Discussion of

"Deep Learning in Asset Pricing"

by Luyang Chen, Markus Pelger and Jason Zhu

Daniele Bianchi
University of Warwick

2019 LBS Summer Finance Symposium

Motivation

Why different stocks have different excess returns $R^e_{i,t+1}$?

Motivation

Why different stocks have different excess returns $R^e_{i,t+1}$?

Fundamental pricing equation

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

Motivation

Why different stocks have different excess returns $R^e_{i,t+1}$?

Fundamental pricing equation

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

Stochastic Discount Factor (SDF)

$$M_{t+1} = f\left(I_t\right)$$

key object that summarizes all pricing information.

Motivation

Why different stocks have different excess returns $R^e_{i,t+1}$?

Fundamental pricing equation

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

Stochastic Discount Factor (SDF)

$$M_{t+1} = f\left(I_t\right)$$

key object that summarizes all pricing information.

Issues and challenges:

- I_t could be of large dimension.
- $f\left(\cdot\right)$ is unknown and possibly non-linear.
- M_{t+1} difficult to estimate,
  i.e., returns are noisy.

Motivation

Why different stocks have different excess returns $R^e_{i,t+1}$?

Fundamental pricing equation

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

Stochastic Discount Factor (SDF)

$$M_{t+1} = f\left(I_t\right)$$

key object that summarizes all pricing information.

Issues and challenges:

- I_t could be of large dimension.
- $f\left(\cdot\right)$ is unknown and possibly non-linear.
- M_{t+1} difficult to estimate, i.e., returns are noisy.

This paper:

✓

✓

✓

Interesting !!!

Machine Learning

Why machine learning?

- It is **THE method** for out-of-sample forecasting in statistics
- It is designed to extract high-dimensional and non-linear statistical relations

Machine Learning

Why machine learning?

- It is **THE method** for out-of-sample forecasting in statistics
- It is designed to extract high-dimensional and non-linear statistical relations

In its general/simplest form:

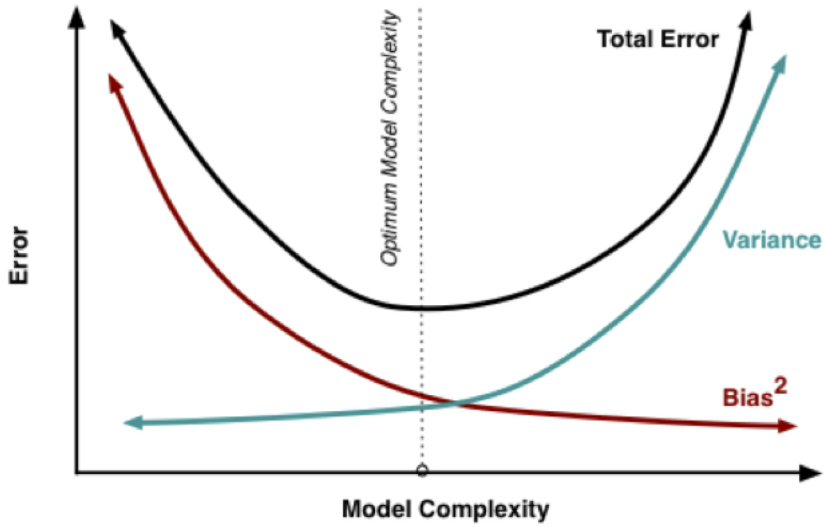
$$\min_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}; \cdot) = \underbrace{\mathcal{L}(\boldsymbol{\theta})}_{\text{Loss Function}} + \underbrace{\phi(\boldsymbol{\theta}_0; \cdot)}_{\text{Penalty Term}}$$

Machine Learning

Why machine learning?

- It is **THE method** for out-of-sample forecasting in statistics
- It is designed to extract high-dimensional and non-linear statistical relations

The **key idea** is simple: Variance vs. Bias



In its general/simplest form:

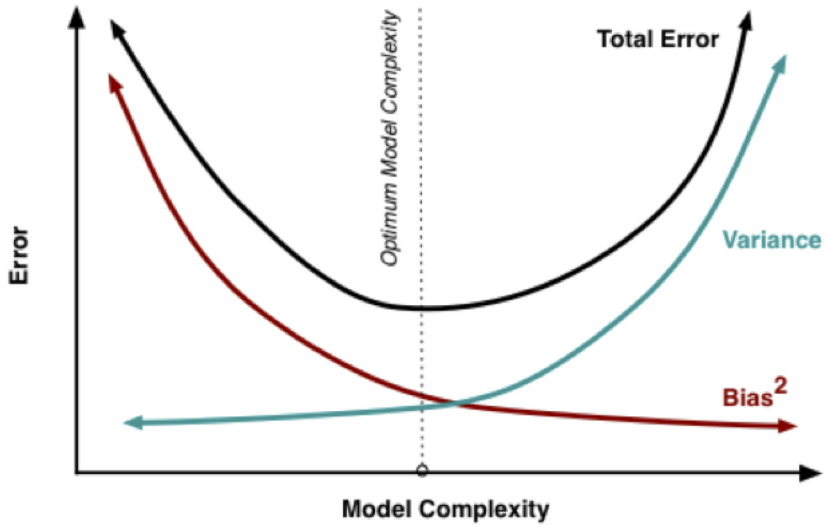
$$\min_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}; \cdot) = \underbrace{\mathcal{L}(\boldsymbol{\theta})}_{\text{Loss Function}} + \underbrace{\phi(\boldsymbol{\theta}_0; \cdot)}_{\text{Penalty Term}}$$

Machine Learning

Why machine learning?

