Deep Learning in Asset Pricing

Luyang Chen †  Markus Pelger ‡  Jason Zhu ‡

†Institute for Computational & Mathematical Engineering, Stanford University
‡Department of Management Science & Engineering, Stanford University

March 7, 2019
Doctoral Seminar
Efficient markets: Asset returns dominated by unforecastable news

⇒ Financial return data has very low signal-to-noise ratio

⇒ This paper: Including financial constraints (no-arbitrage) in learning algorithm significantly improves signal
Motivation: Asset Pricing

The Challenge of Asset Pricing

- One of the most important questions in finance:

  **Why are asset prices different for different assets?**

- **No-Arbitrage Pricing Theory**: Stochastic discount factor SDF (also called pricing kernel or equivalent martingale measure) explains differences in risk and asset prices.

- Fundamental question: What is the SDF?

- Challenges:
  - SDF should depend on all available economic information: Very large set of variables
  - Functional form of SDF unknown and likely complex
  - SDF needs to capture time-variation in economic conditions
  - Risk premium in stock returns has a low signal-to-noise ratio
Goals of this paper:

General non-linear asset pricing model and optimal portfolio design

⇒ Deep-neural networks applied to all U.S. equity data and large sets of macroeconomic and firm-specific information.

Why is it important?

1. Stochastic discount factor (SDF) generates tradeable portfolio with highest risk-adjusted return
   (Sharpe-ratio=expected excess return/standard deviation)

2. Arbitrage opportunities
   - Find underpriced assets and earn “alpha”

3. Risk management
   - Understand which information and how it drives the SDF
   - Manage risk exposure of financial assets
Contribution of this paper

Contribution

This Paper: Estimate the SDF with deep neural networks

Crucial innovation: Include no-arbitrage condition in the neural network algorithm and combine four 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. Dimension reduction: Regularization through no-arbitrage condition
5. Signal-to-noise ratio: No-arbitrage conditions increase the signal to noise-ratio

⇒ General model that includes all existing models as a special case
Contribution of this paper

Empirical Contributions

- Empirically outperforms all benchmark models.
- Optimal portfolio has out-of-sample annual Sharpe ratio of 2.1.
- Non-linearities and interaction between firm information matters.
- Most relevant firm characteristics are price trends, profitability, and capital structure variables.
Literature (Partial List)

- Deep-learning for predicting asset prices
  - Feng, Polson and Xu (2019)
  - Gu, Kelly and Xiu (2018)
  - Feng, Polson and Xu (2018)
  - Messmer (2017)
  - Predicting future asset returns with feed forward network
    - Gu, Kelly and Xiu (2019)
    - Heaton, Polson and Witte (2017)
  - Fitting asset returns with an autoencoder

- Linear or kernel methods for asset pricing of large data sets
  - Feng, Giglio and Xu (2017): Risk-premium lasso
  - Freyberger, Neuhierl and Weber (2017): Group lasso
  - Kelly, Pruitt and Su (2018): Instrumented PCA
The Model

No-arbitrage pricing

- \( R_{i,t+1}^e \) = excess return (return minus risk-free rate) at time \( t + 1 \) for asset \( i = 1, \ldots, N \)

- Fundamental no-arbitrage condition:
  for all \( t = 1, \ldots, T \) and \( i = 1, \ldots, N \)

  \[ \mathbb{E}_t[M_{t+1} R_{i,t+1}^e] = 0 \]

- \( \mathbb{E}_t[.] \) 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

  \[ \mathbb{E}[M_{t+1} R_{t+1,i}^e, I_t] = 0 \]

  for any \( \mathcal{F}_t \)-measurable variable \( I_t \)
No-arbitrage pricing

- Without loss of generality SDF is projection on the return space

\[ M_{t+1} = 1 - \sum_{i=1}^{N} w_{i,t} R_{i,t+1}^{e} \]

⇒ Optimal portfolio \( \sum_{i=1}^{N} w_{i,t} R_{i,t+1}^{e} \) has highest conditional Sharpe-ratio

- Portfolio weights \( w_{i,t} \) are a general function of macro-economic information \( I_{t} \) and firm-specific characteristics \( I_{i,t} \):

\[ w_{i,t} = w(I_{t}, I_{i,t}) \]

⇒ Need non-linear estimator with many explanatory variables!

