HODL Strategy or Fantasy?

480 Million Crypto Market Simulations and the Macro-Sentiment Effect

Weikang Zhang Alison Watts

Abstract

Crypto enthusiasts claim that buying and holding crypto assets yields high returns, often citing Bitcoin's past performance to promote other tokens and fuel fear of missing out. However, understanding the real risk–return trade-off and what factors affect future crypto returns is crucial as crypto becomes increasingly accessible to retail investors through major brokerages. We examine the HODL strategy through two independent analyses. **First**, we implement 480 million Monte Carlo simulations across 378 non-stablecoin crypto assets, net of trading fees and the opportunity cost of 1-month T-bills. We find strong evidence of survivorship bias and extreme downside concentration. At the 2–3 year horizon, the median excess return is -28.4% and $\mathrm{CVaR}_{1\%} \approx 1.08$, implying tail scenarios wipe out the principal after deducting all costs, while the top-quartile mean reaches $+1{,}326.7\%$. These results challenge the "HODL" narrative: across a broad set of crypto assets, simple buy-and-hold loads extreme downside risk onto most investors, and the "miracles" mostly belong to the luckiest quarter. **Second**, using Bayesian multi-horizon local projection, we find that endogenous predictors (realized risk-return metrics) exhibit economically negligible population-level effects with limited cross-basket consistency. In contrast, macro-finance factors, particularly the 24-week exponential moving average of the Fear & Greed Index, demonstrate persistent long-horizon effects and high cross-basket stability: where significant, a one-standard-deviation shock reduces forward top-quartile mean excess returns by 15–22 percentage points and median returns by 6–10 percentage points over 1–3 year horizons. These findings imply that realized risk-return distributions offer limited generalizable guidance for investment decisions; macro-sentiment conditions emerge as the dominant indicators of future outcomes.

1 Introduction

"HODL", a popular crypto slang term for buy-and-hold investing, has become a widely promoted strategy in cryptocurrency markets. But does this approach in fact deliver superior long-run outcomes, or does it instead expose typical retail investors to greater risk?

Understanding the risk-return characteristics of cryptocurrency buy-and-hold strategies has become increasingly important as the global cryptocurrency market cap has surpassed \$3 trillion[1] and major brokerages such as Robinhood and Fidelity now offer crypto trading on their platforms. Prior research has documented that cryptocurrencies exhibit extreme tail risk (Gkillas and Katsiampa, 2018; Osterrieder and Lorenz, 2017; Nzokem and Maposa, 2024) and respond to macroeconomic shocks (Ma et al., 2022; Karau, 2023), but two critical questions remain unanswered. First, what are the market-wide risk and return distributions of buy-and-hold outcomes when accounting for transaction costs, opportunity costs, and flexible holding horizons? Second, should investors rely on current realized risk-return distributions to estimate their future outcomes? This paper addresses

[1] As of November 2025; source: CoinGecko, `https://www.coingecko.com/en/charts`.

both questions through large-scale Monte Carlo simulation and Bayesian multi-horizon local projection analysis.

Crypto influencers often cite Bitcoin's success story to promote other tokens, but how far is the distance between the average investor and the luckier top-quartile group? We implement 480 million Monte Carlo simulations across eight baskets—seven major tokens (BTC, ETH, ADA, BNB, DOGE, LINK, XRP) and one market-wide basket randomly sampling from 378 non-stablecoin cryptocurrencies. In the market-wide basket at the longest holding horizon (731–1095 days), the median excess return is −28.4%, while the mean among the top 25% reaches +1,326.7%. Beyond 181 days, more than 50% of scenarios result in losses exceeding 10%, with tail risk ($\text{CVaR}_{1\%}$) approaching total capital loss after deducting opportunity costs and transaction fees. This disparity exposes a dangerous form of survivorship bias embedded in the cryptocurrency investment narrative.

To address whether investors can rely on current realized risk-return distributions to estimate future outcomes, we implement a Bayesian multi-horizon local projection framework that jointly models endogenous predictors (realized risk-return metrics) and stability-selected macro-finance factors with distinct hierarchical shrinkage structures across forecast horizons. Our analysis reveals a fundamental asymmetry: while some basket-specific endogenous predictors exhibit strong effects, a Bayesian random-effects meta-analysis shows their population-level influence remains economically negligible with limited cross-basket consistency. In contrast, macro-finance sentiment indicators demonstrate both persistent long-horizon effects and high cross-basket stability. Where significant, the 24-week exponential moving average of the Fear & Greed Index emerges as the most stable predictor across baskets, reducing forward mean excess returns of the top 25% by 15–22 percentage points and median returns by 6–10 percentage points per standard-deviation shock over 1–3 year horizons. This asymmetry implies that realized risk-return distributions offer limited generalizable guidance for long-term investment decisions. Sentiment effects, however, exhibit substantial heterogeneity: meme coins such as DOGE display approximately 3 times greater sensitivity than established cryptocurrencies.

