Deep Learning in Asset Pricing

French Association of Asset and Liability Manager (AFGAP)

Luyang Chen, Markus Pelger and Jason Zhu

Stanford University

Motivation: Building Investment Strategies with Machine Learning

Why machine learning for investment?

- Extracts signals from a large information set;
- Easily captures complex nonlinear relationships among variables;
- Works well out-of-sample and is robust to overfitting.

Best overall prediction $\neq$ the best input for portfolio creation

Typical machine learning portfolios:

1. Prediction problem: Find signal to predict future returns
   - linear regressions,
   - machine learning (e.g. deep learning)
2. Portfolio design based on the prediction output
   - long-short strategies (highest minus lowest decile, risk factors)
   - mean-variance optimization

So far statistics and machine learning often focus unconditionally on step 1.

This paper:

- Extract signal that is optimal for portfolio construction

Asset Pricing for Academics and Practitioners

Two perspectives on the same fundamental problem with the same solution.

- Stochastic Discount Factor (SDF) = optimal portfolio with highest Sharpe ratio
- Test assets = investment strategies
- Asset pricing model = explains mean returns by exposure to risk factor(s)
- Pricing errors = unexplained performance of investment strategies (alphas)

$\Rightarrow$ This paper constructs the optimal portfolio and asset pricing model

Relevance for academic finance research: Asset pricing (AP)

- Understand source and size of risk premium
- Understand which information is relevant for the explaining average returns

Relevance for Practitioners: Investment

- Optimal portfolios with attractive risk-return trade-off
- Predict returns of assets
- Identify mispricing = alpha opportunities in markets
- Risk management

Motivation: Asset Pricing

Fundamental Problem of Asset Pricing

- Crucial question in finance:
  Why are asset prices different for different assets?
- No-Arbitrage Pricing Theory:
  Stochastic discount factor (SDF) explains differences in asset prices.
- Fundamental Question: What is the SDF?

Challenges:

- Big Data: SDF should depend on all available economic information
- Non-parametric: Functional form of SDF is unknown and likely complex
- Dynamics: SDF needs to capture time-variation in economic conditions
- Weak signal: Risk premium in stock returns has a low signal-to-noise ratio

Can Machine Learning Help?

- Machine-learning methods very flexible and deal with big data, but ...
- Asset returns in efficient markets dominated by unforecastable news

⇒ This paper: Disciplining learning algorithm with no-arbitrage constraint strongly improves signal

Conceptional Challenges in Asset Pricing

What is the functional form of the SDF based on the information set?

- Conventional example: Fama-French 5 factor model
- Problem: Linear form misspecified, 100 more potential characteristics
- Our solution: General non-parametric model with variable selection

What are the test assets?

- Conventional example: 25 Fama-French double-sorted portfolios
- Problem: Selected SDF might only work on these test assets
- Our solution: All stocks and all possible characteristic based portfolios

What are the states of the economy?

- Conventional example: NBER recession indicators
- Problem: 100 of macroeconomic time-series with complex dynamics
- Our solution: Extract a small number of state processes using complete dynamics of a large number of macroeconomic time-series

Contribution of this paper

- This Paper: Estimate the SDF with deep neural networks
- Crucial innovation: Include no-arbitrage condition in the neural network algorithm and combine three neural networks in a novel way
- Key elements of estimator:
  1. Non-linearity: Feed-forward network captures non-linearities
  2. Time-variation: Recurrent (LSTM) network finds a small set of economic state processes
  3. Pricing all assets: Generative adversarial network identifies the states and portfolios with most unexplained pricing information
  4. Signal-to-noise ratio: No-arbitrage conditions improve the risk premium signal

$\Rightarrow$ General model that includes all existing models as a special case

Contribution of this paper

1. Empirically outperforms all benchmark models out-of-sample.
    - Optimal portfolio has out-of-sample annual Sharpe ratio of 2.6.
    - Our model explains 8% of variation in individual stocks
    - Our model explain over 90% of average returns for characteristic managed portfolios
2. Insight into the structure of the SDF
    - Non-linear interactions between firm information matter.
    - Characteristics in isolation approximately linear.
    - Macroeconomic states matter.
    - SDF structure stable over time
      (25 years of test data without refitting)
    - All "classical" firm characteristics relevant with price trends and trading frictions as the most important
3. Economic constraints matter
    - Off-the-shelf machine learning methods perform worse.
    - Machine learning combined with economic model structure works significantly better

Model

Fundamental no-arbitrage condition:

$$\mathbb{E}_t[M_{t+1} R^e_{i,t+1}] = 0$$

for all $t = 1, ..., T$ and $i = 1, ..., N$

- $R^e_{i,t+1} = R_{i,t+1} - R_f$ = excess return at time $t+1$ for asset $i = 1, ..., N$
- $\mathbb{E}_t[.]$ expected value conditioned on information set at time t
- M_{t+1} stochastic discount factor SDF at time $t+1$.

Conditional moments imply infinitely many unconditional moments

$$\mathbb{E}[M_{t+1} R^e_{t+1,i} I_t] = 0$$

for any $\mathcal{F}_t$-measurable variable I_t

Equivalent Factor Model Representation

Without loss of generality SDF is projection on the return space

$$M_{t+1} = 1 - \sum_{i=1}^{N} w_{i,t} R^e_{i,t+1}$$

- SDF portfolio $F_{t+1} = \sum_{i=1}^{N} w_{i,t} R^e_{i,t+1}$ has highest conditional Sharpe-ratio.
- Portfolio weights $w_{i,t}$ are a general function of macro-economic information I_t and firm-specific characteristics $I_{i,t}$:

$$w_{i,t} = w(I_t, I_{i,t}).$$

$\Rightarrow$ Need non-linear estimator with many explanatory variables!

No-arbitrage condition is equivalent to factor representation:

$$R^e_{t+1} = \beta_t F_{t+1} + e_{t+1}.$$

Objects of interest:

- The SDF portfolio F_t and its portfolio weights $w_{i,t}$.
- The risk loadings $\beta_{i,t} = \frac{\text{cov}_t(R^e_{i,t+1}, F_{t+1})}{\text{var}_t(F_{t+1})}$.
- The unexplained residual $\hat{e}_t = \left(I_N - \beta_{t-1}(\beta_{t-1}^\top \beta_{t-1})^{-1}\beta_{t-1}^\top\right) R^e_t$.

Estimation

Estimation

General Method of Moments Objective (g is given):

$$\min_{M} \sum_{i} \| \sum_{t} M(I_t, I_{i,t}) R^e_{i,t+1} g(I_t, I_{i,t}) \|^2$$

- Estimate SDF weights $w(\cdot)$ to minimize no-arbitrage moment conditions for conditioning variables $g(I_t, I_{i,t})$.
- We use a feed forward network to estimate $w_{i,t}$ for given $g(I_t, I_{i,t})$
- Finance intuition: $R^e_{t+1} g(I_t, I_{i,t})$ form characteristic managed portfolios
- Example: g might build size/value portfolios as test assets
- Problem of finding optimal "instruments" = choice of test assets
- Problem: Model implies infinite # of moment conditions.
  Imposing all is infeasible, hard to know which ones to select.
- Solution: Generative Adversarial Network (GAN) chooses informative g

Generative Adversarial Network (GAN)

$$\min_{M} \max_{g} \sum_{i} \| \sum_{t} M(I_t, I_{i,t}) R^e_{i,t+1} g(I_t, I_{i,t}) \|^2$$

For a candidate SDF M the adversary g constructs the test assets (and states) where M has difficulty pricing:

- Two networks play a zero-sum game and are alternative updated:
  1. SDF Network constructs SDF the M with smallest pricing errors for g.
  2. Conditional Network generates conditioning variables g with largest pricing errors for M.
- Example: If M is Fama-French 5 SDF, g constructs momentum portfolios.

$\Rightarrow$ Find economic states and test assets with the most pricing information.

Generative Adversarial Network (GAN)

$$\min_{M} \max_{g} \sum_{i} \| \sum_{t} M(I_t, I_{i,t}) R^e_{i,t+1} g(I_t, I_{i,t}) \|^2$$

For a candidate SDF M the adversary g constructs the test assets (and states) where M has difficulty pricing:

- Two networks play a zero-sum game and are alternative updated:
  1. SDF Network constructs SDF the M with smallest pricing errors for g.
  2. Conditional Network generates conditioning variables g with largest pricing errors for M.
- Example: If M is Fama-French 5 SDF, g constructs momentum portfolios.

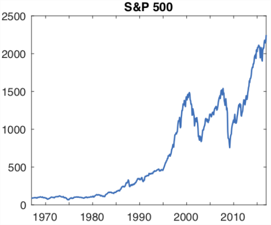
$\Rightarrow$ Find economic states and test assets with the most pricing information.

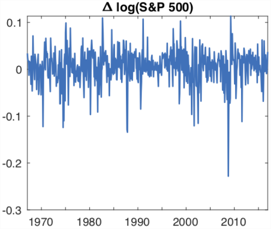
Econometrics perspective:

- Conventional GMM: optimal instruments based on efficiency.
  - Not feasible for large number of potential parameters.
  - Assumes test assets identify SDF parameters.
- Our GAN: optimal instruments based on robustness.
  - Feasible for large set of instruments and parameters.
  - Finds test assets that identify SDF parameters.

