ADAPTIVE LEARNING UNDER SIGNATURE-GENETIC ALGORITHM

A PREPRINT

Ramy Sukarieh
Economics Department
Paris 1 Panthéon-Sorbonne University
FR, Paris 12 Pl. du Panthéon
ecoramy@gmail.com

Raphael Douady
Economics Department
Paris 1 Panthéon-Sorbonne University
FR, Paris 12 Pl. du Panthéon
rdouady@gmail.com

June 5, 2025

ABSTRACT

We employ polymodels and long short-term memory (LSTM) models to identify a small set of optimal explanatory variables for predicting stock returns in the following period. In a previous study, this objective was approached using a signature genetic algorithm in a multivariate setting. In this paper, we address the same goal from a univariate perspective. We begin with the polymodels approach [1], where each factor's returns is individually evaluated against stock returns using a fourth-order polynomial regression. This constitutes the univariate phase of the polymodel methodology. From this analysis, the five factors with the highest out-of-sample adjusted R^2 are selected and combined to produce a stock return forecast. Polymodels are particularly effective in capturing extreme movements while limiting overfitting [2], as each factor is regressed independently against stock returns. Subsequently, we use signature features to independently extract information from each factor. These features are then evaluated using long short-term memory (LSTM) networks to determine their predictive power. The top-performing factors identified in this step are finally jointly used within an LSTM model to forecast stock returns.

***K*eywords** Polymodels · Signatures · Long Short-Term Memory Models

1 Polymodels and Nonlinear Multi-Factor Models

Nonlinear factor models have significantly advanced the understanding of asset pricing by capturing complex relationships between valuation/economic variables and stock returns. The SigGA model we constructed in the previous paper yielded portfolios' information ratios at least twice as good as the benchmark S&P 500. Roger G. Clarke, de Silva, and Thorley (2024) [3] examine the nonlinear relationships between stock characteristics and returns for five equity market factors: value, momentum, small size, low beta, and profitability. Analyzing data from 1964 to 2023 for the largest 1,000 US stocks, the authors find that accounting for nonlinearity enhances information ratios for factor portfolios. The capital asset pricing model (CAPM), introduced by Sharpe (1964) and Lintner (1965), established a linear relationship between expected returns and systemic risk (beta). Despite critiques, CAPM remains widely used due to its simplicity. The arbitrage pricing theory (APT), developed by Ross (1976), emerged as an alternative to CAPM, incorporating multiple factors under a no-arbitrage framework. The Fama-French three-factor model (1993) extended APT by including size and value factors alongside market risk. It became a cornerstone in asset pricing literature. Subsequent enhancements include the Carhart's four-factor model (1997), adding momentum and the Fama-French five-factor model (2015), which incorporates profitability and investment factors [4]. Bandurski and Postek (2023) [5] found statistically significant nonlinear dependencies in the modified Fama and French three-factor model. Zhu et al. (2018) [6] introduce the groupwise interpretable basis selection (GIBS) algorithm to estimate an adaptive multi-factor asset pricing model. The authors demonstrate that this approach outperforms traditional linear factor models, such as the Fama-French five-factor model, in both in-sample fitting and out-of-sample predictive performance [1].

1.1 The Polymodels Description

The mathematical concept of polymodels generalizes the one-dimensional univariate regression to multidimensional, multivariable analysis. Theoretically, a polymodel entails three steps: a) A nonlinear basis formulation. b) A supervised learning regression of each risk factor against the dependent variable, e.g, stock returns [2]. c) A statistical measure, typically an evaluation function such as expectation or variance (value-at-risk/expected shortfall). Let:

- $y \in \mathbb{R}^T$: observed response variable (e.g., stock return vector), over T observations
- $x^{(j)} \in \mathbb{R}^T$: the j-th explanatory variable (factor), for $j = 1, \ldots, p$
- $\hat{y}^{(j)}$: the predicted response from factor j

Polymodels treat each explanatory factor $x^{(j)}$ as a univariate nonlinear function. We choose the polynomial form:

$$Y_t = P_n^{(j)}\left(X_t^{(j)}\right) \tag{1}$$

Where $P_n^{(j)}(x)$ is a polynomial of degree n for factor j, of the form (drop subsript t to simplify notation):

$$P_n^{(j)}(x) = \beta_0^{(j)} + \beta_1^{(j)}x + \beta_2^{(j)}x^2 + \cdots + \beta_n^{(j)}x^n = \sum_{i=0}^{n} \beta_i^{(j)} x^i \tag{2}$$

The polynomial 1 is estimated by solving a regression problem, using the Vandermonde matrix $\mathbf{X}^{(j)}$ with rows containing $[1, x, x^2, \ldots, x^n]$

$$X^{(j)} = \begin{bmatrix} 1 & x_1^{(j)} & \left(x_1^{(j)}\right)^2 & \cdots & \left(x_1^{(j)}\right)^n \\ 1 & x_2^{(j)} & \left(x_2^{(j)}\right)^2 & \cdots & \left(x_2^{(j)}\right)^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_T^{(j)} & \left(x_T^{(j)}\right)^2 & \cdots & \left(x_T^{(j)}\right)^n \end{bmatrix} \in \mathbb{R}^{T \times (n+1)} \tag{3}$$

So the predicted response from factor j is:

$$\hat{y}^{(j)} = P_n^{(j)}\left(x^{(j)}\right) + \epsilon^{(j)} = \mathbf{X}^{(j)}\beta^{(j)} + \epsilon^{(j)} \tag{4}$$

However, this formulation is ill-conditioned due to co-linearity in the powers of x, and autocorrelation in the variable type (non-traded factors). We must obtain orthogonal bases before estimation. We are interested in polynomials of degree $n = 4$ enough to capture the skewness and kurtosis statistics (3^{rd} and 4^{th} moments). A basis that can be easily inverted, integrated, and differentiated is paramount to building multivariate models from univariate models. We orthogonalize each risk factor by the Hermite probabilist function 7. The resulting basis $\varphi_n(\cdot) := \mathcal{H}_n(\cdot)$ consists of orthogonal vectors $[H_0, \ldots, H_n]$ that correspond to the matrix Q in $V_x = QR$ Gram-Schmidt decomposition, with dimension $(T \times (n+1))$, where T is the total number of observations, and $(n+1)$ columns of a constant and polynomial terms:

$$\mathcal{H}_n(x) = [H_0 \quad H_1 \quad \ldots \quad H_n] \tag{5}$$

[1] This section was developed with the assistance of ChatGPT (OpenAI), which provided summaries and structured overviews of academic papers based on publicly available sources.

[2] Scaled and standardized returns

Definition 1.1. A sequence of polynomials $\{\varphi_i(x)\}_{i=0}^{n}$ in one variable and real coefficients where deg $\varphi_n = n$, and $\varphi_0 \neq 0$, is orthogonal on $[a,b]$ with respect to weight function $w(x) > 0$, if the inner product $\langle \varphi_n, \varphi_m \rangle = 0$ for $n \neq m$ or $\langle \varphi_n, \varphi_m \rangle = h_n$ nonzero constant, and $\delta_{nm} = 1$ otherwise. If $h_n = 1 \; \forall n$, the set of functions is ***orthonormal*** $\forall \, m, n = 1, 2 \ldots, n \leq m-2$. The Hermite orthogonal polynomials defined on $[a,b] = (-\infty, \infty), \; w(x) = e^{-x^2/2}$:

$$\langle \varphi_n, \varphi_m \rangle = \int_a^b \varphi_n(x)\varphi_m(x)w(x)dx = \int_{-\infty}^{\infty} e^{\frac{-x^2}{2}} H_n(x)H_m(x)dx = \begin{cases} n!\sqrt{2\pi}\delta_{nm} \neq 0 & n = m \\ 0 & n \neq m \end{cases} \tag{6}$$

The orthogonal **Hermite basis** is generated by:

$$H_n(x) = (-1)^n e^{\frac{x^2}{2}} \frac{d^n}{dx^n}(e^{\frac{-x^2}{2}}) = \left(x - \frac{d}{dx}\right)^n .1 \qquad\qquad \varphi_n(x) = \frac{1}{(n!\sqrt{2\pi})^{1/2}} H_n(x) \tag{7}$$

For $w(x) = e^{-x^2/2}$, $[a,b] = [-\infty, \infty]$, the first five orthogonal Hermite polynomials $H_n(x)$ are shown below. Divide by $\sqrt{n!}$ to obtain the orthonormal Hermite, which improves numerical stability and removes correlation between basis features:

$$H_{e0}(x) = 1 \tag{8}$$
$$H_{e1}(x) = x \tag{9}$$
$$H_{e2}(x) = (x^2 - 1) \tag{10}$$
$$H_{e3}(x) = (x^3 - 3x) \tag{11}$$
$$H_{e4}(x) = (x^4 - 6x^2 + 3) \tag{12}$$

1.2 Polymodels Estimation with Regularization and Cross-Validation

For each factor j, define the design matrix $\phi^{(j)} \in \mathbb{R}^{T \times (n+1)}$ where each column is the Hermite polynomial $H_i^{(j)}[x_t] \in \mathcal{H}$, evaluated at $t = 1, \ldots, T$, and $\alpha^{(j)} \in \mathbb{R}^{n+1}$ as vector of coefficients. The orthogonal polynomial basis expansion:

$$\hat{y}_t = \sum_{i=0}^{n} \alpha_i^{(j)} H_i^{(j)}(x_t) + \epsilon_t^{(j)} \tag{13}$$

where:

$$\mathcal{H}^{(j)} = \begin{bmatrix} H_0^{(j)}(x_1) & H_1^{(j)}(x_1) & \cdots & H_n^{(j)}(x_1) \\ H_0^{(j)}(x_2) & H_1^{(j)}(x_2) & \cdots & H_n^{(j)}(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ H_0^{(j)}(x_T) & H_1^{(j)}(x_T) & \cdots & H_n^{(j)}(x_T) \end{bmatrix} \tag{14}$$

$$y^{(j)} = \phi^{(j)}\alpha^{(j)} + \epsilon^{(j)} \tag{15}$$

The least squares estimator $\hat{\alpha}^{(j)}$ minimizes the residual sum of squares and satisfies the normal equations. Ridge regression adds an ℓ_2-penalty to the least squares problem, helping control overfitting by penalizing the large coefficients:

$$\min_{\alpha^{(j)}} \left\| y - \phi^{(j)}\alpha^{(j)} \right\|^2 + \lambda \left\| \alpha^{(j)} \right\|^2 \tag{16}$$

The ridge solution of the regularized least squares estimator is $\hat{\alpha}_\lambda^{(j)}$, where $\lambda > 0$ is a regularization hyperparameter, and Ω is a diagonal square matrix of weights proportional to the degree of the polynomial is given by:

$$\hat{\alpha}_\lambda^{(j)} = \left(\phi^{(j)\top}\phi^{(j)} + \lambda\Omega\right)^{-1} \phi^{(j)\top} y \tag{17}$$

Partition the sample size into five sub-samples (folds), where each sub-sample consists of a training (in-sample) and a testing (out-of-sample) set denoted by ν_i and κ_i, respectively. Re-estimate the polynomial coefficients $\hat{\alpha}_\lambda^{(j)}$ using the training data, and select those associated with the lowest out-of-sample mean squared error (MSE). This is called the cross-validated out-of-sample estimation, where the error is given by:

$$\eta_\kappa^{(j)}(\lambda) = \frac{1}{\ell} \sum_{i=1}^{\ell} \left[\frac{y_\kappa - \phi_\kappa^{(j)}(x) \hat{\alpha}_\nu^{(j)}(x, \lambda)}{1 - \mathrm{tr}^{(j)}[I_\kappa]/\ell} \right]^2 \tag{18}$$

I_κ is a correction matrix that accounts for model complexity and regularization, given by:

$$I_\kappa(x, \lambda) = \phi_\kappa^{(j)} \left(\phi_\kappa^{(j)\top} \phi_\kappa^{(j)} + \lambda \Omega \right)^{-1} \phi_\kappa^{(j)\top}$$

2 Polymodels Application

we aim to characterize the signal properties associated with the distinct types of the selected five factors—namely, fundamental variables, macroeconomic-financial indicators, and market-based ETFs—while also evaluating the incremental benefits of the optimal degree of polynomial expansion within the polymodel system. Consider a stochastic process X_t that satisfies the following equation, where $V(x)$ is a potential function, e.g., $V(X) = \frac{1}{4}X^4 + \frac{1}{2}X^2$. Then:

$$dX_t = -\nabla V(X_t)\, dt \tag{19}$$

$$\nabla V(X) = X^3 - X$$

In the polynomial generalization, take $V(x)$ as the Hermite basis in [7], equation 19 becomes a polynomial mean-reverting process, where the drift is a polynomial function pulling X_t toward a complex equilibrium. Let:

$$V(x) = \sum_{k=1}^{n} a_k x^k \quad \Rightarrow \quad \nabla V(x) = \sum_{k=1}^{n} k a_k x^{k-1}$$

$$dX_t = -\left(\sum_{k=1}^{n} k a_k X_t^{k-1} \right) dt \tag{20}$$

The polymodel framework provides both theoretical rigor and a benchmark for comparison with the analysis presented in the previous chapter [7]. While polymodels and the signatures from rough path theory differ in their mathematical formulation, they are conceptually aligned in capturing nonlinear dependencies. The present polymodel employs univariate polynomial expansions using orthogonal bases—such as Hermite (HE) or Gram-Schmidt (GS)—over discrete time steps. In contrast, the signature transform provides a functional basis for continuous paths, where linear forms on the signature serve as the analog of polynomial functions for time-series data embedded in path space [8].

2.1 Anomalous Alpha Extraction

Ex-ante (predicted) returns are either positive, negative, or zero. Typically, positive ex-ante returns indicate overweight positions in stocks and vice versa. We define an anomalous alpha from the inverse signal of systematic model predictions across time. The signals are derived from fundamentals or other non-price and price features (e.g., fundamental ratios, macro/financial, market), suggesting a model-inverted value strategy of a contrarian effect relative to the "sign" of the ex-ante returns. This mis-specified predictive model reveals structural inefficiencies, where the signal's inverse captures persistent mean-reversion or overreaction. If the inversion consistently works, it leads to an *informational inefficiency exploitation* strategy. Such a strategy exploits informational inefficiencies in model-predicted returns by systematically inverting return signals. The alpha generated seems to come from a different signal space—potentially related to market inefficiencies the model identifies (and inverts). This suggests the presence of an anomalous alpha and challenges traditional interpretations of model accuracy.

Notably, the alpha signals generated by the model exhibit an inverse return-ordering effect: ex-ante negative predicted returns are systematically associated with portfolios that subsequently outperform the market, while ex-ante positive predicted returns lead to underperforming portfolios. This pattern implies that the predictive model may be uncovering a form of mispricing or transient inefficiency, where contrarian interpretation of the signals (i.e., inverting their ranking) yields superior performance.

Importantly, this behavior does not imply a strict anti-momentum or anti-market stance in the statistical sense—there is no negative beta or correlation, Panel (d) Figure 6 Appendix G. Rather, the strategy may be characterized as a low-factor-exposure, signal-inversion alpha strategy, whose construction leverages nonlinear patterns via orthogonal polynomial regressions. Its distinct return dynamics set it apart from conventional factor portfolios and momentum-driven allocations.

