New Constructs Core Earnings data
White paper

Vinesh Jha, CEO
Yunan Liu, PhD

ExtractAlpha

August 2021

Abstract

Our research finds that New Constructs' Core Earnings – corporate earnings which have been adjusted, using a combination of human and machine learning inputs, to account for transitory shocks and for earnings from activities which are not central to a company's business activities – represent a more accurate and persistent measure of firm's profitability than traditional metrics of a company's profitability. The difference between Core Earnings and reported net income – the *distortion* – can significantly explain future net income even after we have considered analyst consensus forecasts and accounting accruals.

We further show that distortion can be a viable trading signal. A long/short top/bottom decile monthly-rebalanced portfolio, which is long low-distortion stocks and short high-distortion stocks, generates a 10.1% annualized return and Sharpe ratio of 1.44 over the 2015-2021 period.

In addition, most of the return is due to stock idiosyncratic returns ('alpha') rather than factor or sector tilts. After accounting for Fama-French 5 factors, momentum, short-term reversal, and 12 sectors, the signal's alpha is 9.3% (only slightly lower than the raw return of 10.1%).

About ExtractAlpha

ExtractAlpha is an independent research firm dedicated to providing unique, curated, actionable data sets to institutional investors. ExtractAlpha applies their extensive experience in quantitative analysis and the design of investment analytics products to interesting new data sets and tools. Their rigorously built quantitative models are designed for institutional investors to gain a measurable edge over their competitors. ExtractAlpha also partners with top data firms to identify investment value in their data sets and help investors profit from these unique new sources of information.

ExtractAlpha's founding team held senior positions in the original research and sales teams at StarMine and at top quantitative hedge fund groups including Morgan Stanley PDT.

© 2021 ExtractAlpha Research. Confidential - do not copy or distribute.

About New Constructs and the Core Earnings data

New Constructs

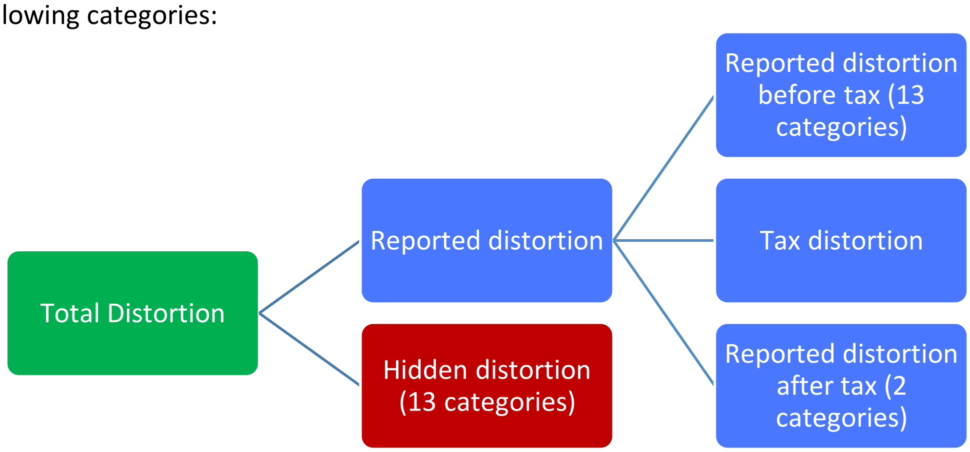
New Constructs (NC) is a research firm that specializes in deep AI-empowered analyses of corporate financial filings, providing users with unique insights about public firms' real economic earnings. NC examines the entirety of the financial filings, including tables, charts, MD&A, and footnotes to extract relevant items that affects a company's earnings. To ensure accuracy and consistency, NC employs analysts and experts to classify relevant items first and, then, trains an AI algorithm to automatically extract and tag new information. Whenever the machine comes across an item that it has not seen previously, it will resort to human intervention to classify and learn. Using this methodology, which is described in detail in Rouen, So, and Wang (2020)[1], NC processes over 150,000 financial filings covering over 5,000 public companies in the U.S.