⇒ Use a feed forward network to estimate \( w_{i,t} \)
The Model

Equivalent factor model representation

- No-arbitrage condition is equivalent to

\[ \mathbb{E}_t[R^e_{i,t+1}] = \frac{\text{cov}_t(R^e_{i,t+1}, F_{t+1})}{\text{var}_t(F_{t+1})} \cdot \mathbb{E}_t[F_{t+1}] \]

\[ = \beta_{i,t} \mathbb{E}_t[F_{t+1}] \]

with factor \( F_t = 1 - M_t \).

\[ \Rightarrow \] Without loss of generality we have a factor representation

\[ R^e_{t+1} = \beta_t F_{t+1} + \epsilon_{t+1} \]
The Model

Objects of Interest

We use different approaches to estimate:

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

Asset Pricing Performance Measure

- Sharpe ratio of SDF factor: $SR = \frac{\hat{E}[F_t]}{\sqrt{\text{Var}(F_t)}}$
- 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 $R^2$: $\text{XS-R}^2 = 1 - \frac{\left(\frac{1}{N} \sum_{i=1}^{N} \frac{T_i}{T} \left(\frac{1}{T_j} \sum_{t \in T_i} \hat{e}_{i,t+1}\right)^2\right)}{\left(\frac{1}{N} \sum_{i=1}^{N} \frac{T_i}{T} \left(\frac{1}{T_j} \sum_{t \in T_i} \hat{R}_{i,t+1}\right)^2\right)}$
Objective Function for Estimation

- Estimate SDF portfolio weights $w(.)$ to minimize the no-arbitrage moment conditions.
- For a set of conditioning variables $\hat{I}_t$ the loss function is

$$L(\hat{I}_t) = \frac{1}{N} \sum_{i=1}^{N} \frac{T_i}{T} \left( \frac{1}{T_i} \sum_{t=1}^{T_i} M_{t+1} R_{i, t+1} \hat{I}_t \right)^2.$$ 

- Allows unbalanced panel.
- How can we choose the conditioning variables $\hat{I}_t = f(l_t, l_{i, t})$ as general functions of the macroeconomic and firm-specific information?

$\Rightarrow$ Generative Adversarial Network (GAN) chooses $\hat{I}_t$!
Determining Moment Conditions

- Two networks play zero-sum game:
  1. one network creates the SDF $M_{t+1}$
  2. other network creates the conditioning variables $\hat{I}_t$

- Iteratively update the two networks:
  1. for a given $\hat{I}_t$ the SDF network minimizes the loss
  2. for a given SDF the conditional networks finds $\hat{I}_t$ with the largest loss (most mispricing)

⇒ Intuition: find the economic states and assets with the most pricing information
Transforming Macroeconomic Time-Series

- **Problems** with economic time-series data
  - Time-series data is often non-stationary ⇒ transformation necessary
  - Business cycles can affect pricing ⇒ assuming Markovian structure of the pricing kernel not sufficient
  - Redundant information ⇒ large number of predictors prove to negatively impact model performance

- **Solution**: Recurrent Neural Network (RNN) with Long-Short-Term Memory (LSTM) cells
  - Transform all macroeconomic time-series into a low dimensional vector of stationary state variables
Model Architecture

**SDF Network:**
Update parameters to minimize loss

**State RNN**
- \( I_t \)
- \( \tilde{I}_t \)

**Feed Forward Network**
- \( w_{i,t} \)
- \( I_{i,t} \)

**Construct SDF**
- \( M_{t+1} \)

**Moment RNN**
- \( \bar{I}_t \)
- \( \tilde{I}_{i,t} \)

**Feed Forward Network**
- \( \bar{I}_{i,t} \)

**Loss Calculation**
- \( L \)
- \( R_{t+1}^{e} \)

**Iterative Optimizer with GAN**
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) ⇒ $l_{i,t}$
normalized to cross-sectional quantiles

178 macroeconomic variables
(124 from FRED, 46 cross-sectional median time-series for characteristics, 8 from Goyal-Welch) ⇒ $l_t$