- It is **THE method** for out-of-sample forecasting in statistics
- It is designed to extract high-dimensional and non-linear statistical relations

The **key idea** is simple: Variance vs. Bias



In its general/simplest form:

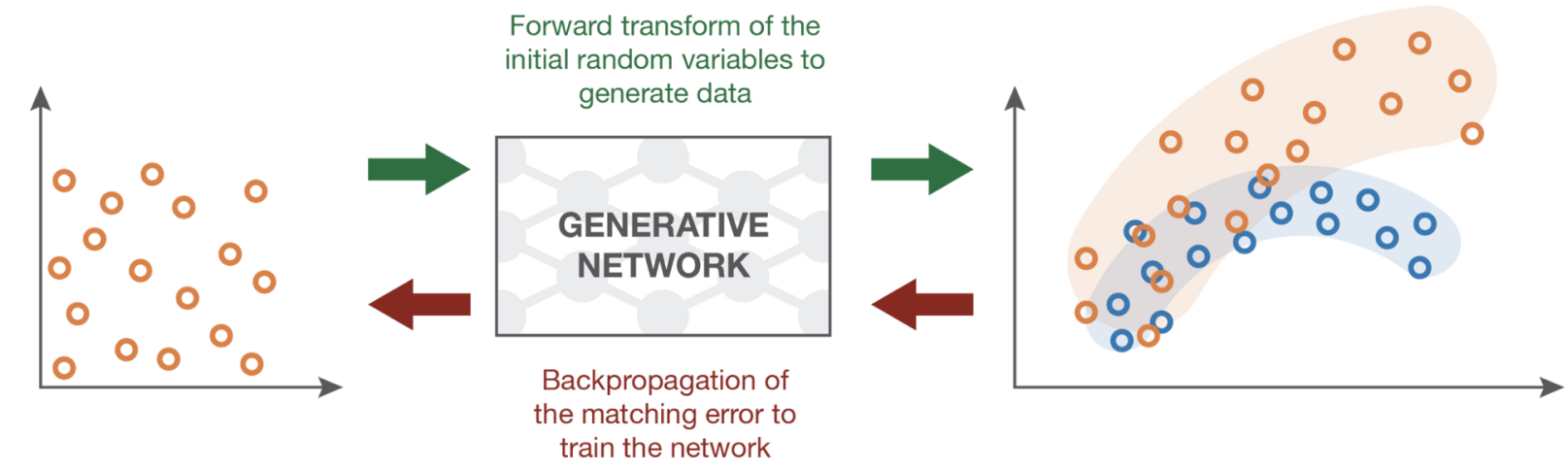
$$\min_{\theta} \mathcal{L}(\theta; \cdot) = \underbrace{\mathcal{L}(\theta)}_{\text{Loss Function}} + \underbrace{\phi(\theta_0; \cdot)}_{\text{Penalty Term}}$$

Contribution of this paper:

- **Economic structure on $\mathcal{L}(\theta)$**
- **Clever optimization strategy**

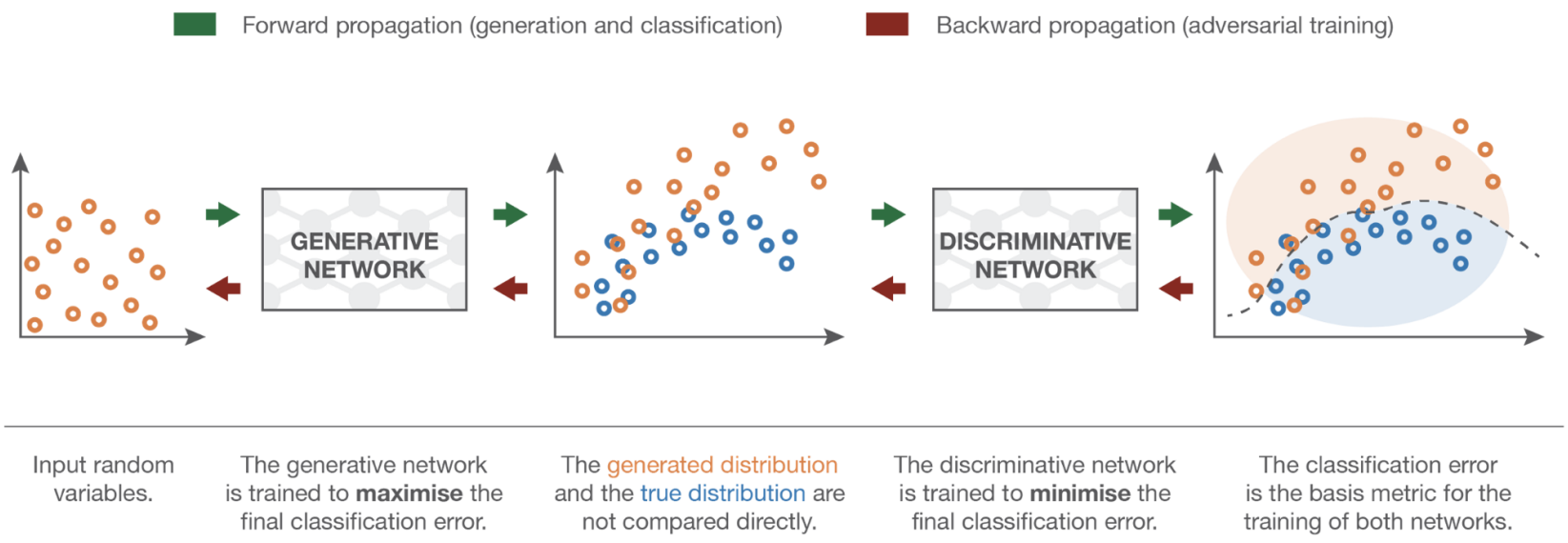
Generative Adversarial Network

Standard Neural Network (e.g., classification)



