ON SEMIPARAMETRIC SDF ESTIMATORS

For private equity fund data

Christian Tausch

October 12, 2020

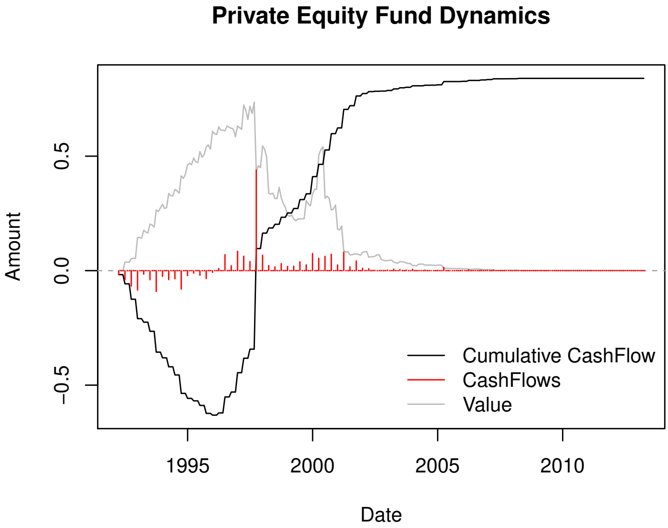
AssetMetrix

Agenda

1. Introduction
2. Semiparametric estimation summary (1. paper, [Tausch, 2020])
3. Model combination methodology (2. paper, [Tausch Pietz, 2020])
4. Empirical results (2. paper, [Tausch Pietz, 2020])
5. Conclusion

1 INTRODUCTION

Stochastic Discount Factor Methods for Non-Traded Cash Flows - The Case of Private Equity


Private Equity Fund Dynamics
Amount
0.5
0.0
−0.5
1995
2000
2005
2010
Date
Cumulative CashFlow
CashFlows
Value

Private equity fund $i = 1, 2, \ldots, n$ is characterized by:

Net Asset Value $NAV_{i,t}$ (fund value proxy)

Net Cash Flow $CF_{i,t}$ (fund distributions minus contributions)

Vintage Year V_i (fund inception year)

Public market is given by:

SDF $\Psi_{\tau,t} > 0$ (stochastic discount factor from t to τ)

Risk-free Rate r_t (from period $t - 1$ to t)

Factor Return $F_{j,t} \geq 0$ (zero-net-investment return from $t - 1$ to t)

Time is discrete $t = 1, 2, \ldots, T$.

Stochastic discount factors (SDFs)

- General pricing framework in empirical finance.
- SDFs allow to move cash flows in time.

We can calculate the time-τ price of a time-t cash flow by

$$P_{\tau,t,i} = \mathbb{E}\left[\Psi_{\tau,t} \cdot CF_{t,i}\right] \qquad \forall \quad \tau, t, i \tag{1}$$

where the SDF $\Psi_{\tau,t} = \Psi_{\tau,t}(\theta)$ depends on the parameter vector θ.

If τ and t are both in the past, the realized price is given by

$$P_{\tau,t,i} = \Psi_{\tau,t} \cdot CF_{t,i} \qquad \forall \quad \tau, t, i \tag{2}$$

with $\tau \leq t$ or $\tau \geq t$.

2 SEMIPARAMETRIC ESTIMATION SUMMARY

- Empirical asset pricing usually uses **semiparametric** approaches to determine the 'optimal' parameter vector θ of the SDF $\Psi_{\tau,\mathrm{t}}(\theta)$.
- Semiparametric means we impose **no distributional assumptions** on the random variable Ψ.
- The parameter vector θ contains no distributional parameters (like μ, σ for a normal distribution).
- We want to **parsimoniously** explain asset returns (cash flows).
- When testing SDFs, we want to test if a given SDF satisfactorily prices the assets and not if the asset returns are (e.g.) normally distributed.
- Parametric estimation (like maximum likelihood) is usually more **efficient** (unbiased with smaller variance) when we know the underlying distribution.