T-bill rates from Kenneth-French website

Training/validation/test split is 20y/5y/25y
## Benchmark models

<table>
<thead>
<tr>
<th></th>
<th>Benchmark models</th>
</tr>
</thead>
</table>
| 1 | **LS & EN - Linear factor models:**  
The optimal portfolio weights $w_t = l_t \theta$ is linear in characteristics. We minimize loss function  
\[
\frac{1}{2} \left\| \frac{1}{T} \tilde{R}_t^K \top 1 - \frac{1}{T} \tilde{R}_t^K \top \tilde{R}_t^K \theta \right\|^2 + \lambda_1 \| \theta \|_1 + \frac{1}{2} \lambda_2 \| \theta \|_2^2.
\]
  $	ilde{R}_t^K = l_t \top \tilde{R}_t^e$ are $K$ portfolios weighted by characteristics $l_t$.  
| 2 | **FFN - Deep learning return forecasting** (Gu et al. (2018)):  
  - Predict conditional expected returns $\mathbb{E}_t[R_{i,t+1}]$  
  - Empirical loss function for prediction  
\[
\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (R_{i,t+1} - g(l_t, l_{i,t}))^2
\]
  - Use only simple feedforward network for forecasting |
### Results - Cross Section of Individual Stock Returns

**Table:** Performance of Different SDF Models

<table>
<thead>
<tr>
<th>Model</th>
<th>SR Train</th>
<th>SR Valid</th>
<th>SR Test</th>
<th>EV Train</th>
<th>EV Valid</th>
<th>EV Test</th>
<th>Cross-Sectional $R^2$ Train</th>
<th>Cross-Sectional $R^2$ Valid</th>
<th>Cross-Sectional $R^2$ Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>LS</td>
<td>1.35</td>
<td>0.80</td>
<td>0.45</td>
<td>0.09</td>
<td>0.04</td>
<td>0.03</td>
<td>0.03</td>
<td>0.04</td>
<td>0.02</td>
</tr>
<tr>
<td>EN</td>
<td>1.01</td>
<td>0.95</td>
<td>0.47</td>
<td>0.15</td>
<td>0.07</td>
<td>0.06</td>
<td>0.04</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td>FFN</td>
<td>0.30</td>
<td>0.28</td>
<td>0.36</td>
<td>0.16</td>
<td>0.07</td>
<td>0.06</td>
<td>0.01</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>GAN</td>
<td>3.26</td>
<td>0.97</td>
<td>0.60</td>
<td>0.21</td>
<td>0.10</td>
<td>0.08</td>
<td>0.01</td>
<td>0.05</td>
<td>0.05</td>
</tr>
</tbody>
</table>
## Results - Cross Section of Individual Stock Returns

### Table: SDF Factor Portfolio Performance

<table>
<thead>
<tr>
<th>Model</th>
<th>SR Train</th>
<th>SR Valid</th>
<th>SR Test</th>
<th>Max Loss Train</th>
<th>Max Loss Valid</th>
<th>Max Loss Test</th>
<th>Max Drawdown Train</th>
<th>Max Drawdown Valid</th>
<th>Max Drawdown Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>FF-3</td>
<td>0.27</td>
<td>-0.09</td>
<td>0.19</td>
<td>-2.45</td>
<td>-2.85</td>
<td>-4.31</td>
<td>7</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>FF-5</td>
<td>0.48</td>
<td>0.40</td>
<td>0.22</td>
<td>-2.62</td>
<td>-2.33</td>
<td>-4.90</td>
<td>4</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>LS</td>
<td>1.35</td>
<td>0.80</td>
<td>0.45</td>
<td>-1.82</td>
<td>-1.50</td>
<td>-3.67</td>
<td>2</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>EN</td>
<td>1.01</td>
<td>0.95</td>
<td>0.47</td>
<td>-3.22</td>
<td>-2.21</td>
<td>-5.99</td>
<td>2</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>FFN</td>
<td>0.30</td>
<td>0.28</td>
<td>0.36</td>
<td>-3.88</td>
<td>-4.93</td>
<td>-4.07</td>
<td>7</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>GAN</td>
<td>3.26</td>
<td>0.97</td>
<td>0.60</td>
<td>-0.09</td>
<td>-1.01</td>
<td>-4.48</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
Results - Cross Section of Individual Stock Returns

Figure: Cumulated Normalized SDF Portfolio
Results - Size Effect

Figure: GAN SDF Weight $\omega$ and Size (LME)

⇒ SDF portfolio is not predominantly investing in small stocks.
# Results - Sharpe Ratio for Forecasting Approach