The investment strategy developed in this study is analyzed against both the broad market index and the momentum factor [3]. Specifically, the top-performing portfolio, the (Gram-Schmidt) long polymodel portfolio in Table 10 Appendix G, has an estimated annualized return of 40% and a risk of 18% with a tracking error of 6% relative to the IVV. The strategy's performance harvests alpha and exhibits dominance over the index, especially when comparing the CDFs.

To construct the portfolios in Table 10, we run polymodels on each factor independently, with approximately 200 factors per stock using a universe of 101 stocks. The factors and stocks used in this study are listed in Appendices C, D, E, F. All stocks are fitted to a polynomial equation of degree 4 under two orthogonal bases—HE and GS. The signal, ex-ante predicted return, is computed based on five factors with the highest (positive) out-of-sample adjusted R^2. Subsequently, stocks are ranked based on their signals from top (positive) to bottom (negative), where the top 18 stocks equally weighted form the short portfolio (sell or underweight), and the bottom 18 stocks also equally weighted form the long portfolio (buy or overweight).

Definition 2.1. $\forall\, Y_i,\ i = 1, \dots, I \in \mathbb{N},$ and selected ordered subsets $\mathcal{J}_i \subseteq \mathcal{J} \in X : X^{(J_i)} := Y_i,\ j = 1, \dots, J \in \mathbb{N},$ denote $Z(\cdot)$ the Hermite (8) or monomials x^n transformation centered around the empirical mean $\bar{x} = (1/T)\sum_{t=1}^{T} x_t,$ and $\hat{\alpha}_\lambda^{(j)}$ the estimated coefficients defined in (17) and (18) of regressions $j \in \mathcal{J}$ pertaining only to the top-k factors such that:

$$\text{adj-}R_j^2 > 0 \quad \text{and} \quad \text{adj-}R_j^2 \in \text{Top-}k \text{ among all factors, (e.g., k = 5)}$$

The signal is defined by:

$$s_{T+1} = \Phi(x^{(1)}, \dots, x^{(k)}) - y_{T+1} \tag{21}$$

Where $\Phi = (1/K)\sum_{k=1}^{K} Z^{(k)}\alpha^{(k)}\ \forall t = 1, \dots, T$, and observation $(T+1) \notin \Phi$:

$$\bar{\hat{y}}_{T+1} = \Phi(x^{(1)}, \dots, x^{(J)}) = \begin{cases} Z_n^{(1)}(\bar{x}, \kappa)\, \hat{\alpha}_\kappa^{(1)}(x, \lambda) & \text{for } x^{(1)} \in \mathcal{J}_1 \\ Z_n^{(2)}(\bar{x}, \kappa)\, \hat{\alpha}_\kappa^{(2)}(x, \lambda) & \text{for } x^{(2)} \in \mathcal{J}_2 \\ \quad\vdots & \\ Z_n^{(k)}(\bar{x}, \kappa)\, \hat{\alpha}_\kappa^{(k)}(x, \lambda) & \text{for } x^{(k)} \in \mathcal{J}_5 \end{cases} \tag{22}$$

Panel (a) Figure 1 shows the probability density function (PDF) of the polymodels long portfolio Table 10 derived under GS basis versus the IVV index. Panel (b) compares the long portfolio to the momentum MTUM index. The portfolio's PDF consistently delivers higher ex-post returns, as seen from the right-shifted PDF and CDF. The broader PDF signals higher return dispersion (higher risk), but with more upside. The CDF of the portfolio is below the index for most of the range, and extends more to the right, indicating more mass in the higher (positive) return range [4]. Given this distributional difference, it's likely that the portfolio is manifesting superior risk-adjusted performance rather than mimicking the index. With a Sharpe ratio of 2.23 greater than 0.83 (IVV) and 1.41 (MTUM), the long portfolio might be harvesting idiosyncratic alpha.

Panel (c) Figure 1 shows the short (or underweight) polymodels portfolio using the Hermite basis versus the IVV index. It has similar characteristics to the long portfolio but directionally opposite, noticeably fatter left tail, and riskier than the benchmark ETF, with a tendency toward more extreme losses. This is typical for short-side strategies, where the asymmetry of return potential (limited upside, large downside) naturally shows up as left-skewed and fat-tailed.

Panel (d) Figure 1 shows that the long portfolio has lower peak density at the mean, but shifted to the right, suggesting higher average returns, better upside potential, and lower downside risk. The narrower (less spread) PDF, right-skewed, with a fatter right tail and thinner left tail implies more frequent large gains and less frequent extreme losses than the index, For example, looking at the CDF, a 25% return corresponds to roughly 93th percentile of the portfolio, but nearly 100th percentile for the index.

[3] IVV: iShares Core S&P 500 ETF, and MTUM: iShares MSCI USA Momentum Factor ETF

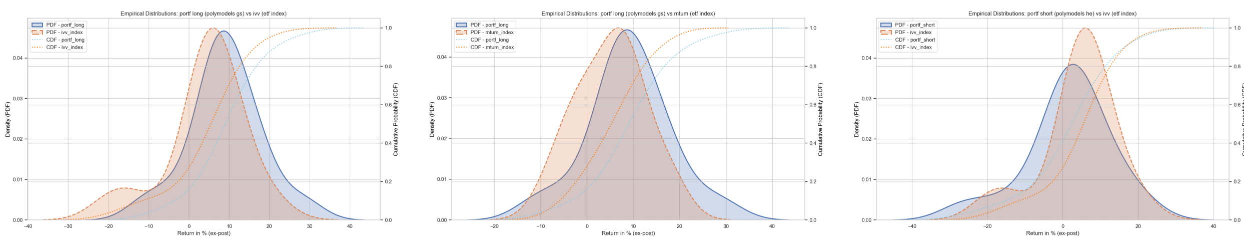
[4] Portfolio's CDF below index for most of the range $\rightarrow$ accumulates probability mass slower $\rightarrow$ higher return distribution.

Panel (a) Figure 6 shows the out-of-sample signal forecast used to construct the long portfolio (deg4) discussed above. Stocks are selected based on the most positive and most negative signal forecast at $(T+1)$ with equal weights, such as:

$$W_{T+1} = \begin{cases} W_{T+1} > 0 & \text{if } s_{T+1} \leq \tau_{-} < 0 \\ W_{T+1} < 0 & \text{otherwise,} \end{cases} \tag{23}$$

Where τ_{-} is the threshold above which the signal is disregarded for the negative signal case, and vice versa. A positive weight indicates a stock buy order (long), and a negative weight means a stock sell order (short). Recall, one motivation for this section is to obtain coefficient estimates and compute signal forecasts only from factors that have positive out-of-sample adj-R^2.

Panel (b) Figure 6 shows the averages over time of the five factors (per stock) with the highest out-of-sample adjusted R^2, with an overall average adj-R^2 of 0.35. Panel (c) Figure 6 illustrates the out-of-sample hit ratio of 58%, which adds statistical rigor to the strategy, where most of the hit ratios fell within the $51\% - 59\%$ range, three within the $71\% - 85\%$ range, and four within the $44\% - 48\%$. The Tables 13 and 14, in Appendix H, show the values of the plots illustrated in Figure 6.



(a) Long Portf. vs. IVV (b) Long Portf. vs. MTUM (c) Short Portf. vs. IVV

Tail Count	% Return < -15%	% Return > 25%
portf_long	0.00%	6.25%
ivv_index	6.25%	0.00%
mtum_index	0.00%	0.00%

(d) Tail proportion comparison: portf_long vs. IVV, MTUM

Figure 1: Performance plots and tail risk summary.

3 A Conceptual Introduction to Adaptive Methods

3.1 Rethinking Asset Modeling

From Stochastic Processes to Data-Driven Adaptability A common practice in both academic and industry practices is to assume that an asset price follows a specified stochastic process—selected for its mathematical tractability or economic interpretability, and often either linear or nonlinear. The former might be a diffusion model such as the log-transformed geometric Brownian motion (GBM), used in the Black-Scholes framework for option pricing [9], while the latter might be a mean-reverting process like the Ornstein–Uhlenbeck or Cox–Ingersoll–Ross (CIR) model for interest rates and stochastic volatility [10], see example 1. These frameworks yield theoretical implications that can be empirically tested, and from the implied distributions—such as probability density functions (PDFs) and cumulative distribution functions (CDFs)—a range of risk measures can be derived, including Value-at-Risk (VaR) and expected shortfall [11]. A fundamental implication of this paradigm is the uniformity it imposes: all assets within the same class (e.g., equities) are often modeled using the same type of stochastic process, regardless of their individual characteristics or the dynamic regimes in which they operate. This can lead to misestimation of return and mispricing of risk, especially under changing market conditions, structural breaks, or asset-specific anomalies.

In contrast, the Signature Genetic Algorithm (SigGA) framework offers a flexible, data-driven, and adaptive alternative. By integrating signatures from rough path theory [12, 13]—a mathematically rigorous way to encode the order and structure of time series paths—with the model selection power of evolutionary algorithms[14, 15], SigGA shifts the modeling burden from rigid a priori assumptions to empirical learning. Rather than assuming a fixed functional form, SigGA evaluates combinations of explanatory variables and signature terms to discover whether linear, nonlinear, or hybrid models better capture an asset's behavior. This adaptive capability enables the model to detect and respond to different regimes. For instance, a particular equity might exhibit a primarily linear relationship to macroeconomic indicators during periods of low volatility, but shift to nonlinear dynamics when markets become turbulent. The model's use of supervised learning and evolutionary search allows it to adaptively identify such conditions, selecting appropriate model structures accordingly.

In practice, this approach challenges the "one-size-fits-all" mindset of classical modeling. While models like Black-Scholes are elegant and analytically tractable, they may fall short in capturing real-world complexities such as volatility clustering, regime switches, and path-dependent features—characteristics that rough path signatures are particularly adept at encoding. By detecting such conditions, SigGA enables more effective and tailored financial modeling, which may improve forecasting and help uncover informational inefficiencies. As financial systems grow increasingly complex and data-rich, the SigGA model represents a promising step toward data-driven model selection and individualized asset behavior analysis.

As a result, the SigGA framework enables the creation of a new sub-classification layer within traditional asset types—such as equities—by introducing a functional taxonomy based on linearity and nonlinearity. For instance, beyond conventional groupings like industry sectors, stocks can be further categorized according to the functional form that best describes their behavior. These classifications emerge from data-driven criteria, where the appropriate modeling form (linear, nonlinear, or hybrid) is determined through the combination of data-slicing techniques and the fitness evaluation of supervised learning variable sets as guided by the SigGA process. Figure 2 and Table 17 Appendix I compare traditional financial modeling to SigGA approach.

Example 1. For example, the GBM model assumes that asset prices S_t evolve according to:

$$dS_t = \mu S_t \, dt + \sigma S_t \, dW_t,$$

which is nonlinear in price but results in log-normality due to the linearity of the logarithmic process. Nonlinear mean-reverting models, such as the CIR model:

$$dX_t = \kappa(\theta - X_t) \, dt + \sigma \sqrt{X_t} \, dW_t,$$

capture state-dependent volatility and are capable of reflecting the empirical features of financial time series, such as heteroskedasticity and persistence[5]

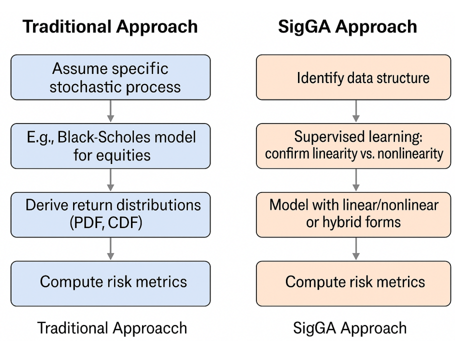


Figure 2: Traditional vs. SigGA Approaches

3.2 Motivation for SigGA as an Adaptive Mechanism

Much of the information learned in Section 3.3 below has already been inferred from the results of the SigGA model developed in Sukarieh[6] et al. [7]. Indeed, Table 1 below of the SigGA output by signature levels (degree) shows that both the slide and dyadic window methods point to degree 2 as optimal. The sliding windows method's adj-R^2 of D2 outperformed D3 by about 5% of the times, and D1 by about 2% of the times. Similarly, for the dyadic windows method, these statistics go up to 7% and 15% in favor of D2 over D3 and D1, respectively. On the contrary, the expanding windows method favors D1 over D3 and D2 by a frequency of 2% and 8%, respectively.

[5]Developed with assistance of ChatGPT (OpenAI); provided summaries, structured overviews of the adpative learning section.
[6]Sections 3.2 & 4.2: Learning Poly Order from SigGA. Data Grouping: Fundamental, Macroeconomic, Market Integration.

Window \ Sig Level	D3 > D1	D3 > D2	D2 > D1
Slide	54	45	52
Expand	48	57	42
Dyadic	61	43	65
ave.	**54**	**48**	**53**

Table 1: Signature levels [D1, D2,D3]

Figure 31 in Appendix [7] presents the adjusted R^2-sorted plots of the three windowing methods across signature levels 1, 2, and 3. These visualizations confirm the patterns observed in Table 1. Notably, the performance of degrees 2 and 3 appears nearly indistinguishable—an observation consistent with the findings in Section 3.3, particularly in relation to the out-of-sample adjusted R^2, as well as the forecast signals and hit ratios in Figures 6.

These results suggest that assuming nonlinearity alone may be insufficient for modeling this dataset; in fact, incorporating linear structures may be necessary under certain conditions. Such conditions could depend on the choice of windowing method or data-slicing technique employed in model coefficient estimation. Additionally, when present, the nonlinearity of second-order dominates higher orders, as we have deduced in Section 3.3, and Tables 12,13 Appendix H, with the caveat that greater than 2 degrees are not irrelevant. While the coexistence of linear and nonlinear structures is not inherently surprising, prior awareness of this interplay can assist analysts in detecting informational inefficiencies and selecting more appropriate or effective modeling strategies.

Given a SigGA output, the rules for a set of data spans a linear or nonlinear space can therefore be conditioned on the data-slicing mechanics, ***windowing methods in signature analysis***. We consider three primary data slicing methods used in conjunction with signature features from rough path theory: **expanding**, **sliding**, and **dyadic window**. Each method emphasizes different structural aspects of the underlying path and is suited for distinct modeling objectives.

Expanding Window: This method starts from a fixed initial point and gradually expands the time interval forward. At each time step, the model is trained on all data up to that point. In signature analysis, this allows the signature to encode increasingly rich path features over a growing interval, capturing long-term dependencies and global structures. However, it may overweight early patterns and adapt slowly to structural changes. It is often preferred in applications with assumed stationarity or slow regime drift, such as macroeconomic modeling.

Sliding Window: The sliding window maintains a fixed-size interval that moves forward through time. At each step, the model is retrained on the most recent data segment. When computing signatures over this fixed window, the method emphasizes recent path geometry and local behavior. This makes it effective for capturing short-term patterns and adapting to regime shifts, although it may ignore long-term dependencies. It is commonly used in financial time series forecasting and environments requiring adaptive behavior.

