	0.072	0.113	0.640
NATRES	0.056	**0.085**	0.033	**0.085**	0.389
PD	0.093	0.101	0.070	0.101	0.694
RE	0.089	0.183	0.066	0.183	0.362
VC	0.118	0.204	0.094	0.205	0.459
MKT	0.110	0.155	0.086	0.155	0.555

Table 2: PITCHBOOK 2023: Annualized average returns, standard deviations (annualized by the square root of time formula), and Sharpe ratios (i.e. the ratio of mean.R-RF to stdv.R-RF) implied by the five-factor models from table 1. The underlying monthly returns are based on MSCI World style indices (in USD) from 1996-01-31 to 2023-02-28.

The average coefficients of the 20 SDF model returns are displayed in table 1. For all fund types, the market beta estimates (MKT-RF) reasonably range from 0.48

for NATRES to 1.4 for BO. All other factors' coefficient estimates (SMB, HML, HDY-MKT, QLT-MKT) feature absolute values smaller than one. In summary, coefficient averaging generates heterogeneous factor models for the fund types under investigation, which consistently appear plausible and suitable. Figure 2 exhibits the associated cumulative returns implied by the average factor model. Here, we see that the BO and DD strategies outperform other fund types in the sample period from 1997 to 2023. The lowest cumulative factor model returns are observed for NATRES and RE. For VC we can observe a peak in the dot-com bubble of the early 2000s followed by a long period of public market underperformance until the year 2020. Table 2 shows that many five-factor models can outperform the MSCI World in terms of annualized return and Sharpe ratio. Interestingly, fund type INF exhibits the highest Sharpe ratio of 0.887. Sharpe ratios bigger than 0.6 are further observed for the debt-related fund types DD, MEZZ, and PD. Surprising, many popular fund types obtain smaller Sharpe ratios than the market ratio of 0.555 (BO, NATRES, RE and VC). In summary, all SDF models from table 1 seem reasonable for public market benchmarking and risk mapping purposes.

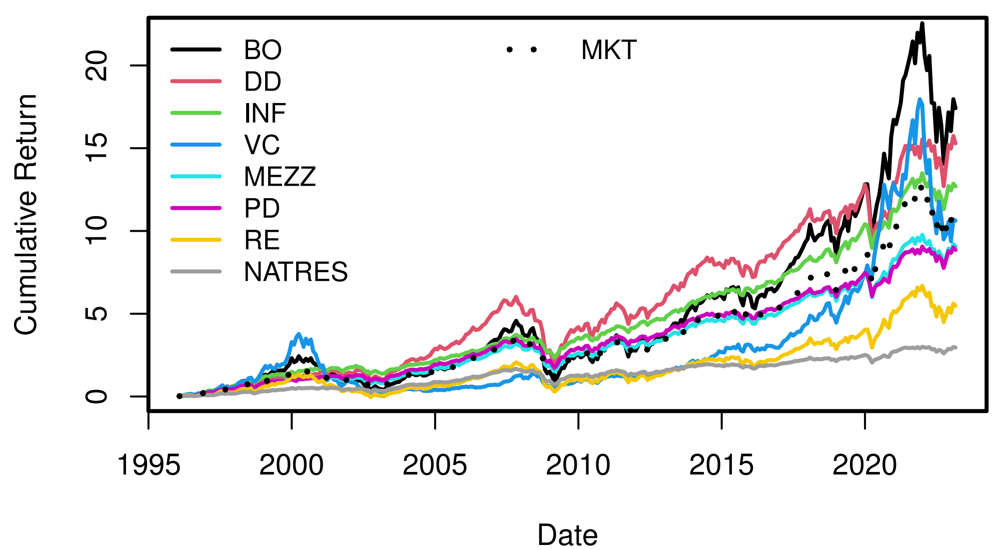


Figure 2: PITCHBOOK 2023: Cumulative USD returns implied by the five-factor models from Table 1 from 1996-01-31 until 2023-02-28.

