

New Constructs' Core Earnings Data

White Paper

Vinesh Jha, CEO

Yunan Liu, PhD

ExtractAlpha

October 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 remove the impact of transitory shocks and earnings from activities which are not central to a company's business activities – represent a more accurate and persistent measure of company's profitability than traditional profitability metrics.

The difference between Core Earnings and reported net income – the *Earnings Distortion* – can significantly explain future net income even after we have considered analyst consensus forecasts and accounting accruals. Additionally, we find that Core Earnings forecasts - next quarter's Core Earnings, forecasted using data from prior quarters - is a more effective benchmark for gauging signal in analyst consensus forecasts, compared to forecasted net income.

We further show that the *Core Earnings Signal* consisting of Earnings Distortion (Reported and Hidden) and the difference between Core Earnings forecasts and analyst consensus forecasts (*Analyst vs Core Earnings Forecasts*) is a viable trading signal. A long/short top/bottom decile monthly-rebalanced portfolio, which is long stocks with large negative Earnings Distortion and/or large positive growth expectations from the sell-side relative to our Core Earnings forecast (short stocks with large positive distortion and large negative growth expectations), 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.

About New Constructs and the Core Earnings Data

New Constructs (NC)

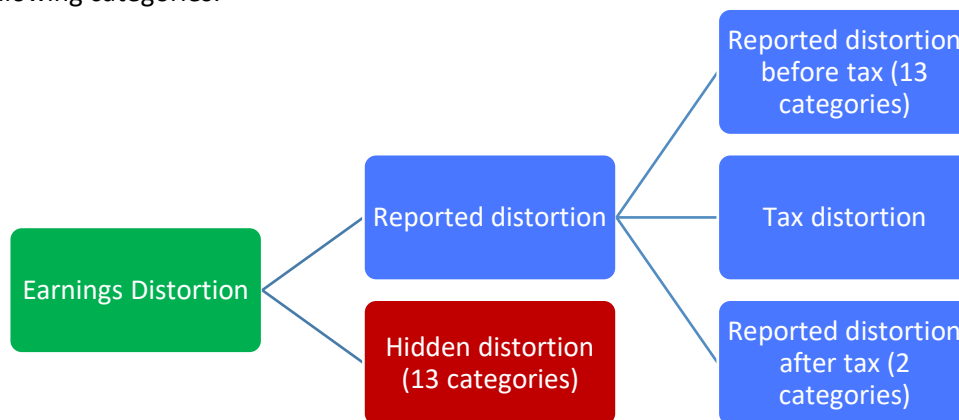
New Constructs (NC) is a research firm that specializes in deep AI-empowered analyses of corporate financial filings, providing users with unique insights about public firms’ real economic earnings. NC examines the entirety of the financial filings, including tables, charts, MD&A, and footnotes to extract 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 270,000 financial filings covering over 5,500 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:

$$\text{Core Earnings} + \text{Earnings Distortion} = \text{Net Income}$$

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



1. **Reported distortions** are transitory gains/losses reported on income statements, such as those related to discontinued operations, M&A expenses etc.
2. **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](#).
3. **Analyst vs Core Earnings Forecasts**: We also use a forecast of next quarter’s Core Earnings as a benchmark against analysts’ consensus earnings forecasts to more accurately gauge the sell-side’s estimates for growth.