Dyadic Window: This method recursively splits the time series into dyadic intervals (halves, quarters, etc.) and computes signatures over each nested sub-interval. It facilitates multi-scale analysis and enables the capture of nonlinear, fractal, or irregular structures. Dyadic windows are particularly suited for rough path modeling, where signatures need to reflect features at multiple resolutions. Despite being computationally heavier, this method is powerful for model selection, detection of roughness, or regime shifts in high-frequency or irregular data.

The learning adaptability of SigGA—an integration of rough path signatures with genetic algorithms for dimension reduction—can be articulated through three main implementation features:

1. **Model structure flexibility**: It facilitates, and often enhances, the decision-making process between linear and nonlinear representations. SigGA facilitates data-driven decisions between linear and nonlinear representations.

2. **Joint variable selection**: It jointly identifies sets of univariate or multivariate supervised-learning variables under both linear or nonlinear forms.

3. **Temporal resolution control**: It enables both global and local characterizations of the data, allowing flexible modeling choices that capture long-term trends or short-term irregularities[8].

[7] Sukarieh et al. Appendix K: Comparison of Window Methods Across Degrees

[8] Developed with the assistance of ChatGPT (OpenAI); provided summaries and structured overviews of the window methods.

Method	Memory Scope	Nonlinearity Detection	Adaptivity	Use Case Example
Expanding Window	Long-term (global)	Limited	Low	Macro trend estimation
Sliding Window	Short-term (local)	Moderate	High	Short-term financial forecasting
Dyadic Window	Multiscale	High	Medium	Roughness detection, regime shifts

Table 2: Comparison of windowing methods for signature-based time series modeling.

3.3 Learning Nonlinearity from Polymodels

3.3.1 A Concrete Univariate Evidence

How do you decide on a stock's optimal representation as a linear or nonlinear form? Is there a need for a linear representation? And, how do you determine the optimal polynomial order? The obvious choice is to regress every stock against each factor using polynomial equations of degrees 1 through 4, then make factor selections based on the best out-of-sample adj-R^2, with a portfolio construction's signal computed as the average of the top five factors. Table 12 Figure 9 Appendix H show the results of such analysis. Portfolio deg2 outperforms portfolios deg1, deg3, and deg4 in terms of risk-adjusted return, as evidenced by the highest Sharpe ratio (2.67), lower downside risk, and a return distribution with heavier right tails in Figures 3, 4, 5. The empirical PDFs and CDFs further show that deg2 captures more upside potential, with its CDF dominating the others in the upper tail. While full second-order stochastic dominance is not observed across the entire return spectrum, partial SSD is evident in higher-return regions, supporting the conclusion that deg2 and, to a certain extent deg3 are the most attractive portfolios under a wide class of risk-averse preferences.

The out-of-sample R^2 and hit ratio Table 13 Appendix H Figure 6 of degrees 1 through 4 show little preference for deg2 portfolio over deg3, not much preference for deg4 over deg2, while deg1 portfolio outperforms when it comes to hit ratio. Furthermore, the pairwise win counts of how often each polynomial degree outperforms the others (based R^2) provided in Table 15, Figure 10 Appendix H show dominance of deg2 over deg3 and deg4 portfolios, but not deg1. In summary, in the context of our dataset, by running incremental-order polynomial regressions, summary Table 16 Appendix H, we learned the following:

- ***Second Order Polynomials***: sufficient to capture nonlinearity and optimal balance between risk and reward. Data and analysis exhibit a strong presence of second-order polynomials.
- ***Higher Order Polynomials***: help minimize overall risk and obtain better statistics for a smaller reward. Data and analysis do not support a strong presence of higher-order polynomials.
- ***Linear Form***: helps enhance return expectations' directions, i.e., indiscriminate use of linear or nonlinear models distorts signal forecasts. Linear forms might be necessary to identify informational inefficiencies and optimize overall return calculations. Data and analysis show linear tendencies.

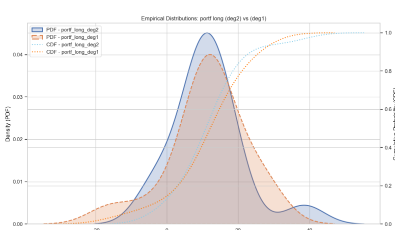


Figure 3: deg 2 vs. deg 1

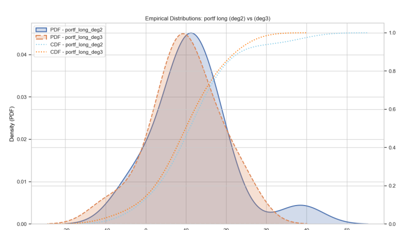


Figure 4: deg 2 vs. deg 3

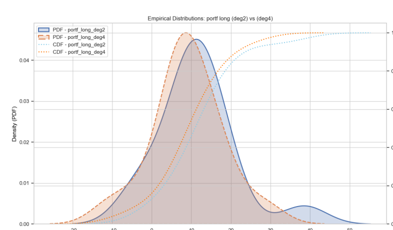
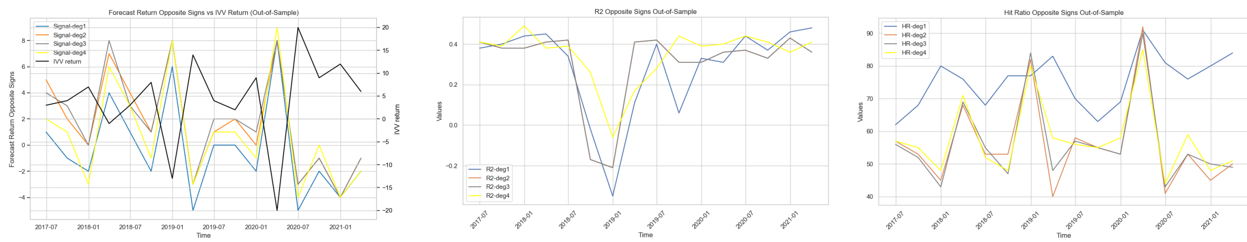


Figure 5: deg 2 vs. deg 4

Such informational inefficiencies or structural issues are illustrated by high hit ratios that do not readily translate into higher expected returns. Despite its failure to consider the joint correlation structure between factors, the aforesaid serves as a benchmark model when evaluating the statistical inferences deduced from the SigGA model.

3.3.2 By Feature Group

The analysis in this section is conducted on three distinct datasets to examine their inherent nonlinearity and temporal dynamics. The data, defined in Appendices C, D, E, capture key financial filings required by the U.S. SEC from



(a) Forecast signal (OOS) (b) R^2 (OOS) (c) Hit ratio (OOS)

Figure 6: Out-of-sample evaluation: forecast signal, adjusted R^2, and hit ratio.

publicly traded companies; company-specific financial ratios [9] known as "fundamental data, as well as country-specific macroeconomics indicators, and market[10] data:

- **Fundamental data** (e.g., balance sheets, income, and cash flow statements)
- **Macroeconomic data** (e.g., interest rates, inflation, business activities)
- **Market data** (e.g., ETFs such as gold, equities, bond indices, style, stize)

06-2017 : 03-2021	**deg1**	**deg2**	**deg3**	**deg4**
Sum Ret				
Macroeconomics	147	156	170	126
Fundamental	184	183	160	155
Market Data ETF	165	145	152	131
Combined	**182**	**183**	**157**	**144**
R^2				
Macroeconomics	-0.13	-0.17	-0.17	-0.10
Fundamental	-0.05	-0.13	-0.09	-0.08
Market Data ETF	-0.33	-0.43	-0.46	-0.45
Combined	**0.29**	**0.31**	**0.34**	**0.35**
Hit Ratio				
Macroeconomics	73	54	54	50
Fundamental	74	60	60	62
Market Data ETF	71	51	51	54
Combined	**75**	**56**	**56**	**58**

Table 3: Performance metrics across polynomial degrees

The results in Table 3, Table 18 Appendix J confirm what we have already observed in section 3.3.1. The Second-order polynomial is optimal. It provides the best balance of return, risk, and model fit. Deg3 follows closely but starts showing diminishing or inconsistent results. Deg4 shows diminishing returns and may not justify the additional complexity, while deg1 gives a strong hit ratio but with higher risk, lower explanatory power, and lower IR. Additionally, higher degrees lead to lower deviation (TE) from the benchmark, lower risk, higher return per unit of risk, and higher IR, Figure 8.

[9]Source: ratios/indicators from Quandl/NASDAQ.
[10]Source: yahoo finance to obtain ETFs

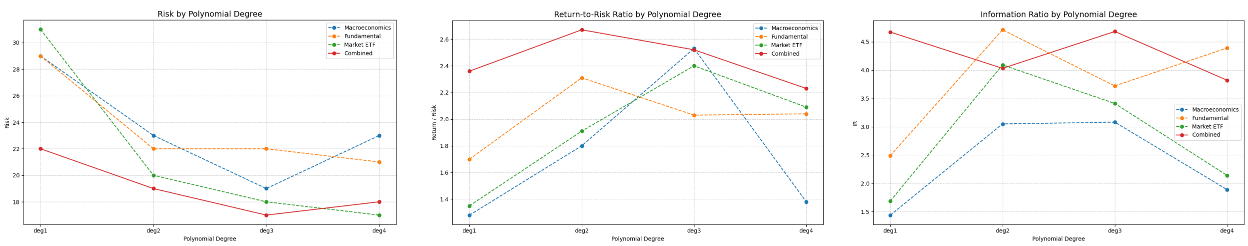
Information Inefficiencies Across Polynomial Degrees

The performance metrics across polynomial degrees indicate the presence of information inefficiencies within individual factor groups (macroeconomic, fundamental, and market data). Specifically, individual groups exhibit low or negative explanatory power, as reflected by negative R^2 values, along with inferior risk-adjusted returns and information ratios.

However, once these information streams are *aggregated* into a combined feature set, we observe a substantial improvement in predictive performance:

- The R^2 becomes strongly positive and increases monotonically across signature degrees.
- Risk-adjusted return measures, such as the Return-to-Risk ratio and the Information Ratio (IR), significantly outperform their individual counterparts.

This divergence suggests that each data group provides non-redundant, complementary information about future returns—implying that markets fail to efficiently synthesize these data streams in isolation. The resulting combined models, which for each stock are based on the best five factors by R^2, outperform naive separate feature-group or univariate baselines, highlighting exploitable inefficiencies in the informational structure of the market.



(a) Long Portf. Risk (b) Long Portf. Ret/Risk (c) Long Portf. IR

Figure 7: Ex-post, Annualized risk, ret/risk, information ratio.

3.4 Learning Nonlinearity from LSTM

To be continued, To be continued, To be continued...

3.5 Learning Nonlinearity from Signatures

Before discussing the results, we will go over the most important properties of signatures from rough path theory, especially as applied in contexts like time series analysis and machine learning. Levin et al.[16] state: "the signature of the path, has a universality that allows one to characterise the functional relationship summarising the conditional distribution of the dependent response". According to Futter at al.[17] the universality property states that linear functionals on the signature are dense in the space of continuous functions on compact sets of paths. The signature transform is a universal, non-parametric representation of a path. For continuous paths of bounded variation (and extended to rough paths) the signature uniquely determines the path up to tree-like equivalence. This makes the signature a complete invariant under such equivalence—meaning no other features are needed in theory. The universality (feature completeness) of the signature transform refers to its ability to approximate arbitrary continuous functions on path spaces, capturing essentially all relevant information about a path, and making it a powerful and canonical feature set for path-dependent functionals and for driving differential equations. This universality stems from the signature's ability to encode multiplicative interactions in paths and its factorial decay, which ensures robust convergence properties. The mathematical description of this property, grounded in rough path theory and kernel methods, is provided hereafter.

This sequence $S(X)$ of tensors captures all path information up to tree-like equivalence, which refers to paths that traverse the same shapes but with possibly different speed or retracing, or reparametrizations that preserve and encode all the essential information about the path's shape and ordering of events, but not necessarily the speed. More precisely, for two paths $X, Y : [0, T] \to \mathbb{R}^d$ of bounded variation, their signatures $\mathrm{Sig}(X)$ and $\mathrm{Sig}(Y)$ are equal if and only if X and Y are tree-like equivalent, a crucial property of faithfulness or uniqueness. This injectivity is a consequence of the Chen's identity, which relates the signature of concatenated paths, and the fact that the signature is invariant under parameterization. *Tree-like equivalence* means that the paths X and Y can be transformed into each other by adding or removing "tree-like excursions" (sub-paths that retrace themselves, contributing trivially to the signature). Tree-like equivalence is an equivalence relation, meaning it is reflexive, symmetric, and transitive [18]. The concept of tree-like equivalence extends Chen's theorems on the uniqueness [19] of iterated integrals for piecewise regular paths to finite-length paths. It essentially means that two paths are equivalent if they have the same "signature" or "Chen series", which is a certain representation of the path using iterated integrals. Mathematically, a path X is tree-like if it can be parameterized such that it starts and ends at the same point and its signature is trivial (i.e., $S(X) = (1, 0, 0, \dots)$).

The theorem implies that the signature provides a complete feature set for distinguishing paths up to tree-like equivalence. This is significant because it shows that the signature captures all relevant information about a path's behavior, making it a universal feature map for applications in stochastic analysis, machine learning, and data science (e.g., time-series analysis). Moreover, the signature satisfies a universality property in the sense that any continuous function on the space of paths (modulo tree-like equivalence) can be approximated by linear functionals on the signature; the linearity in the Signature Space. Nonlinear functionals of the path can often be well-approximated by linear functionals in signature space. This property is the theoretical basis for using linear models (e.g., linear regression, PLS) on signature features.

Invariance to Time parametrization: The signature is invariant to the speed at which the path is traversed (as long as the time ordering is preserved). This is crucial in finance and other domains where time scales can vary irregularly.

Homomorphism: The signature of a concatenation of two paths is the tensor product of the individual signatures: $S(X * Y) = S(X) \otimes S(Y)$. This algebraic property allows recursive and modular modeling of path behavior.

Truncation: Although the full signature is infinite-dimensional, in practice we truncate at a finite level n. The level corresponds to the maximum number of iterated integrals included. Truncating gives a computationally feasible and systematically improvable approximation.

Coordinate-Free & Basis-Agnostic: The signature transform is intrinsic to the path and does not depend on a particular coordinate system or basis. This makes it especially attractive for modeling complex multidimensional data streams.

Machine Learning: Signatures provide a compact and expressive feature set that captures the effect of temporal ordering, unlike classical features like mean, variance, etc. Efficient machine learning, signature methods often outperform traditional time-series feature extraction when temporal dependencies matter.

Definition 3.1. Let $X : [0, T] \to \mathbb{R}^d$ be a continuous path of bounded variation $BV(\mathbb{R}^d)$. The signature of X is defined as the infinite sequence of iterated integrals i.e., the collection of all iterated integrals of X over the interval $[0, T]$. Formally, for a path $X = (X^1, \dots, X^d)$, the signature is defined as:

$$S(X)_{[0,T]} = \left(1, \int_0^T dX_{t_1}, \int_{0<t_1<t_2<T} dX_{t_1} \otimes dX_{t_2}, \int_{0<t_1<\cdots<t_k<T} dX_{t_1} \otimes \cdots \otimes dX_{t_k}, \dots\right) \in T((\mathbb{R}^d)) \quad (24)$$

where each term is an element in the tensor algebra $T((\mathbb{R}^d)) = \bigotimes_{k=0}^{\infty} (\mathbb{R}^d)^{\otimes k}$, $(\mathbb{R}^d)^{\otimes k}$ is the kth tensor power of $\mathbb{R}^d$. For rough paths with finite p-variation, $p \geq 1$, the signature is defined via Lyons' Extension Theorem [20].

