University of Toronto
School of Continuing Studies
3253 - Machine Learning

Can factor investing disrupt the hedge fund industry?

Mirroring "Global Macro" Hedge Funds performance through factor-based predictive modelling and ensemble learning

Presented by:
Gabriel Ignacio
Jason Zhang

Wigglesworth, R. (2018). Can factor investing kill off the hedge fund? *Financial Times*. Retrieved from https://www.ft.com/content/2b3e2eaa-6fe6-11e8-92d3-6c13e5c92914

Hedge Funds

- Wealthy individuals or institutional investors
- Significant diversification benefits
- "Absolute returns" (positive risk-adjusted returns)

<u>Has anyone invested in hedge funds before?</u>

Sources:
Edwards, F. R., & Gaon, S. (2003). Hedge Funds: What Do We Know? *Journal of Applied Corporate Finance*, 15(4), 58-71.
Fung, W., & Hsieh, D. A. (1999). A primer on hedge funds. *Journal of Empirical Finance*, *6*(1999), 309-331.

US Global Macro Hedge Funds

3 trillion USD
CAGR 8%-10%
5-10 years

BRIDGEWATER

14,000 US firms
Top fund uses global macro strat
George Soros - $1billion overnight

Each of the top 35 US hedge funds **>** entire Canadian hedge funds industry (200 hedge funds, 200 billion in AUM)

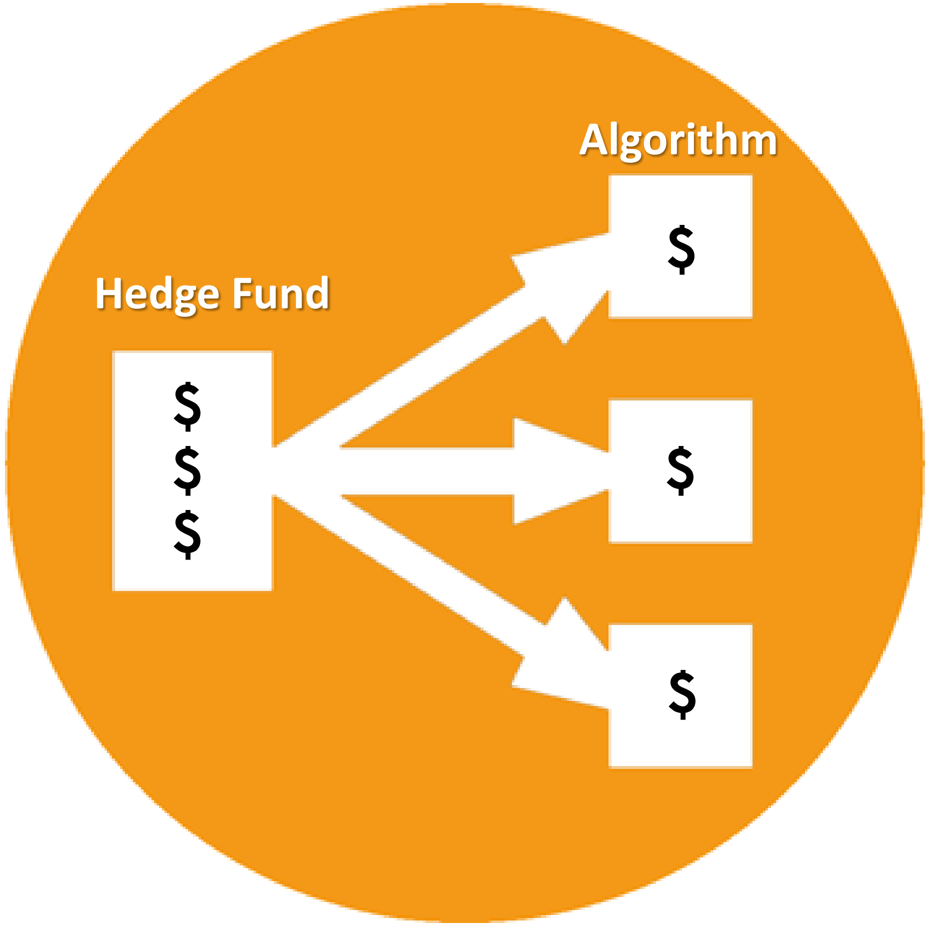
Sources:
Banquier, S. (2019). Overview of Prime Brokerage. Presented at Smith School of Business, Queen's University.
Fung, W., & Hsieh, D. A. (2011). The risk in hedge fund strategies: Theory and evidence from long/short equity hedge funds. *Journal of Empirical Finance, 18*(2011), 547-569.
.

Research Question

Are "Global Macro" Hedge Funds replicable using the factor-based frameworks for non-accredited investors (like you and me)?