Our findings challenge both pillars of the HODL narrative. First, the market-wide risk-return distribution does not justify buy-and-hold for most investors: median outcomes are deeply negative, tail risk approaches total loss, and the gap between typical and top-quartile returns reflects survivorship bias rather than attainable outcomes. Second, past realized distributions offer limited generalizable guidance for future outcomes: endogenous risk-return metrics exhibit economically negligible population-level effects, while macro-finance sentiment indicators—particularly the Fear & Greed Index—demonstrate persistent long-horizon effects with high cross-basket stability. These results suggest investors should focus on macro-sentiment conditions and macroeconomic factors rather than extrapolate from historical performance or selective success stories.

2 Literature Review

Empirically testing "HODL" is challenging because retail investors have heterogeneous beliefs about market cycles, different holding periods, and different levels of tolerance for extreme tail outcomes. To simulate cryptocurrency returns, two branches of simulation approaches are common. The first branch is parametric price-path Monte Carlo: in this type of simulation, researchers usually assume an explicit stochastic data-generating process (DGP), then simulate full price paths to evaluate risk and return distributions. For example, Likitratcharoen et al. (2021) implement a parametric price-path Monte Carlo by assuming Bitcoin follows a geometric Brownian motion and simulating 1,000,000 daily returns with pseudo-random normal shocks to compute VaR. The second branch is nonparametric simulation: rather than specifying an explicit DGP, studies generate synthetic long-

horizon paths by stationary block-bootstrap resampling of historical returns (Politis and Romano, 1994) to preserve serial dependence and non-Gaussian features.

From the perspective of retail investors, there are a few questions often asked about cryptocurrency investments: What are the expected returns? What is the downside risk? Would a longer holding horizon decrease risks? Regarding expected returns, empirical evidence suggests that cryptocurrency returns are systematically related to market capitalization. Li et al. (2020), in research across more than 1,800 cryptocurrencies , find that returns are related to size : small-cap coins earn significantly higher subsequent returns than large-cap coins.

When turning to the question of downside risk, researchers have employed extreme value theory and related statistical methods to quantify the tail risk properties of cryptocurrencies. Gkillas and Katsiampa (2018) apply extreme value theory to the return tails of five major cryptocurrencies and, using Value-at-Risk (VaR) and Expected Shortfall (ES) as tail risk measures, find that Bitcoin Cash is the riskiest, whereas Bitcoin and Litecoin exhibit the lowest tail risk. Compared with major fiat (G10) currencies, Bitcoin's returns exhibit heavier tails, much higher volatility, and substantially larger losses in extreme events (Osterrieder and Lorenz, 2017). In another study on crypto tail risk relative to traditional markets, Aubain Nzokem and Daniel Maposa show that, compared with the S&P 500, Bitcoin's daily returns are markedly more heavy-tailed, making extreme moves more likely (Nzokem and Maposa, 2024).

The third question is whether longer holding horizons reduce risk. This has been examined in both traditional and cryptocurrency markets. For example, using a stationary block bootstrap on 39 developed markets (1841–2019), Anarkulova et al. (2022) show that while average real payoffs rise with horizon, there is still about a 12% chance of an inflation-adjusted loss over 30 years. In the context of holding horizon and holding risks in the cryptocurrency market, Conlon et al. (2024) find that, on average, Bitcoin behaves as a short-horizon hedge for USD exchange rates; however, at longer horizons it can co-move positively with large USD losses, so a longer holding horizon does not guarantee reduced risk in crypto holdings. Another study analyzes the comovement between Bitcoin and the U.S. stock market and finds right-tail dependence that is stronger at long holding horizons; this co-move effect decreases significantly as the horizon shortens from yearly to monthly (Maghyereh and Abdoh, 2021).

