Nowcasting Net Asset Values: The Case of Private Equity*

Gregory W. Brown† Eric Ghysels‡ Oleg Gredil§

First draft – June, 2019
This version – December 27, 2019

*We are grateful to the Private Equity Research Consortium and the Institute for Private Capital for support and data access. We thank Mike Aguilar, Barry Griffiths, Ian Roberts, Erkko Etula, Farshid Asl, Wendy Hu as well as seminar participants at Chapel Hill and Oxford Private Equity Research Symposiums, Goldman Sachs SQAA group.

†University of North Carolina at Chapel Hill, Kenan-Flagler Business School and Frank Hawkins Kenan Institute of Private Enterprise, gregwbrown@unc.edu, (919) 962-9250

‡Department of Finance, Kenan-Flagler Business School, Department of Economics, University of North Carolina at Chapel Hill, Frank Hawkins Kenan Institute of Private Enterprise, and Centre for Economic Policy Research (CEPR), eghysels@unc.edu, (919) 962-9810

§Tulane University, Freeman School of Business, ogredil@tulane.edu, (504) 314-7567

Nowcasting Net Asset Values: The Case of Private Equity

Abstract

We apply advances in analysis of mix frequency and sparse data to estimate "unsmoothed" private equity (PE) Net Asset Values (NAVs) at the weekly frequency for individual funds. Using simulations and a large sample of buyout and venture funds, we show that our method yields superior estimates of fund asset values than a simple approach based on comparable public asset and as-reported NAVs. Our method easily accommodates additional data on PE fund portfolios, such as individual holdings, relevant mergers and acquisitions, secondary trades with fund stakes. The method is easily extended to other illiquid portfolios that are subject to appraisal bias while generating irregular and infrequent cash flows. We find significant variation in systematic and idiosyncratic risk exposures across PE funds and through time. In particular, the risk-return profile based on the samples from the 1990s is not representative of currently operating funds.

Keywords: Private Equity, Venture Capital, Leveraged Buyouts, Institutional Investors

JEL Classification: G23, G24, G30

1 Introduction

Valuing illiquid assets is hard but often necessary to provide critical input for investment decisions.[1] Private equity (PE) investments are a prime example of such a setting whereby various stakeholders have to rely on infrequently and strategically self-reported Net Asset Values (NAVs) by fund managers while the secondary markets for most fund stakes are relatively undeveloped and likely reflect the marginal utility to trade of an unrepresentative investor.[2] However, at least for some PE funds, one has many observations that are plausibly informative of the market value of the fund portfolio and updated at higher than quarterly frequencies. These may include characteristics of individual assets held, comparable private transactions, public returns, the history of fund cash flows, NAV reports and occasional secondary trades, etc. This paper develops a method of how to use these types of data *jointly* in a unified statistical framework to (i) learn about risk, return, and reporting quality characteristics of individual funds, and (ii) to estimate unbiased asset values at the relevant data arrival frequency (such as weekly or daily)—i.e., to *nowcast* PE fund NAVs. Nowcasting is the prediction of the present, the very near future and the very recent past in economics. We expand on the methods used in the literature as nowcasting NAVs is more complex than macroeconomic variables like GDP growth (the most studied example).

We model fund "true" asset values and, correspondingly, "true" returns as the latent factor in a sparse-data State Space Model (SSM) estimated at high frequencies (relative to quarterly fund reporting) at an individual PE fund level. Our setup relies on the well-developed SSM machinery (see, e.g. Durbin and Koopman, 2012) but allows incorporating salient properties of asset return such as heteroskedastic volatility. We also show that our results are robust to autocorrelation in the unobserved error terms and the fund idiosyncratic returns not being conditionally Gaussian. Most importantly, our approach allows for evaluation of the model performance not only via simulated samples but also using real PE

[1] See Brito (1977); Longstaff (2001); Gârleanu (2009); Ang et al. (2014) among others.

[2] See Ewens et al. (2013); Brown et al. (2019) and Nadauld et al. (2016).

fund data arriving after the model estimation period. In particular, we evaluate our models by examining out-of-sample realized cash flows versus the levels implied by the model. This provides additional confidence in the SSM parameters that we estimate for each of 1,745 funds, or 62% of our sample, independently. For 65% to 80% (depending on the metric) of these funds the model outperforms a naïve but intuitive approach whereby reported NAVs are extrapolated using cash flows and returns of a comparable public asset.

We show that, with partial imputation of auxiliary parameters using peer funds, we can estimate fund-level SSMs for over 94% of the sample funds with a 73-86% improvement rate relatively to the naïve approach. The peer funds are defined as having the same strategy and/or industry, a vintage within one year and having the same size tercile as the estimand fund. Our performance assessment metrics are based on the idea that if the fund's true returns were observed and used to discount fund cash flows (and the true net asset values), the fund to-date PMEs (Kaplan and Schoar, 2005) computed using these series would be equal to one. Furthermore, we show that, given fund-level SSM estimates, the in-sample performance is a good predictor of the out-of-sample performance, yielding useful criteria for model selection and forecast combination. Using our estimate of fund NAVs reduces the out-of-sample forecast error of subsequent cash flows by 55-58% relative to the as-reported NAVs of a typical fund, with filtered weekly returns featuring zero autocorrelation and realistic standard deviations (e.g., 33% to 38% per year).

There are two pillars to our approach. One is that we use the individual fund-level data on cash flows and NAVs *simultaneously* to identify how NAVs are "smoothed" while jointly estimating the fund's systematic and idiosyncratic risk exposures. This is different from existing work featuring the complete fund-level data that either (i) relies on reported NAVs and attempts to remove autocorrelation induced by smoothing via a distributed lag market-model/ panel-AR(p) (Woodward and Hall, 2004; Ewens et al., 2013; Goetzmann et al., 2018), or (ii) disregards the NAV information altogether and relies solely on funds' realized cash flow data (see, inter alia, Driessen et al., 2012; Franzoni et al., 2012; Buchner and Stucke, 2014;

Ang et al., 2018). For a typical fund with say 40 NAV reports and 25 distributions over a 10-year life we gain 22 degrees of freedom as we only need to estimate at most 3 additional parameters. The second pillar is that we unsmooth weekly returns cumulative since a fund's inception (rather than per quarter returns, see also Ang 2014).[3] This is more consistent with microfoundations for the smoothing bias whereby appraisers look back for relevant events *within* the past quarter to arrive at their valuation assessment.

A key challenge for studies featuring deal-level datasets has been addressing selection bias in observed deal realizations (typically successful ones) relative to the overall pool of investments (Cochrane, 2005; Korteweg and Sorensen, 2010). This challenge is mostly resolved in our framework since we use the complete history of fund-level investments. In our analysis, deal-level data merely help to better identify the fund's structural risk–return parameters and the inference about the cumulative return path. In other words, deal-level data significantly reduce noise for fund SSM estimates and potentially reduce the number of parameters that require estimation but do not induce selection bias. However, even without portfolio company data, model performance is improved by augmenting the analysis with high frequency comparable benchmark returns.[4]

The PE fund sample for this study is provided by Burgiss (see Brown et al., 2015, for description) and consists of fund-level cash flows between each fund and its investors, fund NAV reports, as well as data on fund strategy, vintage year, and industry. We do not observe the information on the fund individual holdings but demonstrate in simulations that SSM precision improves by a factor of 2-3 for a typical fund even when keeping the comparable asset match quality unchanged. Observing fund portfolio-level holdings and their contribution to fund-level NAVs allows us to compute (rather than estimate jointly with other parameters) the fund distribution density trend and its variance. Utilizing real-

[3] Note that "detectability" of the smoothing at quarterly frequency implies its "detectability" at weekly frequency, but not vice verse. For a weekly smoothing at 0.70-rate, quarterly AR(1) is 0.01 but the variance is only 40% of the unsmoothed series. Thus, regressions at low frequencies never sufficiently unsmooth NAVs.

[4] This occurs even though it requires three additional parameters to estimate (or impute)—the intercept, the slope, and the variance of the comparable asset relatively to high frequency fund return estimates.

time fund holdings data has additional benefits: (i) the construction of a more informative comparable public benchmark and (ii) incorporating valuations in recent M&A deals, both public and private, that feature targets similar to the fund current holdings.

In addition to our methodological contribution, we generate important new findings about fund-level risk and return. For example, we find fund-level systematic risk (beta) to be lower than many other recent studies. The average (median) is around 1.05 (1.09) for buyout funds and 1.20 (1.18) for venture funds, as opposed to averages around 1.25 and 1.80, respectively, in Ang et al. (2018) even though we used their estimates from that study to set up our profile grid for beta estimates.

Part of this difference results from the fact that most studies do not actually estimate the market beta of a fund over its life time but rather the beta of the panel of time-aggregated fund cash flows. The distinction is important when cash flows from funds are clustered in time and contribute greater weight to the panel (e.g. venture in late 90s). Panel-based estimates generally will not reflect the beta of an average fund in this case. We provide regression analyses to support this claim. Moreover, we show that "nudging estimation" to feature higher systematic risk exposures, as well as ignoring the heterogeneity in other parameters in the cross section of funds, results in inferior nowcasting performance of the fund-level SSMs, in-sample and out-of-sample. In other words, higher systematic risk assumptions are less consistent with the fund-level cash flow realizations.

We find notable variation in systematic and idiosyncratic risk exposures across vintage years. While we do confirm betas in the vicinity of two for half of venture funds incepted during late 90s, three-quarters of venture funds incepted after 2004 have beta estimate of less than 1.4. However, the idiosyncratic risk of venture funds—which we measure as a multiple of GARCH(1,1)-filtered volatility of matching industry excess returns—has rebounded to a median of 3× from the 2× lows during 2000–2002. As far as temporal trends in buyout funds, we note an increase in systematic risk relatively to the levels in the 90s, and a spike in idiosyncratic risk during 2003-2005. The trends in PE fund abnormal returns are broadly consistent with those documented in prior literature.

While, in principle, our method supports pooling cash flows and NAV data across funds, we find that this actually complicates the analysis for a few reasons. First, unlike the cash flow methods described above, our method does not require a balanced panel of capital calls and distributions over different calendar time intervals to correlate with the sample of risk factor realizations. In fact, such a pooling implies strong assumptions—e.g., that funds' systematic and idiosyncratic risk exposures are exogenous from the risk factor realizations. Second, we find that averaging several funds with different data coverage quality tends to result in less precise estimates. Instead, it is more efficient and transparent to impute parameters that are not well-identifiable for a particular fund with peer-based aggregates or, even, theoretically motivated values. 'Better data' in our context means (i) distributions spanning multiple quarters, (ii) many NAV reports with variation from the previous quarter, and (iii) a substantially realized fund. This allows us to identify the fund risk exposures from the appraisal bias (i.e. NAV smoothing). However, provided that there are at least some meaningful distributions and consistency in NAV reports, we can still recover the fund-specific systematic risk, average abnormal return, and "unsmoothed" NAV estimates by treating some parameters as fixed.

Our paper contributes to recent studies that aim at developing "matrix pricing" for the portfolios of PE funds using either secondary market trades in fund stakes coupled with a selection model as in Boyer et al. (2018) or a replication of the PE fund returns with holding-level matches to public comparable firms as in Stafford (2017). Our methodology effectively nests both approaches while allowing for fund-level heterogeneity in risk exposures and taking a more structural approach to modeling the fund return process during the periods in which the data are not observable.[5] We also complement contemporaneous work by Gupta and Van Nieuwerburgh (2019), who use machine learning tools to estimate multi-factor time-varying exposures from a panel of PE fund cash flows, as opposed to the

[5] Matching to the transaction-level public benchmarks is embedded in the construction of fund's comparable asset, utilizing secondary price data amounts to augmenting the observations vector—see Section A.2.1.

bottom-up approach that matches on portfolio companies' characteristics. We note that both approaches can be used to construct the comparable public asset for our method.

With regards to methods, our study relates to the nowcasting macroeconomics literature which relies on SSM methods for the purpose of real-time monitoring of real activity (in particular GDP growth, see for example the survey of Bańbura et al., 2013), and the literature on forecast combination for the purpose of nowcasting (see, e.g. Andreou et al., 2013; Kuzin et al., 2013). The closest to our context application of the Kalman filter is in Korteweg and Sorensen (2010) who use venture-backed company valuation over several financing rounds and filter the latent value process in-between those rounds.

2 Nowcasting Model

Nowcasting is a term coined in meteorology and pertains to the prediction of the present, the very near future and the very recent past. It has recently become popular in economics as a method for assessing the current state of an economy. For example, official gross domestic product (GDP) estimates are only determined after a long delay (and are even then subject to subsequent revisions) and nowcasting methods can be utilized to generate more timely estimates for policy and investment decisions.[6] Nowcasting is intrinsically a mixed frequency data problem as the object of interest is a low frequency data series - observed say quarterly - whereas more real-time information (daily, weekly or monthly) during the quarter can be used to assess and potentially continuously update the state of the low frequency series, or put differently, *nowcast* the series of interest.

There are two types of approaches. The first involves state space models featuring latent variables, whereas others are based on regression models.[7] The fact that a critical part of our data, i.e. fund cash flows, is observed irregularly, directs our choice towards the SSM

[6] See Ghysels et al. (2017) for a recent discussion on data revisions and its impact on financial markets.

[7] The SSMs are estimated via maximum likelihood and using Kalman filtering or smoothing to nowcast the series of interest. See Nunes (2005) and Giannone et al. (2008) for early examples. In Appendix A.1 we provide a short intro to SSM. The second type of models involves MIDAS regressions, originated by Ghysels et al. (2006), which are regression models designed to handle data sampled at different frequencies.

approach. We note that nowcasting can be performed without specifying a model. For example, PE practitioners often assume that the value of fund assets is equal to the manager as-reported net asset values on quarter-ends and use a public benchmark return and fund interim cash flows to interpolate values for dates of interest inside the quarter or extrapolate beyond the date of latest reported net asset value. This is a reasonable approach but it makes some very strong assumptions. We will therefore refer to it as a *Naïve nowcast* of fund asset values and use it to evaluate the performance of our model-based nowcasts.[8]

2.1 SSM Representation of a PE Fund

We start with the characterization of the return and cash flow generating processes for a finite-life (usually around 10 years) PE fund. We assume that time t pertains to a weekly sampling scheme and use the following notations:

- C_t – week t cash flows from investors to the fund (a.k.a., Capital Calls),
- NAV_t – fund net asset values reported at the end of week t,
- D_t – fund distributions to its investors on week t,
- R_{mt} – the systematic risk factor gross return during week t,
- R_{ct} – the average gross return on publicly traded assets comparable to fund in week t,
- h_t – the variance of idiosyncratic return on assets similar to those held by the fund t,
- R_t – the gross return on fund asset for week t,
- V_t – the true value of fund assets at the end of week t.

Note that some of these series are latent. For example, the V_t series are not observed and, hence, neither are the R_t series. Other series are often missing at the weekly sampling frequency, for example NAV_t is reported by the fund GPs approximately every 13 weeks.

We assume that R_{mt} and R_{ct} are observed (i.e. not missing) for every week t, and focus on a case of scalar R_{ct}—i.e., the relevant asset returns are aggregated to a single index.

[8] Formally this approach is given by equation A.10 in Appendix, also referred to as "R_{ct}-interpolated NAVs". We also show that it corresponds to a particular parameterization of the fund-level SSM.

Accordingly, we proxy for h_t with a GARCH(1,1)-filtered variance of the idiosyncratic returns of R_{ct} (from projection on R_{mt}).

Throughout the paper, we use capital letters to denote the levels of variables (e.g., D_t for distribution amount) and lowercase to denote the natural logarithm thereof ($r_t = log(R_t)$, $d_t = log(D_t)$, etc.). A lower case letter followed by $(\cdot)$ refers to scalar-valued function, e.g. involving data, such as $\lambda(\cdot)_t$ pertaining to λ at time t.

2.2 Model Structure and Parameters

We assume the following relations describe the fund return- and cash flow processes:

$$R_t = \Big(\alpha + \beta(R_{mt} - 1) + 1\Big)e^{\eta_t}, \qquad \eta_t \sim \mathcal{N}(0, F^2 \cdot h_t), \tag{1}$$

$$R_{ct} = e^{r_t \cdot \beta_c + \psi + \eta_{ct}}, \qquad \eta_{ct} \sim \mathcal{N}(0, F_c^2 \cdot h_t), \tag{2}$$

$$V_t = V_{t-1}R_t - D_t + C_t, \qquad V_0 = C_0 - D_0 > 0 \tag{3}$$

$$R_{0:t} = \prod_{\tau=1}^{t} R_\tau \equiv V_t \cdot M_t, \tag{4}$$

$$\bar{r}_{0:t} = \Big(1 - \lambda(\cdot)_t\Big)r_{0:t} + \lambda(\cdot)_t \bar{r}_{0:t-1}, \tag{5}$$

$$\text{NAV}_t = e^{\bar{r}_{0:t} - m_t + \epsilon_{nt}}, \qquad \epsilon_{nt} \sim \mathcal{N}(0, \sigma_n^2) \tag{6}$$

$$D_t = \delta(\cdot)_t \Big(V_t + D_t\Big)e^{\epsilon_{dt}} \quad \text{iff } D_t > 0, \qquad \epsilon_{dt} \sim \mathcal{N}(0, \sigma_d^2), \tag{7}$$

governed by the following set of parameters collectively referred to as θ:

- β — fund systematic risk exposure;
- α — the fund average return per week unexplained by the systematic risk;
- β_c and ψ — respectively, the slope and intercept from a regression of the comparable asset weekly log returns on those of the fund;
- F and F_c — the scales for the conditional variance h_t of distributions from which the innovations in returns, η_t and η_{ct}, are independently drawn from;
- λ — the structural parameter(s) of the fund appraisal bias function $\lambda(\cdot)_t$;

- δ — the structural parameter(s) of the fund distribution rule $\delta(\cdot)_t$ conditional on a distribution event on week t;
- σ_n and σ_d – respectively, the standard deviations on the NAV reporting noise and distribution rule.

Applying log transformations to equations (1) through (7), except for (5) which is already in log form, yields a SSM as detailed in Appendix Section A.2.[9] As follows from equations (1) and (5) NAV appraisal bias is modeled as pertaining to both, systematic and idiosyncratic, components of fund returns. Also, because t indexes weeks, the model accounts for returns *between* the quarterly report dates. Accordingly, the predictable component of the appraisal bias is represented by the deterministic function $\lambda(\cdot)_t$ which, in our baseline analysis, we assume to have the following form:

$$\lambda(\cdot)_t := \lambda \cdot (1 - w_t) \quad , \tag{8}$$

where $\lambda \in (.01, 0.99)$ is a scalar parameter reflecting the exponential moving average weight and $w_t \in [0, 1]$ are the observed fractions of the fund gross assets that the cash flow in period t comprises in the *naïve nowcast* of fund asset value. This functional form implies a "first-in-first-out" assumption about the fund investment–divestment cycle. All else equal, funds that made substantial distributions recently would be assumed to have "less stale" NAVs. Meanwhile, the weight of returns occurring before period $\tau (0 < \tau < t)$ in NAV_t is limited by the fraction of fund capital invested before period τ.

