

Industry Signals and the Dynamics of Stock Market Predictability

Michael Ellington Sam Pybis Yawen Zheng

University of Liverpool

Manchester Metropolitan University

Durham University

October 2026

Research question and motivation

Research question

Can lagged industry returns improve *real-time forecasts* of aggregate US stock market returns?

- **Hong, Torous and Valkanov (2007):** several industries lead the market, consistent with gradual information diffusion.
- **Tse (2015):** predictive relationships become much weaker in an extended sample.
- **Our contribution:** compare direct return regressions with two component-based forecasts, then assess forecast combinations and economic value.

Central idea

Predictive information may be harder to detect when persistent fundamentals and shorter-run repricing are forecast together.

Data and forecast evaluation

Item	Design
Industry signals	40 fully observed Fama–French industry portfolios
Sample	July 1926 to December 2024 (monthly)
Market variables	Prices, dividends, earnings and Treasury bill rate
First out-of-sample forecast	January 1946 (234-month initial estimation window)
Evaluation	Recursive one-month-ahead forecasts through 2024
Benchmark	Recursively updated historical mean

- All estimates and forecast inputs use information available at the forecast origin.
- Inference uses Clark–West tests and moving-block bootstrap checks.

Forecast designs I: Direct regression and RCD

Direct regression (DR)

$$r_{t+1}^{\text{tot}} = \alpha_i + \beta_i X_t^i + \varepsilon_{t+1}$$

Return-component decomposition (RCD)

$$r_{t+1} = g_{t+1} + u_{t+1}, \quad g_{t+1} = ey_{t+1} + dp_{t+1}$$

$$\boxed{\hat{r}_{t+1|t}^{\text{RCD},i} = g_t + \hat{\beta}_{i,t} \tilde{X}_t^i}$$

- Use lagged fundamentals to forecast the persistent component, g .
- Use recursively demeaned industry returns to forecast the capital-gain component, u (without an intercept).

Forecast designs II: Sum of the parts

Adapted sum-of-the-parts (SOP) forecast

$$\hat{r}_{t+1|t}^{\text{SOP},i} = DP_{m,t} + GE20_t + \widehat{GM}_{t+1|t}^i - Rf_{t+1}$$

- ① **Dividend yield:** $DP_{m,t} = (D12_t/12)/P_t$.
- ② **Earnings growth:** $GE20_t$ is the trailing 240-month average of monthly earnings growth.
- ③ **Valuation-multiple growth:** predict GM_{t+1} using lagged industry return X_t^i .

Key difference from DR

Industry returns forecast *price-earnings multiple growth*, not the entire market return.

Adapted from Ferreira and Santa-Clara (2011); implemented recursively using a simple-return approximation.

Out-of-sample results: Individual industries

Forecast	Mean R^2_{OS} (%)	Positive	CW significant
DR	-0.51	5/40	3/40
RCD	0.59	34/40	34/40
RCD ^s	0.79	37/40	37/40
SOP	0.94	39/40	39/40
SOP^s	1.04	40/40	40/40

- The direct approach does not beat the historical mean on average.
- Both component-based designs improve accuracy across most industries.
- The strongest individual SOP forecast is **Books**: $R^2_{OS} = 1.73\%$.

Mean R^2_{OS} and industry counts: manuscript Table 7, Panel C, 1946 start. CW: positive R^2_{OS} and one-sided Clark–West significance at 10%. Superscript *s* denotes shrinkage.

Forecast combinations: Pooling industry signals

Rule	DR	RCD	RCD ^s	SOP	SOP ^s
Mean	-0.11	0.92	1.02	1.30	1.29
Trimmed mean	-0.09	0.94	1.02	1.32	1.29
Best (equal)	0.09	0.88	1.14	1.49	1.28
Best (decay)	-1.10	-0.36	-0.09	0.87	1.32

- Forecast pooling reinforces the benefits of component-based modelling.
- SOP has the highest recorded combination R_{OS}^2 : **1.49%**.
- Gains are *not uniform* across weighting rules or necessarily greater than the best individual forecast.

Out-of-sample R_{OS}^2 in %; selected rows from manuscript Table 4. Mean, trimmed mean and best (equal) SOP results are significant at 1% by Clark–West tests.

Economic value: Certainty-equivalent gains

Forecast	0 bp	25 bp
DR	0.73	-0.39
RCD	2.23	1.21
RCD ^s	2.19	1.39
SOP	2.14	1.32
SOP^s	2.09	1.52

Trading assumptions

- Mean-variance investor
- Risk aversion $\gamma = 5$
- Equity weight in $[0, 1.5]$
- Five-year rolling variance

- Component-based forecasts retain positive mean utility gains after 25 bp trading costs.

Cross-industry means, calculated from manuscript Tables 5 and 6; gains are annualised percentage points over historical-mean timing.

Robustness: The gains vary over time

Subperiod	1946–75	1976–2000	2001–24
SOP mean R_{OS}^2 (%)	2.38	−0.84	1.24
SOP ^s mean R_{OS}^2 (%)	2.50	−0.58	1.17

- **Business cycle:** SOP mean R_{OS}^2 is 4.09% in recessions, versus 0.07% in expansions.
- **Forecast start:** SOP^s mean R_{OS}^2 is 1.04%, 1.04% and 1.05% with 1946, 1950 and 1955 starting dates.
- **Interpretation:** industry signals are more informative in some market conditions; the effect is not constant over time.

Manuscript Table 7, Panels A–C; recession/expansion classifications use lagged NBER recession indicators.

Conclusions

Main finding

Industry signals are more useful when the components of aggregate market returns are forecast separately.

- ① Direct industry-to-market regressions offer little reliable out-of-sample accuracy.
- ② RCD and SOP designs improve forecast accuracy across industries; SOP gives the strongest overall results.
- ③ Statistical gains translate into positive portfolio utility after allowing for trading costs.

The results are consistent with gradual information diffusion, while the variation across subperiods cautions against assuming a stable forecasting relation.