NC's Core Earnings

Core Earnings, unlike net income, excludes all items from transitory shocks and ancillary business activities identified by NC. Those items, collectively defined as *earnings distortion*, are usually non-recurring. Therefore, Core Earnings is more persistent over time and less susceptible to earnings management, thereby providing a clearer picture about firms, true financial profitability. The relationship between net income and Core Earnings is:

$$\boldsymbol{Core\ Earnings} + \boldsymbol{Total\ Distortion} = \boldsymbol{Net\ Income}$$

If the total distortion is positive, it means that the company reports more earnings in the financial statement than its actual Core Earnings as defined by NC. The total distortion is further broken down into the following categories:



- **Reported distortions** are transitory gains/losses that investors can find reported on income statements, such as those related to discontinued operations, M&A expenses etc.
- **Hidden distortions** are transitory gains/losses that cannot be found on the income statement and are available only in financial footnotes or other disclosures. For examples of the difference in Hidden and Reported distortions, click here.

[1] Rouen, Ethan and So, Eric C. and Wang, Charles C. Y., Core Earnings: New Data and Evidence (November 20, 2020). Harvard Business School Accounting & Management Unit Working Paper No. 20-047, October 2019, Journal of Financial Economics (JFE), Forthcoming, Available at SSRN: https://ssrn.com/abstract=3467814 or http://dx.doi.org/10.2139/ssrn.3467814

© 2021 ExtractAlpha Research. Confidential - do not copy or distribute.

In the below studies, we also examine **analyst distortion** – that is, the difference between Core Earnings and consensus sell-side analyst earnings forecasts.

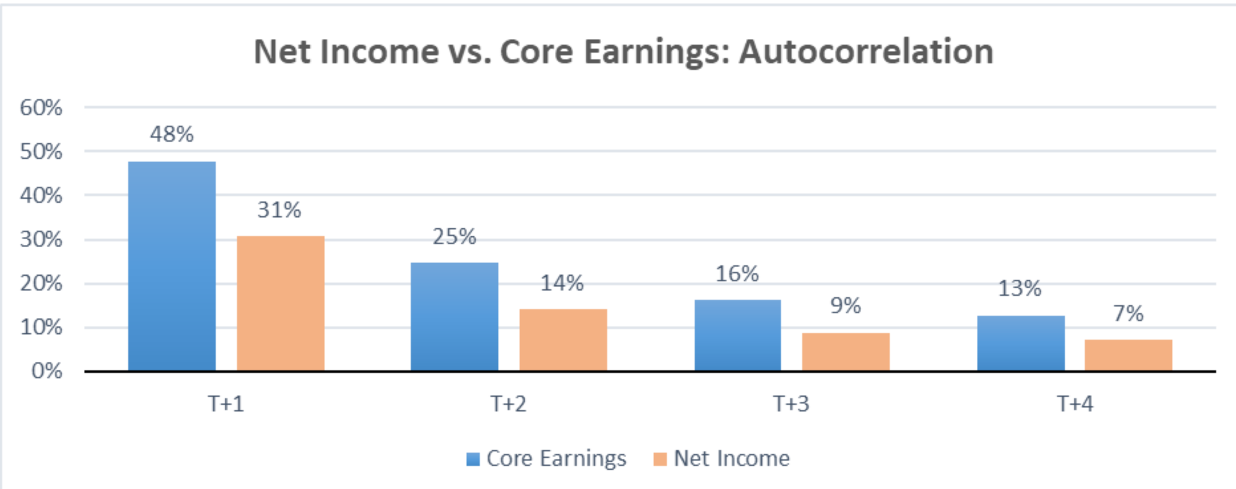
For our studies of persistence, we use NC's trailing 12 month (TTM) Core Earnings values, in order to take advantage of their longer history. However, for the subsequent backtests, we use NC's quarterly data.

