Semiparametric SDF Estimators for Pooled, Non-Traded Cash Flows

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

Alexander Bohnert
Munich University of Applied Sciences
Am Stadtpark 20, D-81243 Munich
alexander.bohnert@hm.edu

January 20, 2026

Keywords

Stochastic Discount Factor, Semiparametric Estimation, Spatial Dependence, Generalized Method of Moments (GMM), Private Equity, Illiquid Assets

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.

Semiparametric SDF Estimators for Pooled, Non-Traded Cash Flows

Abstract

This paper analyzes stochastic discount factor estimation methodologies suited for pooled, non-traded cash flow streams such as the fund-level cash flows of private equity funds. The asymptotic inference framework for our semiparametric nonlinear least squares estimator draws on a spatial notion, i.e., the idea that the economic distance between distinct private equity funds can be measured. The empirical and Monte Carlo simulation results indicate (i) that our method can improve the popular Generalized Method of Moments approach of Driessen et al. (2012), but (ii) that estimator variance for typical data sizes is still high. Thus, we conjecture that traditional semiparametric extremum estimators like the one described by us shall be exclusively used for single-factor models until considerably more vintage year information for private equity funds is available.

1 Introduction

Private equity has outgrown its niche, sitting today on more than $9 trillion in assets under management, yet rigorous asset-pricing tools have not kept pace with this ascent. The empirical analysis and risk assessment of private equity and other non-traded cash flows remain fundamentally challenging due to the absence of market-based valuations and the inherent frictions of private markets (i.e., under incomplete information). Unlike public assets with trusted and tradeable valuations (in liquid secondary markets), private equity investments generate irregular, infrequently observed cash flows for which standard return-based asset pricing techniques are unsuitable.

We address this gap by proposing a semiparametric stochastic discount factor (SDF) estimator tailored to fund-level cash flows that refines the SDF estimators of Driessen et al. (2012) and Korteweg and Nagel (2016). Our nonlinear least squares estimator stems from the class of Least-Mean-Distance (LMD) estimators, which are easier to handle than traditional Generalized Method of Moments (GMM) approaches (Pötscher and Prucha, 1997). Our LMD estimator arguably both simplifies and generalizes the GMM methodology of Driessen et al. (2012), where we provide the asymptotic inference framework that was missing in the original paper. The asymptotic inference formulations rely on the concept of spatial (near-epoch) dependence between funds as proposed by Korteweg and Nagel (2016). In this context, it is paramount to quantify the economic distance between funds by a measure like absolute vintage year difference or cash flow overlap[1].‘

For fund-level cash flow data of private equity funds, we document an asymptotic bias term for cash-flow-based SDF estimators like Driessen et al. (2012) and Korteweg and Nagel (2016) that persists also in large samples. The bias term arises due to the pooled nature of fund-level cash flows and more specifically because of

[1]However, this economic inter-fund *distance* refers **not** to the term Least-Mean-*Distance* estimator.

the different starting dates of the underlying deals (in the fund investment period). Our estimator offers simple averaging over multiple discounting dates as one option to control (but not eliminate) the bias term.

In the empirical application of our new estimator, we test simple linear and exponentially affine SDF models that can draw on the five return factors associated with the q^5 investment factor model recently proposed by Hou et al. (2020). Based on a Spatial Heteroskedasticity and Autocorrelation Consistent (SHAC) covariance matrix estimator, we calculate asymptotic standard errors for the model coefficients. Moreover, we assess the small-sample variance of coefficient estimates and the out-of-sample performance of the different SDF models by hv-block cross-validation, which accounts for the inter-vintage-year dependence of private equity funds (Racine, 2000). We test one- and two-factor models for the most prominent private capital fund types Private Equity (PE) and Venture Capital (VC). The empirical two-factor model results are rather disappointing; not more than the single-market-factor model results seem reasonable given the high estimator variance.

The paper is structured as follows. Section 2 introduces our semiparametric LMD estimator and its corresponding asymptotic inference framework. Section 3 applies the method to estimate q^5-investment-factor SDFs for various private equity fund types using simulated and real-world cash flows. Section 4 concludes.

2 Methodology

Our general SDF estimation framework is similar to that of Driessen et al. (2012) and Korteweg and Nagel (2016); the subtle but important differences are discussed in Section 2.7.

In a nutshell, the Driessen et al. (2012) estimator determines the optimal alpha and beta parameters of a linear factor model by minimizing the net present value (NPV) of all N fund cash flows.

$$\min_{\alpha,\beta} = \sum_{i=1}^{N} [NPV_i(\alpha,\beta)]^2 \tag{1}$$

with NPV for the ith fund

$$NPV_i(\alpha,\beta) = \sum_{t\in\{t_{0,i},t_{0,i}+1,\dots\}} \frac{\text{CashFlow}_{t,i}}{\prod_{s=t_{0,i}}^{t}(1+r_{\text{free},s}+\alpha+\beta r_{\text{market},s})}$$

In this Methodology section, we will show:

1. Why this estimator is usually asymptotically biased for pooled cash flows (Section 2.1).
2. How to modify the estimator to be able to control the bias term (Sections 2.2 and 2.3).
3. How we can estimate asymptotic standard errors for the parameters alpha and beta (Sections 2.5 and 2.6).

2.1 Asset Pricing for Pooled Cash Flows

Let fund $i = 1, 2, \ldots, n$ be characterized by its net cash flows $CF_{t,i}$ (i.e., distributions minus contributions) and its net asset values $NAV_{t,i}$ with discrete time index $t = 0, 1, 2, \ldots, T$. To increase the mathematical tractability of the problem, we assume a deal-by-deal data generating processes (DGP) for CF where each fund deal consists exactly of one investment and one divestment cash flow in combination with a simple return model for the multi-period deal returns. This means the fund-level cash flow process $(CF_{i,t})_{t=0,1,\ldots,T}$ is an aggregation of deal-level cash flow pairs consisting of one negative at deal inception and at least one positive cash flow later $CF_{t,i} = \sum_j^J cf_{j,i,t}$.

Assumption 1. *Deal-level data generation process:*
1. Each fund i consists of J underlying deals.
2. Each deal is characterized by exactly one, negative investment cash flow, denoted by $\mathrm{Inv}_{i,j}$, which occurs at time $t^{\mathrm{Inv}}_{i,j} \in \{0, 1, \ldots, T-1\}$. It holds $\mathrm{Inv}_{i,j} < 0$.
3. Each deal is characterized by a positive divestment cash flow stream, denoted by $(\mathrm{Div}_{i,j,k})_{k=1,\ldots,K}$, which occur at after the investment cash flow $t^{\mathrm{Div}}_{i,j,k} > t^{\mathrm{Inv}}_{i,j}$ for all k. It holds $\mathrm{Div}_{i,j,k} > 0$.
4. The cumulative fund cash flows are generated by summarizing over all deal-level cash flows, i.e., $\sum_{t=0}^T CF_{i,t} = \sum_{j=1}^J \left(\mathrm{Inv}_{i,j} + \sum_k^K \mathrm{Div}_{i,j,k}\right)$ for all i.

From asset pricing theory, we know that we can use a stochastic discount factor Ψ_t to price each underlying deal

$$\mathbb{E}\left[\delta_{i,j} \,\middle|\, \mathcal{F}_{t^{\mathrm{Inv}}_{i,j}}\right] = 0 \quad \forall\ i, j \tag{2}$$

where we denote the deal-level pricing error by

$$\delta_{i,j} := \mathrm{Inv}_{i,j} + \sum_k^K \frac{\Psi_{t^{\mathrm{Div}}_{i,j,k}}}{\Psi_{t^{\mathrm{Inv}}_{i,j}}} \mathrm{Div}_{i,j,k} \tag{3}$$

For the pooled, fund-level cash flow stream, we assume that the true fund valuation $V_{i,\tau}$ is not observable for us

$$V_{i,\tau} := \mathbb{E}\left[\sum_{t=\tau}^T \frac{\Psi_t}{\Psi_\tau} CF_{i,t} \,|\mathcal{F}_\tau\right] \quad \forall\ \tau \geq \min_j \mathrm{Inv}_{i,j} \tag{4}$$

to acknowledge the stale-pricing problem inherent to private capital fund net asset values (NAVs). In other words, we only trust fund cash flows but not fund NAVs in private markets[2].

In this realistic setting, the absence of observable market valuations considerably contributes to the difficulty of our pricing problem. However, also the existence of only pooled cash flows, instead of granular deal-by-deal cash flows, introduces issues for pricing approaches like Driessen et al. (2012) that discount to the fund inception

[2] In the empirical section, we empirically treat the most recent NAV as the final distribution cash flow for non-liquidated funds.

date (see Equation **??**). In the following, we will demonstrate why we cannot easily price pooled fund-level cash flow streams without introducing **inevitable bias terms** (even for fund inception date $\tau = \min_j t^{\text{Inv}}_{i,j}$). Generally, all deal-level cash flows will produce the following "bias term"

$$\mathbb{E}\left[\frac{\Psi_{t^{\text{Inv}}_{i,j}}}{\Psi_\tau}\delta_{i,j}\,|\mathcal{F}_\tau\right] = \text{Cov}\left[\frac{\Psi_{t^{\text{Inv}}_{i,j}}}{\Psi_\tau},\ \delta_{i,j}\right] \quad \forall\ \tau < t^{\text{Inv}}_{i,j} \tag{5}$$

Only for the trivial case of Equation 2, where the investment date coincides with the discounting date $\tau = t^{\text{Inv}}_{i,j}$, the covariance term necessarily equals zero.

For the fund-level cash flows, we therefore introduce the following proposition.

Proposition 1. *Price of a pooled cash flow stream at fund inception:*

$$\mathbb{E}\left[\sum_{t=\tau}^{T}\frac{\Psi_t}{\Psi_\tau}CF_{i,t}\,|\mathcal{F}_\tau\right] = \sum_{j=1}^{J}\text{Cov}\left[\frac{\Psi_{t^{\text{Inv}}_{i,j}}}{\Psi_\tau},\ \delta_{i,j}\right]\ \forall\ i \tag{6}$$

with fund inception date $\tau =\ \min_j \text{Inv}_{i,j}$.

We can simply proof the proposition as follows.

Proof. We start with using Point 4 of Assumption 1 which states that fund-level cash flows are the sum of deal-level cash flows. Further, we stipulate in the proposition that no deal-level cash flow occurs before τ. Thus, the expected value of discounted fund-level cash flows needs to equal the expected value of discounted deal-level cash flows.

$$\mathbb{E}\left[\sum_{t=\tau}^{T}\frac{\Psi_t}{\Psi_\tau}CF_{i,t}\,|\mathcal{F}_\tau\right] = \mathbb{E}\left[\sum_{j=1}^{J}\frac{\Psi_{t^{\text{Inv}}_{i,j}}}{\Psi_\tau}\delta_{i,j}\,|\mathcal{F}_\tau\right] \tag{7}$$

Using Equation 5, we can rewrite

$$\mathbb{E}\left[\sum_{j=1}^{J}\frac{\Psi_{t^{\text{Inv}}_{i,j}}}{\Psi_\tau}\delta_{i,j}\,|\mathcal{F}_\tau\right] = \mathbb{E}\left[\sum_{j=1}^{J}\text{Cov}\left[\frac{\Psi_{t^{\text{Inv}}_{i,j}}}{\Psi_\tau},\ \delta_{i,j}\right]|\mathcal{F}_\tau\right] \tag{8}$$

Linearity of expectations then yields the result we want to proof.

$$\mathbb{E}\left[\sum_{j=1}^{J}\text{Cov}\left[\frac{\Psi_{t^{\text{Inv}}_{i,j}}}{\Psi_\tau},\ \delta_{i,j}\right]|\mathcal{F}_\tau\right] = \sum_{j=1}^{J}\text{Cov}\left[\frac{\Psi_{t^{\text{Inv}}_{i,j}}}{\Psi_\tau},\ \delta_{i,j}\right] \tag{9}$$

□

Corollary 1. *If and only if all deal investment dates coincide with the fund inception date, i.e.,* $\text{Inv}_{i,j} = \min_j \text{Inv}_{i,j} \forall j$*, we have*

$$\mathbb{E}\left[\sum_{t=\tau}^{T}\frac{\Psi_t}{\Psi_\tau}CF_{i,t}\,|\mathcal{F}_\tau\right] = 0 \tag{10}$$

for the standard case where we do not assume independence between Ψ *and* $\delta_{i,j}$*.*

The asset pricing problem for realistic private equity fund cash flows is aggravated by bridge financing, management fees and carry cash flows that further distort the deal-level cash flows.

2.2 Least-Mean-Distance estimator

In this subsection, we introduce a new SDF estimator designed to analyzed the effect of different discounting dates τ on the bias and variance of the estimated SDF parameters. In the previous subsection, we demonstrated that for a pooled cash flow stream, consisting of at least two deals with different investment start dates, **no correct discounting date exists**. Thus our general idea is to rather average over multiple suitable discounting date candidates τ than to try to select only one candidate for the "best" discounting date (which is empirically unknown).

Henceforth, we assume that the underlying transactions within a private equity fund cannot be distinguished individually, and that only the funds total (pooled) cash flows are observable. The stochastic discount factor $\Psi_{\tau,t}$ is used to calculate the time-τ "price" $P_{\tau,t,i}$ of a **single** time-t cash flow of any given PE fund i

$$P_{\tau,t,i} := \Psi_{\tau,t} \cdot CF_{t,i} = \frac{\Psi_t}{\Psi_\tau} \cdot CF_{t,i} \qquad \forall \quad \tau, t, i \tag{11}$$

with multi-period SDF $\Psi_t = \prod_{k=0}^{t} M_k$. As SDFs are commonly parameterized by a vector $\theta \in \mathbb{R}^p$, i.e., $\Psi_{t,\tau} \equiv \Psi_{t,\tau}(\theta)$, our goal is to find an estimation method for the optimal θ. We denote this optimal/best/true parameter vector as θ_0. We call the numerator Ψ_t the discount part of the multi-period SDF $\Psi_{\tau,t}$ (used for present value calculations) and the denominator Ψ_τ the compound part (used for future value calculations). For each fund i and all points τ within a common fund lifetime, the empirical "pricing error" $\epsilon_{\tau,i}$ of **all** fund cash flows is calculated as time-τ "present value"

$$\epsilon_{\tau,i} := \sum_{t=1}^{T} P_{\tau,t,i} \qquad \forall \quad \tau, i \tag{12}$$

We use the terms "price", "pricing error'" and "net present value" in quotation marks to acknowledge the theoretical asset pricing problem which can arise for pooled cash flows and has been described in the previous subsection.