5.2 Idiosyncratic returns for Buyout funds

In this section, we apply the idiosyncratic return estimation method described in subsection 4.3 for BO funds. Here we use a starting point an ensemble of the MSCI market factors similar to the one described in Table 1. Now we however only include fund-size-weighted datasets to reduce the long computation time that comes with this brute-force approach[4]. Specifically, we use an ensemble of 40 two-factor models

[4]Estimation for the reduced Buyout ensemble takes around three days on a 2023 MacBook with 12 CPU cores and parallel processing.

$4 \times 2 \times 5$ with (i) second factors (HML, SMB, HDY-MKT, QLT-MKT) (ii) quadratic and least absolute deviance loss function $L()$, and (iii) maximum months 120, 150, 180, 210, 240. The average factor loadings of the five-factor model are fortunately very similar to the ones from Table 1: MKT-RF (1.46), HML (-0.09), SMB (-0.10), HDY-MKT (-0.15), and QLT-MKT (-0.09).

In the second step, we apply componetwise L_2 boosting (CLB) with 200 iterations for each two-factor ensemble, a damper factor of 0.33, and return bounds of plus and minus 100%. For each of the 40 ensembles, we start with a vector of 307 zeros for e (i.e. 307 monthly observations from 1998-03-31 until 2022-06-30). After all CLB algorithms have been terminated, we average over all 40 error term series and public factor models to obtain our final estimate for e and β. Since only 46/307=15% monthly errors are filled with values other than zero in the final e vector, our estimated error term series can be considered relatively sparse. On a quarterly basis we still have 43/96=45% non-zero error terms from 1998 until 2022. Yet two quarters exhibit quite extreme idiosyncratic returns of around +30% or larger as depicted in Figure 3.

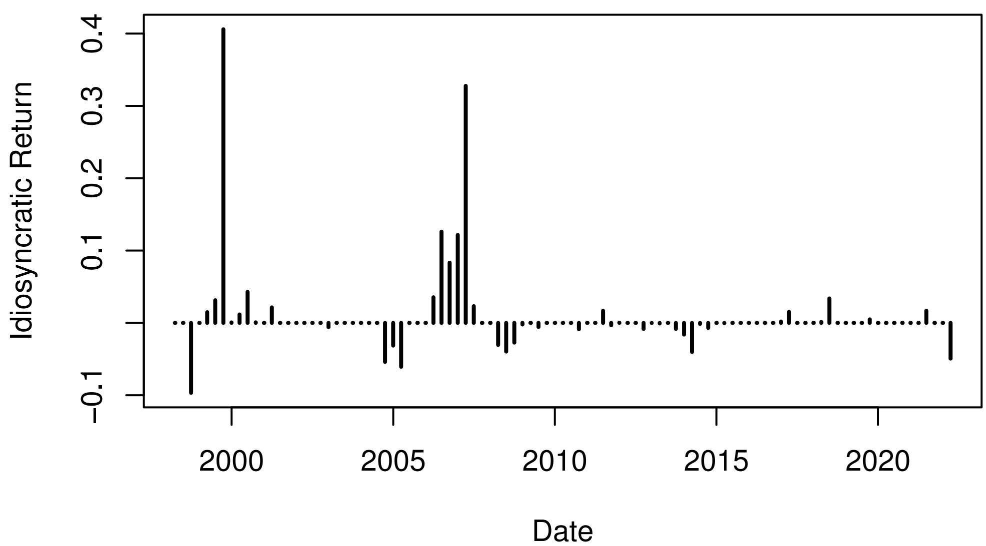


Figure 3: Idiosyncratic returns estimated by componentwise L_2 boosting for fund type BO in the period from 1998-03-31 until 2022-03-31.

For illustration and comparison with other BO return series, we map the monthly returns to quarterly returns in Figures 3 and 4 and analyze the period from 1998-03-31 until 2022-03-31. Figure 4 nicely depicts that adding the error term estimates to our public factor model closely aligns the Average 2-Factor Models + Error series to the NAV returns of Cambridge Associates (CA), Preqin and Pitchbook.[5] The Pearson correlation coefficients between our Public Factors + Error series and the

[5]Because of missing data, we fill the first quarters of the Pitchbook series with CA returns (from 1998-03-31 until 1999-09-30) and the first quarters of the Preqin series with CA returns (from 1998-03-31 until 2007-12-31).