Forecasting properties of Core Earnings

Persistence

We find that Core Earnings is more persistent over time as compared to net income; its autocorrelation with next year's value is 48%, which is noticably higher than the 31% for net income. In addition, the autocorrelation with values further into the future is also stronger for core earnings.

Figure 1: Autocorrelation of net income and Core Earnings



Predicting next period net income

We next show that distortion has a significant negative relationship with next-period net income. Table 1 shows the regression results where we model next-period net income against a group of explanatory variables. All variables have been scaled by the total number of shares outstanding.

The results in Regression Model 1 show that after controlling for the current-year net income, total accruals, and the difference between net income and EBITDA, total distortion has a significant negative relationship with next-period net income. Economically, 1 unit of distortion corresponds to 0.516 units lower future reported earnings. Model 2 further controls for the analyst forecast. As expected, the predictive power of distortion decreases by some extent since some of the distortion might have been factored into analysts' forecast. Nonetheless, the negative relationship is still very significant at 0.26 with a T statistic of -17.9.

© 2021 ExtractAlpha Research. Confidential - do not copy or distribute.

Table 1: Regression results for next period net income

Regression Model 1			**No. Observations: 45504, R-squared: 0.445**	
	Coefficient	STD Error	T Stats	P>|t|
Intercept	0.131	0.016	8.42	0.000
Net Income	0.733	0.004	174.39	0.000
Total Distortion	-0.516	0.011	-47.00	0.000
Total Accruals	0.000	0.006	0.03	0.977
Net Income - EBITDA	0.030	0.005	6.07	0.000

Regression Model 2			**No. Observations: 29244, R-squared: 0.482**	
	Coefficient	STD Error	T Stats	P>|t|
Intercept	-0.078	0.02	-3.88	0.000
Net Income	0.386	0.008	46.38	0.000
Analyst Consensus	0.560	0.01	53.62	0.000
Total Distortion	-0.255	0.014	-17.91	0.000
Total Accruals	0.057	0.007	8.08	0.000
Net Income - EBITDA	0.085	0.006	13.16	0.000

Backtesting Methodology

Signal construction

Total distortion	Hidden distortion	Analyst distortion
Total distortion scaled by market capitalization, 4-quarter moving average	Hidden distortion scaled by market capitalization, 4-quarter moving average	Forecasted next quarter Core Earnings (Holt-Winters method[2]) minus analysts' earnings forecasts, scaled by absolute value of Core Earnings

NC distortion signal
Equal-weighted average of the ranking of the above features, then reversed to be positively related to expected return

The figure above illustrates our feature engineering process. We identified three key features which are predictive of future stock returns. Although hidden distortion is already a component of total distortion, we find it to be the most influential distortion component. The analyst distortion measures the difference between Core Earnings and analyst forecasts. Table 2 shows that the correlation among them is low. The final signal we put into our backtest is an equally weighted average of the three components'

[2] The Holt-Winters method is a type of time-series model that comprises exponential smoothing for three components: level, trend and seasonal adjustment. It is suitable for forecasting time series data that exhibit both trend and seasonal behaviors. Both trend and seasonal components can be modelled as additive or multiplicative. In this study, we choose the 'additive' approach.

© 2021 ExtractAlpha Research. Confidential - do not copy or distribute.

ranking. We choose equal weighting to avoid the risk of overfitting or data mining, and we reverse the sign of the distortion so that we are going long stocks with lower distortion and going short stocks with higher distortion.