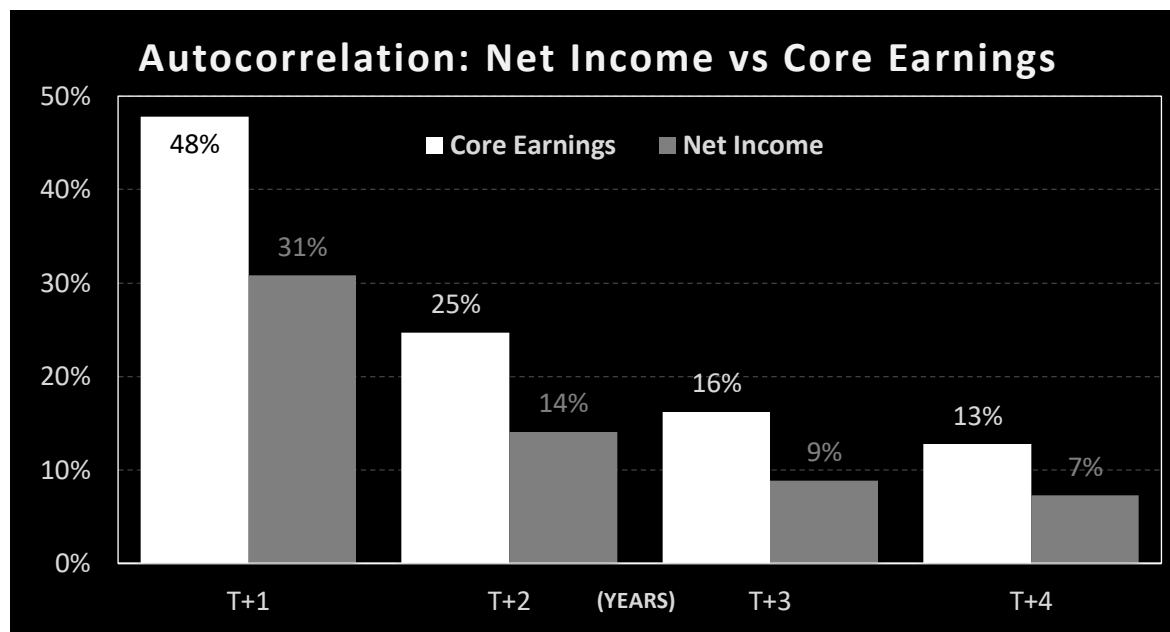
¹ Rouen, Ethan and So, Eric C. and Wang, Charles C. Y., [Core Earnings: New Data and Evidence](#) (November 20, 2020). Harvard Business School Accounting & Management Unit Working Paper No. 20-047, October 2019. [Journal of Financial Economics](#) (JFE).

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 noticeably 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



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

Predicting next period net income

We next show that distortion has a significant negative relationship with next-period net income. [Table 1](#) (next page) 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, Earnings 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.

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
Earnings 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
Earnings 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

Analyst vs Core Earnings Forecasts (benchmarking growth expectations with forecasts of Core Earnings)

Forecasts of next quarter’s Core Earnings can be used as a benchmark from which analyst consensus forecasts can be gauged; the resulting metric is a measure of sell-side growth expectations that is adjusted for earnings distortion:

- To forecast next quarter’s Core Earnings, we use Triple Exponential Smoothing ([Holt Winters method](#)) to account for both trend and seasonality in the quarterly Core Earnings data.
- See component *Analyst vs Core Earnings* in [Table 2](#) of the next section - backtesting methodology - for an explanation of this property.
- [Table 5](#) in the Historical Performance section shows the improved results from benchmarking analyst consensus forecasts using forecasts of Core Earnings, as opposed to using forecasts of net income.
- [Table 6](#) in the Historical Performance section shows the signal’s performance with and without Analyst vs Core Earnings Forecasts component.

Backtesting Methodology

Signal construction

Table 2: Signal construction

ITEM		DEFINITION	LONG (GOOD)	SHORT (BAD)	ECONOMIC RATIONALE
COMPONENTS	Earnings Distortion	Earnings distortion, scaled by market capitalization, 4-quarter moving average	Negative (Core Earnings is larger than reported net income)	Positive (Reported net income is larger than Core Earnings)	Inflated net income, due to a large and positive distortion, means net income is overstated and is therefore unsustainable and unlikely to persist. Alternatively, a negative distortion means net income is understated.
	Hidden Distortion	Hidden distortion, scaled by market capitalization, 4-quarter moving average			
	Analyst vs Core Earnings Forecasts	Core Earnings forecast for next quarter (Holt Winters method) MINUS sell-side analysts consensus earnings forecast for same quarter, scaled by absolute value of Core Earnings	Negative (Consensus forecast larger than Core Earnings)	Positive (Core Earnings larger than consensus forecast)	
SIGNAL	Core Earnings Signal (raw)	Equal weighted rank average of the 3 components above (as rankings)	Negative	Positive	
	Core Earnings Signal (final)	Inverse of the raw signal above	Positive signal	Negative signal	