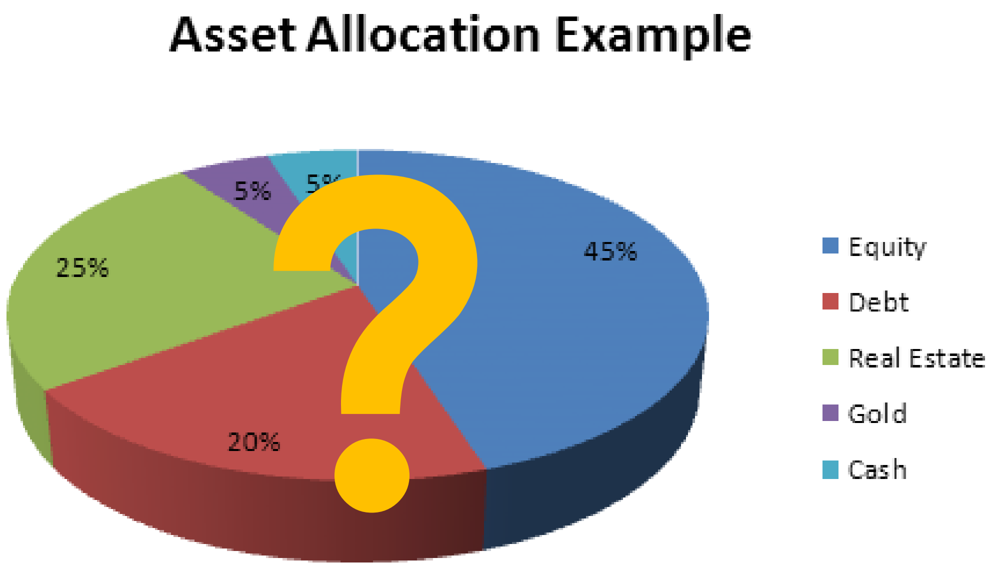
Objectives

- **Dynamic linkages**: can market variables predict GMHF performance?
- **Neutrality**: or do dynamic investment strategies (i.e. manager skill) play a factor?
- **Implication**: systematically replicate results without Hedge Fund fees



Methodology

- Gather and prepare the data – target variables and feature variables
- Predict the target using feature variables using machine learning
- Identify the best factor-based framework to replicate GMHF Performance
- Use feature selection to determine portfolio weightage per feature
- Inform non-accredited investors of index trackers to mirror GMHF returns



How much $ do you allocate per index or asset class to mirror the performance of a GMHF?

Data Exploration

Target: Global Macro Hedge Funds Mean Index Returns (GMHF)

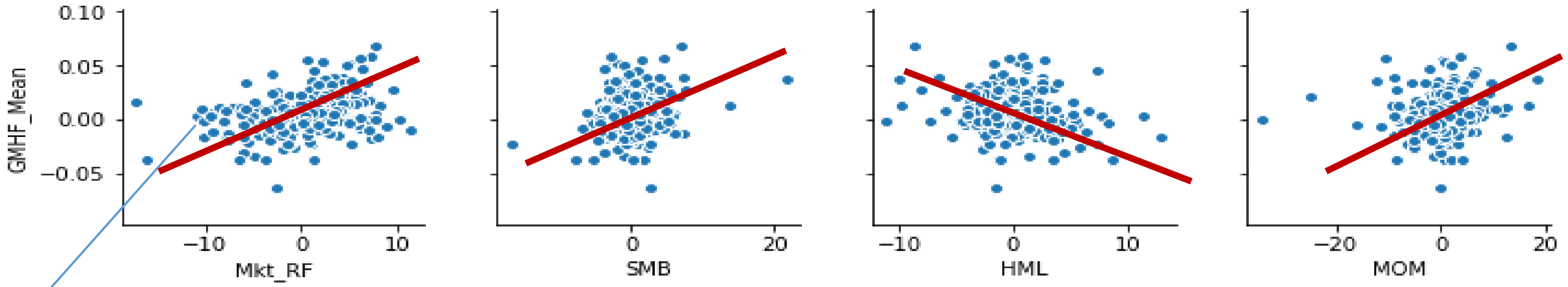
Fama-French-Carhart 4 Factor Model
Market returns (Mkt), size factor (SMB), value factor (HML) and momentum (MOM)

Risk-based Factor Model
Bond Trending following risk factors
Equity-oriented risk factors
Bond-oriented risk factors

Macroeconomic Factor Model
GDP, inflation rate, interest rates, unemployment rates, etc

Observations: 1994-2018, 300 monthly data points
Source: Hedge Fund Research Institute, Ken French Library, David Hsieh Library, US government data

Data Exploration - 4 factor model



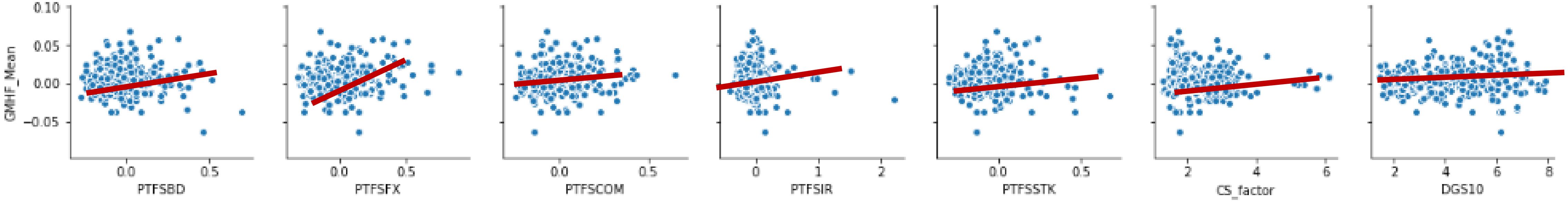
Linear correlation

Target

Features

	GMHF_Mean	Mkt_RF	SMB	HML	MOM
count	300.000000	300.000000	300.000000	300.000000	300.000000
mean	0.005155	0.625467	0.112667	0.149233	0.431233
std	0.017692	4.296355	3.280816	3.026984	4.926031
min	-0.064000	-17.230000	-16.870000	-11.100000	-34.390000
25%	-0.005525	-1.960000	-1.972500	-1.442500	-1.300000
50%	0.003950	1.185000	0.025000	-0.060000	0.480000
75%	0.015100	3.377500	1.912500	1.682500	2.825000
max	0.068200	11.350000	21.710000	12.900000	18.360000