Table 2: Correlation matrix

	Total Distortion	Hidden Distortion	Analyst Distortion
Total Distortion	1.000		
Hidden Distortion	0.161	1.000	
Analyst Distortion	-0.006	0.071	1.000

Backtest settings

Investible universe	EA's investible universe, which requires at least US$100m in market capitalization, US$1m in average daily trading volume, and $4 in nominal stock price. This universe is updated point in time and is without survivorship bias.
Rebalancing frequency:	Monthly
Time of trade	Our signal measurement date is the last day of each month based on all available information. The distortion features are usually available two days after the filing date of the quarterly financial report (10Q or 10K). Only the records with filing date 3 or more days before the measurement date are incorporated in the current rebalancing cycle. Our trade execution is at the market close on the first trading day of the next month.
Long-short percentiles:	Unless otherwise specified, our baseline long-short model is to go long the top 10% and to go short the bottom 10%. We also test a more extreme setting where we go long the top 5% and go short the bottom 5%. For long-short portfolios constructed within each sector, we choose the top 20% / bottom 20% to ensure a sufficient number of stocks in the long and short buckets

Sample coverage

The signal is built from the quarterly distortion data which is available from 2012. Because our total and hidden distortion signals require 4 quarter of data for smoothing, their backtesting sample will start from 2013. For analyst distortion, we use the first 3 years of data to initialize the Holt-Winters model. Hence, its backtesting sample will start from 2015.

© 2021 ExtractAlpha Research. Confidential - do not copy or distribute.

Historical returns performance

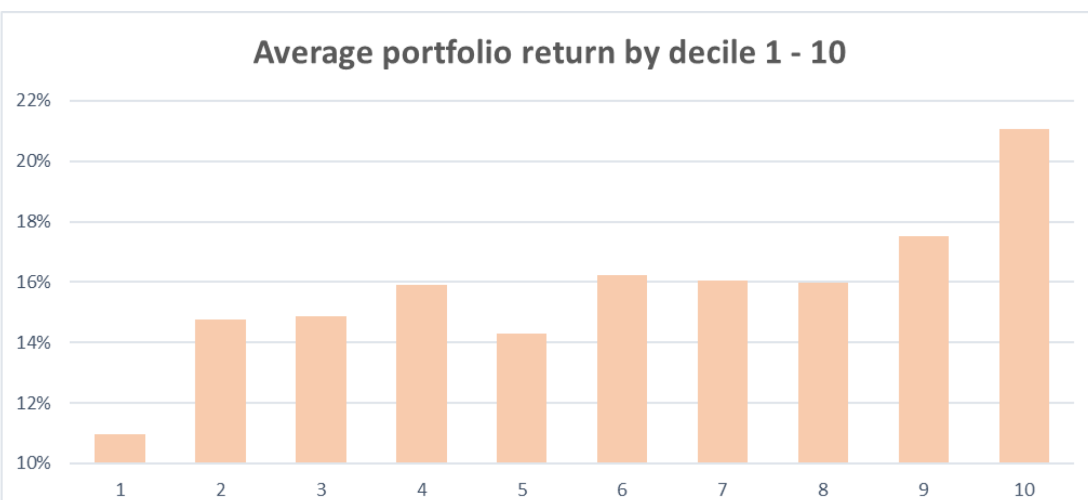
Table 3 shows the decile long-short portfolio returns and Sharpe ratios for total distortion, hidden distortion, and analyst distortion. We will go long those stocks that report large negative distortion and go short those that report large positive distortion. All three features produce positive alpha over their sample periods, and the performance is particularly strong in recent years. Additionally, their performances are not perfectly synchronized, suggesting potential benefits of combining them to form a blended signal.

Table 3: Long-short portfolio returns of the three distortion features

	Total Distortion		Hidden Distortion		Analyst Distortion	
	Annual Return	Sharpe Ratio	Annual Return	Sharpe Ratio	Annual Return	Sharpe Ratio
Full Sample	6.6%	0.69	5.2%	0.89	4.7%	0.61
2013	12.2%	2.65	3.8%	0.80		
2014	-7.1%	-1.41	-5.0%	-1.11		
2015	-11.0%	-1.56	-1.1%	-0.22	0.8%	0.16
2016	19.8%	1.46	8.1%	1.15	5.7%	0.95
2017	6.2%	0.91	-0.2%	-0.05	3.9%	0.87
2018	-3.2%	-0.65	1.4%	0.24	3.9%	0.95
2019	7.4%	1.05	1.8%	0.39	12.5%	2.28
2020	15.3%	0.91	22.8%	2.99	1.3%	0.09
2021	42.3%	3.20	30.2%	3.31	5.1%	0.85