To better analyze the impact of different discount date τ on the estimator's bias and variance, we define the (w_i-weighted) average pricing error $\bar{\epsilon}_i$ that averages over the set $\mathcal{T}_i$

$$\bar{\epsilon}_i = w_i \cdot \frac{1}{\text{card}(\mathcal{T}_i)} \sum_{\tau \in \mathcal{T}_i} \epsilon_{\tau,i} \qquad \forall \quad i \tag{13}$$

where $\mathcal{T}_i$ gives the set of discounting dates τ for fund i which is more thoroughly described in the next Subsection 2.3. Additionally, each fund i is characterized by its vintage year which can be expressed by $v_i = \min(\mathcal{T}_i) \in \{1, 2, \ldots, V\}$, where V denotes the maximum vintage year used in a given data set. Finally, the scalar weighting factor w_i can be (i) one divided by the fund's invested capital for equal weighting of funds, (ii) one divided by the vintage year sum of invested capital for vintage year weighting, (iii) the scalar one for fund-size weighting, or (iv) some macroeconomic deflator.

To find the optimal value for θ, we select an estimator from the broad class of extremum estimators.

Definition 1. *Extremum estimator (Newey and McFadden, 1994, Equation 1.1): An estimator $\hat{\theta}$ is an extremum estimator if there is an objective function $Q_n(\theta)$ such that*

$$\hat{\theta} = \max_{\theta} \; Q_n(\theta)$$

for $\theta \in \Theta$ where Θ is the set of all possible parameter values.

Concretely, our Least Mean Distance (LMD) estimator (Pötscher and Prucha, 1997, Equation 7.1) minimizes the average loss of $\bar{\epsilon}$

$$\hat{\theta} = \arg\min_{\theta\in\Theta} \; Q_n(\theta) \quad \text{with} \quad Q_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} L\left(\bar{\epsilon}_i\right) \tag{14}$$

where $L : \mathbb{R}^1 \to \mathbb{R}^1$ denotes a loss function, e.g., $L(x) = (x-0)^2$ in the case of nonlinear least squares. Throughout the paper, the weighted average fund pricing error $\bar{\epsilon}_i \equiv \bar{\epsilon}_i(\theta)$ is regarded as nonlinear random function of the SDF parameter θ.

2.3 Future and present value dates: the set $\mathcal{T}$

This subsection helps to explain the importance of the set $\mathcal{T}$ from Equation 13. Initially, we introduced averaging over $\mathcal{T}$ to counter the "exploding alpha" issue, first described by Driessen et al. (2012), for cash flow streams with a very small initial cash flow. The exploding alpha problem depicts the mathematical fact that in a net present value formula, a discount factor with a very large alpha term discounts all cash flows (after the first one) close to zero. Thus, in this degenerate situation, the beta factors become irrelevant – an infinite alpha almost perfectly prices the cash flow stream. Even more importantly, our simulation study in Section 3.3 indicates that averaging over $\mathcal{T}$ seems to decrease the asymptotic bias of the estimator empirically.

A discounting date $\tau \in \mathcal{T}_i$ is a discretionary point in time where all fund cash flows are discounted to. The cardinality $\text{card}(\mathcal{T}_i) = |\mathcal{T}_i|$ gives the number of discounting dates used for the ith fund. The smallest possible set $\mathcal{T}_i$ contains just the fund's starting date; in this case, $\text{card}(\mathcal{T}_i)$ consequently is one. This corresponds to a typical NPV calculation in finance and is also used by Driessen et al. (2012) and Korteweg and Nagel (2016). In contrast, the largest candidate set for $\mathcal{T}$ contains all time periods bigger than the fund's starting date until now. In Subsection 3.3, we study the optimal set size of $\mathcal{T}$ by Monte Carlo simulations. There we show in our example that controlling for the optimal size of $\mathcal{T}$ can control the asymptotic bias and variance of the original Driessen et al. (2012) estimator that just discounts all cash flows to the fund inception date. As we average over $\mathcal{T}_i$ in Equation 13 we call $\bar{\epsilon}_i$ the $\mathcal{T}_i$-averaged pricing error, as visualized in Figure 1.

2.4 Cross-sectional unit: individual fund vs. portfolio of funds

According to the classical value-additivity assumption in Hansen and Richard (1987), SDF models invariably shall hold for all pooled or unpooled assets. As discussed

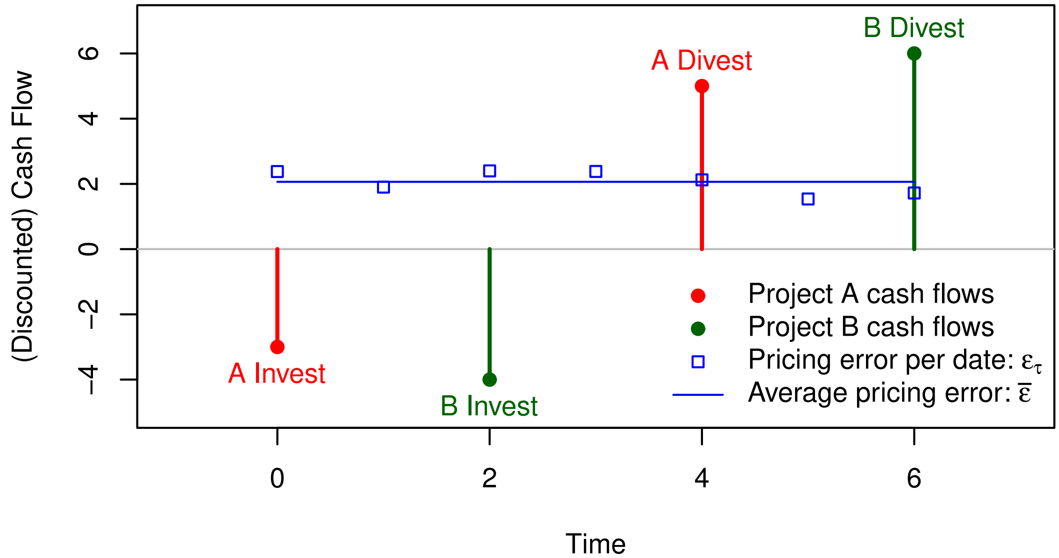

Figure 1: How to calculate and interpret the average pricing error? The time index t is relevant for the net cash flows (black dots). The time index τ is used for the pricing error, i.e., the sum of discounted net cash flows (blue boxes). The weighted average of these "pricing errors" gives the average pricing error $\bar{\epsilon}$ as defined in Equation 13 (solid blue line). In this example, card$(\mathcal{T}_i) = 7$, i.e., the number of blue boxes.

in Subsection 2.1, it is best to use the underlying deals as test assets for SDF estimation to avoid the bias terms caused by pooled fund cash flows. For the second best alternative, it is theoretically not important if the test assets for our SDF are portfolio or individual fund cash flows when the investment dates are the same (see Corollary 1). Empirically it makes a difference, and there are arguments both for and against portfolio formation.

In the risk premium literature, portfolio formation mainly helps to attenuate the errors-in-variables bias connected to two-pass asset pricing methods (Jegadeesh et al., 2019; Pukthuanthong et al., 2019). As this is no issue in our case, we could use individual funds. Cochrane (2011) argues that portfolio sorting (seen as an auxiliary nonparametric regression that imposes linearity on the relationship between returns and characteristics) shall be replaced by multivariate panel models due to the curse of dimensionality. Following the same nonparametric regression viewpoint, Cattaneo et al. (2019) derive a nonparametric framework where the optimal number of portfolio sorts acts as a data-dependent tuning parameter that grows with sample size. Generally, the larger the portfolios, the easier any given SDF can price their cash flows since fewer test assets remain.

In the case of private equity funds, the pooling of fund cash flows helps to counter GP financial engineering[3], which might both change and mask the true risk profile

[3]GPs may use bridge credit facilities below the hurdle rate to boost the fund's internal rate of return. This increases the probability of observing funds with only positive or only negative cash flows. However,

of observed LP cash flows. Especially for private equity funds, portfolio formation based on vintage years is compelling due to its time-series-like indexing as done by Driessen et al. (2012). This procedure also offers substantial computational benefits as it drastically decreases the number of cross-sectional units. Further, as stated in Ang et al. (2020), portfolio formation allows more precise factor loading estimates due to decreasing idiosyncratic risk, but at the expense of sacrificing cross-sectional information. Finally, small (or fixed) T and large N set-ups may face finite sample problems (Raponi et al., 2020).

Assumption 2. *For each vintage year, we pool fund cash flows to form n_v portfolios that serve as cross-sectional units. Thus, $n = \sum_{v=1}^{V} n_v$. The two boundary cases are (i) single fund portfolios and (ii) just one portfolio per vintage year.*

Without loss of generality, we refer to our cross-sectional units as funds, although this corresponds to a special case of our portfolio concept. In the simulation study in Subsection 3.3, we compare both boundary cases (i) individual funds and (ii) vintage year portfolios.

Thinking more broadly, we could even imagine more extreme boundary cases: (iii) on the one hand, we could pool *all* fund cash flows to form only *one* global moment condition for private equity similar to Korteweg and Nagel (2016) and accept potential under-identification; (iv) on the other hand, we could operate on underlying deal level like Buchner (2014, 2016a,b) and use gross-of-fee cash flows.

2.5 Asymptotic framework

To allow for multiple funds from the same vintage year in Assumption 2, we employ an auxiliary "spatial" notion as originally proposed by Korteweg and Nagel (2016). The spatial viewpoint is only a technical means to switch from time-series-like to more panel-data-like indexing. Unlike typical panel data, we do not follow multiple subjects over time, but for each point in time, we exclusively observe multiple new cross-sectional units (i.e., funds from that vintage year). This unusual two-dimensional indexing causes problems in the PE literature as it neatly fits neither in the (i) time-series, (ii) cross-sectional, nor (iii) panel data literature. Thus, we generally consider $\bar{\epsilon}$ from Equation 13 as random field (cf. Figure 2). In our case, it is convenient to interpret the fund vintage year v_i as second dimension in our pricing error random field, i.e., $\bar{\epsilon}_i \equiv \bar{\epsilon}_{i,v_i}$. ' Yet, in this section, we mainly follow the time-series asymptotic framework of Pötscher and Prucha (1997) since our "spatial" distance measure (between vintage years) is time, and adaption to our case is thus straightforward. If we observe only one fund per vintage year (or, equivalently, form vintage year portfolios), we will easily see that the framework of Pötscher and Prucha (1997) with time-series indexing can be applied without any major modification.

we want to avoid (the possibility of) cross-sectional units that exhibit just cash flows with the same algebraic sign. Realistic SDFs never can price these cash flow streams.

2.5.1 Vintage year asymptotics

We assume that the "spatial" (i.e., economic) distance between cross-sectional units, i.e., private equity funds/portfolios, can be measured quantitatively[4]. Our "cross-sectional" asymptotic theory lets the number of funds go to infinity $n \to \infty$. To expose our SDF to many distinct covariate realizations (economic conditions), we also want the number of vintages to increase asymptotically.

Assumption 3. *Vintage year asymptotics:*

1. *The number of vintage years $V \to \infty$ as $n \to \infty$.*
2. *The number of funds per vintage year is bounded by some positive constant.*
3. *The maximal fund lifetime is also bounded by a positive constant.*
4. *The economic distance between fund i and j is measured by the vintage year difference $d_{i,j} = |v_i - v_j| + \rho_0 1_{i \neq j}$ with minimum distance $\rho_0 > 0$.*

In terms of the spatial estimation literature, this assumption postulates increasing domain asymptotics and rules out so-called infill asymptotics (cf. Figure 2). The minimum distance term ρ_0 is a means to ensure these increasing domain asymptotics (Jenish and Prucha, 2012, Assumption 1). Infill asymptotics corresponds to the assumption of Driessen et al. (2012) that the number of funds per vintage tends to infinity.

GMM estimators typically have a fixed number of moment conditions. Thus, GMM estimators, where the number of moment conditions is allowed to grow with sample size, require special attention (Han and Phillips, 2006; Newey and Windmeijer, 2009). In many cases, it is probably most convenient to limit the maximum to a finite number of moment conditions (i.e., not each vintage year should form a moment condition). In this paper, we employ nonlinear least squares estimators since they do not suffer from this "number of moment condition" issue.

2.5.2 Law of large numbers

The global moment condition underlying our estimation approach is that the expected value of $\bar{\epsilon}$ shall be as close as possible to zero if we use the optimal SDF parameter θ_0 (it is nonzero due to Proposition 1). To approach this expected value, we rely on a spatial (cross-sectional) law of large numbers instead of applying a time-series law of large numbers. Here, we want to explicitly acknowledge the statistical dependence of pricing errors from adjacent vintage years.

Assumption 4. *Uniform Law of Large Numbers (ULLN) for random fields (Jenish and Prucha, 2009, Equation 6):*
The (i) time-trend and (ii) dependence structure of $\bar{\epsilon}$ shall allow

$$\sup_{\theta \in \Theta} |Q_n(\theta) - \mathbb{E}[Q_n(\theta)]| \xrightarrow{p.} 0 \quad as \quad n \to \infty$$

where $Q_n(\theta)$ is given by Equation 14.

[4] Generally, the economic distance measure could include multiple dimensions, e.g., temporal, geographic, and industry sector proximity. This could be an interesting topic for future research.

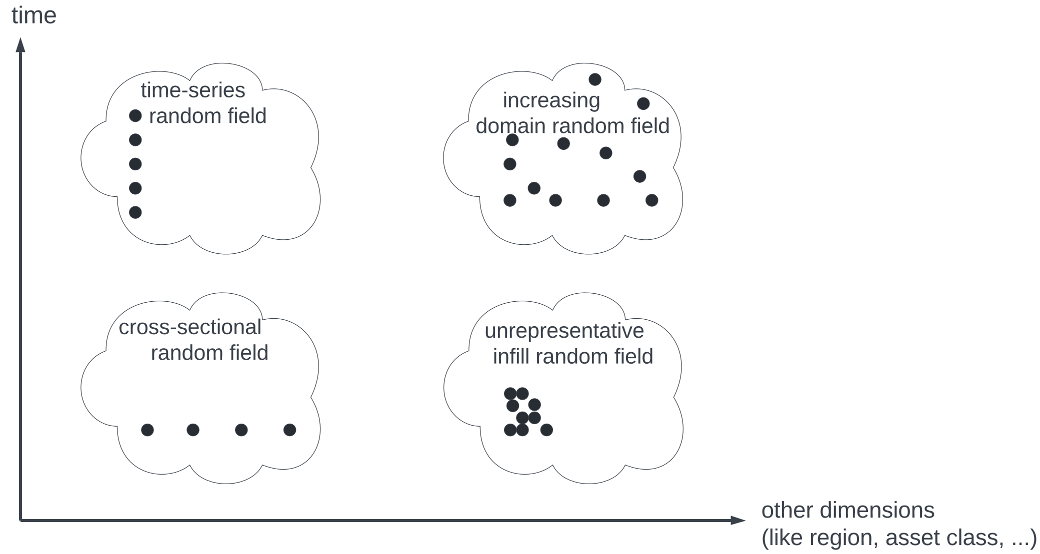

Figure 2: Visualization of generic random field types. Each black dot marks a different observation i of the cash flow data. Importantly, the time axis does **not** correspond to the index t in $CF_{t,i}$ (rather to vintage years v_i). Comparing the four choices, we want to avoid an infill random field but prefer our data to constitute an increasing domain random field. The infill random field is even asymptotically "too clustered" or better "too unrepresentative" to allow for meaningful estimation and inference. The time-series and cross-sectional random fields correspond to the standard cases in the literature but could turn out too restrictive for a general approach. By smart design (like portfolio formation), we often can map an increasing domain random field to simpler time-series or cross-sectional versions.