CA and Pitchbook series are 79% and 72%, respectively. The quarterly average return is relatively high in all three series with 4.7% (Public Factors + Error), 3.6% (CA), and 3.5% (Pitchbook). However, the standard deviation is considerably larger in our Public Factors + Error series with 15.4% compared to 5.7% and 5.1% observed for CA and Pitchbook NAV returns, respectively. For comparison, the MSCI Market [Public Factors] return exhibits a 2.7% [3.8%] average quarterly return with a standard deviation of 8.8% [13.9%]. Interestingly in Figure 4, the public factor model returns of the average 2-factor model and the 5-factor model exhibit a very similar return pattern that slightly outperforms the MSCI Market return. Adding our sparse error term vector e to the factor model generates a new return time series that is closer aligned with both NAV return time series. However, as can be seen in Figure 6, the observed autocorrelation function values for lags 1 and 2 are considerably smaller in our Average 2-Factor Models + Error series with 12.4% and 12.0% than in the CA (35.6% and 28.7%), Preqin (44.2% and 25.2%) and Pitchbook (41.3% and 32.7%) cases. Our boosting procedure thus can generate return proxies that do not suffer from the smoothing and staleness problems associated with NAV returns.

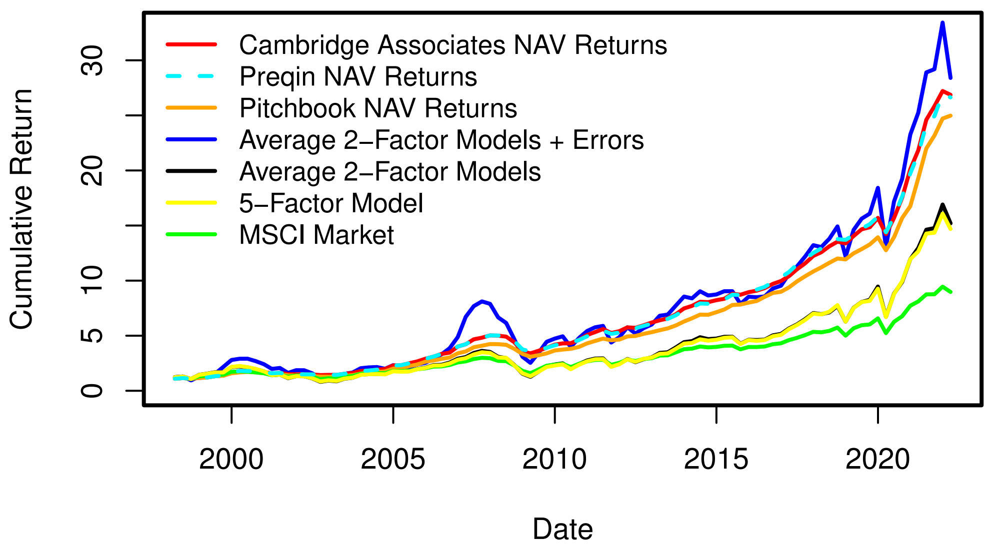
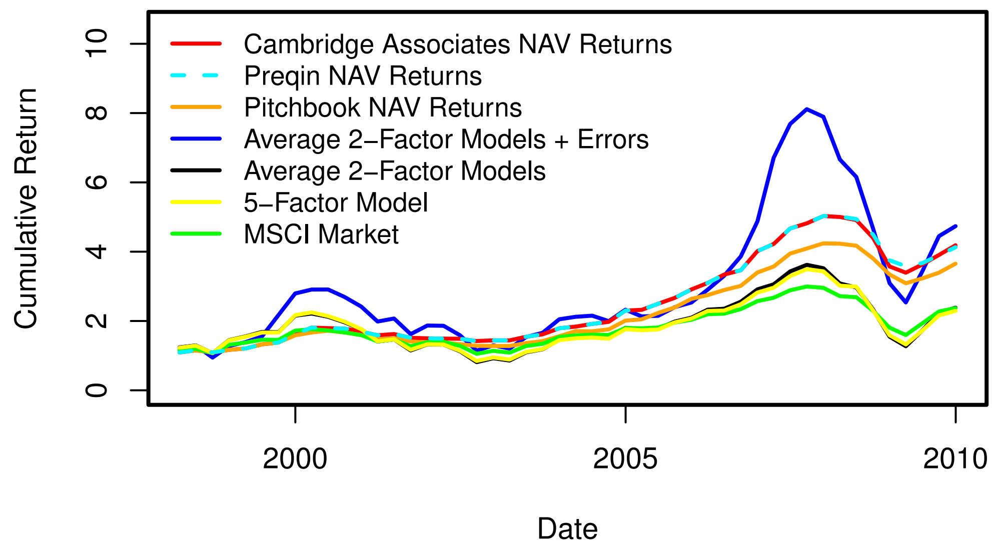


