Innovation and Industry Selection

Vinesh Jha
ExtractAlpha
SmartSpace 3F, Suite 8
Cyberport 3
100 Cyberport Road
Hong Kong SAR
vinesh@extractalpha.com

This draft: 28 February 2019

Vinesh Jha is the Chief Executive Officer of ExtractAlpha Research Limited

Innovation and Industry Selection

28 February 2019

ABSTRACT

We use a novel dataset of company-level innovation measures to identify the most innovative industries based on counts of their applications for foreign worker visas and their patent applications and grants. We are able to build portfolios which overweight these innovative industries and which generate economically significant excess returns, especially in the 2013-2018 period, with low turnover. The results do not appear to be fully explained by risk factors, and the same innovation measures do not predict returns at the single stock level.

Keywords: innovative efficiency, asset allocation

JEL codes: O3, G12, G14

Valuing a company's intangibles is a difficult and uncertain task which requires analysis of information beyond standard public financial filings. As a result, companies in innovation-led sectors are particularly hard to value, and company-level innovation metrics are often noisy when it comes to forecasting either operating performance or market outcomes.

Although there is some academic evidence that innovation, as measured by R&D expenditures, patents, or hiring for example, leads to outperformance (Chan et al, 2001; Hirshleifer et al, 2013), that evidence is somewhat limited. The findings could be underwhelming for several reasons. A small cap stock may "punch above its weight" in terms of its patent activity. Over-patenting could take place, with certain companies flooding the patent office with applications, resulting in a U-shaped return profile by innovation intensity (Bassa Mama, 2018). Further, the simple existence of a patent does not imply that that patent has or will have value to the assignee.

In addition, finding the "correct" company-level scalar for innovation metrics such as the number of patents applied for or granted can be difficult. Scaling by market capitalization can induce an unwanted value bias at the stock level; scaling by total assets will typically unduly punish large but innovative companies; and scaling by R&D expenditures may punish exactly those companies which are investing up front in innovation.

In this research note we propose an alternate application of innovation metrics, namely using innovation intensity to predict the relative returns of *industries*. Using industry aggregates removes many of the scalar issues and also allows us to identify which industries are generally innovating, without requiring us to attach the innovation to a particular company's outcome. For example, small numbers of knowledge workers across many companies in an industry may indicate an innovation trend but may not be as visible for any single company.

I. Data

Most prior research uses Research and Development expenditures from traditional databases such as Compustat to measure innovation. By contrast, we focus on nontraditional data not derived from company financials. To measure innovation, we use ExtractAlpha's ESGEvents Library, which is a database of company-level interactions with a number of different government bodies and regulators, including the Consumer Financial Protection Bureau (CFPB), The Environmental Protection Agency (EPA), the Occupational Health and Safety Administration (OSHA), the Consumer Product Safety Commission (CPSC), the U.S. Senate, the Federal Election Commission (FEC), the department of Labor (DOL), the U.S. Treasury Bureau of the Fiscal Service, and the U.S. Patent and Trademark Office (USPTO). The ESGEvents Library leverages ExtractAlpha's proprietary data collection and name matching algorithm to match company events in these databases to publicly traded securities. The ESGEvents Library dataset is collected at a monthly frequency by the midpoint of each month. For each government data source, company names and event dates are collected. Company names are then matched to FactSet's historical database of security names, using a four-step algorithm:

1) We create hard-coded special cases for large companies with names that are often unconventionally spelled or have synonyms or common acronyms, such as IBM, which is often written out in full as International Business Machines
2) We manipulate the names in both the FactSet list and the government data list by removing punctuation and capitalization, collapsing abbreviations (so that, for example, INTL and INTERNATIONAL are considered equivalently), and removing terms which often won't appear in the source file's company name set, such as LTD, CORPORATION, and so on
3) At this point, many of the parsed and cleaned company names will match between the two data sets
4) We then take any names which are still unmatched and consider a variety of fuzzy matching criteria:
    a. Is there one or more word in a given master/source name pair which is identical?
    b. If so, is it a rare word like VIACOM or a common word like TECHNOLOGIES?
    c. Are the lengths of the master name and source name similar?
    d. Are there words which are similar, but not identical, in a given master/source name pair?
    e. Is the master name similar overall to the source name in spelling, and does it start with the same letters?