Macroeconomic Dynamics: Finding Hidden Macroeconomic States

Macroeconomic time-series with standard transformation: S&P500 price/return





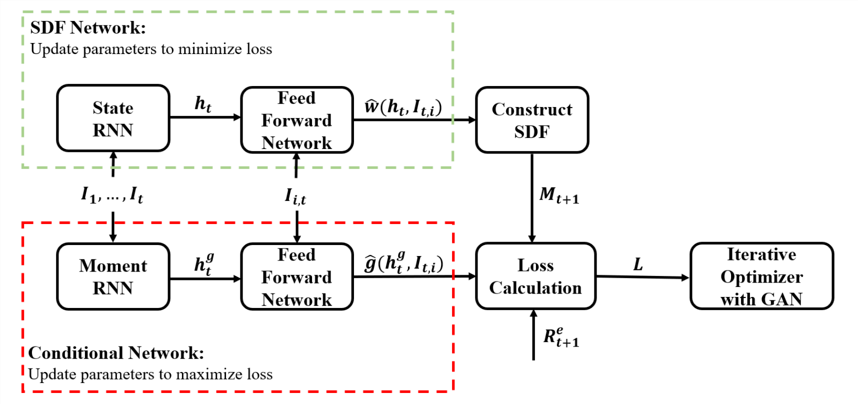
Problems with economic time-series data:

- Time-series data is often non-stationary ⇒ transformation necessary
- States depend on dynamics! ⇒ last differenced observation uninformative
- Macro time-series strongly dependent ⇒ low dimensional structure

Solution: Long-Short-Term Memory (LSTM) Cell Recurrent Neural Network:

- Transform all macroeconomic time-series into a low dimensional vector of stationary state variables
- Intuition: Non-linear hidden state space model combined with non-linear factor model

Model Architecture



Empirical Results

Data

- 50 years of monthly observations: 01/1967 - 12/2016.
- Monthly stock returns for all U.S. securities from CRSP
  (around 31,000 stocks)
  Use only stocks with with all firm characteristics
  (around 10,000 stocks)
- 46 firm-specific characteristics for each stock and every month
  (usual suspects) $\Rightarrow I_{i,t}$
  normalized to cross-sectional quantiles
- 178 macroeconomic variables
  (124 from FRED, 46 cross-sectional median time-series for characteristics, 8 from Goyal-Welch) $\Rightarrow I_t$
- T-bill rates from Kenneth-French website
- Training 20 years, validation 5 years, test 25 years

Benchmark Models

1. Linear model: SDF portfolio weights $w_t = I_t\theta$ linear in characteristics
   Intuition: Mean-variance optimization on characteristic managed long-short factors $\tilde{R}_{t+1} = I_t^\top R_{t+1}^e$.
   LS: Linear regression $\hat{\theta} = \left(\tilde{R}^\top \tilde{R}\right)^{-1} \tilde{R}^\top 1$
   EN: Elastic net regularization (Kozak, Nagel and Santosh (2019)):
   $$\min_\theta \left\| \frac{1}{T}\tilde{R}^\top 1 - \frac{1}{T}\tilde{R}^\top \tilde{R}\theta \right\|_2^2 + \lambda_1\|\theta\|_1 + \lambda_2\|\theta\|_2^2.$$

2. FFN: Deep learning return forecasting (Gu, Kelly and Xiu (2019)):
   - Predict conditional expected returns $\mathbb{E}_t[R_{i,t+1}] = \beta_{t,i}\mathbb{E}_t[F_{t+1}]$.
   - Conditional mean proportional to SDF loading $\beta_{t,i}$
   - Empirical loss function for prediction
   $$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(R_{i,t+1} - \mu(I_t, I_{i,t})\right)^2$$
   - Use only simple feedforward network for forecasting

Evaluation

Objects of Interest:

- The SDF portfolio F_t
- The risk loadings β_t
- The unexplained residual $\hat{e}_t = (I_N - \beta_{t-1}(\beta_{t-1}^\top \beta_{t-1})^{-1} \beta_{t-1}^\top) R_t^e$

Asset Pricing Performance Measure

- Sharpe ratio of SDF portfolio: $SR = \frac{\hat{\mathbb{E}}[F_t]}{\sqrt{\widehat{Var}(F_t)}}$
- Explained variation: $EV = 1 - \frac{\left(\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N_t}\sum_{i=1}^{N_t}(\hat{e}_{i,t+1})^2\right)}{\left(\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N_t}\sum_{i=1}^{N_t}(R_{i,t+1}^e)^2\right)}$
- Cross-sectional mean R^2: XS-R^2 $= 1 - \frac{\frac{1}{N}\sum_{i=1}^{N}\frac{T_i}{T}\left(\frac{1}{T_i}\sum_{t\in T_i}\hat{e}_{i,t+1}\right)^2}{\frac{1}{N}\sum_{i=1}^{N}\frac{T_i}{T}\left(\frac{1}{T_i}\sum_{t\in T_i}\hat{R}_{i,t+1}\right)^2}$

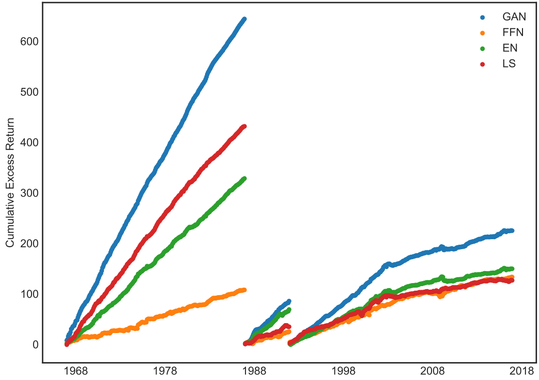
Cross Section of Individual Stock Returns

	SR			EV			Cross-Sectional R^2		
Model	Train	Valid	Test	Train	Valid	Test	Train	Valid	Test
LS	1.80	0.58	0.42	0.09	0.03	0.03	0.15	0.00	0.14
EN	1.37	1.15	0.50	0.12	0.05	0.04	0.17	0.02	0.19
FFN	0.45	0.42	0.44	0.11	0.04	0.04	0.14	-0.00	0.15
GAN	2.68	1.43	0.75	0.20	0.09	0.08	0.12	0.01	0.23

- Our model GAN, forecasting FFN, linear EN and LS
- Annual out-of-sample Sharpe ratio SR for GAN 2.6
- GAN explains twice as much (8%) of the variation in individual stocks
- GAN has explains higher fraction of average returns
- Linear model (EN) outperforms forecasting (FFN) $\Rightarrow$ no-arbitrage matters!

Optimal Portfolio Performance

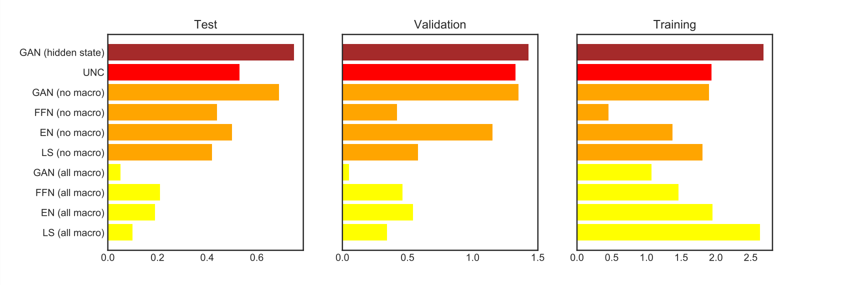
Cumulative SDF factor returns



⇒ GAN portfolio outperforms benchmark models

Performance of Models with Different Macroeconomic Variables

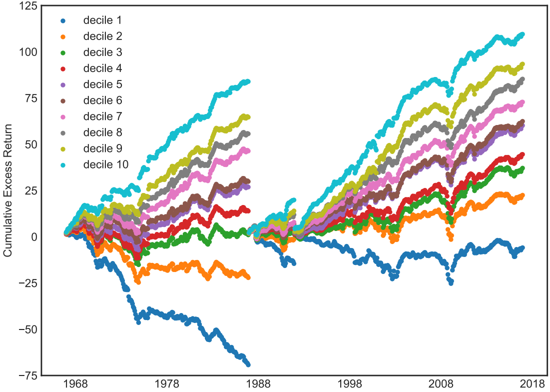
Sharpe Ratio of SDF for different inclusions of macroeconomic information.



- GAN (hidden states) is our reference model
- no macro uses only firm characteristics
- all macro uses standard transformation of macroeconomic time-series without LSTM

⇒ Macroeconomic hidden states matter!

