

Monthly Persistence of Cross-Industry Return Dispersion and Its Relevance to Hierarchical Trading

Adrian Wells, Gavin Russell, Tyler Sullivan

Abstract

Hierarchical trading separates broad portfolio decisions from the implementation of positions, but evidence about one layer cannot automatically validate the other. This article examines that boundary using a reproducible secondary analysis of ten public industry return portfolios. For January 2000 through December 2025, monthly cross-industry dispersion averages 3.550 percentage points. A first-order autoregression estimates a persistence coefficient of 0.360, with a heteroskedasticity and autocorrelation consistent 95 percent confidence interval of 0.251 to 0.468. These estimates describe an observable market state; they do not measure trading profitability, stock selection quality, or execution efficiency. The analysis uses the Hierarchical Reinforced Trader as a motivating architecture and identifies the additional observations required to evaluate its decision layers. Efficient adaptation methods can support text-based state construction, provided that information timing and adaptation costs enter the evaluation. A useful trading claim must connect a documented state to an implementable decision and then to an outcome net of the costs of obtaining and executing that decision.

Keywords: hierarchical trading; industry dispersion; public return data; temporal dependence; execution costs; reproducibility.

1. The unit of evidence in a hierarchical trader

An investment architecture can be internally coherent while its evaluation leaves a gap between portfolio intent and market implementation. A high-level component may decide which exposures to increase, whereas a lower component converts those directions into portfolio weights under operational constraints. Hierarchical Reinforced Trader provides a concrete motivation for examining this distinction (Zhao and Welsch, 2024). The revised version consulted here describes sparse high-level directions and risk-aware lower-level allocation. It should not be silently evaluated as though an earlier description of its learning components were the same system. Version identity belongs in the specification of the empirical object.

The central question of this article is deliberately narrower than whether a hierarchical trader makes money. It asks whether a simple public measure of market heterogeneity exhibits temporal structure, and what that finding would establish for evaluating a decision hierarchy. Dispersion is relevant because identical aggregate market returns can conceal very different cross-industry opportunities and risks. However, dispersion alone does not identify which industry will outperform next month. Nor does it indicate whether a particular account can trade a desired exposure at the price assumed by an evaluation. Treating those implications as separate hypotheses is necessary for a credible empirical design.

An explicit division of evidence helps avoid a common shortcut. State description concerns the information available before a decision. Selection concerns the mapping from that information to relative positions. Execution concerns how the positions become transactions. Portfolio evaluation concerns the resulting sequence of net wealth. A study may contribute to any of these stages, but its reported outcome must identify the stage actually observed. This article contributes only to state description and to the design of subsequent tests; it does not present a trained reinforcement learning agent or a simulated trading account.

2. Data, transformations, and a modest regression

The public input is the value-weighted monthly return block in Kenneth French's ten-industry portfolio file (French, 2026). The downloaded archive was constructed from the August 2026 database release. We retain the 312 months from January 2000 through December 2025, thereby avoiding an incomplete final calendar year. Each month contains ten industry portfolio returns expressed in percent. The selected block contains no coded missing observations in this window. These are broad portfolio returns rather than individual stock orders, quote histories, news records, or account-level holdings. The difference determines which empirical questions can be answered.

For month t , define D_t as the sample standard deviation of the ten industry returns, using a denominator of nine. D_t is measured in percentage points. An equally weighted cross-sectional mean is also retained as a descriptive field, but it is not the return of the source's value-weighted aggregate market. Choosing the sample standard deviation rather than a population denominator is a transparent scaling convention. It does not convert the ten industries into independent random draws from all possible sectors. Comparisons with other dispersion measures must preserve the denominator and the portfolio definitions.

The estimating equation is $D_t = \alpha + \rho D_{(t-1)} + u_t$. One initial observation is lost when constructing the lag, leaving 311 regression observations. Ordinary least squares supplies the point estimates. We report Newey-West covariance estimates with six monthly lags and the finite-sample correction enabled (Newey and West, 1987). This choice acknowledges dependence and changing variance in the regression residuals, without assuming that it removes all specification uncertainty. The lag length is a documented analytical choice rather than a value selected to maximize statistical significance. The accompanying script preserves the complete transformation from the downloaded monthly block.

The model is intentionally small. Adding many macroeconomic predictors, technical indicators, or news embeddings would introduce further timing and model-selection decisions before the meaning of the basic series was established. The regression therefore estimates an in-sample descriptive association. It has no rolling training window, untouched forecast period, portfolio construction rule, or trading-cost model. Its confidence interval addresses uncertainty under the specified covariance procedure; it does not provide a probability that a future trading strategy will be profitable. Robust covariance estimation changes inference about a fitted equation, not the economic content of its dependent variable (White, 1980).