The last two steps are meant to address misspellings and slight variations in company names, and involve employing an "edit distance" algorithm which assigns costs to various letter-changing steps required to change one string into another string. This ensures that large but frequently-misspelled company names such as Weyerhaeuser are still present in the database.

The parameters of the above procedure were tuned to minimize false positives and false negatives, but there will still be "matchable" records which go unmatched, and incorrect matches which end up in the dataset. The dataset does not include data from subsidiaries of publicly traded companies.

We then have the following data points matched to publicly traded securities, representing three of the eleven data sets with the ESGEvents Library:

- Department of Labor (DOL)

- Number of total workers for which the company has applied for H1B Visas in the prior year
  - Number of Permanent H1B Visas for which the company has applied in the prior year
- U.S. Patent and Trademark Office (USPTO)
  - Number of company patent applications in the prior year
  - Number of patents granted to the company in the prior year

For each of these, we sample monthly and look both at the one-year level and the year-over-year change. Note that because the Department of Labor updates their visa application data only quarterly, we apply a 180 calendar day lag from the date of the application, whereas the U.S. Patent and Trademark Office data is updated weekly and therefore a 15 day lag is applied. We restrict our attention to stocks in ExtractAlpha's U.S. research universe, which consists of stocks with market capitalizations of $100 million or more, $1 million average daily trading volume or more, and a nominal price of $4 or greater.

At the beginning of each calendar month, we take each of our four measures, and their year over year changes, totaling eight metrics at the company level. Because we are interested in industry level rollups, we need to aggregate to the industry level. We start with industries from FactSet's 130-level industry classification scheme. These industries are overly granular for our purposes, so we accumulate them into 45 coarser industries. Next, for each metric, we take its sum for each industry. We also sum the market caps from the prior month end for every stock in that industry, and then scale the summed metrics by the summed market cap. This gives us eight measures of innovation per unit market cap per industry. We then rank the innovation measures across industries each month to put them on a zero to one scale:

$$IMRANK(I) = Rank\left(\frac{\sum_{s \in I} IM(s)}{\sum_{s \in I} MCAP(s)}\right)$$

Where

IM(s) is one of our four our innovation measure for each stock, or its YoY change,
MCAP(s) is the market cap for each stock,
And stock *s* is in industry *I*.

II. Portfolios

We can now create long-only portfolios of industries which are tilted more heavily towards the more innovative industries. We start with a benchmark weight for each industry, simply based on market capitalization. We then linearly adjust this weight so that the lowest-innovation industry in a given month gets zero weight, a median-innovation industry gets its benchmark weight, and the highest-innovation industry gets double its benchmark weight. We then adjust the tilted weights so that they sum to 1.

Next, for each of our eight metrics, we compare the tilted portfolio to the benchmark portfolio via its annualized excess return and Information Ratio (IR), where IR is the annualized excess return divided by the annualized standard deviation of excess returns. For this exercise, we ignore transaction costs; a relatively inexpensive version of this monthly rebalanced portfolio could be constructed using low-fee ETFs.

Our initial in-sample analysis consists of months from 2003 to 2015. We later examine an out-of-sample period from 2016 through November 2018.

Factor		**Excess Return**	**Information Ratio**
Level	H1B visa	0.45%	0.17
	Permanent visa	0.39%	0.14
	Patent application	0.46%	0.28
	Patent grant	0.53%	0.30
YoY change	H1B visa	0.29%	0.20
	Permanent visa	-0.78%	(0.47)
	Patent application	-0.02%	(0.02)
	Patent grant	0.61%	0.50

Here we see that most of the factor tilts lead to positive excess returns. Year over year change in permanent visas is the major exception. It's possible that this is because the visa applications are essentially *already* a change variable: change in the number of expected employees.

Next, we create a composite Innovation Score by simply averaging the ranked scores across six of the eight factors, excluding year over year changes in both types of visa applications. Note that the composite score is not likely to perform as well as the sum of its underlying factors, because they are correlated. The visa scores are 90% correlated with each other, the patent scores are 95% correlated with each other, and the visa scores are 60% correlated to the patent scores – indicating that all of these metrics capture different aspects of the same underlying innovation characteristic.

The composite score gets us an in-sample annual excess return of 47 basis points with an Information Ratio of 0.21. These numbers are modest, but so are our industry tilts relative to the benchmark. A more aggressive utilization of these factors could lead to stronger outperformance.