Definition 3.2 (Tree-Like)**.** A p-geometric rough path is tree-like if there exists a continuous function $h : [0, 1] \to [0, \infty]$ such that $h(0) = h(1) = 0$, and for all $j = 1, \dots, p$ and all $0 \leq s \leq t \leq 1$,

$$\left\| X^j_{s,t} \right\|^p \leq h(t) + h(s) - 2 \inf_{u \in [s,t]} h(u) \quad (25)$$

where X^j denotes the j-th tensor component of X. A geometric rough path X satisfies $S(X)_{[0,1]} = (1, 0, \dots)$ iff X is tree-like [21].

Theorem 3.1 (Uniquenss)**.** *Let $\mathcal{X}$ be a space of continuous paths $X : [0, T] \to \mathbb{R}^d \in BV(\mathbb{R}^d)$. Let $S : \mathcal{X} \to T((\mathbb{R}^d))$ be the signature map that associates each path X with its signature $S(X)_{[0,T]}$, then $S(X) = 1$ if and only if X is tree-like. The signature map is injective up to tree-like equivalence, meaning that two paths $X, Y \in \mathcal{X}$ have the same signature, i.e., $S(X)_{[0,T]} = S(Y_{[0,T]}$ if and only if X and Y are tree-like equivalent, denoted by $X \sim Y$ [21].*

Theorem 3.2 (Universal Approximation [16])**.** *Let $\mathcal{P}^p([0, T]), \mathbb{R}^d)$ denote the space of p-rough paths. On compact subsets $K \subset \mathcal{P}^p$, the signature is universal in the following sense: For any continuous function $F : K \to \mathbb{R}$ and $\epsilon > 0$, there exists a linear functional $L : T((\mathbb{R}^d)) \to \mathbb{R}$ on the signature such that:*

$$\sup_{X \in K} \left| F(X) - L(S(X)_{[0,T]}) \right| < \epsilon \quad (26)$$

This follows from the Stone-Weierstrass theorem applied to the algebra of linear functionals on $T((\mathbb{R}^d))$, guaranting that any continuous function on the path space can be approximated by linear combinations of signature terms.

Example 2. Feature Completeness Theorem. Let $\mathcal{P}$ be the space of continuous, bounded variation paths starting at $0 \in \mathbb{R}^d$, and let $\sim$ be a tree-like equivalence. Then:

$$S : P/\sim \longrightarrow T((\mathbb{R}^d))$$

is injective. That is,

$$S(X) = S(Y) \Longleftrightarrow X \sim Y \tag{27}$$

Hence, the signatuee transform is a complete and universal feature map on the quotient space $\mathcal{P}/\sim$.

Corollary 3.2.1 (Universality Property)**.** *For any compact set $K \subset \mathcal{X}$ and any continuous functional $f : K \to \mathbb{R}$, there exists a linear function $\ell : T((\mathbb{R}^d))$ such that $f(X) \approx \ell(S(X)) \, \forall \, X \in K$, with arbitrary precision.*

Lemma 3.3. *The signature $S(X)$ decays factorially: $\|S^n(X)\| \leq \frac{C^n}{n!}$, ensuring convergence of kernel expansions.*

The previous section establishes the benchmark model for this study. Recall that we used polynomials of degree 1 through 4 to construct concise stock portfolios based on predicted signals computed over five factors, which have been selected based on the highest R^2. To construct a structured approach from an academic, methodological, and practical viewpoint, we transition from discrete to continuous time, i.e., from time-series to path-signatures, and from polynomials to long short-term memory models (LSTM), Table 4. Polynomials, especially with mean-reverting terms like $(x - \mu)^2$ are structurally biased towards reverting behaviors—fitting for contrarian logic. LSTM, with its memory of past trends, is naturally suited to momentum/trend-following strategies. This design is not arbitrary, but conceptually aligned with financial intuition. It is both relevant and strategically valuable to compare fundamentally different model classes: nonlinear regression (polynomials) vs. sequence models (LSTM)—with and without signature transforms—because this helps create a broader comparative benchmarking without the genetic algorithms, while strengthening the robustness of the analysis towards adaptive learning, Figure 12, Appendix K. This framework demonstrates the generalization of signature utility as a universal feature representation for paths, evaluated under both traditional regression (polynomials) and deep learning (LSTM) models, which can help build a stronger case for their utility across modeling paradigms.

It is important to note that we are not just adding models, but rather creating a methodological data-intensive pipeline that is interpretable and modular through multiple advanced mathematical and machine learning techniques geared towards adaptive modeling logic (SigGA mapped to mean-reversion $\to$ polynomial or momentum $\to$ LSTM), which is informed and innovative from capital markets and financial mathematics perspectives. The framework covers interdisciplinary breadth as it integrates stochastic analysis (rough paths) with machine learning/artificial intelligence (PLS, LSTM, GA), econometrics (R^2, model comparison), and finance (factor modeling, momentum/contrarian logic). It has potentials to produce good stock predictions, especially in relative terms (ranking stocks, portfolio construction), provided two things are handled well:

1. Factor Quality: prediction quality will depend heavily on the relevance and informativeness of the input features (fundamental, macroeconomic, and market inputs are diverse and structured).
2. Robustness & Regime Awareness: by combining:
    - Signature features (to encode path structure)
    - Adaptive model selection (GA choosing best signature configuration)
    - Portfolio-type assignment (contrarian vs. trend)

However, even the best models won't forecast levels (or returns) precisely; rather, outperformance comes from better relative ranking and allocation. For regime-switching environments where pure trend or mean-reversion models often fail, deploying such adaptive learning models for portfolio selection, not point prediction, can lead to stronger and more consistent results than static models.

Table 4: Overview of Experimental Models

Model Class	**Description**	**Purpose**
Polynomial (deg 1–4)	Raw factor time series into polynomials	Baseline nonlinear prediction
[Signatures-PLS] + Poly	Sig-transformed to PLS into polynomials	Nonlinear feature enhancement
LSTM	Raw factor time series into LSTM	Sequence-based Nonlinear prediction
[Signatures-PLS] + LSTM	Sig-transformed to PLS into LSTM	Nonlinear feature enhancement

3.5.1 Signatures + Poylmodels Univariate Analysis

In addition to factor selection [11], that we accomplish through the use of signatures transform, we discuss informational inefficiencies of non-adaptive models. We want to choose five factors from a universe of 200+ factors per stock. To start, convert inputs from time series of returns to paths via "augmentation" or "embedding". Then, to compute the signature terms, partition paths into windows. For a rigorous and detailed treatment of these methods, see Adelaine Fermanian 2021 [8]: "augmentations remove the signature invariance to translation and/or parametrization", and "windows to get finer-scale information, where a global window $W(x) = (x)$ that takes the entire time series is the simplest". In this section, we use a few combinations of augmentation methods with the slide, expand and dyadic windows:

Time → basepoint augmentations: As per A. Fermanian [8] "the time parameter guarantees the uniqueness of the signature Hambly and Lyons, 2010 [21] and it adds information about the parametrization of the time series. Basepoint transformation makes the signature sensitive to translations Yang et al., 2017 [22]; Wu et al., 2021 [23]." Then:

$$\begin{aligned}\phi^{b,t}(x) &= \phi^b \circ \phi_t(x) \\ &= ((0,\mathbf{0}),(t_1,x_1),(t_2,x_2),\dots,(t_n,x_n)), \quad t_0=0,\ x_0=0 \\ &\in S(\mathbb{R}^{d+1})\end{aligned} \tag{28}$$

Lead-lag → time augmentations: Designed to embed temporal order and variation structure of a 1D (or multidimensional) time series into a higher-dimensional path suitable for signature transform. Lead channel: carries the “future” point in the time increment, X_{t_i}. Lag channel: carries the “past” point $X_{t_{i-1}}$. This ensures that areas under the curve in the signature capture/preserve the order via t_i and the interaction of increments via lead-lag (key for rough paths). Denote $x_{t_i} \in \mathbb{R}^d$ the original path, $t_i \in \mathbb{R}$ the timestamp associated with the sample:

$$\begin{aligned}\phi^{\text{Time+LL}} : S(\mathbb{R}^d) &\to S(\mathbb{R}^{2d+1}) \\ x &\mapsto (t_i, x^{\text{lead}}_{t_i}, x^{\text{lag}}_{t_{i-1}}), \quad i=1,\dots,T\end{aligned} \tag{29}$$

Edge handling at boundaries $i=0, X_{-1}=X_0 \mid 0$ & $i=T+1, X_T=X_{T+1} \to$ adds a (the last) datapoint to sample.

Signature Windowing Schemes Let $X=(X_t)_{t=0}^T \in \mathbb{R}^d$ be a multivariate time series (or path), and let $\text{Sig}^{(m)}$ denote the level-m signature transform. The path is segmented into windows, and signatures are computed on each segment.

1. Sliding Window (fixed width, moving step): Given a window size w and step size s, define the set of windowed segments as:

$$\mathcal{W}^{\text{slide}} = \{X_{t_i:t_i+w} \mid t_i = 0, s, 2s, \dots, T-w\}$$

For each window $X_{t_i:t_i+w}$, compute the signature:

$$\text{Sig}^{(m)}(X_{t_i:t_i+w})$$

2. Expanding Window (increasing horizon): Starts from a fixed origin $t_0=0$, and grows over time:

$$\mathcal{W}^{\text{expand}} = \{X_{0:t_i} \mid t_i \in \{s, 2s, \dots, T\}\}$$

Signatures are computed on each expanding prefix:

$$\text{Sig}^{(m)}(X_{0:t_i})$$

3. Dyadic Window (multiscale hierarchy): Dyadic windows segment the path using intervals of dyadic length 2^{-k} for $k=0,1,\dots,K$. Assuming unit interval $[0,1]$ is rescaled to the path domain $[0,T]$, define the dyadic partition at scale k:

$$\mathcal{P}_k = \left\{ \left[\frac{j}{2^k}T, \frac{j+1}{2^k}T\right) \,\middle|\, j=0,1,\dots,2^k-1 \right\}$$

Then for each interval $[a,b) \in \mathcal{P}_k$, compute:

$$\text{Sig}^{(m)}(X_{a:b})$$

This yields a **multiscale collection** of signatures, which are then concatenated:

$$\bigcup_{k=0}^{K} \left\{ \text{Sig}^{(m)}(X_{a:b}) \,\middle|\, [a,b) \in \mathcal{P}_k \right\}$$

[11] different than feature selection of signature coefficients, which we do through the partial least squares (PLS).

Generalized Signature Method A. Fermanina 2021 [8], chap 3, section 3.3.5 refers to the procedure of combining these different elementary operations to impact the final feature set as the *generalized signature method*: $x \mapsto z_{i,j}$. Following the author, we describe our signature pipeline:

Let input $x = (x_1, \ldots, x_n) \in \mathbb{R}^{n \times d}$ denote a discrete univariate (or multivariate) time series. We interpret this as a path in $\mathbb{R}^d$ by constructing a piecewise-linear interpolation, denoted $\tilde{x} \in S(\mathbb{R}^d)$, suitable for signature computations.

1. Define the augmented path $\phi(x) \in S(\mathbb{R}^{d'})$ where ϕ^i denotes the i-th augmentation: time, lead-lag, basepoint:

$$\begin{aligned} \phi^{\text{aug}} : S(\mathbb{R}^d) &\to S(\mathbb{R}^{d'})^p \\ x &\mapsto (\phi^1(x), \ldots, \phi^p(x)) \end{aligned} \tag{30}$$

2. Windowing function with w windows per augmentation (e.g., sliding, dyadic, or expanding windows) by:

$$\begin{aligned} W : S(\mathbb{R}^{d'}) &\to S(\mathbb{R}^{d'})^w \\ \phi^i(x) &\mapsto \left(W^1(\phi^i(x)), \ldots, W^w(\phi^i(x))\right) \end{aligned} \tag{31}$$

3. Signature transform where $\mathcal{S}^N$ is the signature of depth N, producing M features:

$$\mathcal{S}^N : S(\mathbb{R}^{d'}) \to \mathbb{R}^M \tag{32}$$

4. Partial least squares (PLS) dented ρ where $c \ll M$ is the number of PLS components retained:

$$\rho_c : \mathbb{R}^M \to \mathbb{R}^c \tag{33}$$

5. Polymodel that is a polynomial (e.g., contrarian) model of degree D:

$$\Phi^D : \mathbb{R}^c \to \mathbb{R} \tag{34}$$

6. Full composition where the set $\{z_{i,j}\}$ defines the full family of transformed and modeled outputs for all augmentations and windowed signature paths. For all $i = 1, \ldots, p$ and $j = 1, \ldots, w$, we compute:

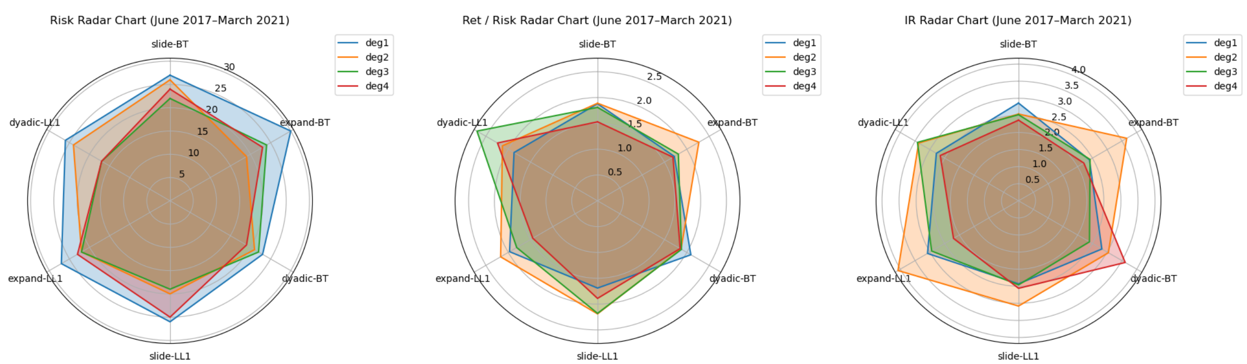
$$z_{i,j} = \left(\Phi^D \circ \rho_c \circ \mathcal{S}^N \circ W^j \circ \phi^i\right)(x) \tag{35}$$

Empirical Results We analyzed portfolio performance across signature degrees 1 to 4 for the period from June 2017 to March 2021. The results reveal a consistent trend: returns generally decline with higher signature degrees, while risk also tends to decrease, suggesting that deeper signature transforms reduce noise but may also filter out useful signal. Notably, portfolios constructed using degree 2 and degree 3 signatures deliver the best balance of risk and return. For instance, dyadic-LL1 at degree 3 achieves a return-to-risk ratio of 2.70 and an information ratio of 3.42, while expand-LL1 at degree 2 attains the highest information ratio overall at 4.08.