Specifically, we could assume (as a so-called primitive condition) the random field $\bar{\epsilon}$ to be spatial near-epoch dependent with respect to fund vintage years (Jenish and Prucha, 2009, 2012), i.e., two funds with distance $d_{i,j} > D$ are assumed to be independent.

To satisfy the time trend part (i) of this law of large number assumption, the weighting factor w, introduced in Equation 13, can be used to make $\bar{\epsilon}$ stationary. Spatial near-epoch dependence with respect to fund vintage years formalizes the simple idea that two fund pricing errors $\bar{\epsilon}$ with a small absolute vintage year difference are supposed to be dependent since they are exposed to the same macroeconomic conditions. In contrast, two funds with a large absolute vintage year difference can be assumed independent.

2.5.3 Consistency

The estimator $\hat{\theta}$ shall converge in probability to the true parameter value θ_0 as the number of distinct vintage years in our data set goes to infinity. Multiple funds for a specific vintage year are not necessarily required but provide additional information that we want to exploit if available.

Lemma 1. *A modified version of (Newey and McFadden, 1994, Theorem 2.1) holds, i.e., if there is a function $Q_0(\theta)$ such that*

1. *Identification: $Q_0(\theta)$ is uniquely minimized at θ_0,*
2. *Boundedness: Θ is compact,*
3. *Continuity: $Q_0(\theta)$ is continuous,*
4. *Uniform convergence: $\hat{Q}_n(\theta)$ converges uniformly in probability to $Q_0(\theta)$,*

then $\hat{\theta} \xrightarrow{p} \theta_0$ as $n \to \infty$.

Proof. The general proof is given in (Newey and McFadden, 1994, Chapter 2) for a max instead of min extremum estimator. Thus, we only recapitulate the four conditions required by the lemma in our specific context.

1. Obviously, we have to replace "maximized at θ_0" by "minimized at θ_0" compared to the exposition of (Newey and McFadden, 1994, Chapter 2). Then, we need to first show that $\mathbb{E}\left[\bar{\epsilon}(\theta_0)\right] = x$, where x is the minimum bias achievable, see Proposition 1. Secondly, we know $Q_0(\theta_0) = \mathbb{E}\left[L\left(\bar{\epsilon}(\theta_0)\right)\right] \geq 0$, e.g., for $L(x) = x^2$ we have $Q_0(\theta) = \mathbb{E}\left[\left(\bar{\epsilon}(\theta)\right)^2\right] = \mathrm{Var}\left[\bar{\epsilon}(\theta)\right] + \left(\mathbb{E}\left[\bar{\epsilon}(\theta)\right]\right)^2$ where the second summand can be perceived as bias term that is zero for θ_0. The variance term $\mathrm{Var}\left[\bar{\epsilon}(\theta)\right]$ for a simplified DGP is analyzed by Corollary **??**.
2. Compactness of Θ can be assured by lower and upper bounds for all parameters that can be justified by economic reasoning. In our case, e.g., a market beta factor of ten seems implausible for PE funds because of the implied risk and return expectations.
3. Continuity of the limit is a quiet weak and thus a standard regularity condition.
4. The second standard regularity condition is given by Assumption 4 which satisfies the definition of uniform convergence in probability (Newey and McFadden, 1994, Section 2.1) . To make this obvious, we can write $\hat{Q}_n(\theta) = Q_n(\theta) = n^{-1}\sum_{i=1}^{n} L\left(\bar{\epsilon}_i\right)$ and $Q_0(\theta) = \mathbb{E}\left[Q_n(\theta)\right]$ and compare it to Assumption 4.

□ □

2.5.4 Central limit theorem

To assess the large-sample significance of our parameter estimates (as done in the following Subsection 2.6), we want to describe the asymptotic distribution of the parameter vector as a normal distribution.

Proposition 2. *With estimator consistency established in Lemma 1, and the five (technical) conditions from (Newey and McFadden, 1994, Theorem 3.1) satisfied, it holds*

1. $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma)$ *as* $V, n \to \infty$ *with covariance matrix* Σ*, and*

2. *The covariance matrix* Σ *can be characterized by Pötscher and Prucha (1997, Theorem 11.2.b, Theorem H.1) (as outlined in the next Section 2.6).*

Proof. The extended proof of Proposition 2 may be derived in analogy to the GMM case in (Jenish and Prucha, 2012, Theorem 4) that shows that the general structure of the Pötscher and Prucha (1997) framework also applies to the spatial near-epoch dependent case. Alternatively (and easier), the estimator from Equation 14 can be clearly formulated as extremum estimator in alignment with our Definition 1. In consequence, (Newey and McFadden, 1994, Theorem 3.1), which generally describes the asymptotic normality of extremum estimators, is directly applicable to obtain the stated result. Thus, all details of the proof can be found in the original reference (Newey and McFadden, 1994, Chapter 3). □ □

2.6 Large sample inference

In this subsection, we demonstrate how to empirically apply Proposition 2 to obtain the asymptotic standard errors for our estimator from Equation 14. In the time-series, near-epoch-dependent LMD literature, the covariance matrix Σ can be characterized according to Pötscher and Prucha (1997, Theorem 11.2.b, Theorem H.1):

$$\Sigma = H^{-1}\Lambda(H^{-1})^\top$$

with expected Hessian matrix converging to H as $n \to \infty$

$$\mathbb{E}\left[\nabla_{\theta\theta} Q_n\right] \xrightarrow{p} H$$

and the expected covariance matrix of gradients converging to Λ as $n \to \infty$

$$n \cdot \mathbb{E}\left[\nabla_\theta Q_n (\nabla_\theta Q_n)^\top\right] \xrightarrow{p} \Lambda$$

Here, the gradient vector $\nabla_\theta Q_n$ is denoted as column vector. We define the corresponding finite sample estimators analogously to Pötscher and Prucha (1997, Chapters 12, 13.1), and numerically approximate the first and second partial derivatives by finite differences[5]. Specifically, we use the following central difference approximations (with "small" δ) (Eu, 2017, Algorithm 2):

$$f_x(x, y) \approx \frac{f(x + \delta, y) - f(x - \delta, y)}{2\delta}$$

$$f_{xx}(x, y) \approx \frac{f(x + \delta, y) + f(x - \delta, y) - 2f(x, y)}{\delta^2}$$

$$f_{xy}(x, y) \approx \frac{f(x + \delta, y + \delta) + f(x - \delta, y - \delta) - f(x + \delta, y - \delta) - f(x - \delta, y + \delta)}{4\delta^2}$$

[5]As an alternative to finite differences the widespread Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm can be applied to approximate the Hessian (Nocedal and Wright, 2006, Section 6.1).

The Hessian term $\hat{H}$ is relatively straightforward

$$\hat{H} = \frac{1}{n}\sum_{i=1}^{n} \nabla_{\theta\theta} L\left(\epsilon_i\right)$$

Due to the spatial near-epoch dependence, the involved and computationally expensive part is to consistently estimate $\hat{\Lambda}$ by a Spatial Heteroskedasticity and Autocorrelation Consistent (SHAC) covariance matrix estimator (Kim and Sun, 2011, Equation 2)

$$\hat{\Lambda} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} k_{i,j}\left[\nabla_{\theta} L\left(\epsilon_i\right)\left(\nabla_{\theta} L\left(\epsilon_j\right)\right)^{\top}\right] \tag{15}$$

We define the kernel weight k as

$$k_{i,j} \equiv K\left(\frac{d_{i,j}}{b_n}\right)$$

with kernel function $K : \mathbb{R} \to [0,1]$ satisfies $K(0) = 1$, $K(x) = K(-x)$, $\int_{-\infty}^{\infty} K^2(x)dx < \infty$, and $K(\cdot)$ continuous at zero and at all but a finite number of other points. A common choice is the Bartlett kernel $K_{BT}(x) = \max(0, 1 - |x|)$; see equation 2.7 in Andrews (1991) for other popular kernel choices. This means absolute vintage year differences larger than the bandwidth (or truncation) parameter $b_n = D$ are considered independent and are thus excluded from the $\hat{\Lambda}$ estimation formula.

In large samples, the vector of parameter standard errors can thus be estimated by

$$\mathrm{SE}(\hat{\theta}) = \sqrt{\mathrm{diag}\left[n^{-\frac{1}{2}}\hat{H}^{-1}\hat{\Lambda}(\hat{H}^{-1})^{\top}(n^{-\frac{1}{2}})^{\top}\right]} = \sqrt{\mathrm{diag}\left[\hat{H}^{-1}\hat{\Lambda}(\hat{H}^{-1})^{\top}\right]\cdot\frac{1}{n}} \tag{16}$$

The Wald test statistic for linear hypotheses $H_0 : R\theta = r$ and $H_1 : R\theta \neq r$ is constructed as (where, H_0 and H_1 are hypotheses and not Hessian terms H)

$$W = (R\hat{\theta} - r)^{\top}\left[R\frac{\hat{H}^{-1}\hat{\Lambda}(\hat{H}^{-1})^{\top}}{n}R^{\top}\right]^{-1}(R\hat{\theta} - r) \overset{H_0}{\sim} \chi_q^2$$

where $\hat{\theta}$ is the $p \times 1$ parameter vector, R is a $q \times p$ matrix, and r is a $q \times 1$ vector. Usually, we select R as $p \times p$ identity matrix, and r as $p \times 1$ vector (e.g., of zeros). Under the null hypothesis, W is chi-squared distributed with q degrees of freedom. As large values of W indicate the rejection of H_0, the corresponding p-value is calculated as $1 - F_{\chi_q^2}(W)$ where $F_{\chi_q^2}$ is the cumulative distribution function of a chi-squared random variable with q degrees of freedom.

However, given the limited amount of available private equity data (typically the oldest vintages start in the 1980s), asymptotic characterizations of Σ and $\mathrm{SE}(\hat{\theta})$ are of limited importance. In empirical applications, the small sample behavior of an estimation method for private equity data is more relevant than its asymptotic theory. Moreover, the standard asymptotic distribution associated with an estimator is generally not valid for post-model-selection inference, i.e., if a model selection procedure is applied to find the best model from a collection of competitors (Leeb and Pötscher, 2005).

2.7 Comparison to similar estimators

Our Least-Mean-Distance (LMD) estimator introduced in Section 2.2 belongs to the class of semiparametric nonlinear M-estimators as defined in Pötscher and Prucha (1997) which are extremum estimators. To gain more flexibility and avoid unneeded complexity, we intentionally opt against the most prominent semiparametric nonlinear M-estimator framework, i.e., classical time-series Generalized Method of Moments (GMM) (Hansen, 1982, 2012). A classical GMM approach requires the construction of stationary, ergodic time-series of moment conditions that are used to empirically estimate the expected value of pricing errors in Equation 12. The stationarity requirement of classical time-series GMM limits (i) more elaborate weighting schemes for w, like fund-size weighting, and (ii) the usage of fund cash flows from non-realized vintages.

2.7.1 Comparison to Driessen et al. (2012)

The Driessen et al. (2012) approach is most closely related to our methodology.

One important difference is that we select a simpler and more flexible LMD estimator instead of a cross-sectional GMM approach. In our view, the choice of the more complex cross-sectional GMM just causes some conceptional issues, whereas the underlying formulas are basically the same as for our LMD estimator[6]. As a first limitation, they have to regard vintage-year portfolios as their cross-sectional units; we can also use individual funds. In this context, we also question their statement that “to identify β, it is essential that the different FoFs are exposed to different market returns” since it is perfectly fine to perceive their estimator as cross-sectional approach[7]. Second, the Driessen et al. (2012) asymptotic theory assumes the number of funds (or deals) per vintage year portfolio to go to infinity. To comply with standard GMM assumptions, the number of vintage years, which corresponds to the number of moment conditions in their approach, *must be considered fixed* and thus cannot grow asymptotically (Han and Phillips, 2006; Newey and Windmeijer, 2009). For a typical LMD estimator (e.g., nonlinear least squares), this constraint does not exist. Our asymptotic theory lets both (i) the number of vintage years and (ii) the number of funds go to infinity but bounds the number of funds per vintage year.

Further, Driessen et al. (2012) discount all fund cash flows just to the first cash flow date (like in a classical net present value calculation). In contrast, we additionally average over all dates within $\mathcal{T}_i$ to tackle the problem arising from pricing pooled cash flows, which we thoroughly analyzed in subsection 2.1. Although Driessen et al. (2012) describe their estimator as a one-step GMM approach, we consider it a special case of our LMD estimator. Specifically, Equation 14 from our methodology is a generalization of equation 3 from their paper. Consequently, if someone accepts the assumptions from Subsection 2.5, our large sample inference framework from Subsection 2.6 applies to their case without any significant modification. Finally, Driessen et al. (2012) apply simple cross-sectional bootstrapping to obtain standard

[6]The formulas are this similar because Driessen et al. (2012) use the identity matrix as GMM weighting matrix and skip the second GMM step.

[7]In a classic cross-sectional regression, we only have one market return realization.

errors; in contrast, in Subsection 3.2, we use a cross-validation technique that is adapted to the near-epoch dependence of the PE fund data.

2.7.2 Comparison to Korteweg and Nagel (2016)

Korteweg and Nagel (2016), first of all, realized the usefulness of employing an auxiliary spatial framework to establish asymptotic inference results for a fund-level panel dataset of private equity funds. To account for the cross-sectional dependence between funds, they measure the economic distance between two private equity funds (by the degree of cash flow overlap). Concretely, their asymptotic inference framework draws on the spatial HAC estimator of Conley (1999); our spatial HAC framework uses Pötscher and Prucha (1997); Kim and Sun (2011); Jenish and Prucha (2012). However, they ultimately utilize a classical GMM estimator, thus a time-series law of large numbers. Specifically, we obtain the estimator of (Korteweg and Nagel, 2016, Equation 18) in our framework if we replace $Q_n(\theta)$ in Equation 14 by Equation 17.

$$Q_n(\theta) = L\left(\frac{1}{n}\sum_{i=1}^{n}\bar{\epsilon}_i\right) \quad \text{with} \quad L(x) = x^\top W x = x^\top I x \tag{17}$$

with identity matrix I as weighting matrix W. In accordance with classical GMM, the function $\frac{1}{n}\sum_{i=1}^{n}\bar{\epsilon}_i : \mathbb{R}^{n\times T}\times\Theta \to \mathbb{R}^m$ should be perceived as multidimensional where the dimensionality of the function output corresponds to the number of moment conditions.

Time-series GMM estimators inherently bear the risk of under-identification if the corresponding time-series is constructed by pooling all fund cash flows from a given fund type. Exactly this happens in Equation 17 with $m = 1$ where we consequentially obtain a GMM estimator with just one moment condition[8]. To counter under-identification, additional characteristic-based fund portfolios could be formed to increase the number of moment conditions per fund type; also, random portfolios combined with bootstrapping could make sense. Yet, Korteweg and Nagel (2016) take another approach and introduce the concept of Generalized Public Market Equivalent (GPME), which elegantly avoids the under-identification issue. Firstly, a public market SDF model is estimated by pricing public trading strategies that shall replicate PE funds instead of directly pricing the observed PE fund cash flows. Only in a second step these public market SDF models are applied to evaluate private equity fund cash flows.