We then apply this composite Innovation measure to the full data sample, where it exhibits performance as shown in Exhibit 1.

Exhibit 1. Performance of industry tilt portfolios

		In Sample (2003-2015)		Out of sample (2016-2018)		Full sample (2003-2018)	
Factor		**Excess Return**	**Information Ratio**	**Excess Return**	**Information Ratio**	**Excess Return**	**Information Ratio**
Level	H1B visa	0.45%	0.17	1.86%	0.66	0.71%	0.27
	Permanent visa	0.39%	0.14	2.12%	0.76	0.72%	0.26
	Patent application	0.46%	0.28	1.32%	0.68	0.62%	0.36
	Patent grant	0.53%	0.30	1.17%	0.60	0.65%	0.37
YoY change	H1B visa	0.29%	0.20	0.89%	0.41	0.40%	0.25
	Permanent visa	-0.78%	(0.47)	0.25%	0.15	-0.58%	(0.35)
	Patent application	-0.02%	(0.02)	1.49%	1.02	0.26%	0.19
	Patent grant	0.61%	0.50	1.21%	1.22	0.73%	0.61
Composite Innovation score		**0.47%**	**0.21**	**2.00%**	**0.87**	**0.75%**	**0.34**

Here we can see that our results hold up out of sample, and that innovative industries have outperformed significantly since 2013, though prior to about 2009 there was little difference between more- and less-innovative industries.

III. Robustness checks

As a robustness check, we try restricting to just sectors which are innovation-led, and remove stocks which are categorized in the Energy, Finance, Industrials, Materials, and Miscellaneous sectors. After applying this filter, our full sample excess return is 51 basis points with an Information Ratio of 0.23, slightly weaker than our overall results – which indicates that innovation is important even outside of these sectors.

We may wish to know whether our composite Innovation Score is redundant with standard risk factors. The cross-sectional correlations between our score and various risk factors, aggregated from the stock up to the industry level, is shown in Exhibit 2.

Exhibit 2. Correlation between composite Innovation Score and common risk factors

	value	momentum	volatility	growth	leverage	yield	size
Correlation	(0.30)	0.02	0.18	0.07	(0.27)	(0.37)	(0.19)

Where the risk factors are defined using the following descriptors, which are then standardized and normalized cross-sectionally within our research universe:

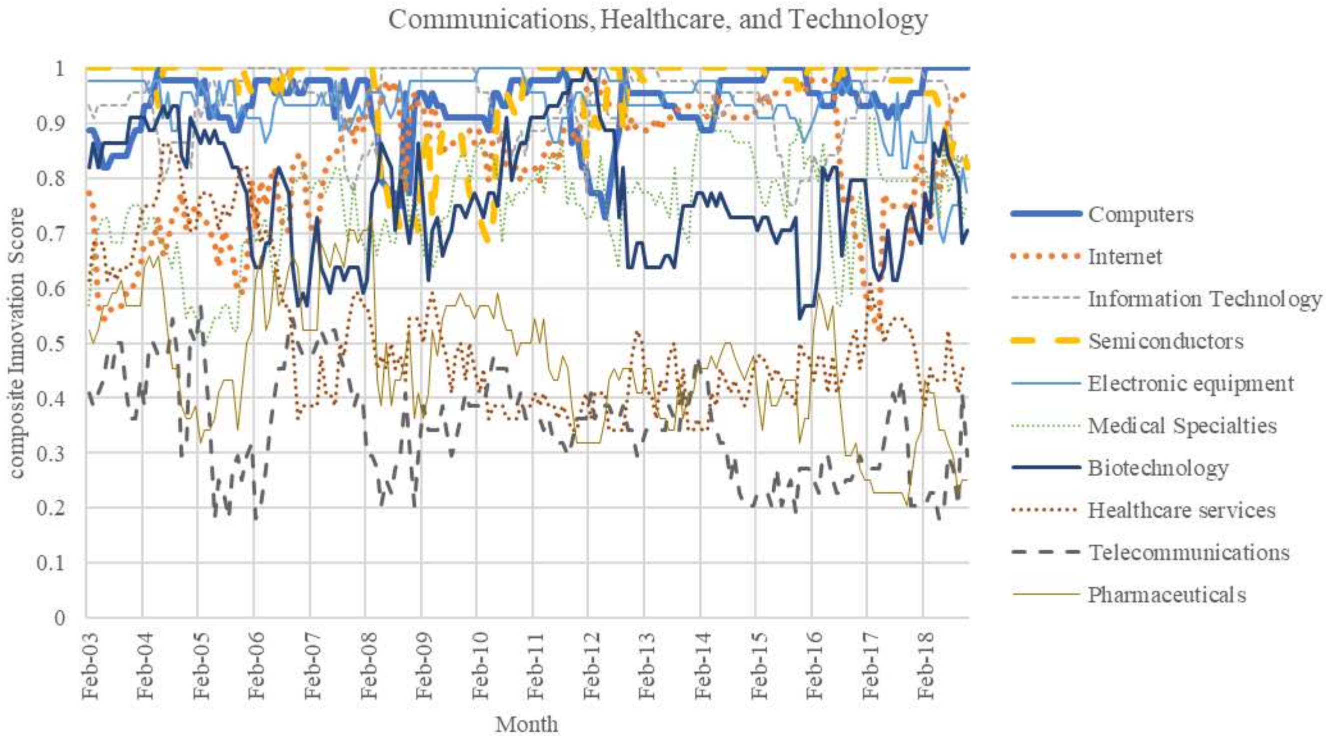
Value: Earnings yield and sales yield
Momentum: 12 month total return, excluding the prior 1 month
Volatility: standard deviation of daily returns over the prior 12 months
Growth: Year over year earnings and sales growth
Leverage: Debt to equity ratio
Yield: Dividend yield
Size: Log Market Capitalization and Total Assets