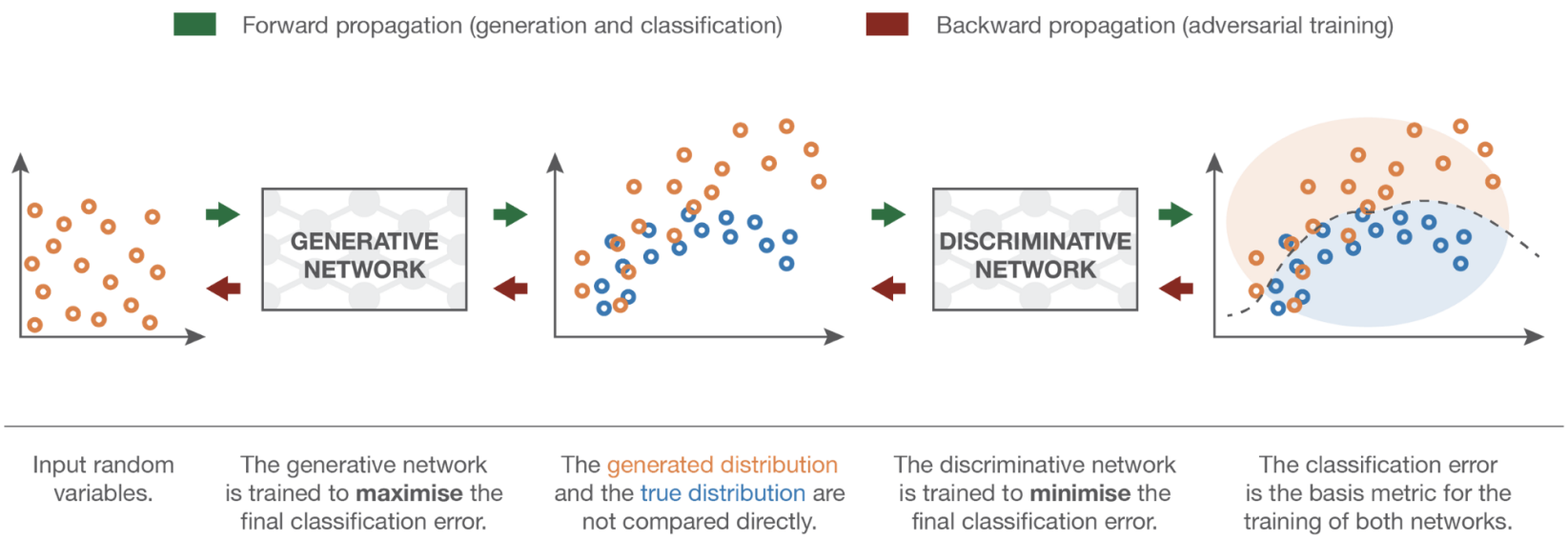
Generative Adversarial Network

Generative Adversarial Network (e.g., classification)



Generative Adversarial Network

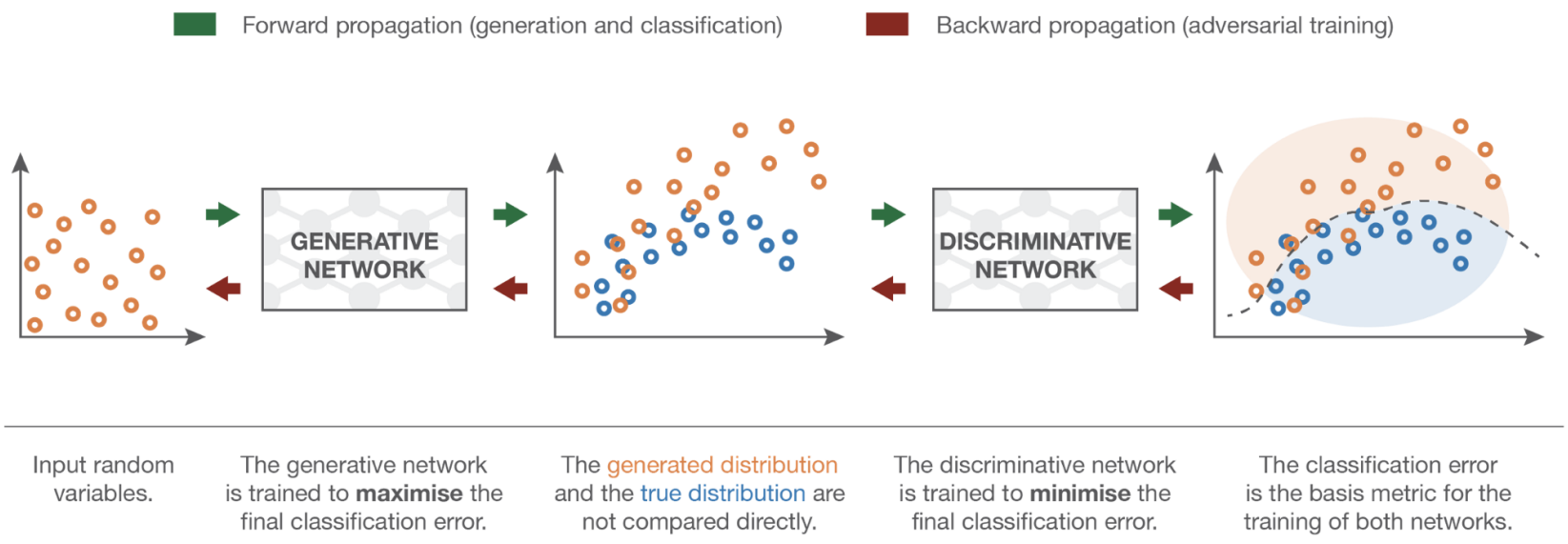
Generative Adversarial Network (e.g., classification)