Given these differences, our approach may not be perceived as a straightforward generalization of the Korteweg and Nagel (2016) framework. In contrast, our LMD estimator generalizes the Driessen et al. (2012) method. Table 1 summarizes the most prominent distinctions between the three approaches.

[8]In contrast, our estimator corresponds to the opposite edge case with asymptotically an infinite number of LMD "moment conditions" (units to price) as we let $n \to \infty$.

	Driessen et al. (2012)	Korteweg and Nagel (2016)	Our approach
M-estimator	Cross-sectional Generalized Method of Moments	Time-series Generalized Method of Moments	Least-Mean-Distance
Pricing error averaging	No	No	Yes
Cash flows priced	PE cash flows	public cash flows	PE cash flows
Asymptotics	cross-sectional #funds $\rightarrow \infty$	time-series #vintages $\rightarrow \infty$	spatial # of both $\rightarrow \infty$
Inference	bootstrap	spatial HAC	cross-validation & spatial HAC
Cross-sectional unit	vintage year portfolio	single fund	testing both
SDF	simple linear	exponentially affine	testing both

Table 1: Comparison to similar estimation frameworks.

3 Empirical application

3.1 Data

We use the Preqin cash flow data set as of 14th April 2022 that is well known in the academic private equity literature (Harris et al., 2014; Korteweg and Nagel, 2016; Ang et al., 2018). To keep the empirical analysis clear and concise, we filter the dataset for Private Equity (PE) funds.

The Private Equity sample contains 3004 distinct funds spreading over 37 vintage years. The region distribution is as follows: 1983 (1), 1985 (2), 1986 (4), 1987 (5), 1988 (7), 1990 (2), 1991 (4), 1992 (9), 1993 (13), 1994 (20), 1995 (16), 1996 (20), 1997 (24), 1998 (59), 1999 (46), 2000 (72), 2001 (51), 2002 (41), 2003 (45), 2004 (64), 2005 (117), 2006 (154), 2007 (176), 2008 (183), 2009 (83), 2010 (92), 2011 (146), 2012 (136), 2013 (142), 2014 (182), 2015 (152), 2016 (186), 2017 (143), 2018 (157), 2019 (190), 2020 (140), 2021 (120). The region distribution is as follows: Africa (21), Asia (173), Australasia (41), Europe (749), Latin America & Caribbean (34), Middle East (17), North America (1968). The strategy distribution is as follows: Balanced (71), Buyout (1449), Co-Investment (108), Co-Investment Multi-Manager (67), Direct Secondaries (34), Fund of Funds (696), Growth (386), Secondaries (174), Turnaround (19).

For this subset of Private Equity funds, 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 final cash flow. We explicitly refrain from excluding the most recent vintage years. Thus, the minimum vintage year is 1983 (just for PE), and the maximum is 2021.

The public market factors that enter our SDF draw on the US data set of the recently popularized q^5 investment factor model sourced from `http://global-q.org/factors.html` (Hou et al., 2015, 2020). Their five-factor model includes the market excess return (MKT), a size factor (ME), an investment factor (IA), a return on equity factor (ROE), and an expected growth factor (EG).

3.2 Model and estimator specifications

We test a simple linear SDF model similar to Driessen et al. (2012)

$$\Psi^{\mathrm{SL}}_{\tau,t}(\theta) = \frac{\prod_{h=0}^{\tau}\left(1+\alpha+r_h+\sum_j \ \beta_j \ F_{j,h}\right)}{\prod_{h=0}^{t}\left(1+\alpha+r_h+\sum_j \ \beta_j \ F_{j,h}\right)} \tag{18}$$

and an exponential affine SDF model adapted from Korteweg and Nagel (2016)

$$\Psi^{\mathrm{EA}}_{\tau,t}(\theta) = \exp\left[\sum_{h=0}^{\tau} X_h \sum_{h=0}^{t} -X_h\right] \tag{19}$$

with

$$X_h = \alpha + \log(1+r_h) + \sum_{j\in J} \beta_j \cdot \log(1+F_{j,h})$$

with (arithmetic) risk-free return $r = R_{rf} - 1$, (arithmetic) zero-net-investment portfolio returns F_j, and parameter vector $\theta = (\alpha, \beta)$. To avoid overfitting, we just test six simple SDF models that contain {MKT} alone or {MKT} plus {ME or IA or ROE or EG or Alpha}. In Equation 14, we use the quadratic loss function $L(x) = x^2$.

To assess the parameter significance, we compute the asymptotic standard errors as outlined in Subsection 2.6. For the Bartlett kernel's bandwidth $b_n = D$ we select 12 years, i.e., funds with absolute vintage year differences larger than 12 years are assumed to be independent.

Additionally, we want to test the finite - or, more honestly, small - sample parameter significance and the out-of-sample performance of our SDF models. To account for the dependency between funds from adjacent vintage years caused by overlapping fund cash flows, we draw on *hv*-block cross-validation (Racine, 2000). Therefore, we form three partitions for several vintage year groups. As larger validation sets are preferred for model selection, the validation set (*v*-block) always contains funds of three neighboring vintage years (e.g., 2000, 2001, 2002). To reduce the dependency between training and validation set, we remove all funds from three-year-adjacent vintage years, i.e., the *h*-block (e.g., 1997, 1998, 1999, 2003, 2004, 2005). Funds from the remaining vintage years enter the training set and are thus used for model estimation (e.g., 1985-1996, 2006-2019). We apply ten-fold cross validation using the ten validation sets described in Table 2. This means we replace the bootstrap standard error calculation of Driessen et al. (2012) by *hv*-block cross-validation since the new method (i) accounts for near-epoch-dependence, (ii) focuses directly on the out-of-sample performance of the SDF models, and (iii) is computationally cheaper.

3.3 Simulation study

Our Monte Carlo experiments examine the following questions related to the bias and variance of our estimation methodology in finite samples.[9] Is it beneficial to

[9]As each simulation study it more investigates the ability to identify the assumed data generating process than the corresponding SDF model.

training.before	h-block.before	v-block	h-block.after	training.after
estimation	remove	validation	remove	estimation
start-1984	1985,1986,1987	1988,1989,1990	1991,1992,1993	1994-end
start-1987	1988,1989,1990	1991,1992,1993	1994,1995,1996	1997-end
start-1990	1991,1992,1993	1994,1995,1996	1997,1998,1999	2000-end
start-1993	1994,1995,1996	1997,1998,1999	2000,2001,2002	2003-end
start-1996	1997,1998,1999	2000,2001,2002	2003,2004,2005	2006-end
start-1999	2000,2001,2002	2003,2004,2005	2006,2007,2008	2009-end
start-2002	2003,2004,2005	2006,2007,2008	2009,2010,2011	2012-end
start-2005	2006,2007,2008	2009,2010,2011	2012,2013,2014	2015-end
start-2008	2009,2010,2011	2012,2013,2014	2015,2016,2017	2018-end
start-2011	2012,2013,2014	2015,2016,2017	2018,2019,2020	2021-end

Table 2: Partitions used for hv-block cross-validation.

use vintage-year portfolios instead of individual funds? Which SDF model performs better when we also use the corresponding data generating process (i.e., assume correct model specification)? How is estimator precision affected by varying numbers of vintage years and cross-sectional units? Which is the optimal set of discounting dates $\mathcal{T}$?

We use historical q-investment factors from 1986 to 2005 and simulate 20 funds for each of these 20 vintage years. Each fund contains 15 deals with equal investment amounts and exactly one divestment cash flow. Deals are entered within the first five years of the fund lifetime following a discrete uniform distribution and afterward held between one to ten years again uniformly distributed. The deal returns are generated by the simple linear or exponential affine SDF models described in Equations 18 and 19. In the base case, we just use the MKT factor with $\beta_{\text{MKT}} = 1$ and in each month, add a normal i.i.d. error term with standard deviation $\sigma = 0.2$ and zero mean. Additionally, we test an intercept term α of -0.25% per month and a high β_{MKT} of 2.5. In the exponential affine case, we adjust the lognormally distributed error mean to zero by subtracting $0.5\sigma^2$. If a negative return exceeds -100%, the company defaults with a zero exit cash flow. In contrast, the error term in the simulations of Driessen et al. (2012) is more well-behaved as it follows a shifted lognormal distribution that, even with arbitrarily high error term variance, just allows for returns below say -99%, if the market return is close to its lower bound (see equation 9 in their online appendix). In our base case, the set of discounting dates $\mathcal{T}$ contains all months from the first cash flow to the maximum month 180. To assess our estimator's bias and variance, we simulate 1000 test scenarios for vintage year portfolios and only 200 test cases when using individual funds due to the higher computational costs of simulating the individual fund cash flows.

Cross-sectional unit i: As presumed in Subsection 2.4, vintage year portfolio results appear to have lower bias and variance when compared to individual funds. For the simple linear SDF and maximum month 180, the mean and standard deviation of the coefficient estimate $\hat{\beta}_{\text{MKT}}$ is 1.016 (0.2) for the vintage year portfolio and 1.096 (0.376) for individual funds. More results are depicted in Figure 3. However,

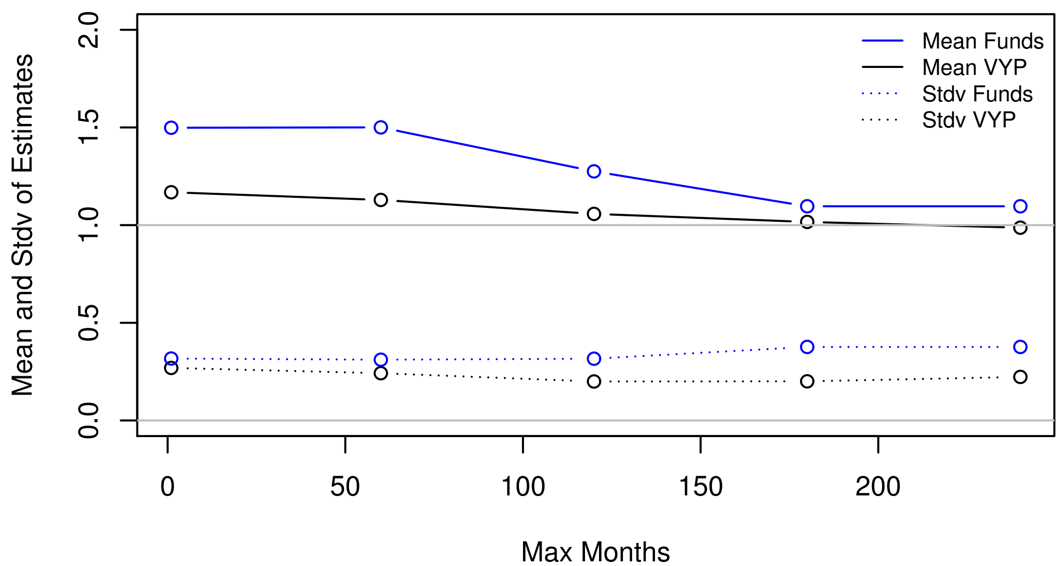

Figure 3: Simulation results comparing individual funds vs. vintage year portfolios (VYPs) with true $\beta = 1$ and simple linear SDF (200 simulation iterations).

for individual funds, we only simulate 200 iterations due to the high computational cost.

This finding has two important implications: On the one hand, vintage year portfolio formation can substantially decrease our estimator's bias and variance. On the other hand, it also dramatically reduces the number of cross-sectional units and consequentially impairs the importance of asymptotic results. These considerations may explain the choice of Korteweg and Nagel (2016) to use individual funds as cross-sectional units in their asymptotic SHAC framework to obtain smaller standard error estimates.

SDF model Ψ: In our base case with vintage year portfolios, the exponential affine SDF shows a mean and standard deviation of 1.011 (0.175) compared to the 1.016 (0.2) achieved by the simple linear SDF. Generally, the exponential affine SDF model and the simple linear SDF model exhibit similar bias and variance, cf. panels A and B in Table 4. Figure 4 visualizes the true $\beta = 1$ case, which shows that the estimation results are not overly sensitive to the choice of the SDF model.

Moreover, the perceived superiority of exponential affine SDFs is probably rather theoretical than practical as other proponents also emphasize their universality mainly from a mathematical perspective without providing supportive empirical or simulation results (Gourieroux and Monfort, 2007; Bertholon et al., 2008).

Varying vintages V and portfolio sizes n/V: To test the effect of varying data sizes available for MKT factor estimation, we in/decrease the (i) number of vintage years and (ii) the number of funds per vintage year (cf. Table 3). Here we

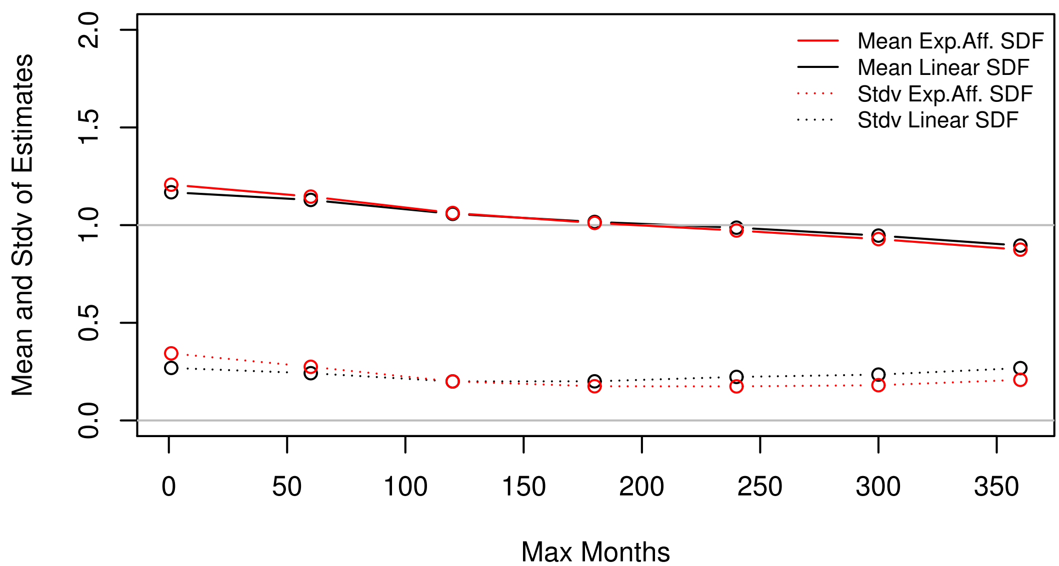

Figure 4: Simulation results comparing exponentially affine and simple linear SDF with true $\beta = 1$ and vintage year portfolios (1000 simulation iterations).

use vintage year portfolios and the simple linear SDF. For our simple data generating process, increasing the number of deals/funds per vintage year portfolio appears to decrease the estimator's variance more effectively than adding more vintage years. However, the bias is almost the same for all tested specifications. Generally, we seem to need many new data points to ensure a reasonable variance of our estimator.