Data Exploration – Risk Based Model

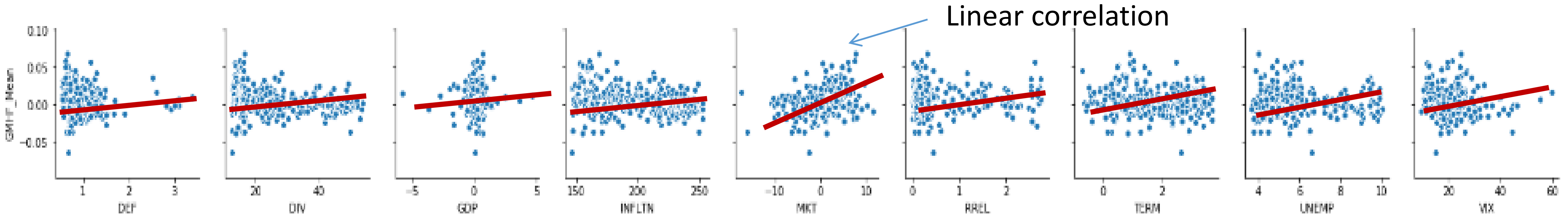


Target

Features

	GMHF_Mean	PTFSBD	PTFSFX	PTFSCOM	PTFSIR	PTFSSTK	CS_factor	DGS10
count	300.000000	300.00000	300.000000	300.000000	300.000000	300.000000	300.000000	300.000000
mean	0.005155	-0.01890	-0.008167	-0.005833	-0.020800	-0.048400	2.421133	4.133867
std	0.017692	0.15202	0.193697	0.140519	0.249809	0.141559	0.786240	1.636767
min	-0.064000	-0.27000	-0.320000	-0.250000	-0.420000	-0.300000	1.300000	1.460000
25%	-0.005525	-0.13000	-0.150000	-0.100000	-0.140000	-0.150000	1.770000	2.667500
50%	0.003950	-0.05000	-0.050000	-0.030000	-0.070000	-0.070000	2.335000	4.095000
75%	0.015100	0.04000	0.082500	0.060000	0.040000	0.020000	2.860000	5.422500
max	0.068200	0.69000	0.900000	0.650000	2.220000	0.670000	6.100000	7.910000

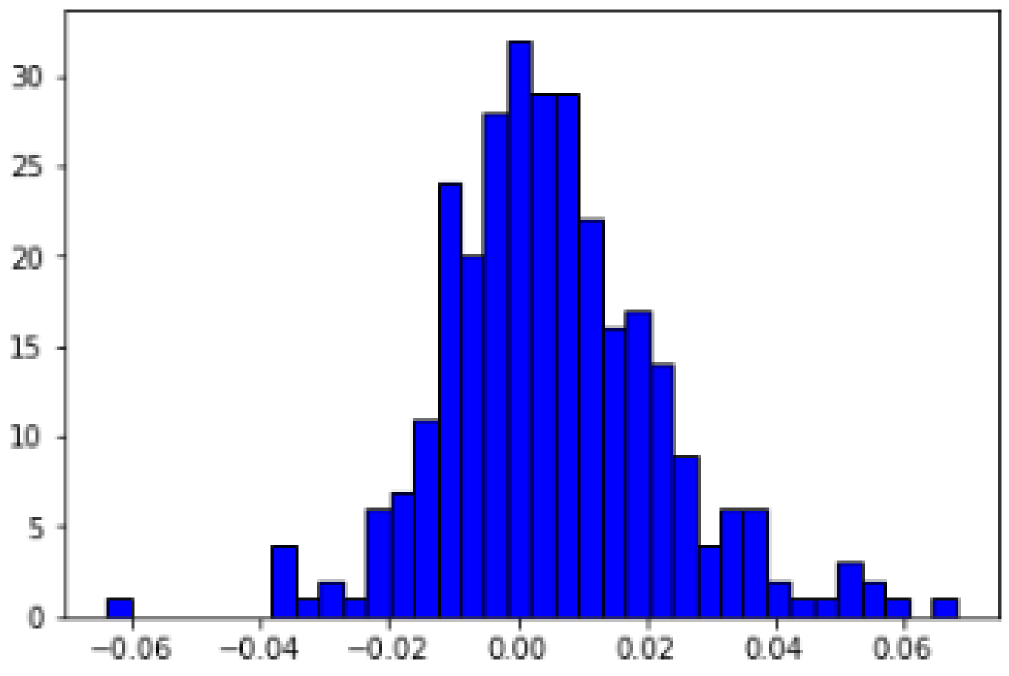
Data Exploration – Macroeconomic Model



	GMHF_Mean	DEF	DIV	GDP	INFLTN	MKT	RREL	TERM	UNEMP	VIX
count	300.000000	300.000000	300.000000	300.000000	300.000000	300.000000	300.000000	300.000000	300.000000	300.000000
mean	0.005155	0.958833	25.117967	0.150367	199.963367	0.625467	0.620500	1.701100	5.788000	19.699133
std	0.017692	0.415017	11.252852	0.720316	31.924226	4.296355	0.764919	1.090485	1.631193	7.668713
min	-0.064000	0.550000	12.620000	-5.800000	146.300000	-17.230000	0.000000	-0.700000	3.700000	9.510000
25%	-0.005525	0.690000	16.035000	-0.040000	170.975000	-1.960000	0.120000	0.882500	4.600000	13.870000
50%	0.003950	0.870000	22.230000	0.190000	201.850000	1.185000	0.280000	1.685000	5.400000	17.970000
75%	0.015100	1.072500	29.727500	0.342500	231.070000	3.377500	0.792500	2.602500	6.125000	23.662500
max	0.068200	3.380000	53.750000	4.710000	252.790000	11.350000	2.760000	3.680000	10.000000	59.890000

Target : Global Macro Hedge Funds Mean Index Returns (GMHF)

GMHF_Mean	1.000000
Mkt_RF	0.343204
MKT	0.343204
PTFSFX	0.218802
SMB	0.203621
RF	0.176255
PTFSCOM	0.167985
DGS10	0.151699
MOM	0.148669
PTFSSTK	0.084910
RREL	0.080355
GDP	0.019099
PTFSBD	0.017395
DEF	-0.045166
PTFSIR	-0.061360
UNEMP	-0.063736
VIX	-0.089781
CS_factor	-0.090315
TERM	-0.093911
CMA	-0.148520
HML	-0.177371
DIV	-0.185010
INFLTN	-0.196122
RMW	-0.259977