Figure 4: Comparison between the total returns for fund type BO implied by our two-factor ensemble and our two-factor ensemble plus the error term from Figure 3. Both series are contrasted against the NAV Return indices provided by Cambridge Associates, Preqin and Pitchbook and the MSCI stock market index in the period 1998-03-31 until 2022-03-31.



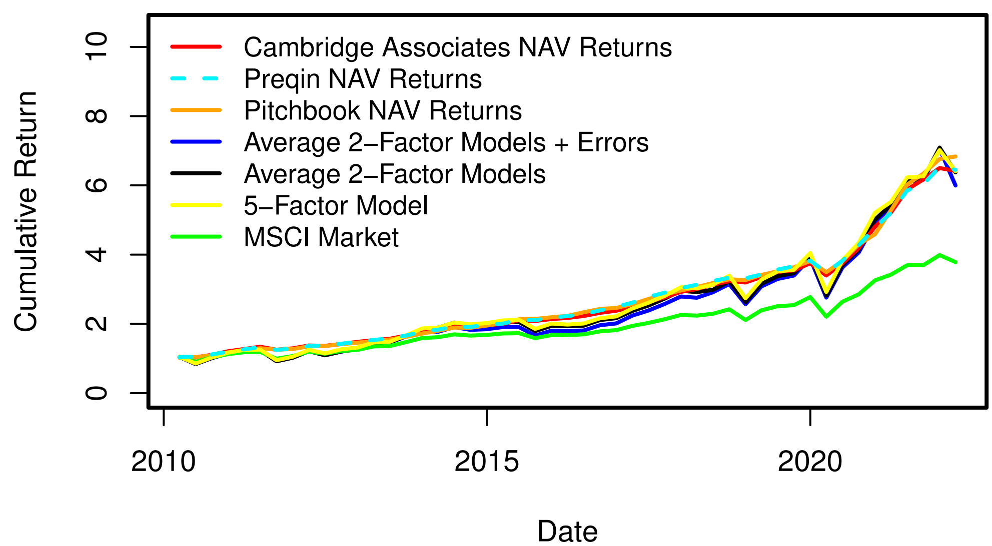


Figure 5: In these two subplots, we split the full time series from Figure 4 into a pre-2010 and post-2010 period.