Table: Sharpe Ratio of Long-Short Portfolios with FFN

<table>
<thead>
<tr>
<th>Quantile</th>
<th>SR (Train)</th>
<th>SR (Valid)</th>
<th>SR (Test)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(i) Equally-Weighted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1%</td>
<td>1.08</td>
<td>0.75</td>
<td>0.65</td>
</tr>
<tr>
<td>5%</td>
<td>1.26</td>
<td>1.15</td>
<td>0.70</td>
</tr>
<tr>
<td>10%</td>
<td>1.11</td>
<td>1.22</td>
<td>0.65</td>
</tr>
<tr>
<td>25%</td>
<td>1.03</td>
<td>1.20</td>
<td>0.56</td>
</tr>
<tr>
<td>50%</td>
<td>0.96</td>
<td>1.16</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>(ii) Value-Weighted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1%</td>
<td>0.77</td>
<td>0.55</td>
<td>0.41</td>
</tr>
<tr>
<td>5%</td>
<td>0.79</td>
<td>0.77</td>
<td>0.39</td>
</tr>
<tr>
<td>10%</td>
<td>0.59</td>
<td>0.46</td>
<td>0.32</td>
</tr>
<tr>
<td>25%</td>
<td>0.46</td>
<td>0.09</td>
<td>0.19</td>
</tr>
<tr>
<td>50%</td>
<td>0.42</td>
<td>0.23</td>
<td>0.18</td>
</tr>
</tbody>
</table>

⇒ Long-short portfolio is based on extreme quantiles.
Results - Predictive Performance

Figure: Cumulative Excess Return of Decile Sorted Portfolios by GAN

⇒ Risk loading predicts future stock returns.
### Table: Explained Variation and Pricing Errors for Short-Term Reversal Sorted Portfolios

<table>
<thead>
<tr>
<th>Decile</th>
<th>ST_REV</th>
<th>Explained Variation (EV)</th>
<th>Cross-Sectional $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Elastic Net</td>
<td>FFN</td>
</tr>
<tr>
<td>1</td>
<td>0.91</td>
<td>0.92</td>
<td>0.91</td>
</tr>
<tr>
<td>2</td>
<td>0.95</td>
<td>0.96</td>
<td>0.95</td>
</tr>
<tr>
<td>3</td>
<td>0.94</td>
<td>0.96</td>
<td>0.95</td>
</tr>
<tr>
<td>4</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>5</td>
<td>0.91</td>
<td>0.92</td>
<td>0.91</td>
</tr>
<tr>
<td>6</td>
<td>0.85</td>
<td>0.88</td>
<td>0.92</td>
</tr>
<tr>
<td>7</td>
<td>0.69</td>
<td>0.78</td>
<td>0.88</td>
</tr>
<tr>
<td>8</td>
<td>0.48</td>
<td>0.61</td>
<td>0.81</td>
</tr>
<tr>
<td>9</td>
<td>0.19</td>
<td>0.32</td>
<td>0.64</td>
</tr>
<tr>
<td>10</td>
<td>-0.03</td>
<td>-0.11</td>
<td>0.29</td>
</tr>
</tbody>
</table>

Overall: 0.70 0.72 0.81 0.87 0.89 0.95

Explained variation and pricing errors for decile-sorted portfolios based on Short-Term Reversal (ST_REV).
Table: Explained Variation and Pricing Errors for Momentum Sorted Portfolios

<table>
<thead>
<tr>
<th>Decile</th>
<th>Elastic Net</th>
<th>FFN</th>
<th>GAN</th>
<th>Elastic Net</th>
<th>FFN</th>
<th>GAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.22</td>
<td>0.25</td>
<td>0.48</td>
<td>0.29</td>
<td>0.30</td>
<td>0.71</td>
</tr>
<tr>
<td>2</td>
<td>0.49</td>
<td>0.52</td>
<td>0.72</td>
<td>0.73</td>
<td>0.82</td>
<td>0.98</td>
</tr>
<tr>
<td>3</td>
<td>0.68</td>
<td>0.73</td>
<td>0.86</td>
<td>0.90</td>
<td>0.97</td>
<td>1.00</td>
</tr>
<tr>
<td>4</td>
<td>0.81</td>
<td>0.85</td>
<td>0.91</td>
<td>0.95</td>
<td>1.00</td>
<td>0.99</td>
</tr>
<tr>
<td>5</td>
<td>0.89</td>
<td>0.90</td>
<td>0.92</td>
<td>1.00</td>
<td>1.00</td>
<td>0.98</td>
</tr>
<tr>
<td>6</td>
<td>0.92</td>
<td>0.90</td>
<td>0.89</td>
<td>1.00</td>
<td>0.99</td>
<td>0.98</td>
</tr>
<tr>
<td>7</td>
<td>0.91</td>
<td>0.89</td>
<td>0.86</td>
<td>0.99</td>
<td>0.99</td>
<td>0.98</td>
</tr>
<tr>
<td>8</td>
<td>0.88</td>
<td>0.88</td>
<td>0.84</td>
<td>0.98</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>9</td>
<td>0.84</td>
<td>0.85</td>
<td>0.82</td>
<td>0.99</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>10</td>
<td>0.80</td>
<td>0.79</td>
<td>0.77</td>
<td>1.00</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>Overall</td>
<td>0.61</td>
<td>0.63</td>
<td>0.73</td>
<td>0.86</td>
<td>0.87</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Explained variation and pricing errors for decile-sorted portfolios based on Momentum (r12.2).
## Results - ST_REV and r12_2 Double Sorted Portfolios