Applied models	Hypothesis
regression	How much could the businessmen get in terms of ROI
classification	Could they gain or lose in their GMHF investment

Model Building Process

Feature Selection	Numeric Variables
Fama-French-Carhart	'Mkt_RF','SMB','HML','MOM'
Risk-based	'PTFSBD','PTFSFX','PTFSCOM','PTFSIR','PTFSSTK','DGS10','CS_factor'
Macroeconomic	'DEF','DIV','GDP','INF','MKT','RREL','TERM','UNEMP','VIX'

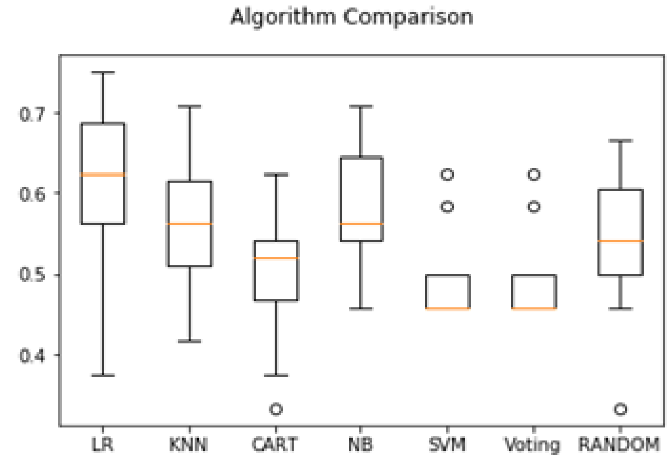
Algorithms Comparison	Regression (RMSE)	Classification (Accuracy)
Linear Algorithms	LinearRegression	LogisticRegression
Non-Linear Algorithms	KNeighborsRegressor, SVR	KNeighborsClassifier, DecisionTreeClassifier, GaussianNB, SVC
Emsemble Algorithms	GradientBoostingRegressor, RandomForestRegressor	VotingClassifier, RandomForestClassifier

Making pipeline	1. Scaling 2. PCA 3. Parameters tuning (GridSearchCV) for the best algorithm

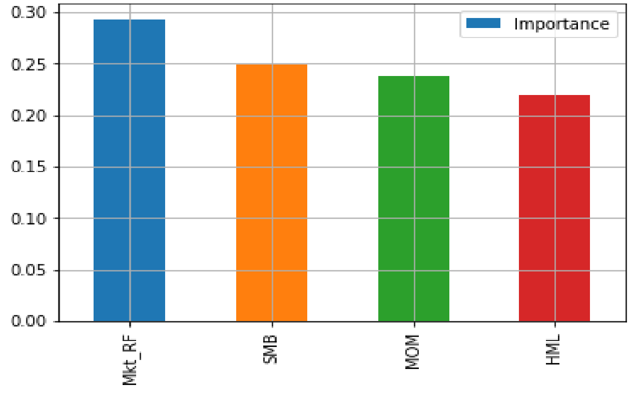
Feature Importance	Use Algorithm RandomForest method best_estimator.feature_importances

Classification

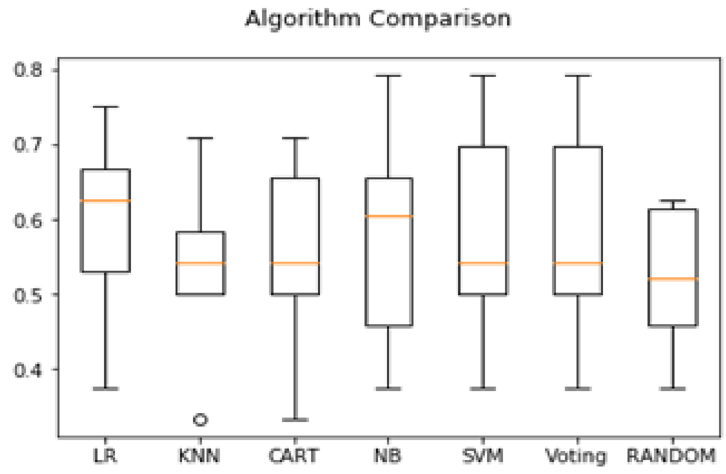
Fama-French-Carhart



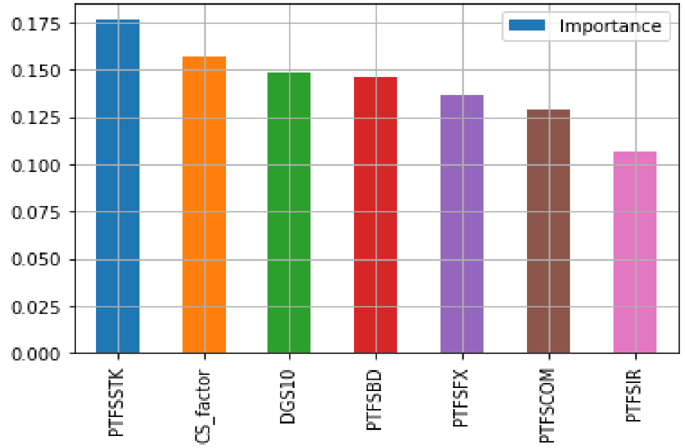

```
Best parameter (CV score=0.637):
{'logistic__C': 0.0001, 'logistic__penalty': 'l2', 'pca__n_components': 4}
 LG Model test Accuracy::  0.7666666666666667
```