Note that, because equation (5) pertains to cumulative (i.e., since-inception) returns of the fund—$R_{0:t}$, we actually model a smoothing of NAVs rather than smoothing of periodic returns. This is a particularly relevant perspective as it allows us to interpret the reported NAVs as a weighted-average of valuation snapshots taken at different points in the past. Furthermore, as follows from the equation, most of the weight may come from weeks *within* the past quarter (rather than from ending values of the previous quarters) corresponding

[9] In addition, Appendix A.2.1 contains examples providing further intuition and discusses generalizations (e.g., incorporating longer-term or event-specific "memory"). Likewise, Appendix A.2.2 provides further generalizations (e.g., incorporating PE secondary market prices).

to a realistic assumption about how the actual appraisal of fund assets is performed. This assumption distinguishes our approach from the extant literature that attempts to unsmooth quarterly returns inferred from NAV changes.[10] However, since PE funds normally incur interim capital calls and distributions, cumulative returns are generally not equal to the asset values (scaled by initial investment). Therefore we introduce a mapping function—M_t (which m_t is a log of), implicitly defined by equations (4) and (6). It is a step-function which accounts for the history of cash flows and valuations and is further discussed in Section 2.4. For the distribution rule $\delta(\cdot)_t$, we assume:

$$\delta(\cdot)_t := \min(0.99,\, \delta \cdot t_y) \quad , \tag{9}$$

where $\delta \in (0, 1)$ is a scalar parameter, whereas t_y is the time (in years) elapsed since the fund inception. Constraining the domain of $\delta(\cdot)_t$ has clear economic intuition—until it is known that the fund is resolved, it has to be that D_t is non-negative and $V_t > \delta(\cdot)_t(V_t + D_t)e^{\epsilon_{dt}}$.

The economic interpretation for δ and σ_d is, respectively, the trend and the noise of the distribution density. Given equal number of distribution and intervals between them, funds with higher δ will tend to return capital faster. Overall there is relatively little insight that can be learned from $(\delta,\, \sigma_d)$. One can think of these as just auxiliary parameters relating distributions—which reveal the true (i.e., unsmoothed value) of the parts of the fund portfolio—and the reported NAVs, therefore allowing for inference about λ and σ_n). If the fund individual investments' valuation reports and exit proceeds were observed, then $\delta(\cdot)_t$ need not be estimated. Instead it could be computed as a fraction of the previous quarter portfolio-level NAVs that the sold assets had comprised. Furthermore, under the assumption that the excess return (relative to the overall fund portfolio) is zero during the exit quarter, ϵ_{dt} is also zero, which makes σ_d a redundant parameter too. In that case, $d_t + m_t$ maps without uncertainty to the true to-date return $r_{0:t}$.

[10]See e.g. Ewens et al. (2013); Goetzmann and Phalippou (2019). Appendix shows that unsmoothing of the intra-quarter appraisal bias cannot be done using the inter-quarter returns only.

2.3 Other Estimands of Interest

Given the data observed for week $t = 1, ..., T$ and parameters θ, we can apply the Kalman filter to obtain estimates of fund returns over each week.[11] We then apply the mapping function M_t in equation (4) to obtain the estimate of the fund asset values. We use the following notation to emphasize that these estimates are functions of a particular θ and the ending period T: $\hat{R}_t^{\theta,T}$, $\hat{R}_{\tau:t}^{\theta,T}$, and $\hat{V}_t^{\theta,T}$, which are conditional on θ and data up to period T. Note that the parameters in θ are not necessarily obtained through econometric estimation. As discussed later, they can also be determined via imputations, or values taken from the prior literature.

In some cases the returns and asset value nowcasts do not depend on whether one can actually estimate the parameter vector θ. For example, the *naïve nowcast* that we described earlier is given by a θ in which $\beta_c = 1$, $F_c = \sigma_n = 0$, and (α, β, F) are given from a regression of R_{ct} on R_{mt}. In other words, the filtering step is completely independent from the parameter estimation step that we discuss in the next section.

For each fund and θ of interest, we compute:

1. Variances and autocorrelations of the filtered weekly returns between $t = 1$ and $t = T$ using the standard estimators at weekly and quarterly frequency;

2. $PME(\theta, T)_{0:t}$ Kaplan and Schoar PMEs on a to-date basis (Brown et al., 2019) that utilize filtered returns and asset values along with the complete history of fund cash flows realized up to period t.

3. $PME(\theta, T)_{\tau:t}$ PMEs on the to-date basis in which capital calls up to period $\tau < t$ are replaced with $\hat{V}_\tau^{\theta,\tau}$ and $\hat{R}_{0:t}^{\theta,T}$ are replaced with $\hat{R}_{\tau:t}^{\theta,T}$.

While implausibly high (or low) variance and autocorrelations of filtered return in (1) provide straightforward diagnostics of the model misspecification and the sensitivity to different parametrizations, variables in (2)–(3) are the building blocks to nowcasting performance

[11] We use backward recursion as given by equation (A.3) in the Appendix to utilize information before and after a given week to infer (i.e. filter) the return over that week as well the cumulative return up to it.

metrics. They are based on the simple idea that the NPV of cash flows will be zero if true fund returns are used for discounting. Specifically, we define the following three metrics, so that each can be interpreted as a fund-specific pricing error:

1. **In-Sample RMSE** – mean squared difference of $PME(\theta, T)_{0:t}$ from one over the period $t = \tau_0, ..., \tau$ where τ_0 and τ are within the span of data used to estimate θ;
2. **Pseudo Out-of-Sample RMSE (POS)** – mean squared deviation of $PME(\theta, T)_{0:t}$ from one over periods $t = \tau, ..., T$ such that no fund-specific data beyond week $\tau - 1$ is used to estimate θ;
3. **Fully Out-of-Sample RMSE (FOS)** – mean squared deviation of $PME(\theta, T)_{\tau:T}$ from one such that no fund-specific data beyond week $\tau - 1$ is used to estimate θ.

The advantage of the FOS metric is that it directly compares the model-predicted asset values (with the information up to that date only) with subsequent realizations of the cash flows without a risk of over-fitting to the data used in either filtering or parameter estimation (which we discuss next). The FOS metric may be ambiguous to interpret for funds that are far from being resolved and lacks power when NAV reporting noise, σ_n, is very high.[12] In contrast, the in-sample metric pertains to the structural stability of the model under the estimated parameters and averages-out the noise. However, it might not be informative if the parameter estimation (or return filtering) involves a minimand proportional to the square of logPME. In that sense, the POS metric is a compromise that, at the very least, enables a comparison of the pricing errors under different θ for a given fund.

2.4 Parameter Estimation

We estimate the parameter vector θ for each fund in our sample using a combination of two methods: (i) maximum likelihood (ML) and (ii) (partial) imputation.

As discussed in the introduction, our method exploits simultaneously cash flow and re-

[12] FOS will be biased towards rejecting our SSM estimates in favor of 'R_{ct}-interpolated'-NAVs for unresolved funds. To see this, consider a situation in which there were no distributions after τ so that the 'R_{ct}-interpolated'-NAVs ratio will be exactly as in equation (A.10) yielding a zero error.

ported NAV information. It allows us to obtain fund-specific estimates of the parameter vector for a majority of funds, independently from data and/or parameters from other funds. Nevertheless, the span of distributions and innovations in NAV reports for many funds are simply insufficient to identify all parameters independently (the following section provides specific criteria). In such cases we resort to imputation. Another reason for adopting this strategy is that some parameters might not vary much across fund groups, so there are efficiency gains from imputing parameters across peer funds or using estimates reported in the existing literature. We also find that a partial imputation approach for parameters dominates the pooling of the data across funds.

In the following subsections, we describe key steps and features of the two approaches we consider: (1) fund-specific estimates and (2) partially imputed methods. In both, we obtain standard error estimates for every parameter in θ using the numerical Hessian method as in Miranda and Fackler (2004). We conclude the section with a synopsis of Monte Carlo simulation results that illustrate parameter estimation efficiency and nowcasting performance across different degrees of model misspecification and data quality.[13]

The important feature of PE fund data is that distributions almost never occur on the day of NAV reports. This fact allows us to identify σ_d from σ_n—the non-persistent noise parameter on NAV reports. Meanwhile, the sparseness and irregularity in the distribution and NAV vectors mitigates the effect of the autocorrelation in the observed residuals that arises due to the possible misspecification of both $\delta(\cdot)_t$ and $\lambda(\cdot)_t$. We scrutinize this claim in our simulation and empirical analysis.

2.4.1 Fund-specific estimates

It is well established that identifying just α and β in PE is difficult even when data is pooled across many funds (e.g., see Ljungqvist and Richardson, 2003; Driessen et al., 2012;

[13] In the Appendix we provide (i) technical details about the estimation (Section A.3), (ii) a description of the Monte Carlo setup and results (Section A.4), as well as (iii) a single-fund case study that illustrates the intuition behind the key steps and assumptions (Section A.4.1).

Ang et al., 2018). To estimate the 10-dimensional parameter vector θ by fund, we therefore proceed in two steps. The goal of the first step is to find the fund-specific α and β that fit the SSM well while satisfying an additional NPV-based restriction. We pursue this goal via a profile likelihood method augmented with a penalty function (see, e.g. Ghysels and Qian, 2019, for details). In the second step, we iteratively estimate the remaining 8 parameters in θ and the value-to-return mapping function $\hat{m}_t$.

To implement the first step, we build a 15-by-15 grid of plausible values for α and β based on range of estimates obtained from the prior literature. Specifically, in case of a buyout fund, β ranges between 0.668 and 1.831 with a mean of 1.25 and corresponding 15 probability quantiles—0.01, 0.05, .125, ..., 0.875, 0.95, 0.99; guided by estimates in Ang et al. (2018).[14] We center the α-range at the annualized estimate of fund-level excess returns from the Kaplan and Schoar (2005) PME method which happens to be 20.5% per year or 0.38% per week (see Appendix equation (A.11) for details). We consider seven equally spaced values up to +/-3.5% per year from the each side of the α-range center.

For each of the 225 (α, β)-pairs on the grid we (i) obtain the likelihood score corresponding to the MLE estimate of $(\delta, \lambda, F, \sigma_n, \sigma_d)$ in which (α, β) are kept fixed at the respective grid values and r_{ct} is dropped from the observations vector; (ii) evaluate the pricing error that the (α, β)-pair implies against the fund cash flows.[15]

In the second step, we estimate the remaining 8 parameters (collected in the subvector ϑ), via maximum likelihood while keeping α and β at their optimized profile levels. We use R_{ct}*-interpolated NAVs* as the estimate for V_t (see equation (A.8) for details) to construct the initial asset-to-value mapping. We then iteratively update the mapping function utilizing equations (A.14–A.16) and re-estimate ϑ using the new mapping until the change in those ϑ parameters from the previous iteration is below the threshold (eq. A.17).

[14] Following Ang et al. (2018), $\beta \sim \mathcal{N}(1.25, 0.25)$, $\mathcal{N}(1.80, 0.30)$, and $\mathcal{N}(0.77, 0.23)$ for buyout, venture, and real estate funds respectively.

[15] Regarding step (i), the last row in $\tilde{Z}_t$ and $\tilde{H}_t$ of the matrix representation of SSM in (A.1) are dropped while the transition equation remains unchanged. The comparable asset helps with nowcasting but is redundant for R_m-related parameter identification. Regarding step (ii) –see equation A.12.

2.4.2 Partially imputed estimates

Here again we proceed in several steps. First, we estimate α, β, and λ. However, now we profile α and λ on a 15-by-7 grid, whereby the range for α is determined as before while λ is centered around that of the peer fund median. Recall that λ is essentially the exponential moving-average weight, constrained to be below 1 and above 0. We therefore, define the range of λ in terms of its probit function's z-score distance from the center.[16]

As for β, we treat it as a free parameter in the first step but constrained to vary within just one standard deviation (using again Ang et al. (2018) estimates by fund type) from that of the peer funds (henceforth, β-anchor). We continue applying the penalty function to the profiled likelihood grid. However, now the penalty measures the distance from zero for the autocorrelation of filtered returns (see Section 2.3) obtained from a nowcast with only R_{mt}. In doing so, we mitigate forcing too high or too low smoothing intensity upon the fund from its peers. We find this approach to be more informative about the (β, λ) pair location (than the PME-based penalty discussed above) when the fund exhibits only several quarters with meaningful distributions.

In the second step, we iterate the return-to-value mapping function and the remaining free parameters until convergence as we do with the fund-specific estimates, however we always keep F and σ_n fixed to the median of the peer funds. Therefore, this EM-like procedure (cfr. Dempster et al. 1977) involves only 5 parameters (rather than 8 under the fund-specific estimation approach).

In both steps, the peer fund estimates are defined as medians of the respective parameter estimated independently. The next section examines several ways to define peer funds. We include the fund's own fund-specific estimates to compute the peer group median. For β, we additionally consider setting the center of the distribution to be equal to the estimates in the literature by fund type (Ang et al., 2018).

[16] We consider a range from -0.75 to $+0.75$ with 0.25 increment.

2.5 Simulation Experiment Summary

An extensive simulation study is conducted to examine the performance of our method with regards to both parameter estimation and nowcasting performance. One of the focus areas in this simulation experiment is the consequences of the model misspecification relatively to the true data generating process. For example, we introduce discontinuous jumps in the funds' idiosyncratic return processes contrary to the assumption that those are conditionally Gaussian, and examine cases in which the assumed distribution function (equation 9) is different from the true one.

For brevity, we defer the details to Appendix A.4 where we examine a single case step-by-step process and report extensive summaries for a realistic panel of simulated funds. Below we summarize the main takeaways.

1. Nearly all parameters are consistently estimated despite realistically high idiosyncratic noise in fund returns (unconditionally 40% per year); the typical estimation error on fund-specific estimates (α, β, λ) is (0.002,0.359,0.054). Cross-sectional variation in α and β is statistically detectable within as few funds as a hundred.

2. The only key parameter that is affected by non-Gaussian jumps in the true return process is the volatility scale F, which attenuates to a median of 1.5x from the true (assumed) level of 2.0, while the noise level in NAV reports is over-estimated with a 0.055 median relatively to a true value of 0.050. However, this bias vanishes if deal-level NAVs are utilized (i.e. $\delta(\dot{})_t$ and σ_d are computed rather than estimated), while the efficiency of α and β estimation approaches that of the weekly OLS regressions estimates that assume fund returns are observable). Reasonable misspecification of the distribution function has virtually no effect on the consistency and efficiency of key parameter estimates.

3. Under the assumption that the true parameters are similar to those of the peer funds, partial imputation discussed in Section 2.4.2 reduces the estimation error on β and λ by a factor 1.2 to 1.6 but leads to slightly large estimation errors for α.

4. Using the FOS metric, the median nowcast error is about 0.12 in the baseline fund-specific case (0.07 if deal-level data are utilized) versus 0.28 with naïve nowcast, yielding an improvement of 74%. With respect to the POS metric, the model estimates result in smaller nowcast errors for 85-95% of simulated funds and are on average 10-times smaller than those from naïve nowcasts. Partial imputation of parameters also yields a modest improvement in the nowcasting performance.

Overall, the simulations suggest that parameter estimation and nowcasting performance are promising using our methods and that the results are not particularly sensitive to distributional and functional form misspecifications.

3 Empirical Results

In this section, we report results for parameter estimates and nowcasting performance for the sample of private equity funds provided by Burgiss (see Brown et al., 2015, for description). The data include fund-level history of cash flows between each fund and its investors, fund NAVs reports for most quarters, as well as time-invariant data such as fund strategy, vintage year, and industry based on 8 Global Industry Classification sectors. We use this industry designation to select a comparable asset as from the S&P500 [MSCI global] subindex for funds that are classified as focused on North America [Rest of the World] investments. We use the CRSP value-weighted index for market returns.

3.1 PE Fund Sample

Table 1 reports basic summary statistics separately for 1,437 buyout and 1,142 venture funds in Panels A and B, respectively.[17] This is a subset of the Burgiss Manager universe as of December 2017. It includes the funds that satisfy the following criteria:

- vintage year is between 1983 and 2008,

[17] We classify all none-venture equity-focused (i.e. not debt or mezzanine) funds as buyouts. We examine subcategories later in our analysis.

- operated for at least 23 quarters,
- reported NAV values for at least 10 quarters,
- made at least two distributions,
- have at least two peer funds for which we obtain fund-specific estimates (Section 2.4.1).

The peers are defined as funds of the same type (i.e. buyout or venture) with vintage years within a year of the fund of interest. For feasibility of fund-specific estimates, we additionally require at least five quarters during which the distributions exceed 5% of fund size and tighten the requirement for the fund resolution status so that the latest reported NAV does not exceed half of its cumulative distributions.

These filters drop 119 buyout and 57 venture funds from our sample. We note that the excluded funds tend to have lower returns and smaller size. For example, the excluded venture funds had median size/PME of $27mln/0.76 as opposed to $163mln/0.81 for the included funds. These differences are even larger for the buyout sample at $49/0.95 for excluded and $399/1.06 for included. Nevertheless, once adjusted for vintage averages, the differences in performance are statistically insignificant (t-stat of -1.15 for the combined sample). Furthermore, the excluded funds comprise just less than 2% and 1%, respectively, of buyout and venture funds and of capital committed. We therefore argue that our parameter estimates are largely representative of the PE fund universe available for U.S. institutional investors.

The last rows of each panel show a cash flow characteristic that is important for the application of our method (see, e.g, the case study in Appendix Section A.4.1), the number of quarters with meaningful distributions. It follows that a quarter of buyout funds have 5 or fewer quarters with distributions that exceed 5% of fund size. For venture funds the corresponding number is only 4. This reduces the feasibility of fund-specific estimates for the venture sample to a greater extent. We focus on relatively mature funds so we can evaluate the nowcasting performance.[18]

[18] Appendix Table A.2 reports summary statistics for PE Real Estate funds, which we also examine in our study and apply similar filters to.

3.2 Fund-specific estimates

Table 2 reports summary statistics for fund-specific estimates of SSM parameters and the assessment of the associated nowcasts. As described in Sections 2.4.1 and 3.1, identification of these utilizes only the data from the respective fund (besides any matched public benchmarks) and requires a relatively high number of quarters with substantial distributions made by the fund. We are able to obtain such fund-specific parameter estimates for 1,745 PE fund which corresponds to 62% of the fund sample described in Table 1. In subsequent sections, we will break down selected parameter and nowcasting metrics by fund-type and provide a more granular analyses via multivariate regressions.

3.2.1 Parameters

Panel A of Table 2 displays selected statistics for the 10-dimensional parameter vector θ. We observe that the average [median] α of PE funds with fund-specific estimates is 4.8 [0.036]% per year with an interquartile range of 13.7%. As alluded to earlier, it is not necessarily representative of a typical fund's risk-adjusted return since these funds are likely to have a greater fraction of successful deals that resulted in large distribution amounts happening at different points of their lives. The loading on the public market return is 1.19 on average, ranging from 0.668 at the 10th percentile to 1.979 at the 90th. It is somewhat lower than the 1.25 [1.80] mean βs for buyout and venture from Ang et al. (2018), which we used to build our profiling grids. This suggests that the penalized Likelihood attains its maximum at lower values for β..

The third row of Panel A reports the scale coefficient estimate on our proxy of time-varying idiosyncratic risk. It shows that the standard deviation of idiosyncratic returns for a typical PE fund is 2.59 times the GARCH(1,1)-filtered idiosyncratic volatility of the matched industry index. However, we note the large dispersion for this parameter. While it is equal or smaller than 0.18 for 10% of the funds, it is greater than 5.49 for a quarter of funds. In fact, for over 10% of funds it hits the upper bound of 10 that we imposed in each estimation.

This suggest that the idiosyncratic volatility parameter is hard to identify at the individual fund level. However, it seems plausible to assume F is similar across funds within the same strategy-, vintage-, and size cohort and we utilize this assumption subsequently.

The next two rows of Panel A report summaries of parameters that characterize reporting quality by PE funds. First, as follows from the percentile ranks of λ, PE fund NAVs indeed tend to exhibit a very high degree of appraisal smoothing. The median of 0.951 implies that public market valuations from five or more weeks ago comprise 77.4% (= 0.951^5) of a typical fund NAV report. This exponential weight is consistent with the AR(2) coefficients of about (0.5,0.3) on quarterly return series derived from reported NAVs. Our results suggest substantially less smoothing for about the bottom quarter of funds—e.g., λ is 0.869 or less, suggesting persistence of only about 0.16 or less at quarterly frequency.