	Base	Big n/V	Big V	Big V	Small V	Small V
Start vintage	1986	1986	1967	1967	1986	1996
End vintage	2005	2005	2005	2005	1995	2005
#Funds per vintage	20	40	10	20	20	20
Mean β_{MKT}	1.011	1.020	0.993	1.015	1.027	0.934
Stdv β_{MKT}	0.187	0.133	0.263	0.227	0.232	0.418

Table 3: Simulation study for varying number of vintages and number of funds per vintage. We use vintage year portfolios, the simple linear SDF with true $\beta_{\text{MKT}} = 1$, maximum month 180, and 500 simulation iterations.

Size of set $\mathcal{T}$: The results in Table 4 indicate that we can control the asymptotic bias by an appropriate choice of the set $\mathcal{T}$. For the one-factor model, the bias almost vanishes when we average over all discounting dates in the maximal fund lifetime of 180 months. For smaller or larger sets for $\mathcal{T}$, we find increasing bias terms. Recall that using the minimal set for $\mathcal{T}$, i.e., discounting all cash flows just to the fund inception date, corresponds exactly to the Driessen et al. (2012) approach. Thus, the original Driessen et al. (2012) methodology might achieve a suboptimal asymptotic bias since it does not average pricing errors over multiple discounting dates.

The same finding also holds when we limit the maximal fund lifetime to ten years by reducing the maximum deal holding period from ten to five years. Here, under correct model specification with $\beta_{\text{MKT}} = 1$, the smallest bias is obtained for maximum month 120, for max. month 60 we get 1.028 (0.116), for max. month 120, we get 1.005 (0.116), and for max. month 180 we get 0.969 (0.13). However, this simulation results only hold for the one-factor model (MKT) reported here. For multi-factor models, smaller max. month values can yield the lowest bias term. Thus, our simulation study could not reveal a formula how to determine the optimal set of $\mathcal{T}$; it seems to be data and SDF model dependent.

In Table 4 for both true and false model specifications, the α standard deviation is very high compared to its mean value. This may indicate it is rather delicate to empirically determine private equity's historical outperformance by our semiparametric estimator.

To conclude, our simulations study rationalizes two key practices from the Driessen et al. (2012) paper: (i) vintage year portfolio formation helps to improve estimator precision, and (ii) increasing the number of funds per vintage seems to be more effective in controlling estimator variance that increasing the number of vintages[10]. However, our examples with correct specification cannot support the assumption of Korteweg and Nagel (2016) that (iii) the exponential affine SDF is (clearly) superior to the simple liner SDF in a multi-period framework; actually, their bias and variances are quite equal. Moreover, our simulation study suggests that (iv) averaging pricing errors over multiple dates strikingly reduces the bias inherent to the original procedure of Driessen et al. (2012) that just discounts all cash flows to the fund inception date. Actually, choosing the set $\mathcal{T}$ according to the fund lifetime seems to decrease the bias (and to a lesser extent also the variance) more effectively than all other measures combined.

[10]Finding (ii) may explain the choice of Driessen et al. (2012) to employ an asymptotic law that lets the number of deals/funds per vintage tend to infinity.

Panel A: simple linear SDF

Model==DGP	True	False		False	True	
MaxMonth	$\beta = 1$	$\alpha = 0$	$\beta = 1$	$\beta = 2.5$	$\alpha = -0.25$	$\beta = 2.5$
1 - mean	1.168	1625%	0.003	2.023	5879%	-16.711
1 - stdv	0.269	2792%	9.968	0.342	866%	13.347
60 - mean	1.129	0.138%	0.933	2.103	-0.086%	2.285
60 - stdv	0.242	0.245%	0.363	0.302	0.253%	0.406
120 - mean	1.058	0.112%	0.906	2.063	-0.085%	2.239
120 - stdv	0.200	0.214%	0.313	0.253	0.239%	0.385
180 - mean	1.016	0.041%	0.965	2.052	-0.161%	2.370
180 - stdv	0.200	0.172%	0.334	0.277	0.173%	0.403
240 - mean	0.987	-0.053%	1.077	2.072	-0.277%	2.589
240 - stdv	0.223	0.162%	0.361	0.326	0.118%	0.375
300 - mean	0.946	-0.149%	1.175	2.080	-0.357%	2.714
300 - stdv	0.235	0.174%	0.377	0.398	0.114%	0.366
360 - mean	0.895	-0.245%	1.269	2.048	-0.461%	2.859
360 - stdv	0.268	0.201%	0.399	0.551	0.140%	0.386

Panel B: exponential affine SDF

Model==DGP	True	False		False	True	
MaxMonth	$\beta = 1$	$\alpha = 0$	$\beta = 1$	$\beta = 2.5$	$\alpha = -0.25$	$\beta = 2.5$
1 - mean	1.207	203%	1.276	2.256	692%	1.704
1 - stdv	0.344	314%	0.710	0.290	13%	1.666
60 - mean	1.146	0.126%	0.941	2.264	-0.018%	2.277
60 - stdv	0.275	0.264%	0.386	0.256	0.370%	0.473
120 - mean	1.062	0.107%	0.908	2.221	0.009%	2.205
120 - stdv	0.200	0.237%	0.333	0.187	0.357%	0.448
180 - mean	1.011	0.027%	0.971	2.182	-0.136%	2.358
180 - stdv	0.175	0.211%	0.366	0.168	0.344%	0.505
240 - mean	0.972	-0.088%	1.095	2.144	-0.441%	2.723
240 - stdv	0.174	0.224%	0.406	0.178	0.317%	0.503
300 - mean	0.928	-0.202%	1.203	2.083	-0.717%	2.985
300 - stdv	0.181	0.253%	0.426	0.254	0.340%	0.513
360 - mean	0.874	-0.319%	1.304	1.685	-1.095%	3.272
360 - stdv	0.208	0.291%	0.447	0.772	0.374%	0.586

Table 4: Simulation study to compare the simple linear with the exponential affine SDF and to determine the optimal size of the set $\mathcal{T}$. Here, we always use vintage year-portfolios and 1000 simulation iterations. For better readability, $\beta_{\text{MKT}} = \beta$. For the unity and high beta model, we test true and false model specifications (with and without the monthly α term).

3.4 Empirical results

Following the conclusions from the previous subsection, we use vintage-year portfolios to estimate simple linear SDF models with maximum month 180. Asymptotic inference results for the full dataset are exhibited in Table 6 for fund-size weighting and in Table 8 for equal weighting. The results for hv-block cross-validation are displayed in Table 7 for fund-size weighting and in Table 9 for equal weighting. We generally analyze the results in a two-step procedure: For a given model specification, we use the cross-validation error (i.e., the average out-of-sample error) to select the best model for each fund type but analyze the corresponding coefficient estimates from the asymptotic inference tables (estimated on the entire data set). Therefore, for each fund type the SDF models in the asymptotic inference Tables 6 and 8 are sorted by the corresponding cross-validation error. Throughout this subsection, we define the statistical significance of coefficient estimates in terms of a t-ratio $\hat{\theta}[SE(\hat{\theta})]^{-1}$ greater than 1.96.

		MKT Factor			Second Factor		
Weighting	Inference	Coef	SE	SE.indep	Coef	SE	SE.indep
fund-size	asymptotic	0.75	27.06	19.73	0.80	28.95	20.94
fund-size	cross-validation	0.85	0.38	-	0.59	0.51	-
equal	asymptotic	0.76	26.75	16.16	0.76	11.25	6.69
equal	cross-validation	0.84	0.34	-	0.62	0.50	-

Table 5: Top-level overview over Table 6 to 9: Averages of absolute values of coefficient estimates and standard errors (SEs). We see that asymptotic SEs are much higher than the SEs obtained by cross-validation.

Table 5 helps to get a rough overview of Table 6 to 9 as it summarizes their absolute column means. Conspicuously, asymptotic standard errors (SEs) seem enormously high and, moreover, contain colossal outliers. The standard errors implied by $hv-$block cross-validation are considerably smaller than the asymptotic SEs and seem to lie within a plausible range. When just looking at asymptotic standard errors of the second factors, fund-size weighting exhibits substantially larger SEs than fund equal-weighting. Assuming independence between funds from different vintages decreases asymptotic SEs by approximately 30-40% compared to a realistic kernel bandwidth of $D = 12$. But even these independent SEs rarely imply statistical significance coefficient estimates with t-ratios bigger than 1.96. In Table 6 with fund-size weighting, just one out of 36 models exhibit asymptotically significant MKT and second-factor estimates. In the case of equal-weighting, Table 8 also shows just one asymptotically significant model out of 36.

In summary, the results reveal weak two-factor models with MKT plus a second q-investment factor. Likewise, the simulation results from the previous subsection indicate a rather high variance associated with our semiparametric estimator (given the amount of data typically available). Thus, we recommend focusing on single MKT factor models even when their asymptotic t-ratios are below 1.96. At least the hv-block cross-validation standard deviations imply significant one-factor MKT models for fund types PE, VC, PD, INF. In contrast, RE is just significant for equal weighting, and NR is insignificant for both weighting schemes.

Focus on PE and VC estimates Here, we briefly summarize the one-factor MKT and the two-factor Alpha model estimates for fund types PE (i.e., mainly Buyout and Growth) and VC. For PE, all one-factor MKT model β_{MKT} estimates fall in the range from 1.13 to 1.28. If we add an α term, all β_{MKT} estimates decrease to the range 0.61 to 0.77 with annualized α coefficients of approximately positive 4-5% per year. For VC, the one-factor MKT model β_{MKT} estimates are in the range from 0.80 to 1.14. If we add an α term, all β_{MKT} estimates strongly increase to the range 1.81 to 2.06 with annualized α coefficients of approximately negative 6-7% per year. These results at least weakly indicate - given their insignificant asymptotic standard errors - that PE funds outperform public markets with a market beta coefficient of less than one, which suggests low market risk. On the other hand, VC underperforms public markets with market beta coefficients of roughly two, which implies high market risk. So, even Driessen et al. (2012) use the problematic Thomson Venture Economics (TVE) dataset for their empirical analysis[11], we obtain similar quantitative and qualitative results using Preqin data: (i) the market beta of VC seems to be higher than that of PE, and (ii) VC, in contrast to PE, appears to exhibit a negative abnormal performance α[12].

As a robustness check, we reestimate all SDF models on a dataset that just contains funds from vintages older or equal than 2011. Interestingly, the PE and VC results regarding β_{MKT} and α can be qualitatively and also quantitatively confirmed on this 'mostly-liquidated' dataset[13].

[11] Harris et al. (2014) discuss the potential downward bias of the TVE dataset.

[12] Similarly, (Metrick and Yasuda, 2010, Exhibit 4.6) find high beta coefficients (1.63 and 2.04) and small to negative alphas (-2.11% and 0.13%) for VC funds in their lag-return regression.

[13] All R code and data is available in an online repository. `https://github.com/quant-unit/Fundwise_SDF/tree/master/r_project`

	MKT Factor			Second Factor			
Type	Estim.	SE	SE.indep	Factor	Estim.	SE	SE.indep
PE	0.709	2.470	1.153	EG	0.807	4.693	1.960
PE	0.770	7.976	3.348	ROE	1.540	5.140	3.499
PE	1.126	1.003	0.868	MKT	1.126	1.003	0.868
PE	0.644	1.234	0.585	Alpha	0.003	0.036	0.013
PE	1.121	1.023	0.897	ME	0.074	2.021	0.915
PE	1.158	1.125	1.068	IA	-0.338	2.499	1.259
VC	1.053	4.150	2.733	IA	-1.959	2.100	1.767
VC	1.114	3.861	2.894	ME	-1.383	5.102	2.211
VC	1.806	11.391	4.279	Alpha	-0.006	0.124	0.046
VC	0.801	704.455	561.598	MKT	0.801	704.455	561.598
VC	1.429	8.073	3.219	ROE	-1.306	18.055	6.919
VC	1.507	17.322	6.966	EG	-0.904	15.344	5.737

Table 6: Asymptotic inference using vintage-year portfolios with fund-size weighting, max month 180, and kernel bandwidth $D = 12$. Standard Errors (SE) are calculated by Equation 16.

	MKT Factor		Second Factor			
Type	Mean	SD	Factor	Mean	SD	CV error
PE	0.867	0.276	EG	0.720	0.137	112808
PE	0.927	0.305	ROE	1.375	0.420	126801
PE	1.276	0.296	MKT	1.276	0.296	151964
PE	0.772	0.238	Alpha	0.004	0.002	154805
PE	1.317	0.396	ME	0.236	0.664	209319
PE	1.311	0.370	IA	0.014	0.703	210650
VC	1.045	0.126	IA	-1.890	0.238	11858
VC	1.172	0.126	ME	-1.448	0.263	13301
VC	1.930	0.356	Alpha	-0.005	0.001	17723
VC	0.804	0.363	MKT	0.804	0.363	21852
VC	1.527	0.517	ROE	-0.972	0.679	26680
VC	1.646	0.678	EG	-0.644	0.556	32730

Table 7: *hv*-block cross-validation using vintage-year portfolios with fund-size weighting and max month 180.

	MKT Factor			Second Factor			
Type	Estim.	SE	SE.indep	Factor	Estim.	SE	SE.indep
PE	0.775	0.638	0.550	EG	0.667	5.558	2.125
PE	0.610	1.064	0.387	Alpha	0.004	0.006	0.002
PE	0.826	20.352	8.308	ROE	1.087	33.514	12.143
PE	1.134	1.050	0.694	MKT	1.134	1.050	0.694
PE	1.146	1.001	0.638	IA	-0.386	1.909	0.813
PE	1.134	1.048	0.702	ME	-0.014	1.797	0.736
VC	1.181	24.418	16.693	ME	-1.277	4.928	4.352
VC	1.137	7.259	6.057	IA	-1.553	3.716	2.139
VC	1.956	4.189	1.520	Alpha	-0.006	0.335	0.117
VC	1.034	2.205	1.758	MKT	1.034	2.205	1.758
VC	1.488	1.801	0.941	ROE	-1.148	4.060	1.424
VC	1.535	2.821	1.336	EG	-0.754	3.626	1.260

Table 8: Asymptotic inference using vintage-year portfolios with fund-equal weighting, max month 180, and kernel bandwidth $D = 12$. Standard Errors (SE) are calculated by Equation 16.

	MKT Factor		Second Factor			
Type	Mean	SD	Factor	Mean	SD	CV error
PE	0.886	0.262	EG	0.614	0.217	101444
PE	0.719	0.205	Alpha	0.004	0.001	105842
PE	0.948	0.267	ROE	0.975	0.407	110926
PE	1.250	0.262	MKT	1.250	0.262	127589
PE	1.247	0.274	IA	-0.183	0.598	157037
PE	1.281	0.323	ME	0.048	0.644	169552
VC	1.250	0.153	ME	-1.292	0.234	16305
VC	1.183	0.169	IA	-1.507	0.327	16449
VC	2.052	0.257	Alpha	-0.006	0.001	18666
VC	1.138	0.341	MKT	1.138	0.341	25321
VC	1.610	0.431	ROE	-0.946	0.426	26618
VC	1.688	0.505	EG	-0.616	0.331	30392

Table 9: *hv*-block cross-validation using vintage-year portfolios with fund-equal weighting and max month 180.