These findings suggest that degree 2–3 signatures strike a critical balance between complexity and overfitting, enabling more robust factor extraction. Methods based on dyadic and expanded local linear transforms (LL1) consistently outperform alternatives, pointing to the power of combining rough path theory with locally adaptive approximations. For practical applications, we recommend focusing on signature levels 2 and 3, particularly using dyadic or expand transforms within the LL1 framework, to construct optimal portfolios.

06-2017 : 03-2021	deg1	deg2	deg3	deg4
Sum Ret				
slide-BT	186	178	146	143
expand-BT	191	160	161	146
dyadic-BT	176	147	154	134
slide-LL1	166	160	152	172
expand-LL1	193	170	149	129
dyadic-LL1	178	183	161	136
Combined	**193**	**183**	**161**	**172**
R^2				
slide-BT	0.35	1.03	0.78	1.16
expand-BT	0.32	0.81	0.60	0.77
dyadic-BT	0.37	0.82	0.79	0.83
slide-LL1	0.35	1.00	0.93	1.04
expand-LL1	0.34	0.80	0.75	0.73
dyadic-LL1	0.37	0.91	1.03	1.16
Combined	**0.32**	**0.80**	**0.60**	**0.73**
Hit Ratio				
slide-BT	73	52	54	57
expand-BT	73	55	53	54
dyadic-BT	73	55	54	54
slide-LL1	73	58	55	55
expand-LL1	73	56	54	55
dyadic-LL1	73	56	55	56
Combined	**73**	**58**	**55**	**57**

Table 5: Performance metrics across polynomial degrees by portfolio method



(a) Long Portf. Risk (b) Long Portf. Ret/Risk (c) Long Portf. IR

Figure 8: Ex-post, Annualized risk, ret/risk, information ratio.

3.5.2 Signatures + LSTM Univariate Analysis

To be continued, To be continued, To be continued

4 Adaptive Learning

This section presents a comparative analysis between classical nonlinear modeling approaches using polynomial regression (referred to as polymodels) and modern pathwise modeling via signatures from rough path theory. The goal is to examine the extent to which signature transforms—particularly when combined with genetic algorithms (SigGA)—can better capture the underlying structure and predictive power of different financial variable types.

The Signature Genetic Algorithm is an Adaptive Learner, not just a variable selector—it evolves in response to fitness based on PLS. This positions it squarely within adaptive learning frameworks, especially in financial contexts where the "signal" shifts over time. This section compares Polymodels vs. SigGA $\equiv$ Static vs Adaptive: Polymodels (even up to Quartic degree) are static function approximations; SigGA helps emphasize the contrast in adaptive capability rather than just nonlinearity. This comparison strengthens the argument that non-adaptive models miss temporal and structural complexities. SigGA learns both the variables and the appropriate signature levels dynamically, making it adaptive in both structure and feature representation.

To be continued, To be continued, To be continued

References

[1] A. Cherny, Douady, and S. Molchanov, "On measuring nonlinear risk with scarce observations," in *Finance Stoch (2010) 14: 375–395*. Springer-Verlag 2009, 2009, pp. DOI 10.1007/s00 780–009–0107–y, received: 25 March 2008 / Accepted: 15 January 2009 / Published online: 7 November 2009.

[2] T. Barrau, "Predicting financial markets, developments and applications of polymodels theory," in *Université Paris 1 Panthéon-Sorbonne*. Center d'economie de la Sorbonne, 18 Dec 2020, ecole Doctoral ED-465.

[3] R. Clarke, H. D. Silva, and S. Thorley, "Non-linear factor returns in the u.s. equity market," in *BYU Marriott School of Business, Allspring Global Investments*. Financial Analysts Journal, volume 80, issue 3, 2024, [10.1080/0015198X.2024.2351020].

[4] P. AI and Perplexity, "This summary was generated with the assistance of perplexity ai," in *An artificial intelligence platform*, 5 April 2025.

[5] J. Bandurski and L. Postek, "Nonlinear dependencies in the fama and french three-factor model," in *Faculty of Economic Sciences, University of Warsaw, Poland*. Investment Analysts Journal, Vol 52, Issue 2, 27 Apr 2023.

[6] L. Zhu, S. Basu, R. A. Jarrow, and M. T. Wells, "High-dimensional estimation, basis assets, and the adaptive multi-factor model," in *preprint*. arXiv:1804.08472, 2018.

[7] R. Sukarieh, E. Meltzer, and R. Douady, "Dimension reduction signature genetic algorithm," in *Université Paris 1 Panthéon-Sorbonne*. Center d'economie de la Sorbonne, 05 Apr 2025, ecole Doctoral ED-465.

[8] A. Fermanian, "Learning time-dependent data with the signature transform," in *En vue de l'obtention du doctorat de Sobonne Université, Ecole Doctoral Sciences Mathematiques de Paris Center*. Laboratoire de Probabilités Statistique and Modélisation, Ph.D. Thesis, 15 Oct 2021.

[9] F. Black and M. Scholes, "The pricing of options and corporate liabilities," *Journal of Political Economy*, vol. 81, no. 3, pp. 637–654, 1973.

[10] J. C. Cox, J. E. Ingersoll, and S. A. Ross, "A theory of the term structure of interest rates," *Econometrica*, vol. 53, no. 2, pp. 385–407, 1985.

[11] P. Jorion, *Value at Risk: The New Benchmark for Managing Financial Risk*, 3rd ed. McGraw-Hill, 2006.

[12] T. Lyons, "Differential equations driven by rough signals," *Revista Matemática Iberoamericana*, vol. 14, no. 2, pp. 215–310, 1998.

[13] I. Chevyrev and T. Lyons, "Characteristic functions of measures on geometric rough paths," *Annals of Probability*, vol. 44, no. 6, pp. 4049–4082, 2016.

[14] J. H. Holland, *Adaptation in Natural and Artificial Systems*. University of Michigan Press, 1975.

[15] M. Mitchell, *An Introduction to Genetic Algorithms*. MIT Press, 1998.

[16] D. Levin, T. Lyons, and H. Ni, "Learning from the past, predicting the statistics for the future, learning an evolving system," in *University of Oxford*. arXiv:1309.0260v6 [q-fin.ST], March 23, 2016.

[17] O. Futter, B. Horvath, and M. Wiese, "Signature trading: A path-dependent extension of the mean-variance framework with exogenous signals," in *Imperial College London, University of Oxford, University of Kaiserslautern, Department of Mathematics*. arXiv:2308.15135v2 [q-fin.PM], 30 Aug 2023.

[18] D. Lee, "The surface signature and rough surfaces," in *Functional Analysis*. arXiv:2406.16857v2 [math.FA], 1 Aug 2024.

[19] F. Galuppi, G. Moreno, and P. Santarsiero, "Path signatures of ode solutions," in *University of Warsaw, Università di Bologna*. arXiv:2505.13234v1 [math.AG], 19 May 2025, license: CC BY-NC-ND 4.0.

[20] T. J. Lyons, M. Caruana, and T. Lévy, "Differential equations driven by rough paths," in *Springer*, 2004, ecole d'Eté de Probabilités de Saint-Flour XXXIV.

[21] B. Hambly and T. J. Lyons, "Uniqueness for the signature of a path of bounded variation and the reduced path group," in *Annals of Probability*. arXiv:math/0507536v2 [math.CA], 19 Dec 2006, mathematical Institute, Oxford University.

[22] S. Lai, L. Jin, and W. Yang, "Online signature verification using recurrent neural network and length-normalized path signature," in *South China University of Technology, School of Electronic and Information Engineering*. arXiv:1705.06849v1 [cs.CV], 19 May 2017.

[23] Y. Wu, H. Ni, and T. J. Lyons, "Signature features with the visibility transformation," in *University College of London, University of Oxford, Loughborough University*. arXiv:2004.04006v4 [cs.LG], 8 Oct 2020.

A Appendix Scaled Levels & Standardized Returns

Scale all input levels and return data to the same base. Denote the assets and integer time indices by $i \in \{1, 2, \ldots, I\}$ and $t \in \{1, 2, \ldots T\}$. Denote standardized returns on stock i by $R_i, \forall\, i \in \mathcal{I} = \{1, 2, \ldots, I\}$ and standardized returns on factor j by $F_j, \forall j \in \mathcal{J} = \{1, 2, \ldots, J\}$. The natural logarithm returns of stocks and factors r, f respectively. T corresponds to the most recent observation and is indexed to the number of observations. Set $Y_i \equiv R_i$ and $X_j \equiv F_j$. For each stock and risk factor, we compute a divider as follows:

$$
\begin{aligned}
p_{i,t=1:T} &= levels, prices \\
l_{i,t=1:T} &= |p_{i,t=1:T}| \\
d_{i=1:I,T} &= 10^{log_{10}(\max{(l_{i,t=1:T})})} \\
L_{i,t=1:T} &= 100 + \frac{p_{i,t=1:T}}{d_{i=1:I,T}}
\end{aligned}
$$

$$r_{i,t} = \ln(L_{i,t}) - \ln(L_{i,t-1}) \qquad f_{i,t} = \ln(L_{i,t}) - \ln(L_{i,t-1}) \tag{36}$$

$$R_{i,t} = \frac{r_{i,t} - \overline{\mu}}{\sigma} \qquad F_{i,t} = \frac{f_{i,t} - \overline{\mu}}{\sigma} \tag{37}$$

$$\sigma = \sqrt{\frac{1}{\tau}\sum_{i=1}^{T}(x_{i,t} - \overline{\mu})^2} \qquad \overline{\mu} = \frac{1}{\tau}\sum_{t=1}^{T} x_{i,t}$$

B Hermite Transformation:

For $w(x) = e^{-x^2/2}$, $[a, b] = [-\infty, \infty]$, the first five orthogonal Hermite polynomials $H_n(x)$ are:

$$H_{e0}(x) = 1 \tag{38}$$
$$H_{e1}(x) = x \tag{39}$$
$$H_{e2}(x) = (x^2 - 1) \tag{40}$$
$$H_{e3}(x) = (x^3 - 3x) \tag{41}$$
$$H_{e4}(x) = (x^4 - 6x^2 + 3) \tag{42}$$

Adjusted Hermite: The values of variables in the powers of x increase exponentially with i, $(x^i, i = 1, \ldots n)$. To adjust for aberration in variable (tail) values, interpolate the Hermites of large negative and positive observations by the first-order derivative of the functions $H_n(x)$. Derive thresholds using $\{\alpha, \beta\}$ the constant and slope Pareto parameters of both tails $\pm \left[e^{\alpha}/(1-p)\,\right]^{-1/\beta}$. Then interpolate the bottom and top 20% i.e., $1 - p = 0.2$ observations, that is, $\forall\, x \in (-x < -\tau)$ and $(x > \rho)$ compute $\overleftarrow{H}$ and $\overrightarrow{H}$ respectively. The general forms of the interpolated $\overleftarrow{H}_n(-x < -\tau)$ and Hermite polynomial $H_n(x) := H_{(x<-\tau)} \cup H_{(x\geq-\tau\ \&\ x\leq\rho)} \cup H_{(x>\rho)}$:

$$\overleftarrow{H}_n(-x) = \Big[H_n(-\tau) - H_n^{'}(-\tau) \times (-\tau)\Big] + H_n^{'}(-\tau) \times (-x) \tag{43}$$

$$\overleftarrow{H}_0(-x) = 1 \tag{44}$$
$$\overleftarrow{H}_1(-x) = -x \tag{45}$$
$$\overleftarrow{H}_2(-x) = -\tau^2 - 1 + 2\tau x \tag{46}$$
$$\overleftarrow{H}_3(-x) = 2\tau^3 - 3\tau^2 x + 3x \tag{47}$$
$$\overleftarrow{H}_4(-x) = -3\tau^4 + 6\tau^2 + 3 + 4\tau^3 x - 12\tau x \tag{48}$$

$$H_n(x) = \overleftarrow{H}_n(x) \cup \overline{H}_n(x) \cup \overrightarrow{H}n(x) \tag{49}$$

C Appendix Fundamental Data Definition

Idx	Indicator	Title	Description
1	cashnequsd	Cash and Equivalents (USD)	[Balance Sheet] [CASHNEQ] in USD, converted by [FXUSD].
2	debtusd	Total Debt (USD)	[Balance Sheet] [DEBT] in USD, converted by [FXUSD].
3	equityusd	Shareholders Equity (USD)	[Balance Sheet] [EQUITY] in USD, converted by [FXUSD].
4	taxassets	Tax Assets	[Balance Sheet] A component of [ASSETS] representing tax assets and receivables.
5	ppnenet	Property, Plant & Equipment Net	[Balance Sheet] A component of [ASSETS] representing the amount after accumulated depreciation, depletion and amortization of physical assets used in the normal conduct of business to produce goods and services and not intended for resale.
6	inventory	Inventory	[Balance Sheet] A component of [ASSETS] representing the amount after valuation and reserves of inventory expected to be sold, or consumed within one year or operating cycle, if longer.
7	intangibles	Goodwill and Intangible Assets	[Balance Sheet] A component of [ASSETS] representing the carrying amounts of all intangible assets and goodwill as of the balance sheet date, net of accumulated amortization and impairment charges.
8	investments	Investments	[Balance Sheet] A component of [ASSETS] representing the total amount of marketable and non-marketable securities, loans receivable, and other invested assets.
9	receivables	Trade and Non-Trade Receivables	[Balance Sheet] A component of [ASSETS] representing trade and non-trade receivables.
10	accoci	Accumulated Other Comprehensive Income	[Balance Sheet] A component of [EQUITY] representing the accumulated change in equity from transactions and other events and circumstances from non-owner sources, net of tax effect, at period end. Includes foreign currency translation items, certain pension adjustments, unrealized gains and losses on certain investments in debt and equity securities.
11	retearn	Accumulated Retained Earnings (Deficit)	[Balance Sheet] A component of [EQUITY] representing the cumulative amount of the entity's undistributed earnings or deficit. May only be reported annually by certain companies, rather than quarterly.
12	taxliabilities	Tax Liabilities	[Balance Sheet] A component of [LIABILITIES] representing outstanding tax liabilities.
13	deferredrev	Deferred Revenue	[Balance Sheet] A component of [LIABILITIES] representing the carrying amount of consideration received or receivable on potential earnings that were not recognized as revenue, including sales, license fees, and royalties, but excluding interest income.
14	deposits	Deposit Liabilities	[Balance Sheet] A component of [LIABILITIES] representing the total of all deposit liabilities held, including foreign and domestic, interest and non-interest bearing. May include demand deposits, saving deposits, Negotiable Order of Withdrawal, and time deposits, among others.
15	payables	Trade and Non-Trade Payables	[Balance Sheet] A component of [LIABILITIES] representing trade and non-trade payables.
16	assetsnc	Assets Non-Current	[Balance Sheet] Amount of non-current assets, for companies that operate a classified balance sheet. Calculated as the difference between Total Assets [ASSETS] and Current Assets [ASSETSC].

