Alpha Go Everywhere: Machine Learning and International Stock Returns

Darwin Choi, Wenxi (Griffin) Jiang, Chao Zhang

CUHK Business School, CUHK Business School, University of Oxford

ABFER 10th Annual Conference 2023

May 22, 2023

International Stock Returns

- Variables related to firm characteristics (such as size, value, and momentum) predict international stock returns
    - Fama & French (1998, 2012, 2017), Hou, Karolyi, Kho (2011), Rouwenhorst (1998)
- Global equity markets are partially (but not fully) integrated
    - Karolyi & Stulz (2003), Lewis (2011)

International Stock Returns

- Variables related to firm characteristics (such as size, value, and momentum) predict international stock returns
  - Fama & French (1998, 2012, 2017), Hou, Karolyi, Kho (2011), Rouwenhorst (1998)
- Global equity markets are partially (but not fully) integrated
  - Karolyi & Stulz (2003), Lewis (2011)
- Not fully integrated: The predictive power of different firm characteristics varies across markets

International Stock Returns

- Variables related to firm characteristics (such as size, value, and momentum) predict international stock returns
  - Fama & French (1998, 2012, 2017), Hou, Karolyi, Kho (2011), Rouwenhorst (1998)
- Global equity markets are partially (but not fully) integrated
  - Karolyi & Stulz (2003), Lewis (2011)
- Not fully integrated: The predictive power of different firm characteristics varies across markets
- Not fully segmented: Information obtained from the firm characteristics in a foreign market may be relevant for local stocks

International Stock Returns

- How do we form predictions in *many* different markets without pre-specifying the models?

International Stock Returns

- How do we form predictions in *many* different markets without pre-specifying the models?
- How do foreign firm characteristics improve the return predictability of local models?

International Stock Returns

- How do we form predictions in *many* different markets without pre-specifying the models?
- How do foreign firm characteristics improve the return predictability of local models?
- We use machine learning techniques
  - Capable of detecting non-monotonic relationships and complex interactions between returns and many characteristics even without local knowledge
    - Say, the price limit rules in China and the investor clientele in India)

International Stock Returns

- How do we form predictions in *many* different markets without pre-specifying the models?
- How do foreign firm characteristics improve the return predictability of local models?
- We use machine learning techniques
  - Capable of detecting non-monotonic relationships and complex interactions between returns and many characteristics even without local knowledge
    - Say, the price limit rules in China and the investor clientele in India)
  - Able to explore various foreign variables and their interactions with local variables, which may not be ideal in a linear setting

Our Paper

- We use stock characteristics to predict international stock returns in 31 markets
  - Gu, Kelly, Xiu (GKX 2020) examine the U.S. market with 94 stock characteristics, 8 macroeconomic variables, and 74 sector dummies (920 predictors $= 94 \times (8 + 1) + 74$)
  - We use GKX's methods and set of potential hyperparameter values
  - Data availability is lower internationally. We use 36 features
  - (Our U.S. results are similar to those in GKX)
- Following GKX, we compare
  - Linear (OLS, OLS with Huber loss, LASSO, RIDGE, ENET)
  - Tree models (RF and GBRT+H)
  - Neural network (NN) models with 1–5 hidden layers (NN1–NN5)

Main Findings

- We train and validate our models using U.S. data and apply them on each of the 31 markets ("U.S.-trained models")
  - A stringent out-of-sample test: hyperparameters and parameters estimated from the U.S.
  - NNs and tree models outperform linear models in forming profitable portfolios and predicting return rankings (but not in terms of R^2; Kelly, Malamud, Zhou, 2021)

Main Findings

- We train and validate our models using U.S. data and apply them on each of the 31 markets ("U.S.-trained models")
  - A stringent out-of-sample test: hyperparameters and parameters estimated from the U.S.
  - NNs and tree models outperform linear models in forming profitable portfolios and predicting return rankings (but not in terms of R^2; Kelly, Malamud, Zhou, 2021)
- We construct market-specific models and compare them with the U.S.-trained models
  - We train and validate our models separately for each country
  - NNs achieve even stronger results
  - Trees underperform linear models, especially in markets where the number of observations is low

Main Findings

- Local NN models outperform U.S.-Trained NN models
  - Bigger improvements in profitability when
    - The local NN is less similar to the U.S.-trained NN, measured by the central kernel alignment (CKA) similarity index (Kornblith et al., 2019)
    - The NN has more hidden layers

Main Findings

- Local NN models outperform U.S.-Trained NN models
  - Bigger improvements in profitability when
    - The local NN is less similar to the U.S.-trained NN, measured by the central kernel alignment (CKA) similarity index (Kornblith et al., 2019)
    - The NN has more hidden layers
- We examine cross-market integration
  - U.S. market returns predict other markets' returns (Rapach, Strauss, Zhou, 2013)
  - Local NNs can be further improved by adding U.S.-based variables
    - U.S. characteristic gaps: $95^{th} - 5^{th}$ of a stock characteristic (Cohen, Polk, Vuolteenaho, 2003; Huang, 2021)
  - The variable importance of the U.S. characteristic gaps increases with the market integration metrics constructed by Bekaert et al. (2011) and Akbari, Ng, Solnik (2020)

Data and Methodology

- 36 variables include past returns, market cap, volume, past returns of the industry, and accounting information: DataStream
- US data from CRSP and China data from CSMAR
- We require at least 100 stocks for at least 3 years: 31 markets + U.S.