As described in the prior methodology section, we combine these three distortion features using a static, equal-weighted approach, with the higher-distortion stocks being on the short side and the lower-distortion stocks being on the long side. Figure 2 shows the average annualized return of deciles portfolios formed using this signal. In general, the signal is positively related to future stock returns and the relationship is monotonic. The relationship is much stronger on the top and bottom decile, where the difference in annualized return is over 10%.

Figure 2: Average return by decile



© 2021 ExtractAlpha Research. Confidential - do not copy or distribute.

Blended portfolios

Table 4 shows the detailed performance metrics of the long-short portfolio. Overall, the annualized return is 10.1% with Sharpe ratio of 1.44. If we form a more extreme portfolio using the top 5% and bottom 5%, the annualized return is 13.0% with nearly identical Sharpe ratio of 1.43. The basic decile portfolio is sufficiently diversified, with about 220 stocks in the long and short side on average. Since the rebalancing frequency is monthly, the daily average turnover is very low at only 1.2%.

We further show the signal's performance by year. With the exception of 2015, the signal generates positive returns for all years and its performance is particularly strong in years 2020-21. Furthermore, the signal seems to work better among small caps (a group roughly analogous to Russell 2000 constituents), partly because market information for them is more opaque and so their distortion is less likely to be anticipated by the market.

Table 4: Long-short portfolio returns of the NC distortion signal

	Annual Return	Sharpe Ratio	Max Drawdown	Daily Turnover	No. Stocks Long	No. Stocks Short	No. Days
Full Sample							
Long-short 90/10	10.1%	1.44	-12.1%	1.2%	220	219	1609
Long-short 95/5	13.0%	1.43	-11.0%	1.5%	110	109	1609
By year - long-short 10/90							
2015	-2.3%	-0.43	-7.1%	1.2%	218	217	252
2016	13.9%	1.72	-3.5%	1.1%	220	219	252
2017	6.8%	1.27	-4.3%	1.1%	225	224	251
2018	0.3%	0.06	-9.8%	1.1%	226	225	251
2019	6.2%	1.05	-4.1%	1.2%	220	219	252
2020	23.5%	2.43	-9.7%	1.5%	215	215	253
2021 (up to 20210524)	41.5%	4.25	-3.1%	1.7%	206	205	98
By market cap - long-short 10/90							
Large Cap	-0.4%	-0.05	-23.3%	1.9%	41	40	1609
Mid Cap	1.4%	0.15	-19.7%	3.2%	42	41	1609
Small Cap[3]	13.4%	1.69	-14.5%	1.9%	137	136	1609

[3] The Large Cap refers to the largest 500 stocks by market cap in our investible universe, Mid Cap refers to the next 500 stocks by market cap ranking and Small Cap refers to all remaining stocks in the universe.

© 2021 ExtractAlpha Research. Confidential - do not copy or distribute.

Figure 3 presents the cumulative performance of the signal and its three constituent features. For the long-short 95%/5% portfolio, its cumulative return is over 80% and it shows no significant drawdown.

Figure 3: Cumulative return of the three features and the NC distortion signal

Sector portfolios

Table 5 shows the signal's performance if we form the long-short portfolio within each sector. To ensure a reasonably diversified portfolio, we adjust our settings to long-short 80%/20% (hence a smaller magnitude of return than previously shown).

Among all sectors, the signal performs well in consumer discretionary, materials, and commercial services, but not in healthcare, finance, and utilities. If we exclude the worst-performing 5 sectors and rerun the backtest, the signal generates an annualized return of 14.1% (Table 6) which is much larger than the 10.1% generated when using the full universe.