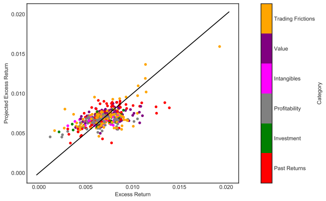
Predictive Performance



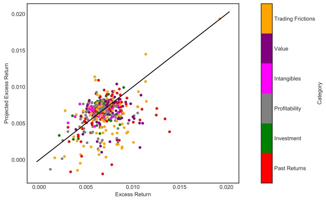
Cumulative excess returns of β sorted decile portfolios for GAN

⇒ Risk loadings predict future stock returns.

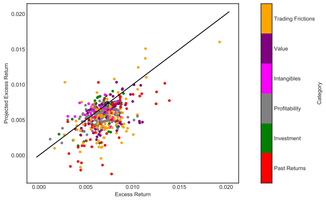
Asset Pricing on Sorted Portfolios



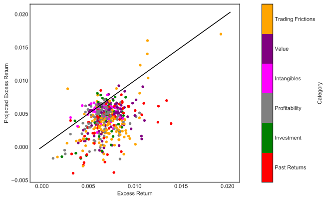
(a) GAN



(b) FFN



(c) EN



(d) LS

Predicted and average returns for value weighted characteristic sorted portfolios.

- Out-of-sample results for 46 characteristic sorted decile portfolios
- GAN always has cross-sectional $R^2 > 90\%$ for each 46 decile portfolios

$\Rightarrow$ GAN explains better the cross-section of average returns

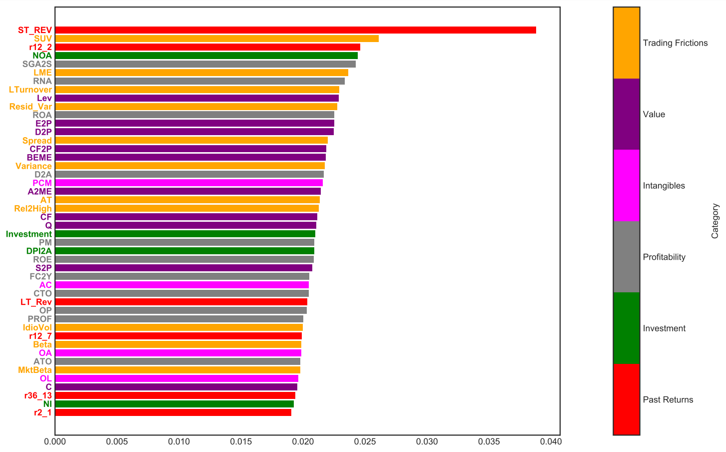
SDF Factor and Fama-French Factors

	Mkt-RF	SMB	HML	RMW	CMA	intercept
Regression Coefficients	0.00	0.00	-0.04	0.08***	0.04	0.76***
	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)	(0.06)
Correlations	-0.10	-0.09	0.01	0.17	0.05	-

Out-of-sample correlation and regression of GAN SDF on Fama-French 5 factors.

$\Rightarrow$ Fama-French factors do not span GAN SDF.

Characteristic Importance

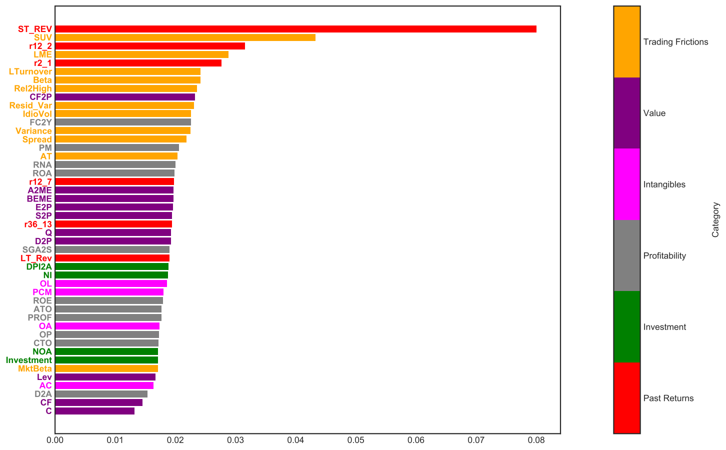


GAN characteristic importance ranking in terms of average absolute gradient

⇒ Price trends and trading frictions most relevant

⇒ All categories represented among top 20 variables

Characteristic Importance

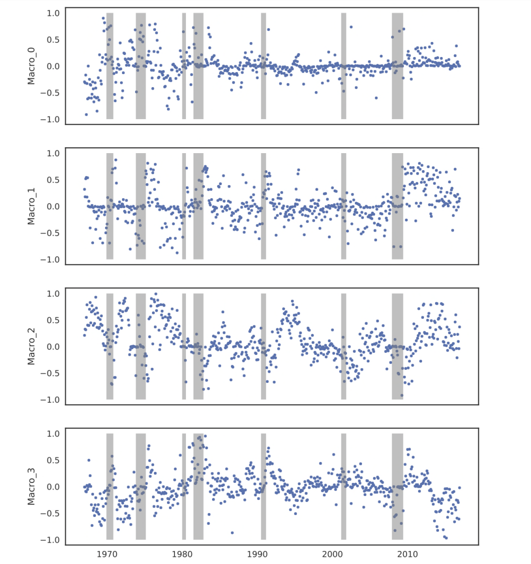


FFN characteristic importance ranking in terms of average absolute gradient

⇒ Simple forecasting approach mainly selects price trends, volatility and illiquidity (consistent with Gu, Kelly and Xiu (2019))

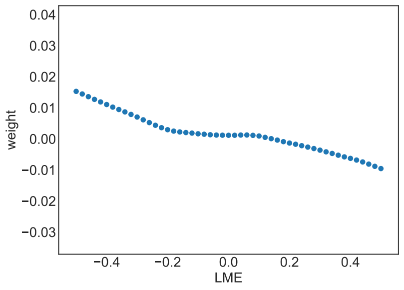
⇒ Does FFN mainly fit illiquid small cap stocks?

Macroeconomic Hidden States

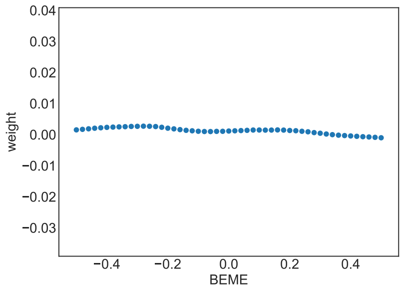


GAN Macroeconomic hidden states process. Gray areas mark NBER recession periods.

SDF Weights



(a) Size



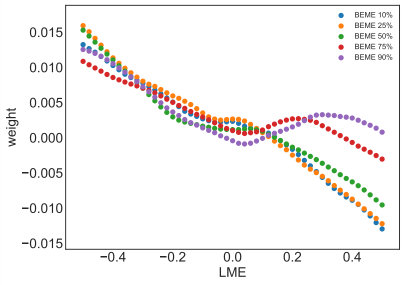
(b) Book to Market Ratio

SDF weight as a function of size and book to market ratio

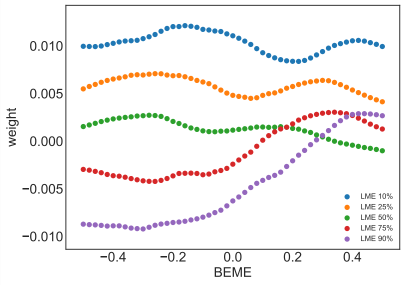
SDF weight ω as 1-dimensional function keeping other covariates at their mean

$\Rightarrow$ Size and book to market have close to linear effect!

SDF Weights



(a) Size



(b) Book to Market Ratio

Conditional weight as a function of size and book to market ratio

SDF weight ω as 2-dimensional function keeping other covariates at their mean

$\Rightarrow$ Complex interaction between multiple variables!

Robustness of Model Fit

1. Market capitalization
   - Evaluate and/or estimate models without small cap stocks
   - GAN robust qualitatively to removing small cap stocks
   - FFN and EN sensitive to removing small cap stocks
     $\Rightarrow$ potential overfitting of small, illiquid stocks for FFN and EN
2. Tuning parameters
   - Compare GAN models with best validation tuning parameters
   - All benchmark criteria essentially identical on test data ($\Delta < 3\%$)
   - SDF time-series of GAN models highly correlated (around 90%)
   - Variable importance and SDF weights very similar
3. Time stability
   - Fit GAN on rolling window $\Rightarrow$ time-varying SDF weight $\omega_t(I_t, I_{i,t})$
   - SDF of constant and time-varying GAN strongly correlated (78%)
   - Variable importance and SDF weights very similar
   - Slightly better test performance for benchmark criteria ($\Delta \approx 10\%$)

$\Rightarrow$ Robust model fit that captures economic structure

Conclusion

Conclusion

Methodology

- Novel combination of deep-neural networks to estimate the pricing kernel
- Key innovation: Use no-arbitrage condition as criterion function
- Time-variation explained by macroeconomic states and firm characteristics
- Test assets with most pricing information selected by adversarial approach
- General asset pricing model that includes all other models as special cases

Empirical Results

- GAN outperforms benchmark models.
- Non-linearities matter for the interaction.
- Characteristics in isolation approximately linear.
- Macroeconomic states matter.
- SDF predicts future returns and explains cross-sectional average returns
- SDF structure stable over time.
- SDF efficient portfolio highly profitable.
- GAN framework is complementary to conditional multi-factor models