- We divide the sample period into 3 parts
  1. Train: Estimate the model subject to a particular set of hyperparameter values
  2. Validate: Construct forecasts and calculate objective functions based on the estimated model from the training sample. Iteratively search for hyperparameters that optimize the objective functions
  3. Test: Not used for estimation or tuning. "Out-of-sample"

Data and Methodology

- Standardize all predictors in each month in each market to zero mean and unit standard deviation
- Outcome variable: Excess return in USD in the next month (denominated in USD, in excess of the corresponding market return)
  - We focus on the return predictability in the cross-section

Data and Methodology

- Standardize all predictors in each month in each market to zero mean and unit standard deviation
- Outcome variable: Excess return in USD in the next month (denominated in USD, in excess of the corresponding market return)
  - We focus on the return predictability in the cross-section
- We evaluate the models using
  - Annualized Sharpe Ratio of the long–short portfolio (long top decile, short bottom decile of predicted returns)
  - Out-of-sample R^2: $1 - \frac{\sum_{(i,t)\in \text{Test}} (r_{i,t} - \hat{r}_{i,t})^2}{\sum_{(i,t)\in \text{Test}} r_{i,t}^2}$
  - Rank Correlation: Spearman rank correlation between predicted and realized returns
  - Decile Score Distance: Distance between the long portfolio's actual return deciles and the short portfolio's
  - Sum of the Squares of the partial Derivatives (SSD): Measuring variable importance

U.S.-Trained Models: Summary

	Sharpe Ratio (EW)			Sharpe Ratio (VW)			R^2_{oos}		
	Tree−Linear	NN−Linear	NN−Tree	Tree−Linear	NN−Linear	NN−Tree	Tree−Linear	NN−Linear	NN−Tree
difference	0.41	0.65	0.25	-0.05	0.37	0.42	-1.17	0.01	1.18
# of +	26	30	25	15	27	29	8	17	30
fraction of +	0.84	0.97	0.81	0.48	0.87	0.94	0.26	0.55	0.97

	Rank Correlation			Decile Score Distance		
	Tree−Linear	NN−Linear	NN−Tree	Tree−Linear	NN−Linear	NN−Tree
difference	1.69	1.75	0.06	0.15	0.22	0.07
# of +	26	29	15	26	28	22
fraction of +	0.84	0.94	0.48	0.84	0.90	0.71

U.S.-Trained Models: Summary

	Sharpe Ratio (EW)			Sharpe Ratio (VW)			R^2_{oos}		
	Tree−Linear	NN−Linear	NN−Tree	Tree−Linear	NN−Linear	NN−Tree	Tree−Linear	NN−Linear	NN−Tree
difference	0.41	0.65	0.25	-0.05	0.37	0.42	-1.17	0.01	1.18
# of +	26	30	25	15	27	29	8	17	30
fraction of +	0.84	0.97	0.81	0.48	0.87	0.94	0.26	0.55	0.97

	Rank Correlation			Decile Score Distance		
	Tree−Linear	NN−Linear	NN−Tree	Tree−Linear	NN−Linear	NN−Tree
difference	1.69	1.75	0.06	0.15	0.22	0.07
# of +	26	29	15	26	28	22
fraction of +	0.84	0.94	0.48	0.84	0.90	0.71

Machine learning models, NNs in particular, outperform linear models in forming profitable portfolios and predicting return rankings (but not in terms of R^2, probably because of extreme values)

Market-Specific Models: Summary

		Sharpe Ratio (EW)			Sharpe Ratio (VW)		
		Tree−Linear	NN−Linear	NN−Tree	Tree−Linear	NN−Linear	NN−Tree
All markets	*difference*	0.17	0.60	0.44	-0.21	0.41	0.61
	# of +	19	26	25	12	25	28
	fraction of +	0.59	0.81	0.78	0.38	0.78	0.88

		Rank Correlation			Decile Score Distance		
		Tree−Linear	NN−Linear	NN−Tree	Tree−Linear	NN−Linear	NN−Tree
All markets	*difference*	-0.04	1.08	1.12	0.02	0.20	0.19
	# of +	18	24	23	17	25	23
	fraction of +	0.56	0.75	0.72	0.53	0.78	0.72