17	assets	Total Assets	[Balance Sheet] Sum of the carrying amounts as of the balance sheet date of all assets that are recognized. Major components are [CASHNEQ], [INVESTMENTS], [INTANGIBLES], [PPNENET], [TAXASSETS], and [RECEIVABLES].
18	liabilities	Total Liabilities	[Balance Sheet] Sum of the carrying amounts as of the balance sheet date of all liabilities that are recognized. Principal components are [DEBT], [DEFERREDREV], [PAYABLES], [DEPOSITS], and [TAXLIABILITIES].
19	assetsc	Current Assets	[Balance Sheet] The current portion of [ASSETS], reported if a company operates a classified balance sheet that segments current and non-current assets.
20	debtc	Debt Current	[Balance Sheet] The current portion of [DEBT], reported if the company operates a classified balance sheet that segments current and non-current liabilities.
21	investmentsc	Investments Current	[Balance Sheet] The current portion of [INVESTMENTS], reported if the company operates a classified balance sheet that segments current and non-current assets.
22	liabilitiesc	Current Liabilities	[Balance Sheet] The current portion of [LIABILITIES], reported if the company operates a classified balance sheet that segments current and non-current liabilities.
23	debtnc	Debt Non-Current	[Balance Sheet] The non-current portion of [DEBT] reported if the company operates a classified balance sheet that segments current and non-current liabilities.
24	investmentsnc	Investments Non-Current	[Balance Sheet] The non-current portion of [INVESTMENTS], reported if the company operates a classified balance sheet that segments current and non-current assets.
25	liabilitiesnc	Liabilities Non-Current	[Balance Sheet] The non-current portion of [LIABILITIES], reported if the company operates a classified balance sheet that segments current and non-current liabilities.
26	ebitdausd	EBITDA (USD)	[Metrics] [EBITDA] in USD, converted by [FXUSD].
27	ps1	Price to Sales Ratio	[Metrics] An alternative calculation method to [PS], that measures the ratio between a company's [PRICE] and its [SPS].
28	pe1	Price to Earnings Ratio	[Metrics] An alternative to [PE] representing the ratio between [PRICE] and [EPSUSD].
29	ebt	Earnings before Tax	[Metrics] Earnings Before Tax is calculated by adding [TAXEXP] back to [NETINC].
30	ev	Enterprise Value	[Metrics] Enterprise value is a measure of the value of a business as a whole, calculated as [MARKETCAP] plus [DEBTUSD] minus [CASHNEQUSD].
31	fcf	Free Cash Flow	[Metrics] Free Cash Flow is a measure of financial performance calculated as [NCFO] minus [CAPEX].
32	fcfps	Free Cash Flow per Share	[Metrics] Free Cash Flow per Share is a valuation metric calculated by dividing [FCF] by [SHARESWA].
33	grossmargin	Gross Margin	[Metrics] Gross Margin measures the ratio between a company's [GP] and [REVENUE].
34	invcap	Invested Capital	[Metrics] Invested capital is an input into the calculation of [ROIC], and is calculated as: [DEBT] plus [ASSETS] minus [INTANGIBLES] minus [CASHNEQ] minus [LIABILITIESC]. Please note this calculation method is subject to change.
35	bvps	Book Value per Share	[Metrics] Measures the ratio between [EQUITY] and [SHARESWA].

36	evebitda	Enterprise Value over EBITDA	[Metrics] Measures the ratio between [EV] and [EBITDAUSD].
37	evebit	Enterprise Value over EBIT	[Metrics] Measures the ratio between [EV] and [EBITUSD].
38	de	Debt to Equity Ratio	[Metrics] Measures the ratio between [LIABILITIES] and [EQUITY].
39	pb	Price to Book Value	[Metrics] Measures the ratio between [MARKETCAP] and [EQUITYUSD].
40	tbvps	Tangible Assets Book Value per Share	[Metrics] Measures the ratio between [TANGIBLES] and [SHARESWA].
41	ps	Price Sales (Damodaran Method)	[Metrics] Measures the ratio between a company's [MARKETCAP] and [REVENUEUSD].
42	ebitdamargin	EBITDA Margin	[Metrics] Measures the ratio between a company's [EBITDA] and [REVENUE].
43	netmargin	Profit Margin	[Metrics] Measures the ratio between a company's [NETINCCMN] and [REVENUE].
44	marketcap	Market Capitalization	[Metrics] Represents the product of [SHARESBAS], [PRICE] and [SHAREFACTOR].
45	sps	Sales per Share	[Metrics] Sales per Share measures the ratio between [REVENUEUSD] and [SHARESWA].
46	payoutratio	Payout Ratio	[Metrics] The percentage of earnings paid as dividends to common stockholders. Calculated by dividing [DPS] by [EPSUSD].
47	currentratio	Current Ratio	[Metrics] The ratio between [ASSETSC] and [LIABILITIESC], for companies that operate a classified balance sheet.
48	tangibles	Tangible Asset Value	[Metrics] The value of tangible assets calculated as the difference between [ASSETS] and [INTANGIBLES].
49	workingcapital	Working Capital	[Metrics] Working capital measures the difference between [ASSETSC] and [LIABILITIESC].
50	sharesbas	Shares (Basic)	[Entity] The number of shares or other units outstanding of the entity's capital or common stock or other ownership interests, as stated on the cover of related periodic report (10-K/10-Q), after adjustment for stock splits.
51	ncff	Net Cash Flow from Financing	[Cash Flow Statement] A component of [NCF] representing the amount of CF (outflow) from financing activities, from continuing and discontinued operations. Principal components of financing CF are: issuance (purchase) of equity shares, issuance (repayment) of debt securities, and payment of dividends, other cash distributions.
52	ncfi	Net Cash Flow from Investing	[Cash Flow Statement] A component of [NCF] representing the amount of CF (outflow) from investing activities, from continuing and discontinued operations. Principal components of investing CF are: capital (expenditure) disposal of equipment [CAPEX], business (acquisitions) disposition [NCFBUS] and investment (acquisition) disposal [NCFINV].
53	ncfo	Net Cash Flow from Operations	[Cash Flow Statement] A component of [NCF] representing the amount of CF (outflow) from operating activities, from continuing and discontinued operations.
54	ncfdiv	Dividends Payment, Other Cash Distributions	[Cash Flow Statement] A component of [NCFF] representing dividends and their equivalents paid on common stock, and restricted stock units.

55	ncfcommon	Issuance (Purchase) of Equity Shares	[Cash Flow Statement] A component of [NCFF] representing the net CF (outflow) from common equity changes. Includes additional capital contributions from share issuances and exercise of stock options, and outflow from share repurchases.
56	ncfdebt	Issuance (Repayment) of Debt Securities	[Cash Flow Statement] A component of [NCFF] representing the net cash inflow (outflow) from issuance (repayment) of debt securities.
57	ncfbus	NCF - Business Acquisitions, Disposals	[Cash Flow Statement] A component of [NCFI] representing the NCF (outflow) associated with the acquisition, disposal of businesses, joint-ventures, affiliates, and other named investments.
58	ncfinv	NCF - Investment Acquisitions, Disposals	[Cash Flow Statement] A component of [NCFI] representing the NC inflow (outflow) associated with the acquisition, disposal of investments, including marketable securities and loan originations.
59	capex	Capital Expenditure	[Cash Flow Statement] A component of [NCFI] representing the net cash inflow (outflow) associated with the acquisition, disposal of long-lived, physical, intangible assets that are used in the normal conduct of business to produce goods and services and are not intended for resale. Includes cash inflows/outflows to pay for construction of self-constructed assets, software.
60	sbcomp	Share Based Compensation	[Cash Flow Statement] A component of [NCFO] representing the total amount of noncash, equity-based employee remuneration. This may include the value of stock or unit options, amortization of restricted stock or units, and adjustment for officers' compensation. As noncash, this element is an add back when calculating net cash generated by operating activities using the indirect method.
61	ncfx	Effect of Exchange Rate Changes on Cash	[Cash Flow Statement] A component of Net Cash Flow [NCF] representing the amount of increase (decrease) from the effect of exchange rate changes on cash and cash equivalent balances held in foreign currencies.
62	depamor	Depreciation, Amortization, and Accretion	[Cash Flow Statement] A component of operating CF representing the aggregate net amount of depreciation, amortization, and accretion recognized during an accounting period. As a non-cash item, the net amount is added back to net income when calculating cash provided by or used in operations using the indirect method.
63	ncf	NCF - Change in Cash Cash Equivalents	[Cash Flow Statement] Principal component of the CF statement representing the amount of increase (decrease) in cash and cash equivalents. Includes [NCFO], investing [NCFI] and financing [NCFF] for continuing and discontinued operations, and the effect of exchange rate changes on cash [NCFX].
64	ebitusd	Earning Before Interest, Taxes (USD)	[Income Statement] [EBIT] in USD, converted by [FXUSD].
65	epsusd	Earnings per Basic Share (USD)	[Income Statement] [EPS] in USD, converted by [FXUSD].
66	netinccmnusd	Net Income Common Stock (USD)	[Income Statement] [NETINCCMN] in USD, converted by [FXUSD].
67	revenueusd	Revenues (USD)	[Income Statement] [REVENUE] in USD, converted by [FXUSD].

68	rnd	Research and Development Expense	[Income Statement] A component of [OpEx] representing the aggregate costs incurred in a planned search or critical investigation aimed at discovery of new knowledge with the hope that such knowledge will be useful in developing a new product or service.
69	sgna	Selling, General and Administrative Expense	[Income Statement] A component of [OpEx] representing the aggregate total costs related to selling a firm's product and services, as well as all other general and administrative expenses. Direct selling expenses (for example, credit, warranty, and advertising) are expenses that can be directly linked to the sale of specific products. Indirect selling expenses are expenses that cannot be directly linked to the sale of specific products, for example telephone expenses, Internet, and postal charges. General and administrative expenses include salaries of non-sales personnel, rent, utilities, communication, etc.
70	gp	Gross Profit	[Income Statement] Aggregate revenue [REVENUE] less cost of revenue [COR] directly attributable to the revenue generation activity.
71	taxexp	Income Tax Expense	[Income Statement] Amount of current income tax expense (benefit) and deferred income tax expense (benefit) pertaining to continuing operations.
72	intexp	Interest Expense	[Income Statement] Amount of the cost of borrowed funds accounted for as interest expense.
73	opex	Operating Expenses	[Income Statement] Operating expenses represents the total expenditure on [SGnA], [RnD] and other operating expense items, it excludes [CoR].
74	opinc	Operating Income	[Income Statement] Operating income is a measure of financial performance before the deduction of [INTEXP], [TAXEXP] and other Non-Operating items. It is calculated as [GP] minus [OPEX].
75	cor	Cost of Revenue	[Income Statement] The aggregate cost of goods produced and sold and services rendered during the reporting period.
76	consolinc	Consolidated Income	[Income Statement] The portion of profit or loss for the period, net of income taxes, which is attributable to the consolidated entity, before the deduction of [NetIncNCI].
77	shareswa	Weighted Average Shares	[Income Statement] The weighted average number of shares or units issued and outstanding that are used by the company to calculate [EPS], determined based on the timing of issuance of shares or units in the period.
78	shareswadil	Weighted Average Shares Diluted	[Income Statement] The weighted average number of shares or units issued and outstanding that are used by the company to calculate [EPSDil], determined based on the timing of issuance of shares or units in the period.

D Appendix Economics and Financial Data Definition

Index	Indicator Code	Indicator Name
1	ADP	Adp Employment Change
2	BCONF	Business Confidence
3	BOT	Balance Of Trade
4	BP	Building Permits
5	BR	Bankruptcies
6	CA	Current Account
7	CARS	Car Registrations
8	CBBS	Central Bank Balance Sheet

9	CCONF	Consumer Credit
10	CCPI	Core Consumer Prices
11	CF	Capital Flows
12	CFNAI	Chicago Fed National Activity Index
13	CHJC	Challenger Job Cuts
14	CJC	Continuing Jobless Claims
15	CNCN	Consumer Confidence
16	COR	Crude Oil Rigs
17	COSC	Crude Oil Stocks Change
18	CP	Corporate Profits
19	CPIC	Inflation Rate
20	CPMI	Chicago PMI
21	CPPI	Core Pce Price Index
22	CSP	Consumer Spending
23	CU	Capacity Utilization
24	DINV	Changes In Inventories
25	DPINC	Disposable Personal Income
26	DUR	Durable Goods Orders
27	EHS	Existing Home Sales
28	EMPST	Ny Empire State Manufacturing Index
29	EXPX	Export Prices
30	EXVOL	Exports
31	FACT	Factory Orders
32	FBI	Foreign Bond Investment
33	FDI	Foreign Direct Investment
34	FER	Foreign Exchange Reserves
35	FINF	Food Inflation
36	FOET	Factory Orders Ex Transportation
37	GAGR	GDP Annual Growth Rate
38	GASSC	Gasoline Stocks Change
39	GBVL	Government Budget Value
40	GCP	GDP Constant Prices
41	GD	GDP Deflator
42	GFCF	Gross Fixed Capital Formation
43	GGR	GDP Growth Rate
44	GPAY	Government Payrolls
45	GREV	Government Revenues
46	GSP	Government Spending
47	GYLD	Government Bond 10y
48	HSTT	Housing Starts
49	IMPX	Import Prices
50	IMVOL	Imports
51	ISMNYI	Ism New York Index
52	JCLM	Initial Jobless Claims
53	JOBOFF	Job Offers
54	JVAC	Job Vacancies
55	LC	Labour Costs
56	LPS	Loans To Private Sector
57	LUNR	Long Term Unemployment Rate
58	M0	Money Supply M0
59	M1	Money Supply M1
60	M2	Money Supply M2
61	MANWG	Wages In Manufacturing
62	MKT	Stock Market
63	MORTG	Mortgage Rate
64	MP	Manufacturing Production

65	MPAY	Manufacturing Payrolls
66	NAHB	Nahb Housing Market Index
67	NATGSC	Natural Gas Stocks Change
68	NFIB	Nfib Business Optimism Index
69	NHS	New Home Sales
70	NLTTF	Net Long Term Tic Flows
71	NMPMI	Non Manufacturing PMI
72	NO	New Orders
73	OIL	Crude Oil Production
74	OPT	Economic Optimism Index
75	PCEPI	Pce Price Index
76	PFED	Philadelphia Fed Manufacturing Index
77	PHS	Pending Home Sales
78	PPIC	Producer Prices Change
79	PROD	Productivity
80	PSAV	Personal Savings
81	RSM	Retail Sales Mom
82	RSY	Retail Sales Yoy
83	TOT	Terms Of Trade
84	TOUR	Tourist Arrivals
85	TVS	Total Vehicle Sales
86	UNR	Unemployment Rate
87	UNRY	Youth Unemployment Rate
88	WAGE	Wages