Appendix

Firm specific characteristics

Past Returns	Investment	Profitability	Intangibles	Value	Trading Frictions
Momentum	Investment	Operating profitability	Accrual	Book to Market Ratio	Size
Short-term Reversal	Net operating assets	Profitability	Operating accruals	Assets to market cap	Turnover
Long-term Reversal	Change in prop. to assets	Sales over assets	Operating leverage	Cash to assets	Idiosyncratic Volatility
Return 2-1	Net Share Issues	Capital turnover	Price to cost margin	Cash flow to book value	CAPM Beta
Return 12-2		Fixed costs to sales		Cashflow to price	Residual Variance
Return 36-13		Profit margin		Dividend to price	Total assets
		Return on net assets		Earnings to price	Market Beta
		Return on assets		Tobin's Q	Close to High
		Return on equity		Sales to price	Spread
		Expenses to sales		Leverage	Unexplained Volume
		Capital intensity			Variance

Literature (Partial List)

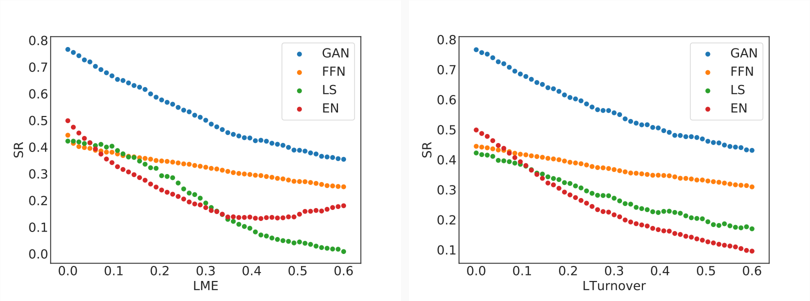
- Deep-learning for predicting asset prices
  - Gu, Kelly and Xiu (2020)
  - Feng, Polson and Xu (2020)
  - Bianchi, Büchner and Tamoni (2019)

  ⇒ Predicting future asset returns with feed forward network
- Neural networks for no-arbitrage pricing
  - Bansal and Viswanathan (1993): Non-linear SDF
- Deep-learning autoencoder
  - Gu, Kelly and Xiu (2020)
  - Heaton, Polson and Witte (2017)

  ⇒ Low dimensional non-linear factor structure
- Linear methods for asset pricing of large data sets
  - Kelly, Pruitt and Su (2019): Instrumented PCA
  - Lettau and Pelger (2020): Risk-premium PCA
  - Kozak, Nagel and Santosh (2019): Mean-variance with regularization
- Tree-based learning for general non-linear interactions
  - Bryzgalova, Pelger and Zhu (2020): Asset-Pricing Trees

Machine Learning Investment: Trading Friction Trade-Offs

Out-of-sample Sharpe ratios with trading frictions.

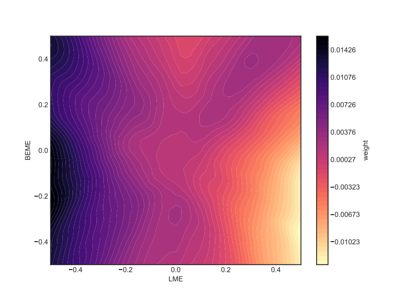


- Portfolio weights ω set to zero if either the market capitalization (`LME`) or turnover (`Lturnover`) is below a cross-sectional quantile.
- Trade-off between trading-frictions and achievable Sharpe ratios (lower bound)
- Standard protocol for most machine learning portfolios:
    1. Extract signal from predicting returns
    2. Form portfolios based on signal (long-short or mean-variance efficient)

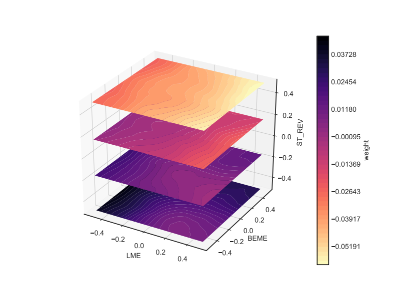
$\Rightarrow$ This paper: Extract signal for optimal portfolio design.

$\Rightarrow$ Next step in Bryzgalova, Pelger and Zhu (2020):
Extract signal for optimal portfolio design under constraints.

SDF Weights



(a) Size and Book to Market Ratio



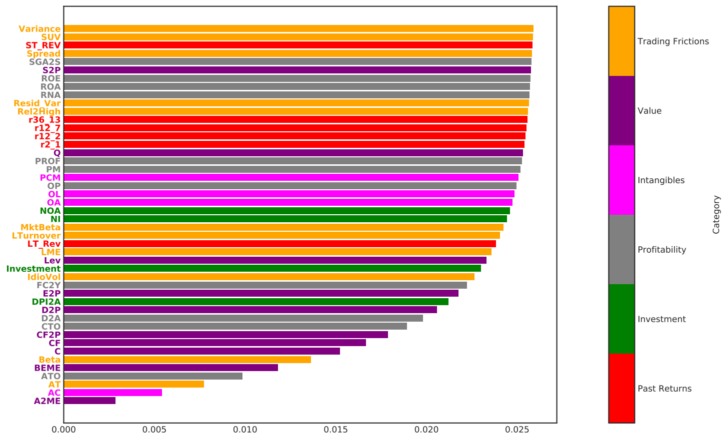
(b) Size, Book To Market and Short-Term Reversal

Conditional weight as a function of size and book to market ratio

SDF weight ω as 3-dimensional function keeping other covariates at their mean

$\Rightarrow$ Complex interaction between multiple variables!

Optimal Test Assets

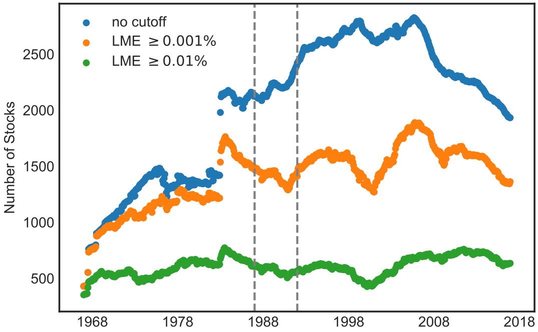


GAN adversarial characteristic importance ranking with average absolute gradient

⇒ Robust instruments (test assets) include all major categories

⇒ Size and book-to-market not sufficient

Robustness to Market Capitalization



Number of stocks per month for

1. the total sample
2. stocks with market cap larger than 0.01% of total market cap.
3. stocks with market cap larger than 0.001% of total market cap.

Robustness to Market Capitalization

	SR			EV			Cross-Sectional R^2		
Model	Train	Valid	Test	Train	Valid	Test	Train	Valid	Test
Size $\geq$ 0.001% of total market cap									
LS	1.44	0.31	0.13	0.07	0.05	0.03	0.14	0.03	0.10
EN	0.93	0.56	0.15	0.11	0.09	0.06	0.17	0.05	0.14
FFN	0.42	0.20	0.30	0.11	0.10	0.05	0.19	0.08	0.18
GAN	2.32	1.09	0.41	0.23	0.22	0.14	0.20	0.13	0.26
Size $\geq$ 0.01% of total market cap									
LS	0.32	-0.11	-0.06	0.05	0.07	0.04	0.13	0.05	0.09
EN	0.37	0.26	0.23	0.09	0.12	0.07	0.17	0.08	0.14
FFN	0.32	0.17	0.24	0.13	0.22	0.09	0.22	0.15	0.26
GAN	0.97	0.54	0.26	0.28	0.34	0.18	0.27	0.23	0.32

Different SDF Models Evaluated on Large Market Cap Stocks

Robustness to Market Capitalization

Model	SR			EV			Cross-Sectional R^2		
	Train	Valid	Test	Train	Valid	Test	Train	Valid	Test
LS	1.91	0.40	0.19	0.08	0.06	0.04	0.18	0.05	0.12
EN	1.34	0.92	0.42	0.13	0.13	0.07	0.23	0.09	0.19
FFN	0.37	0.19	0.28	0.13	0.13	0.07	0.21	0.10	0.21
GAN	3.57	1.18	0.42	0.24	0.23	0.14	0.23	0.13	0.26

Different SDF models estimated and evaluated on large market cap stocks (size larger than 0.001% of the total market capitalization).

