

# A fundwise stochastic discount factor estimator for private equity funds

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

March 17, 2020

## Keywords

Stochastic discount factor, Private equity fund, Fund level data set

## Acknowledgements

I thank Hsin-Chih Ma, Stefan Mittnik, Daniel Schalk, and all participants of the LMU econometrics research seminar SS 2019 for helpful discussions and support.

## Declaration of interest

The author reports no conflict of interest. The author alone is responsible for the content and writing of the paper.

# A fundwise stochastic discount factor estimator for private equity funds

## Abstract

This paper proposes a simple stochastic discount factor estimation methodology suited for private equity fund-level cash flow data.

## 1 Introduction

Do investments in private equity funds offer abnormal returns to fund investors when risk-adjusted to public market factors? Or more profoundly, do we need yet another Stochastic Discount Factor (SDF) approach to evaluate private equity fund cash flows?

To answer these questions, we propose a simple SDF model estimation methodology that incorporates private equity fund-level cash flow data. The basic idea is that the expected value of discounted fund net cash flows is zero when an appropriate SDF model is applied. Given this setting, our estimator can be regarded as the simplified version of the approaches of Driessen et al. (2012) and Korteweg and Nagel (2016).

## 2 Methodology

### 2.1 Fundwise least-mean-distance estimator

Let fund  $\chi = 1, 2, \dots, n$  be characterized by its (net) cash flows  $CF_{t,\chi}$  and its net asset values  $NAV_{t,\chi}$  with discrete time index  $t = 1, 2, \dots, T$ . The stochastic discount factor  $\Psi_{t,\tau}$  can be used to calculate the time- $\tau$  present value  $P_{t,\tau,\chi}$  of a time- $t$  cash flow of any given PE fund  $\chi$

$$P_{t,\tau,\chi} = \Psi_{t,\tau} \cdot CF_{t,\chi}$$

As SDFs are commonly parameterized by a vector  $\theta$ , i.e.,  $\Psi_{t,\tau} \equiv \Psi_{t,\tau}(\theta)$ , our goal is to find an estimation method for the optimal  $\theta$ . For each fund  $\chi$  and all points  $\tau$  within a common fund lifetime, the pricing error  $\epsilon_{\tau,\chi}$  of all fund cash flows is calculated as

$$\epsilon_{\tau,\chi} = \sum_{t=1}^T P_{t,\tau,\chi} \quad \forall \quad \tau, \chi \quad (1)$$

Our least-mean-distance estimator minimizes the sum of all  $\tau, \chi$  pricing error losses

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \sum_{\chi=1}^n \sum_{\tau \in W_\chi} L[w_{\tau,\chi} \epsilon_{\tau,\chi}] \equiv \arg \min_{\theta \in \Theta} \sum_{i=1}^N L[w_i \epsilon_i] \quad (2)$$

where  $L$  denotes a loss function, e.g.,  $L[x] = (x - 0)^2$ , and  $N$  gives the cardinality of  $\epsilon_{\tau,\chi}$ . The weighting factor  $w_i$  can be one divided by invested capital for equal weighting of funds, a positive constant for fund-size weighting, or e.g., some macroeconomic deflator.  $W_\chi$  gives the set of relevant present value times  $\tau$  for fund  $\chi$  and can be thought of all quarterly dates

within the usual fund lifetime of ten to fifteen years. Each fund  $\chi$  is characterized by its vintage year which can be expressed by  $v_\chi = \min(W_\chi) \in 1, 2, \dots, V$ , where  $V$  denotes the maximum vintage year used in a given data set.