We observe that the average [median] NAV noise—the standard deviation in NAVs that the SSM cannot attribute to past or future returns—is 4.6 [3.1]%. For some funds it appears to be implausibly high or low—e.g., 15.6% per the 90th percentile, the lower bound of 0.1% for over a quarter of funds. Thus, this parameter also appears hard to identify reliably. Just as with the idiosyncratic risk levels, we subsequently use the fact that these are plausibly similar across peer funds.[19]

Panel A also reports δ and σ_d which govern the fund distribution process. While auxiliary in nature, they need to be estimated for each fund to reflect the heterogeneity in the distribution magnitudes (given the assumed functional form of $\delta(\cdot)$). Here we just note that for the vast majority of funds both parameters are well within the upper (0.9 and 4) or lower (0.001) bounds imposed on the estimation.

The last three rows of Panel A report parameters that govern the mapping of the comparable asset to the latent fund returns at the weekly frequency, as per equation (2), representing the intercept, slope, and the variance scale. Because returns are in logs and the regression is essentially a reverse one (i.e. factor series on fund series), some caution is required with

[19] Naturally they relate to fund investments' type, portfolio diversification and operating period.

their interpretations. In particular, even though the average and median ψ are positive at about 3% per year, it does not mean that the funds underperformed those industry benchmarks since the intercept is shifted upward by the $0.5\text{var}(\eta_{ct})$-term. Moreover, the average [median] β_c of 0.79[.812] does not imply that r_t are less risky than r_{ct}. Instead, β_c is *inversely* proportional to the loading of r_t on r_{ct}. Finally, all else equal, lower levels of F_c estimates can indicate two situations: (i) better match of the comparable asset to the fund, and/or (ii) overestimated risk level of the fund.[20] For a given variance of r_{ct} and r_{mt}, the variance of η_{ct} has to go down if either β or F increases. Therefore, examining the nowcasting performance sheds light on whether we are in situation (i) or (ii).

3.2.2 Properties of filtered returns

Panel B of Table 2 compares the autocorrelations and variances of fund returns obtained from a naïve nowcast that uses reported NAVs with those from fund-specific SSM estimates. We compare series at two frequencies—weekly and quarterly. For the naïve nowcast, the quarterly series correspond to NAV-changes adjusted for the value of within-quarter cash flows, discounted at the industry benchmark returns. Whereas the weekly naïve nowcast assumes the asset values track the industry benchmark on top of the interpolated between-quarter change in PME (as from equation A.10). We see that those series suggest a rather high positive autocorrelation of 41–56% for the typical fund and over 71–89% for a tenth of the sample. Interestingly though, for another tenth of the sample we observe no meaningful correlation in naïvely nowcasted returns. Meanwhile the variances subpanel, suggests that the annualized standard deviation of returns for the middle half of PE funds is between 16.7% and 32.8%.

The SSM-based return nowcasts are obtained via the Kalman filter at weekly frequency as explained in Section 2.3. We then aggregate these filtered estimates to the quarterly frequency by summation of the log returns within each quarter. Panel B shows that auto-

[20] To see this, approximate the log of equation (1) with $r_t = \alpha^* + \beta r_{mt} + \eta_t$, and substitute it in the log of equation (2). This results in $r_{ct} = \psi + \beta_c\alpha^* + \beta_c\beta r_{mt} + \beta_c\eta_t + \eta_{ct}$.

correlation of SSM-based fund returns is virtually zero at both frequencies on average (and for the median fund) and is within 10% of zero at the weekly frequency for the middle two quartiles. Meanwhile, the variance of filtered returns is a factor of 2 to 2.6 higher than under the naïve nowcast, suggesting the annualized standard deviations between 24.5% and 54.7% for the middle two-quartiles.

While these results suggest the SSM-filtered returns are more representative of the true return process than one inferred from reported NAVs, we note that for a substantial minority of funds the SSM nowcast exhibit fairly extreme autocorrelations. For example, one-fifth of the fund sample exhibits the autocorrelation of filtered returns in excess of +0.47 or below -0.37 at the quarterly frequency. This is likely driven by the model (e.g., $\lambda(\cdot)_t$ and $\delta(\cdot)_t$) or parameter misspecification, and can be addressed on the case-by-case basis in practice. We also note that usage of the backward induction with the Kalman filter likely underestimates the variance of the latent return series, and, hence, overestimates the magnitude of the correlations.[21] To examine the scope of this bias, the last row of Panel C reports the variance estimate of fund returns derived from the fund parameter and benchmark variance estimates. Indeed, we see that these are a factor of 1.5 higher than with filtered returns, suggesting that the typical PE fund total risk is about 44%, if measured by annualized standard deviation.

3.2.3 Nowcasting performance

Panel C of Table 2 reports fund-level nowcasting performance metrics introduced in Section 2.3. First, we note that the out-of-sample RMSEs are similar to the in-sample RMSEs and are positively correlated, as evident from Figure 1 (especially on the pseudo-out-of-aample basis in Panel A). Thus, the in-sample pricing error is a useful predictor for nowcasting performance as suggested by the simulations discussed in Section 2.5.

Second, SSM-based RMSEs are notably smaller than those from the naïve nowcast, again as suggested by simulations. For a typical fund, the POS RMSE reduces from 0.34 to

[21] The fact that the autocorrelation magnitudes are notably smaller at weekly frequency is also consistent with this explanation since the bias gets cumulated within a quarter.

0.097 while the fully-out-of-sample values drop from 0.25 to 0.146 and are sizable across all percentiles. Figure 2 shows that improvements are largest for funds with relatively large naïve nowcast errors and that the SSM-based nowcasts are materially inferior to the reported NAVs only when the latter happen to be particularly good—e.g., to the left of log MSE of -5 on the x-axis, which corresponds to the RMSE of less than 8.2%.

Across the buyout and venture sample, SSM-based nowcasts with fund-specific parameter estimates result in smaller POS [FOS] RMSE than the naïve nowcast for 80% [65%] of individual funds [untabulated]. Our analysis also indicates that the gains are higher when a portfolio of funds is concerned since the SSM-based nowcast errors are less correlated than those from the naïve approach.

3.3 Partially imputed estimates by fund type

Next, we examine partially imputed parameter estimates as described in Section 2.4.2 and compare the nowcast performance with that generated by fund-specific estimates. Most importantly, the partial imputation enables us to obtain parameter estimates for approximately 95% of the sample funds and, therefore, provide a plausibly representative picture of the institutional PE fund universe.

We consider three ways to define peer funds. First, in what we will refer to as *Peer-imputed* we set F, σ_n as well as the centers of β-anchor and λ-grid to the median values of the respective fund-specific parameter values for funds of the same strategy (buyout of venture) incepted in the same or adjacent vintage year. Additionally, if there are three or more peer funds in the same size tercile, we limit the peer group to those peers funds only. We then attempt to further narrow the peer group to at least three funds in the same industry of specialization.

Second, we take the opposite approach and set the four imputed values to the buyout or venture fund average in what we refer to as *Average-imputed*. Finally, we reset the β-anchor to the strategy-specific estimate from Ang et al. (2018) in what we refer to *Literature-*

imputed, while keeping λ-grid, F, and σ_n same as in the *Average-imputed* case.

Accordingly, we compare the SSM nowcast performance metrics introduced in Section 2.3 under these four different sets of parameters. This exercise sheds light on three important questions: (i) is there a meaningful time-series variation in PE fund risk-return profile, (ii) how big is the noise reduction benefit from restricting the fund-specific parameters, (iii) are our systemic risk exposure estimates more consistent with the fund-level cash flow observations than previously established in the literature?[22]

3.3.1 Buyout funds

Table 3 reports selected parameters estimates and nowcast performance metrics for the three imputation methods alongside with the fund-specific estimates. The latter are available for 1,105 funds, while the partial imputation increases the coverage to 1,391 funds (96.7% of the sample per Panel A of Table 1). For the remaining 3.3% the MLE does not converge.

In Panel A, we focus on how key parameter estimates, namely α and β, very across estimation methods. We observe that the average and median α estimates of 2.6 to 4.0% are lower for peer- and average-imputed estimates in comparison to the 5.2-5.3% in fund-specific estimates sample. However, these are still notably higher than the near-zero values for the literature-imputed estimates. These also stand out with notably higher βs at 1.55/1.59 on average/median, especially against the 1.04–1.10 range for mean and medians in fund-specific and peer-imputed cases.

Additionally, in Panel A we examine the smoothing parameter λ, as well as the variance scale on comparable asset parameter, F_c, to support the diagnostics discussed in the closing paragraph of Section 3.2.1. We note that F_c levels go up in peer-imputed versus fund-specific and partially drop back with the average- and literature-imputed estimates. Meanwhile, we observe no meaningful variation in the smoothing rate across estimation types in means but a 3-5pp increase in percentiles below 75th as we go from the peer- to literature-imputed.

[22] Ang et al. (2018) encompass the previous literature estimates as priors for Bayesian MCMC.

Panel B reports quarterly autocorrelations and variances of filtered returns in comparison with those from the naïve approach. The latter are computed for the partially-imputed sample and suggest similar levels of autocorrelation 40.5% as the fund-specific estimates sample reported in Table 2 (which includes venture funds), and a total level of risk corresponding to an annualized standard deviation of returns of 23% for a typical fund. We note that fund-specific and peer-imputed returns have very similar summary statistics, with the typical autocorrelation being a bit closer to zero for peer-imputed (0.026 versus 0.074) and the total risk corresponding to 33-34% return standard deviation per year. Meanwhile, the filtered returns from average- and literature-imputed estimates have a factor in the range of 1.5–1.7 higher variances, which translates into annualized standard deviations of 41–44% for the median fund and over 51% for the funds in the top quartile of total risk. As for the autocorrelations, the interquartile ranges is similar to that of fund-specific and peer-imputed but shifted downwards by about 10pp and 15pp, respectively, for average- and literature-imputed.

Finally, Panel C of the table reports summary statistics for the nowcasting performance and reveals that the partially-imputed estimates have lower nowcast errors and higher improvement rates relative to the naïve nowcast on both POS and FOS basis. However, performance monotonically declines as we move from peer- to literature-imputed estimates. For example, the median FOS RMSE with literature-imputed β-anchor is 0.173 and 27% higher than peer-imputed estimates but nevertheless 42% lower than the naïve nowcast.

Overall, Table 3 suggests the following conclusions. Peer-imputed estimates dominate other partially-imputed estimates and benefit from a noise suppression relative to the fund-specific estimates for a majority of cases. Nonetheless, heterogeneity in parameters appears genuine across buyout funds.[23] Forcing higher levels of fund systematic risk tends to not reduce the autocorrelation magnitudes but, in contrast, seems to amplify the degree of model misspecification across buyout funds.

[23] In untabulated analyses, we find that the performance gap between peer-imputed and fund-specific estimates vanishes with funds for which in-sample RMSE from SSM is better then from the naïve approach.

3.3.2 Venture funds

Table 4 reports the same analyses as in Table 3 but for the venture fund sample. Fund-specific estimates are available for 640 funds, while the partial imputation increases the coverage to 1,052 funds (92.1% of the venture sample).

From Panel A, we see that differences between the fund-specific and peer-imputed in α as well as β estimates for venture funds appear more stark than those in the buyout sample. At 1.196, the average peer-imputed β is 0.18 lower than that of fund-specific, albeit the difference in medians is less than half of that. Most notably, the mean [median] α estimates at 4.8 [0.5]% for fund-specific case are sharply higher than the 0.1 [-2.7]% levels for peer-imputed case, and especially so when compared to -3.6 [-6.4]% in the literature-imputed case. As with buyouts, the latter case is characterized with more than 0.5 higher βs.

Turning to the filtered return analyses in Panel B, we observe a similar pattern as with buyout funds. Namely, there is a notable and roughly equal reduction in autocorrelations relative to the naïve method, and a shift towards negative levels for average- and literature-imputed relative to the fund-specific and peer-imputed. The gap in variances of filtered returns is less pronounced than in buyout sample. For venture, the literature-imputed estimates correspond to the median standard deviation of 46% annualized as opposed to 35-38% range with fund-specific and peer-imputed parameters.

In Panel C of Table 4, we observe again a similar pattern as in the corresponding panel for buyout funds: peer-imputed parameters result in better nowcasts but further reducing variance in F and σ_n and/or nudging higher βs appears counter productive. We also note that, while SSM RMSEs are higher for the venture sample and do not shrink as much relatively to naïve nowcast as they do with buyout funds, overall the improvement rate is actually slightly higher for venture funds. For example, 74.6% of venture funds have smaller FOS RMSE from SSM with peer-imputed parameters than from a naïve nowcast for the respective fund, as opposed to 72.8% of buyout funds.

It it important to note however that a β closer-to-one does not warrant better nowcasting performance. This is evident from Table A.2 in the Appendix, which reports a similar analyses for REPE funds. There again peer-imputed estimates result in best nowcasting performance with average β being indistinguishable from the literature-imputed levels and further below one for a typical fund.

3.4 Discussion

3.4.1 Why do we find lower betas?

Tables 3–4 provide evidence that our estimates of systematic risk level are lower while also more consistent with cash flows realizations of individual funds. This raises a question of why other studies found higher β levels for largely same PE funds. One potential explanation is that funds that have higher risk produce larger cash flows, especially, if higher risk is associated with higher abnormal returns which appears consistent with differences between α and β levels in the subsamples with fund-specific estimates and without.

Table 5 tests this hypothesis more formally by regressing fund-level β estimates on fund characteristics. In specification 1, we regress β on α using fund-specific estimates in Panel A and peer-imputed in Panel B, while also controlling for fund industry and subtype fixed effects (discussed below). Both panels reveal a strong positive association between PE funds excess returns and systematic risk levels. Because β and α are jointly estimated, this regression result is prone to be affected by an omitted variable bias reflecting the joint estimation error. We note that such a bias would likely be reducing a positive relationship insofar that an understated β would result in higher α given the on average positive market returns. Nevertheless, in the remaining specifications we use explanatory variables that are independent from α estimate—namely, fund size, cumulative distributions and the number thereof—while not seeking to establish a causal relationship.

Our preferred specification is 6, where we include cash flow characteristics as well additionally control for industry and market variance over the life-time of the fund. In both panels, we see that the fund β increases in fund size and the value of distributions it made.

This explains why we find lower average and median βs. Prior studies have been answering a different question—what is the β of the panel of fund cash flows (rather than that of an average or typical fund).[24]

Figure 3 plots industry and fund subtype fixed effects from specification 6 of both panels. We see that the industry effects are in concordance with the general intuition that utilities, non-cyclical consumer goods and health care exhibit less systematic risk relatively to the 'Consumer Discretionary'-goods baseline, in contrast to 'Energy', 'Telecoms' and 'Information Technology'. A similar pattern consistent with the economic intuition is observed for fund subtype fixed effects relatively to 'buyout' funds group that has the average beta of just over one, according to both sets of estimates. Interestingly, we see that late-stage venture funds have a positive fixed effect as high as early-stage funds for both fund-specific and peer-imputed estimates, and above that of generalists and expansion capital.

It is also worth noting that prior studies' inference about β may reflect the marginal utility to trade of investors in private companies or funds, rather than the risk of the underlying project cash flows. Consistent with this hypothesis, findings of higher level of systematic risk tend to rely on round-to-round valuation change of venture-backed companies (Korteweg and Sorensen, 2010) and secondary fund trades in buyouts (Boyer et al., 2018). Additionally, the contracts between portfolio firms and funds as well as between a fund and its investors likely further attenuate the risk at fund level. In both, the decision-makers (e.g. entrepreneurs and fund GPs) get an option-like payout which not only reduces the upside variation in fund cash flows but subsidizes them on the downside through claw back provisions and state-contingent cash flow rights (see, e.g., Kaplan and Strömberg, 2003)

3.4.2 Trends in PE fund risk

The notable differences in nowcasting performance between the peer- and average-imputed estimates, suggest a meaningful variation in PE fund risk profiles. We depict this via a

[24]In untabulated analyses, we find that these results do not hold for the idiosyncratic risk level which falls in fund size and insignificantly relates to the fund's cumulative distributions.

vintage-year box plot of α, β, and F in Figure 4.

Panel A of the figure plots the results for the buyout sample. We see two clear spikes in the idiosyncratic risk exposures, during 1990-93 and then during 2002-05, which also happened to be the best vintages. The levels for the 2007-08 vintages appear moderate by historical standards with excess returns turning slightly negative. As for systematic risk, the cross-sectional variation appears to dominate the time series trend. Nevertheless, there appears to be a shift towards consistently above-one median levels post-2000. However, three quarters of buyout funds have had β estimated at below 1.25 for each vintage since 2005.

Panel B of the figure plots the results for venture sample. Here we note a clear spike in systematic risk exposures during the 1995–98 vintages, with top-quartile β being in vicinity of 2.0. We then observe a rapid decline towards sub-one β for more than half of the funds in 2002 vintage and then a rebound towards 1.1–1.3 medians more recently. Nevertheless three quarters of venture funds have had β estimated at below 1.4 for each vintage since 2005 with αs starting exceed zero for about half of funds. Interestingly, there is a clear rebound in the idiosyncratic risk levels taken post-2002 even though the 2007-08 vintage seem to have scaled back to under $3\times$ the level of public equities.

4 Conclusion

This paper develops a new method that provides reliable estimates of PE fund asset values at high frequency while shedding light on rich temporal and cross section variation in manager risk and reporting quality characteristics even when only fund-level cash flows and NAVs are utilized. We show how this framework can be extended to incorporate other types of data, such as portfolio-level deals information, secondary transactions with fund stakes, etc. These new insights are of critical importance in light of continued growth of PE and other illiquid assets in the institutional portfolios and the increasingly high regulatory interest to the risks arising from these investments. Our results suggest that the risk-return profile of PE based on the samples from the 1990s are likely not representative of the currently operating funds.

Table 1. Fund sample summary statistics

This table reports summary statistics for the sample of buyout (Panel A), and venture funds (Panel B) incepted between 1983 and 2008. For the fund to be included in our sample, it has to meet three criteria: (i) have been operating for at least 6 years since inception, (ii) have at least 20 NAV reports, (ii) have at least two peer funds which meet criteria for the estimation in which SSM parameters are obtained independently from other funds (see Section 3.1 for details). # of 'Capital Calls', 'Distributions', 'NAV reports' count the respective data point for each fund during at most 12 years of fund operations ('Fund life'). Data beyond 12th year of fund life are omitted. '% Unresolved' is the ratio of fund latest NAV report to the sum thereof with the cumulative distributions. The last two rows of each panel count the number of quarters with non-zero distributions or only when those exceed 5% of fund size.