© 2021 ExtractAlpha Research. Confidential - do not copy or distribute.

Table 5: Long-short portfolio returns by sector

	Annual Return	Sharpe Ratio	Daily Turnover	No. Stocks Long	No. Stocks Short	No. Days
By sector - long-short 80/20						
Consumer Discretionary	10.8%	1.00	1.2%	60	59	1609
Materials	10.4%	0.91	1.3%	28	28	1609
Commercial Services	6.2%	0.51	1.2%	24	23	1609
Technology	5.8%	0.69	1.1%	67	66	1609
Energy	3.3%	0.19	1.7%	25	24	1609
Consumer Non-Durables	2.0%	0.14	1.0%	17	16	1595
Utilities	1.8%	0.14	1.1%	14	13	1609
Finance	1.5%	0.21	1.1%	105	104	1609
Healthcare	-0.1%	-0.01	1.2%	50	49	1609
Industrials	-0.1%	-0.01	1.1%	37	36	1609
Transportation	-16.5%	-0.88	1.6%	11	10	1417

Table 6: Long-short portfolio returns for samples excluding non-performing sectors

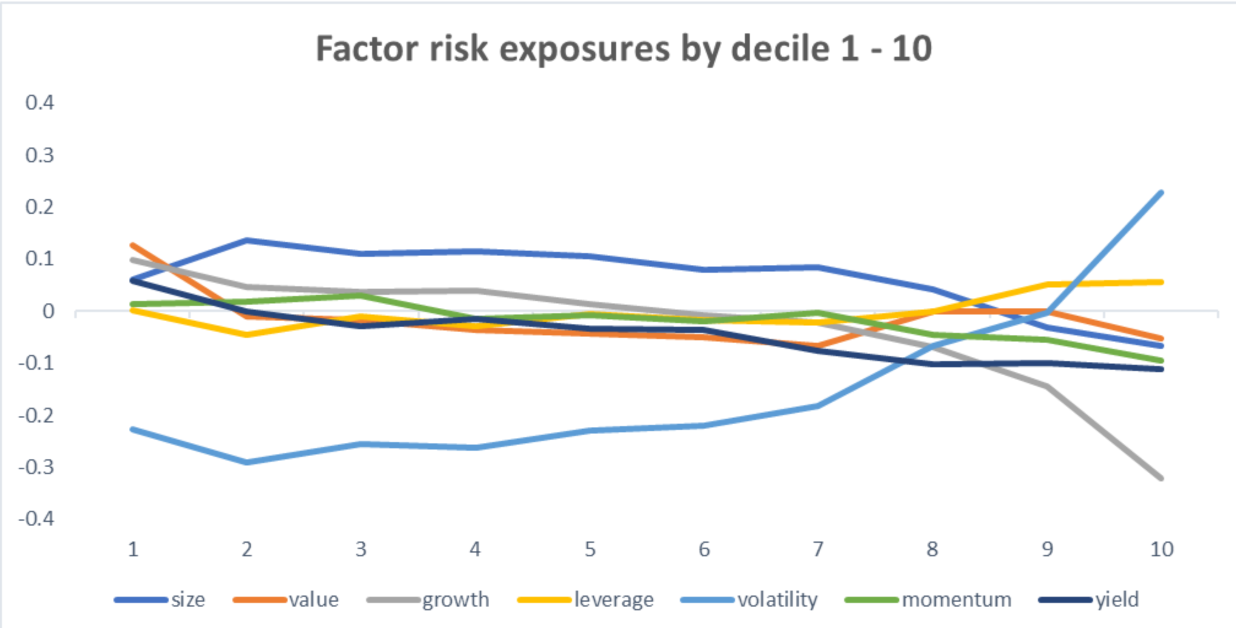
	Annual Return	Sharpe Ratio	Max Drawdown	Daily Turnover	No. Stocks Long	No. Stocks Short	No. Days
Full Sample							
Long-short 90/10	14.1%	1.54	-10.4%	1.4%	112	111	1609
By year							
2015	0.9%	0.15	-5.3%	1.3%	114	113	252
2016	22.4%	2.15	-4.3%	1.2%	114	113	252
2017	3.7%	0.62	-4.5%	1.3%	115	114	251
2018	1.7%	0.32	-9.6%	1.3%	115	114	251
2019	14.2%	2.39	-3.1%	1.3%	111	110	252
2020	23.6%	1.61	-9.8%	1.7%	109	108	253
2021	60.6%	4.44	-4.0%	2.0%	103	102	98