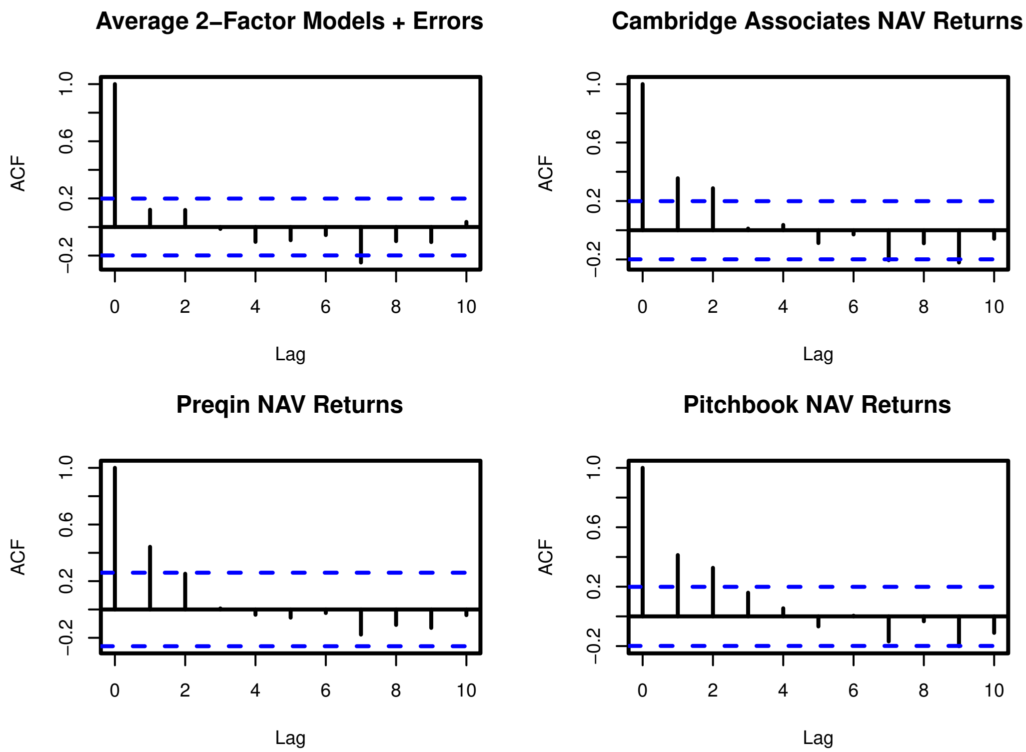


Figure 6: Comparison between the autocorrelation functions of our total return time series and the NAV-return time series of Cambridge Associates, Preqin and Pitchbook.

In Figure 5, we split Figure 4 in a pre-2010 and a post-2010 period. We observe a big difference between the Average 2-Factor Models and the Average 2-Factor Models + Errors series in the period before 2010. However after 2010, all public factor model returns (with and without errors) seem closely aligned with the CA and Pitchbook NAV return indices. Certainly, this result does not come by surprise as Figure 3 displays all larger error terms before 2010. From these analyses it remains unclear if the time-consuming error term estimation is always necessary since the average of the two-factor models alone can adequately price BO cash flows in the more recent past. In appendix **??**, we receive a similar result for VC returns.

Finally, table 3 reports the summary statistics for the error terms to address the question if Buyout (and other fund types) outperformed the public market on a risk-adjusted basis. The extremely high values for skewness and kurtosis challenge the validity of using the mean of the error term series as measure for the expected outperformance, i.e., $\alpha = E[e]$. Supposedly, the median gives a more realistic approximation for practical purposes since asymptotic behavior cannot be expected for time series of length 100.

Type	Mean	Stdv.	Skew.	Kurt.	Percentage filled
BO	1.12 %	6.44 %	432 %	2081 %	44 % = 45 / 102
VC	1.33 %	12.36 %	553 %	3827 %	62 % = 61 / 99
RE	1.06 %	16.84 %	-9 %	1799 %	51 % = 50 / 98
INF	0.91 %	11.15 %	404 %	2579 %	36 % = 34 / 94
NATRES	5.77 %	41.31 %	794 %	6887 %	36 % = 36 / 99
PD	1.05 %	5.70 %	420 %	2166 %	44 % = 34 / 77

Table 3: Quarterly error term return summary for various fund segments. The percentage filled column gives us the number of quarters with non-zero error terms as determined by the sparse componentwise L_2 boosting algorithm; the remaining unfilled quarters show zero idiosyncratic returns. Thus, the percentage filled column is a measure of sparseness for our CLB method. The median error term is 0 for all reported fund segments due to this sparseness.

6 Conclusion

In summary, this paper introduces a straightforward ensemble method designed to translate private equity cash flows into a public factor model and an error term time series. The versatility of this method is evident in its applicability, serving two primary purposes: (i) exploring the public market factor exposures of various private equity fund types or strategies and (ii) analyzing specific private equity portfolios. By enabling the use of factor loadings and error terms in integrated risk management, this approach establishes a standardized foundation for comparing the risk exposure of public and private asset portfolios.

