Explaining 2020 Stock Returns with Different Industry Classification Systems

•••

GICS vs NAICS vs SIC

Task

Which Industry Classification System can best explain and predict stock returns for:

1. Initial pandemic downturn (Jan 2020 - Mar 2020)
2. Subsequent recovery (Apr 2020 - Dec 2020)

Data:

- 2020 Stock Returns for 2700 companies in the Russell 3000 index
- Codes for Industry Classification Systems

Classification Systems

Comparing Three Classification Systems:

- Global Industry Classification Standard (GICS)
- North American Industry Classification System (NAICS)
- Standard Industrial Classification (SIC)

Four Levels of Specification:

- Most broad - (GIC Sector, NAICS Sector, SIC Division)
- (GIC Industry Group, NAICS Sub-Sector , SIC Major Group)
- (GIC Industry, NAICS Industry Group, SIC Industry Group)
- Most Specific - (GIC Sub-Industry, NAICS Industry, SIC Industry)

Classification Systems

NAICS CODE

Main Sector	51	Information
Subsector	511	Publishing industries
Industry Group	5111	Newspaper, book and directory publishers
Industry	51119	Other publishers
U.S. Industry	511191	Greeting card publishers

Data Pre-Processing

1. Linking codes to text descriptions with lookup tables:

GIC Sector	
Code	Name
10	Energy
15	Materials
20	Industrials
25	Consumer Discretionary
30	Consumer Staples
35	Health Care
40	Financials
45	Information Technology
50	Telecommunication Services
55	Utilities
60	Real Estate

2. Merging descriptions with codes from Russell 3000 dataset:

Number of unique classifications in dataset □

	GICS	NAICS	SIC
0	11	18	10
1	24	80	67
2	66	230	198
3	146	386	215

Number of unique possible classifications □

	GICS	NAICS	SIC
0	11	20	10
1	25	113	83
2	68	398	416
3	157	973	1007

Regression Results

- GICS
- NAICS
- SIC

Num_Unique	Adj_rsqrd_Early	Adj_rsqrd_Late	Sig_Coef_Early	Sig_Coef_Late
11	0.115457	0.092359	11.0	10.0
24	0.122945	0.102984	23.0	22.0
66	0.147931	0.136639	56.0	50.0
146	0.157788	0.183491	112.0	84.0

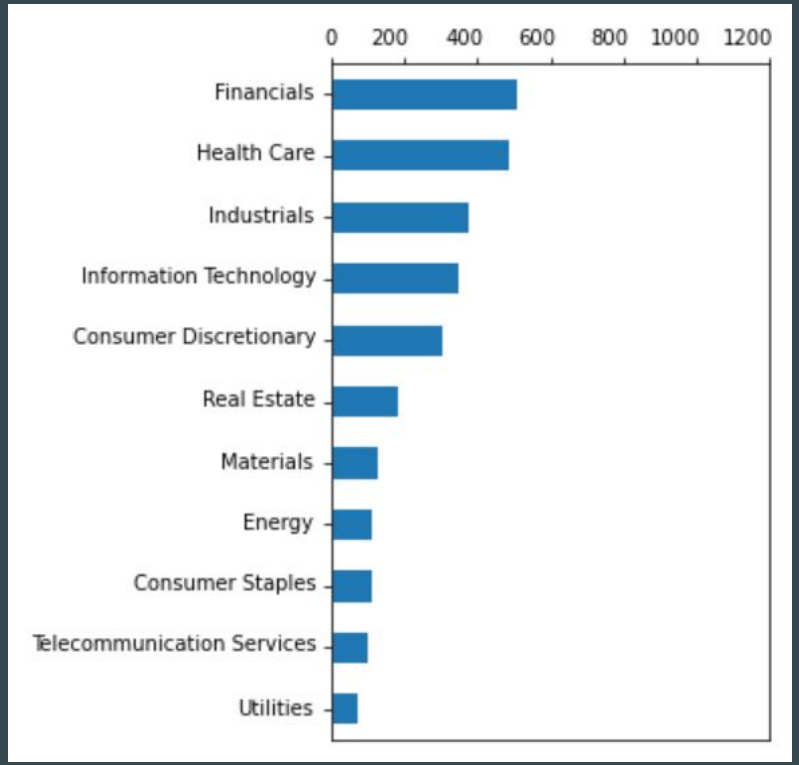
Num_Unique	Adj_rsqrd_Early	Adj_rsqrd_Late	Sig_Coef_Early	Sig_Coef_Late
18	0.065346	0.069405	17.0	14.0
80	0.118474	0.123370	66.0	58.0
230	0.113097	0.149253	132.0	104.0
386	0.088525	0.170758	171.0	115.0