	[Driessen et al., 2012]	[Korteweg and Nagel, 2016]	[Tausch, 2020]
M-estimator	Least-Mean-Distance	Generalized Method of Moments	Least-Mean-Distance
Pricing error averaging	No	No	Yes
Cash flows priced	PE cash flows	public cash flows	PE cash flows
Asymptotics	cross-sectional #funds $\rightarrow \infty$	time-series #vintages $\rightarrow \infty$	spatial # of both $\rightarrow \infty$
Inference	bootstrap	spatial HAC	cross-validation & spatial HAC
Cross-sectional unit	vintage year portfolio	single fund	testing both
SDF	simple linear	exponentially affine	testing both

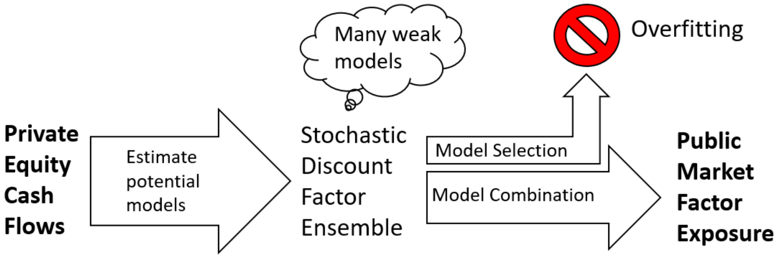
Table: Comparison of similar SDF estimation frameworks.

Results of [Tausch, 2020]:

- Simulation results indicate high small-sample variance even for simple data generating processes.
- Empirical estimation for simple two-factor models reveals (using public q^5-investment factors of [Hou et al., 2020]):
  - ⚠️ very high asymptotic standard error estimates (for vintage year portfolios),
  - hv-block cross-validation standard errors are smaller,
  - model selection remains challenging when confronted with a large set of competing models.
- 💡 Model combination instead of model selection?

3 MODEL COMBINATION

3.1 MODEL COMBINATION IDEA



Transform weak model ensemble to strong multi-factor model.

3.2 WHY SO MANY/WEAK/DIFFICULT?

- Why so many models?
  - Many public market factor candidates.
  - Many potential estimators, loss functions, hyperparameters.
  - Many different proprietary private data sets.
- Why so weak models?
  - Sparse private equity fund data ($\leq$ 40 vintages).
  - Near-epoch dependency by overlapping fund cash flows.
  - Multi-factor models almost surely overfit.
- Why model selection is difficult?
  - Model uncertainty especially high for weak models.
  - Limited data may encourage data snooping.
  - Correct post model selection inference is generally hard.

The weighted pricing error obtained by SDF model averaging is defined as

$$\epsilon_{\tau,i}^{(M^*)} = \sum_{m=1}^{M^*} w_m \sum_{t=1}^{T} \Psi_{\tau,t}^{(m)} CF_{t,i} \tag{3}$$

with model weight $w_m \geq 0$ and all weights sum to one $\sum_m^{M^*} w_m = 1$. The ensemble size is M^*.

- Forecast combination puzzle: Often $w_m = \frac{1}{M*}$ outperforms more 'advanced' weighting schemes.
- Model combination can be perceived as diversification strategy to minimize the risk of selecting an invalid model (i.e., investing everything in the wrong replication strategy).

4 EMPIRICAL RESULTS

We use a **simple linear** SDF model as in [Driessen et al., 2012]

$$\Psi^{\mathrm{SL}}_{\tau,t}(\theta) = \prod_{h=1}^{t} \left(1 + \alpha + r_h + \sum_j \beta_j F_{j,h}\right)^{-1} \prod_{h=1}^{\tau} \left(1 + \alpha + r_h + \sum_j \beta_j F_{j,h}\right) \tag{4}$$

with (arithmetic) risk-free return r, (arithmetic) zero-net-investment portfolio returns F_j, and parameter vector $\theta = (\alpha, \beta)$.