This innovative analysis represents a valuable addition to traditional methods for characterizing the style and risk of private equity investments, providing investors with additional insights. Our error term analysis, particularly for Buyout funds, reveals a temporal shift in the significance of the error term in describing total returns. Before 2010, the error term played a crucial role, whereas after 2010, the public factor model alone proved sufficient to price Buyout cash flows. Moreover, our error term estimates exhibit too extreme higher moments to be adequately described by lower order summary statistics like the mean value alone. This, in turn, casts doubts on the usefulness of the exptected outperformance term $\alpha = E[e]$ in the context of private capital returns. Notably, our empirical application yields total return time series for private equity segments with significantly lower autocorrelations than traditional NAV-return time series. However, our methodology for the error term estimation seems to sometimes yield unreliable results at the beginning and end of the error term time series when relatively small cash flows need to be priced by our SDF.

Looking forward, there are promising avenues for refining our approach. Firstly, enhancing the (i) simple ensemble approach for the public factor model and (ii) brute-force componentwise L_2 boosting for the error term can be achieved through more efficient machine learning algorithms (Bühlmann, 2012). Additionally, integrating cross-validation techniques can enhance the reliability of identifying valid

SDF models compared to our current quantile-based procedure, as briefly mentioned in this article. Finally, to validate the robustness of our findings, further studies should replicate our results using alternative private and public data sources beyond Pitchbook and MSCI data. These potential enhancements and validations are essential steps toward strengthening the credibility and generalizability of our proposed methodology.

References

Ang, A., Chen, B., Goetzmann, W. N., and Phalippou, L. (2018). Estimating private equity returns from limited partner cash flows. *Journal of Finance*, 73(4):1751–1783.

Arlot, S. and Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics surveys*, 4:40–79.

Bachoc, F., Leeb, H., and Pötscher, B. M. (2019). Valid confidence intervals for post-model-selection predictors. *The Annals of Statistics*, 47(3):1475–1504.

Bai, J. and Ng, S. (2009). Boosting diffusion indices. *Journal of Applied Econometrics*, 24(4):607–629.

Bühlmann, P. (2006). Boosting for high-dimensional linear models. *Annals of Statistics*, 34(2):559–583.

Bühlmann, P. (2012). Bagging, boosting and ensemble methods. In *Handbook of computational statistics*, pages 985–1022. Springer.

Bühlmann, P. and Van De Geer, S. (2011). *Statistics for high-dimensional data: methods, theory and applications.* Springer Science & Business Media.

Claeskens, G., Magnus, J. R., Vasnev, A. L., and Wang, W. (2016). The forecast combination puzzle: A simple theoretical explanation. *International Journal of Forecasting*, 32(3):754–762.

Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of Finance*, 66(4):1047–1108.

Coqueret, G. and Guida, T. (2023). *Machine Learning for Factor Investing: Python Version.* CRC Press.

Dimson, E. (1979). Risk measurement when shares are subject to infrequent trading. *Journal of financial economics*, 7(2):197–226.

Driessen, J., Lin, T.-C., and Phalippou, L. (2012). A new method to estimate risk and return of nontraded assets from cash flows: the case of private equity. *Journal of Financial and Quantitative Analysis*, 47(3):511–535.

Feng, G., Giglio, S., and Xiu, D. (2020). Taming the factor zoo: A test of new factors. *The Journal of Finance*, 75(3):1327–1370.

Ferson, W. (2019). *Empirical Asset Pricing: Models and Methods*. MIT Press, 1 edition.

Gredil, O. R., Griffiths, B., and Stucke, R. (2023). Benchmarking private equity: The direct alpha method. *Journal of Corporate Finance*, 81:102360.

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273.

Hansen, B. E. (2014). Model averaging, asymptotic risk, and regressor groups. *Quantitative Economics*, 5(3):495–530.

Hansen, L. P. and Richard, S. F. (1987). The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. *Econometrica*, pages 587–613.

Hansen, P. R., Lunde, A., and Nason, J. M. (2011). The model confidence set. *Econometrica*, 79(2):453–497.

Huang, H. and Lee, T.-H. (2010). To combine forecasts or to combine information? *Econometric Reviews*, 29(5-6):534–570.

Hüther, N., Schmid, L., and Steri, R. (2023). Credit market equivalents and the valuation of private firms.