3. Results and their permissible interpretation

Table 1 reports the completed analysis. Mean monthly dispersion is 3.550 percentage points, while the median is 2.951. The maximum is 12.782 percentage points in April 2009. These are cross-industry differences within a month, not the volatility of a single industry's return over the entire sample. The fitted persistence coefficient is 0.3598, and the model explains approximately 13.0 percent of the observed in-sample variation in dispersion. A positive relationship between adjacent months is visible in this specification, but substantial variation remains outside the fitted linear component.

Table 1. Public industry-return dispersion, January 2000-December 2025. Confidence limits apply to the autoregressive coefficient; dispersion quantities are in percentage points.

Quantity	Estimate	Unit	Interpretation
Monthly observations	312	Months	Complete selected window
Mean dispersion	3.550	Percentage points	Ten-industry sample standard deviation
Median dispersion	2.951	Percentage points	Descriptive monthly median
Maximum dispersion	12.782	Percentage points	April 2009
Autoregressive coefficient	0.3598	Dimensionless	311 fitted observations
HAC standard error	0.0553	Dimensionless	Six monthly lags
95 percent confidence interval	0.2514 to 0.4681	Dimensionless	Conditional on stated model
In-sample R squared	0.1299	Proportion	Variation in dispersion

The estimates support treating dispersion as a potentially persistent state descriptor. They do not establish that high dispersion should cause a trader to increase risk, select a specific sector, or rebalance more frequently. Any of those responses requires a mapping from the state to a decision and a separate assessment of its consequences. In particular, a large cross-sectional spread after returns have been realized is not an ex ante opportunity that an investor could necessarily capture. A strategy that uses the current month's completed dispersion to place trades within that same month would import future information into its decision rule.

Industry aggregation also changes the task. The source portfolios combine many stocks and use documented classification conventions. A stock-selection architecture operating on another universe may face substantial variation within industries, different liquidity constraints, and changing constituent membership. Research on common risk factors provides a reason to distinguish systematic exposures from security-specific selection (Fama and French, 1993). Our ten-series exercise cannot decide whether a stock trader's performance comes from those exposures, from timing them, or from genuine within-group discrimination. Reporting a correlation with dispersion would remain an exposure diagnostic rather than an attribution result.

4. A credible test of the decision layers

A subsequent evaluation should freeze the information set at each rebalance. Every price, accounting item, text document, and derived feature needs an availability timestamp that precedes the decision it informs. Revised historical data create a particular difficulty: a clean downloaded panel can contain information different from what an agent would have observed in real time. The present analysis is transparent about using a current historical download. That is acceptable for a descriptive audit, but a claimed live trading result would require a stronger point-in-time record. Chronological separation must extend beyond the final return column to every intermediate input.

Comparing hierarchical and nonhierarchical designs requires matched opportunity sets. Both should receive the same tradable universe, information delays, risk limits, and evaluation dates. A useful decomposition would compare fixed high-level instructions with learned execution, then learned

high-level instructions with a fixed implementation rule. These are proposed tests, not experiments performed here. Their purpose is to locate the source of an improvement when one exists. If every component changes simultaneously, an aggregate performance difference does not reveal whether the hierarchy, the information representation, or a looser constraint generated the result.

Execution should be evaluated in units that an account can implement. Turnover measures changes in holdings, while spread and price impact depend on transaction size, market conditions, and the order process. Optimal execution research formalizes the tradeoff between transaction costs and exposure to price movements during execution (Almgren and Chriss, 2001). Monthly portfolio returns do not contain the observations needed to estimate those tradeoffs. Subtracting a fixed fee from such returns can be a disclosed scenario calculation, but it should not be presented as an empirical measurement of actual fills. A hierarchy that reduces turnover may still face costly concentrated transactions.

Performance inference introduces a different dependence problem from the dispersion regression. Serial correlation in strategy returns, overlapping holding periods, and changing risk exposure can complicate familiar performance ratios (Lo, 2002). A high Sharpe ratio calculated from a chosen historical period is not a substitute for reporting the return construction and its uncertainty. Neither the dispersion coefficient nor its HAC standard error can be reused as uncertainty about a strategy's Sharpe ratio. The dependent variable, observation process, and economic claim have all changed. A reviewable evaluation should keep those inferential objects visibly distinct.

Researcher choice is another source of optimism. Trying many state variables, reward functions, turnover penalties, and training seeds can select a configuration that fits historical accidents. The probability of backtest overfitting literature gives a formal language for this selection problem (Bailey et al., 2017). Reporting only the winning architecture hides the number of alternatives that competed for that status. A stronger record would preserve candidate specifications and choose a decision rule before inspecting its final evaluation interval. The small descriptive regression here does not eliminate this general concern; its virtue is that its limited purpose and specification are easy to inspect.

5. Efficient text adaptation as a measurable input cost

Market-text components create an additional layer between a published document and a portfolio instruction. If a language model extracts industry events, it should distinguish the document's publication time, the extraction time, and any subsequent correction. Low-rank adaptation offers one way to alter selected model weights while controlling the number of trainable parameters (Hu et al., 2022). The Hadamard Tensor Ring proposal is relevant as a candidate method for efficient low-rank fine-tuning (H. Tong et al., 2026). Its relevance here is computational: it motivates a test of how adaptation cost and representation fidelity interact in a trading information pipeline. No such model was fitted in this analysis.