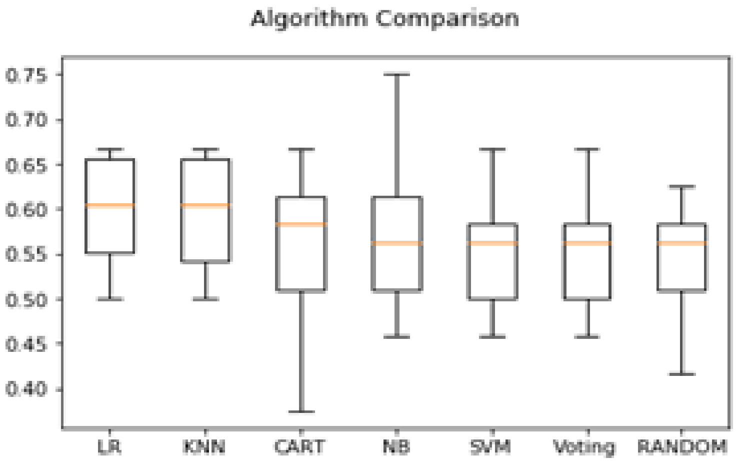
Risk-based



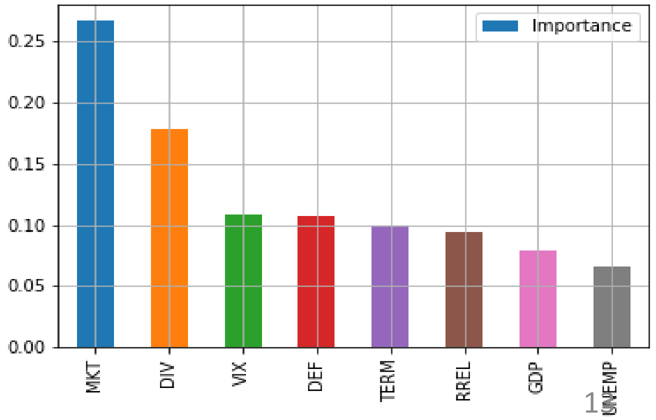

```
Best parameter (CV score=0.625):
{'logistic__C': 35.564803062231285, 'logistic__penalty': 'l2', 'pca__n_components': 2}
 LG Model test Accuracy::  0.6833333333333333
```



Macroeconomic

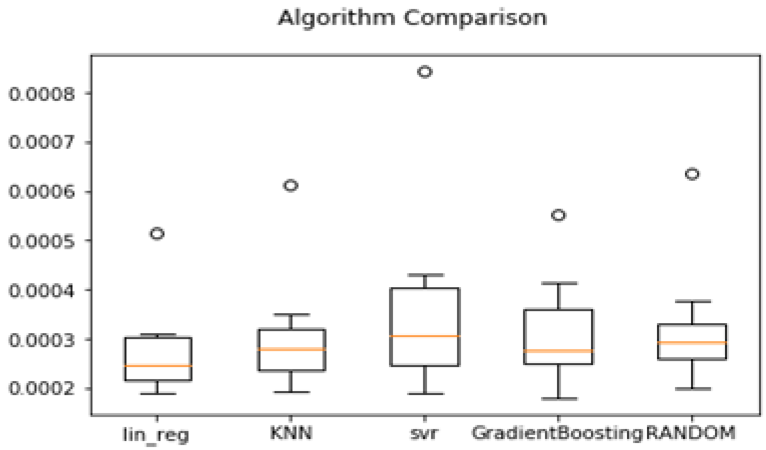



```
Best parameter (CV score=0.646):
{'logistic__C': 0.013257113655901081, 'logistic__penalty': 'l2', 'pca__n_components': 3}
 LG Model test Accuracy::  0.6333333333333333
```