Asset Pricing on Sorted Portfolios

ST_REV	EN	FFN	GAN	EN	FFN	GAN
Decile	Explained Variation			Alpha		
1	0.84	0.74	0.77	-0.18	-0.21	-0.13
2	0.86	0.81	0.82	0.00	-0.05	0.00
3	0.80	0.82	0.84	0.13	0.04	0.06
4	0.69	0.80	0.82	0.16	0.03	0.03
5	0.58	0.68	0.71	0.13	-0.03	-0.04
6	0.43	0.66	0.75	0.22	0.05	0.01
7	0.23	0.64	0.77	0.20	0.03	-0.02
8	-0.07	0.49	0.67	0.23	0.03	-0.05
9	-0.25	0.29	0.58	0.30	0.09	-0.01
10	-0.24	-0.04	0.35	0.47	0.38	0.18
	Explained Variation			Cross-Sectional R^2		
All	0.43	0.58	0.70	0.45	0.79	0.94

Explained variation and pricing errors for short-term reversal sorted portfolios

- Out-of-sample results for value weighted decile portfolios.
- GAN explains extreme quantiles better

Asset Pricing on Sorted Portfolios

	Explained Variation			Cross-Sectional R^2		
Charact.	EN	FFN	GAN	EN	FFN	GAN
ST_REV	0.43	0.58	0.70	0.45	0.79	0.94
SUV	0.42	0.75	0.83	0.64	0.97	0.99
r12_2	0.26	0.27	0.54	0.66	0.71	0.93
NOA	0.58	0.69	0.78	0.94	0.96	0.95
SGA2S	0.52	0.63	0.73	0.93	0.95	0.96
LME	0.83	0.78	0.86	0.96	0.95	0.97
RNA	0.50	0.48	0.69	0.93	0.87	0.96
...	...	...	...	...	...	...
CF2P	0.46	0.47	0.66	0.90	0.89	0.99
BEME	0.70	0.75	0.82	0.97	0.94	0.98
Variance	0.48	0.27	0.61	0.74	0.72	0.90
...	...	...	...	...	...	...

Explained variation and pricing errors for decile sorted portfolios

- Out-of-sample results for all value weighted decile portfolios.
- GAN always explains more variation than other approaches.
- GAN always has cross-sectional $R^2 > 90\%$.

Sensitivity of Forecasting (FFN) to Size

Quantile	SR (Train)	SR (Valid)	SR (Test)
(i) Equally-Weighted			
1%	1.24	0.65	0.66
5%	1.36	1.10	0.71
10%	1.30	1.31	0.67
25%	1.19	1.20	0.57
50%	1.09	1.26	0.52
(ii) Value-Weighted			
1%	0.98	0.35	0.39
5%	0.89	0.71	0.42
10%	0.70	0.45	0.32
25%	0.55	0.28	0.17
50%	0.43	0.20	0.15

Sharpe Ratio of Long-Short Portfolios with FFN

Risk Measures for SDF Factor

	SR			Max Loss			Max Drawdown		
Model	Train	Valid	Test	Train	Valid	Test	Train	Valid	Test
FF-3	0.27	-0.09	0.19	-2.45	-2.85	-4.31	7	10	10
FF-5	0.48	0.40	0.22	-2.62	-2.33	-4.90	4	3	7
LS	1.80	0.58	0.42	-1.96	-1.87	-4.99	1	3	4
EN	1.37	1.15	0.50	-2.22	-1.81	-6.18	1	3	5
FFN	0.45	0.42	0.44	-3.30	-4.61	-3.37	6	3	5
GAN	2.68	1.43	0.75	0.38	-0.28	-5.76	0	1	5

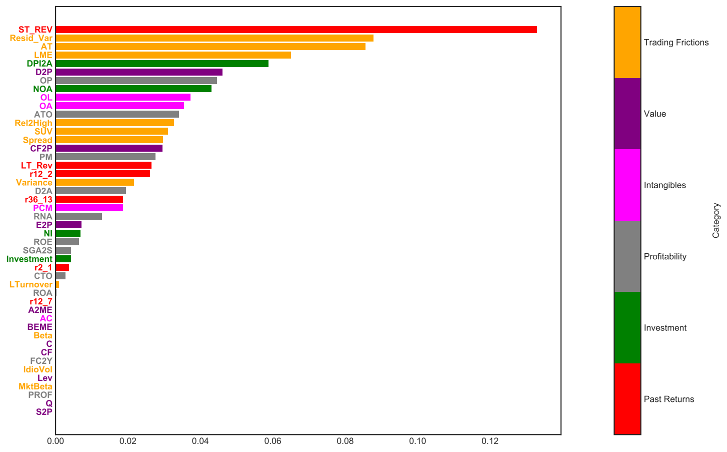
$\Rightarrow$ GAN lower or similar risk measured by max loss or drawdown but higher Sharpe ratio

Turnover

	Long Position			Short Position		
Model	Train	Valid	Test	Train	Valid	Test
LS	0.25	0.22	0.24	0.64	0.55	0.61
EN	0.36	0.35	0.35	0.83	0.83	0.84
FFN	0.69	0.63	0.65	1.38	1.29	1.27
GAN	0.47	0.40	0.40	1.05	1.04	1.02

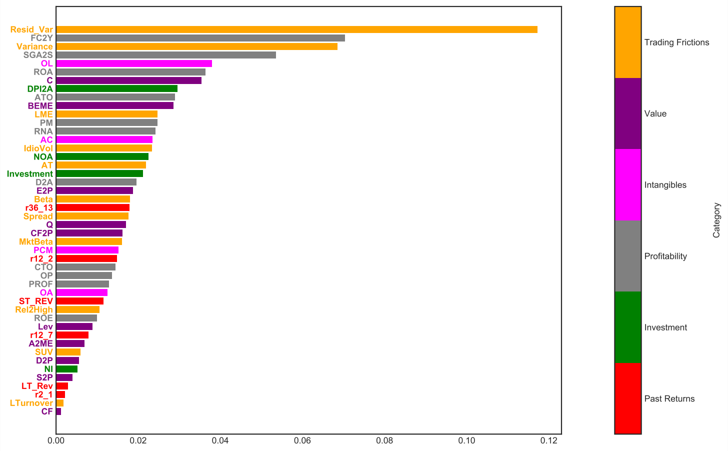
Turnover for positions with positive and negative weighs for the SDF factor portfolio.

Characteristic Importance



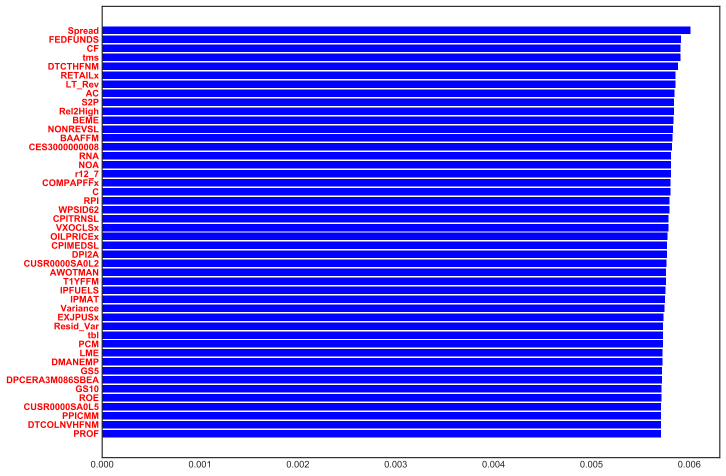
EN characteristic importance ranking in terms of average absolute gradient

Characteristic Importance



LS characteristic importance ranking in terms of average absolute gradient

Characteristic Importance



GAN variable importance ranking of the 178 macroeconomic variables

Performance of Alternative GAN Models

Model	SR			EV			Cross-Sectional R^2		
	Train	Valid	Test	Train	Valid	Test	Train	Valid	Test
GAN 1	2.78	1.47	0.72	0.18	0.08	0.07	0.12	0.01	0.21
GAN 2	3.02	1.39	0.77	0.18	0.08	0.07	0.12	0.00	0.22
GAN 3	2.55	1.38	0.74	0.22	0.11	0.09	0.17	0.04	0.25
GAN 4	2.44	1.38	0.77	0.19	0.08	0.07	0.11	0.01	0.22
GAN Rolling	N/A	N/A	0.88	N/A	N/A	0.08	N/A	N/A	0.24
GAN No Frict	2.94	1.37	0.77	0.20	0.10	0.08	0.14	0.01	0.23

Performance for alternative GAN models.

- GAN 1, 2, 3 and 4 are the four best GAN models on the validation data from an independent re-estimation.
- GAN Rolling is re-estimated every year on a rolling window of 240 months.
- GAN No Frict is estimated without trading frictions and past returns for the conditioning function g.

⇒ GAN is robust to tuning parameters, time-variation and limits to arbitrage.

Correlation with Alternative GAN Models

	GAN	GAN 1	GAN 2	GAN 3	GAN 4	GAN Rolling	GAN No Frict
GAN	1	0.84	0.87	0.84	0.80	0.70	0.78
GAN 1	0.84	1	0.88	0.92	0.89	0.79	0.89
GAN 2	0.87	0.88	1	0.87	0.88	0.73	0.83
GAN 3	0.84	0.92	0.87	1	0.89	0.74	0.86
GAN 4	0.80	0.89	0.88	0.89	1	0.78	0.84
GAN Rolling	0.70	0.79	0.73	0.74	0.78	1	0.78
GAN No Frict	0.78	0.89	0.83	0.86	0.84	0.78	1

Correlation of Benchmark GAN SDF with SDF of Alternative GAN Estimations.