Estimate four two-factor models (using linear SDF from equation 4):
MKT-RF $\times$ {SMB or HML or HDY-MKT or QLT-MKT} with

- MKT-RF: MSCI Market Return Minus Risk-free Rate
- SMB: MSCI Small Cap Minus MSCI Large Cap Return
- HML: MSCI Value Minus MSCI Growth Return
- HDY-MKT: MSCI High Dividend Yield Minus MSCI Market Return
- QLT-MKT: MSCI Quality Minus MSCI Market Return

For each of these four models, we generate $2 \times 2 \times 5$ estimates by varying (i) a quadratic and last absolute deviance loss function, (ii) equal- and fund-size-weighted cash flows, and (iii) maximum months in $\{120, 150, 180, 210, 240\}$ for $\mathcal{T}$.

Finally, simply **average** the $4 \times 2 \times 2 \times 5$ model **coefficients**.

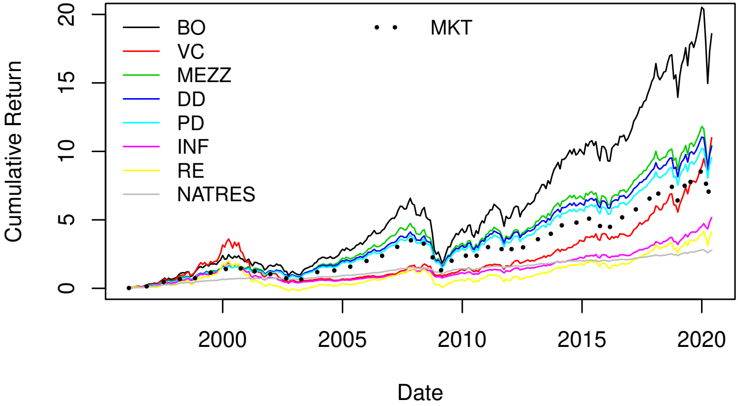
The public factor exposure of private equity

Type	MKT-RF	HML	SMB	HDY-MKT	QLT-MKT
BO	1.33 (0.15)	-0.15 (0.12)	0.2 (0.03)	0.3 (0.1)	0.21 (0.05)
DD	0.96 (0.09)	-0.11 (0.04)	0.21 (0.01)	0.14 (0.1)	0.16 (0.05)
INF	0.71 (0.22)	-0.37 (0.06)	-0.33 (0.13)	-0.47 (0.35)	0.36 (0.11)
MEZZ	1.08 (0.13)	0.06 (0.1)	0.14 (0.04)	0.16 (0.1)	0.06 (0.11)
NATRES	0.36 (0.27)	-0.04 (0.22)	-0.02 (0.22)	0.16 (0.36)	0.11 (0.17)
PD	0.96 (0.08)	-0.07 (0.04)	0.16 (0.03)	0.06 (0.09)	0.15 (0.04)
RE	1.14 (0.44)	-0.3 (0.16)	-0.42 (0.13)	-0.91 (0.15)	-0.4 (0.1)
VC	1.02 (0.67)	-0.61 (0.11)	-0.42 (0.03)	-0.75 (0.14)	0.84 (0.61)
MKT	1	0	0	0	0

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

Private equity: Preqin cash flow data. Public: MSCI style indices.

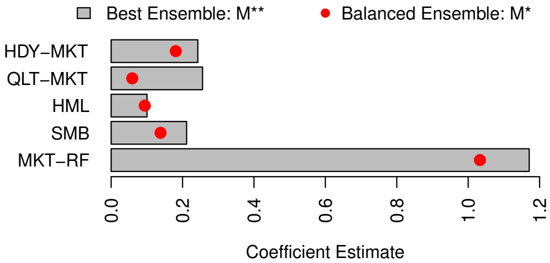
Did private equity outperform the public market portfolio?