This paper: The adversarial network "picks" the moment conditions that matter the most for pricing.

Generative Adversarial Network

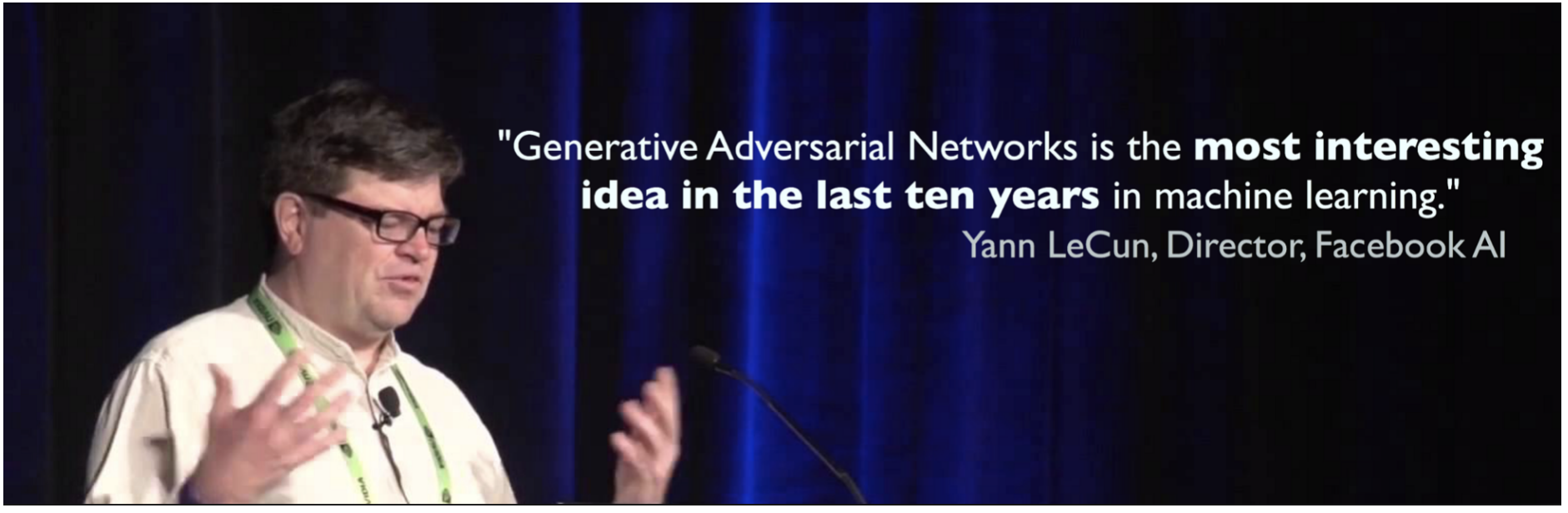
Generative Adversarial Network (e.g., classification)



This paper: The adversarial network "picks" the moment conditions that matter the most for pricing.

N.B: Conceptually similar to a Markov Chain Monte Carlo, e.g., RJ-MCMC, Metropolis-Hastings, etc.

Generative Adversarial Network



Main Results

GAN (somewhat) outperforms:

- A standard Deep Neural network (see Gu et al. 2018)
- A linear version of GAN
- A penalized linear model (elastic net)
- **Side comment:** why not others?

Main Results

GAN (somewhat) outperforms:

- A standard Deep Neural network (see Gu et al. 2018)
- A linear version of GAN
- A penalized linear model (elastic net)
- **Side comment:** why not others?

Some of the key insights:

- Not all non-linearities are alike, only interactions matter
- Macroeconomic states matter (at the right frequency)

My Comments:

My comments:

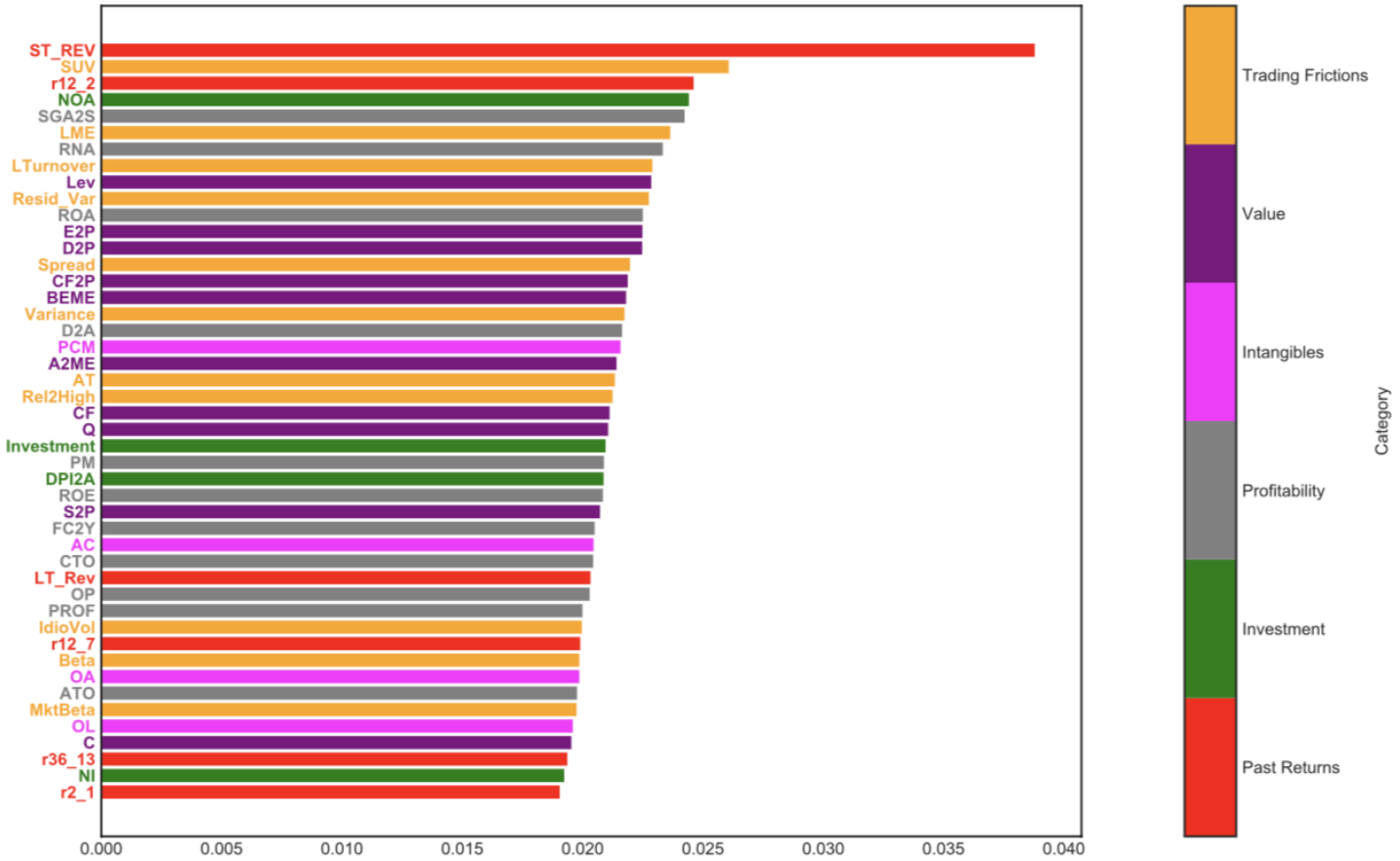
1. Variables importance
2. Out-of-sample results
3. Model dynamics
4. Macroeconomic variables