4 Conclusion

Theoretically, our Least-Mean-Distance estimator can be easily generalized to estimate SDF models for all kinds of non-traded cash flows. Practically, semiparametric estimators commonly exhibit problematic small sample behavior. Given the amount of currently available private equity fund data, our estimator's variance seems quite large, even for simple SDF model specifications. Specifically, our Monte Carlo simulation results prompt us to conclude that the closely related Driessen et al. (2012) estimator may exhibit more bias and variance than originally assumed in their paper. Especially, the variance of α estimates seems to be too high to allow reliable abnormal performance conclusions. Fortunately, we show that at least the bias can be easily reduced by averaging pricing errors over all dates within the fund lifetime.

In the data-sparse private equity domain with only 20-40 cross-sectional units (i.e., vintage year portfolios) currently available for estimation, asymptotic inference seems not to be overly useful. Thus, we strongly advise always challenging asymptotic inference results by resampling or cross-validation techniques adapted to the dependence structure of overlapping fund cash flows. However, even these conclusions should be double-checked to avoid unreasonable instances, e.g., when hv-block cross-validation chooses dubious models with negative MKT factor estimates. Unfortunately, using individual funds instead of vintage year portfolios, which yields smaller asymptotic standard errors, constitutes no viable resolution as individual funds show considerably larger small-sample bias and variance in our Monte Carlo example. Since, in our empirical analyses, basically all two-factor models' asymptotic standard errors appear statistically insignificant, we conjecture that naive versions of our SDF estimator shall be exclusively used for a single-MKT-factor model until considerably more vintage year information for private equity funds is available.

If someone wants to estimate more complex SDF models that incorporate additional factors, more structure is needed. These can be parametric assumptions for the data generating process (Ang et al., 2018) or to extract additional information from intermediate net asset values (Gredil et al., 2020; Brown et al., 2021). A first "modern" approach to the same problem is applying machine learning techniques that regularize/shrink all coefficients other than the MKT factor. Secondly, given the high estimator variance revealed in the simulation study, statistical learning methods that create a strong learner by combining multiple weak learners seem also worth considering (boosting, bagging, or model averaging).

Incorporate fees and carry model to DGP for SDF estimation for PE gross-fee performance

Finally, we point to the potentially most interesting topic for future research. Our simulation study indicates that the estimator's bias and variance can be controlled by an appropriate choice for the set $\mathcal{T}$. This set averages the pricing error over multiple discounting dates. In simpler terms, an identification method that utilizes a future value concept instead of net present values obtains more favorable results in our case. The bias in our simulation study is minimal when the set of discounting dates corresponds to the fund lifetime. A parsimonious but general model that allows for misspecification and can explain this $\mathcal{T}$-averaging effect from a mathematical perspective would be highly appreciated.

References

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica: Journal of the Econometric Society*, pages 817–858.

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.

Ang, A., Liu, J., and Schwarz, K. (2020). Using stocks or portfolios in tests of factor models. *Journal of Financial and Quantitative Analysis*, 55(3):709–750.

Baker, K. R. and Trietsch, D. (2013). *Principles of sequencing and scheduling.* John Wiley & Sons.

Bansal, R. and Yaron, A. (2004). Risks for the long run: A potential resolution of asset pricing puzzles. *The journal of Finance*, 59(4):1481–1509.

Bertholon, H., Monfort, A., and Pegoraro, F. (2008). Econometric asset pricing modelling. *Journal of Financial Econometrics*, 6(4):407–458.

Brown, G. W., Ghysels, E., and Gredil, O. (2021). Nowcasting net asset values: The case of private equity. working paper (as of 2021-03-19).

Buchner, A. (2014). The alpha and beta of private equity investments. working paper (as of 2014-10-24).

Buchner, A. (2016a). Dealing with non-normality when estimating abnormal returns and systematic risk of private equity: A closed-form solution. *Journal of International Financial Markets, Institutions and Money*, 45:60–78.

Buchner, A. (2016b). Risk-adjusting the returns of private equity using the capm and multi-factor extensions. *Finance Research Letters*, 16:154–161.

Campbell, J. Y. and Cochrane, J. H. (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of political Economy*, 107(2):205–251.

Cattaneo, M. D., Crump, R. K., and Farrell, M. H. (2019). Characteristic-sorted portfolios: Estimation and inference. *Review of Economics and Statistics.*

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

Conley, T. G. (1999). Gmm estimation with cross-sectional dependence. *Journal of Econometrics*, 92(1):1–45.

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.

Eu, C. N. L. (2017). *Numerical Analysis in Nonlinear Least Squares Methods and Applications*. PhD thesis, Curtin University.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.

Farnsworth, H., Ferson, W., Jackson, D., and Todd, S. (2002). Performance evaluation with stochastic discount factors. *Journal of Business*, 75(3):473–503.

Franzoni, F., Nowak, E., and Phalippou, L. (2012). Private equity performance and liquidity risk. *The Journal of Finance*, 67(6):2341–2373.

Giommetti, N. and Jørgensen, R. (2021). Risk adjustment of private equity cash flows. http://dx.doi.org/10.2139/ssrn.3980121.

Gourieroux, C. and Monfort, A. (2007). Econometric specification of stochastic discount factor models. *Journal of Econometrics*, 136(2):509–530.

Gredil, O., Sorensen, M., and Waller, W. (2020). Evaluating private equity performance using stochastic discount factors. working paper (as of 2020-02-03).

Han, C. and Phillips, P. C. (2006). Gmm with many moment conditions. *Econometrica*, 74(1):147–192.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054.

Hansen, L. P. (2012). Proofs for large sample properties of generalized method of moments estimators. *Journal of Econometrics*, 170:325–330.

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.

Harris, R. S., Jenkinson, T., and Kaplan, S. N. (2014). Private equity performance: What do we know? *The Journal of Finance*, 69(5):1851–1882.

Hou, K., Xue, C., and Zhang, L. (2015). Digesting anomalies: An investment approach. *Review of Financial Studies*, 28(3):650–705.

Hou, K., Xue, C., and Zhang, L. (2020). Replicating anomalies. *The Review of Financial Studies*, 33(5):2019–2133.

Hüther, N., Schmid, L., and Steri, R. (2022). Credit market equivalents and the valuation of private firms. Unpulished working paper.

Jegadeesh, N., Noh, J., Pukthuanthong, K., Roll, R., and Wang, J. (2019). Empirical tests of asset pricing models with individual assets: Resolving the errors-in-variables bias in risk premium estimation. *Journal of Financial Economics*, 133(2):273–298.

Jenish, N. and Prucha, I. R. (2009). Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of Econometrics*, 150(1):86–98.

Jenish, N. and Prucha, I. R. (2012). On spatial processes and aysmptotic inference under near-epoch dependence. *Journal of Econometrics*, 170(1):178–190.

Kim, M. S. and Sun, Y. (2011). Spatial heteroskedasticity and autocorrelation consistent estimation of covariance matrix. *Journal of Econometrics*, 160(2):349–371.

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

Korteweg, A. and Nagel, S. (2024). Risk-adjusted returns of private equity funds: a new approach.

Leeb, H. and Pötscher, B. M. (2005). Model selection and inference: Facts and fiction. *Econometric Theory*, 21:21–59.

Li, H., Xu, Y., and Zhang, X. (2016). Hedge funds performance evaluation under the stochastic discount factor framework. *Journal of Financial and Quantitative Analysis*, 51(1):231–257.

Metrick, A. and Yasuda, A. (2010). *Venture capital and the finance of innovation.* John Wiley and Sons.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245.

Newey, W. K. and Windmeijer, F. (2009). Generalized method of moments with many weak moment conditions. *Econometrica*, 77(3):687–719.

Nocedal, J. and Wright, S. (2006). *Numerical optimization.* Springer Science & Business Media.

Pötscher, B. M. and Prucha, I. R. (1997). *Dynamic nonlinear econometric models: Asymptotic theory.* Springer Science & Business Media.

Pukthuanthong, K., Roll, R., and Subrahmanyam, A. (2019). A protocol for factor identification. *The Review of Financial Studies*, 32(4):1573–1607.

Racine, J. (2000). Consistent cross-validatory model-selection for dependent data: hv-block cross-validation. *Journal of Econometrics*, 99:39–61.

Raponi, V., Robotti, C., and Zaffaroni, P. (2020). Testing beta-pricing models using large cross-sections. *The Review of Financial Studies*, 33(6):2796–2842.

Taleb, N. N. (2020). Statistical consequences of fat tails: Real world preasymptotics, epistemology, and applications. *arXiv preprint arXiv:2001.10488.*

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

Tausch, C. and Pietz, M. (2024). Machine learning private equity returns. *The Journal of Finance and Data Science*, 10:100141.

MCM	Market **Loading**	Market **SE**	Market **SE Ind.**	Second **Name**	Second **Loading**	Second **SE**	Second **SE Ind.**
1	-614.990			Alpha	729.827		
30	1.760	1.776	0.810	Alpha	0.005	0.010	0.005
60	1.899	0.953	0.470	Alpha	0.003	0.001	0.000
120	1.576	21.343	9.443	Alpha	-0.001	0.139	0.062
150	1.511	28.882	12.761	Alpha	-0.002	0.198	0.086
180	1.555	9.033	3.913	Alpha	-0.003	0.073	0.030
1	1.770	0.628	0.343	EG	1.071	0.401	0.219
30	2.050	1.892	0.864	EG	0.628	1.116	0.523
60	2.075	4.400	2.096	EG	0.297	2.601	1.246
120	1.735	5.949	2.902	EG	-0.340	4.326	2.066
150	1.665	4.396	2.149	EG	-0.498	3.528	1.667
180	1.625	2.468	1.234	EG	-0.540	2.218	1.023
1	1.981	1.862	0.992	IA	0.879	3.282	1.713
30	2.181	2.067	1.085	IA	0.860	3.326	1.716
60	2.134	5.600	2.778	IA	0.368	8.684	4.306
120	1.629	5.579	2.813	IA	-0.697	10.840	5.417
150	1.501	1.543	0.821	IA	-0.940	2.500	1.251
180	1.445	1.663	0.886	IA	-0.995	2.487	1.246
1	1.929	0.979	0.496	ME	0.949	1.929	1.155
30	2.296	0.724	0.420	ME	1.400	3.010	1.443
60	2.422	0.480	0.358	ME	1.643	3.646	1.684
120	1.578	0.832	0.546	ME	1.318	2.606	1.789
150	1.379	0.873	0.677	ME	1.647	5.247	2.960
180	1.253	1.627	0.876	ME	1.683	7.103	3.692
1	1.767	0.990	0.522	MKT	1.767	0.990	0.522
30	1.899	1.048	0.531	MKT	1.899	1.048	0.531
60	2.042	1.108	0.557	MKT	2.042	1.108	0.557
120	1.505	0.943	0.517	MKT	1.505	0.943	0.517
150	1.239	0.908	0.511	MKT	1.239	0.908	0.511
180	1.115	0.890	0.498	MKT	1.115	0.890	0.498
1	1.843	0.908	0.471	ROE	1.331	0.819	0.420
30	2.093	1.596	0.773	ROE	0.960	1.325	0.648
60	2.082	13.672	6.664	ROE	0.319	10.764	5.247
120	1.743	4.481	2.252	ROE	-0.485	3.992	1.975
150	1.660	2.378	1.197	ROE	-0.660	2.535	1.176
180	1.594	2.103	1.102	ROE	-0.690	2.242	1.045

Table 10: Asymptotic results for PE Funds. This table reports asymptotic parameter estimates for linear two-factor models using Fund-Size Weighted Vintage-Year Portfolios. Standard errors are estimated by a SHAC estimator (Equation 16) with a Bartlett kernel bandwidth of $D = 12$ vintage years to correct for dependence between overlapping funds. MCM (Maximum Compounding Month) denotes the maximum cash flow compounding horizon.

MCM	Market **Loading**	Market **SE**	Market **SE Ind.**	Second **Name**	Second **Loading**	Second **SE**	Second **SE Ind.**
1	-284.624			Alpha	867.416		
30	1.613	1.984	0.884	Alpha	0.005	0.011	0.005
60	1.787	0.182	0.074	Alpha	0.004	0.001	0.001
120	1.550	0.860	0.358	Alpha	-0.000	0.004	0.002
150	1.527	29.159	12.176	Alpha	-0.002	0.198	0.081
180	1.562	9.786	4.003	Alpha	-0.003	0.081	0.032
1	1.588	0.579	0.317	EG	1.354	0.323	0.189
30	1.908	2.010	0.942	EG	0.657	1.241	0.593
60	1.995	4.611	2.188	EG	0.322	2.848	1.362
120	1.721	5.911	2.779	EG	-0.294	4.559	2.073
150	1.663	4.012	1.893	EG	-0.457	3.410	1.505
180	1.615	2.496	1.231	EG	-0.505	2.235	0.932
1	1.931	2.276	1.218	IA	0.919	4.061	2.142
30	2.119	2.900	1.525	IA	0.817	4.724	2.455
60	2.086	24.149	12.056	IA	0.347	38.176	19.060
120	1.623	3.598	1.778	IA	-0.688	7.106	3.379
150	1.508	1.413	0.759	IA	-0.934	2.570	1.153
180	1.448	1.346	0.749	IA	-1.002	2.525	1.134
1	1.900	0.954	0.502	ME	1.033	1.857	1.148
30	2.220	0.832	0.418	ME	1.246	2.659	1.317
60	2.288	0.669	0.382	ME	1.404	3.219	1.469
120	1.586	1.061	0.571	ME	1.079	1.672	1.189
150	1.405	0.840	0.627	ME	1.405	3.342	2.552
180	1.310	1.678	1.174	ME	1.591	8.674	5.240
1	1.709	0.949	0.510	MKT	1.709	0.949	0.510
30	1.891	1.013	0.538	MKT	1.891	1.013	0.538
60	2.020	1.085	0.548	MKT	2.020	1.085	0.548
120	1.537	0.902	0.493	MKT	1.537	0.902	0.493
150	1.314	0.933	0.539	MKT	1.314	0.933	0.539
180	1.190	0.900	0.531	MKT	1.190	0.900	0.531
1	1.744	0.773	0.473	ROE	1.617	0.692	0.415
30	1.977	1.945	0.983	ROE	0.952	1.686	0.852
60	2.021	10.651	5.332	ROE	0.372	8.933	4.466
120	1.732	4.125	2.005	ROE	-0.452	4.058	1.898
150	1.661	2.193	1.095	ROE	-0.634	2.476	1.085
180	1.594	2.131	1.087	ROE	-0.675	2.206	0.957

Table 11: Asymptotic results for BO Funds. This table reports asymptotic parameter estimates for linear two-factor models using Fund-Size Weighted Vintage-Year Portfolios. Standard errors are estimated by a SHAC estimator (Equation 16) with a Bartlett kernel bandwidth of $D = 12$ vintage years to correct for dependence between overlapping funds. MCM (Maximum Compounding Month) denotes the maximum cash flow compounding horizon.