After understanding the risk and return distribution and their dynamics over time, a natural question arises: What factors may affect these distributions? Although macroeconomic shocks are known to influence cryptocurrency returns (Ma et al., 2022; Karau, 2023), a comparative impulse-response framework that jointly quantifies the effect sizes of endogenous risk-return metrics versus macro-finance factors across multiple forecast horizons has not been explored. A natural framework for studying such shock transmission is impulse response analysis. Local Projections (LP; Jordà (2005)) estimate impulse responses without fully specifying a multivariate dynamic system. At the population level, with unrestricted lag structures and the same shock identification, LPs and VARs target the same impulse responses; differences are primarily finite-sample bias–variance trade-offs. Later research by Kilian and Kim (2011) shows that, in small samples, LP intervals are often wider and have less accurate coverage than bias-adjusted (bootstrap) VAR intervals. To improve LP estimation, Barnichon and Brownlees (2019) note that LP impulse responses can be highly variable and jagged in practice and propose Smooth Local Projections, penalized B-spline (ridge) smoothing across horizons, to deliver smoother, more precise impulse responses without altering identification. Starting from the local-projection setup, Ferreira et al. (2025) introduce Bayesian Local Projections (BLP), which shrink LP coefficients through hierarchical informative priors to handle the LP, VAR bias, variance trade-off while preserving LPs' robustness to misspecification. With informative priors, BLP delivers uncertainty comparable to a BVAR with standard macro priors, and its posterior mean can be viewed as an optimally weighted combination of LP and

(iterated) VAR information at each horizon.

When applying impulse response methods to high-dimensional settings with many potential predictors, variable selection becomes crucial for two reasons. First, macroeconomic variables are often highly correlated, making it difficult to identify individual effects. Second, identifying a parsimonious set of predictors enhances interpretability and improves the stability of coefficient estimates. Two prominent approaches help identify the relevant predictors among many candidates: regularizing priors and stability selection. To handle high-dimensional Bayesian models, the horseshoe global–local prior adapts to unknown sparsity by aggressively shrinking near-zero coefficients toward zero while preserving large signals (Carvalho et al., 2010). Empirically, horseshoe-type priors, with a calibrated global scale and the regularized horseshoe, lead to more reliable estimation and more stable posterior computation (Piironen and Vehtari, 2017b,a).

A complementary, resampling-based approach to identifying robust predictors is stability selection, which addresses selection instability in high-dimensional, correlated settings. In correlated, high-dimensional problems, CV with unstable selectors leads to noisy error curves and unstable selections (Breiman, 1996). Following Breiman's instability–stabilization view, stability selection repeatedly applies a sparse selector to many half-sample subsamples and retains variables with high selection frequency, and it provides error-control through an upper bound on the expected number of false selections (Meinshausen and Bühlmann, 2010; Shah and Samworth, 2013). This idea extends naturally to joint-sparsity settings: recent studies pair multitask group-Lasso with stability selection, yielding more stable feature selection under LD-driven collinearity (Nouira and Azencott, 2022, 2025).

Empirical studies suggest cryptocurrencies are not insulated from macroeconomic shocks. On FOMC days, a 1-bp unexpected tightening (measured by the two-year Treasury yield) sends Bitcoin down about 0.25% immediately, with a larger, persistent multi-day decline; the effect is strongest in bull markets (Ma et al., 2022). Since late 2020, unexpected U.S. monetary tightenings depress Bitcoin prices (Karau, 2023).

Beyond macroeconomic shocks, sentiment-driven dynamics also play a significant role in cryptocurrency markets. Investor sentiment—both rational and irrational—significantly and positively drives Bitcoin returns, especially after COVID-19, with patterns associated with FOMO (fear of missing out) (Güler, 2023). Relatedly, Li et al. (2021) find that a higher prior MAX—the largest daily return in the past month—predicts higher subsequent returns.

2.1 Research Gap and Contributions

The literature above identifies two simulation approaches for cryptocurrency returns (parametric price-path Monte Carlo and nonparametric block-bootstrap), demonstrates that cryptocurrencies respond to macro-finance shocks, and introduces Bayesian local projection methods for studying impulse responses.

However, two critical gaps remain. First, while prior simulation studies either assume stylized parametric processes or apply block-bootstrap to historical returns, existing work does not simultaneously combine flexible holding horizons (spanning 1 to 1,095 days), realistic transaction costs benchmarked against risk-free returns, and broad cryptocurrency coverage. Prior studies focus on Bitcoin or a few high-cap tokens, leaving survivorship bias and downside risk concentration across the broader market unquantified. We address this by simulating 480 million episodes across eight baskets, including an aggregate basket of 378 non-stablecoin cryptocurrencies, to capture both token-specific and market-wide risk distributions.