Num_Unique	Adj_rsqrd_Early	Adj_rsqrd_Late	Sig_Coef_Early	Sig_Coef_Late
10	0.051015	0.064907	8.0	8.0
67	0.106131	0.114422	56.0	48.0
198	0.109120	0.115062	128.0	87.0
215	0.098662	0.136705	96.0	75.0

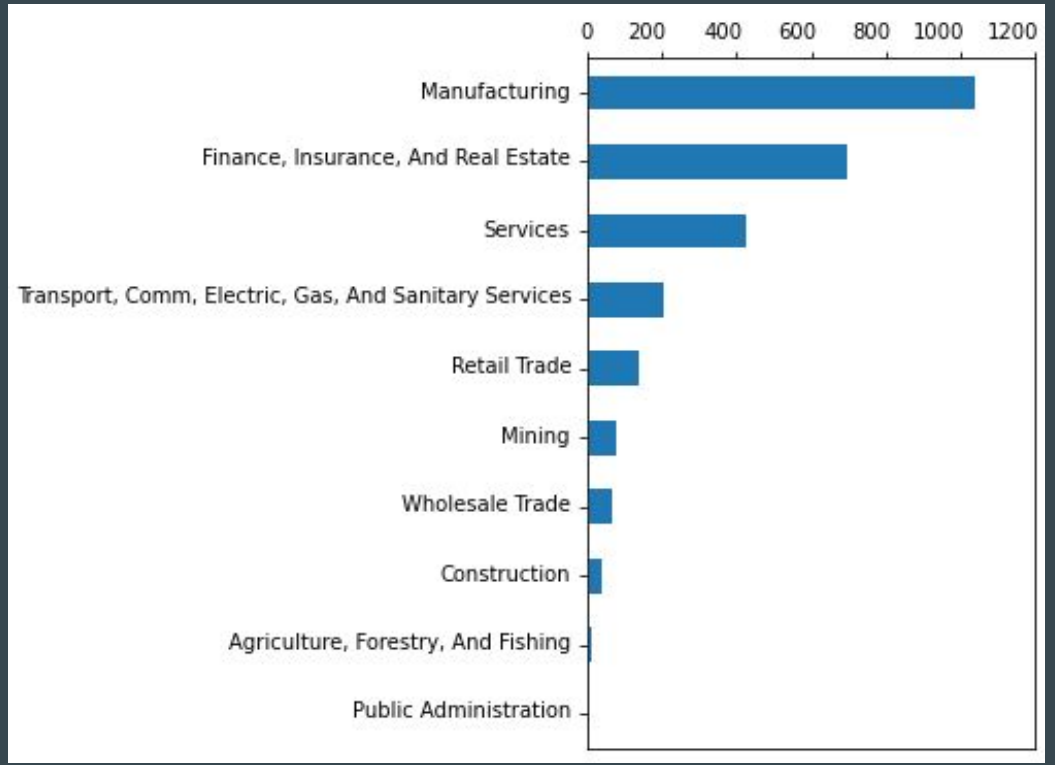
- GICS outperforms other two systems at nearly every level
- More specific indicators generally explain more variation than broad indicators

Number of Companies in Dataset - by system

GICS Sector



SIC Division



- GICS indicators are much more balanced than SIC

Broad vs Specific Indicators - GICS

- When more specificity helps:

Industry	Regression Coefficient
35_Health Care	**-0.139677**
35102010_Health Care Distributors	**0.029810**
35102020_Health Care Facilities	**-0.350752**

- More specific indicators can account for differences in sub-industries

Broad vs Specific Indicators - GICS

- When more general indicators help:
  - Out-of-sample prediction

GIC Level	Cross-Val RMSE
Sector	**0.285**
Group	**0.284**
Industry	**0.296**

- More specific indicators can have trouble generalizing when number of companies in an industry is small

Conclusion

- Industry Matters
  - Differences in return outcomes can be partially explained by a company's industry
- Classification System Matters
  - GICS indicators were more balanced and explained more variation in return outcomes than NAICS or SIC
- Level of Specificity Matters
  - What level is best? – Depends on situation/problem