<table>
<thead>
<tr>
<th>ST_REV</th>
<th>r12_2</th>
<th>Explained Variation (EV)</th>
<th>Cross-Sectional $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Elastic Net</td>
<td>FFN</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.58</td>
<td>0.70</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>0.85</td>
<td>0.86</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>0.90</td>
<td>0.91</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>0.85</td>
<td>0.89</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>0.80</td>
<td>0.86</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.48</td>
<td>0.54</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0.79</td>
<td>0.81</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>0.87</td>
<td>0.86</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.80</td>
<td>0.83</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>0.79</td>
<td>0.82</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.24</td>
<td>0.26</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0.60</td>
<td>0.69</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0.81</td>
<td>0.83</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>0.86</td>
<td>0.85</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>0.78</td>
<td>0.77</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>-0.13</td>
<td>-0.22</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>0.20</td>
<td>0.41</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>0.54</td>
<td>0.71</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>0.72</td>
<td>0.80</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>0.68</td>
<td>0.67</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>-0.51</td>
<td>-0.81</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>-0.17</td>
<td>-0.06</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>0.18</td>
<td>0.38</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>0.35</td>
<td>0.44</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.43</td>
<td>0.44</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td>0.49</td>
<td>0.52</td>
</tr>
</tbody>
</table>
Figure: Correlation between SDF Factors for Different Models

(a) Whole Time Horizon

(b) Test Period
Table: GAN-SDF Factor and Fama-French 5 Factors

<table>
<thead>
<tr>
<th>Regression Coefficients</th>
<th>Mkt-RF</th>
<th>SMB</th>
<th>HML</th>
<th>RMW</th>
<th>CMA</th>
<th>intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.01)</td>
<td>0.07***</td>
<td>0.01</td>
<td>0.03</td>
<td>0.13***</td>
<td>-0.01</td>
<td>0.38***</td>
</tr>
<tr>
<td>(0.02)</td>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.03)</td>
<td></td>
</tr>
<tr>
<td>Correlation</td>
<td>0.14</td>
<td>-0.11</td>
<td>0.23</td>
<td>0.31</td>
<td>0.04</td>
<td>-</td>
</tr>
</tbody>
</table>

Out-of-sample correlation and regression of GAN SDF factor on the Fama-French 5 factors. The regression intercept is the monthly time-series pricing error of the SDF portfolio for the Fama-French model. Standard errors are in parenthesis.
Results - Characteristic Importance by GAN
Results - Macroeconomic Hidden State Processes
Figure: Weight as a function of Size (LME) and Book-to-Market Ratio (BEME)

⇒ Size and value have close to linear effect
Results - SDF Weights

Relationship between Weights and Characteristics

Figure: Weight as a function of Short-Term Reversal (ST\_REV)

⇒ non-linear effect
Results - SDF Weights

Relationship between Weights and Characteristics

Figure: Weight as a function of Size (LME) and Book-to-Market Ratio (BEME)

⇒ Size and value have non-linear interaction!
Results - SDF Weights

Relationship between Weights and Characteristics

Figure: Weight as a function of Size (LME), Book-to-Market Ratio (BEME) and Short-Term Reversal (ST_REV).

⇒ Complex interaction between multiple variables!
Simulation Setup

Motivation

We illustrate with simulations that

- the no-arbitrage condition in GAN is necessary to find the SDF in a low-signal to noise setup
- the flexible form of GAN is necessary to correctly capture the interactions between characteristics
- the LSTM-RNN is necessary to correctly incorporate macroeconomic dynamics in the pricing kernel
Simulation Setup

Setup

- Excess returns follow a no-arbitrage model with SDF factor $F$
  \[ R_{i,t+1}^e = \beta_{i,t} F_{t+1} + \epsilon_{i,t+1}. \]

- The SDF factor follows $F_t \sim i.i.d. \mathcal{N}(\mu_F, \sigma_F^2)$.