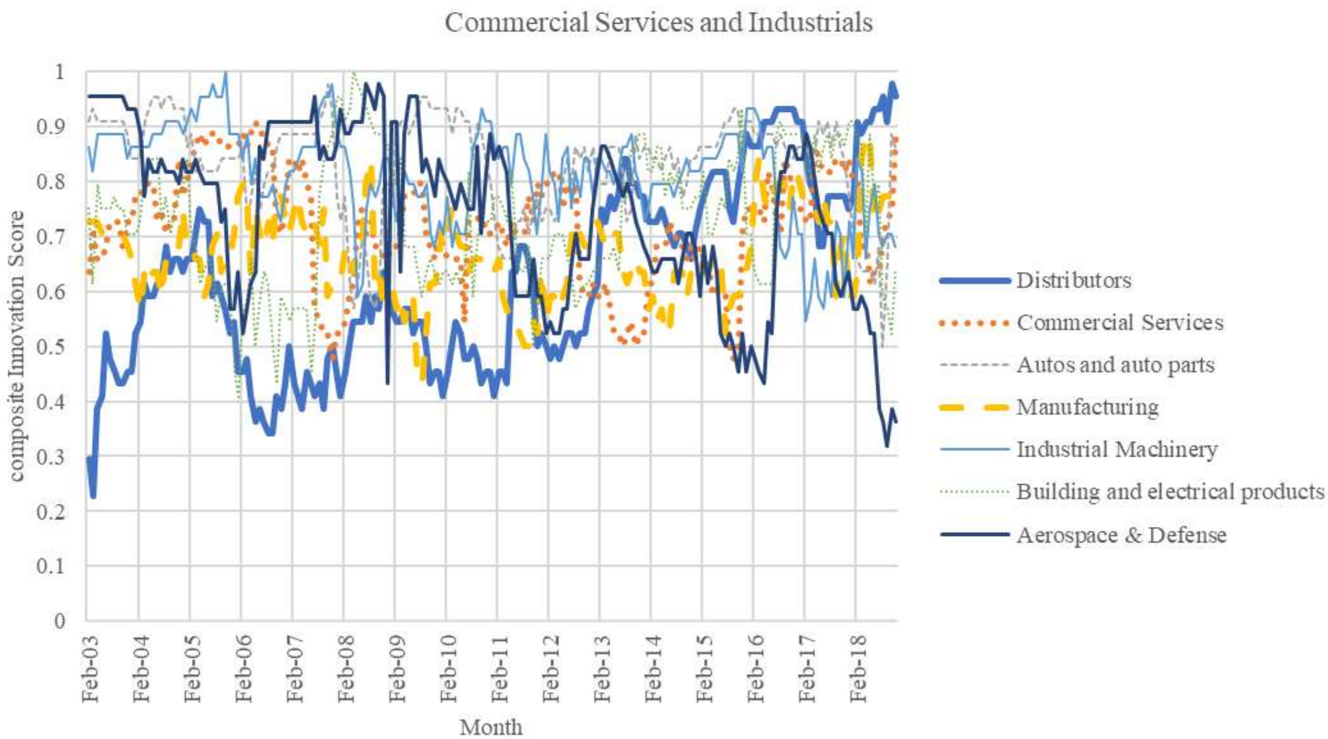
Innovative industries tend to be composed of stocks which are smaller, more volatile, higher growth, and with higher valuations and lower dividend yields than other industries. Surprisingly, they do not exhibit higher momentum.