E Appendix Exchange Traded Funds ETFs Definitions

No.	Ticker	ETF Name
1	IVV	iShares Core S&P 500 ETF
2	IJK	iShares S&P Mid-Cap 400 Growth ETF
3	OEF	iShares S&P 100 ETF
4	IWO	iShares Russell 2000 Growth ETF
5	IUSG	iShares Core S&P U.S. Growth ETF
6	IWV	iShares Russell 3000 ETF
7	IWB	iShares Russell 1000 ETF
8	ITOT	iShares Core S&P Total U.S. Stock Market ETF
9	IJJ	iShares S&P Mid-Cap 400 Value ETF
10	IVW	iShares S&P 500 Growth ETF
11	IWN	iShares Russell 2000 Value ETF
12	IOO	iShares Global 100 ETF
13	IUSV	iShares Core S&P U.S. Value ETF
14	IYY	iShares Dow Jones U.S. ETF
15	IWR	iShares Russell Mid-Cap ETF
16	AGG	iShares Core U.S. Aggregate Bond ETF
17	IJH	iShares Core S&P Mid-Cap ETF
18	IVE	iShares S&P 500 Value ETF
19	IJT	iShares S&P Small-Cap 600 Growth ETF
20	IWP	iShares Russell Mid-Cap Growth ETF
21	IWM	iShares Russell 2000 ETF
22	IJR	iShares Core S&P Small-Cap ETF
23	DVY	iShares Select Dividend ETF
24	IJS	iShares S&P Small-Cap 600 Value ETF
25	IWS	iShares Russell Mid-Cap Value ETF
26	EPP	iShares MSCI Pacific ex Japan ETF
27	EEM	iShares MSCI Emerging Markets ETF

Continued on next page

Table 8 – *Continued from previous page*

No.	Ticker	ETF Name
28	ILF	iShares Latin America 40 ETF
29	EWA	iShares MSCI Australia ETF
30	EWO	iShares MSCI Austria ETF
31	EWK	iShares MSCI Belgium ETF
32	EWC	iShares MSCI Canada ETF
33	EWQ	iShares MSCI France ETF
34	EWG	iShares MSCI Germany ETF
35	EWH	iShares MSCI Hong Kong ETF
36	EWI	iShares MSCI Italy ETF
37	EWJ	iShares MSCI Japan ETF
38	JPXN	iShares JPX-Nikkei 400 ETF
39	EWN	iShares MSCI Netherlands ETF
40	EWS	iShares MSCI Singapore ETF
41	EWP	iShares MSCI Spain ETF
42	EWD	iShares MSCI Sweden ETF
43	EWL	iShares MSCI Switzerland ETF
44	EWU	iShares MSCI United Kingdom ETF
45	FXI	iShares China Large-Cap ETF
46	EWZ	iShares MSCI Brazil ETF
47	EWM	iShares MSCI Malaysia ETF
48	EWW	iShares MSCI Mexico ETF
49	EZA	iShares MSCI South Africa ETF
50	EWY	iShares MSCI South Korea ETF
51	EWT	iShares MSCI Taiwan ETF
52	SHY	iShares 1-3 Year Treasury Bond ETF
53	IEF	iShares 7-10 Year Treasury Bond ETF
54	TLT	iShares 20+ Year Treasury Bond ETF
55	TIP	iShares TIPS Bond ETF
56	LQD	iShares iBoxx $ Investment Grade Corporate Bond ETF
57	IYK	iShares U.S. Consumer Staples ETF
58	IYC	iShares U.S. Consumer Discretionary ETF
59	IXC	iShares Global Energy ETF
60	IYE	iShares U.S. Energy ETF
61	IXG	iShares Global Financials ETF
62	IYG	iShares U.S. Financial Services ETF
63	IYF	iShares U.S. Financials ETF
64	IXJ	iShares Global Healthcare ETF
65	IBB	iShares Biotechnology ETF
66	IYH	iShares U.S. Healthcare ETF
67	IYT	iShares U.S. Transportation ETF
68	IYJ	iShares U.S. Industrials ETF
69	IYM	iShares U.S. Basic Materials ETF
70	IYR	iShares U.S. Real Estate ETF
71	ICF	iShares Cohen & Steers REIT ETF
72	IGM	iShares Expanded Tech Sector ETF
73	IGV	iShares Expanded Tech-Software Sector ETF
74	SOXX	iShares Semiconductor ETF
75	IYW	iShares U.S. Technology ETF
76	IXP	iShares Global Comm Services ETF
77	IYZ	iShares U.S. Telecommunications ETF
78	IDU	iShares U.S. Utilities ETF

F Appendix: Stocks Used in Study

No.	Ticker	Stock Name
1	AAPL	Apple Inc.
2	ABT	Abbott Laboratories
3	ADBE	Adobe Inc.
4	ACN	Accenture plc
5	AIZ	Assurant Inc.
6	AMAT	Applied Materials Inc.
7	AMGN	Amgen Inc.
8	AMZN	Amazon.com Inc.
9	APA	APA Corporation
10	AVB	AvalonBay Communities Inc.
11	BAC	Bank of America Corp
12	BBY	Best Buy Co. Inc.
13	BEN	Franklin Resources Inc.
14	BLL	Ball Corporation
15	BLK	BlackRock Inc.
16	BMY	Bristol-Myers Squibb Co.
17	BXP	Boston Properties Inc.
18	CBRE	CBRE Group Inc.
19	C	Citigroup Inc.
20	CMCSA	Comcast Corp
21	COST	Costco Wholesale Corp
22	COG	Coterra Energy Inc.
23	CRM	Salesforce Inc.
24	CSCO	Cisco Systems Inc.
25	CVS	CVS Health Corp
26	CVX	Chevron Corp
27	DE	Deere & Company
28	DIS	The Walt Disney Company
29	DLTR	Dollar Tree Inc.
30	DTE	DTE Energy Co.
31	DVA	DaVita Inc.
32	ECL	Ecolab Inc.
33	ED	Consolidated Edison Inc.
34	FITB	Fifth Third Bancorp
35	FLIR	Teledyne FLIR LLC
36	FRT	Federal Realty Investment Trust
37	GE	General Electric Co.
38	GPS	The Gap Inc.
39	GOOGL	Alphabet Inc.
40	HD	The Home Depot Inc.
41	HFC	HollyFrontier Corp
42	HON	Honeywell International Inc.
43	IBM	International Business Machines Corp
44	INTC	Intel Corp
45	INTU	Intuit Inc.
46	IVV	iShares Core S&P 500 ETF
47	J	Jacobs Engineering Group Inc.
48	JNJ	Johnson & Johnson
49	JPM	JPMorgan Chase & Co.
50	KIM	Kimco Realty Corp
51	KO	The Coca-Cola Company
52	KR	The Kroger Co.
53	LEG	Leggett & Platt Inc.
54	LEN	Lennar Corporation
55	LLY	Eli Lilly and Company
56	MCD	McDonald's Corp
57	MDT	Medtronic plc

Continued on next page

Table 9 – *Continued from previous page*

No.	Ticker	Stock Name
58	MMM	3M Company
59	MRK	Merck & Co. Inc.
60	MTD	Mettler-Toledo International Inc.
61	NEE	NextEra Energy Inc.
62	NFLX	Netflix Inc.
63	NKE	Nike Inc.
64	NOV	NOV Inc.
65	NVDA	NVIDIA Corp
66	ORLY	O'Reilly Automotive Inc.
67	PBCT	People's United Financial Inc.
68	PE	Parsley Energy Inc.
69	PEP	PepsiCo Inc.
70	PFE	Pfizer Inc.
71	PG	Procter & Gamble Co.
72	PH	Parker-Hannifin Corporation
73	PRGO	Perrigo Company plc
74	PVH	PVH Corp
75	QCOM	Qualcomm Inc.
76	RMD	ResMed Inc.
77	RL	Ralph Lauren Corporation
78	ROL	Rollins Inc.
79	RTX	Raytheon Technologies Corp
80	SBUX	Starbucks Corp
81	SEE	Sealed Air Corp
82	SIVB	SVB Financial Group
83	T	AT&T Inc.
84	TGT	Target Corp
85	TMO	Thermo Fisher Scientific Inc.
86	TSCO	Tractor Supply Co.
87	TXN	Texas Instruments Inc.
88	UNH	UnitedHealth Group Inc.
89	UNM	Unum Group
90	UNP	Union Pacific Corp
91	UPS	United Parcel Service Inc.
92	VFC	VF Corporation
93	VIAC	ViacomCBS Inc.
94	VNO	Vornado Realty Trust
95	VZ	Verizon Communications Inc.
96	WFC	Wells Fargo & Co.
97	WELL	Welltower Inc.
98	WMT	Walmart Inc.
99	WY	Weyerhaeuser Company
100	XOM	Exxon Mobil Corp
101	ZBRA	Zebra Technologies Corp

G Appendix: Anomalous Alpha Extraction with Polymodels

Polymodel	Hermite		Gram-Schmidt		IVV	MTUM
Date	**Short**	**Long**	**Short**	**Long**	**Index**	**Index**
2017-06-30	1	7	-1	9	3	9
2017-09-30	2	5	1	6	4	10
2017-12-31	7	9	4	8	7	11
2018-03-31	-6	2	-7	4	-1	-4
2018-06-30	-1	8	0	5	3	5
2018-09-30	4	10	3	9	8	-4
2018-12-31	-15	-6	-19	-3	-13	0
2019-03-31	14	14	10	17	14	8
2019-06-30	0	7	0	9	4	6
2019-09-30	4	4	1	4	2	0
2019-12-31	3	13	4	14	9	9
2020-03-31	-26	-19	-29	-10	-20	-8
2020-06-30	12	24	16	28	20	18
2020-09-30	-4	20	3	19	9	1
2020-12-31	21	16	21	14	12	16
2021-03-31	7	15	11	11	6	5
Total	23	129	18	144	68	81
Ret p.a.	3	34	2	40	16	21
Risk p.a.	22	20	24	18	19	15
Ret/risk	0.14	1.66	0.06	2.23	0.83	1.41
TE p.a.	10	6	9	6	0	13

Table 10: Polymodel (deg[4], feature= 5) Portfolios

Polymodel Portfolios	Short-HE	Long-HE	Short-GS	Long-GS
Beta (MTUM)	0.51	0.50	0.46	0.58
Correl (MTUM)	0.77	0.69	0.77	0.70
Beta (IVV)	0.78	0.91	0.75	1.03
Correl (IVV)	0.90	0.96	0.94	0.95
Beta (P-IVV, MTUM)	0.46	-0.07	0.86	-0.64
Correl (P-IVV, MTUM)	0.31	-0.03	0.51	-0.27

Table 11: Beta and correlation with IVV and MTUM

H Appendix: Learning Nonlinearity From Polymodels: A Concrete Univariate Evidence

Poly, GS	deg $= 1$		deg $= 2$		deg $= 3$		deg $= 4$	
Date	**Short**	**Long**	**Short**	**Long**	**Short**	**Long**	**Short**	**Long**
2017-06-30	-4	10	-1	9	1	10	-1	9
2017-09-30	-2	11	1	9	3	8	1	6
2017-12-31	1	13	6	11	7	10	4	8
2018-03-31	-4	6	-3	4	-6	3	-7	4
2018-06-30	-3	12	0	9	-1	7	0	5
2018-09-30	-2	12	5	12	4	13	3	9
2018-12-31	-19	-4	-16	-3	-16	-1	-19	-3
2019-03-31	4	17	12	18	11	18	10	17
2019-06-30	-6	9	-1	11	0	8	0	9
2019-09-30	-2	6	2	6	-1	4	1	4
2019-12-31	0	15	7	12	4	15	4	14
2020-03-31	-23	-16	-31	-3	-28	-10	-29	-10
2020-06-30	10	27	14	38	15	25	16	28
2020-09-30	-7	15	-1	15	3	16	3	19
2020-12-31	5	28	17	18	19	22	21	14
2021-03-31	-4	22	12	18	6	9	11	11
Total	-58	182	25	183	22	157	18	144
Ret p.a.	-15	51	3	52	3	44	2	40
Risk p.a.	16	22	23	19	22	17	24	18
Ret/risk	-0.91	2.36	0.15	2.67	0.13	2.52	0.06	2.23
VaR p.a.	-41	-14	-40	-7	-38	-7	-43	-10
TE p.a.	7	8	9	9	7	6	9	6
IR p.a.	-4.75	4.67	-1.43	4.03	-1.93	4.68	-1.00	4.15

Table 12: Portfolios of deg[1], deg[2], deg[3], deg[4], factors= 5, VaR at 5%

Date	R2-d1	R2-d2	R2-d3	R2-d4	HR-d1	HR-d2	HR-d3	HR-d4
2017-06-30	0.4	0.4	0.4	0.4	62	57	56	57
2017-09-30	0.4	0.4	0.4	0.4	68	53	52	55
2017-12-31	0.4	0.4	0.4	0.5	80	45	43	48
2018-03-31	0.5	0.4	0.4	0.4	76	68	69	71
2018-06-30	0.3	0.4	0.4	0.4	68	53	55	52
2018-09-30	-0.0	-0.2	-0.2	0.3	77	53	47	48
2018-12-31	-0.3	-0.2	-0.2	-0.1	77	82	84	80
2019-03-31	0.1	0.4	0.4	0.2	83	40	48	58
2019-06-30	0.4	0.4	0.4	0.3	70	58	57	56
2019-09-30	0.1	0.3	0.3	0.4	63	55	55	55
2019-12-31	0.3	0.3	0.3	0.4	69	53	53	58
2020-03-31	0.3	0.4	0.4	0.4	91	92	90	85
2020-06-30	0.4	0.4	0.4	0.4	81	41	43	44
2020-09-30	0.4	0.3	0.3	0.4	76	53	53	59
2020-12-31	0.5	0.4	0.4	0.4	80	45	50	48
2021-03-31	0.5	0.4	0.4	0.4	84	50	49	51
Average	0.29	0.31	0.31	0.35	75	56	56	58

Table 13: R^2 and Hit Ratio for Degrees 1 to 4 (rounded to nearest 0)

Table 14: Signal Values per Polynomial Degree vs. IVV Benchmark

Date	Signal-deg1	Signal-deg2	Signal-deg3	Signal-deg4	IVV
2017-06-30	1	5	4	2	3
2017-09-30	-1	2	3	1	4
2017-12-31	-2	0	0	-3	7
2018-03-31	4	7	8	6	-1
2018-06-30	1	4	3	3	3
2018-09-30	-2	1	1	-1	8
2018-12-31	6	8	8	8	-13
2019-03-31	-5	-3	-3	-3	14
2019-06-30	0	1	2	1	4
2019-09-30	0	2	2	1	2
2019-12-31	-2	0	1	-1	9
2020-03-31	8	8	8	9	-20
2020-06-30	-5	-3	-3	-4	20
2020-09-30	-2	-1	-1	0	9
2020-12-31	-4	-4	-4	-4	12
2021-03-31	-2	-1	-1	-2	6
Average	**-0.30**	**1.73**	**1.78**	**0.85**	**4.24**

The Pairwise wins Table 15 counts how often each polynomial degree outperforms the others (based on adjusted R^2)

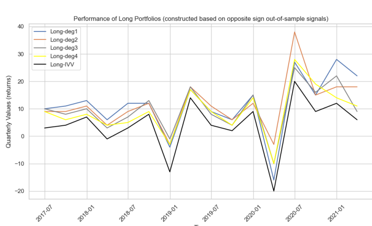


Figure 9: Actual Returns (ex-post)

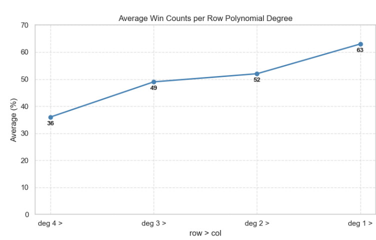


Figure 10: Adjusted R^2 Count

row > col	deg 1	deg 2	deg 3	deg 4	ave.
deg 4 >	34	38	37	n/a	36
deg 3 >	39	46	n/a	63	49
deg 2 >	40	n/a	54	62	52
deg 1 >	n/a	60	61	66	63

Table 15: Pairwise wins by degree (R^2)

Empirical analysis across polynomial degrees reveals that the deg $= 2$ model consistently outperforms alternatives in both return distribution characteristics and risk-adjusted performance. Notably, deg2 exhibits a heavier right tail, greater return dispersion, and a secondary mode [12] in its empirical distribution, suggesting the ability to capture extreme upside outcomes and high-return subgroups that are not as well represented by other models. Compared to deg1, deg2 delivers nearly identical annualized returns (52% vs. 51%) but at significantly lower volatility (19% vs. 22%), resulting in a superior Sharpe Ratio (2.67 vs. 2.36). When evaluated against deg3 and deg4, deg2 again dominates, offering higher returns (52% vs. 44% and 40%) at only marginally higher volatility. This efficiency in balancing return and risk supports the inference that deg2 second-order stochastically dominates deg1, deg3, and deg4 — a conclusion further reinforced by cumulative distribution function (CDF) crossings observed around the 10–12% mark. Collectively, these findings establish deg2 as the most effective degree for polynomial modeling in this context, achieving optimal trade-offs between complexity, explanatory power, and performance stability [13].