MCM	Market **Loading**	Market **SE**	Market **SE Ind.**	Second **Name**	Second **Loading**	Second **SE**	Second **SE Ind.**
1	-614.990			Alpha	729.827		
30	1.726	1.866	0.823	Alpha	0.005	0.010	0.005
60	1.879	2.428	1.022	Alpha	0.004	0.013	0.006
120	1.563	21.164	8.925	Alpha	-0.000	0.143	0.061
150	1.511	0.442	0.188	Alpha	-0.002	0.002	0.001
180	1.540	9.693	3.997	Alpha	-0.003	0.081	0.032
1	1.691	0.539	0.295	EG	1.308	0.309	0.177
30	2.023	1.858	0.854	EG	0.670	1.109	0.521
60	2.077	4.183	1.977	EG	0.343	2.506	1.192
120	1.721	5.843	2.739	EG	-0.294	4.492	2.041
150	1.647	4.133	1.937	EG	-0.453	3.619	1.602
180	1.601	2.473	1.203	EG	-0.500	2.351	0.992
1	1.988	2.337	1.235	IA	0.856	4.066	2.118
30	2.203	2.687	1.361	IA	0.847	4.255	2.122
60	2.156	8.339	4.028	IA	0.393	12.881	6.216
120	1.626	3.425	1.701	IA	-0.662	6.531	3.169
150	1.496	1.412	0.754	IA	-0.904	2.583	1.166
180	1.436	1.382	0.743	IA	-0.968	2.544	1.147
1	1.952	0.980	0.489	ME	0.948	1.899	1.116
30	2.292	0.803	0.394	ME	1.221	2.755	1.339
60	2.368	0.590	0.359	ME	1.415	3.228	1.506
120	1.587	0.998	0.535	ME	1.106	1.609	1.129
150	1.396	1.045	1.066	ME	1.441	7.730	5.357
180	1.293	3.914	2.630	ME	1.577	15.957	10.092
1	1.770	0.966	0.512	MKT	1.770	0.966	0.512
30	1.940	1.045	0.527	MKT	1.940	1.045	0.527
60	2.070	1.104	0.552	MKT	2.070	1.104	0.552
120	1.536	0.876	0.485	MKT	1.536	0.876	0.485
150	1.296	0.905	0.536	MKT	1.296	0.905	0.536
180	1.173	0.887	0.521	MKT	1.173	0.887	0.521
1	1.824	0.785	0.459	ROE	1.491	0.714	0.407
30	2.088	1.734	0.855	ROE	0.988	1.451	0.718
60	2.096	6.020	2.998	ROE	0.396	4.976	2.471
120	1.733	4.032	1.954	ROE	-0.447	3.935	1.839
150	1.646	2.396	1.177	ROE	-0.624	2.823	1.259
180	1.580	2.096	1.054	ROE	-0.662	2.350	1.033

Table 12: Asymptotic results for GroBO Funds. This table reports asymptotic parameter estimates for linear two-factor models using Fund-Size Weighted Vintage-Year Portfolios. Standard errors are estimated by a SHAC estimator (Equation 16) with a Bartlett kernel bandwidth of $D = 12$ vintage years to correct for dependence between overlapping funds. MCM (Maximum Compounding Month) denotes the maximum cash flow compounding horizon.

MCM	Market **Loading**	Market **SE**	Market **SE Ind.**	Second **Name**	Second **Loading**	Second **SE**	Second **SE Ind.**
1	-166.932			Alpha	513.048		
30	2.131	2.859	1.374	Alpha	0.003	0.012	0.006
60	2.207	3.348	1.532	Alpha	0.002	0.013	0.007
120	2.396	20.431	8.107	Alpha	-0.004	0.105	0.040
150	2.483	9.617	3.742	Alpha	-0.006	0.055	0.020
180	2.601	6.104	2.431	Alpha	-0.008	0.041	0.014
1	1.782	0.398	0.243	EG	1.440	0.203	0.134
30	2.203	1.858	1.020	EG	0.494	0.952	0.540
60	2.227	2.435	1.251	EG	0.379	1.237	0.657
120	2.118	10.313	5.379	EG	-0.330	5.599	2.815
150	2.082	3.478	1.811	EG	-0.626	2.285	0.992
180	2.124	2.527	1.333	EG	-0.828	1.864	0.722
1	2.144	17.381	12.646	IA	0.343	17.772	13.076
30	2.392	7.679	4.938	IA	0.487	7.562	4.707
60	2.409	16.410	10.341	IA	0.344	15.354	9.521
120	1.938	1.933	1.231	IA	-0.936	1.295	0.919
150	1.763	1.480	0.986	IA	-1.360	0.978	0.677
180	1.707	1.439	0.955	IA	-1.657	0.737	0.413
1	2.093	0.817	0.453	ME	-0.479	2.923	2.145
30	2.394	0.587	0.352	ME	0.500	2.362	1.205
60	2.478	0.629	0.361	ME	0.742	2.304	1.182
120	1.940	1.121	0.759	ME	0.187	4.462	2.805
150	1.643	3.316	2.261	ME	-0.137	10.178	6.942
180	1.412	1.442	0.980	ME	-0.308	2.785	1.667
1	2.101	1.528	0.934	MKT	2.101	1.528	0.934
30	2.308	1.851	0.989	MKT	2.308	1.851	0.989
60	2.353	1.931	1.031	MKT	2.353	1.931	1.031
120	1.922	1.395	0.865	MKT	1.922	1.395	0.865
150	1.658	1.103	0.736	MKT	1.658	1.103	0.736
180	1.448	1.083	0.842	MKT	1.448	1.083	0.842
1	2.027	3.096	1.856	ROE	0.517	2.210	1.335
30	2.294	2.680	1.468	ROE	0.466	1.770	0.981
60	2.333	4.191	2.180	ROE	0.243	2.723	1.425
120	2.139	4.301	2.348	ROE	-0.614	3.109	1.593
150	2.053	2.562	1.434	ROE	-0.922	2.056	0.968
180	2.052	1.820	1.046	ROE	-1.148	1.579	0.647

Table 13: Asymptotic results for VC Funds. This table reports asymptotic parameter estimates for linear two-factor models using Fund-Size Weighted Vintage-Year Portfolios. Standard errors are estimated by a SHAC estimator (Equation 16) with a Bartlett kernel bandwidth of $D = 12$ vintage years to correct for dependence between overlapping funds. MCM (Maximum Compounding Month) denotes the maximum cash flow compounding horizon.

MCM	Market **Loading**	Market **SE**	Market **SE Ind.**	Second **Name**	Second **Loading**	Second **SE**	Second **SE Ind.**
1	-166.932			Alpha	513.048		
30	0.591	5.852	2.795	Alpha	0.012	0.038	0.018
60	0.536	10.957	5.928	Alpha	0.011	0.074	0.039
120	0.558	24.771	15.790	Alpha	0.007	0.159	0.101
150	0.448	0.218	0.144	Alpha	0.006	0.002	0.001
180	0.260	0.160	0.094	Alpha	0.005	0.001	0.001
1	2.063	1.046	0.568	EG	5.417	0.570	0.314
30	1.503	9.431	5.433	EG	1.286	10.426	6.119
60	1.224	15.941	9.496	EG	0.968	18.087	10.881
120	1.059	7.361	4.912	EG	0.603	10.456	7.110
150	0.851	4.221	3.362	EG	0.513	4.896	4.172
180	0.587	2.832	1.905	EG	0.498	2.439	2.344
1	1.601	4.063	3.175	IA	2.507	31.045	21.799
30	1.699	1.466	0.643	IA	1.241	4.324	4.139
60	1.462	1.227	0.617	IA	0.829	4.755	4.085
120	1.223	0.961	0.662	IA	0.558	7.836	5.161
150	1.021	0.882	0.601	IA	0.018	10.419	4.876
180	0.753	854.362	392.124	IA	-0.305	12257.183	5624.389
1	1.542	1.564	1.122	ME	2.765	4.318	1.897
30	1.716	1.231	0.894	ME	2.316	4.618	1.958
60	1.524	1.104	1.012	ME	2.164	4.712	2.274
120	1.338	1.661	1.411	ME	1.728	6.098	3.345
150	1.115	2.488	1.999	ME	1.613	7.149	4.618
180	0.833	123.732	89.688	ME	1.476	158.289	115.372
1	1.230	1.175	0.636	MKT	1.230	1.175	0.636
30	1.373	1.161	0.610	MKT	1.373	1.161	0.610
60	1.386	1.163	0.611	MKT	1.386	1.163	0.611
120	1.225	1.175	0.636	MKT	1.225	1.175	0.636
150	1.021	1.157	0.735	MKT	1.021	1.157	0.735
180	0.740	1.647	0.997	MKT	0.740	1.647	0.997
1	1.909	0.463	0.170	ROE	3.538	0.717	0.437
30	1.662	0.198	0.130	ROE	2.024	1.828	0.886
60	1.365	0.851	0.540	ROE	1.641	4.199	2.441
120	1.116	16.478	10.048	ROE	1.062	50.688	31.055
150	0.905	2.079	1.419	ROE	0.892	4.914	3.470
180	0.664	1.876	1.088	ROE	0.636	1.608	1.484

Table 14: Asymptotic results for RE Funds. This table reports asymptotic parameter estimates for linear two-factor models using Fund-Size Weighted Vintage-Year Portfolios. Standard errors are estimated by a SHAC estimator (Equation 16) with a Bartlett kernel bandwidth of $D = 12$ vintage years to correct for dependence between overlapping funds. MCM (Maximum Compounding Month) denotes the maximum cash flow compounding horizon.

MCM	Market **Loading**	Market **SE**	Market **SE Ind.**	Second **Name**	Second **Loading**	Second **SE**	Second **SE Ind.**
1	1.164	5.015	2.231	Alpha	0.001	0.036	0.017
30	1.051	9.753	4.299	Alpha	0.001	0.073	0.033
60	0.984	0.100	0.042	Alpha	0.000	0.000	0.000
120	0.790	6.656	2.617	Alpha	-0.000	0.108	0.044
150	0.804	7.028	2.756	Alpha	-0.000	0.114	0.046
180	0.804	6.995	2.743	Alpha	-0.000	0.113	0.046
1	1.226	3.992	1.753	EG	0.281	6.579	2.902
30	1.094	21.925	11.905	EG	0.284	28.164	15.171
60	0.967	16.551	10.687	EG	0.182	25.376	16.701
120	0.764	0.684	0.849	EG	-0.002	4.346	2.164
150	0.771	0.683	0.852	EG	-0.021	4.635	2.293
180	0.768	0.683	0.851	EG	-0.012	4.493	2.229
1	1.333	1.851	0.993	IA	0.509	5.769	4.609
30	1.158	1.618	1.253	IA	0.416	11.111	10.815
60	1.010	0.922	0.919	IA	0.352	3.824	4.135
120	0.782	16.204	20.589	IA	0.695	83.557	106.164
150	0.773	7.815	10.013	IA	0.676	39.938	51.169
180	0.768	6.230	8.014	IA	0.669	31.653	40.718
1	1.321	1.511	0.685	ME	0.519	0.725	1.079
30	1.147	1.116	0.719	ME	0.829	1.112	2.087
60	1.007	0.845	0.762	ME	0.947	1.352	2.582
120	0.765	1.018	1.335	ME	1.021	2.566	4.426
150	0.753	0.985	1.277	ME	0.941	2.625	4.195
180	0.747	0.960	1.232	ME	0.866	2.700	4.011
1	1.272	1.346	0.624	MKT	1.272	1.346	0.624
30	1.137	1.391	0.940	MKT	1.137	1.391	0.940
60	1.012	0.832	0.848	MKT	1.012	0.832	0.848
120	0.763	0.772	0.837	MKT	0.763	0.772	0.837
150	0.764	0.772	0.837	MKT	0.764	0.772	0.837
180	0.764	0.772	0.837	MKT	0.764	0.772	0.837
1	1.318	9.615	4.490	ROE	0.655	22.987	10.666
30	1.129	13.357	8.632	ROE	0.398	27.563	17.815
60	1.004	0.748	0.813	ROE	0.065	8.269	6.262
120	0.814	1.004	0.917	ROE	-0.274	6.789	3.267
150	0.829	0.815	0.851	ROE	-0.347	10.581	5.147
180	0.833	0.826	0.852	ROE	-0.368	9.608	4.639

Table 15: Asymptotic results for PD Funds. This table reports asymptotic parameter estimates for linear two-factor models using Fund-Size Weighted Vintage-Year Portfolios. Standard errors are estimated by a SHAC estimator (Equation 16) with a Bartlett kernel bandwidth of $D = 12$ vintage years to correct for dependence between overlapping funds. MCM (Maximum Compounding Month) denotes the maximum cash flow compounding horizon.

MCM	Market Loading	Market SE	Market SE Ind.	Second Name	Second Loading	Second SE	Second SE Ind.

Table 16: Asymptotic results for DD Funds. This table reports asymptotic parameter estimates for linear two-factor models using Fund-Size Weighted Vintage-Year Portfolios. Standard errors are estimated by a SHAC estimator (Equation 16) with a Bartlett kernel bandwidth of $D = 12$ vintage years to correct for dependence between overlapping funds. MCM (Maximum Compounding Month) denotes the maximum cash flow compounding horizon.

MCM	Market Loading	Market SE	Market SE Ind.	Second Name	Second Loading	Second SE	Second SE Ind.

Table 17: Asymptotic results for MEZZ Funds. This table reports asymptotic parameter estimates for linear two-factor models using Fund-Size Weighted Vintage-Year Portfolios. Standard errors are estimated by a SHAC estimator (Equation 16) with a Bartlett kernel bandwidth of $D = 12$ vintage years to correct for dependence between overlapping funds. MCM (Maximum Compounding Month) denotes the maximum cash flow compounding horizon.

MCM	Market Loading	Market SE	Market SE Ind.	Second Name	Second Loading	Second SE	Second SE Ind.

Table 18: Asymptotic results for FOF Funds. This table reports asymptotic parameter estimates for linear two-factor models using Fund-Size Weighted Vintage-Year Portfolios. Standard errors are estimated by a SHAC estimator (Equation 16) with a Bartlett kernel bandwidth of $D = 12$ vintage years to correct for dependence between overlapping funds. MCM (Maximum Compounding Month) denotes the maximum cash flow compounding horizon.

MCM	Market Loading	Market SE	Market SE Ind.	Second Name	Second Loading	Second SE	Second SE Ind.

Table 19: Asymptotic results for SEC Funds. This table reports asymptotic parameter estimates for linear two-factor models using Fund-Size Weighted Vintage-Year Portfolios. Standard errors are estimated by a SHAC estimator (Equation 16) with a Bartlett kernel bandwidth of $D = 12$ vintage years to correct for dependence between overlapping funds. MCM (Maximum Compounding Month) denotes the maximum cash flow compounding horizon.