Second, while prior work documents crypto responses to macroeconomic and sentiment shocks (Ma et al., 2022; Karau, 2023; Güler, 2023), to our knowledge there is no published framework that

jointly (i) models multiple risk–return targets, (ii) contrasts basket-specific endogenous distribution measures with macro-finance predictors, and (iii) evaluates these relations across horizons. We develop a Bayesian multi-horizon local projection model to estimate impulse responses of four risk-return metrics at six forecast horizons (30, 90, 180, 365, 730, and 1,095 days), quantifying the relative strength and persistence of basket-specific risk-return responses to these two predictor types.

3 Data

We first compiled the top 800 crypto assets by market capitalization from CoinMarketCap[2]. For each asset, we mapped the CoinMarketCap symbol to a Yahoo Finance ticker of the form SYMBOL-USD (e.g., BTC-USD), and, if unavailable, attempted the raw SYMBOL as a fallback. Using the yfinance Python package, we then downloaded the longest available daily data, specifically the High, Low, Close, and Volume, for each asset.

Data cleaning We removed assets whose first available date was on or after 2024-01-01. We then removed stablecoins by computing each token's average daily Close and the standard deviation of Close over the sample; tokens with an average daily Close within $[0.97, 1.03]$ USD and a standard deviation ≤ 0.03 were classified as stablecoins and excluded. We removed tokens whose average daily volume over the last 365 calendar days of the sample was $<$ \$100,000 and tokens whose latest observation was before 2025-04-26. For missing data, we trimmed each token's history to begin on the first date on which High, Low, Close, and Volume are all observed. Tokens that still contained any missing values thereafter were dropped. As a final quality screen, we excluded any token with ≥ 10 days where High $= 0$ or Volume $= 0$, or ≥ 10 days with Volume $<$ \$500.

Beyond the crypto asset panel, we also collect supplementary macro-financial indicators. We collect the Crypto Fear & Greed Index[3] via Alternative.me's API, aggregate the daily series to a weekly mean (Monday–Sunday), and assign each week's timestamp to the Monday at the start of that week.

We collect macro-finance series from FRED, take the last available business day each week (Friday, or Thursday when Friday is missing) as the weekly value, stamp it on the Monday of that week, and list the associated codes and series names in Appendix A.2.1.

We also compute the log return of Bitcoin. Let P_t^{BTC} be the closing price on the last day of week t (Sunday; if missing, use the last available trading day that week). The weekly log return is

$$r_t^{\mathrm{BTC}} = \ln\left(\frac{P_t^{\mathrm{BTC}}}{P_{t-1}^{\mathrm{BTC}}}\right).$$

For alignment, we reindex each r_t^{BTC} to the Monday that starts week t.

4 Monte Carlo Simulation

Our Monte Carlo simulation approximates the full distribution of one-time buy–hold–sell outcomes that a crypto investor could experience without perfect foresight. We executed 10,000,000 simulations for each of six holding-period intervals per basket—480,000,000 in total across eight baskets.

[2] `https://coinmarketcap.com`, homepage pages 1–8; accessed April 2025.

[3] A market–sentiment gauge for crypto (primarily Bitcoin), scaled 0–100 where 0 denotes "Extreme Fear" and 100 "Extreme Greed"; source: Alternative.me, `https://alternative.me/crypto/fear-and-greed-index/`.

Table 1: Notation for the Simulation Section

Symbol	Description
ℓ, h	Lower/upper bound of allowed holding horizons (days)
τ_i	Holding period of episode i (days)
T	Length of the selected basket-price sample (days)
s_i, e_i	Buy and sell day indices for episode i
c_i	Index of the selected coin in the basket
$H_{t,c}, L_{t,c}$	Daily high and low prices of coin c on day t
P_i^{buy}	Simulated buy price for episode i
P_i^{sell}	Simulated sell price for episode i
r_t^{rf}	Daily simple risk-free rate
γ_t	Cumulative log risk-free return up to day t
R_i^{rf}	Risk-free return for episode i's holding period
ϕ	One-way proportional trading fee
G_i	Net return of episode i after fees
X_i	Excess return of episode i over cash

4.1 Draw a Holding Horizon

We pre-set six inclusive horizon intervals: 1–30, 31–90, 91–180, 181–365, 366–730, 731–1095. For each simulated episode, the simulator randomly picks an integer from the lower bound to the upper bound of the chosen interval. This design gives the model flexible holding periods: the shortest interval $[1, 30]$ captures investors who treat crypto almost like a checking-account balance, ready to move in and out, whereas the longest interval $[731, 1095]$ represents holders who worry less about short-run fluctuation and focus on multi-year returns.