Comment 1 on Variables Importance

Variables Importance

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

Figure 13. Characteristic Importance by GAN

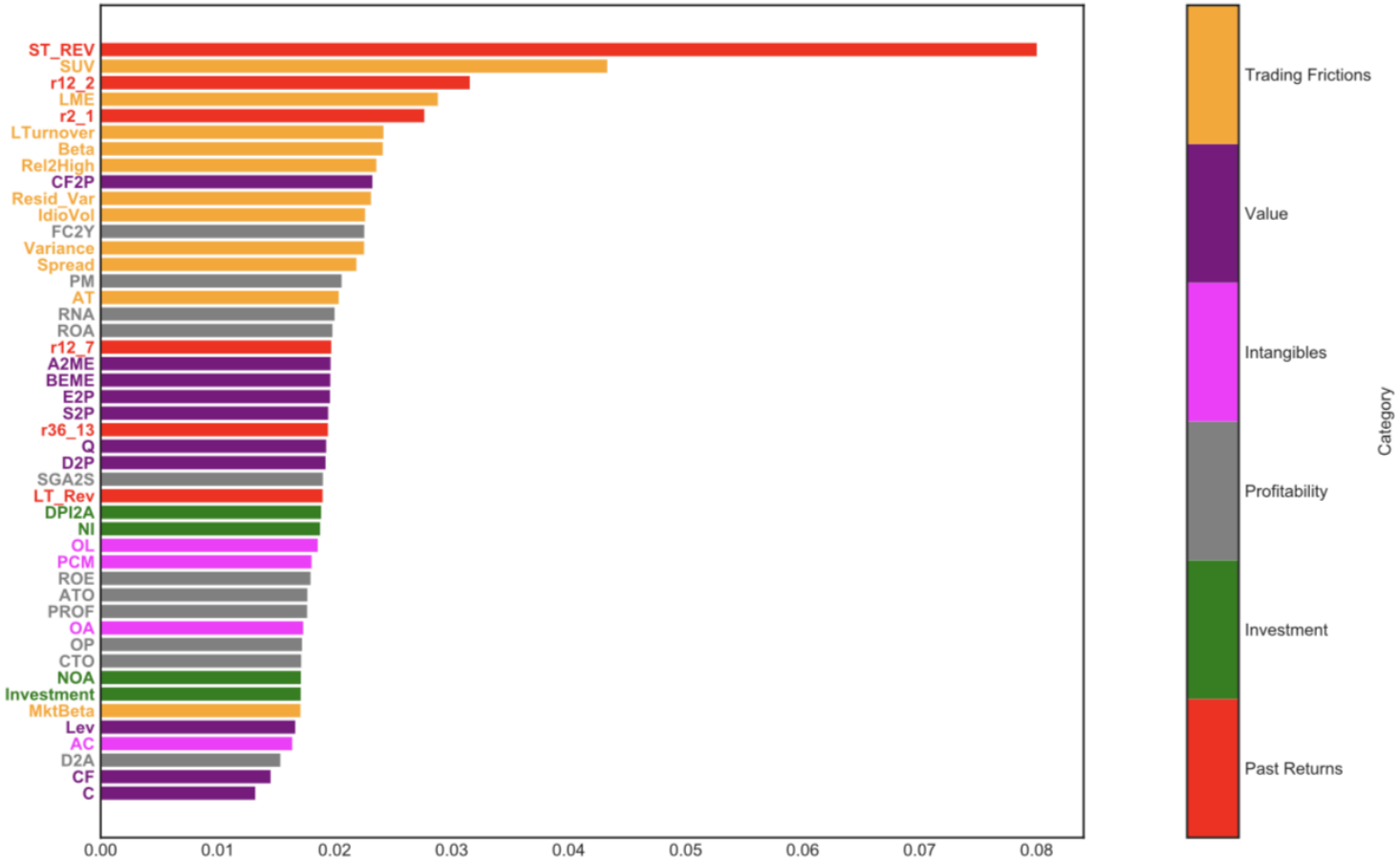


Comment 1 on Variables Importance

Variables Importance

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

Figure 14. Characteristic Importance for FFN



Comment 1 on Variables Importance

Comments:

- GAN seems to generate a pricing kernel which looks (close to) an equally-weighted average of characteristics
- Unlike other methods there is not clear group ranking
- Can we rationalize these results in some way? No group dominating implicitly means we are not able to appeal to a single theory underlying
- Different sensitivity analysis may give better results

Comment 1 on Variables Importance

Review and comparison of methods to study the contribution of variables in artificial neural network models

Muriel Gevrey [a,*], Ioannis Dimopoulos [b], Sovan Lek [a]

[a] *CESAC UMR 5576, CNRS-University Paul Sabatier, 118, route de Narbonne, 31062 Toulouse cedex, France*
[b] *Department of Health and Welfare Unit Administration, Technological Educational Institute of Kalamata, Antikalamos, 24100 Kalamata, Greece*

Comment 1 on Variables Importance

Review and comparison of methods to study the contribution of variables in artificial neural network models

Muriel Gevrey [a,*], Ioannis Dimopoulos [b], Sovan Lek [a]

[a] CESAC UMR 5576, CNRS-University Paul Sabatier, 118, route de Narbonne, 31062 Toulouse cedex, France
[b] Department of Health and Welfare Unit Administration, Technological Educational Institute of Kalamata, Antikalamos, 24100 Kalamata, Greece

---

A Unified Approach to Interpreting Model Predictions

Scott M. Lundberg
Paul G. Allen School of Computer Science
University of Washington
Seattle, WA 98105
slund1@cs.washington.edu

Su-In Lee
Paul G. Allen School of Computer Science
Department of Genome Sciences
University of Washington
Seattle, WA 98105
suinlee@cs.washington.edu

Comment 2 on the Out-of-Sample Results

Make clearer (economically) where GAN gives you advantage

Table III Performance of Different SDF Models

	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

Monthly Sharpe Ratio (SR) of the SDF factor, explained time series variation (EV) and cross-sectional mean R^2 for the GAN, FFN, EN and LS model.

Comment 2 on the Out-of-Sample Results

Make clearer (economically) where GAN gives you advantage

Table VIII Different SDF Models Evaluated on Large Market Cap Stocks

	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

Monthly Sharpe Ratio (SR) of the SDF factor, explained time series variation (EV) and cross-sectional mean R^2 for the GAN, FFN, EN and LS model. The model is estimated on all stocks but evaluated on stocks with market capitalization larger than 0.01% or 0.001% of the total market capitalization.

Comment 2 on the Out-of-Sample Results

Make clearer (economically) where GAN gives you advantage

Table VI Explained Variation and Pricing Errors for Decile Sorted Portfolios

	EN	FFN	GAN	EN	FFN	GAN	EN	FFN	GAN	EN	FFN	GAN
	Short-Term Reversal						Momentum					
Decile	Explained Variation			Alpha			Explained Variation			Alpha		
1	0.84	0.74	0.77	-0.18	-0.21	-0.13	0.04	-0.06	0.33	0.37	0.39	0.11
2	0.86	0.81	0.82	0.00	-0.05	0.00	0.12	0.10	0.52	0.25	0.18	-0.01
3	0.80	0.82	0.84	0.13	0.04	0.06	0.19	0.25	0.66	0.14	0.05	-0.06
4	0.69	0.80	0.82	0.16	0.03	0.03	0.28	0.34	0.73	0.15	0.08	-0.02
5	0.58	0.68	0.71	0.13	-0.03	-0.04	0.37	0.46	0.80	0.19	0.09	0.02
6	0.43	0.66	0.75	0.22	0.05	0.01	0.45	0.58	0.78	0.02	-0.03	-0.09
7	0.23	0.64	0.77	0.20	0.03	-0.02	0.62	0.69	0.68	0.01	0.01	-0.05
8	-0.07	0.49	0.67	0.23	0.03	-0.05	0.58	0.71	0.64	-0.03	-0.04	-0.09
9	-0.25	0.29	0.58	0.30	0.09	-0.01	0.55	0.70	0.58	0.08	0.04	-0.03
10	-0.24	-0.04	0.35	0.47	0.38	0.18	0.51	0.53	0.53	0.24	0.29	0.19
	Explained Variation			Cross-Sectional R^2			Explained Variation			Cross-Sectional R^2		
All	0.43	0.58	0.70	0.45	0.79	0.94	0.26	0.27	0.54	0.66	0.71	0.93