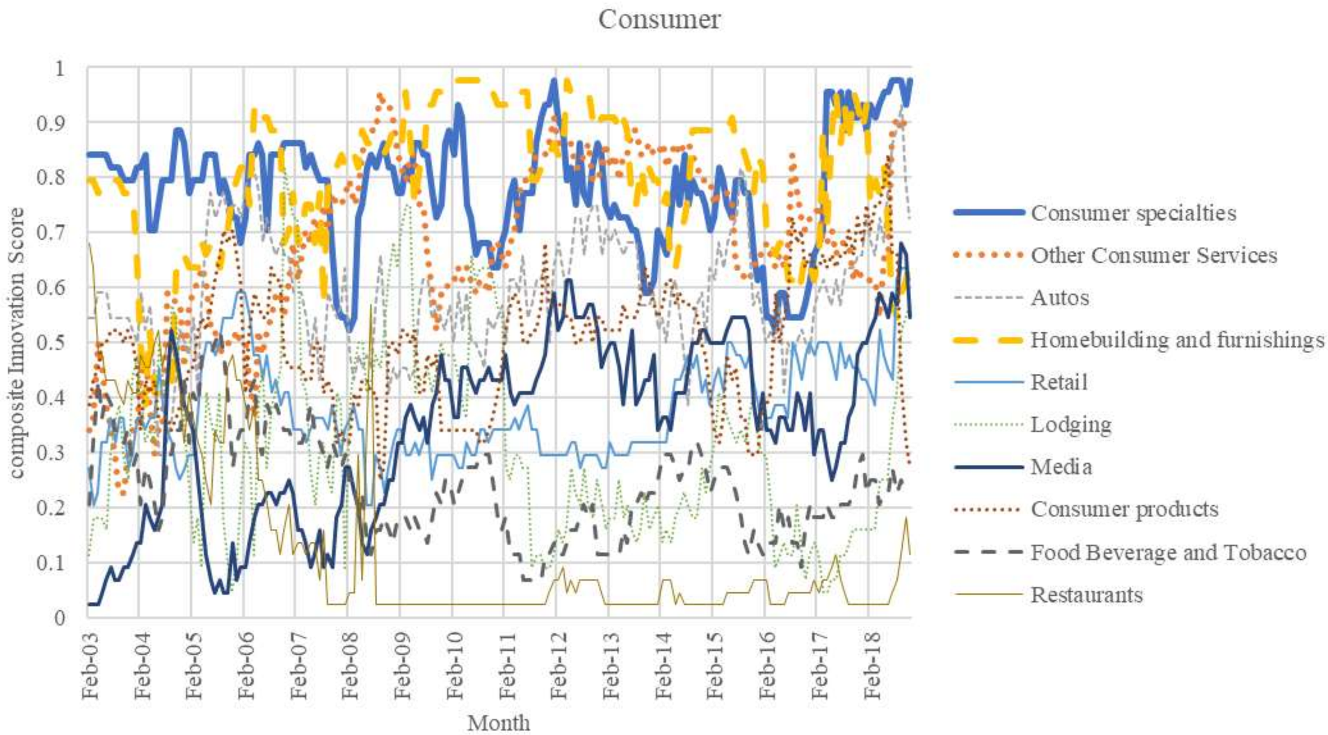
When selecting industries, some allocation models look at momentum, value, and growth factors. If we perform a simple cross-sectional residualization of our Innovation Score and build portfolios out of the residualized score, we see full sample excess returns of 50bps per annum with an IR of 0.32. So it seems that perhaps a third of the value of the innovation score may be explained by these factors.