Karolyi, G. A. and Van Nieuwerburgh, S. (2020). New methods for the cross-section of returns. *The Review of Financial Studies*, 33(5):1879–1890.

Korteweg, A. and Nagel, S. (2016). Risk-adjusting the returns to venture capital. *Journal of Finance*, 71(3):1437–1470.

L'Her, J.-F., Stoyanova, R., Shaw, K., Scott, W., and Lai, C. (2016). A bottom-up approach to the risk-adjusted performance of the buyout fund market. *Financial Analysts Journal*, 72(4):36–48.

Madhavan, A. and Sobczyk, A. (2019). Fine-tuning private equity replication using textual analysis. *Journal of Financial Data Science*, 1(1):111–121.

Mittnik, S., Robinzonov, N., and Spindler, M. (2015). Stock market volatility: Identifying major drivers and the nature of their impact. *Journal of banking & Finance*, 58:1–14.

Moral-Benito, E. (2015). Model averaging in economics: An overview. *Journal of Economic Surveys*, 29(1):46–75.

O'Doherty, M. S., Savin, N. E., and Tiwari, A. (2017). Hedge fund replication: A model combination approach. *Review of Finance*, 21(4):1767–1804.

Phalippou, L. (2014). Performance of buyout funds revisited? *Review of Finance*, 18(1):189–218.

Phalippou, L. and Gottschalg, O. (2009). The performance of private equity funds. *The Review of Financial Studies*, 22(4):1747–1776.

Porter, S. C. and Porter, S. (2019). Introducing objective benchmark-based attribution in private equity. *Journal of Financial Data Science*, 1(1):130–140.

Qian, W., Rolling, C. A., Cheng, G., and Yang, Y. (2019). On the forecast combination puzzle. *Econometrics*, 7(3):1–26.

Rapach, D. E., Strauss, J. K., and Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *The Review of Financial Studies*, 23(2):821–862.

Schapire, R. E. (1990). The strength of weak learnability. *Machine learning*, 5(2):197–227.

Smith, J. and Wallis, K. F. (2009). A simple explanation of the forecast combination puzzle. *Oxford Bulletin of Economics and Statistics*, 71(3):331–355.

Stafford, E. (2017). Replicating private equity with value investing, homemade leverage, and hold-to-maturity accounting. *Working Paper (May 20, 2017)*.

Tancar, R. and Viebig, J. (2008). Alternative beta appliedan introduction to hedge fund replication. *Financial markets and portfolio management*, 22(3):259.

Tausch, C. (2019). Quadratic hedging strategies for private equity fund payment streams. *The Journal of Finance and Data Science*, 5(3):127–139.

Weisang, G. (2014). Factor selection in dynamic hedge fund replication models: A bayesian approach. In *Bayesian Model Comparison*. Emerald Group Publishing Limited.

White, H. (2000). A reality check for data snooping. *Econometrica*, 68(5):1097–1126.

A Comparison to competing approaches

In this section, we compare our two-step methodology to potentially simpler competitor models highlighting advantages as well as disadvantages associated with each model. In this context, it is paramount to explain which general intricacies arise when researchers want to estimate a regression model $R = \beta X + e$ with private equity returns R as dependent variable and a linear factor model βX plus error term e on the right-hand side.

The first fundamental challenge is accurately measuring private equity returns R, as fund NAVs are impacted by intermediate cash flows and stale pricing, both of which are difficult to avoid for closed-end funds operating in illiquid markets. The second challenge is estimating a factor model βX when only incomplete return time series or cash flow data are available. The third problem is to derive the error term e that comes with a factor model βX given the same fundamental problem.

A.1 Return measurment

The most straightforward alternative to our model would be a "precise" measurement of "true" PE fund returns. Given well-behaved PE return time-series, factor model estimation and error term calculation can be simply done by a linear Ordinary Least Squares (OLS) regression. However, in practice true PE returns can be only approximated by so-called Horizon Internal Rate of Returns (Horizon IRRs) or NAV returns (or variants thereof).