Panel A. Buyout funds

Number of funds: 1,437

	mean	sd	skew	p5	p10	p25	p50	p75	p90	p95
Vintage year	2001.6	6.0	-1.1	1983	1989	1998	2004	2006	2008	2008
Fund size (USD mil.)	939.0	1798.3	5.2	20	49	172	398.5	874	3671	9165
Fund life (years)	11.2	1.0	-1.3	8.04	9.17	10.5	12.0	12.0	12.0	12.0
IRR(%)	11.6	15.7	1.0	-24.2	-9.43	3.60	9.91	17.9	36.4	64.2
Money Multiple	1.7	0.9	3.3	0.22	0.58	1.20	1.57	2.04	3.12	5.26
PME (v. CRSP VW)	1.2	0.6	1.9	0.15	0.39	0.78	1.06	1.43	2.18	3.14
# Capital Calls	31.9	19.9	1.1	3	6	17	28.5	44	68	94
# Distributions	30.3	22.0	2.9	3	7	16	26	38	69	107
# NAV reports	19.7	2.7	-0.8	12	14	18	20	22	23	24
% Resolved	83.8	17.7	-1.7	17.8	44.2	77.8	90.1	96.1	99.6	99.9
# Quarters w/ Dist.> 0	8.5	4.8	0.4	0	1	5	8	12	17	20
w/ Dist > 5% of Fund	5.7	3.7	0.9	0	1	3	5	8	12	17

Panel B. Venture funds

Number of funds: 1,142

	mean	sd	skew	p5	p10	p25	p50	p75	p90	p95
Vintage year	1999.9	6.9	-0.9	1981	1985	1997	2000	2006	2008	2008
Fund size (USD mil.)	254.5	323.6	5.1	10	25	74	163	316	743	1530
Fund life (years)	11.4	1.0	-1.9	8.15	9.27	11.1	12.0	12.0	12.0	12.0
IRR(%)	12.6	40.1	5.6	-31.5	-16.1	-3.07	5.16	14.3	63.7	197.9
Money Multiple	2.0	2.8	6.7	0.13	0.31	0.82	1.34	2.04	5.72	15.3
PME (v. CRSP VW)	1.2	1.7	6.4	0.10	0.20	0.52	0.81	1.24	3.25	10.1
# Capital Calls	23.0	16.1	1.9	2	4	12	19	30	52	81
# Distributions	18.7	13.8	1.8	1	3	9	16	25	47	64
# NAV reports	20.4	2.5	-1.2	12	15	19	21	22	23	24
% Resolved	71.7	24.8	-0.9	6.10	20	56	78.2	93	99.3	99.9
# Quarters w/ Dist.> 0	4.9	3.7	1.0	0	0	2	4	7	12	16
w/ Dist > 5% of Fund	3.7	3.0	1.2	0	0	1	3	5	10	13

Table 2. Summary of *fund-specific* estimates

This table reports summary statistics parameter estimates and nowcasting performances metrics obtained with fund-level SSMs for the sample of buyout and venture funds described in Table 1. Panel A reports SSM parameters that are estimated independently for each fund as described in Section 2.4.1 using at least (most) four (seven) years. The return autocorrelations and variances in Panel B are computed from the full history of fund returns estimates. The nowcasting performance metrics in Panel C are as discussed in Section 2.3 utilize at least [most] one [five] remaining years of fund life.

	Number of fund-estimates: 1,745							
	mean	sd	skew	p10	p25	p50	p75	p90
Panel A. Parameter estimates								
Main parameters:								
α: Excess return (p.a.)	0.048	0.14	2.7	-0.127	-0.035	0.036	0.102	0.265
β: Systematic risk	1.19	0.41	1.1	0.668	0.962	1.102	1.399	1.979
F: Idiosync Volty ($\times$)	3.77	3.21	0.9	0.180	1.364	2.591	5.495	10.00
λ: NAV smoothing bias	0.88	0.20	-2.9	0.357	0.869	0.951	0.983	0.990
σ_n: NAV report noise	0.046	0.058	2.4	0.001	0.001	0.031	0.059	0.156
δ: Dist intensity trend	0.030	0.10	27.2	0.004	0.010	0.019	0.035	0.079
σ_d: Dist intensity noise	2.56	0.52	0.5	1.792	2.187	2.512	2.875	3.496
Parameter mapping to Comparable asset:								
ψ: log intercept to Fund (p.a.)	0.026	0.11	-0.9	-0.148	-0.041	0.034	0.097	0.184
β_i: beta on Fund return	0.53	0.34	1.1	0.123	0.262	0.469	0.722	1.156
F_c: IdVol vs Fund return ($\times$)	0.79	0.42	0.6	0.100	0.546	0.812	0.998	1.488
Panel B. Filtered return properties								
Autocorrelations:								
Reported NAVs (quarterly)	0.38	0.25	-0.9	-0.050	0.239	0.411	0.559	0.712
SSM estimates (quarterly)	0.042	0.25	0.1	-0.369	-0.133	0.051	0.201	0.472
SSM estimates (weekly)	0.013	0.19	2.0	-0.141	-0.101	-0.060	0.052	0.422
Naive nowcast (weekly)	0.51	0.29	-0.3	0.029	0.266	0.557	0.767	0.895
Variances:								
Reported NAVs (quarterly)	0.048	0.23	13.8	0.003	0.007	0.014	0.027	0.103
SSM estimates (quarterly)	0.085	0.37	27.2	0.006	0.015	0.031	0.072	0.263
SSM estimates (weekly)	0.0037	0.006	6.2	$<$0.001	0.001	0.002	0.004	0.013
SSM parameter-based (weekly)	0.0067	0.007	1.1	0.001	0.001	0.003	0.010	0.020
Panel C. Nowcasted performance assessment								
In-Sample RMSE	0.19	0.19	2.3	0.041	0.077	0.128	0.222	0.637
POS RMSE SSM	0.16	0.19	2.5	0.021	0.050	0.097	0.195	0.571
POS RMSE Naive	0.54	0.55	1.8	0.061	0.178	0.341	0.657	2.032
FOS RMSE Naive	0.35	0.34	2.2	0.021	0.117	0.250	0.448	1.025
FOS RMSE SSM	0.24	0.33	4.3	0.011	0.065	0.146	0.290	0.734

Table 3. Buyout fund estimates comparison

This table reports summary statistics on fund-level SSM parameter estimates (Panel A), fund quarterly return estimates (Panel B) and nowcasting performances metrics (Panel C) for buyout funds incepted between 1983 and 2008. Section 2.3 defines the nowcasting performance metrics. Row titles indicate the method that parameter estimates were obtained with. Sections 2.4.1–2.4.2 describe the key differences between fund-specific and partially imputed parameter estimations. Section 3.3 describes the three different imputation methods we applied.

	Number of fund-level estimates: fund-specific—1,105, partially-imputed—1,391							
	mean	p25	p50	p75	mean	p25	p50	p75
	Panel A. Selected parameters							
	α (p.a.)				β			
Fund-specific	0.052	−0.017	0.053	0.114	1.098	0.839	1.040	1.307
Peer-imputed	0.040	−0.039	0.031	0.105	1.050	0.802	1.089	1.290
Average-imputed	0.036	−0.045	0.026	0.105	1.132	0.803	1.303	1.303
Literature-imputed	0.006	−0.080	−0.002	0.083	1.551	1.250	1.593	1.750
	F_c				λ			
Fund-specific	0.838	0.570	0.842	1.030	0.859	0.843	0.945	0.980
Peer-imputed	1.479	0.770	1.007	1.596	0.886	0.828	0.920	0.965
Average-imputed	1.072	0.710	0.918	1.217	0.880	0.849	0.937	0.963
Literature-imputed	0.989	0.686	0.867	1.119	0.889	0.861	0.967	0.967
	Panel B. Fund return properties							
	Quarterly return autocorrelation				Quarterly return variance			
Naive nowcast	0.362	0.216	0.405	0.548	0.063	0.007	0.013	0.029
Fund-specific	0.067	−0.100	0.074	0.234	0.076	0.013	0.027	0.070
Peer-imputed	0.057	−0.142	0.026	0.222	0.058	0.017	0.028	0.049
Average-imputed	−0.032	−0.222	−0.065	0.112	0.069	0.029	0.043	0.066
Literature-imputed	−0.059	−0.247	−0.090	0.075	0.075	0.034	0.049	0.073
	Panel C. Nowcast performance assessment							
	POS RMSE				FOS RMSE			
Naive nowcast	0.472	0.174	0.329	0.597	0.345	0.110	0.250	0.463
Fund-specific	0.148	0.044	0.084	0.177	0.222	0.054	0.124	0.267
Peer-imputed	0.185	0.049	0.099	0.212	0.174	0.047	0.105	0.207
Average-imputed	0.201	0.060	0.119	0.229	0.186	0.051	0.111	0.213
Literature-imputed	0.208	0.067	0.126	0.241	0.201	0.053	0.118	0.221
	POS Improvement Rate				FOS Improvement Rate			
Fund-specific	0.781	–	–	–	0.657	–	–	–
Peer-imputed	0.785	–	–	–	0.728	–	–	–
Average-imputed	0.761	–	–	–	0.714	–	–	–
Literature-imputed	0.747	–	–	–	0.696	–	–	–

Table 4. Venture fund estimates comparison

This table reports summary statistics on fund-level SSM parameter estimates (Panel A), fund quarterly return estimates (Panel B) and nowcasting performances metrics (Panel C) for venture funds incepted between 1983 and 2008. Section 2.3 defines the nowcasting performance metrics. Row titles indicate the method that parameter estimates were obtained with. Sections 2.4.1–2.4.2 describe the key differences between fund-specific and partially imputed parameter estimations. Section 3.3 describes the three different imputation methods we applied.

	Number of fund-level estimates: fund-specific—640, partially-imputed—1,052							
	mean	p25	p50	p75	mean	p25	p50	p75
Panel A. Key parameter estimates								
	α				β			
Fund-specific	0.048	−0.062	0.005	0.080	1.379	1.102	1.102	1.548
Peer-imputed	0.001	−0.114	−0.027	0.058	1.196	0.802	1.178	1.402
Average-imputed	−0.007	−0.122	−0.039	0.047	1.325	1.029	1.297	1.629
Literature-imputed	−0.036	−0.155	−0.064	0.023	1.720	1.500	1.500	2.100
	F_c				λ			
Fund-specific	0.717	0.495	0.753	0.955	0.904	0.901	0.960	0.985
Peer-imputed	0.900	0.577	0.829	1.073	0.909	0.862	0.934	0.969
Average-imputed	0.755	0.536	0.726	0.916	0.908	0.856	0.965	0.980
Literature-imputed	0.722	0.544	0.704	0.864	0.917	0.906	0.966	0.981
Panel B. Fund return properties								
	Quarterly return autocorrelation				Quarterly return variance			
Naive nowcast	0.380	0.257	0.414	0.560	0.060	0.007	0.013	0.029
Fund-specific	0.009	−0.165	0.019	0.172	0.100	0.019	0.036	0.076
Peer-imputed	0.037	−0.125	0.014	0.182	0.074	0.020	0.031	0.050
Average-imputed	−0.015	−0.172	−0.037	0.109	0.081	0.031	0.045	0.064
Literature-imputed	−0.039	−0.194	−0.060	0.083	0.091	0.038	0.052	0.075
Panel C. Nowcast performance assessment								
	POS RMSE				FOS RMSE			
Naive nowcast	0.551	0.198	0.358	0.636	0.386	0.156	0.301	0.509
Fund-specific	0.180	0.061	0.112	0.208	0.237	0.075	0.162	0.298
Peer-imputed	0.151	0.059	0.099	0.176	0.188	0.060	0.136	0.241
Average-imputed	0.168	0.075	0.124	0.199	0.218	0.072	0.158	0.286
Literature-imputed	0.183	0.082	0.137	0.221	0.233	0.081	0.173	0.297
	POS Improvement Rate				FOS Improvement Rate			
Fund-specific	0.811	–	–	–	0.633	–	–	–
Peer-imputed	0.864	–	–	–	0.746	–	–	–
Average-imputed	0.831	–	–	–	0.721	–	–	–
Literature-imputed	0.809	–	–	–	0.697	–	–	–

Table 5. Fund characteristics and *Systematic risk* exposure

This table regresses PE fund-level β estimates on its α-estimates as well as selected fund characteristics. Panel A uses the sample of fund-specific estimates, while Panel B uses peer-imputed estimates. The sample includes PE funds described in Table 1 as well as REPE funds per Table A.2. ***/**/* denotes significance at 1/5/10% confidence level, t-statistics robust to error clustering at vintage level are reported in parentheses.

Panel A. Fund-specific estimates

	(1)	(2)	(3)	(4)	(5)	(6)
α estimate	0.252***					
	(3.84)					
log(Fund Size)		0.0156**			0.0364***	0.0353***
		(2.04)			(4.30)	(4.14)
log($\sum$Distributions/Size)			0.0205		0.0712***	0.0710***
			(1.11)		(3.39)	(3.38)
log(# of Distributions)				−0.0457***	−0.0977***	−0.0955***
				(−2.75)	(−4.96)	(−4.79)
Industry variance						−0.0087
						(−1.18)
Market variance						−0.0167
						(−0.75)
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Fund type FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,953	1,953	1,953	1,953	1,953	1,953
R^2	0.252	0.247	0.246	0.249	0.258	0.259

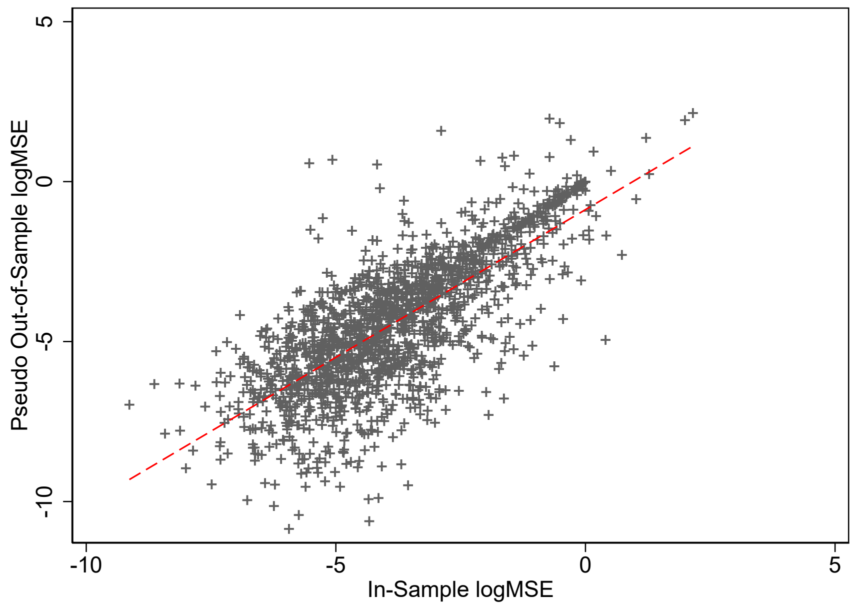
Panel B. Peer-imputed estimates

	(1)	(2)	(3)	(4)	(5)	(6)
α estimate	0.122***					
	(3.50)					
log(Fund Size)		0.0255***			0.0330***	0.0317***
		(5.12)			(5.96)	(5.70)
log($\sum$Distributions/Size)			0.0117		0.0332***	0.0331***
			(1.44)		(3.22)	(3.21)
log(# of Distributions)				0.00280	−0.0340***	−0.0318***
				(0.34)	(−3.06)	(−2.84)
Industry variance						−0.0120**
						(−2.57)
Market variance						−0.0319*
						(−1.87)
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Fund type FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,758	2,758	2,758	2,758	2,758	2,758
R^2	0.155	0.156	0.150	0.149	0.161	0.164

Figure 1. Predictability of nowcasting performance

This figure compares the logs of in-sample nowcasting error (x-axis) with those out-of-sample (y-axis) as defined in Section 2.3 using fund-specific estimates summarized in Table 2. A positive slope indicates that that SSM nowcasting precision is more predictable. Table 1 describes the sample.

Panel A: Pseudo Out-of-Sample



Panel B: Fully Out-of-Sample

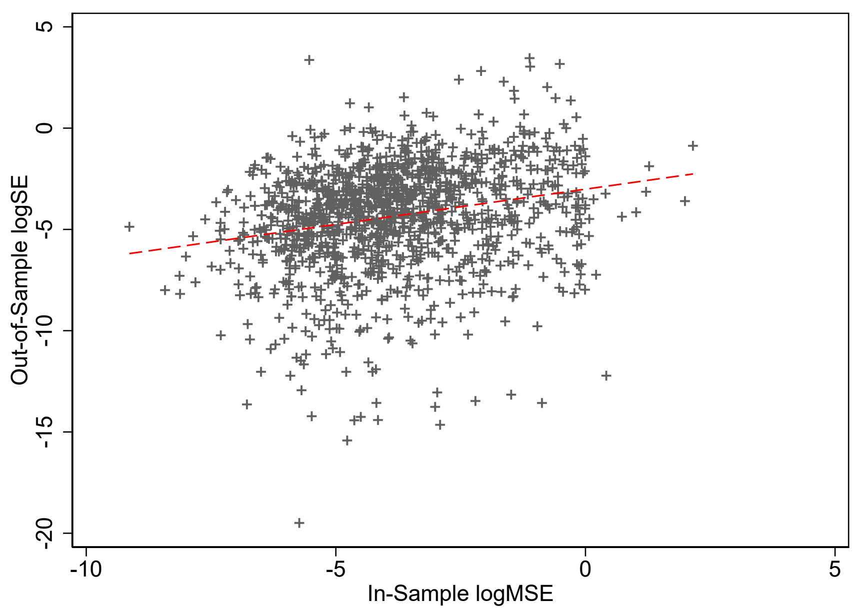
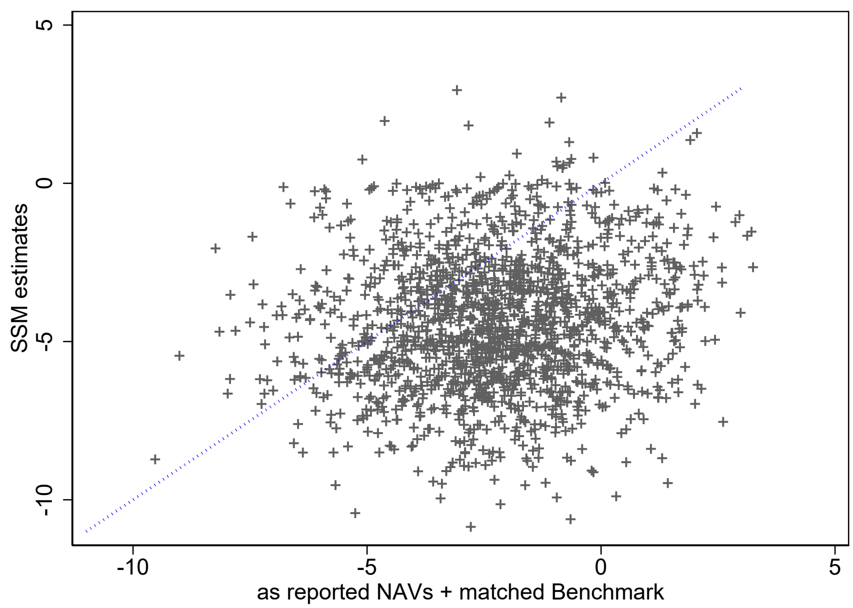


Figure 2. Nowcasting performance assessment

This figure compares nowcasting errors from SSM (y-axis) using fund-specific estimates summarized in Table 2 with those from a naïve method based on as=reported NAVs and matched benchmark returns. The metrics are defined in Section 2.3 and are based on the idea that the NPV of fund cash flows has to be zero if true returns are used to discount them.