Market-Specific Models: Summary

		Sharpe Ratio (EW)			Sharpe Ratio (VW)		
		Tree−Linear	NN−Linear	NN−Tree	Tree−Linear	NN−Linear	NN−Tree
	difference	0.17	0.60	0.44	-0.21	0.41	0.61
All markets	# of +	19	26	25	12	25	28
	fraction of +	0.59	0.81	0.78	0.38	0.78	0.88
		Rank Correlation			**Decile Score Distance**		
		Tree−Linear	NN−Linear	NN−Tree	Tree−Linear	NN−Linear	NN−Tree
	difference	-0.04	1.08	1.12	0.02	0.20	0.19
All markets	# of +	18	24	23	17	25	23
	fraction of +	0.56	0.75	0.72	0.53	0.78	0.72

NNs continue to outperform

Market-Specific Models: Summary

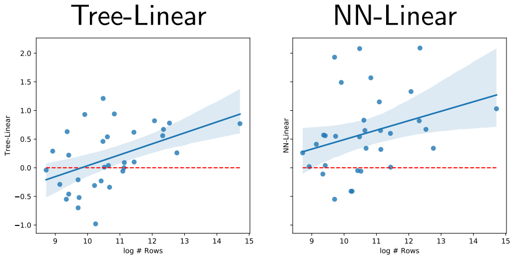
		Sharpe Ratio (EW)			Sharpe Ratio (VW)		
		Tree−Linear	NN−Linear	NN−Tree	Tree−Linear	NN−Linear	NN−Tree
All markets	*difference*	0.17	0.60	0.44	-0.21	0.41	0.61
	# of +	19	26	25	12	25	28
	fraction of +	0.59	0.81	0.78	0.38	0.78	0.88

		Rank Correlation			Decile Score Distance		
		Tree−Linear	NN−Linear	NN−Tree	Tree−Linear	NN−Linear	NN−Tree
All markets	*difference*	-0.04	1.08	1.12	0.02	0.20	0.19
	# of +	18	24	23	17	25	23
	fraction of +	0.56	0.75	0.72	0.53	0.78	0.72

NNs continue to outperform

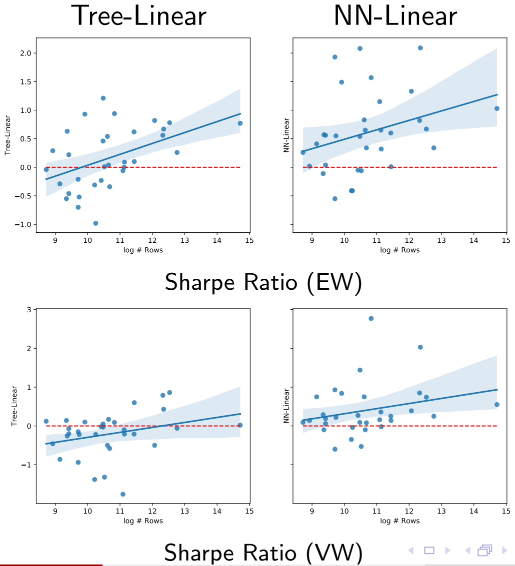
Trees show signs of overfitting, especially in markets where the number of observations is low

Market-Specific Models: Performance vs Log # Obs

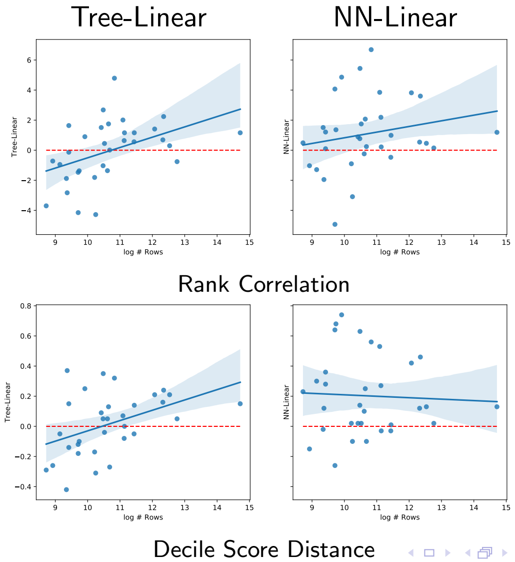


Sharpe Ratio (EW)

Market-Specific Models: Performance vs Log # Obs



Market-Specific Models: Performance vs Log # Obs



Market-Specific Models vs U.S.-Trained Models

		NN1	NN2	NN3	NN4	NN5
Sharpe Ratio (EW)	difference	0.74	0.77	0.75	0.69	0.74
	# of +	26	26	26	24	27
	fraction of +	0.84	0.84	0.84	0.77	0.87
Sharpe Ratio (VW)	difference	0.52	0.46	0.40	0.42	0.52
	# of +	24	23	24	23	23
	fraction of +	0.77	0.74	0.77	0.74	0.74