Comment 2 on the Out-of-Sample Results

Make clearer (economically) where GAN gives you advantage

Table VI Explained Variation and Pricing Errors for Decile Sorted Portfolios

	EN	FFN	GAN	EN	FFN	GAN	EN	FFN	GAN	EN	FFN	GAN
	Book-To-Market						Size					
Decile	Explained Variation			Alpha			Explained Variation			Alpha		
1	0.38	0.66	0.70	0.03	-0.12	-0.08	0.80	0.75	0.79	0.09	-0.00	0.10
2	0.48	0.73	0.78	0.10	-0.05	-0.04	0.89	0.89	0.90	-0.11	-0.09	-0.06
3	0.71	0.84	0.86	0.07	-0.03	-0.01	0.91	0.80	0.91	-0.07	0.02	-0.02
4	0.76	0.88	0.89	0.00	-0.07	-0.07	0.90	0.77	0.91	-0.05	0.04	-0.01
5	0.82	0.87	0.88	0.05	0.02	0.01	0.90	0.78	0.91	0.01	0.10	0.04
6	0.77	0.82	0.88	0.06	0.04	0.02	0.88	0.80	0.91	0.03	0.09	0.02
7	0.81	0.81	0.87	0.03	0.08	0.03	0.84	0.81	0.89	0.04	0.05	-0.01
8	0.71	0.59	0.78	0.03	0.12	0.06	0.84	0.85	0.88	0.06	0.03	-0.02
9	0.80	0.72	0.80	-0.02	0.11	0.07	0.77	0.81	0.82	0.06	-0.01	-0.04
10	0.68	0.73	0.79	-0.05	-0.00	0.00	0.32	0.28	0.49	-0.04	-0.15	-0.10
	Explained Variation			Cross-Sectional R^2			Explained Variation			Cross-Sectional R^2		
All	0.70	0.75	0.82	0.97	0.94	0.98	0.83	0.78	0.86	0.96	0.95	0.97

Comment 2 on the Out-of-Sample Results

Make clearer (economically) where GAN gives you advantage

Comments:

- From a **methodological standpoint** it is crystal clear why GAN is useful
- However, not fully clear why (and where) GAN are superior from an **economic standpoint** than standard Neural Networks
- It seems GAN performs better when **small, less liquid, more noisy stocks** are concerned
- Perhaps a separate focus on less liquid stocks could help appreciate GAN better

Comment 3 on Model Dynamics

Model dynamics:

At the outset you state: *"The SDF can have a complex dynamic structure and the risk exposure for individual assets can vary over time."*

Comment 3 on Model Dynamics

Model dynamics:

At the outset you state: *"The SDF can have a complex dynamic structure and the risk exposure for individual assets can vary over time."*

$g(\cdot)$ and $w(\cdot)$ are static

$+$

Model estimated through a fixed train/validation period

Comment 3 on Model Dynamics

Model dynamics:

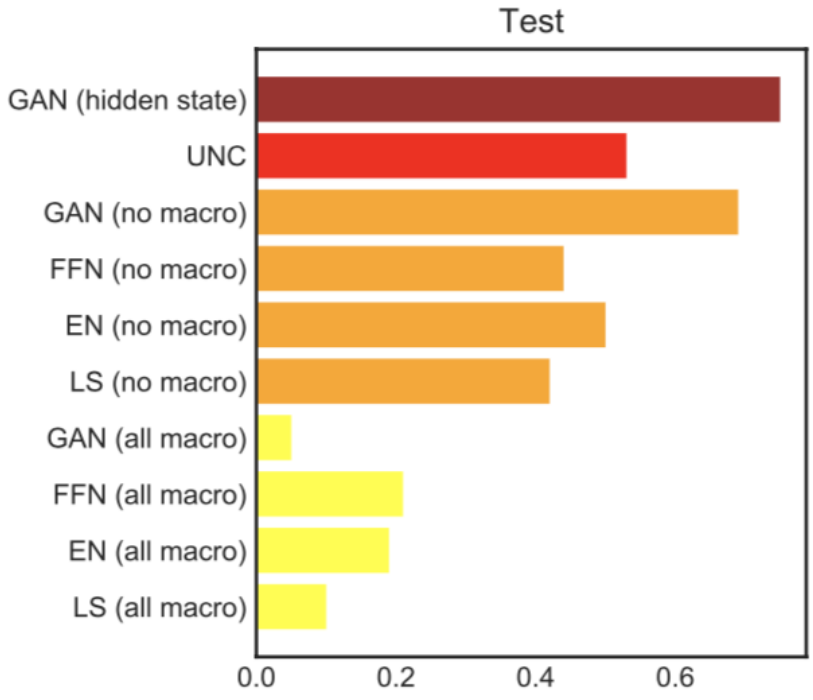
At the outset you state: *"The SDF can have a complex dynamic structure and the risk exposure for individual assets can vary over time."*

$$g(\cdot) \text{ and } w(\cdot) \text{ are static}$$
$$+$$
$$\text{Model estimated through a fixed train/validation period}$$