The table above explains the signal and its components, feature engineering, and signal construction. We identified three key features which are predictive of future stock returns. Some additional notes:

- Although Hidden Distortion is already a component of Earnings Distortion, we find it to be the most influential subcomponent within Earnings Distortion; thus we've added it as a separate component to reflect its importance. Additionally, Hidden Distortion is unique to [NC's data](#); identified by deep AI-empowered analyses of corporate financial filings.
- The Analyst vs Core Earnings Forecasts component is different from Earnings Distortion in that it is distinctly forward looking; its value is derived from benchmarking consensus forecasts against our Core Earnings forecast, therefore providing an undistorted measure of sell-side growth expectations.
- The Analyst vs Core Earnings Forecasts component is independent from the Earnings Distortion based components due to the forecasting nature of the former. For the former, we are strictly considering the absolute difference between forecasts, whereas the latter measures distortion in the present.
- [Table 3](#) (next page) shows that the correlation among the 3 components is low
- The final Core Earnings signal that we backtest is an equally weighted average of the three components' rankings; we choose equal weighting to avoid the risk of overfitting or data mining, and we reverse the sign of the final signal to keep it intuitive (positive signal equates to going long and vice versa).

Table 3: Correlation matrix

	Earnings Distortion	Hidden Distortion	Analyst vs Core Earnings
Earnings Distortion	1.000		
Hidden Distortion	0.161	1.000	
Analyst vs Core Earnings	-0.006	0.071	1.000

Backtest settings

<i>Investible universe</i>	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.
<i>Rebalancing frequency:</i>	Monthly
<i>Time of trade</i>	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.
<i>Long-short percentiles:</i>	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 Core Earnings data which is available from 2012. Because Earnings Distortion and Hidden Distortion require 4 quarter of data for smoothing, their backtesting sample will start from 2013. For the Analyst vs Core Earnings component, we use the first 3 years of data to initialize the Holt-Winters model. Hence, its backtesting sample starts from 2015.

Historical Performance

Long-short returns

Table 4 shows the decile long-short portfolio returns and Sharpe ratios for all 3 components: Earnings Distortion, Hidden Distortion, and Analyst vs Core Earnings. The portfolio is long stocks with large negative earnings distortion and large positive growth expectations from the sell-side relative to our Core Earnings forecast (short stocks with large positive earnings distortion and large negative growth expectations). 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 4: Long-short portfolio returns of the Core Earnings signal's three components

	Earnings Distortion		Hidden Distortion		Analyst vs Core Earnings	
	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

Table 5 shows the difference in results from benchmarking analyst forecasts with Core Earnings forecasts versus net income forecasts.

Table 5: Long-short portfolio returns for Analyst vs Core Earnings Forecasts component only (benchmarking analyst consensus forecasts using Core Earnings forecasts and net income forecasts)

	Analyst consensus forecast vs. Core Earnings forecast		Analyst consensus forecast vs. net income forecast	
	Annual Return	Sharpe Ratio	Annual Return	Sharpe Ratio
Full Sample	5.6%	0.71	4.8%	0.73
2015	-0.4%	-0.08	2.3%	0.45
2016	5.1%	0.90	4.8%	1.00
2017	2.0%	0.46	3.0%	0.75
2018	4.9%	1.20	3.2%	0.73
2019	13.1%	2.29	12.1%	2.65
2020	8.0%	0.51	-0.2%	-0.02
2021	7.9%	1.17	13.8%	1.95