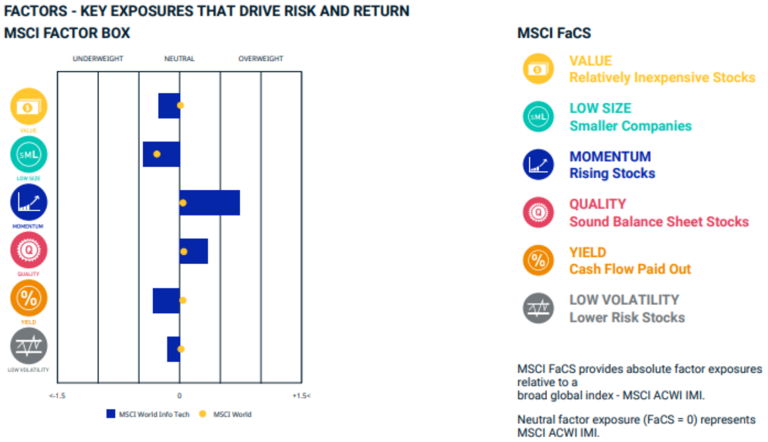
Type	Annualized Return				Sharpe Ratio
	mean.R	stdv.R	mean.R-RF	stdv.R-RF	mean/stdv.R-RF
BO	**0.152**	0.195	**0.125**	0.196	0.641
DD	0.116	0.144	0.091	0.144	0.630
INF	0.085	0.119	0.060	0.119	0.506
MEZZ	0.120	0.162	0.094	0.162	0.581
NATRES	0.057	**0.049**	0.033	**0.049**	**0.671**
PD	0.113	0.143	0.087	0.143	0.610
RE	0.092	0.203	0.067	0.203	0.329
VC	0.124	0.176	0.099	0.176	0.561
MKT	0.107	0.152	0.082	0.152	0.536

Table: Annualized average returns, standard deviations (annualized by the square root of time formula), and Sharpe ratios (i.e., the ratio of mean.R-RF to stdv.R-RF) implied by the five-factor models (1996-01-31 to 2020-05-31).

Bottom-up (fund-by-fund) aggregation of averaged coefficients
for a sample portfolio of 100 private capital funds.



- Balanced Ensemble: all valid SDF models for a given fund.
- Best Ensemble: subset of all valid SDF models with smallest pricing error for a given fund.



Source (2020-10-12): `https://www.msci.com/documents/10199/69aaf9fd-d91d-4505-a877-4b1ad70ee855`

5 CONCLUSION

5.1 SUMMARY, OUTLOOK, THOUGHTS, IDEAS

- Significant semiparametric SDF estimates for private equity funds are hard to obtain. Asymptotic inference not very useful when forming vintage year portfolios.
- Model combination is a straightforward means to form a strong(er) SDF model from a collection of weak competitors.
- Conjecture: Averaging pricing errors over cash flow duration (fund lifetime) may be general feature of an 'optimal' SDF estimator for non-traded cash flows.
- Future research: Effect of taking historical (fixed) public market returns vs simulated scenarios in simulation study: What are the issues? What is optimal?
- Future research: Analyze improved version of the [Korteweg and Nagel, 2016] estimator (simulation-based portfolios avoid under-identification, but compatible with averaging pricing errors?).

- Publish two papers (1. technical and 2. practitioner)? Or merge both to one combined paper?
- 2. paper is joint paper with my boss from AssetMetrix, plan to also use it for marketing a business application.
    - Option a): Submit now to practitioner journal.
    - Option b): Publish now as white paper at company homepage.
- Replace 'spatial' by more catchy name for 'averaged over multiple NPV dates' (in first paper title).
- Personal aim of this project -> understand GMM estimators.

References

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.

Hou, K., Xue, C., and Zhang, L. (2020).
An augmented q^5 model with expected growth.
Review of Finance.

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

Working paper and R code

will be available on my blog

Quant-Unit.com

.

Do you have comments?