## 2.2 Asymptotic assumptions

The estimator in equation 2 exhibits a cross-sectional nature, since it sums over all funds rather than constructing vintage year based time-series. This means, we intentionally opt against a framework compatible with classical, time-series Generalized Method of Moments (GMM) that requires the construction of stationary, ergodic time-series of moment conditions (Hansen, 1982, 2012). These time-series (of random functions) are used to empirically estimate the expected value of equation 1. The stationarity requirement would prevent us from (i) using funds from non-liquidated vintages for model estimation and (ii) weighting vintages by cumulative fund sizes or invested capital. Additionally, our approach automatically down-weights non-fully-invested funds and preserves the unbalanced data structure of empirical single fund data. As the number of funds increases by time, more recent vintages are overweighted by our procedure. As we drop the time-series GMM framework, in turn identification requires a sufficient number of funds from different vintage years in the fund-level data set used for model estimation (Driessen et al., 2012; Korteweg and Nagel, 2016).

Law of large numbers: The global moment condition underlying our estimation approach is that the  $(\tau, \chi)$ -unconditional) expected pricing error shall be zero, if we use the optimal SDF parameter  $\theta_0$ . This also means, instead of applying a time-series law of large numbers, we rely on a cross-section law of large numbers, but acknowledge the statistical dependence of pricing errors with respect to (i) vintage year  $v_\chi$  and (ii) fund lifetimes  $\tau$ . Jenish and Prucha (2012) develop an asymptotic inference framework for near-epoch dependent spatial processes that is suitable for our setting.

**Assumption 1.** *The dependence structure and time-trend of  $\{w\epsilon\}$  shall allow*

$$N^{-1} \sum_{i=1}^N w_i \epsilon_i \xrightarrow{a.s.} E[w\epsilon] \quad \text{as } V \rightarrow \infty$$

Consistency: The estimator  $\hat{\theta}$  converges in probability to the true parameter value  $\theta_0$  as the number of distinct vintage years in our data set goes to infinity. Multiple funds for a specific vintage year are not necessarily required and are thus considered as additional, stabilizing moment conditions.

**Assumption 2.** *Consistency of  $\hat{\theta}$  requires  $\hat{\theta} \xrightarrow{p} \theta_0$  as  $V \rightarrow \infty$ . Thus  $E[w\epsilon] = 0$  if and only if  $\theta = \theta_0$ . The parameter space is compact  $\theta \in \Theta$ .*

Compactness of  $\Theta$  can be assured by lower and upper bounds for certain parameters that are justified by economic reasoning, e.g., a market  $\beta$  of  $\pm 12$  seems implausible for PE funds because of the implied return expectations.

Central limit theorem: To assess the significance of our parameter estimates, we want to describe the asymptotic distribution of the parameter vector as a normal distribution; see (Pötscher and Prucha, 1997, Theorem 11.2) and (Jenish and Prucha, 2012, Theorem 4).

**Assumption 3.**  $\mathbf{D}^{-\frac{1}{2}}\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathbf{I})$  as  $V \rightarrow \infty$  for some Matrix  $\mathbf{D}$  and identity matrix  $\mathbf{I}$ .

Due to the limited amount of available private equity data (typically the oldest vintages start in the 1980s), we refrain from asymptotic closed-form solutions for  $\mathbf{D}$  which rely on the dependence assumptions about  $we$ . In empirical applications, the small sample behavior of an estimation method for private equity data is more important than its asymptotic theory.

## 2.3 Comparison to other approaches

In contrast to our approach, the closely related method of Driessen et al. (2012) forms vintage year portfolios instead of using individual fund moments. Further, they discount all fund cash flows just to the first cash flow date (like in a classical net present value calculation) and we additionally sum over all dates within  $W_\chi$  to alleviate the exploding alpha issue mentioned in Driessen et al. (2012).

Similar to our ansatz, Korteweg and Nagel (2016) draw on a spatial GMM framework to handle cross-sectional dependence (Conley, 1999). However, to obtain their Generalized Public Market Equivalent (GPME) metric they opt for pricing public market cash flows that shall replicate PE funds instead of pricing the observed fund cash flows itself.