Market-Specific Models vs U.S.-Trained Models

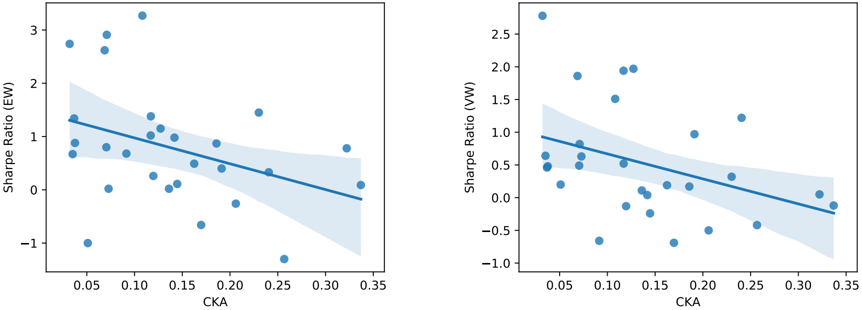
		NN1	NN2	NN3	NN4	NN5
Sharpe Ratio (EW)	difference	0.74	0.77	0.75	0.69	0.74
	# of +	26	26	26	24	27
	fraction of +	0.84	0.84	0.84	0.77	0.87
Sharpe Ratio (VW)	difference	0.52	0.46	0.40	0.42	0.52
	# of +	24	23	24	23	23
	fraction of +	0.77	0.74	0.77	0.74	0.74

Local NN models outperform U.S.-Trained NN models

Market-Specific vs U.S.-Trained Models: CKA Similarity

Centered Kernel Alignment (CKA, Kornblith et al., 2019) Similarity in the last hidden layer

Market-specific NN5 – U.S.-trained NN5



Sharpe Ratio (EW) Sharpe Ratio (VW)

Cross-Market Integration

In each market-specific NN, we add top 10 U.S. characteristic gaps plus their interaction terms with the corresponding local characteristic

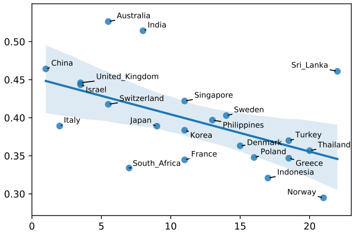
Cross-Market Integration

In each market-specific NN, we add top 10 U.S. characteristic gaps plus their interaction terms with the corresponding local characteristic

Augmented model - Original model (Top 25 markets)

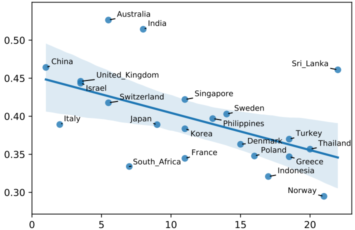
		NN1	NN2	NN3	NN4	NN5	Best NN
Sharpe Ratio (EW)	difference	0.10	0.07	0.12	0.16	0.04	0.12
	# of +	13	12	16	16	13	18
	fraction of +	0.52	0.48	0.64	0.64	0.52	0.72
Sharpe Ratio (VW)	difference	0.07	0.17	0.33	0.31	0.42	0.29
	# of +	15	16	17	21	19	22
	fraction of +	0.60	0.64	0.68	0.84	0.76	0.88

Cross-Market Integration: U.S. Variable Importance

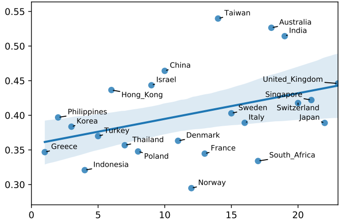


Segmentation (Bekaert et al. (2011))

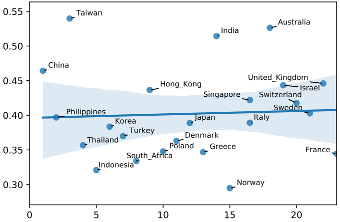
Cross-Market Integration: U.S. Variable Importance



Segmentation (Bekaert et al. (2011))



Economic Integration
(Akbari, Ng, and Solnik (2020))



Financial Integration
(Akbari, Ng, and Solnik (2020))

Conclusion

- In the U.S. and most of the 31 international markets, we obtain high Sharpe Ratios
    - Train and validate models using U.S. data, or separately for each country
    - Neural network models are powerful, suggesting that non-linearity and complex interactions are important
    - Tree models can overfit and underperform linear models, especially when the number of observations is low
- Market-specific models outperform U.S.-trained models
    - Market-specific details are important
- Local NNs can be further improved by adding U.S.-based variables
    - Markets are (partially) integrated