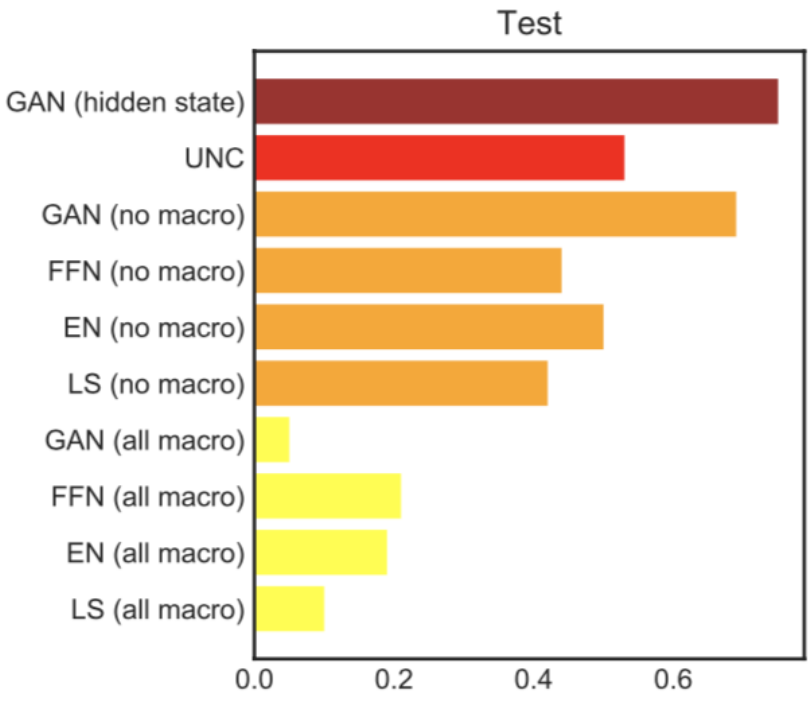
Comments:

- Is not clear what exactly is "dynamic" here
- The model is not recursively estimated, e.g., via rolling window
- Hard to capture any dynamic structure, breaks, in the model/network
- Perhaps recalibrate for different sub-samples
  (and show the sensitivities for each)

Comment 4 on Macroeconomic Variables



Comment 4 on Macroeconomic Variables

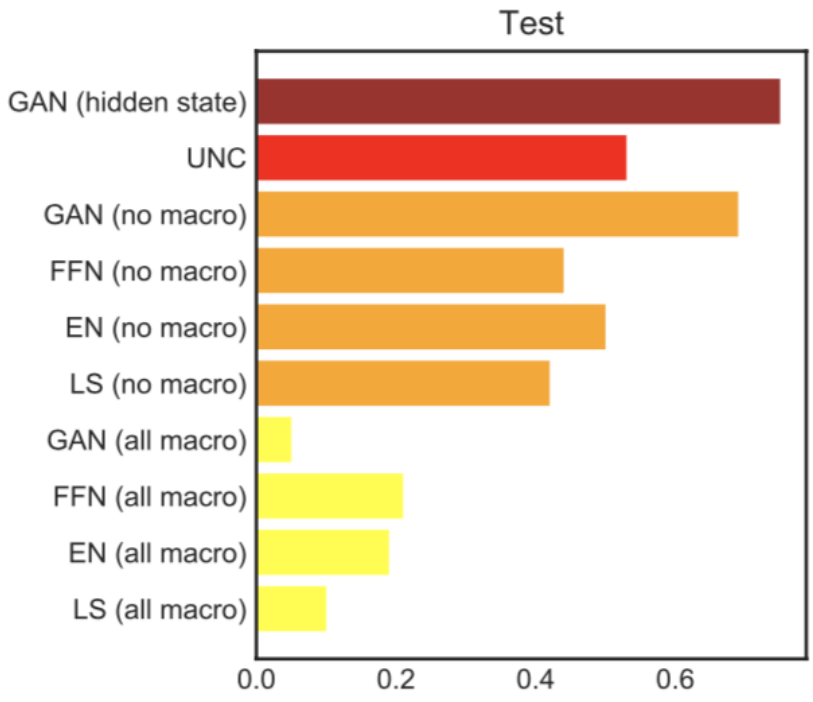


It seems Macro variables are detrimental

+

Summarizing information through LSTM is a cool idea

Comment 4 on Macroeconomic Variables



It seems Macro variables are detrimental

+

Summarizing information through LSTM is a cool idea

Comments:

- Maybe due to some specific Macro variable
- Sensitivity analysis of Macros could help discriminate
- By grouping variables (ensemble networks) macro effects can be stronger

Conclusion

My view: really interesting paper!! (must read)

What I like:

- Economic structure in the loss function (super nice idea)
- GAN is a powerful tool with many potentials in empirical finance

My suggestions for possible improvement:

- Let us know more about the economics of the results
- Revisit the sensitivity analysis
- Make clearer the model dynamics, if any
- Investigate better the role of macro variables