Given inclusive horizon bounds $\ell < h$,

$$\tau_i \sim \text{DiscreteUniform}\{\ell, \ell+1, \ldots, h\} \quad (i = 1, \ldots, n),$$

where τ_i denotes the randomly drawn holding horizon for episode i.

4.2 Determine Buy & Sell Dates

For each holding horizon, the simulator determines the latest possible buying and selling dates. Since Python indexing starts at 0, we subtract 1 to ensure the latest buying date remains within the tokens' data range.

Let T denote the total length of the dataset in trading days. The buy date s_i must be chosen such that it leaves enough days (τ_i) within the selected time horizon, ensuring that the sell date e_i does not exceed the dataset boundaries:

$$0 \;\le\; s_i \;\le\; T - \tau_i - 1, \qquad e_i \;=\; s_i + \tau_i \;<\; T.$$

Accordingly, we randomly select the buy date from a discrete uniform distribution:

$$s_i \;\sim\; \text{DiscreteUniform}\{0, 1, \ldots, T - \tau_i - 1\}, \qquad e_i = s_i + \tau_i.$$

In the situation of $T - \tau_i - 1 < 0$, the draw is discarded.

4.3 Draw a Coin

We construct eight baskets in total. Seven baskets are single-token baskets (`BTC`, `ADA`, `BNB`, `DOGE`, `ETH`, `LINK`, `XRP`), while the `ALL` basket contains the entire dataset comprising $C = 378$ tradable tokens.

For each simulated episode, we select a coin by uniformly sampling from pre-assigned consecutive integer indices:

$$c_i \sim \text{DiscreteUniform}\{0, \ldots, C-1\},$$

where c_i denotes the coin selected for episode i. When the basket contains only a single token ($C = 1$), this selection becomes deterministic, with $c_i = 0$.

4.4 Data Validity Filter

When we draw a coin from the `ALL` basket, some coins may not have values on certain dates, introducing null values into our model. Therefore, we apply a validity filter to ensure that all sampled price points are valid.

Let $H_{t,c}$ and $L_{t,c}$ denote the daily high and low prices, respectively, of coin c on day t. An episode is accepted only if:

$$H_{s_i,c_i} \geq L_{s_i,c_i} > 0, \quad H_{e_i,c_i} \geq L_{e_i,c_i} > 0.$$

If these conditions are not met, the draw is rejected and resampled.

4.5 Loop of Sampling

To ensure exactly n valid episodes are collected per holding-period interval, especially after filtering out invalid data points due to varying token lengths and the potential lack of data for longer horizon bands, we iteratively sample and validate episodes until the target number is reached or until 50 consecutive failed attempts occur. After each sampling iteration, the remaining episodes needed are updated as:

$$\text{needed episodes} = n - (\text{number of valid episodes collected})$$

4.6 Intra-Day Trading

As we defined the buy and sell dates in the previous section, the model ensures that a token's buying date will always be earlier than its selling date. The simulator then draws its buy and sell prices as follows:

$$P_i^{\text{buy}} = L_{s_i,c_i} + U_i^{(1)}(H_{s_i,c_i} - L_{s_i,c_i}), \tag{1}$$

$$P_i^{\text{sell}} = L_{e_i,c_i} + U_i^{(2)}(H_{e_i,c_i} - L_{e_i,c_i}), \tag{2}$$

where $U_i^{(1)}, U_i^{(2)} \overset{\text{i.i.d.}}{\sim} \text{Uniform}(0,1)$.

4.7 Risk-Free Benchmark over the Same Holding Period

To calculate the *excess* return for each crypto investment episode, we benchmark it against the return an investor would have earned by continuously rolling over the 1-month U.S. Treasury bill across the exact same calendar days.

Let r_t^{rf} denote the *daily* risk-free simple return. We first pre-compute the cumulative log returns up to each day t:

$$\gamma_t \;=\; \sum_{s=0}^{t} \ln\!\big(1 + r_s^{\text{rf}}\big).$$

Then, for an episode i with buy date s_i and sell date e_i, the holding-period-specific risk-free return is calculated as:

$$R_i^{\text{rf}} \;=\; \exp\!\big(\gamma_{e_i} - \gamma_{s_i}\big) - 1.$$

4.8 Net Trade Return After Fees

To bring the simulation closer to real-world trading, we apply a proportional fee to *each side* of every spot trading. Although fee schedules vary across exchanges, users typically pay on the order of 0.10 % per side. We therefore set the one-way fee at $\phi = 0.001$[4]

Let ϕ denote that one-way proportional fee. If a given episode buys at P_i^{buy} and later sells at P_i^{sell}, the net (post-fee) return is

$$G_i \;=\; (1-\phi)\,\frac{P_i^{\text{sell}}}{P_i^{\text{buy}}}\,(1-\phi) \;-\; 1.$$

4.9 Excess Return

For every valid simulated episode we compute the *excess* return as

$$X_i \;=\; G_i \;-\; R_i^{\text{rf}}.$$

Here G_i is the net trade return after fees, and R_i^{rf} is the corresponding risk-free benchmark.