- GAN 1, 2, 3 and 4 are the four best GAN models on the validation data from an independent re-estimation.
- GAN Rolling is re-estimated every year on a rolling window of 240 months.
- GAN No Frict is estimated without trading frictions and past returns for the conditioning function g.

⇒ GAN is robust to tuning parameters, time-variation and limits to arbitrage.

IPCA Asset Pricing with Different SDFs

IPCA assumes a K-factor model where the loadings are a linear function of the characteristics:

$$R_i = a_{t,i} + b_{t,i}^\top f_{t+1}^{\text{IPCA}} + \epsilon_i \qquad b_{t,i} = I_{i,t}^\top \Gamma_b.$$

Any multi-factor model assumes that the SDF is spanned by the factors:

$$F = \sum_{k=1}^{K} \omega^f(I_{k,t}, I_t) f_{t+1,k}^{\text{IPCA}}.$$

Fundamental problem: Find factor weights $\omega^f(I_{k,t}, I_t) \in \mathbb{R}^K$ for SDF.

- Combination of GAN and IPCA estimates conditional $\omega^{\text{I-GAN}}$
- Unconditional mean-variance efficient weights $\omega^{\text{I-SR}} = \text{Cov}\left(f_{t+1}^{\text{IPCA}}, f_{t+1}^{\text{IPCA}\top}\right)^{-1} \mathbb{E}\left[f_{t+1}^{\text{IPCA}}\right]$
- Alternative constant weights maximize XS-R^2 or EV: $\omega^{\text{I-XS}}$ and $\omega^{\text{I-EV}}$

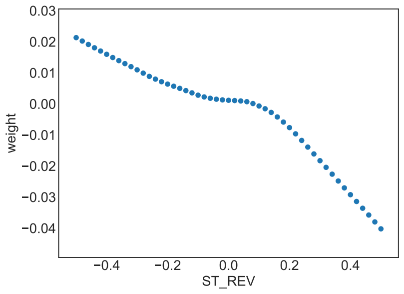
$\Rightarrow$ GAN framework is complementary to multi-factor models and can optimally make use of the additional information incorporated in factors.

IPCA Asset Pricing with Different SDFs

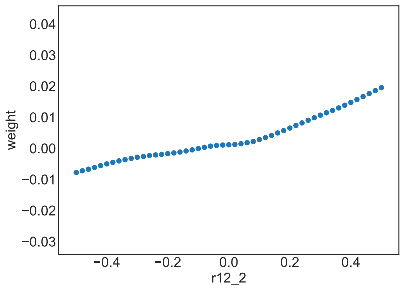
Model	Benchmark	3	4	5	6	7	8	9	10
	SR	0.61	0.71	0.77	0.70	0.79	0.82	0.72	0.81
IPCA GAN	EV	0.05	0.04	0.04	0.05	0.05	0.05	0.04	0.05
$(\omega^{\text{I-GAN}}, \beta^{\text{I-GAN}})$	XS-R^2	0.20	0.19	0.17	0.20	0.18	0.20	0.17	0.21
	SR	0.69	0.79	0.82	0.84	0.83	0.86	0.86	0.94
IPCA Max SR FFN Beta	EV	0.04	0.03	0.03	0.04	0.04	0.04	0.06	0.03
$(\omega^{\text{I-SR}}, \beta^{\text{I-FFN}})$	XS-R^2	0.14	0.13	0.11	0.14	0.14	0.15	0.19	0.14
	SR	0.69	0.79	0.82	0.84	0.83	0.86	0.86	0.94
IPCA Max SR	EV	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
$(\omega^{\text{I-SR}}, \beta^{\text{I-SR}})$	XS-R^2	-0.05	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04
	SR	0.11	0.11	0.15	0.17	0.15	0.15	0.14	0.16
IPCA Max EV	EV	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
$(\omega^{\text{I-EV}}, \beta^{\text{I-EV}})$	XS-R^2	-0.02	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03
	SR	-0.06	0.15	0.12	0.41	0.33	0.37	0.34	0.41
IPCA Max XS-R^2	EV	-0.02	-0.01	-0.02	-0.02	-0.02	-0.01	-0.02	-0.02
$(\omega^{\text{I-XS}}, \beta^{\text{I-XS}})$	XS-R^2	-0.03	0.07	0.06	0.12	0.12	0.13	0.13	0.14
	SR	0.69	0.79	0.82	0.84	0.83	0.86	0.86	0.94
IPCA Multifactor	EV	0.05	0.05	0.06	0.06	0.06	0.06	0.06	0.07
$(b_{t,i} \in \mathbb{R}^K)$	XS-R^2	-0.04	-0.03	-0.02	-0.01	-0.02	-0.01	-0.02	-0.02

Out-of-sample asset pricing results for different SDFs based on IPCA

SDF Weights



(a) Short-Term Reversal

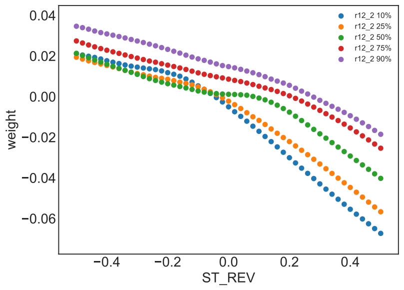


(b) Momentum

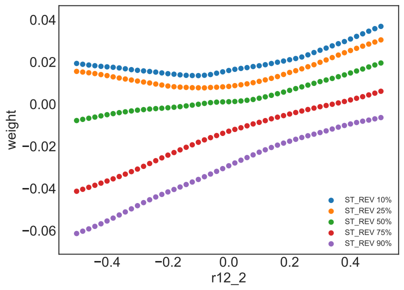
SDF weight as a function of short-term reversal and momentum

SDF weight ω as 1-dimensional function keeping other covariates at their mean
$\Rightarrow$ Short-term reversal and momentum have close to linear effect!

SDF Weights



(a) Short-Term Reversal

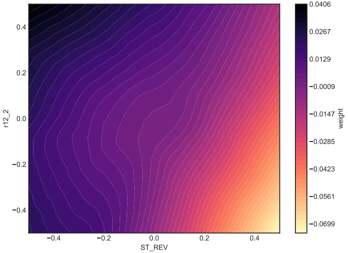


(b) Momentum

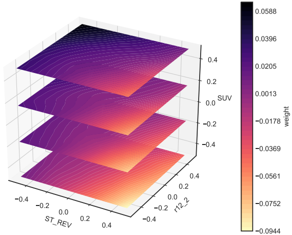
Conditional weight as a function of short-term reversal and momentum

SDF weight ω as 2-dimensional function keeping other covariates at their mean
$\Rightarrow$ Complex interaction between multiple variables!

SDF Weights



(a) Short-Term Reversal and Momentum

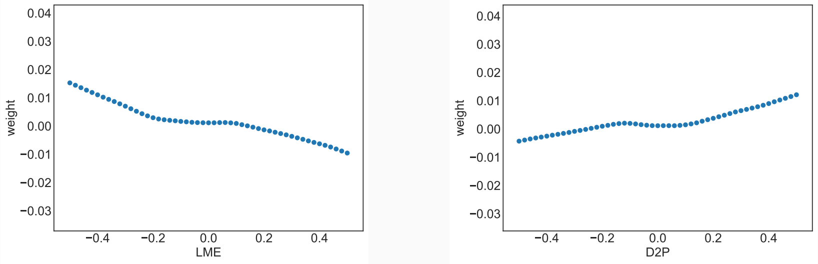


(b) Short-Term Reversal, Momentum and Size

Conditional weight as a function of short-term reversal and momentum

SDF weight ω as 3-dimensional function keeping other covariates at their mean
$\Rightarrow$ Complex interaction between multiple variables!

SDF Weights



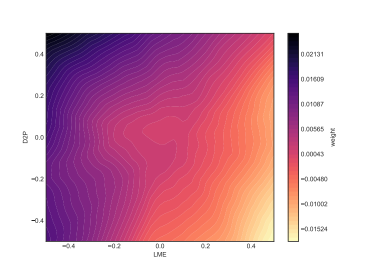
(a) Size

Dividend Yield

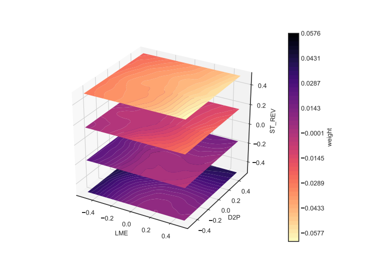
Figure 8: Weight as a function of size and dividend yield

$\Rightarrow$ Size and dividend yield have close to linear effect!

SDF Weights



(a) Size and Dividend Yield

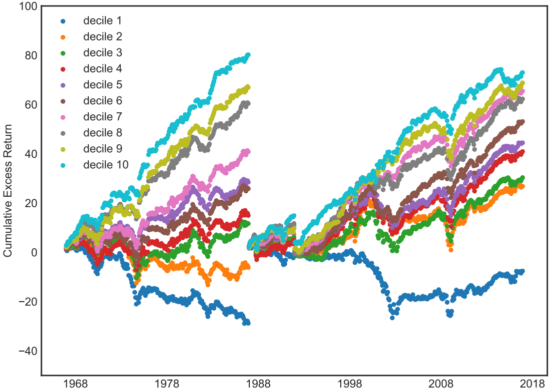


(b) Size, Dividend Yield and Short-Term Reversal

Weight as a function of multiple variables

⇒ Complex interaction between multiple variables!