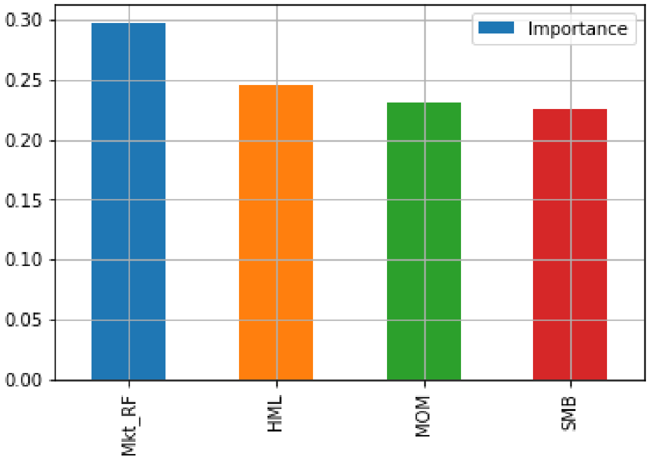


Regression

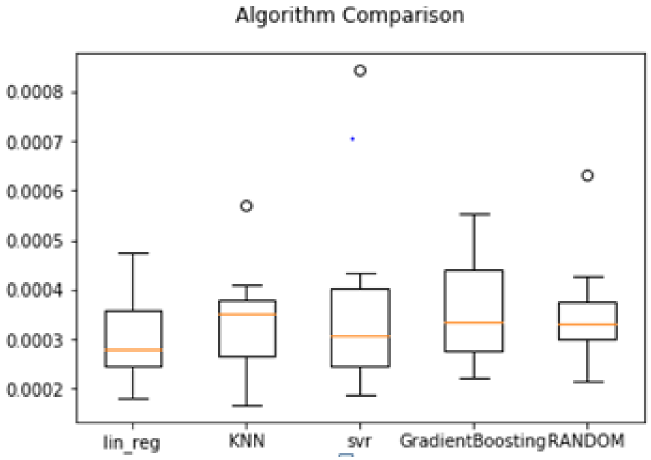
Fama-French-Carhart



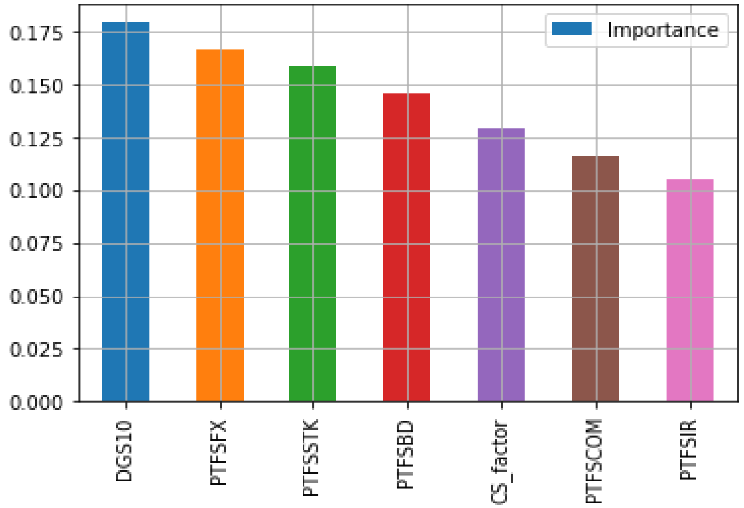

```
Best parameter (CV score=0.074):
{'pca__n_components': 4}
RSME Based On Linear Regression:  0.01536944488278775
```



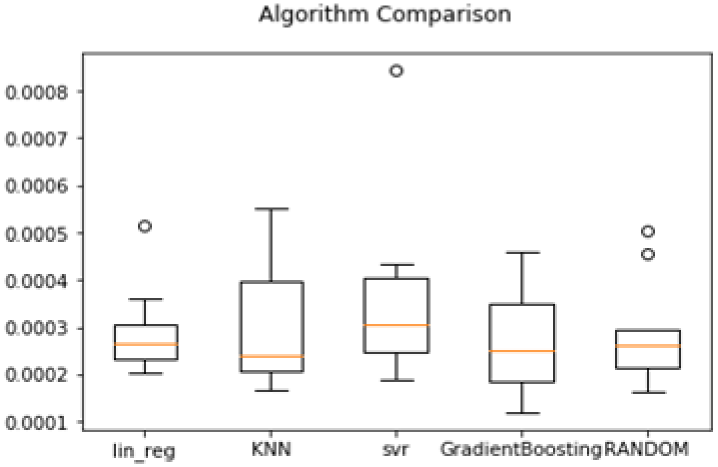
Risk-based



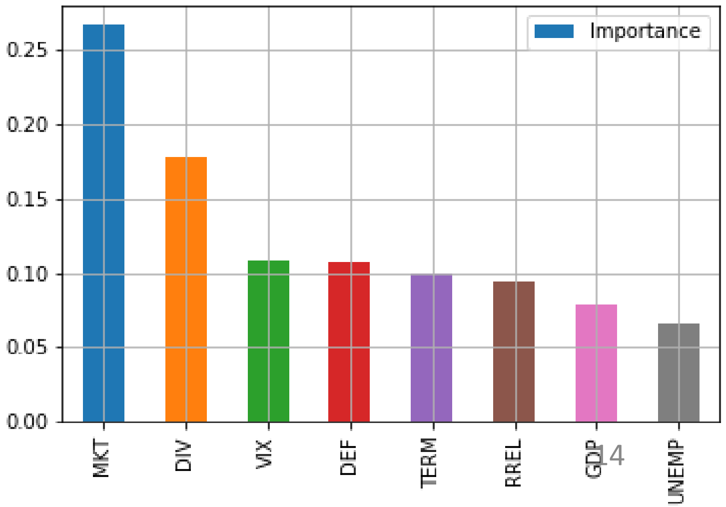

```
Best parameter (CV score=-0.001):
{'pca__n_components': 7}
RSME Based On Linear Regression:  0.017031892214731736
```



Macroeconomic




```
Best parameter (CV score=0.137):
{'GradientBoosting__learning_rate': 0.01, 'GradientBoosting__max_depth': 4, 'GradientBoosting__n_estimators': 100, 'GradientBoosting__subsample': 0.5, 'pca__n_components': 6}
RSME Based On Gradient Boosting:  0.015977497904660194
```



Summary Results: Best Performing Model

- Evaluated each factor model using classification (accuracy) and regression (RMSE)
- The best model is chosen based on accuracy and RMSE

Models	Classification Accuracy	Regression RMSE
Fama-French-Carhart 4-factor model	**77% Logistic Regression**	**0.0153 Linear Regression**
Risk-based factor model	**69% Logistic Regression**	**0.0170 Linear Regression**
Macroeconomic factor model	**63% Logistic Regression**	**0.0159 Gradient Boosting**

Fama-French-Carhart has the best predictive power across the 3 models using logistic regression and Linear Regression. Not surprisingly, AQR also uses the same factor model

Conclusion and Recommendations

- Given models performance, it is possible to predict (and mirror) GMHF performance using the Fama-French-Carhart, risk-based and macroeconomic factor models
- Among the three factor-models, Fama-French-Carhart best explain Global Macro Hedge Funds' performance
- Among all the features (or factors), market performance, foreign exchange look back straddle and small cap stock performance best predict GMHF returns
- Exploiting these indicators might be key to drive alpha
- The model performance could be improved better by doing data engineer work on the dataset. E.g. Data transformation, coefficient analysis, and outlier detection

Appendices


STATES OF AMERICA
A 01035536 B
WASHINGTON, D.C.
10
GLOBAL ASSET MANAGEMENT
LAURENCE KEEN UNIT TRUST MGMT
LAURENTIAN UNIT TST MGMT LTD
MARTIN CURRIE UNIT
ONE CENT