Panel A: Pseudo Out-of-Sample



Panel B: Fully Out-of-Sample

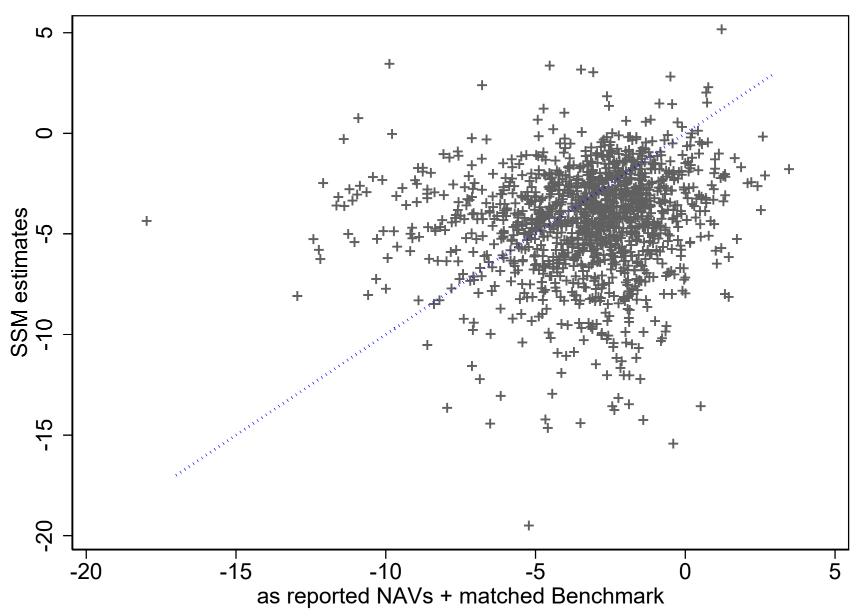
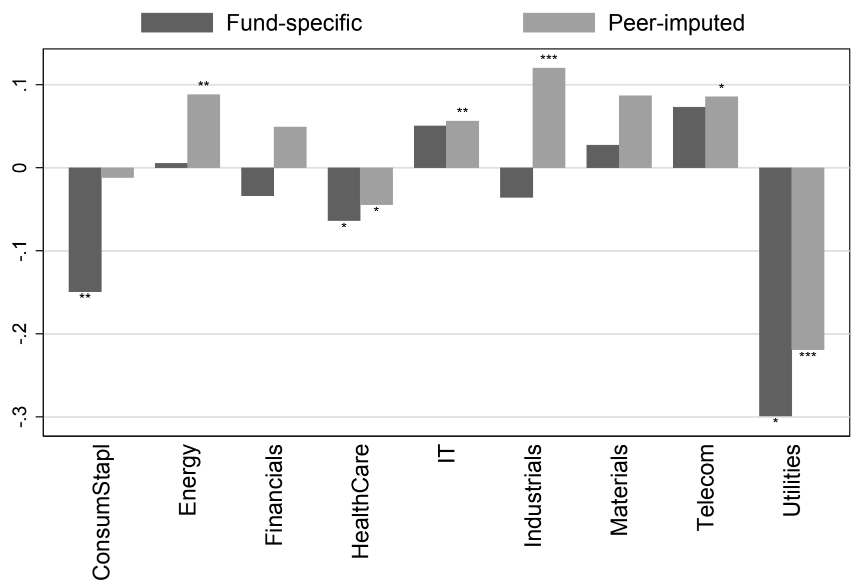


Figure 3. Fund characteristics and *Systematic risk* exposure

This figure plots fixed effect estimates of regression of fund-level β estimates on fund characteristics corresponding to specification 6 of Table 5. The sample includes PE funds described in Table 1 as well as REPE funds per Table A.2. ***/**/* denotes significance at 1/5/10% confidence level with inference robust to clustering at vintage level. The baseline effects are, respectively, 'Consumer discretionary' and 'Buyout'.

Panel A: Industry fixed effects



Panel B: Fund type fixed effects

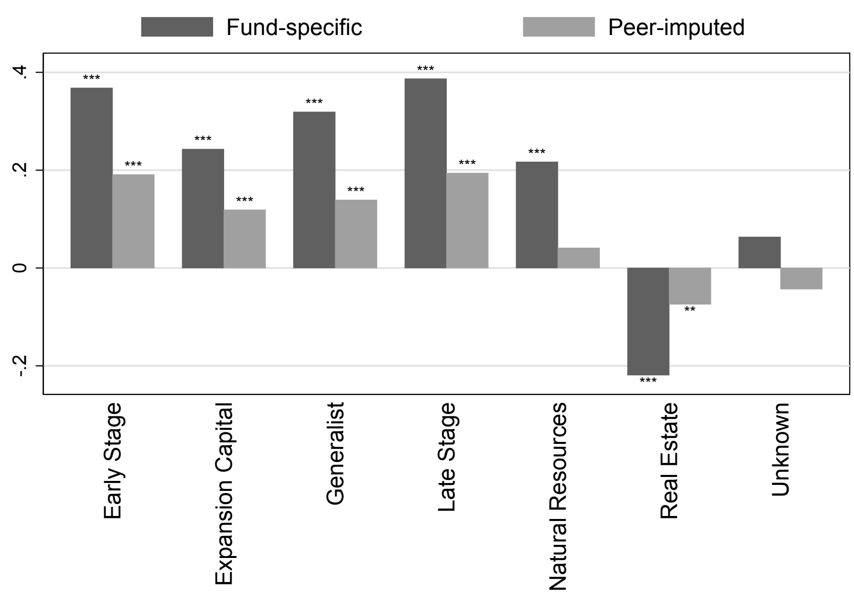


Figure 4. Trends in fund risk and returns

This figure plots the time series and cross-sectional variation of selected fund-level parameters obtained partially-imputed from peers with fund-specific estimates as described in Sections 2.4.2 and 3.3. The "boxes" reflect the inter-quartile range and the median of the respective parameter over buyout (Panel A) or venture (Panel B) funds incepted in the year indicated on the x-axis. Outlier values are not depicted.

Panel A: Buyout funds

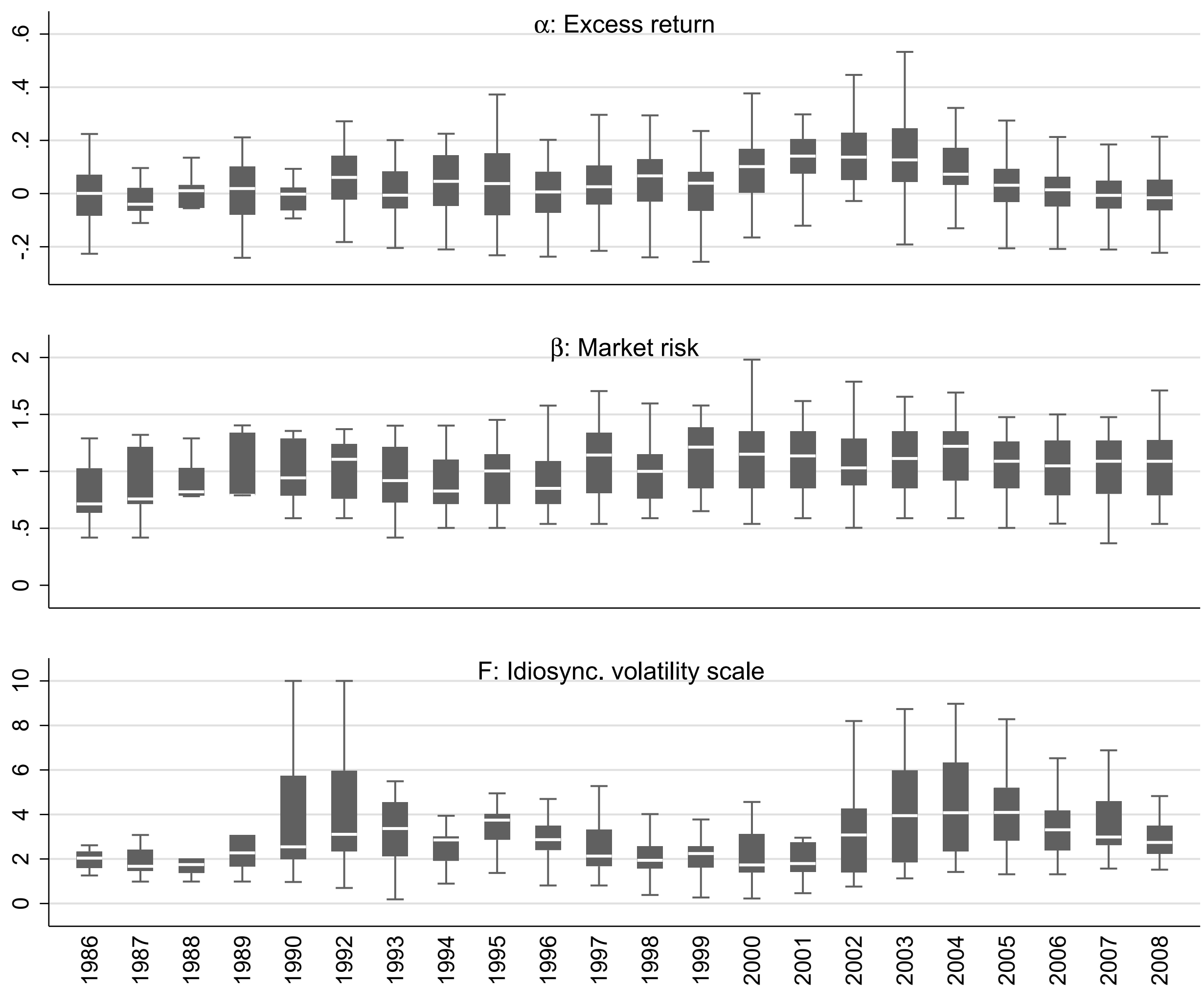
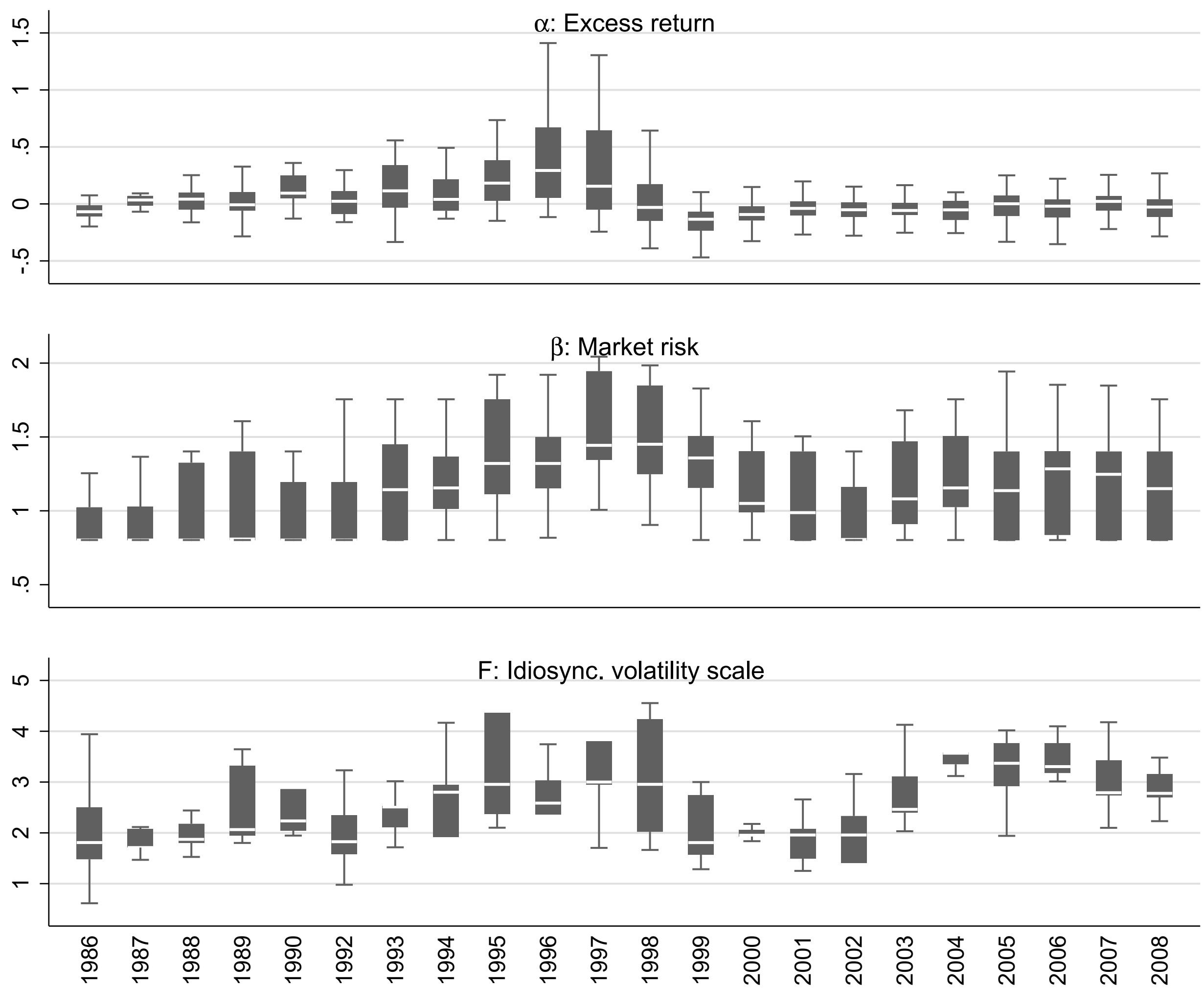


Figure 5. Trends in fund risk and returns—(*Continued*)

Panel B: Venture funds



References

Albuquerque, R. A., Cassel, J., Phalippou, L., Schroth, E. J., 2018. Liquidity provision in the secondary market for private equity fund stakes. Oxford University Working Paper .

Andreou, E., Ghysels, E., Kourtellos, A., 2013. Should macroeconomic forecasters use daily financial data and how? Journal of Business and Economic Statistics 31, 240–251.

Ang, A., 2014. Asset management: A systematic approach to factor investing. Oxford University Press.

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

Ang, A., Papanikolaou, D., Westerfield, M. M., 2014. Portfolio choice with illiquid assets. Management Science 60, 2737–2761.

Bańbura, M., Giannone, D., Modugno, M., Reichlin, L., 2013. Now-casting and the real-time data flow. In: Elliott, G., Timmermann, A. (eds.), *Handbook of Economic Forecasting, Volume 2 Part A*, Elsevier, pp. 195–233.

Boyer, B., Nadauld, T. D., Vorkink, K. P., Weisbach, M. S., 2018. Private equity indices based on secondary market transactions. National Bureau of Economic Research Working Paper .

Brito, N. O., 1977. Marketability restrictions and the valuation of capital assets under uncertainty. Journal of Finance 32, 1109–1123.

Brown, G. W., Gredil, O. R., Kaplan, S. N., 2019. Do private equity funds manipulate reported returns? Journal of Financial Economics 132, 267–297.

Brown, G. W., Harris, R. S., Jenkinson, T., Kaplan, S. N., Robinson, D., 2015. What do different commercial data sets tell us about private equity performance? University of North Carolina Working Paper .

Buchner, A., Stucke, R., 2014. The systematic risk of private equity. Passau University Working Paper .

Cochrane, J. H., 2005. The risk and return of venture capital. Journal of Financial Economics 75, 3–52.

De Jong, P., 1988. The likelihood for a state space model. Biometrika 75, 165–169.

Dempster, A. P., Laird, N. M., Rubin, D. B., 1977. Maximum likelihood from incomplete data via the em algorithm. Journal of the Royal Statistical Society: Series B (Methodological) 39, 1–22.

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

Durbin, J., Koopman, S. J., 2012. Time series analysis by state space methods, vol. 38. Oxford University Press.

Ewens, M., Jones, C. M., Rhodes-Kropf, M., 2013. The price of diversifiable risk in venture capital and private equity. Review of Financial Studies 26, 1854–1889.

Franzoni, F., Nowak, E., Phalippou, L., 2012. Private equity performance and liquidity risk. Journal of Finance 67, 2341–2373.

Gârleanu, N., 2009. Portfolio choice and pricing in illiquid markets. Journal of Economic Theory 144, 532–564.

Ghysels, E., Horan, C., Moench, E., 2017. Forecasting through the rearview mirror: Data revisions and bond return predictability. Review of Financial Studies 31, 678–714.

Ghysels, E., Qian, H., 2019. Estimating MIDAS regressions via OLS with polynomial parameter profiling. Econometrics and Statistics 9, 1–16.

Ghysels, E., Santa-Clara, P., Valkanov, R., 2006. Predicting volatility: getting the most out of return data sampled at different frequencies. Journal of Econometrics 131, 59–95.

Giannone, D., Reichlin, L., Small, D., 2008. Nowcasting: The real-time informational content of macroeconomic data. Journal of Monetary Economics 55, 665–676.

Goetzmann, William N. and, E. G., Phalippou, L., 2019. How alternative are private markets? Yale University Working Paper .

Goetzmann, W. N., Gourier, E., Phalippou, L., 2018. How Alternative Are Alternative Investments? The Case of Private Equity Funds. Yale School of Management Working Paper .

Gornall, W., Strebulaev, I. A., 2017. Squaring venture capital valuations with reality. Journal of Financial Economics, forthcoming .

Gredil, O., Griffiths, B. E., Stucke, R., 2014. Benchmarking private equity: The direct alpha method. Tulane University Working Paper .

Gupta, A., Van Nieuwerburgh, S., 2019. Valuing private equity investments strip by strip. New York University Working Paper .

Kalman, R. E., 1960. A new approach to linear filtering and prediction problems. Journal of Basic Engineering 82, 35–45.

Kaplan, S. N., Schoar, A., 2005. Private Equity Performance: Returns, Persistence, and Capital Flows. Journal of Finance 60, 1791–1823.

Kaplan, S. N., Strömberg, P., 2003. Financial contracting theory meets the real world: An empirical analysis of venture capital contracts. The review of economic studies 70, 281–315.

Kohn, R., Ansley, C. F., 1989. A fast algorithm for signal extraction, influence and cross-validation in state space models. Biometrika 76, 65–79.

Korteweg, A., Nagel, S., 2016. Risk-Adjusting the Returns to Venture Capital. Journal of Finance 71, 1437–1470.

Korteweg, A., Nagel, S., 2017. Risk-Adjusted Returns of Private Equity Funds: A New Approach. University of South California Working Paper .

Korteweg, A., Sorensen, M., 2010. Risk and return characteristics of venture capital-backed entrepreneurial companies. Review of Financial Studies 23, 3738–3772.

Kuzin, V., Marcellino, M. G., Schumacher, C., 2013. Pooling versus model selection for nowcasting GDP with many predictors: Empirical evidence for six industrialized countries. Journal of Applied Econometrics 28, 392–411.

Ljungqvist, A., Richardson, M., 2003. The cash flow, return and risk characteristics of private equity. National Bureau of Economic Research Working Paper .

Longstaff, F. A., 2001. Optimal portfolio choice and the valuation of illiquid securities. Review of Financial Studies 14, 407–431.

Miranda, M. J., Fackler, P. L., 2004. Applied Computational Economics and Finance. MIT press.

Nadauld, T. D., Sensoy, B. A., Vorkink, K., Weisbach, M. S., 2016. The liquidity cost of private equity investments: Evidence from secondary market transactions. National Bureau of Economic Research Working Paper .

Nunes, L. C., 2005. Nowcasting quarterly GDP growth in a monthly coincident indicator model. Journal of Forecasting 24, 575–592.

Stafford, E., 2017. Replicating private equity with value investing, homemade leverage, and hold-to-maturity accounting. Harvard University Working Paper .

Woodward, S. E., Hall, R. E., 2004. Benchmarking the returns to venture. National Bureau of Economic Research Working paper .

Appendix

A.1 Review of State Space Models

Following a slightly modified notation from Durbin and Koopman (2012), the general linear Gaussian SSM can be written as

$$\begin{aligned} y_t - \tilde{\Gamma}_t x_t &= \tilde{Z}_t s_t + \epsilon_t, , \qquad \epsilon_t \sim \mathcal{N}(0, \tilde{H}_t), \\ s_{t+1} &= \tilde{G}_t s_t + \tilde{V}_t \eta_t, , \qquad \eta_t \sim \mathcal{N}(0, \tilde{Q}_t), \qquad t = 1, ..., T, \end{aligned} \tag{A.1}$$

where y_t is a p-by-1 vector of data observable at period t with some (or all) entries possibly being missing, x_t—a n-by-1 vector of observable regressors that explain the variation in y_t, s_t—a m-by-1 vector of unobserved state that explains the residual variation in the data (i.e., $y_t - \tilde{\Gamma}_t x_t$) up to mutually and temporarily uncorrelated vectors of normally distributed shocks ϵ_t and η_t. The first equation in (A.1) is referred to as *observation* or *measurement* while the second equation, which describes the evolution of the latent state, is referred to as *state* or *transition equation.* The initial state value s_0 is either known or drawn from a normal distribution as well while being independent from ϵ_t and η_t.

The matrices $(\tilde{\Gamma}_t, \tilde{Z}_t,\tilde{G}_t, \tilde{V}_t)$ and $(\tilde{H}_t,\tilde{Q}_t)$ can all be time-varying but deterministic functions of observed data and/or some parameters θ. Moreover, $\tilde{G}_{t-1}$ and $\tilde{Z}_t$ can depend on $y_1, ..., y_{t-1}$. It is intuitive to think about $\tilde{Z}_t$ and $\tilde{H}_t$ as measuring sensitivity of the observed data to, respectively, the state level and to the noise normalized to unit variance.