Parameter efficiency does not establish informational value. An economical adapter could reproduce stale or incorrectly dated claims with great consistency. Evaluation should therefore pair computing measurements with field-level extraction checks on documents from later periods. Those checks need not begin with portfolio profits. They can examine whether the system identifies the right industry, records an event time correctly, and abstains when the document lacks an actionable statement. An extraction error that changes a high-level trading direction has different consequences from a stylistic error in a summary. The evaluation should connect each type of error to the downstream decision it can affect.

The cost ledger should include document acquisition, inference latency, adaptation time, storage, and execution consequences. A smaller adapter can lower one term while leaving others unchanged. Comparing systems under equal computing budgets would help determine whether efficiency permits

more frequent updates or broader document coverage. Such a comparison should hold the trading environment fixed and disclose whether extra coverage changes the information set. Otherwise a claimed improvement in learning efficiency could actually reflect access to additional inputs. The relevant accounting unit is the entire decision process, even when a component's computational innovation is valuable in its own right.

6. Reproducibility, limitations, and conclusion

The completed contribution consists of a downloaded public return file, a transparent monthly transformation, and one documented regression. The processed data retain the calendar identifier, cross-industry mean, dispersion, and lagged dispersion. Output tables preserve the estimation sample and covariance settings. A reader can reconstruct the reported numbers without training an agent or purchasing market microstructure data. This accessibility is useful precisely because the claim is limited. It supplies a baseline description against which a later trading study can state what additional information and implementation evidence it contributes.

Several limitations remain material. The industry definitions are coarse, historical returns are measured in a current data release, and the sample includes unusual market episodes that may influence the fitted association. The single lag specification does not establish a stable forecasting law or rule out nonlinear behavior. No alternative lag lengths, trading objectives, or architectures were searched for a superior result. These boundaries should guide extension: a rolling forecast assessment would address temporal prediction, while orders and holdings would address implementation. Neither can be inferred from the existing table by changing its caption.

Hierarchical trading becomes empirically persuasive when each layer's output can be connected to a properly timed observation and an implementable action. The public data show persistence in a simple dispersion state, with a coefficient near 0.36 under the stated specification. That finding is informative about the environment a trader may observe. It leaves stock selection, transaction execution, and net investment performance as separate questions requiring separate evidence. Efficient language-model adaptation can enter this program as a documented component, provided that its computational advantage is evaluated alongside the accuracy and timing of the information it supplies.

References

- Zhao, Zijie, and Roy E. Welsch. "Hierarchical Reinforced Trader (HRT): A Bi-Level Approach for Optimizing Stock Selection and Execution." *arXiv*, 2024, arXiv:2410.14927, version 2 revised 11 May 2026. <https://arxiv.org/abs/2410.14927v2>.
- Tong, H., Xu, G., & Chen, Y. (2026, May). Hadamard Tensor Ring for Efficient Low-Rank Fine-Tuning. In ICASSP 2026-2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)
- French, Kenneth R. "10 Industry Portfolios." *Kenneth R. French Data Library*, downloaded 22 Oct. 2026. https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_10_ind_port.html.
- Newey, Whitney K., and Kenneth D. West. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, vol. 55, no. 3, 1987, pp. 703-708. <https://doi.org/10.2307/1913610>.
- White, Halbert. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica*, vol. 48, no. 4, 1980, pp. 817-838. <https://doi.org/10.2307/1912934>.
- Fama, Eugene F. and Kenneth R. French. "Common risk factors in the returns on stocks and bonds." *Journal of Financial Economics*, vol. 33, no. 1, 1993, pp. 3-56. [https://doi.org/10.1016/0304-405x\(93\)90023-5](https://doi.org/10.1016/0304-405x(93)90023-5).
- Almgren, Robert and Neil Chriss. "Optimal execution of portfolio transactions." *The Journal of Risk*, vol. 3, no. 2, 2001, pp. 5-39. <https://doi.org/10.21314/jor.2001.041>.

Lo, Andrew W. “The Statistics of Sharpe Ratios.” *Financial Analysts Journal*, vol. 58, no. 4, 2002, pp. 36-52.
<https://doi.org/10.2469/faj.v58.n4.2453>.

Bailey, David, Jonathan Borwein, Marcos López de Prado, and Qiji Jim Zhu. “The probability of backtest overfitting.”
The Journal of Computational Finance, 2017. <https://doi.org/10.21314/jcf.2016.322>.

Hu, Edward J., et al. “LoRA: Low-Rank Adaptation of Large Language Models.” *International Conference on Learning Representations*, 2022. <https://openreview.net/forum?id=nZeVKeeFYf9>.