5 Monte Carlo Simulation Results

After conducting 480,000,000 Monte–Carlo episodes, we calculate a series of statistical metrics that include central tendency (mean and median), standard deviation, downside tail risk (VaR and CVaR), non-annualized risk-adjusted return (Sharpe and Sortino ratios), the proportion of episodes with losses $> 10\%$[5], the 75-th-percentile excess return, and the mean excess return above that threshold. These statistical metrics are computed in both overall and weekly aggregations, with minor differences: the overall (non-weekly) aggregation[6] employs a 1% threshold for downside risk and includes additional statistics such as the inter-quartile range, skewness, and kurtosis. In

[4] At the default (VIP 0) tier, major crypto exchanges quote taker fees of approximately 0.10%—Binance 0.10%, Gate.io 0.10%, and OKX 0.10% (maker 0.08%). Figures are taken from the exchanges' official fee schedules (accessed May 2025).

[5] `p_sig_loss` in the code, defined as excess return $< -10\%$.

[6] The overall perspective is structured at two complementary levels: (i) aggregated by holding horizon and basket, and (ii) aggregated by basket, irrespective of holding horizon.

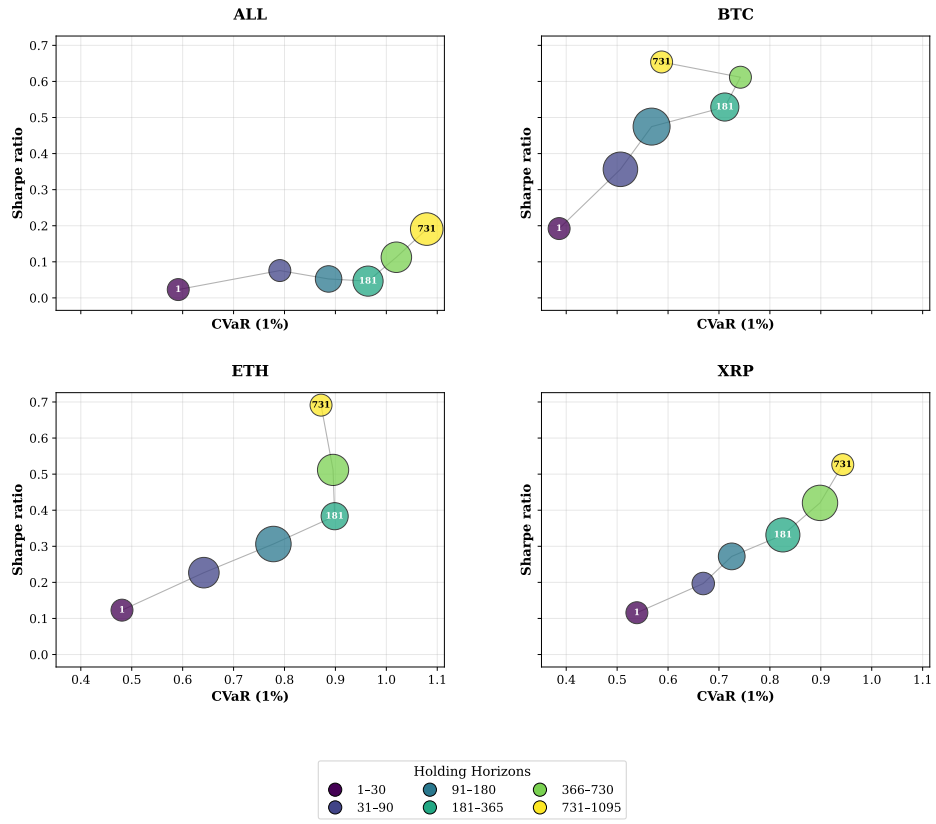


Figure 1: Risk–Return Trade-Off Across Holding Horizons

contrast, the weekly aggregation sets the downside risk threshold at 10%, omitting skewness and kurtosis due to imbalanced weekly sample sizes throughout the dataset. Comprehensive explanations and formulas are detailed in Appendix A[7].