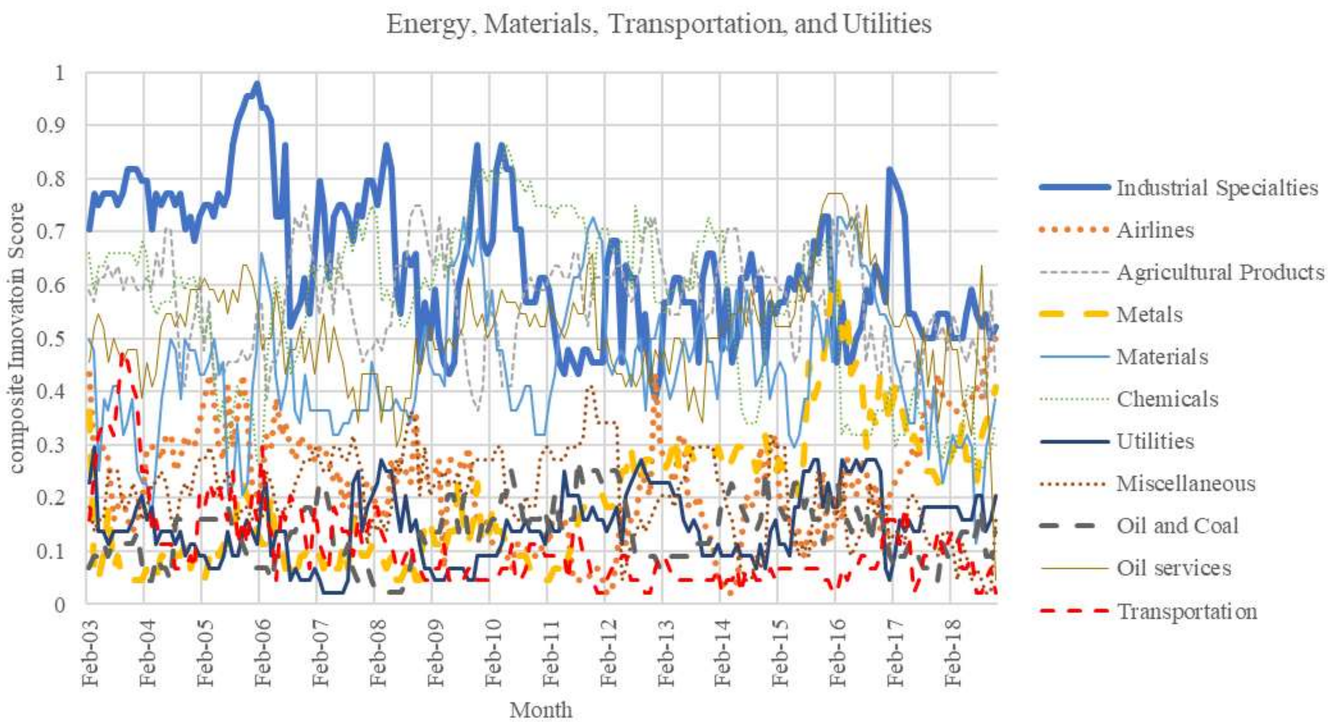
In Exhibit 3 we plot the composite Innovation Scores for each industry over time, by broad sector groupings. Some technology industries such as Computers and Semiconductors remain at or near the top of the rankings throughout time, and Financial industries generally have consistently low innovation scores. The month over month autocorrelation of the level-type scores is 98% or higher, and for the year-over-year type scores it is 88% or higher. This high degree of autocorrelation indicates that the industry tilts could be implemented with very low transaction costs. However, there is variation over time. Distributors and Media companies, for example, have become more innovative over time per our measure.