Frameworks Used

Frameworks	Theoretical	Empirical
Fama-French-Carhart Four Factor Model	Market, size, value and momentum explain stock performance	$(R_i - R_f) = \alpha_i + \beta_{MKT}(R_{MKT} - R_f) + \beta_{SMB}(SMB) + \beta_{HML}(HML) + \beta_{UMD}(UMD) + e_i$
Trend Following Factors	**Trend-Following Risk Factors (3):** -Bond Trend-Following Factor -Currency Trend-Following Factor -Commodity Trend-Following Factor **Equity-oriented Risk Factors (2):** -Equity Market Factor -The Size Spread Factor **Bond-oriented Risk Factors (2):** -The Bond Market Factor -The Credit Spread Factor	$(R_i - R_f) = \alpha_i + \beta_1(PTFSBD) + \beta_2(PTFSFX) + \beta_3(PTFSCOM) + \beta_4(PTFSIR) + \beta_5(PTFSTK) + \beta_6(DGS10) + \beta_7(CS\_factor) + e_i$
Macroeconomic Factors	Default spread, aggregate dividend yield, the growth rate of real Gross Domestic Product (GDP) per capita, inflation rate, equity market index, term spread, short-term interest rate changes, and unemployment rate	$(R_i - R_f) = \alpha_i + \beta_1(DEF) + \beta_2(DIV) + \beta_3(GDP) + \beta_4(INFLTN) + \beta_5(MKT) + \beta_6(RREL) + \beta_7(TERM) + \beta_7(UNEMP) + e_i$
Volatility	Volatility index	$(R_i - R_f) = \alpha_i + \beta_1(VIX) + e_i$

Data Preparation

Fama-French-Carhart 4 Factor Model (combination of market risks)

Variables	Description	# of observations	Timeframe	Corrected to	Source
Mkt-RF	Rm-Rf, the excess return on the market	1,111 monthly data points	July 1926 to Jan 2019	Jan 1994 to Dec 2018	Kenneth R. French - Data Library
SMB	Size factor: small stocks relative to large stocks	1,111 monthly data points	July 1926 to Jan 2019	Jan 1994 to Dec 2018	Kenneth R. French - Data Library
HML	Value factor: value stocks relative to growth stocks	1,111 monthly data points	July 1926 to Jan 2019	Jan 1994 to Dec 2018	Kenneth R. French - Data Library
MOM	Tendency for the stock price to continue rising if it is going up and to continue declining if it is going down	1,105 data points Monthly	Jan 1927 to Jan 2019	Jan 1994 to Dec 2018	Kenneth R. French - Data Library

Data Preparation

Trend-Following factors under Fung and Hsieh (directional / trend risks)

Variables		Description	# of observations	Timeframe	Corrected to	Source
PTFSBD		Return of PTFS Bond lookback straddle	301 data points	Jan 1, 1994 to Jan 1, 2019	Jan 1994 to Dec 2018	David A. Hsieh's Data Library: Hedge Fund Risk Factors
PTFSFX		Return of PTFS Currency Lookback Straddle	301 data points	Jan 1, 1994 to Jan 1, 2019	Jan 1994 to Dec 2018	David A. Hsieh's Data Library:
PTFSCOM		Return of PTFS Commodity Lookback Straddle	301 data points	Jan 1, 1994 to Jan 1, 2019	Jan 1994 to Dec 2018	Hedge Fund Risk Factors
PTFSIR		Return of PTFS Short Term Interest Rate Lookback Straddle	301 data points	Jan 1, 1994 to Jan 1, 2019	Jan 1994 to Dec 2018	David A. Hsieh's Data Library:
PTFSSTK		Return of PTFS Stock Index Lookback Straddle	301 data points	Jan 1, 1994 to Jan 1, 2019	Jan 1994 to Dec 2018	Hedge Fund Risk Factors
The Credit Spread Factor	DBAA	Moody's Baa yield	398 data points 33 years Monthly obs	Jan 1, 1986 to Jan 1, 2019	Jan 1994 to Dec 2018	Federal Reserve Bank of St. Louis
	DGS10	10-year treasury constant maturity yield	686 data points	Jan 1, 1962 to Feb 1, 2019	Jan 1994 to Dec 2018	Federal Reserve Bank of St. Louis
The Bond Market Factor	DGS10	The monthly change in the 10-year treasury constant maturity yield	686 data points	Jan 1, 1962 to Feb 1, 2019	Jan 1994 to Dec 2018	Federal Reserve Bank of St. Louis

Data Preparation

Macroeconomic variables by Bali, Brown and Caglayan (2014)

Variables	Description	# of observations	Timeframe	Corrected to	Source
DEF	default spread measured as the difference between yields on BAA-rated and AAA-rated corporate bonds	1,202 data points 100 years Monthly obs	Jan 1, 1919 to Feb 1, 2019	Jan 1994 to Dec 2018	Federal Reserve Bank of St. Louis
DIV	aggregate dividend yield on the Standard&Poor's (S&P)500Index	1,768 data points	Jan 1871 to Dec 2018	Jan 1994 to Dec 2018	Robert Shiller's online data library
GDP	U.S. monthly Growth rate of real GDP per capita	719 data points	Feb 1, 1959 to Dec 1, 2018	Jan 1994 to Dec 2018	Federal Reserve Bank of St. Louis
INF	INF: monthly inflation rate based on the U.S. consumer price index	865 data points	Jan 1, 1947 to Jan 1, 2019	Jan 1994 to Dec 2018	Robert Shiller's online data library
MKT	MKT: excess return on the value-weighted NYSE/Amex/ Nasdaq (CRSP) equity market index	666 monthly data points	July 1963 to Dec 2018	Jan 1994 to Dec 2018	Kenneth R. French - Data Library
RREL	RREL: relative T-bill rate, defined as the difference between the three-month T-billrateandits12-monthbackwardmoving average	1,022 monthly data points	Jan 1, 1934 to Feb 1, 2019	Jan 1994 to Dec 2018	Federal Reserve Bank of St. Louis
TERM	TERM: term spread measured as the difference Between yields on ten-year and three-month Treasury securities	446 monthly data points	Jan 1, 1982 to Feb 1, 2019	Jan 1994 to Dec 2018	Federal Reserve Bank of St. Louis
UNEMP	UNEMP: the U.S. monthly unemployment rate defined as the number of unemployed as a Percentage of the labor force	854 monthly data points	Jan 1, 1948 to Feb 1, 2019	Jan 1994 to Dec 2018	US Bureau of Labor and Employment Statistics