Figure 1 illustrates the risk–return trade-off across different holding horizons for four baskets based on our simulation results, while the remaining four baskets are shown in Appendix B, Figure 10. The x-axis represents the 1% Conditional Value-at-Risk (CVaR), indicating the average loss experienced in the worst 1% of scenarios, while the y-axis depicts the Sharpe ratio. Both metrics are computed separately for each basket and holding horizon. The size of each bubble represents the proportion of outcomes with losses exceeding 10%. Bubble sizes are calculated for each basket–holding-horizon pair and then rescaled relative to the mean and dispersion of that basket's other horizons; larger bubbles therefore indicate a higher-than-average risk of losses within the same basket, while smaller bubbles indicate comparatively lower risks.[8]

From Figure 1 we can observe that, as the length of the holding horizon increases, the CVaR(1%) climbs markedly for the ALL and XRP baskets. The ETH basket exhibits a similar upward trajectory through the 181–365-day horizon, but its tail risk subsequently moderates, declining from 0.899 to 0.872 in the longest period. Even though BTC performs relatively better than the other baskets in this simulation, its $\text{CVaR}_{1\%}$ falls while the Sharpe ratio increases between the

[7]Refer to Appendix A, Section A.1.3 (§A.1.3) for the definitions of all statistics; see Sections A.1.1–A.1.2 (§A.1.1–§A.1.2) for notation and tail-probability thresholds.

[8]Bubble area $A = \min\{\max[\,400\,(1 + 1.5\,z),\,400],\,2800\}$, with $z = (p_{\text{sig loss}} - \mu)/\sigma$. This design ensures that the bubbles remain visible and emphasizes the risk associated with each holding horizon.

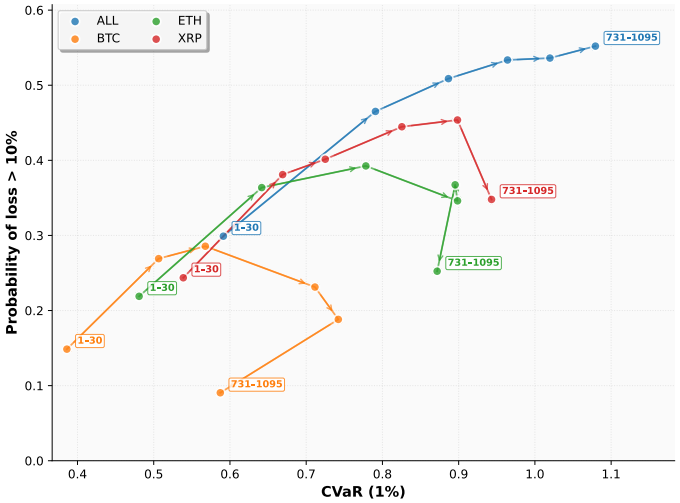


Figure 2: Trade-off between moderate and extreme risks across holding horizons

181–365-day and 731–1095-day horizons; however, the longest horizon still shows a 58.7 % tail loss. One notable point that grabs our attention is the ALL basket. As defined in our Monte-Carlo simulation, this basket randomly selects a single coin from 378 tokens for each episode. This method captures the market-wide risk–return distribution rather than the experience of any single investor. The simulation highlights a critical insight: buy and hold is not a safe strategy for most cryptocurrencies. Specifically, the CVaR at 1% approaches nearly 1 for holding horizons between 181 and 365 days, and exceeds 1 for longer horizons, indicating that the worst 1% of outcomes not only wipe out the principal but also forgo the alternative returns from Treasury bills. Given that each scenario is simulated independently millions of times, these findings reflect systemic market risk rather than a few outliers.

Figure 2 reveals the trade-off between moderate risk (probability of losses greater than 10%) and extreme risk (CVaR 1%). The x-axis represents the CVaR (1%), and the y-axis represents the probability of losses exceeding 10%. Arrows indicate the progression from shorter to longer holding periods, with clearly marked starting and ending periods. We plot four baskets in Figure 2 for clarity, and an additional four are provided in Figure 11.

Some argue that purchasing cryptocurrencies, forgetting about them, and never selling ("HODL"[9]) is an ideal investment strategy. Our analysis addresses this claim from two perspectives. First, we find that for certain baskets (tokens), longer holding horizons, typically from 181–365 days onward, show the probability of moderate losses beginning to decline, while extreme risk (CVaR) rises through mid-range horizons before declining at the longest holding periods. However, this pattern is not generalizable when considering the crypto market as a whole. In fact, observations from the ALL basket show that, as the holding period lengthens, both the probability of experiencing moderate losses and the severity of extreme losses increase significantly. Specifically, the worst 1% group's losses approach 100% of their initial investment, and the probability of experiencing losses greater than 10% exceeds 50% once the holding period extends beyond 181–365 days. Crucially, this scenario is not extreme but represents the overall distribution observed in our data. In summary, cryptocurrency investors should be cautious and avoid relying solely on successful cases like Bitcoin and Ethereum to justify the 'HODL' strategy. In most situations, both moderate and extreme risks rise as the holding period lengthens.