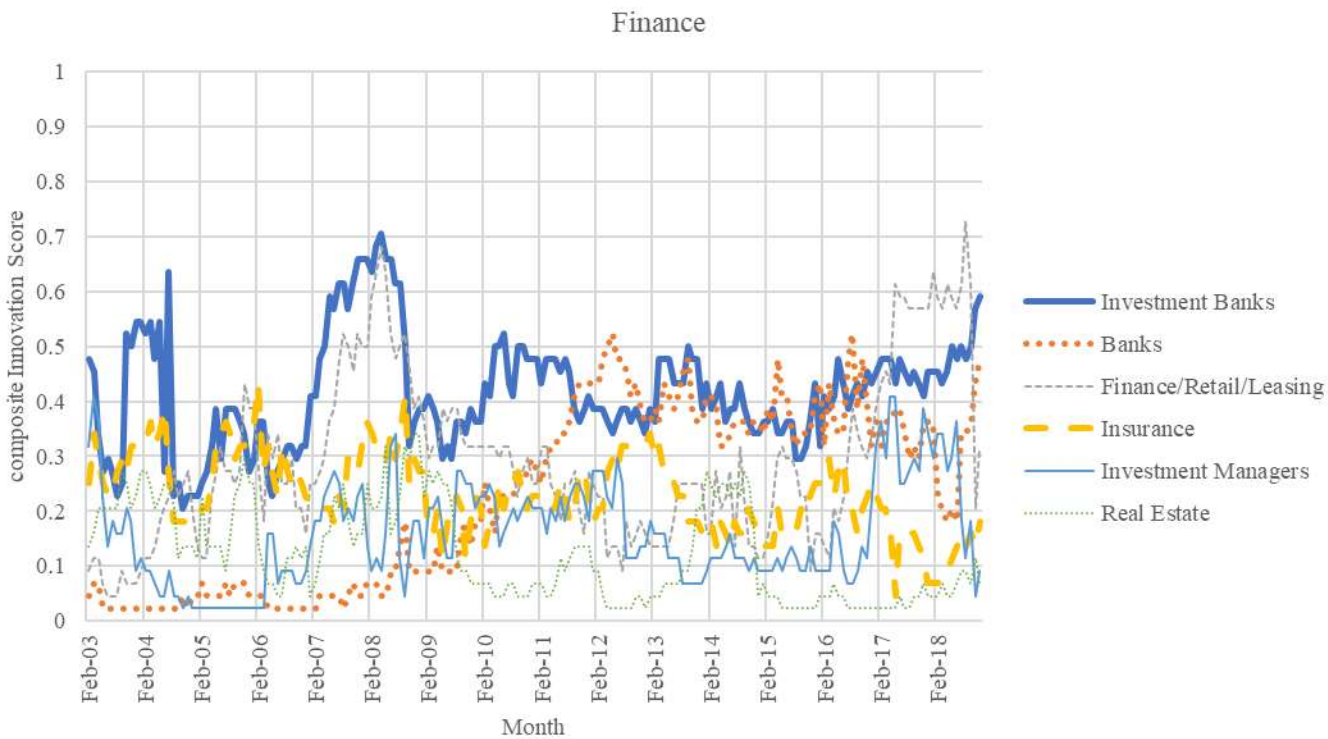
Exhibit 3. Composite Innovation Scores over time by industry


Communications, Healthcare, and Technology
composite Innovation Score
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Feb-03
Feb-04
Feb-05
Feb-06
Feb-07
Feb-08
Feb-09
Feb-10
Feb-11
Feb-12
Feb-13
Feb-14
Feb-15
Feb-16
Feb-17
Feb-18
Month
Computers
Internet
Information Technology
Semiconductors
Electronic equipment
Medical Specialties
Biotechnology
Healthcare services
Telecommunications
Pharmaceuticals


Commercial Services and Industrials
composite Innovation Score
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Feb-03
Feb-04
Feb-05
Feb-06
Feb-07
Feb-08
Feb-09
Feb-10
Feb-11
Feb-12
Feb-13
Feb-14
Feb-15
Feb-16
Feb-17
Feb-18
Month
Distributors
Commercial Services
Autos and auto parts
Manufacturing
Industrial Machinery
Building and electrical products
Aerospace & Defense


Consumer
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
composite Innovation Score
Feb-03
Feb-04
Feb-05
Feb-06
Feb-07
Feb-08
Feb-09
Feb-10
Feb-11
Feb-12
Feb-13
Feb-14
Feb-15
Feb-16
Feb-17
Feb-18
Month
Consumer specialties
Other Consumer Services
Autos
Homebuilding and furnishings
Retail
Lodging
Media
Consumer products
Food Beverage and Tobacco
Restaurants


Energy, Materials, Transportation, and Utilities
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
composite Innovatoin Score
Feb-03
Feb-04
Feb-05
Feb-06
Feb-07
Feb-08
Feb-09
Feb-10
Feb-11
Feb-12
Feb-13
Feb-14
Feb-15
Feb-16
Feb-17
Feb-18
Month
Industrial Specialties
Airlines
Agricultural Products
Metals
Materials
Chemicals
Utilities
Miscellaneous
Oil and Coal
Oil services
Transportation