Predictive Performance



Cumulative excess returns of β sorted value weighted portfolios for GAN

$\Rightarrow$ Risk loadings predicts future stock returns.

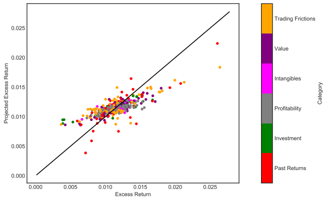
Predictive Performance

	Average Returns		Fama-French 3				Fama-French 5			
	Whole	Test	Whole		Test		Whole		Test	
Decile			α	t	α	t	α	t	α	t
1	-0.12	-0.02	-0.21	-12.77	-0.13	-5.01	-0.20	-11.99	-0.12	-4.35
2	-0.00	0.05	-0.09	-8.79	-0.05	-3.22	-0.09	-8.29	-0.05	-2.68
3	0.04	0.08	-0.04	-5.18	-0.02	-1.40	-0.04	-4.87	-0.01	-1.05
4	0.07	0.09	-0.02	-2.30	-0.00	-0.35	-0.02	-2.86	-0.01	-0.54
5	0.10	0.12	0.01	2.08	0.03	2.46	0.01	1.36	0.03	2.17
6	0.11	0.12	0.02	2.75	0.03	2.85	0.01	1.51	0.02	2.20
7	0.14	0.15	0.05	6.61	0.05	4.39	0.04	5.16	0.04	3.41
8	0.18	0.18	0.08	9.32	0.08	5.83	0.07	8.05	0.07	4.86
9	0.22	0.21	0.11	9.16	0.11	5.71	0.11	8.58	0.11	5.39
10	0.37	0.37	0.24	10.03	0.25	6.27	0.25	10.43	0.27	6.59
10-1	0.48	0.39	0.45	18.50	0.38	10.14	0.46	18.13	0.39	9.96
GRS Asset Pricing Test			GRS	p	GRS	p	GRS	p	GRS	
			39.72	0.00	11.25	0.00	37.64	0.00	10.75	0.00

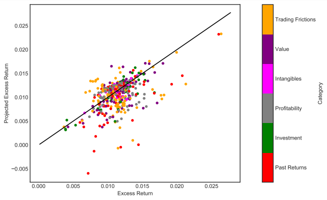
Time Series Pricing Errors for β-Sorted Portfolios

$\Rightarrow$ Standard factor models cannot explain cross-sectional returns of β-sorted portfolios.

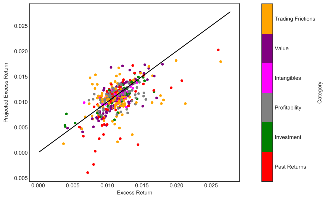
Asset Pricing on Sorted Portfolios



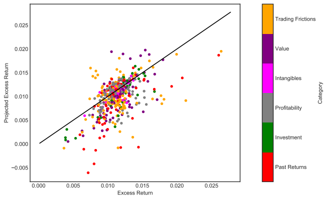
(a) GAN



(b) FFN



(c) EN



(d) LS

Predicted and average excess returns for characteristic sorted decile portfolios.

⇒ GAN explains better the cross-section of average returns (equally weighted)

Asset Pricing on Sorted Portfolios

LME	EN	FFN	GAN	EN	FFN	GAN
Decile	Explained Variation			Alpha		
1	0.80	0.75	0.79	0.09	-0.00	0.10
2	0.89	0.89	0.90	-0.11	-0.09	-0.06
3	0.91	0.80	0.91	-0.07	0.02	-0.02
4	0.90	0.77	0.91	-0.05	0.04	-0.01
5	0.90	0.78	0.91	0.01	0.10	0.04
6	0.88	0.80	0.91	0.03	0.09	0.02
7	0.84	0.81	0.89	0.04	0.05	-0.01
8	0.84	0.85	0.88	0.06	0.03	-0.02
9	0.77	0.81	0.82	0.06	-0.01	-0.04
10	0.32	0.28	0.49	-0.04	-0.15	-0.10
	Explained Variation			Cross-Sectional R^2		
All	0.83	0.78	0.86	0.96	0.95	0.97

Explained Variation and Pricing Errors for Size Sorted Portfolios

Asset Pricing on Sorted Portfolios

r12_2	EN	FFN	GAN	EN	FFN	GAN
Decile	Explained Variation			Alpha		
1	0.04	-0.06	0.33	0.37	0.39	0.11
2	0.12	0.10	0.52	0.25	0.18	-0.01
3	0.19	0.25	0.66	0.14	0.05	-0.06
4	0.28	0.34	0.73	0.15	0.08	-0.02
5	0.37	0.46	0.80	0.19	0.09	0.02
6	0.45	0.58	0.78	0.02	-0.03	-0.09
7	0.62	0.69	0.68	0.01	0.01	-0.05
8	0.58	0.71	0.64	-0.03	-0.04	-0.09
9	0.55	0.70	0.58	0.08	0.04	-0.03
10	0.51	0.53	0.53	0.24	0.29	0.19
	Explained Variation			Cross-Sectional R^2		
All	0.26	0.27	0.54	0.66	0.71	0.93

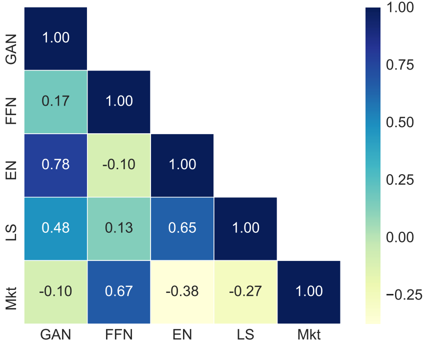
Explained Variation and Pricing Errors for Momentum Sorted Portfolios

Asset Pricing on Sorted Portfolios

BEME	EN	FFN	GAN	EN	FFN	GAN
Decile	Explained Variation			Alpha		
1	0.38	0.66	0.70	0.03	-0.12	-0.08
2	0.48	0.73	0.78	0.10	-0.05	-0.04
3	0.71	0.84	0.86	0.07	-0.03	-0.01
4	0.76	0.88	0.89	0.00	-0.07	-0.07
5	0.82	0.87	0.88	0.05	0.02	0.01
6	0.77	0.82	0.88	0.06	0.04	0.02
7	0.81	0.81	0.87	0.03	0.08	0.03
8	0.71	0.59	0.78	0.03	0.12	0.06
9	0.80	0.72	0.80	-0.02	0.11	0.07
10	0.68	0.73	0.79	-0.05	-0.00	0.00
	Explained Variation			Cross-Sectional R^2		
All	0.70	0.75	0.82	0.97	0.94	0.98

Explained Variation and Pricing Errors for Book-to-Market Ratio Sorted Portfolios

Correlation of SDF Factors



Correlation between SDF Factors for Different Models

⇒ GAN SDF factor has low correlation with the market factor and FFN.

⇒ GAN has highest correlation with its linear special case EN

Simulation Results

Simulation Results - Setup

- Excess returns follow a no-arbitrage model with SDF factor F

$$R^e_{i,t+1} = \beta_{i,t} F_{t+1} + \epsilon_{i,t+1}.$$

- The SDF factor: $F_t \overset{i.i.d.}{\sim} \mathcal{N}(\mu_F, \sigma_F^2)$ with $\sigma_F^2 = 0.1$ and $SR_F = 1$.
- The idiosyncratic component: $\epsilon_{i,t} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_e^2)$ with $\sigma_e^2 = 1$.
- $N = 500$ and $T = 600$. Training/validation/test split is 250,100,250.

Case 1: One characteristic and one macroeconomic state process (LSTM matters):

$$\beta_{i,t} = C^{(1)}_{i,t} \cdot b(h_t), \qquad h_t = sin(\pi * t/24) + \epsilon^h_t.$$

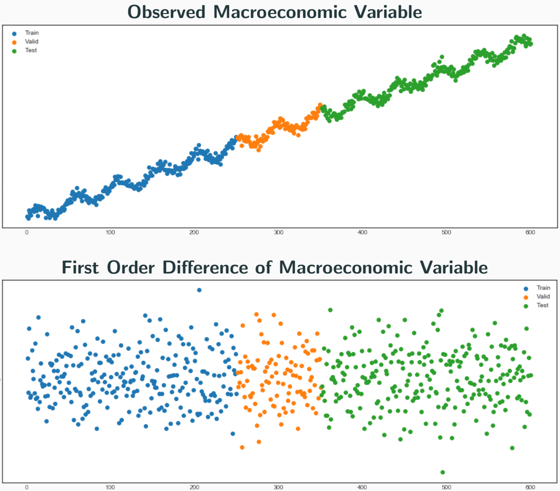
$$b(h) = \begin{cases} 1 & \text{if } h > 0 \\ -1 & \text{otherwise.} \end{cases}$$

- Only observe macroeconomic time-series $Z_t = \mu_M t + h_t$.
- All innovations are i.i.d.: $C^{(1)}_{i,t} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$ and $\epsilon^h_t \overset{i.i.d.}{\sim} \mathcal{N}(0, 0.25)$.