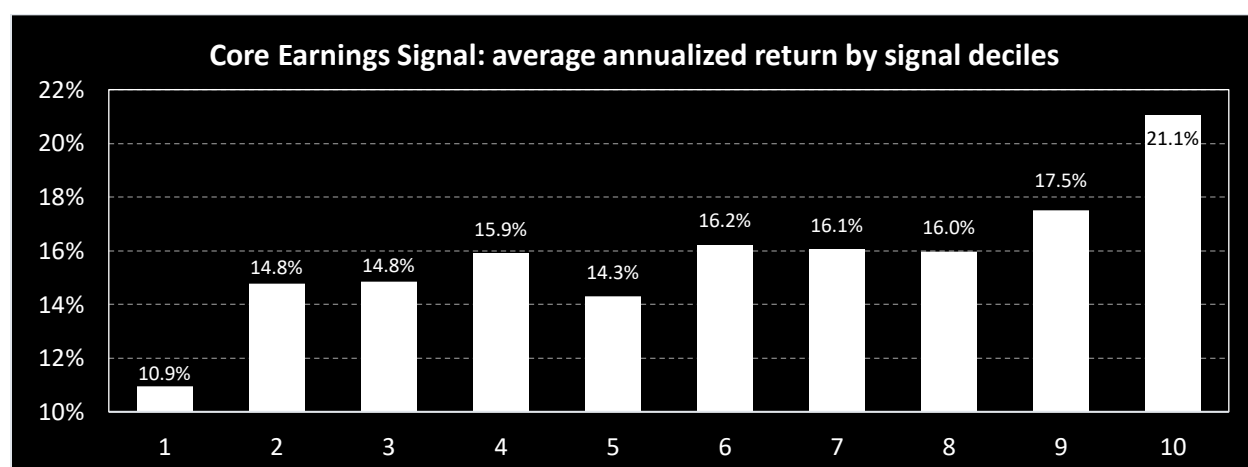
Table 6 shows the difference in results between a signal with and without the Analyst vs Core Earnings Forecasts component.

Table 6: Long-short portfolio returns with and without Analyst vs Core Earnings Forecasts component

	Without Analyst vs Core Earnings Forecasts component (Earnings Distortion + Hidden Distortion only)			All 3 components			Improvement: All 3 components – Earnings Distortion + Hidden Distortion only		
	Annual Return	Sharpe Ratio	Max Drawdown (MDD)	Annual Return	Sharpe Ratio	MDD	Annual Return	Sharpe Ratio	MDD
Full Sample	8.6%	0.93	-24.8%	10.1%	1.44	-12.1%	+1.5%	+0.51	12.1%
By Year									
2015	-4.4%	-0.65	-7.6%	-2.3%	-0.43	-7.1%	+2.1%	+0.22	7.1%
2016	14.7%	1.43	-4.8%	13.9%	1.72	-3.5%	-0.8%	+0.29	3.5%
2017	6.1%	1.03	-3.5%	6.8%	1.27	-4.3%	+0.7%	+0.24	4.3%
2018	-4.2%	-0.63	-12.4%	0.3%	0.06	-9.8%	+4.5%	+0.69	9.8%
2019	4.5%	0.56	-8.6%	6.2%	1.05	-4.1%	+1.7%	+0.49	4.1%
2020	21.2%	1.51	-18.7%	23.5%	2.43	-9.7%	+2.3%	+0.92	9.7%
2021	43.4%	3.62	-4.5%	41.5%	4.25	-3.1%	-1.9%	+0.63	3.1%

As described in the prior [Backtesting Methodology section](#), we combine three components using a static, equal-weighted approach. Figure 2 shows the average annualized return of deciles portfolios formed using the final Core Earnings 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%.