The important advantage of the SSM approach to a problem is that, given θ, the system (A.1) enables extracting the expected value estimates of the latent state series s_t. The most famous application of this method for Gaussian SSMs is the *Kalman filter* (Kalman, 1960) in which the Projection theorem leads to the following standard results derived recursively from $t = 1$ to $t = T$:

$$\begin{aligned} s_{t+1}^{filter} &= \tilde{G}_t s_t^{filter} + \tilde{K}_t \Big(y_t^{resid} - \tilde{Z}_t s_t^{filter} \Big) = \mathrm{E}\Big[s_t \Big| Y_t \Big] \\ \tilde{F}_t &= \tilde{Z}_t \tilde{P}_t \tilde{Z}_t' + \tilde{H}_t = \mathrm{var}\Big(y_t^{resid} - \tilde{Z}_t s_t^{filter} \Big| Y_{t-1} \Big) \\ \tilde{P}_{t|t} &= \tilde{P}_t - \tilde{P}_t \tilde{Z}_t' \tilde{F}_t^{-1} \tilde{Z}_t \tilde{P}_t = \mathrm{var}\Big(s_t^{filter} \Big| Y_t \Big) \\ \tilde{P}_{t+1} &= \tilde{G}_t \tilde{P}_{t|t} \tilde{G}_t' + \tilde{V}_t \tilde{Q}_t \tilde{V}_t' \end{aligned} \tag{A.2}$$

where $\tilde{K}_t = \tilde{G}_t \tilde{P}_t \tilde{Z}_t' \tilde{F}_t^{-1}$ is the Kalman Gain, and $y_t^{resid} := y_t - \tilde{\Gamma}_t x_t$ while Y_t denotes the set of the past observations, $(y_1^{resid\,\prime}, \ldots, y_t^{resid\,\prime})'$. The equations in (A.2) are also called the *Forward recursion* as they estimate of state s_t increments forward each period as new information becomes available. However, depending on the problem under consideration, it may be appropriate to condition on the observations from future periods τ, $t < \tau < T$, as well to form the estimate of the state value as of t. This is known respectively as *Kalman smoother* and *backward recursion.* It can be defined as an adjustment to the estimates from *Forward recursion*:

$$s_t^{smooth} = s_t^{filter} + \tilde{P}_t b_{t-1} = \mathrm{E}\Big[s_t \Big| Y_T \Big], \tag{A.3}$$

where $b_{t-1} = \tilde{Z}_t' \tilde{F}_t^{-1} \Big(y_\tau^{resid} - \tilde{Z}_\tau s_\tau^{filter} \Big) + \Big(\tilde{G}_t - \tilde{K}_t \tilde{Z}_t \Big)' b_t$ and: $\tilde{P}_t b_{t-1} = \sum_{\tau=t}^{T} \mathrm{cov}\Big(s_t^{filter}, y_\tau^{resid} - \tilde{Z}_\tau s_\tau^{filter} \Big) \tilde{F}_\tau^{-1} \Big(y_\tau^{resid} - \tilde{Z}_\tau s_\tau^{filter} \Big)$ since η_t and ϵ_t are mutually and temporarily independent.

A final word about estimation, which uses the maximum likelihood principle. Suppose we collect all the parameters governing the matrices $\tilde{\Gamma}_t$, $\tilde{Z}_t$, $\tilde{G}_t$, $\tilde{V}_t$, $\tilde{H}_t$, $\tilde{Q}_t$, $\tilde{G}_{t-1}$, $\tilde{Z}_t$, $\tilde{Z}_t$, and $\tilde{H}_t$, is denoted $\theta \in \Theta$. We will denote the Gaussian likelihood given the data and parameter vector by $\mathcal{L}((y_t, x_t), t = 1, \ldots, T; \theta)$ (see again Durbin and Koopman (2012) for details of the likelihood fn specification).

A.2 SSM Representation of PE Fund - Details

We proceed with casting the equations (1) through (7) as a SSM per equation (A.1):

$$y_t = [d_{t-1} \quad nav_{t-2} \quad r_{c,t}]', \qquad \tilde{\Gamma}_t = I_{3\times 3}, \qquad x_t = [\hat{m}_{t-1} \quad \hat{m}_{t-2} \quad 0]', \tag{A.4}$$

$$\tilde{Z}_t = \begin{bmatrix} 0 & 1 & 0 & log\left(\frac{\delta(\cdot)_t}{1-\delta(\cdot)_t}\right) \\ 0 & 0 & 1 & 0 \\ \beta_c & 0 & 0 & \psi \end{bmatrix}, \qquad s_t = \begin{bmatrix} r_t \\ r_{0:t-1} \\ \bar{r}_{0:t-2} \\ 1 \end{bmatrix}, \qquad \tilde{H}_t = \begin{bmatrix} \sigma_d & 0 & 0 \\ 0 & \sigma_n & 0 \\ 0 & 0 & F_c\sqrt{\hat{h}_t} \end{bmatrix}, \tag{A.5}$$

$$\tilde{G}_t = \begin{bmatrix} 0 & 0 & 0 & \alpha\beta(\cdot)_{t+1} \\ 1 & 1 & 0 & 0 \\ 0 & \left(1-\lambda(\cdot)_{t-1}\right) & \lambda(\cdot)_{t-1} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad \tilde{V}_t = \begin{bmatrix} F \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \tilde{Q}_t = \sqrt{\hat{h}_{t+1}}, \tag{A.6}$$

where d_t is set to missing values whenever $D_t = 0$ (i.e., all but 20-30 weeks), nav_t is set to missing values whenever the NAV report from GPs is not available (i.e., normally 12 weeks out of 13), and

$$\alpha\beta(\cdot)_t := log\Big(\alpha + \beta(R_{mt} - 1) + 1\Big). \tag{A.7}$$

Let us first discuss the state vector s_t appearing in equation (A.5). The first entry of the state vector is defined as $r_t = \alpha\beta(\cdot)_t + \eta_t$, corresponding to the log of equation (1). The last element of the state vector s_t is a constant, while the other two simply accumulate the realizations of r_t since the fund inception at $t = 0$ until $t-1$ or $t-2$, respectively denoted $r_{0:t-1}$ and $\bar{r}_{0:t-2}$, and then produce the exponentially moving average thereof so that they could be related to the observed fund distributions and NAV reports respectively. Note that this introduces nonstationarity into the state space models, since the cumulative sums are trending series. This is a topic worth discussing here. As highlighted by De Jong (1988) and Kohn and Ansley (1989), among others, while state space models can be applied to nonstationary process, provided that the initialization of the Kalman filter is handled properly. It is worth highlighting that we do *not* face this problem because the start values for $r_{0:t-1}$ and $\bar{r}_{0:t-2}$ are in both cases r_0, a stationary series with well defined unconditional mean and variance. This implies that the initialization of the Kalman filter is standard.

Next we discuss the vector of observables, denoted y_t appearing in equation (A.4). The only entry without missing values for any weeks is the comparable asset return, r_{ct}, which ideally should be a portfolio of the publicly traded assets matched to the fund holdings as of the beginning of week t. The weekly observed process r_{ct} relates directly to that key state variable, i.e. the fund latent return series, r_t, through the loading β_c with a heteroskedastic unobserved error, $\psi + \eta_t$.

The other two elements of y_t, namely d_{t-1} and nav_{t-2}, reflect the up-to-period t history of fund cash flows and, therefore, cannot be mapped to the fund since inception returns directly. Instead, they are converted to returns via the mapping function m_t which is implicitly defined by equations (4) and (6). These mapping functions are step-functions that increment whenever the fund makes distributions or capital call (as follows from equation (3), note that it is undefined for the last period since V_T is zero - in this case, m_{T-1} should be used):

$$m_t - m_{t-1} = log\Big((V_t + D_t - C_t)/V_t\Big). \tag{A.8}$$

However, since the true fund asset values are unobserved, the vector of regressors x_t contains the correspondent lags of the best feasible estimates of m_t that are filtered from an auxiliary SSM parametrized with the same θ as the PE fund SSM described in Section 2.4.

The other latent process featuring in both $\tilde{H}_t$ and $\tilde{Q}_t$ is the stochastic variance factor h_t, which scales distributions of idiosyncratic return shocks: η_t and η_{ct}. We obtain a proxy of it from a weekly frequency GARCH(1,1)-filter of comparable asset returns R_{ct} orthogonalized with respect to the risk-factor returns R_{mt}. We estimate the GARCH models separately and ignore the estimation uncertainty caused by the plug-in method we use for volatility in the model.

A.2.1 NAV smoothing and noise

Equations (5), (6) and (8) assume that the NAV appraisal process is governed by a parametric function $\lambda(\cdot)_t$ and a mean-zero noise ϵ_{nt} parametrized by time-invariant σ_n. It is important to note that smoothing pertains to both systematic and idiosyncratic component of fund returns and accounts for the path of returns *between* the quarterly report dates as t indexes weeks. Following equation 8, $\lambda \in (0.01, 0.99)$ is the structural parameter to be estimated while $w_t \in [0, 1]$ is the observed values reflecting the fraction of the fund assets that the cash flow in period t comprises. Below we provide some examples.

Consider a period when distribution D_t fully resolves the fund so that residual NAV is known to be zero. For this case, $\tilde{w}_t = D_t/(N\hat{A}V_t + D_t) = 1$. Consequently, from equations (5) and (8), it follows that the "smoothed" return to-date equals the true one regardless of λ:

$$\begin{aligned}\bar{r}_{0:t} &= \Big(1 - \lambda(1 - \tilde{w}_t)\Big) r_{0:t} + \lambda(1 - \tilde{w}_t)\bar{r}_{0:t-1} \\ &= \Big(1 - \lambda \cdot 0\Big) r_{0:t} + \lambda \cdot 0 \cdot \bar{r}_{0:t-1}\end{aligned}$$

Suppose that $w_t = 0.5$ instead, which would be the case if C_t doubles the fund capital invested or D_t resolves the fund by half (i.e., $D_t = N\hat{A}V_t$). If the fund GPs put only half of the weight on past performance (i.e. $\lambda = 0.5$), then the smoothed return to-date would equal:

$$\begin{aligned}\bar{r}_{0:t} &= \Big(1 - 0.5(1 - 0.5)\Big) r_{0:t} + 0.5(1 - 0.5)\bar{r}_{0:t-1} \\ &= 0.75 \cdot r_{0:t} + 0.25 \cdot \bar{r}_{0:t-1}\end{aligned}$$

as opposed to $0.5 \cdot r_{0:t} + 0.5 \cdot \bar{r}_{0:t-1}$ for a period without any cash flows and $0.65 \cdot r_{0:t} + 0.35 \cdot \bar{r}_{0:t-1}$ had the GPs been smoothing at the $\lambda = 0.9$ rate (given $w_t = 0.5$). Given $\lambda = 0.9$ and conditional on absent cash flows during the quarter (i.e., 13 weeks), the quarter-ago valuations would be contributing 25.4% ($= 0.9^{13}$) to the current fund NAV reports.

Equation (8) therefore models the fund asset appraisal process to reflect both, the observed return paths (for r_{mt} and r_{ct}) and the unobserved idiosyncratic fund returns, η_t, to be are most consistent with (i) the realized cash flow pattern and (ii) the history of NAV reports. All else equal, funds that made substantial distributions recently would be assumed to have less "stale" NAVs. Meanwhile, the weight of returns occurring before period τ, $0 < \tau < t$ in NAV_t is limited by the fraction of fund capital invested before period τ. While under equation (8) the influence of the past valuations decays at an exponential rate, it is also straightforward to incorporate "longer memory events" in our framework by temporarily shifting $\lambda(\cdot)_t$ up post specific events, such as highly related comparable M&A transaction or IPOs, or by expanding the state-space vector outright.

The Gaussian NAV residual noise parametrized by σ_n is orthogonal to the observed returns and the distribution process. While the temporal independence assumption appears implausible for actual PE funds, we argue that the sparseness in the NAV observation series (i.e., once every 13 weeks at best) and the irregularity of fund cash flows (and, hence, w_t values) mitigate the consequences of possibly non-zero temporal correlation structure in the residuals for the identification of the fund model parameters. The entry corresponding to the NAV observations in matrix y_t is simply missing for most observations and almost never overlaps with a non-missing value for d_t. To better illustrate this, we interpolate ϵ_{nt} between the values drawn for the adjacent quarterly NAV reports to obtain the between quarter estimate of $N\hat{A}V_t$ in a simulation exercise reported in Section A.4. In addition, we investigate the effect of $\lambda(\cdot)_t$ being misspecified in the estimations relatively to the data generating process.[25] One can also consider making σ_n time-varying to reflect the stage of current fund holdings and differences in GP reporting incentives (e.g., see Brown et al., 2019). The latter can be also modeled as a conditional mean of the distribution of ϵ_{nt}. Similarly, it is straightforward to incorporate data about carry rule to further reduce noise in the estimation.

[25] If the econometrician also observes dates and amounts with regards to the individual investments in the fund portfolio, $\lambda(\cdot)_t$ can be further refined to reflect the return histories during the remaining investments only. Therefore, the 'First-In-First-Out'-assumption that expression (8) implements is no longer needed.

A.2.2 Cash flow process

According to equations (7) and (9), the fund distribution process is governed by the parametric function $\delta(\cdot)_t$ and a mean-zero noise ϵ_{dt} parametrized by time-invariant σ_d. By construction, they are capturing the variation in the fund distributions that cannot be explained the asset return path, on the one hand, and the periodically observed NAV series, on the other hand. Also note that fund capital calls do not relate to either $\delta(\cdot)_t$ or ϵ_{dt} but only affect the asset-to-value mapping function, m_t, discussed later in this section.

While equation (9) certainly misspecifies $\delta(\cdot)_t$ relatively to the true conditional expectation (and, therefore inducing potentially heavy temporal dependencies in the residuals), we argue that irregularity in distribution occurrence mitigates the violations of the SSM identifying assumptions discussed in the beginning of this section. As with NAV smoothing process, we scrutinize this conjecture in our simulations discussed below.

Another insightful extension to the baseline setup in expressions (1–7) would be utilization of prices (or quotes) from the secondary market transaction with the fund stakes (e.g., see Nadauld et al., 2016). As are the fund distributions (in fact, they are equivalent from the LPs perspective), these prices would map to the fund unsmoothed returns, but one-to-one:

$$log(\text{SecondaryPrice}_t) + m_t = r_{0:t} + \epsilon_{pt} \quad , \tag{A.9}$$

whereby ϵ_{pt}, the price distortion effects, can be either assumed to be zero, or modeled accordingly to incorporate time-varying demand and supply variation in these still nascent markets (e.g., see Albuquerque et al., 2018). Similarly, one can incorporate venture deals round valuation data—if the fraction of reported fund-level NAVs that the deal corresponds to is known.[26]

A.3 Estimation details

R_{ct}-interpolated NAVs. For any $q^- < t \leq q$ where q and q^- are last weeks of two adjacent quarters 'R_{ct}-interpolated NAVs' are defined as follows:

$$\tilde{V}_t^{ip} = NAV_{t=q^-} e^{(t-q^-)\psi_q} \prod_{q^- < \tau \leq t} R_{c\tau} + \sum_{q^- < \tau \leq t} \Big((C_\tau - D_\tau) e^{(t-\tau)\psi_q} \prod_{\tau < p \leq t} R_{cp} \Big) \tag{A.10}$$

where ψ_q is the quarter-specific drift that makes eq. (A.10) hold if $\tilde{V}_t^{ip}$ is replaced with $NAV_{t=q}$.

α-range center for profiling. For center of the α-range, we compute IRRs from the discounted cash flows as do Gredil, Griffiths, and Stucke (2014):

$$\alpha_{p0} = \arg\min_{\alpha} \sum_{n=0}^{T} \Big((C_n - D_n) \prod_{n < t \leq T} \text{proj}\{R_{ct}|R_{mt}\} \cdot (1+\alpha)^{-n} \Big) \tag{A.11}$$

and approximate fund duration with a weighted average life when IRR solutions are ambiguous. In work contemporaneous to ours, Korteweg and Nagel (2017) show that using average beta estimates improves inference about fund-level abnormal performance under the Generalized PME framework in Korteweg and Nagel (2016) even if fund βs are heterogeneous. Therefore, to obtain the center of the α-grid, we use the projection of fund-matched R_{ct} on R_{mt} for the discount factor construction.

Penalty function For each combination of α and β on the profile grid (i.e., grid cell ij), we obtain the maximum likelihood estimates of ϑ_{ij}^{pout} and standardize the Loglikelihood scores by subtracting the grid-wide min and dividing by the grid-wide standard deviation to obtain $sLL(\alpha_i, \beta_j) \equiv sLL_{ij}$. From each sLL_{ij}, we then subtract a penalty function $sPMEerr_{ij}$ which is the grid-wide stan-

[26] Provided that the contractual terms are comparable or adjusted accordingly (see Gornall and Strebulaev, 2017).

dardized $PMEerr_{ij}$ defined as:

$$PMEerr_{ij} = \Big| PME_{ij} - 1 \Big| \tag{A.12}$$

$$PME_{ij} = \frac{\sum_{t=0}^{T-1} D_t \prod_{\tau=t+1}^{T} \Big(1 + \alpha_i + \beta_j (R_{mt} - 1)\Big) + \hat{V}_T + D_T}{\sum_{t=0}^{T-1} C_t \prod_{\tau=t+1}^{T} \Big(1 + \alpha_i + \beta_j (R_{mt} - 1)\Big) + C_T} \tag{A.13}$$

to obtain pLL_{ij}, where $\hat{V}_T = NAV_T + \hat{bias}_T$ and $\hat{bias}_t$ is the expected appraisal bias at t defined in equation A.14. Picking the supremum value of pLL_{ij} on the grid identifies the corresponding (α_p, β_p)-pair which we use as fixed data in the second step in which we estimate remaining parameters, $\vartheta \in \theta$, as well as mapping-to-value functions and weekly fund returns.

Mapping-to-Parameter iterations. Because our latent state is fund returns and future fund distributions inherently related to today's returns, we argue that the *backward recursion* method is the most natural choice in our context (i.e. nowcasting). We therefore drop the superscript ^{smooth} and instead superscript it the per period and cumulative returns estimates with the set of SSM parameters that they were derived upon, e.g. $r\theta_t$, $r^{\theta}_{0:t}$, $\bar{r}^{\theta}_{0:t}$. It follows from equations (4) and (6) that the expected appraisal bias and the the unbiased net asset values (i.e. V_t) for any week t are:

$$bias(\theta)_t := r^{\theta}_{0:t} - \bar{r}^{\theta}_{0:t} \tag{A.14}$$

$$V_t(\hat{m}_t; \theta)_t = \exp\{r^{\theta}_{0:t} - \hat{m}_t\} \tag{A.15}$$

Meanwhile, for weeks t when NAV reports are observed (i.e., $t \in |t_{eq}|$), the *debiased NAV* values have to satisfy:

$$V_t(NAV_t; \theta)_t = NAV_t e^{-bias(\theta)_t} \Big| \quad t \in |t_{eq}| \tag{A.16}$$

which we use in combination with expression A.8 to obtain $\hat{m}_t$ series. However, this makes the mapping m_t a function of parameters θ and equation (A.15) a fixed point. Furthermore, $\hat{m}_t$ enter the observation vector x_t and, therefore, potentially affects the parameter values under which the SSM logLikelihood is maximized. Therefore, starting with the $\hat{m}_t$ implied by the R_{ct}*-interpolated NAVs* ($\equiv m_t | (r_{c1}, \ldots, r_{cT})$), and the corresponding ϑ estimates, we set a loop over i:

$$\begin{aligned} m(i)_{t=1,\ldots,T} &= m(\vartheta^{argMax}_{i-1}) \\ \vartheta^{argMax}_{i} &= \arg\max_{\vartheta} LL_p(\vartheta; m(i)_{t=1,\ldots,T}) \end{aligned} \tag{A.17}$$

in which $m(i)_t$ and ϑ^{argMax}_i are re-estimated sequentially until the average change in ϑ^{argMax}_i from ϑ^{argMax}_{i-1} is smaller than 5% on average across all ϑ elements, whereby LL_p denotes the SSM likelihood function conditional on $\theta \notin \vartheta$ being fixed at values determined in the first step.