## 3 Empirical estimation

### 3.1 Data

We use the Preqin cashflow data set as of 26th February 2020. We pool all regions and analyze the following fund types: INF ("Infrastructure", 174 funds in data set), NATRES ("Natural Resources", 167), PD ("Private Debt"; 488), PE ("Private Equity"; 2474), RE ("Real Estate"; 810), VC ("Venture Capital"; 985), MEZZ ("Mezzanine"; 147), DD ("Special Situations", "Distressed Debt"; 60+153), BO ("Growth", "Buyout"; 265+1251), FOF ("Fund of Funds"; 589), SEC ("Secondaries"; 131). For these fund types we extract all unweighted cash flow series, which corresponds to fund-size weighting. For non-liquidated funds we treat the latest net asset value as final cash flow.

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

### 3.2 Regularization and blockwise cross-validation

To avoid to determine the parameter significance by asymptotic results, we apply model selection by means of a regularized loss function and estimate the regularization hyperparameter by (blockwise) cross-validation.

Specifically, we use the following lasso-regularized loss function in equation 2

$$L(w\epsilon; \lambda) = (w\epsilon - 0)^2 + \lambda \sum_{j \in J} |\theta_j| \quad (3)$$

where the MKT coefficient is not subject to any regularization,  $J = \{\text{ME}, \text{IA}, \text{ROE}, \text{EG}\}$ . By not penalizing the MKT coefficient, we preferably explain private equity returns by the market return factor.

The lasso parameter  $\lambda$  is determined by  $h\nu$ -block cross validation to account for the dependency in our data set introduced by overlapping fund cash flows for funds from adjacent vintage years (Racine, 2000). Therefore, we form three partitions for several vintage year groups. As larger validation sets are preferred for model selection, the validation set ( $\nu$ -block) always contains funds of three neighboring vintage years (e.g. 2000, 2001, 2002). To reduce the dependency between training and validation set, we remove all funds from three-year-adjacent vintage years, i.e., the  $h$ -block (e.g. 1997, 1998, 1999, 2003, 2004, 2005). Funds from the remaining vintage years enter the training set and are thus used for model estimation (e.g. 1985-1996, 2006-2019). We apply ten-fold cross validation using the ten validation sets described in table 3.2.

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

Table 1: Partitions used for  $h\nu$ -block cross validation.

### 3.3 Results

We apply the simple linear SDF

$$\Psi_{t,\tau}(\theta) = \prod_{h=\tau}^t \left( 1 + r_f + \sum_j \theta_{j,h} \cdot F_{j,h} \right)^{-1} \quad (4)$$

with risk-free return  $r_f$  and zero-net-investment portfolio returns  $F_j$ .

Our general approach is to estimate an ensemble of SDF models by varying the set of relevant fund lifetimes  $W_\chi$  and the maximum vintage year of funds that enter estimation. We use the variation of estimated parameters within the ensemble to gauge their significance. Specifically, we regard relevant fund lifetimes between 40 and 60 quarters, and maximum vintage years between year 2010 and 2015. To avoid overfitting our SDF ensemble just considers two-factor models that contain  $\{\text{MKT}\}$  and  $\{\text{ME or IA or ROE or EG}\}$ .

## 4 Public market factor style analysis

Assume we want to assess the performance of a given (liquidated) fund, and we estimated a total of  $N$  competing SDF models for that given fund type and region  $\{\Psi^{(i)}, i = 1, 2, \dots, N\}$ . After we compute the set of (possible) net present values  $\{\text{NPV}^{(i)}, i = 1, 2, \dots, N\}$ , we can select the subset of  $n$  SDFs that exhibit the smallest squared net present value errors  $\{j : (\text{NPV}^{(j)})^2 < \alpha_n\}$ . In the next step, we linearly combine the collection of  $n$  underlying factor models to obtain one factor model that describes the investment style of that given fund in terms of public market factors (cf. model averaging vs. combination, bucket of models, bake-off contest).