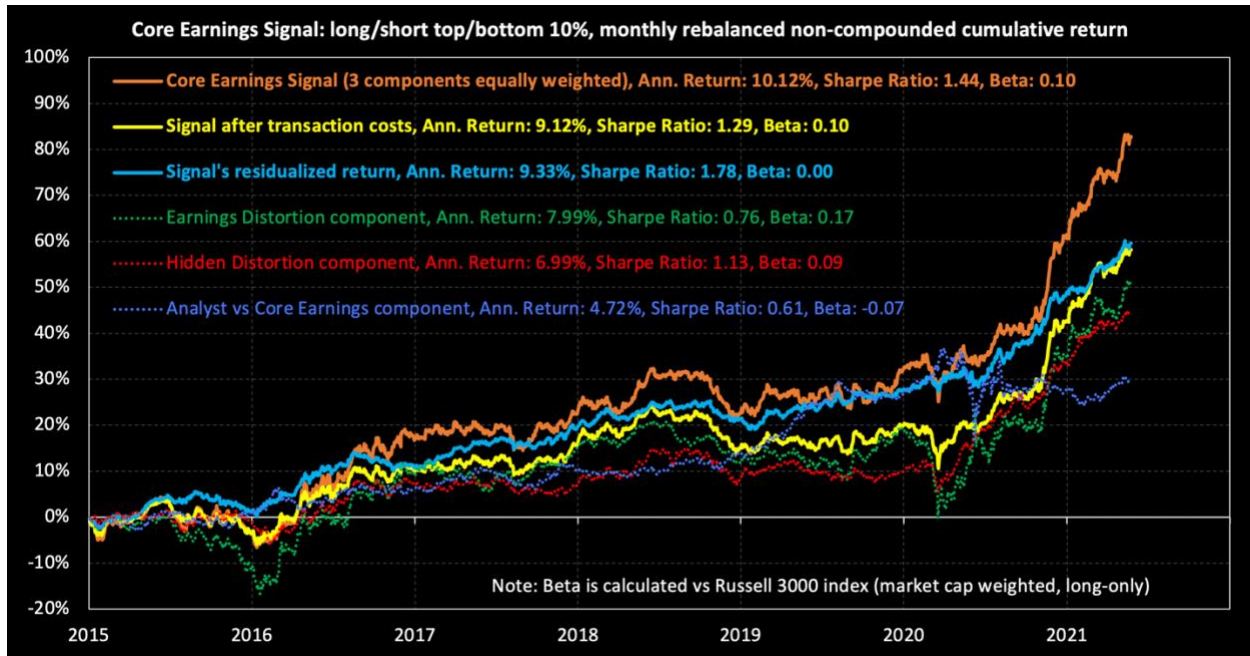
Figure 2: Average return by decile



Blended portfolios

Figure 3 shows the long-short returns of the signal and its components: overall, the annualized return is 10.1% with Sharpe ratio of 1.44. Since the rebalancing frequency is monthly, the daily average turnover is very low at only 1.2%; net return (9.1% p.a.) is very similar to the gross return (10.1% p.a.).

Figure 3: Long-short returns of Core Earnings signal and components



In Table 7, we show the signal’s detailed performance and breakdown by year. The basic decile portfolio is sufficiently diversified, with about 220 stocks in each of the long and short sides on average. If we form a more extreme portfolio using the top 5% and bottom 5%, the annualized return improves to 13.0% with nearly identical Sharpe ratio of 1.43. With the exception of 2015, the signal generates positive returns for all years and its performance is particularly strong in years 2020 and 2021. Furthermore, the signal seems to work better among small caps (a group roughly analogous to Russell 2000 constituents), possibly because market information for them is more opaque and, thus, their distortion is less likely to be anticipated by the market.

Table 7: Long-short portfolio returns of the Core Earnings signal, details and breakdown by year

	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 ²	13.4%	1.69	-14.5%	1.9%	137	136	1609

Sector portfolios

Table 8 (next page) 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 9, next page) which is much larger than the 10.1% generated when using the full universe.

² 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.