A.4 Monte Carlo experiment

In this section, we demonstrate our estimation methodology and its nowcasting performance using simulated data at weekly frequency. We begin with a single-fund case study and then present evidence aggregated over a random sample of such funds.

A.4.1 Single-fund case study

This particular fund operated for 626 weeks. It made 18 capital calls over the first 215 weeks for the total of \$5.25mln (including \$1mln at inception, i.e. week 0) and 25 distributions of various amounts for a total of \$17.71mln at irregular intervals, starting on 110th week.

Panel A of Figure A.1 plots these data over fund life-weeks, along with the "True Asset Values" (i.e. V_t in equation 3)—unobservable with real fund data—as well as the quarterly NAV reports that reflect appraisal smoothing as in equation (8) at λ of 0.90 and the reporting noise volatility, σ_n, of 5.0%. We can see that this process generates both upward and downward biases in reported NAVs (relatively to V_t), depending on the recent path in fund returns, as plotted in Panel C.

Panel B rescales the variables by the fund asset value so that 'D/(D+V)' correspond to $exp\{\delta(\cdot)_t + \epsilon_{dt}\}$ in equation 7 while 'N/V-1' is the overall reporting distortion, $exp\{\bar{r}_{0:t} - r_{0:t} + \epsilon_{nt}\} - 1$ in equations (4–6). The panel also plots similar statistic for weekly series obtained by interpolating as reported NAVs at the rate of comparable asset returns R_{ct} while accounting for the cash flows between the quarters (henceforth, 'R_{ct}-interpolated NAVs', see equation (A.10) below). Essentially, this is a nowcast of fund NAVs that assumes no appraisal bias present and a unit β_c as in equation (2), providing for a natural benchmark to evaluate the SSM-based nowcasts.

The since-inception return corresponding to 'R_{ct}-interpolated NAVs' is plotted against the true since-inception returns ($R_{0:t}$ in equation (4)) in Panel C. By construction, the two series are tightly correlated and occasionally coincide (e.g., around 250th and 350th weeks). Even though the fund is fully resolved by the 626th week and the expected per-period difference in idiosyncratic risk realization—η_{ct} in equation 2—is zero, the series diverge notably towards the end of fund life nonetheless. The reason for this divergence is the assumptions embedded in the interpolation procedure in equation (A.10), namely: (i) reported NAVs reflect the true value of fund assets, and (ii) distributions reduce remaining NAVs one-for-one.

As evident from Panel B and C, the series based on NAV interpolation overshoot whenever the fund makes distributions following the quarter with downward-biased NAVs. Intuitively, the returns during periods following these distributions have to be higher than true ones to "catch-up" with the subsequent NAV report, especially if the latter is less [more] negatively [positively] biased than the previous. These dynamics are present but concealed with since-inception money multiples (also plotted in Panel C), which unlike $R_{0:t}$ are V_{t-1}-weighted per period return, therefore exhibiting very little variation after V_t gets small in front peak values (around week-200 in this case).[27]

Figure A.2 depicts the SSM likelihood scores and the pricing error via 3-dimensional mesh plots on Panel A. Both are normalized by subtracting the grid-wide mean and dividing by standard deviation. It follows from Panel A that the SSM likelihood scores do peak with respect to β around the value of 1.20 but hardly exhibit any curvature along the α-axis. In contrast, the flat region in the surface of pricing errors is off-diagonal.[28] Therefore, subtracting these pricing errors from the SSM likelihood scores—i.e., "penalizing the SSM likelihood"—tends to result in a sharp peak across both dimensions, as plotted on Panel B. In our case, the maximum corresponds to the (α, β) pair of (0.308%,1.15)—close to the OLS estimates of (0.327%,1.26) that one would obtain if weekly $(R_t - 1)$ were regressed on $(R_{mt} - 1)$, especially when compared to those from quarterly OLS estimates feasible with real fund data which are plotted as the blue rectangle in Panel B.[29]

As discussed in the main text, having identified the pair of α and β, we iteratively estimate the asset-to-value mapping ad the remaining parameters. For this case study fund, this mapping-to-parameter convergence took 12 iterations and resulted in the following $(\alpha, \beta, \vartheta)$-vector, whereby standard errors were obtained using numerical differentiation (see, e.g., Miranda and Fackler, 2004):

	α	β	δ	σ_d	λ	σ_n	F	F_c	ψ	β_c
Estimate	.0031	1.154	.0432	1.488	.7918	.0121	2.014	.7853	.0021	.713
SE(Est)	.0043	0.643	.0141	1.349	0.0805	.0014	0.832	.108	.0018	0.182

Panel A of Figure A.3 examines the quality of per period fund return estimates for our case fund by plotting the difference between $\hat{r}_{0t}$ obtained from the SSM parametrized with θ-values tabulated above and the true r_{0t} as a solid red line. In is natural to compare it to a similar difference obtained from the intuitive approach based on 'R_{ct}-interpolated NAVs' plotted with dotted red line here (obtained by subtracting the solid line from the dashed one in Panel C of Figure A.1). It

[27]See Brown et al. (2019) for a discussion on inference about NAV appraisal bias with multiples.

[28]That is typically the case with an upward-trending market return—without regularly spaced intervals, high cumulative returns can be "equally well explained" by either high slope or large intercept.

[29]For example, regressing quarterly 'cash flow'-adjusted quarterly growth in NAV on contemporaneous and lagged market returns yields a beta estimate of at most 0.84 and 0.4 if weekly returns based on 'R_{ct}-interpolated NAVs' are used.

follows that the SSM-based series tend to be closer to the true fund returns during and after the parameter estimation period. The panel also compares three Asset-to-Value mapping functions for our case study fund: the initial guess, the true, and iteratively estimated one as discussed above. The latter two tend to be closer indicating that the SSM provides for a better guess of the true asset values on the cash flow dates.

Panel A of Figure A.3 also conveys the general intuition of our approach—our fund return prediction for a week is a scaled weighted average of r_{ct} and r_{mt} unless there is a NAV report or distribution that justify $\eta_{ct} \neq 0$ (see equation 2).[30] A distribution also shifts the mapping function up in proportion to its size relatively to the remaining asset value estimate.

As for nowcasting performance metrics introduced in Section 2.3, the POS and FOS RMSE for the estimated model-based series are, respectively, 0.066 and 0.063. Both compare favorably with the respective metrics based on the *Naïve nowcast* of fund approach (see A.10).

Finally, we note that Figure A.3 there appears to be no sharp change in the SMM performance in the post-estimation period (397th week). This is to be expected in the case the SSM parameters are stable during the fund life which might not necessary be the case with real-life PE funds. Either way, to the extent the choice of parameter estimation window length is not identifying structural breaks, one can consider the in-sample PME(SSM) values as a predictor of POS and FOS metrics. Specifically, we compute mean squared deviation from one for the during 156 weeks before t_b.

A.4.2 Simulated panel of funds

The case study fund is a draw from the following data generating process. First, we simulate market returns R_{mt} from a normal distribution with GARCH(1,1) volatility process corresponding to the unconditional volatility of 20% per annum.[31] We then create 25 vintage years each featuring four funds that operate for at most 13 years (=676 weeks) and have $\beta \sim \mathcal{N}(1.25, 0.125)$ with respect to R_{mt} and $\alpha \sim \mathcal{N}(0, 0.0022)$, both fixed during fund life. In addition, each fund returns are driven by mean-zero shocks drawn at weekly frequency as a mixture of two heteroskedastic normal shocks, one of which is common within vintage but temporarily independent still, and Poisson jumps (see next paragraph). Meanwhile the benchmark returns, R_{ct}, are obtained by averaging the common components of fund returns within each vintage-week (i.e. $1 + \beta_i(R_{mt} - 1)$ plus the common shock).

To construct fund non-R_{mt} returns, η_t, for each week we draw two independent normal shocks from $\mathcal{N}(0, h_t)$ where $log(h_t) \sim \mathcal{N}(\rho_0 + \rho_1 log(h_{t-1}), vv^2)$. We add the two shocks together assuming that the vintage-wide shock comprises 1/3 and the fund-specific—2/3. This implies that if we use R_{ct}-based estimates in place of $\hat{h_t}$, the true value of the volatility scale F equals 2 while the $\mathrm{E}\eta_t^2 = 0.40$.[32] In addition, in spirit of simulations in Korteweg and Nagel (2017), we assume that each week there is 10% probability of a 10% idiosyncratic increase in fund asset value. We draw these jumps from a Poisson distribution, add them to the normal shocks, and demean the sum to zero over each fund life before adding to $1 + \alpha_i + \beta_i(R_{mt} + 1)$ for each fund-week t to obtain R_t.

Next we construct fund cash flows and asset values. Fund cash flow occurrences are drawn from a Poisson distribution parametrized such that there is on average 15 capital calls lasting through the 225th week and 20 distributions per fund starting with 105th week. The first capital call is equal to 1 with subsequent capital calls (conditional on occurrence) equal to 0.25. Fund distribution process follows equation (9) with δ set to 0.04 while $\sigma_d = 1.20$. The distribution amounts are determined sequentially with fund true asset values following equations (3–7). Each fund life ends (i.e., $t = T$) on either 676th week or the week in which its true asset values drop below 0.1 (i.e. 10% of the initial capital call) with the last distribution set to equal V_T (and the latter replaced with zero).

Finally, the quarterly observed NAVs are constructed using equations (5–6) by setting λ to 0.9 and drawing observation noise, e_{nt}, from $\mathcal{N}(0, 0.05)$ independently for each fund-quarter. When constructing the weekly R_{ct}-*interpolated NAVs* to estimate change in mapping in equation (A.8) and $\lambda(\cdot)_t$-process in equation (8), we interpolate the values of e_{nt} between the adjacent quarters inducing the same autocorrelation pattern in the estimands as if it were real PE fund data.

[30] This can also be a secondary trade in fund stake, or a comparable exit, or a financing round for the portfolio company.

[31] The persistence terms on squared return and lagged variance are 0.10 and 0.85, respectively.

[32] $\rho_0{=}-0.5784, \rho_1{=}0.9, vv{=}0.40 p.a.$, to match the idiosyncratic return volatility estimates in Korteweg and Sorensen (2010).

Table A.1 reports estimation results for the main parameters of interest and the pricing error criteria using the holdout fund-weeks as discussed in Section 2.3. First seven columns of each panel report the means and the medians of parameter estimates across the 100 of simulated funds as well as the first, second, and third quartile of the root mean squared error of the fund estimate relatively to the true parameter value. Last four columns report PME($\hat{V(\theta)}, r(\theta)$), denoted with subscript $fund$, and PME(NAV, r_c), denoted with subscript bmk, for pseudo out-of-sample and fully out-of-sample case. The sample of 100 simulated funds is common across panels, therefore POS_{bmk} and FOS_{bmk} are the same panel-wise reflecting same NAV_t and r_{ct} series.

Panel A reports *Direct Estimates* as in Section 2.4.1 (henceforth, *Baseline*) and reveals that estimated parameters are typically close to the true values, exhibiting minimal bias for $(\beta, \lambda\, \sigma_n)$, in particular. A notable upward bias as well as high levels of RMSE relatively to the true parameter values are evident for σ_d and, perhaps, for δ. Panel A also shows an advantage of using fund return and asset values estimates (over the matched benchmark and as reported NAVs) as both, POS and FOS, PME($\hat{V(\theta)}, r(\theta)$) tend to be closer to 1 than PME(NAV, r_c), while median FOS RMSE at 0.1171 is less than half of 0.2841 corresponding to the naïve approach.[33]

In Panel B, we assume the distribution function $\delta(\cdot)_t$ to be known and feature no uncertainty $\sigma_d = 0$. As discussed in Section A.2.2, this corresponds to a situation in which one can observe the fund individual investments' valuation reports and exit terms. Unsurprisingly, this helps with estimation precision, especially, with respect to the variance parameters—F (σ_n) is no longer downward (upward) biased while the associated RMSEs are smaller by a factor of 4 to 15. Interestingly, the presence of jumps in the fund return process does not prevent consistency in estimating the volatility scale parameter via MLE which assumes conditionally Gaussian returns. The gains in precision are also notable for β estimation whereby RMSEs drop by half, as well as for the smoothing parameter, λ. All these translate in better nowcasting performance as POS(FOS) 75th percentile RMSEs drop to 0.0093(0.1298) from 0.0230(.2368) in Panel A.

In Panel C, we misspecify $\delta(\cdot)_t$ by assuming in the estimation that $\delta(\cdot)_t$ is proportional to the time since last distribution (rather than the time since inception) as in Korteweg and Nagel (2016). As discussed in Section A.2.2, this induces autocorrelation in the error term of the distribution equation 7. It follows that this has virtually no adverse effects on the identification of the model parameters relatively to *Baseline*, other than δ and σ_d themselves. Neither does it materially affect the nowcasting performance relatively to *Baseline*, consistent with the case study evidence.

In Panel D, we implement the partial imputation approach discussed in Section 2.4.2. While we do observe improvements to RMSE on a few parameters including β, these are somewhat predetermined by the homogeneity in the parameters across funds, and do not translate in meaningful improvements with regards to nowcasting.

Figure A.4 plots the individual fund estimates of α and β from our SSM against the corresponding OLS estimates obtain with weekly returns (infeasible with the actual PE data). Panel A shows that profiled α are well aligned with what one could have estimated if PE funds weekly returns were observable, even though slightly downward biased. For β, we continue to observe a positive relation between the two sets of estimates although much greater noise is present.

Finally, Figure A.4 compares the nowcasting performance as per *Baseline* estimations against the naïve approach that utilizes as reported NAVs and matched benchmark returns. Panel C focus on pseudo out-of-sample metric while Panel D—on fully out-of-sample ones, as discussed in Section 2.3. For better visualization, we utilize the logs of variances so the magnitudes are harder to read than RMSEs from Table A.1. However, it is clear from that our SSM approach results in better nowcasts for a significant majority of simulated funds regardless of the metric.

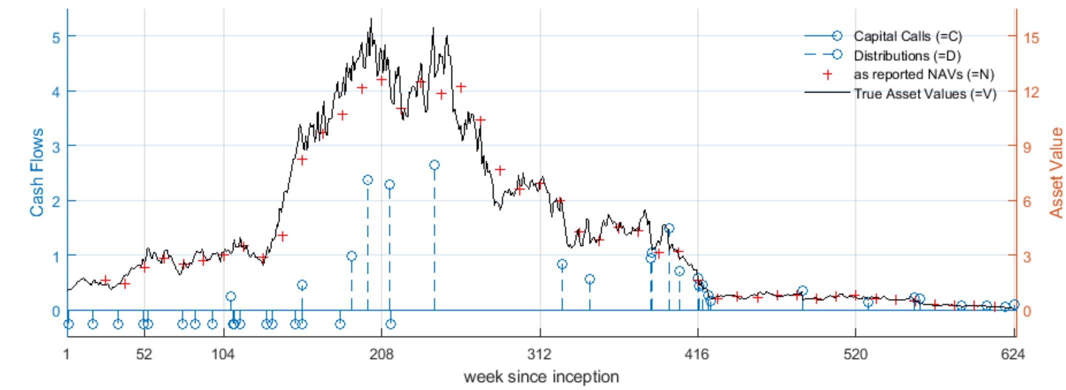
Panel C of Figure A.3 compares the nowcast performance of the SSM parametrized with these θ for our case study by plotting the quarterly RMSEs for NAV estimation error against those from the same SSM parametrized with *fund-specific estimates* referred to as *Baseline*. We see that NAV estimates based on *Partially imputed* θ are often smaller but typically close to the *Baseline*. The improvement is due the fact that the peer-imputed (λ, F, σ_n) are close to the true values than those estimated directly, which may or may not be true with real-life funds. For comparison, Panel C also plots quarterly NAV errors for the case when the distribution function $\delta(\cdot)_t$ need not be estimated which—as discussed in Section A.2.2—could be the individual deal-level transparency at the fund.

[33] In untabulated analysis, we review an estimator which profiles α only. It results in larger nowcasting errors.

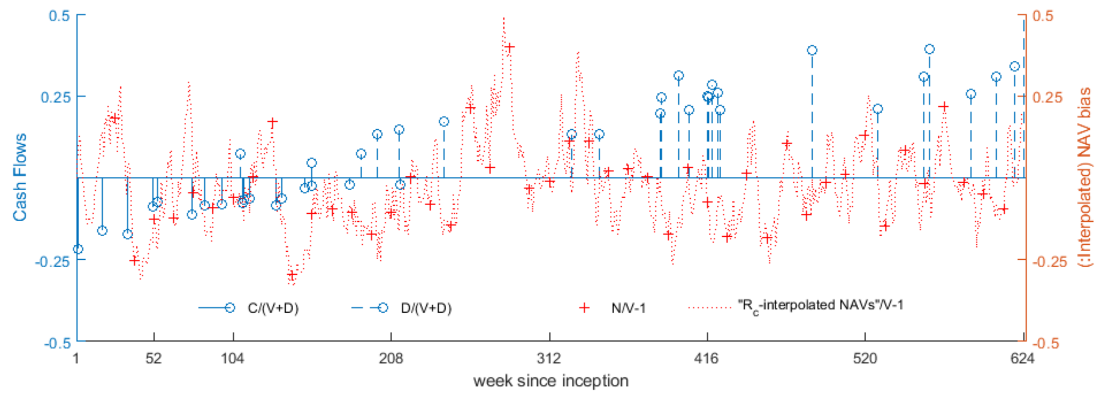
Figure A.1. Case study: Data for a hypothetical fund

This figure reports data for a hypothetical PE fund to illustrate our method applications. See Section A.4.1 for description.

Panel A: Cash flows and reported NAVs normalized by first capital call



Panel B: Cash flows and reported NAVs normalized by Asset Values



Panel C: Cumulative returns and Multiples

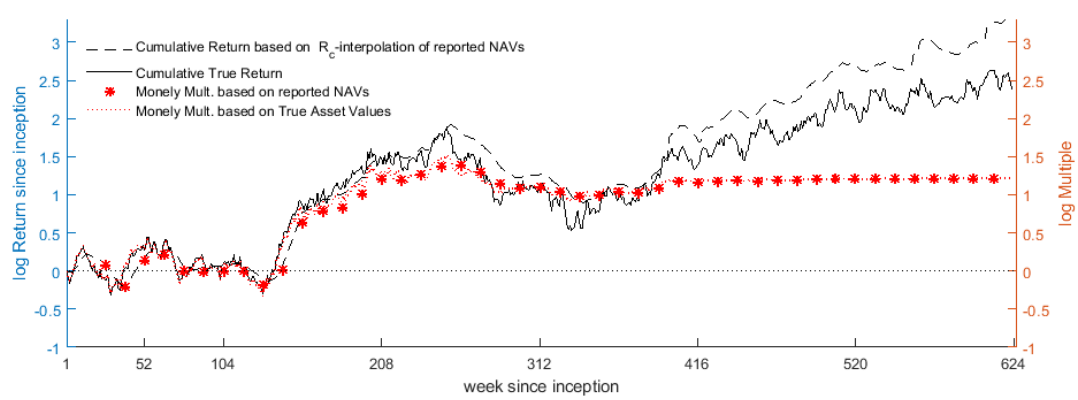
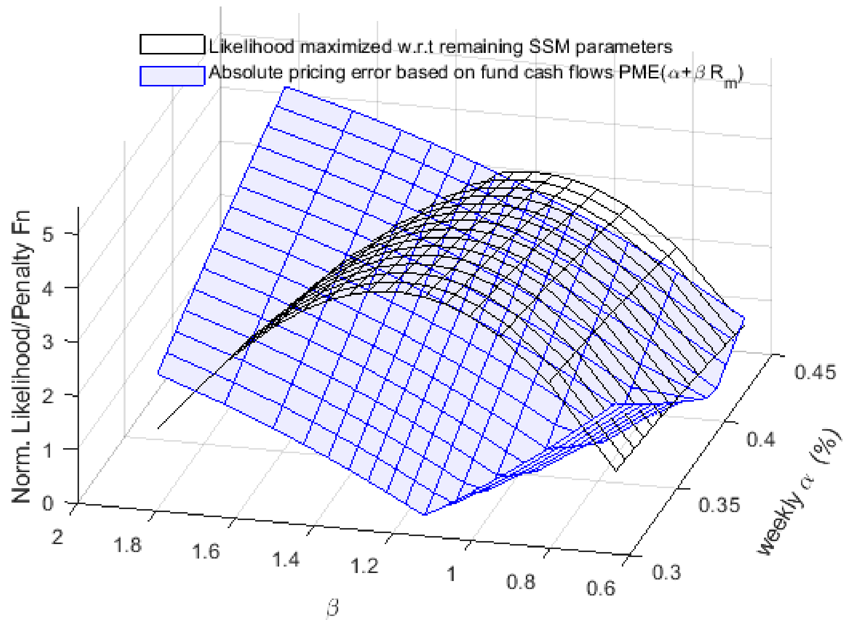


Figure A.2. Case study: Profiled Likelihood

This figure illustrates selected parameters' identification for a hypothetical PE fund to illustrate our method applications. See Section A.4.1 for description.