Horizon IRR is a measure used to evaluate the performance of private equity investments over a specific time period, accounting for both interim cash flows and changes in NAV. Unlike traditional IRR, which focuses on the full lifetime cash flows of an investment, Horizon IRR assesses returns within a given investment horizon, typically over shorter periods like one or several years. This metric is often used for benchmarking and performance comparison across funds during specific periods.

$$\text{IRR}_{horizon} := \arg\min_{r \in \mathbb{R}} \left| NAV_{t_1} - \sum_{t=t_1}^{t_2} \frac{CF_t}{(1+r)^{t-t_1}} - \frac{NAV_{t_2}}{(1+r)^{t_2-t_1}} \right| \stackrel{!}{=} 0$$

The connection between Horizon IRR and NAV returns lies in their shared focus on interim valuations. NAV returns capture the percentage change in a private equity fund's net asset value, adjusted for distributions and contributions, and are often calculated on a quarterly or annual basis.

$$R^{\text{NAV}}_{t_1,t_2} := \frac{NAV_{t_2} - \sum_{t=t_1}^{t_2} CF_t}{NAV_{t_1}}$$

Horizon IRR incorporates NAV returns by considering both the periodic changes in NAV and cash flows (capital calls and distributions) during the investment horizon, providing a more time-weighted performance measure that accounts for both realized and unrealized gains. Unfortunately, the inherent stale pricing of PE funds' NAVs yields autocorrelated NAV returns as previously exemplified by Figure 6. Thus, NAV returns and horizon IRRs can only be considered as proxy returns.

A.2 Factor-model estimation

The Dimson (1979) beta approach – initially developed for public shares that are subject to infrequent trading – is particularly useful for estimating factor loadings in private equity returns, where stale pricing (of NAV appraisals) can lead to asynchronous movements with market factors. Traditional beta estimation assumes contemporaneous correlation between asset returns and factor returns, which may not hold for illiquid, non-traded assets like private equity. The Dimson (1979) beta corrects for this by incorporating lagged factor returns in the OLS regression, capturing delayed or smoothed responses to market-wide movements. This method potentially improves the accuracy of factor loading estimates, reflecting a more realistic sensitivity of private equity returns to systematic risk factors. As dependent variable for a Dimson (1979) regression, we could use NAV returns or Horizon IRRs.

$$R_t^{\text{Dimson}} := \alpha + \sum_{l=0}^{L} \sum_{j} \beta_{j,l} F_{j,t-l} + e_t$$

Unfortunately, this method is not helpful to determine the "true" total PE return as the error terms in the formula above are directly derived from the stale returns that serve as dependent variable in the regression.

Driessen et al. (2012), Korteweg and Nagel (2016), Ang et al. (2018) propose factor-models estimators that use cash flows instead for return time series to completely avoid the (fundamental) stale pricing problem associated with NAVs.

A.3 Error-term derivation

The advanced Ang et al. (2018) ansatz aims at creating a Markov Chain Monte Carlo (MCMC) which equilibrium distribution matches the distribution of latent private equity returns denoted by g_t^e. While estimating the MCMC they further decompose the PE returns by a linear multi-factor model $g_t = \alpha + \beta^{'} F_t + f_t + r_t^{rf}$ where r_t^{rf} denotes the risk-free rate and f_t is perceived as asset-class specific latent factor (i.e., the error term) with mean zero that is orthogonal to the traded factors, F_t. The corresponding factor loadings β need to be subsequently updated in each MCMC iteration after a new candidate for g_t has been drawn via "a standard regression draw" (Ang et al., 2018, internet appendix, p.4). In the final step, each MCMC iteration samples a nuisance parameter, assumed to follow a normal distribution.

In summary, their approach first samples total PE returns by normal draws around the factor-model mean for each date and then – as a second step – updates the public factor model via a multi-variate Least Absolute Deviations (LAD) regression on this total return series. In contrast, we first combine multiple, potentially simpler (uni- or bi-variate) factor models by model averaging and then – in the second step – directly estimate the time-series of idiosyncratic returns via a straightforward but brute-force algorithm. A minor drawback of the Ang et al. (2018) framework is its inherent complexity, which, along with its Bayesian nature, offers researchers considerable flexibility in selecting priors and making subtle design decisions. In other words, several variations of the Ang et al. (2018) algorithm appear reasonable and worth investigating.