IV. Stock selection

A natural follow-up question is whether these same metrics can predict the relative performance of stocks *within* industries. For each of the eight company-level counts – four levels and four changes in H1B visa applications, permanent visa applications, patent applications, and patent grants, and their year over year changes – we scale them by the stock's market capitalization at the beginning of each month. We then rank these by month and by industry, and form dollar-neutral, equally weighted quintile portfolios which are long the top 20% of stocks in each industry and short the bottom 20% of stocks in each industry. Note that although nearly every industry should have some visa and patent data, the same is not true at the stock level; in any given industry there could be many stocks where there are no visa applications or patents in our matched sample. As a result, for our four level indicators we build portfolios in two different ways: one which considers only stocks with nonzero visas and patents, and another which includes the zero-visa and zero-patent stocks in the short portfolio.

Because these portfolios will be exposed to various risk factors, we also try cross-sectionally residualizing the ranks by our full set of common risk factors: Value, Momentum, Size, Growth, Leverage, Volatility, and Dividend Yield.

In exhibit 4 we show the gross (pre-transaction cost) returns to the decile portfolios built on raw and residualized innovation counts at the stock level.

Exhibit 4. Dollar neutral stock selection performance of innovation metrics

		Factor	Number of Companies	Annualized Return	Sharpe Ratio	% Positive Days
Raw	Level, nonzero	H1B visa	1164	4.3%	0.62	52%
		Permanent visa	669	6.8%	0.73	52%
		Patent Application	719	2.8%	0.36	51%
		Patent Grant	791	2.6%	0.37	50%
	Level, with zeroes	H1B visa	2508	3.6%	0.42	51%
		Permanent visa	2509	3.5%	0.53	52%
		Patent Application	2508	3.1%	0.38	51%
		Patent Grant	2503	2.5%	0.31	51%
	YoY Change	H1B visa	952	-1.9%	(0.43)	49%
		Permanent visa	459	-2.9%	(0.43)	49%
		Patent Application	585	0.3%	0.05	51%
		Patent Grant	669	0.8%	0.15	51%
			0	0.0%	-	0
Residualized	Level, nonzero	H1B visa	1164	2.6%	0.64	52%
		Permanent visa	668	4.0%	0.61	52%
		Patent Application	719	2.2%	0.38	52%
		Patent Grant	791	2.6%	0.52	52%
	Level, with zeroes	H1B visa	2507	1.1%	0.26	51%
		Permanent visa	2507	4.0%	0.87	53%
		Patent Application	2507	0.8%	0.19	51%
		Patent Grant	2507	0.4%	0.10	51%
	YoY Change	H1B visa	952	-0.8%	(0.20)	49%
		Permanent visa	459	-3.1%	(0.50)	49%
		Patent Application	585	0.2%	0.04	50%
		Patent Grant	668	0.6%	0.12	51%

The stock-level results, while generally in the expected direction, are small in magnitude for a stock-selection factor which would require some degree of transaction costs to implement, relative to an industry-tilt portfolio which could be implemented more cheaply using ETFs. The factor performance is similar and a bit weaker when we include zero-visa and zero-patent stocks in our short portfolio. Many of the factors are also weakened by residualization.

V. Conclusions

It appears that innovation measures can be used in a novel way to select industries which are likely to outperform. None of these factors exhibit significant ability to differentiate between outperforming and underperforming stocks within an industry. But aggregating them to the industry level and applying industry tilts seems to be a useful and cost-efficient way to systematically incorporate innovation metrics into portfolios.

The current findings may motivate future research which could investigate the types of knowledge workers for whom visas are applied, the success of a company at converting visa applications into approved hires or converting patent applications into patent grants, and the content and level of citations of those patents.

References

H. Basse Mama. Innovative efficiency and stock returns: Should we care about nonlinearity? *Finance Research Letters*, Vol 24 (2018)

L.K.C. Chan, J. Lakonishok, T. Sougiannis. The Stock Market Valuation of Research and Development Expenditures. *J. Finance*, 56 (6) (2001)

D. Hirshleifer, P.-H. Hsu, D. Li. Innovative efficiency and stock returns. *J. Financ. Econ.*, 107 (3) (2013)