## 5 Conclusion

## References

- Conley, T. G. (1999). Gmm estimation with cross-sectional dependence. *Journal of Econometrics*, 92(1):1–45.
- Driessen, J., Lin, T.-C., and Phalippou, L. (2012). A new method to estimate risk and return of nontraded assets from cash flows: the case of private equity. *Journal of Financial and Quantitative Analysis*, 47(3):511–535.
- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054.
- Hansen, L. P. (2012). Proofs for large sample properties of generalized method of moments estimators. *Journal of Econometrics*, 170:325–330.
- Hou, K., Xue, C., and Zhang, L. (2015). Digesting anomalies: An investment approach. *Review of Financial Studies*, 28(3):650–705.
- Hou, K., Xue, C., and Zhang, L. (2020). An augmented  $q^5$  model with expected growth. *Review of Finance*.
- Jenish, N. and Prucha, I. R. (2012). On spatial processes and asymptotic inference under near-epoch dependence. *Journal of Econometrics*, 170(1):178–190.
- Korteweg, A. and Nagel, S. (2016). Risk-adjusting the returns to venture capital. *Journal of Finance*, 71(3):1437–1470.
- Pötscher, B. M. and Prucha, I. R. (1997). *Dynamic nonlinear econometric models: Asymptotic theory*. Springer Science & Business Media.
- Racine, J. (2000). Consistent cross-validatory model-selection for dependent data: hv-block cross-validation. *Journal of Econometrics*, 99:39–61.

Type	Factor	MKT.mean	MKT.t	Q.mean	Q.t
BO	EG	2.03	24.13	2.88	11.45
DD	EG	1.42	37.55	3.92	68.13
FOF	EG	2.11	52.67	3.62	25.28
INF	EG	1.55	25.06	4.42	78.74
MEZZ	EG	1.64	5.78	2.35	7.93
NATRES	EG	-0.40	-18.66	3.26	13.43
PD	EG	1.42	31.92	3.89	66.41
PE	EG	1.97	24.29	3.05	14.23
RE	EG	0.93	20.19	3.85	52.35
SEC	EG	1.80	35.44	4.54	31.48
VC	EG	2.51	26.53	-0.83	-1.70
BO	IA	2.16	31.00	2.71	10.87
DD	IA	1.47	22.84	0.93	5.75
FOF	IA	2.30	149.30	3.39	75.52
INF	IA	1.36	32.58	2.38	40.06
MEZZ	IA	1.67	6.64	1.05	7.64
NATRES	IA	-0.65	-26.58	2.10	55.22
PD	IA	1.47	22.00	0.96	6.62
PE	IA	2.13	30.68	2.65	13.50
RE	IA	0.98	23.60	1.55	28.97
SEC	IA	1.61	74.75	4.25	26.46
VC	IA	2.34	12.20	-1.18	-1.99
BO	ME	1.28	14.14	1.42	30.17
DD	ME	1.03	17.87	0.78	21.14
FOF	ME	1.73	48.83	1.17	11.67
INF	ME	1.33	32.05	0.52	28.39
MEZZ	ME	0.92	3.69	1.24	14.31
NATRES	ME	-0.95	-46.48	0.91	19.26
PD	ME	0.98	17.09	0.86	24.44
PE	ME	1.23	14.70	1.45	33.66
RE	ME	0.11	2.15	1.38	11.82
SEC	ME	0.80	25.22	1.90	14.08
VC	ME	2.71	91.65	-0.94	-5.85
BO	ROE	2.39	20.09	2.45	8.83
DD	ROE	1.95	21.77	2.33	19.64
FOF	ROE	2.28	65.92	2.43	9.97
INF	ROE	1.46	26.34	5.98	50.50
MEZZ	ROE	1.72	5.67	2.04	14.61
NATRES	ROE	-0.16	-4.34	2.58	16.02
PD	ROE	1.93	20.14	2.31	17.42
PE	ROE	2.35	22.04	2.54	10.13
RE	ROE	1.22	19.34	2.69	8.64
SEC	ROE	2.59	22.44	4.32	18.00
VC	ROE	2.40	21.45	-0.87	-3.11