© 2021 ExtractAlpha Research. Confidential - do not copy or distribute.

Risk exposures

Figure 4 shows the average factor risk exposure for deciles 1 to 10. The risk exposure numbers are normalized across the full universe of stocks in order to have a mean of zero and standard deviation of one. Therefore, the magnitudes of risk exposures shown in the chart are all fairly small for all 7 factors, and none of these exposures is greater than 0.3 standard deviations. The signal seems to have a positive relationship with volatility and negative relationship with size and growth.

Figure 4: Average risk exposure by decile



Return attribution

To exclude the factor and industry influences from our backtest results, we conduct a thorough Fama-French regression where we regress our daily decile long-short returns on Fama-French five factors plus momentum, short-term reversal, and 12 sectors. The result in Table 7 shows that after considering all factor and sector correlations, the signal's alpha is 9.3% (only slighter lower than the raw excess return of 10.1%) and is statistically significant with a T-stat of 4.456.

© 2021 ExtractAlpha Research. Confidential - do not copy or distribute.

Table 7: Fama-French regression results

	Coefficient	STD Error	T Stats	P>|t|
Alpha - annualized	0.093	0.021	4.456	0.00
Factor exposure				
Size	0.153	0.016	9.709	0.00
Value	0.008	0.03	0.265	0.79
Profitability	-0.062	0.026	-2.334	0.02
Investment	0.296	0.032	9.397	0.00
Momentum	-0.047	0.012	-3.964	0.00
ST Reversal	*0.005*	*0.009*	*0.509*	*0.61*
Sector Exposure				
Consumer Nondurables	-0.047	0.02	-2.377	0.02
Consumer Durables	0.009	0.007	1.204	0.23
Manufacturing	0.024	0.022	1.105	0.27
Energy	0.074	0.009	8.541	0.00
Chemicals	-0.043	0.019	-2.213	0.03
Business Equipment	0.131	0.019	6.811	0.00
Telecom	-0.021	0.014	-1.472	0.14
Utilities	-0.129	0.011	-11.523	0.00
Wholesale & Retail	0.008	0.019	0.44	0.66
Healthcare	0.020	0.017	1.225	0.22
Finance	-0.025	0.024	-1.055	0.29
Other	0.010	0.027	0.371	0.71

Conclusion

Our research demonstrates that Core Earnings from New Constructs provides investors with a unique and reliable indicator of firm's profitability. It is more persistent over time than net income, and it strongly predicts the next period's financial outcomes even after taking into account accounting accruals and analysts' consensus forecasts.

We also show that the features from the Core Earnings data can be turned into a viable trading signal, which in our case is an equal-weighted composite of the total distortion, hidden distortion and analyst distortion. A decile monthly-rebalanced, market neutral portfolio built from these features has returned 10.1% annually with a Sharpe ratio of 1.44, and its performance was noticeably strong in 2020 and 2021. Using a Fama-French return attribution, we find that most of the long-short return (9.3%, versus the 10.1% total return) is idiosyncratic alpha, rather than due to factor or sector exposures. In particular, its relatively low loading (even negative) on the profitability factor indicates that the signal has its unique predictive power which is not captured by traditional profitability metrics.

© 2021 ExtractAlpha Research. Confidential - do not copy or distribute.