Data Preparation

Volatility Index (market risks)

Variables	Description	# of observations	Timeframe	Corrected to	Source
VVIX Index	CBOE simple proxy for uncertainty, calculated uncertainty betas for individual stocks using the volatility of their implied option volatilities	344 data points Monthly obs	June 30, 1986 to June 29, 2018	Jan 1994 to Dec 2018	Chicago Board Options Exchange (CBOE®) website

Global Macro Hedge Fund Index (Dependent Variable Y)

Variables	Description	# of observations	Timeframe	Corrected to	Source
HFRI GMHF Index	Index mean return on Global Macro Hedge Funds in the US	300 data points	Jan 1, 1994 to Dec 1, 2018	Jan 1994 to Dec 2018	Hedge Fund Research Index

References

Bali, T. G., Brown, S. J., & Caglayan, M. O. (2014). Macroeconomic risk and hedge fund returns. *Journal of Financial Economics, 114*(2014), 1–19.

Banquier, S. (2019). Overview of Prime Brokerage. Presented at Smith School of Business, Queen's University.

Board of Governors of the Federal Reserve System (US) (2019). Total Consumer Credit Owned and Securitized, Outstanding [TOTALSL], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/TOTALSL.

Carhart, M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance, 52*(1997), 1-26.

Chawla, R. K. , & Uppal, S. Household Debt in Canada. Retrieved from https://www150.statcan.gc.ca/n1/pub/75-001-x/2012002/article/11636-eng.htm.

Edwards, F. R., & Gaon, S. (2003). Hedge Funds: What Do We Know? *Journal of Applied Corporate Finance*, 15(4), 58-71.

El Kalak, I., Azevedob, A., & Hudson, R. (2016). Reviewing the hedge funds literature II: Hedge funds' returns and risk management characteristics. *International Review of Financial Analysis 48*(2016) 55-66.

Fama, E. F., & French, K. R. (2015). A Five-Factor Asset Pricing Model. *Journal of Financial Economics, 116*(2015), 1–22.

French, K. (2018). Fama-French 5 Factors (2x3). Retrieved from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_5_factors_2x3.html.

Fung, W., & Hsieh, D. A. (1999). A primer on hedge funds. *Journal of Empirical Finance*, *6*(1999), 309-331.

Fung, W., & Hsieh, D. A. (2011). The risk in hedge fund strategies: Theory and evidence from long/short equity hedge funds. *Journal of Empirical Finance, 18*(2011), 547-569.

Grant, T. (2019). Alternative Investing. Retrieved from http://v1.theglobeandmail.com/partners/free/srinvest/art7_hedge.html.

Hedge Fund Research Index. (2019). HFRI® Indices Performance Tables. Retrieved from https://www.hedgefundresearch.com/family-indices/hfri

References

Hsieh, D. A. (2019). David A. Hsieh's Data Library: Hedge Fund Risk Factors. Retrieved from https://faculty.fuqua.duke.edu/~dah7/HFRFData.htm

Isenberg School of Management. (2018). CISDM Equity Long/Short Index (xlsx). Retrieved from https://www.isenberg.umass.edu/resources/downloads/download-cisdm-equity-longshort-index.

Ji, X., Martin, J., & Yao, Y. (2017). Macroeconomic risk and seasonality in momentum profits. *Journal of Financial Markets 36*(2017)76–90

Krause, T.A. (2019). Hedge fund returns and uncertainty. *North American Journal of Economics and Finance 47* (2019) 597–601.

L'Her, J. F., Masmoudib, T., & Suret, J. M. Evidence to support the four-factor pricing model from the Canadian stock market. *International Financial Markets, Institutions and Money, 14* (2004) 313–328.

Namvar, E., Phillips, B., & Pukthuanthong, K., Raghavendra, P. (2016). Do hedge funds dynamically manage systematic risk? *Journal of Banking & Finance, 64* (2016) 1–15.

Panopoulou, E., & Vrontos, S. (2015).Hedge fund return predictability: To combine forecasts or combine information. *Journal of Banking & Finance, 56* (2015) 103–122.

Patton, A. J. (2009). Are "Market Neutral" Hedge Funds Really Market Neutral? *Review of Financial Studies, 22*(7), 2495-2530.

Racicot, F.E., & Theoret, R. Multi-moment risk, hedging strategies, & the business cycle. *International Review of Economics and Finance 58*(2018) 637–675.

Slavutskaya, A. (2013). Short-term hedge fund performance. *Journal of Banking & Finance, 37* (2013), 4404–4431.

Statisa. (2017). Assets of the Hedge Funds Worldwide. Retrieved from https://www.statista.com/statistics/271771/assets-of-the-hedge-funds-worldwide/.

Stafylas, D., Anderson, K., & Uddin, M. (2017). Recent advances in explaining hedge fund returns: Implicit factors and exposures. *Global Finance Journal, 33*(2017) 69–87

Wigglesworth, R. (2018). Can factor investing kill off the hedge fund? *Financial Times*. Retrieved from https://www.ft.com/content/2b3e2eaa-6fe6-11e8-92d3-6c13e5c92914