MCM	Market **Loading**	Market **SE**	Second **Name**	Second **Loading**	Second **SE**	**CV-Error**
1	-570.184	141.688	Alpha	708.149	68.552	64.207
30	1.845	0.627	Alpha	0.004	0.004	444.452
60	1.941	0.642	Alpha	0.003	0.003	518.588
120	1.649	0.614	Alpha	-0.001	0.004	722.843
180	1.551	0.685	Alpha	-0.003	0.004	842.841
1	1.905	0.463	EG	1.187	1.532	401.096
30	2.122	0.442	EG	0.421	0.544	467.132
60	2.125	0.448	EG	0.163	0.427	554.380
120	1.780	0.338	EG	-0.341	0.373	735.911
180	1.632	0.289	EG	-0.488	0.328	848.835
1	2.068	0.469	IA	0.795	1.212	566.446
30	2.201	0.431	IA	0.477	1.063	645.960
60	2.166	0.425	IA	0.093	0.893	681.517
120	1.658	0.282	IA	-0.680	0.616	706.392
180	1.457	0.224	IA	-0.891	0.507	778.635
1	1.911	0.339	ME	0.621	1.011	559.787
30	2.242	0.416	ME	1.354	0.808	392.441
60	2.385	0.435	ME	1.730	0.546	415.122
120	1.595	0.307	ME	0.904	1.302	786.977
180	1.292	0.251	ME	1.899	0.385	682.764
1	1.776	0.248	MKT	1.776	0.248	292.435
30	1.932	0.328	MKT	1.932	0.328	349.625
60	2.053	0.362	MKT	2.053	0.362	414.842
120	1.491	0.258	MKT	1.491	0.258	546.106
180	1.101	0.211	MKT	1.101	0.211	663.749
1	1.893	0.350	ROE	0.909	0.905	407.286
30	2.148	0.425	ROE	0.668	0.812	466.802
60	2.133	0.421	ROE	0.183	0.597	568.490
120	1.785	0.315	ROE	-0.455	0.517	768.976
180	1.608	0.271	ROE	-0.591	0.485	944.079

Table 20: Cross-validation results for PE Funds. The table reports the validation error and average parameter estimates from *hv*-block cross-validation. The estimation uses Fund-Size Weighted Vintage-Year Portfolios. MCM (Maximum Compounding Month) denotes the maximum cash flow compounding horizon.

MCM	Market **Loading**	Market **SE**	Second **Name**	Second **Loading**	Second **SE**	**CV-Error**
1	-298.743	213.571	Alpha	703.615	137.853	76.367
30	1.642	0.622	Alpha	0.005	0.004	577.778
60	1.777	0.638	Alpha	0.003	0.003	681.805
120	1.571	0.601	Alpha	-0.000	0.004	910.315
180	1.516	0.704	Alpha	-0.002	0.005	1069.645
1	1.765	0.529	EG	1.745	1.838	450.811
30	1.952	0.420	EG	0.494	0.488	555.658
60	2.010	0.422	EG	0.221	0.374	676.421
120	1.738	0.315	EG	-0.290	0.350	931.469
180	1.623	0.297	EG	-0.463	0.322	1095.273
1	2.011	0.525	IA	0.900	1.262	783.770
30	2.095	0.424	IA	0.508	1.005	771.625
60	2.069	0.400	IA	0.125	0.803	808.204
120	1.625	0.261	IA	-0.671	0.561	894.285
180	1.458	0.223	IA	-0.930	0.449	979.862
1	1.865	0.366	ME	0.763	1.021	745.056
30	2.131	0.417	ME	1.159	1.037	629.642
60	2.222	0.413	ME	1.423	0.838	628.518
120	1.581	0.280	ME	0.790	1.192	944.497
180	1.291	0.226	ME	1.456	0.736	900.638
1	1.712	0.252	MKT	1.712	0.252	411.406
30	1.885	0.298	MKT	1.885	0.298	464.976
60	1.994	0.337	MKT	1.994	0.337	557.944
120	1.514	0.242	MKT	1.514	0.242	730.207
180	1.162	0.192	MKT	1.162	0.192	859.795
1	1.799	0.376	ROE	1.258	1.011	525.106
30	2.008	0.410	ROE	0.716	0.755	579.600
60	2.036	0.397	ROE	0.251	0.549	701.095
120	1.748	0.298	ROE	-0.422	0.487	957.961
180	1.609	0.284	ROE	-0.591	0.471	1172.411

Table 21: Cross-validation results for BO Funds. The table reports the validation error and average parameter estimates from *hv*-block cross-validation. The estimation uses Fund-Size Weighted Vintage-Year Portfolios. MCM (Maximum Compounding Month) denotes the maximum cash flow compounding horizon.

MCM	Market Loading	Market SE	Second Name	Second Loading	Second SE	CV-Error
1	-468.531	258.416	Alpha	695.895	73.404	74.480
30	1.751	0.665	Alpha	0.005	0.004	548.486
60	1.863	0.689	Alpha	0.003	0.003	655.800
120	1.580	0.626	Alpha	-0.001	0.004	875.507
180	1.493	0.728	Alpha	-0.002	0.005	1021.586
1	1.855	0.538	EG	1.748	1.889	409.289
30	2.062	0.463	EG	0.500	0.489	529.846
60	2.090	0.470	EG	0.239	0.379	657.233
120	1.742	0.335	EG	-0.290	0.352	895.890
180	1.609	0.312	EG	-0.454	0.327	1057.742
1	2.062	0.527	IA	0.815	1.209	719.449
30	2.183	0.461	IA	0.511	1.023	757.458
60	2.147	0.446	IA	0.152	0.828	802.224
120	1.633	0.284	IA	-0.646	0.578	868.400
180	1.447	0.242	IA	-0.887	0.479	965.232
1	1.911	0.384	ME	0.687	0.961	668.688
30	2.201	0.443	ME	1.137	0.968	574.661
60	2.301	0.449	ME	1.435	0.750	581.292
120	1.578	0.297	ME	0.823	1.085	859.362
180	1.277	0.235	ME	1.507	0.598	833.313
1	1.772	0.272	MKT	1.772	0.272	384.748
30	1.944	0.335	MKT	1.944	0.335	447.609
60	2.052	0.376	MKT	2.052	0.376	540.387
120	1.515	0.256	MKT	1.515	0.256	699.730
180	1.147	0.200	MKT	1.147	0.200	809.030
1	1.868	0.386	ROE	1.100	0.921	489.404
30	2.111	0.451	ROE	0.737	0.756	546.853
60	2.112	0.440	ROE	0.274	0.548	677.356
120	1.754	0.314	ROE	-0.417	0.484	919.884
180	1.597	0.295	ROE	-0.576	0.472	1134.021

Table 22: Cross-validation results for GroBO Funds. The table reports the validation error and average parameter estimates from *hv*-block cross-validation. The estimation uses Fund-Size Weighted Vintage-Year Portfolios. MCM (Maximum Compounding Month) denotes the maximum cash flow compounding horizon.

MCM	Market Loading	Market SE	Second Name	Second Loading	Second SE	CV-Error
1	-228.376	221.884	Alpha	566.708	91.779	15.102
30	2.228	0.328	Alpha	0.002	0.003	100.596
60	2.333	0.419	Alpha	0.001	0.003	133.168
120	2.495	0.286	Alpha	-0.005	0.002	171.080
180	2.679	0.262	Alpha	-0.009	0.001	199.361
1	1.992	0.672	EG	2.002	1.944	86.072
30	2.272	0.476	EG	0.343	0.483	110.729
60	2.328	0.454	EG	0.164	0.508	149.453
120	2.203	0.288	EG	-0.449	0.242	179.057
180	2.219	0.281	EG	-0.896	0.114	210.453
1	2.224	0.408	IA	0.442	1.549	131.725
30	2.447	0.428	IA	0.165	1.193	156.017
60	2.417	0.357	IA	-0.081	1.065	180.724
120	1.957	0.229	IA	-1.151	0.552	178.398
180	1.767	0.232	IA	-1.783	0.322	198.562
1	2.123	0.260	ME	-0.572	0.572	89.808
30	2.429	0.352	ME	0.408	0.526	91.271
60	2.508	0.331	ME	0.867	0.143	111.286
120	1.974	0.275	ME	0.523	0.352	176.732
180	1.416	0.268	ME	0.208	0.998	224.945
1	2.112	0.263	MKT	2.112	0.263	70.890
30	2.317	0.271	MKT	2.317	0.271	83.780
60	2.345	0.275	MKT	2.345	0.275	105.958
120	1.901	0.243	MKT	1.901	0.243	164.143
180	1.373	0.189	MKT	1.373	0.189	179.872
1	2.063	0.409	ROE	0.336	0.877	92.850
30	2.337	0.392	ROE	0.282	0.643	110.977
60	2.403	0.344	ROE	-0.012	0.611	143.661
120	2.202	0.261	ROE	-0.746	0.285	172.333
180	2.133	0.268	ROE	-1.215	0.114	205.599

Table 23: Cross-validation results for VC Funds. The table reports the validation error and average parameter estimates from *hv*-block cross-validation. The estimation uses Fund-Size Weighted Vintage-Year Portfolios. MCM (Maximum Compounding Month) denotes the maximum cash flow compounding horizon.

MCM	Market **Loading**	Market **SE**	Second **Name**	Second **Loading**	Second **SE**	**CV-Error**
1	-257.053	353.090	Alpha	565.145	349.294	388.581
30	0.902	0.661	Alpha	0.008	0.008	543.876
60	0.855	0.677	Alpha	0.007	0.008	573.035
120	0.960	0.705	Alpha	0.004	0.007	688.527
180	0.651	0.907	Alpha	0.002	0.007	680.871
1	2.838	2.162	EG	5.118	4.605	2916.965
30	1.620	0.210	EG	1.008	1.001	526.431
60	1.353	0.190	EG	0.732	0.854	553.360
120	1.198	0.243	EG	0.406	0.623	566.432
180	0.756	0.396	EG	0.253	0.638	592.652
1	1.501	0.362	IA	2.027	2.331	812.315
30	1.636	0.233	IA	0.608	1.373	689.651
60	1.439	0.150	IA	0.368	1.145	547.766
120	1.275	0.243	IA	0.127	1.228	581.677
180	0.826	0.276	IA	-0.613	1.212	586.761
1	1.440	0.264	ME	2.041	1.773	524.857
30	1.593	0.221	ME	1.606	1.820	735.970
60	1.470	0.138	ME	1.369	1.703	653.901
120	1.401	0.260	ME	1.046	1.788	740.613
180	0.991	0.335	ME	0.944	1.799	711.253
1	1.226	0.099	MKT	1.226	0.099	286.364
30	1.379	0.180	MKT	1.379	0.180	291.563
60	1.381	0.112	MKT	1.381	0.112	330.875
120	1.247	0.191	MKT	1.247	0.191	406.365
180	0.779	0.226	MKT	0.779	0.226	465.965
1	2.053	0.815	ROE	2.928	2.116	408.173
30	1.748	0.265	ROE	1.547	1.445	531.495
60	1.479	0.162	ROE	1.249	1.276	561.301
120	1.235	0.229	ROE	0.787	0.836	547.524
180	0.760	0.272	ROE	0.391	0.596	529.247

Table 24: Cross-validation results for RE Funds. The table reports the validation error and average parameter estimates from *hv*-block cross-validation. The estimation uses Fund-Size Weighted Vintage-Year Portfolios. MCM (Maximum Compounding Month) denotes the maximum cash flow compounding horizon.

MCM	Market **Loading**	Market **SE**	Second **Name**	Second **Loading**	Second **SE**	**CV-Error**
1	-102.521	276.790	Alpha	249.752	410.786	244.498
30	1.228	0.393	Alpha	0.001	0.003	281.676
60	1.193	0.401	Alpha	0.000	0.003	328.282
120	1.021	0.320	Alpha	-0.001	0.002	430.718
180	1.056	0.335	Alpha	-0.002	0.002	463.528
1	1.513	0.472	EG	0.257	0.607	307.818
30	1.295	0.402	EG	0.278	0.388	262.234
60	1.171	0.415	EG	0.141	0.266	302.329
120	0.960	0.342	EG	-0.106	0.258	406.895
180	0.951	0.324	EG	-0.143	0.238	424.626
1	1.867	1.058	IA	1.125	1.889	448.152
30	1.390	0.510	IA	0.378	0.878	321.203
60	1.237	0.507	IA	0.330	0.581	344.323
120	0.934	0.343	IA	0.601	0.212	374.724
180	0.893	0.297	IA	0.579	0.205	384.034
1	1.602	0.503	ME	0.492	0.985	349.746
30	1.373	0.500	ME	0.875	0.602	257.392
60	1.281	0.595	ME	1.090	0.369	312.063
120	0.954	0.417	ME	1.109	0.313	382.511
180	0.890	0.334	ME	0.922	0.238	386.511
1	1.353	0.203	MKT	1.353	0.203	189.868
30	1.243	0.249	MKT	1.243	0.249	222.339
60	1.177	0.371	MKT	1.177	0.371	283.094
120	0.925	0.357	MKT	0.925	0.357	391.182
180	0.896	0.315	MKT	0.896	0.315	404.971
1	1.682	0.704	ROE	1.055	1.677	242.145
30	1.358	0.450	ROE	0.428	0.655	267.605
60	1.237	0.439	ROE	0.061	0.542	331.359
120	0.993	0.309	ROE	-0.319	0.469	427.041
180	0.984	0.280	ROE	-0.423	0.468	435.017

Table 25: Cross-validation results for PD Funds. The table reports the validation error and average parameter estimates from *hv*-block cross-validation. The estimation uses Fund-Size Weighted Vintage-Year Portfolios. MCM (Maximum Compounding Month) denotes the maximum cash flow compounding horizon.

MCM	Market **Loading**	Market **SE**	Second **Name**	Second **Loading**	Second **SE**	**CV-Error**

Table 26: Cross-validation results for DD Funds. The table reports the validation error and average parameter estimates from hv-block cross-validation. The estimation uses Fund-Size Weighted Vintage-Year Portfolios. MCM (Maximum Compounding Month) denotes the maximum cash flow compounding horizon.

MCM	Market **Loading**	Market **SE**	Second **Name**	Second **Loading**	Second **SE**	**CV-Error**

Table 27: Cross-validation results for MEZZ Funds. The table reports the validation error and average parameter estimates from hv-block cross-validation. The estimation uses Fund-Size Weighted Vintage-Year Portfolios. MCM (Maximum Compounding Month) denotes the maximum cash flow compounding horizon.

MCM	Market **Loading**	Market **SE**	Second **Name**	Second **Loading**	Second **SE**	**CV-Error**

Table 28: Cross-validation results for FOF Funds. The table reports the validation error and average parameter estimates from hv-block cross-validation. The estimation uses Fund-Size Weighted Vintage-Year Portfolios. MCM (Maximum Compounding Month) denotes the maximum cash flow compounding horizon.

MCM	Market **Loading**	Market **SE**	Second **Name**	Second **Loading**	Second **SE**	**CV-Error**

Table 29: Cross-validation results for SEC Funds. The table reports the validation error and average parameter estimates from hv-block cross-validation. The estimation uses Fund-Size Weighted Vintage-Year Portfolios. MCM (Maximum Compounding Month) denotes the maximum cash flow compounding horizon.