Table 8: 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

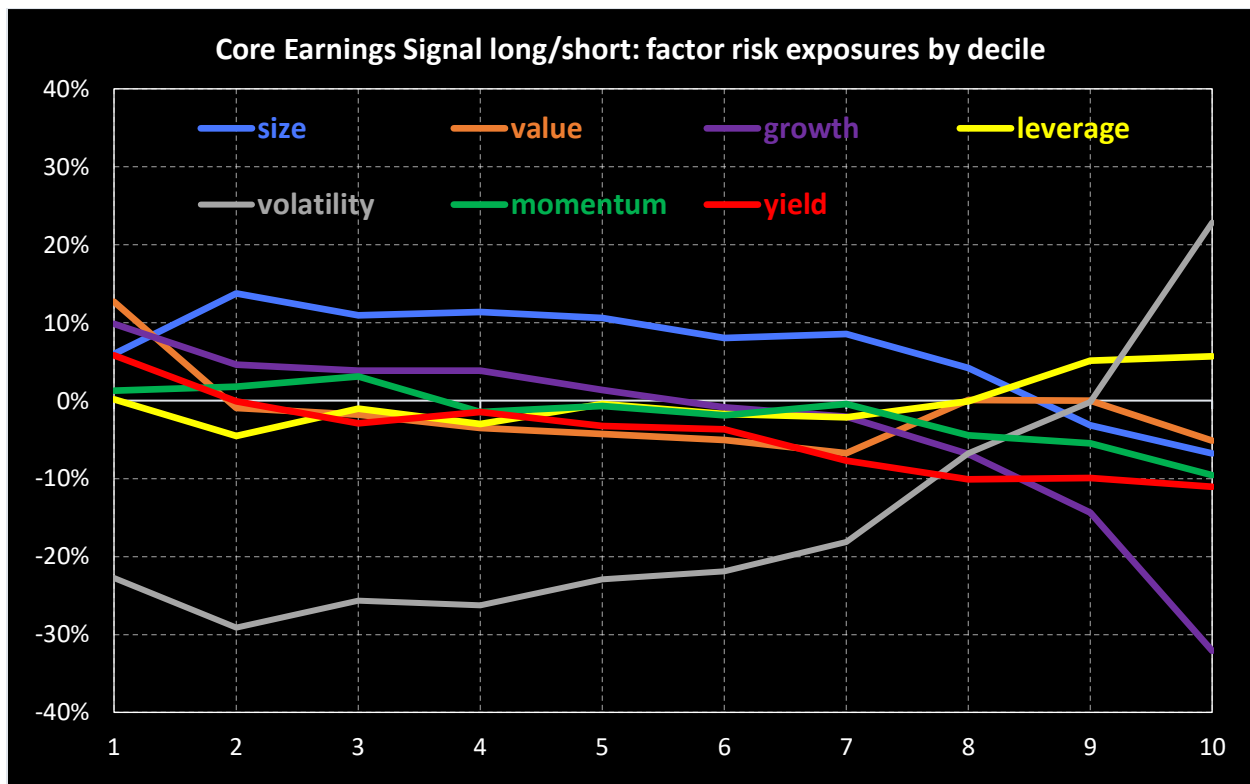
Table 9: Long-short portfolio returns for samples excluding the worst performing 5 sectors

	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

Risk exposures

Figure 4 shows the average factor risk exposure for deciles 1 to 10 of the Core Earnings signal long/short portfolio. We use the [ExtractAlpha Risk Model](#)'s 7 equity risk factors; 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 (long) 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 time series in [Figure 3](#) and details in [Table 10](#) (next page) shows that after removing all factor and sector return components, the signal's residual return ('alpha') is 9.3% p.a. (which is 92% of the raw signal's return at 10.1% p.a.) and is statistically significant with a T-stat of 4.456.

Table 10: Fama-French regression results of Core Earnings signal long-short decile portfolio

	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 Earnings Distortion, Hidden Distortion, and Analyst vs Core Earnings Forecasts components. A decile monthly-rebalanced, dollar-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 is idiosyncratic alpha unexplained by various factors and sectors (9.3% p.a. residual return vs 10.1% p.a. raw return). In particular, its relatively low loading (even negative) on the Fama-French profitability factor indicates that the signal has its unique predictive power which is not captured by traditional profitability metrics.