Panel A: SSM likelihood and the Penalty function



Panel B: Penalized likelihood

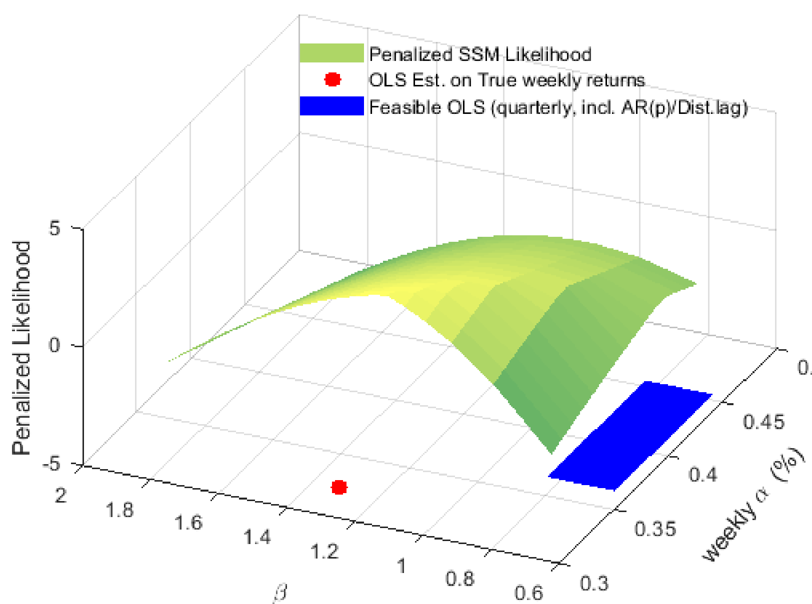
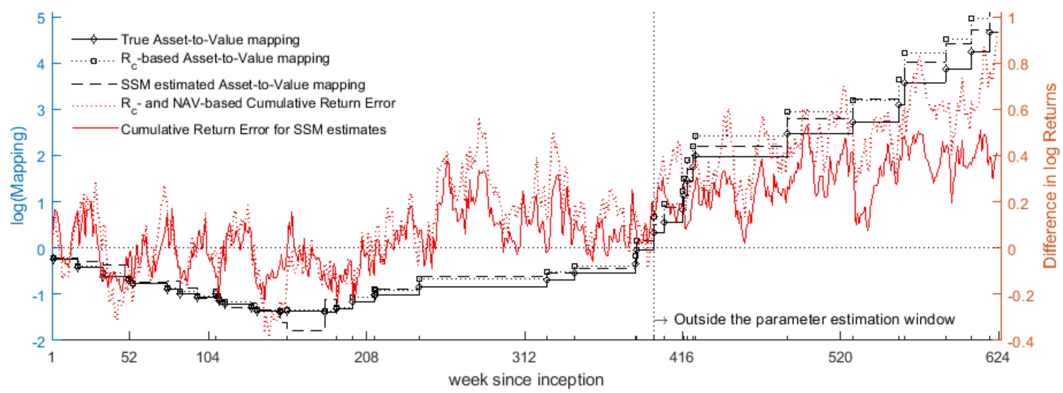


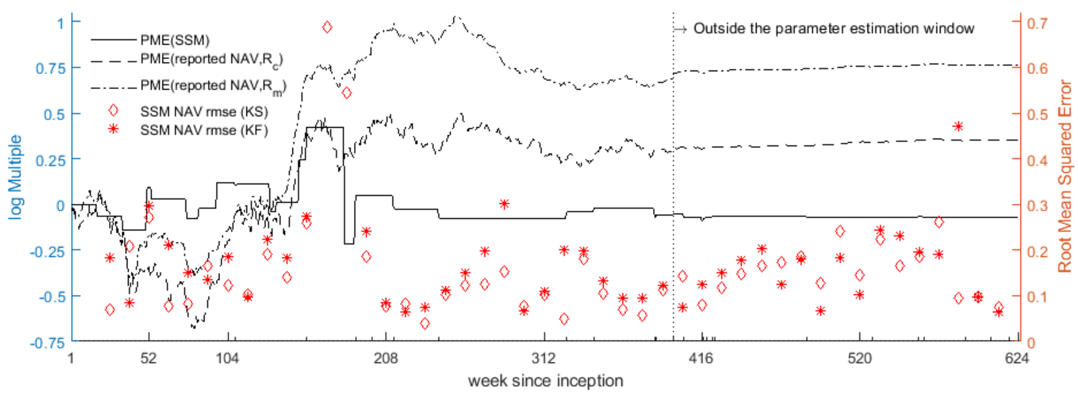
Figure A.3. Case study: Nowcasting results

This figure demonstrates nowcasts obtained from a SSM for a hypothetical PE fund to illustrate our method applications. See Section A.4.1 for description.

Panel A: Return estimates and Mapping to Asset Values



Panel B: Asset Value estimates and PMEs



Panel C: Asset Values estimates from models with fewer parameters

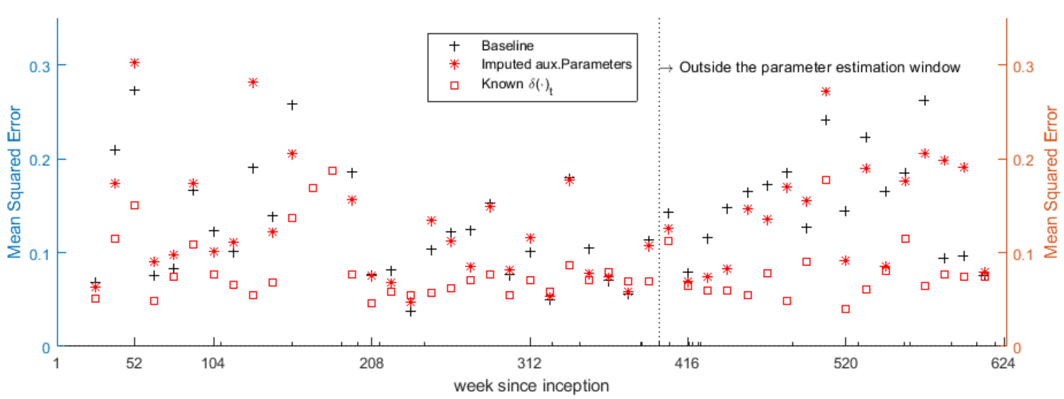
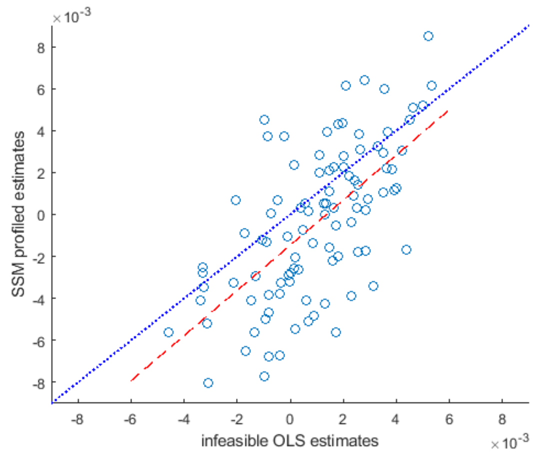


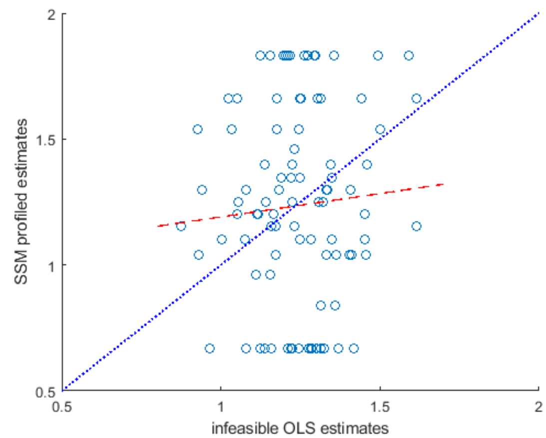
Figure A.4. Monte Carlo experiment: cross-section of α,β and nowcasting performance

This figure reports SSM estimates on simulated fund data described in Section A.4.2. The dotted and dashed lines are, respectively, the 45-degree and the least-squared fit lines. Panel A (B) plots the OLS estimates—which are BLUE but infeasible with the real data—of funds' α (β) on x-axis against the SSM-based estimates on y-axis. A positive association and low variance across y-axis indicate good performance of our method. Whereas Panels C and D plot the log of mean squared error obtained using matched benchmark returns on x-axis against SSM-based estimated fund returns on y-axis. Panel C (D) uses Pseudo (Fully) Out-of-Sample metric discussed in Section 2.3. Observations below the 45-degree line indicate outperformance of our method over a naïve approach that simply uses public benchmark returns to model fund NAVs.

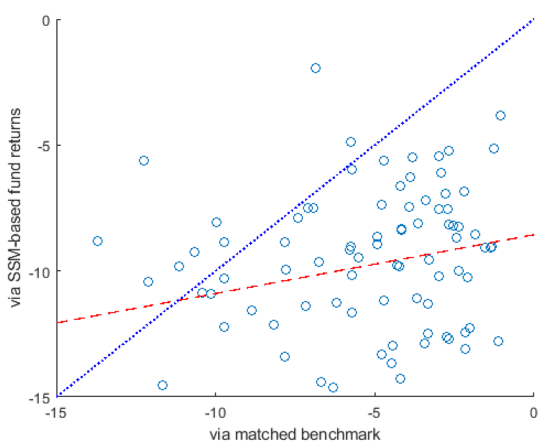
Panel A: α estimates



Panel B: β estimates



Panel C: log MSE POS



Panel D: log MSE FOS

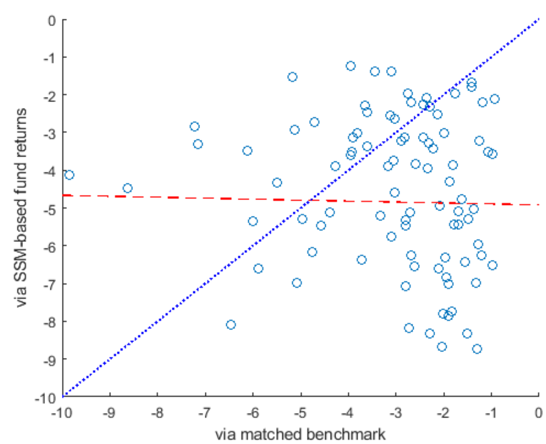


Table A.1. Monte Carlo experiment: parameter estimates and nowcasting performance

This table reports SSM estimates on simulated fund data described in Section A.4.2. Each reports seven key parameters of interest as well Pseudo-Out-of-Sample (POS) and Fully-Out-of-Sample (FOS) nowcasting results as discussed in Section 2.3, whereby subscripts *bmk* and *fund* denote to the matched benchmark return and SSM-filtered fund return series, respectively. 'Est. mean' and 'Est. median' refer to the simulated fund sample mean and median estimate of the respective parameter (or of PME against the respective return series for POS/FOS columns). 'RMSE p25/p50/p75' refer to the respectively percentile of the Root Mean Squared Error relatively to the true parameter value (or 1 for POS/FOS columns). Panel A reports the results for baseline estimation approach, Panel B treats distribution function as known, Panel C misspecifies the estimated distribution function relatively to the simulated data. Panel D reports results for partially matched parameter estimates.

Panel A: Baseline

	α	β	δ	λ	F	σ_d	σ_n	POS_{bmk}	POS_{fund}	FOS_{bmk}	FOS_{fund}
Est. mean	-0.0006	1.2306	0.0485	0.8722	1.7111	2.0340	0.0524	1.4805	1.0422	0.9204	0.9911
Est. median	-0.0002	1.2264	0.0470	0.8955	1.4746	2.0364	0.0547	1.1250	1.0093	0.8547	0.9750
RMSE p25	0.0008	0.1406	0.0133	0.0192	0.4006	0.6812	0.0169	0.0333	0.0022	0.1525	0.0418
RMSE p50	0.0022	0.3591	0.0170	0.0544	0.7147	0.8812	0.0358	0.1250	0.0093	0.2841	0.1171
RMSE p75	0.0038	0.5730	0.0225	0.0802	1.4178	1.0314	0.0490	0.3069	0.0230	0.4070	0.2368

Panel B: Known $\delta(\cdot)_t$

	α	β	δ	λ	F	σ_d	σ_n	POS_{bmk}	POS_{fund}	FOS_{bmk}	FOS_{fund}
Est. mean	-0.0005	1.1978	-	0.8774	2.1717	-	0.0484	1.4805	1.0098	0.9204	0.9952
Est. median	-0.0002	1.2500	-	0.9000	2.0000	-	0.0500	1.1250	1.0037	0.8547	1.0046
RMSE p25	0.0008	0.0961	-	0.0002	0.1000	-	0.0001	0.0333	0.0012	0.1525	0.0313
RMSE p50	0.0022	0.2099	-	0.0086	0.2417	-	0.0026	0.1250	0.0037	0.2841	0.0712
RMSE p75	0.0038	0.3621	-	0.0264	0.5690	-	0.0050	0.3069	0.0093	0.4070	0.1298

Panel C: Misspecified $\delta(\cdot)_t$

	α	β	δ	λ	F	σ_d	σ_n	POS_{bmk}	POS_{fund}	FOS_{bmk}	FOS_{fund}
Est. mean	-0.0006	1.2433	0.2308	0.8743	1.8297	3.8015	0.0502	1.4805	1.0428	0.9204	1.0396
Est. median	-0.0002	1.2264	0.2307	0.9065	1.5543	3.9367	0.0530	1.1250	1.0097	0.8547	0.9828
RMSE p25	0.0008	0.1362	0.1647	0.0223	0.3395	2.4508	0.0169	0.0333	0.0019	0.1525	0.0492
RMSE p50	0.0023	0.3541	0.2007	0.0482	0.8104	2.6415	0.0347	0.1250	0.0097	0.2841	0.1094
RMSE p75	0.0038	0.5644	0.2359	0.0861	1.4166	2.8139	0.0490	0.3069	0.0226	0.4070	0.2364

Panel D: Partially matched parameters (from *Baseline*)

	α	β	δ	λ	F	σ_d	σ_n	POS_{bmk}	POS_{fund}	FOS_{bmk}	FOS_{fund}
Est. mean	-0.0006	1.3216	0.0481	0.8722	1.7111	2.0409	0.0524	1.4805	1.0113	0.9204	0.9811
Est. median	-0.0002	1.3481	0.0460	0.8974	1.6058	2.0509	0.0460	1.1250	1.0058	0.8547	0.9686
RMSE p25	0.0009	0.0837	0.0131	0.0099	0.3305	0.6909	0.0055	0.0333	0.0015	0.1525	0.0584
RMSE p50	0.0024	0.1853	0.0160	0.0324	0.4728	0.8766	0.0118	0.1250	0.0058	0.2841	0.1108
RMSE p75	0.0040	0.3737	0.0216	0.0536	0.8356	1.0426	0.0209	0.3069	0.0141	0.4070	0.2196

Table A.2. Real Estate Private Equity (REPE) funds

This table reports summary statistics for the sample of 486 Real Estate funds in our sample, selected SSM parameter estimates, filtered returns, and nowcasting performance metrics. See Tables 1 and 3 for definitions and details.

Summary statistics:

	mean	sd	skew	p5	p10	p25	p50	p75	p90	p95
Vintage year	2003.2	4.6	-1.1	1989	1994	2000	2005	2007	2008	2008
Fund size (USD mil.)	607.0	860.5	6.1	16	73	206	396.5	684	1725.0	4001
Fund life (years)	10.8	1.4	-1.1	6.40	7.98	9.83	11.2	12.0	12.0	12.0
IRR(%)	4.9	13.7	-0.3	-38.5	-17.6	-2.42	6.06	13.2	25.6	33.6
Money Multiple	1.3	0.6	0.6	0.12	0.36	0.86	1.30	1.62	2.22	3.22
PME (v. CRSP VW)	0.9	0.5	0.9	0.080	0.22	0.56	0.86	1.21	1.75	2.27
# Capital Calls	23.6	18.8	1.7	3	6	11	16	30	63	92
# Distributions	31.9	23.0	1.4	1	5	16	26	45	77	117
# NAV reports	19.1	2.6	-0.6	13	14	18	20	21	23	24
% Resolved	86.2	19.7	-2.0	14.1	39.2	82.3	95.3	99	99.7	100
# Quarters w/ Dist.> 0	9.5	5.8	0.2	0	1	5	9	14	19	22
w/ Dist > 5% of Fund	5.6	4.4	1.0	0	0	2	5	8	14	18

	mean	p25	p50	p75	mean	p25	p50	p75
Key parameter estimates:								
	α				β			
Fund-specific	0.031	−0.031	0.048	0.097	0.796	0.576	0.727	1.148
Peer-imputed	−0.020	−0.097	−0.004	0.084	0.891	0.717	0.922	1.089
Average-imputed	−0.019	−0.091	0.001	0.077	0.912	1.005	1.006	1.006
Literature-imputed	−0.019	−0.092	0.001	0.077	0.893	0.875	1.000	1.000
	F_c				λ			
Fund-specific	0.861	0.788	0.917	1.010	0.817	0.821	0.928	0.976
Peer-imputed	1.257	0.875	0.991	1.144	0.889	0.818	0.923	0.972
Average-imputed	0.969	0.818	0.905	1.070	0.890	0.914	0.947	0.947
Literature-imputed	0.982	0.813	0.924	1.085	0.894	0.922	0.953	0.953
Filtered return properties:								
	Quarterly return autocorrelation				Quarterly return variance			
Naive nowcast	0.385	0.230	0.416	0.593	0.049	0.005	0.009	0.018
Fund-specific	0.001	−0.182	0.015	0.192	0.082	0.010	0.025	0.076
Peer-imputed	0.019	−0.182	−0.006	0.187	0.052	0.019	0.029	0.047
Average-imputed	−0.139	−0.336	−0.168	0.001	0.077	0.042	0.056	0.077
Literature-imputed	−0.130	−0.332	−0.168	−0.004	0.076	0.041	0.055	0.077
Nowcast performance assessment:								
	POS RMSE				FOS RMSE			
Naive nowcast	0.333	0.158	0.295	0.476	0.335	0.147	0.274	0.470
Fund-specific	0.161	0.037	0.075	0.174	0.227	0.052	0.134	0.262
Peer-imputed	0.176	0.048	0.092	0.197	0.161	0.050	0.101	0.194
Average-imputed	0.208	0.072	0.133	0.259	0.215	0.062	0.133	0.255
Literature-imputed	0.203	0.072	0.130	0.246	0.206	0.053	0.131	0.251