[12] Refers to another value (or multiple values) that also appears with a high frequency, but perhaps not as high as the primary mode

[13] The summary and inference comparison table 16 were developed with the assistance of ChatGPT (OpenAI).

Metric	deg2 vs deg1	deg2 vs deg3	deg2 vs deg4	Inference
PDF peak pos.	Similar	Similar	Similar	Comparable location
PDF shape	deg1 wider	deg3 narrower	deg4 narrower	Higher deg $\rightarrow \downarrow$var.
Right tail (+)	deg2 heavier	deg2 heavier	deg2 heavier	deg2 $\rightarrow \uparrow$extreme returns
CDF cross (CP*)	$\sim$10%	$\sim$12%	$\sim$12%	deg2 dominates for CP $>$ x%
Return dispersion	deg2 higher	deg2 higher	deg2 higher	deg2 $\rightarrow \uparrow$spread, not pure noise
2nd mode present	Yes (deg2 only)	Yes (deg2 only)	Yes (deg2 only)	deg2 captures high-return group
Risk profile	deg1 noisier	deg3 more stable	deg4 more stable	deg2 $\rightarrow$ best balance of risk/flexibility
Sharpe Ratio (SR)	2.67 $>$ 2.36	2.67 $>$ 2.52	2.67 $>$ 2.23	deg2 maximizes risk-adjusted return
Ann. Return / Std	52% / 19%	52% $>$ 44%	52% $>$ 40%	deg2 $\rightarrow \uparrow$return at $\downarrow$risk (strong dominance)

Table 16: Empirical Distribution Comparison: Degrees $[2]$ vs. $[1, 3, 4]$ (with Sharpe Ratios and Annualized Returns. *CP: Cumulative Probability)

I Appendix: A Conceptual Introduction to Adaptive Methods

The classical dynamics, while elegant and analytically tractable, implicitly impose model homogeneity. The SigGA avoids the uniform modeling assumption and instead adapts to asset-specific behavior and context. This flexibility empowers SigGA to capture path-dependent behavior, detect informational inefficiencies, and adapt to dynamic market conditions. By shifting the modeling process from rigid stochastic assumptions to adaptive empirical learning, SigGA aligns more naturally with the complexities of modern financial time series.

Aspect	Traditional Modeling	SigGA Approach
Model Assumption	Fixed stochastic process (e.g., GBM, Ornstein-Uhlenbeck)	Data-driven selection of linear, nonlinear, or hybrid models
Asset Classification	By asset type and sector	By asset type, sector, and data-learned functional form (linearity, nonlinearity)
Model Flexibility	Low; predetermined model structure	High; adaptive based on data patterns and learning
Temporal Resolution	Typically global/static	Dynamic: Expanding, Sliding, or Dyadic windows
Dimensionality Reduction	PCA or econometric heuristics	Rough path signatures + Genetic Algorithm (SigGA)
Variable Selection	Manual or via econometrics	Joint selection via fitness evaluation in supervised-learning framework
Use Case Fit	Best when theory-driven assumptions hold	Best when data complexity or irregularity invalidates fixed-process assumptions
Interpretability	High, theory-bound	Moderate to high, depending on selection and output interpretability

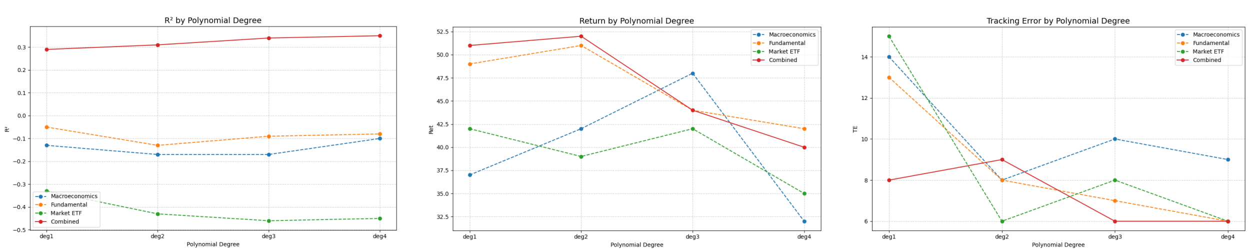
Table 17: Comparison between Traditional Financial Modeling and the SigGA Approach.

J Appendix: By Feature Grouping

06-2017 : 03-2021	deg1	deg2	deg3	deg4
Ret				
Macroeconomics	37	42	48	32
Fundamental	49	51	44	42
Market Data ETF	42	39	42	35
Combined	**51**	**52**	**44**	**40**
Risk				
Macroeconomics	29	23	19	23
Fundamental	29	22	22	21
Market Data ETF	31	20	18	17
Combined	**22**	**19**	**17**	**18**
Ret / Risk				
Macroeconomics	1.28	1.80	2.53	1.38
Fundamental	1.70	2.31	2.03	2.04
Market Data ETF	1.35	1.91	2.40	2.09
Combined	**2.36**	**2.67**	**2.52**	**2.23**
TE				
Macroeconomics	14	8	10	9
Fundamental	13	8	7	6
Market Data ETF	15	6	8	9
Combined	**8**	**9**	**6**	**6**
IR				
Macroeconomics	1.44	3.05	3.08	1.89
Fundamental	2.49	4.71	3.72	4.39
Market Data ETF	1.69	4.09	3.41	2.14
Combined	**4.67**	**4.03**	**4.68**	**3.82**

Table 18: Annualized Performance metrics across Polynomial degrees

Performance metrics suggest the presence of information inefficiencies across polynomial degrees. Combined signals consistently outperform individual groups. The positive R^2 of the combined dataset vs. negative R^2 of the individual datasets suggest that the selected best five factors are most likely drawn from all three datasets.



(a) Out-of-sample R^2 (b) Returns Annualized (ex-post) (c) Tracking Error (TE)

Figure 11: Performance metrics across polynomial degrees 1–4 using top five factors.

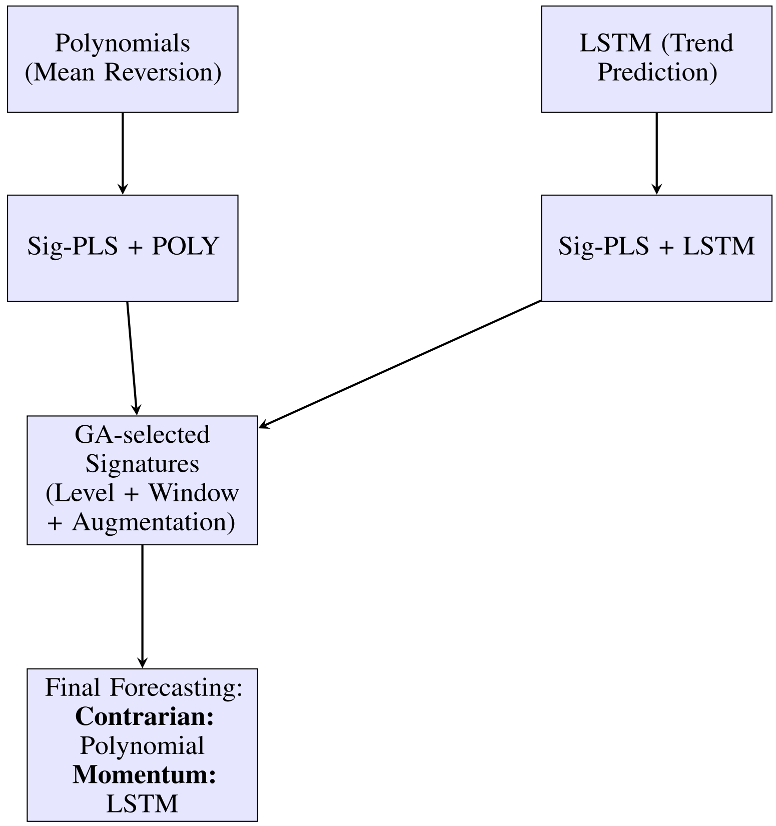


Figure 12: Evolution of Models Toward Adaptive Learning Architecture

K Appendix: Learning Nonlinearity from Signatures

Notation: Path Spaces and Signature Input

The symbol S in $S(\mathbb{R}^{d+1})$ denotes the signature transform space of a path — that is, it's not just any function or point in $\mathbb{R}^{d+1}$, but a path-valued object prepared for signature computation. $S(\mathbb{R}^{d+1})$ has two common interpretations:

1. Space of piecewise continuous paths in $S(\mathbb{R}^{d+1})$: This is the most natural space on which signatures are defined. The notation indicates you're working with augmented paths — not raw data points — that can be input to the signature map $Sig^N : S(\mathbb{R}^{d+1}) \to T^N(\mathbb{R}^{d+1})$, where T^N is the truncated up to level N tensor algebra. $S(\mathbb{R}^{d+1}) := \{\gamma : [0,T] \to \mathbb{R}^{d+1}$ continuous, finance variation$\}$

2. Space of discretized sequences for signature input: In practice, discrete signatures can be computed using Python's *iisignature* or *esig*, you can interpret: $S(\mathbb{R}^{d+1}) := (\mathbb{R}^{d+1})^n$ as the space of discrete paths of length $n \in \mathbb{R}^{d+1}$.

Throughout this work, we denote by $S(\mathbb{R}^{d'})$ the space of discrete time-series paths in $\mathbb{R}^{d'}$, where d' refers to the dimension of the augmented path after applying transformations such as time augmentation, lead-lag, or basepoint. Formally, for a path $x = (x_1, \ldots, x_n) \subset \mathbb{R}^d$, we define an augmentation map $\phi : \mathbb{R}^d \to \mathbb{R}^{d'}$ (e.g., $\phi = \phi^{b,t}$ for basepoint and time augmentation), yielding an augmented path:

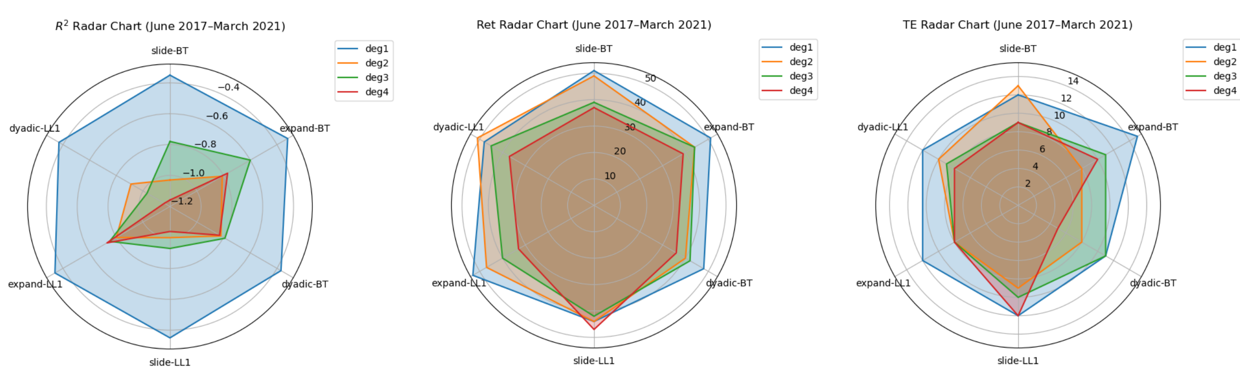
$$\phi(x) = ((t_0, x_0), (t_1, x_1), \ldots, (t_n, x_n)) \in S(\mathbb{R}^{d'}).$$

Here, the set $S(\mathbb{R}^{d'})$ represents the space of **discretely sampled, finite-variation paths** suitable for signature computations. In practice, these are represented as sequences in $(\mathbb{R}^{d'})^n$, and interpolated piecewise-linearly when computing the signature via libraries such as `iisignature`.

The signature transform of depth N, denoted $\mathrm{Sig}^N : S(\mathbb{R}^{d'}) \to T^N(\mathbb{R}^{d'})$, maps such augmented paths into the truncated tensor algebra of order N, encoding their higher-order pathwise interactions.

Table 19: Performance Metrics Across Signature Degrees (06-2017 to 03-2021)

Metric	deg1	deg2	deg3	deg4
Ret				
slide-BT	51	49	39	37
expand-BT	51	44	44	39
dyadic-BT	48	40	42	36
slide-LL1	44	44	42	47
expand-LL1	53	47	40	33
dyadic-LL1	48	51	45	37
Combined	**53**	**51**	**45**	**47**
Risk				
slide-BT	27	26	22	24
expand-BT	30	19	24	23
dyadic-BT	23	21	22	19
slide-LL1	26	20	19	25
expand-LL1	27	22	22	23
dyadic-LL1	26	24	17	17
Combined	**23**	**19**	**17**	**17**
Ret / Risk				
slide-BT	1.88	1.89	1.81	1.53
expand-BT	1.72	2.27	1.81	1.69
dyadic-BT	2.09	1.87	1.88	1.85
slide-LL1	1.69	2.19	2.18	1.89
expand-LL1	1.97	2.17	1.81	1.45
dyadic-LL1	1.87	2.13	2.70	2.24
Combined	**2.09**	**2.27**	**2.70**	**2.24**
TE				
slide-BT	12	13	9	9
expand-BT	15	8	11	10
dyadic-BT	11	8	11	5
slide-LL1	12	9	10	12
expand-LL1	12	8	8	8
dyadic-LL1	12	10	9	8
Combined	**11**	**8**	**8**	**5**
IR				
slide-BT	2.86	2.54	2.52	2.36
expand-BT	2.40	3.66	2.42	2.21
dyadic-BT	2.82	3.04	2.40	3.61
slide-LL1	2.43	3.08	2.46	2.56
expand-LL1	3.08	4.08	2.94	2.20
dyadic-LL1	2.78	3.38	3.42	2.65
Combined	**3.08**	**4.08**	**3.42**	**3.61**



(a) Out-of-sample R^2 (b) Returns Annualized (ex-post) (c) Tracking Error (TE)

Figure 13: Performance metrics across polynomial degrees 1–4 using top five factors.