- The idiosyncratic component $\epsilon_{i,t} \sim i.i.d. \mathcal{N}(0, \sigma_e^2)$.

- $N = 500$ and $T = 600$. Define training/validation/test = 250, 100, 250.

- The SDF factor has $\sigma_F^2 = 0.1$ and $SR_F = 1$. The idiosyncratic noise variance $\sigma_e^2 = 1$. 
Simulation Setup

We consider two different formulations for the risk loadings

1. Two characteristics:

   \[ \beta_{i,t} = C_{i,t}^{(1)} \cdot C_{i,t}^{(2)} \quad \text{with} \quad C_{i,t}^{(1)}, C_{i,t}^{(2)} \sim \mathcal{N}(0, 1). \]

2. One characteristic and one macroeconomic state process:

   \[ \beta_{i,t} = C_{i,t}^{(1)} \cdot b(h_t), \quad h_t = \sin(\pi \cdot t/24) + \epsilon_t^h. \]

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

We observe only the macroeconomic time-series \( Z_t = \mu_M t + h_t \). All innovations are independent and normally distributed:

\[ C_{i,t}^{(1)}, i.i.d. \sim \mathcal{N}(0, 1) \text{ and } \epsilon_t^h, i.i.d. \sim \mathcal{N}(0, 0.25). \]
Simulation Results - Setup I

Loadings $\beta$ with 2 characteristics

(a) Population Model
(b) GAN
(c) FFN
(d) LS
## Simulation Results - Setup I

Table: Performance of Different SDF Models

<table>
<thead>
<tr>
<th>Model</th>
<th>Sharpe Ratio</th>
<th></th>
<th></th>
<th>EV</th>
<th></th>
<th></th>
<th>Cross-sectional $R^2$</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Train</td>
<td>Valid</td>
<td>Test</td>
<td>Train</td>
<td>Valid</td>
<td>Test</td>
<td>Train</td>
<td>Valid</td>
<td>Test</td>
</tr>
<tr>
<td>Population</td>
<td>0.96</td>
<td>1.09</td>
<td>0.94</td>
<td>0.16</td>
<td>0.15</td>
<td>0.17</td>
<td>0.17</td>
<td>0.15</td>
<td>0.17</td>
</tr>
<tr>
<td>GAN</td>
<td>0.98</td>
<td>1.11</td>
<td>0.94</td>
<td>0.12</td>
<td>0.11</td>
<td>0.13</td>
<td>0.10</td>
<td>0.09</td>
<td>0.07</td>
</tr>
<tr>
<td>FFN</td>
<td>0.94</td>
<td>1.04</td>
<td>0.89</td>
<td>0.05</td>
<td>0.04</td>
<td>0.05</td>
<td>-0.30</td>
<td>-0.09</td>
<td>-0.33</td>
</tr>
<tr>
<td>LS</td>
<td>0.07</td>
<td>-0.10</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
</tr>
</tbody>
</table>
Simulation Results - Setup II

**Observed Macroeconomic Variable**

- Train
- Valid
- Test

**First order difference of Macroeconomic Variable**

- Train
- Valid
- Test
Simulation Results - Setup II

True hidden Macroeconomic State

Fitted Macroeconomic State by LSTM
Simulation Results - Setup II

**Table: Performance of Different SDF Models**

<table>
<thead>
<tr>
<th>Model</th>
<th>Sharpe Ratio</th>
<th>EV</th>
<th>Cross-sectional $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Train</td>
<td>Valid</td>
<td>Test</td>
</tr>
<tr>
<td>Population</td>
<td>0.89</td>
<td>0.92</td>
<td>0.86</td>
</tr>
<tr>
<td>GAN</td>
<td>0.79</td>
<td>0.77</td>
<td>0.64</td>
</tr>
<tr>
<td>FFN</td>
<td>0.05</td>
<td>-0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>LS</td>
<td>0.12</td>
<td>-0.05</td>
<td>0.10</td>
</tr>
</tbody>
</table>
## Conclusion

### Summary

- Linear models perform well because when considering characteristics in isolation, the models are approximately linear.
- Non-linearities matter for the interaction.
- Most relevant variables are price trends and liquidity.
- Macroeconomic data has a low dimensional factor structure.
- Pricing all individual stocks leads to better pricing models on portfolios.
- SDF structure stable over time.
- Mean-variance efficient portfolio implied by pricing kernel highly profitable in a risk-adjusted sense.