Case 2: Two interacting characteristics (GAN matters):

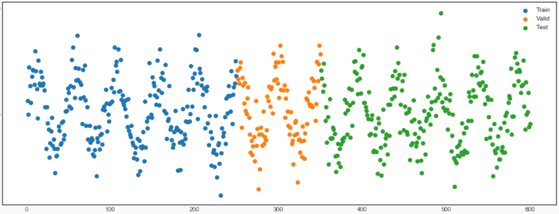
$$\beta_{i,t} = C^{(1)}_{i,t} \cdot C^{(2)}_{i,t} \qquad \text{with} \qquad C^{(1)}_{i,t}, C^{(2)}_{i,t} \overset{i.i.d.}{\sim} \mathcal{N}(0,1).$$

Simulation Results for Case 1 - Observed Macroeconomic Variable

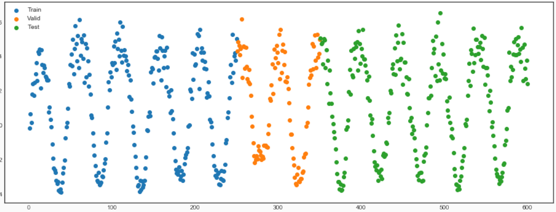


Simulation Results for Case 1- Fitted Macroeconomic State

True Hidden Macroeconomic State



Fitted Macroeconomic State by LSTM



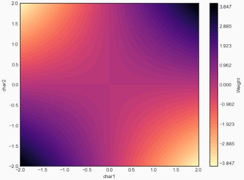
Simulation Results for Case 1 - Evaluation

	Sharpe Ratio			EV			Cross-sectional R^2		
Model	Train	Valid	Test	Train	Valid	Test	Train	Valid	Test
Population	0.89	0.92	0.86	0.18	0.18	0.17	0.19	0.20	0.15
GAN	0.79	0.77	0.64	0.18	0.18	0.17	0.19	0.20	0.15
FFN	0.05	-0.05	0.06	0.02	0.01	0.02	0.01	0.01	0.02
LS	0.12	-0.05	0.10	0.16	0.16	0.15	0.15	0.18	0.14

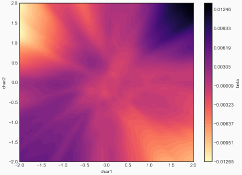
Performance of Different SDF Models for Case 1.

Simulation for Case 2 - Nonlinear Interaction

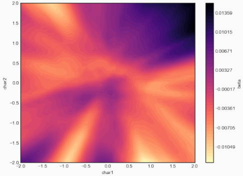
SDF weight ω with 2 characteristics



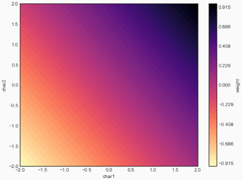
(a) Population Model



(b) GAN



(c) FFN



(d) LS

Simulation for Case 2 - Nonlinear Interaction

	Sharpe Ratio			EV			Cross-sectional R^2		
Model	Train	Valid	Test	Train	Valid	Test	Train	Valid	Test
Population	0.96	1.09	0.94	0.16	0.15	0.17	0.17	0.15	0.17
GAN	0.98	1.11	0.94	0.12	0.11	0.13	0.10	0.09	0.07
FFN	0.94	1.04	0.89	0.05	0.04	0.05	-0.30	-0.09	-0.33
LS	0.07	-0.10	0.01	0.00	0.00	0.00	0.00	0.01	0.01

Performance of Different SDF Models for Case 2

Implementation

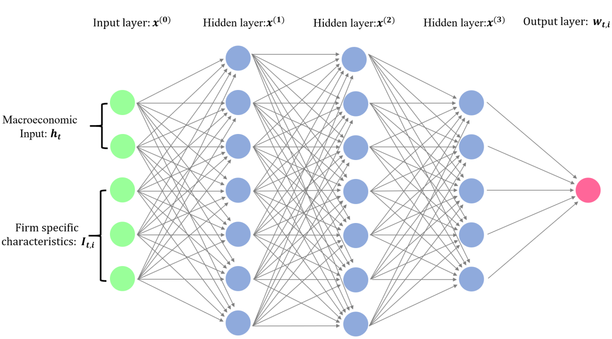
Hyper-Parameter Search

Notation	Hyperparameters	Candidates	Optimal
HL	Number of layers in SDF Network	2, 3 or 4	2
HU	Number of hidden units in SDF Network	64	64
SMV	Number of hidden states in SDF Network	4 or 8	4
CSMV	Number of hidden states in Conditional Network	16 or 32	32
CHL	Number of layers in Conditional Network	0 or 1	0
CHU	Number of hidden units in Conditional Network	4, 8, 16 or 32	8
LR	Initial learning rate	0.001, 0.0005, 0.0002 or 0.0001	0.001
DR	Dropout	0.95	0.95

Selection of Hyperparameters for GAN

1. For each combination of hyperparameters (384 models) we fit the GAN model.
2. We select the five best combinations of hyperparameters on validation data set.
3. For each of the five combinations we fit 9 models with the same hyperparameters but different initialization.
4. We select the ensemble model with the best performance on validation data set.

Feedforward Network

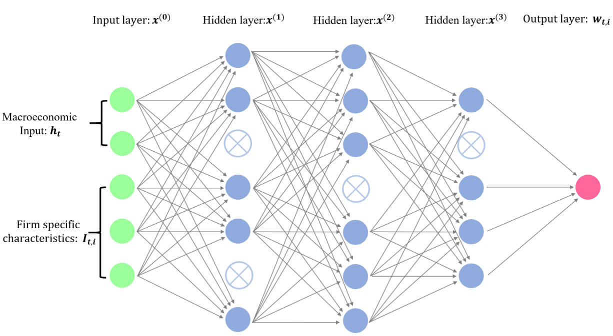


Feedforward Network with 3 Hidden Layers

$$x^{(l)} = \mathrm{ReLU}(W^{(l-1)\top} x^{(l-1)} + w_0^{(l-1)})$$

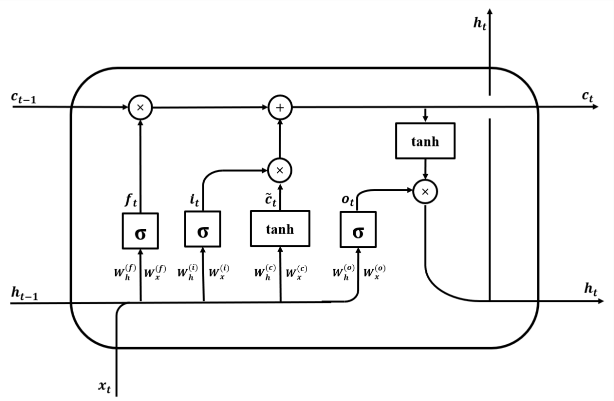
$$y = W^{(L)\top} x^{(L)} + w_0^{(L)}$$

Feedforward Network with Dropout



Feedforward Network with 3 Hidden Layers and Dropout

Long-Short-Term-Memory Cell (LSTM)



Long-Short-Term-Memory Cell (LSTM)

LSTM Cell Structure

At each step, a new memory cell $\tilde{c}_t$ is created with current input x_t and previous hidden state h_{t-1}

$$\tilde{c}_t = \tanh(W_h^{(c)} h_{t-1} + W_x^{(c)} x_t + w_0^{(c)}).$$

The input and forget gate control the memory cell, while the output gate controls the amount of information stored in the hidden state:

$$\begin{aligned} \text{input}_t =& \sigma(W_h^{(i)} h_{t-1} + W_x^{(i)} x_t + w_0^{(i)}) \\ \text{forget}_t =& \sigma(W_h^{(f)} h_{t-1} + W_x^{(f)} x_t + w_0^{(f)}) \\ \text{out}_t =& \sigma(W_h^{(o)} h_{t-1} + W_x^{(o)} x_t + w_0^{(o)}). \end{aligned}$$

The final memory cell and hidden state are given by

$$\begin{aligned} c_t =& \text{forget}_t \circ c_{t-1} + \text{input}_t \circ \tilde{c}_t \\ h_t =& \text{out}_t \circ \tanh(c_t). \end{aligned}$$

Economic Significance of Variables

- We define the sensitivity of a particular variable as the average absolute derivative of the weight w with respect to this variable:

$$\text{Sensitivity}(x_j) = \frac{1}{C} \sum_{i=1}^{N} \sum_{t=1}^{T} \left| \frac{\partial w(I_t, I_{t,i})}{\partial x_j} \right|,$$

  where C a normalization constant.

- A sensitivity of value z for a given variable means that the weight w will approximately change (in magnitude) by $z\Delta$ for a small change of Δ in this variable.

$\Rightarrow$ Generalization of linear slope coefficients!