[9]HODL is a term commonly used in the crypto community meaning to buy-and-hold indefinitely.

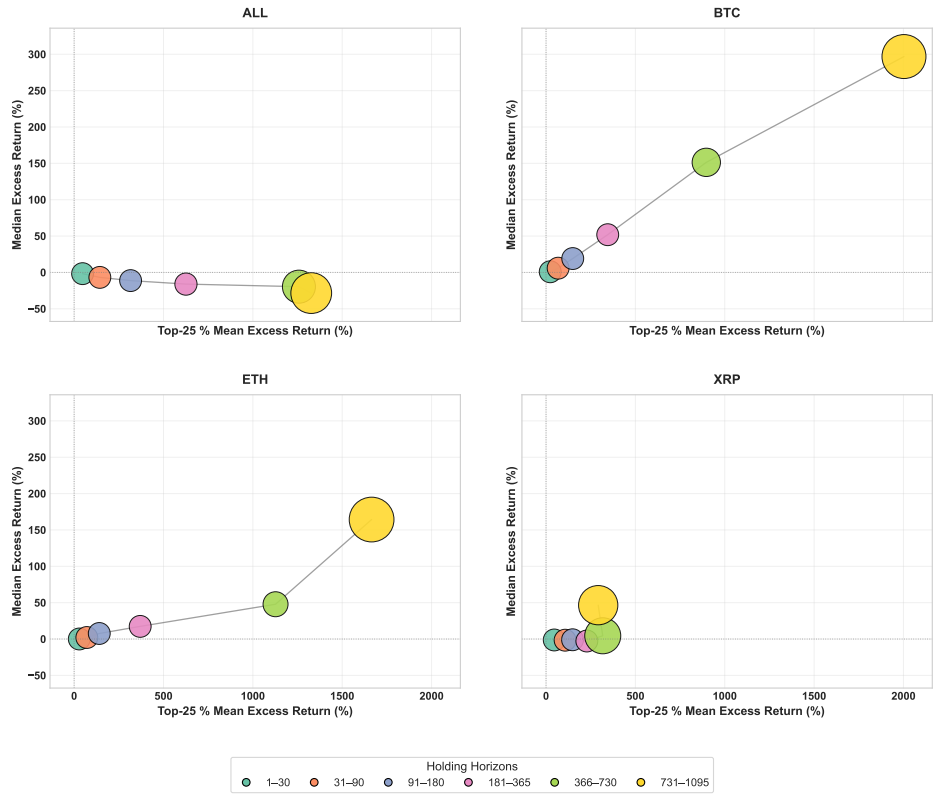


Figure 3: Median vs. top-25% mean excess returns across holding horizons

Figure 3 plots four baskets, showing median excess returns against the mean excess return of the top-25 % observations; the remaining four baskets are presented in Figure 12. The x- and y-axes are expressed in percent. Colours encode holding horizons, whereas bubble area visualises dispersion: within each basket the area is proportional to the inter-quartile range (IQR) of excess returns for that horizon. A larger bubble therefore indicates that the central 50 % of observations are more widely spread (higher volatility), while a smaller bubble indicates a tighter clustering (lower volatility) relative to other horizons in the same basket.[10]

From Figure 3 we observe that, as the holding horizon increases, the top-25 % mean excess return generally rises. XRP shows a slight backward step when moving from the 366–730-day window to the 731–1095-day window, yet its overall trend remains upward. Dispersion (IQR) for each basket tends to increase with holding length, especially across the 366–1095-day span. Viewed token-by-token, this pattern appears to support the "HODL" claim. The ALL basket, however, tells a different story: while its top-quartile mean at the longest horizon (731–1095 days) exceeds 1326.7 %, the median return is -28.4 %. Its larger IQR confirms the wider spread of outcomes. Taken together, these findings suggest that many "HODL missionaries" rely on the performance of the fortunate upper quartile to overstate the typical rewards available to the average investor across a broad universe of cryptocurrencies.

[10]Bubble area is computed as $A = \min\{\max[\,400\,(1+1.5\,z),\,400],\,2800\}$, where $z = (\mathrm{IQR} - \mu_{\mathrm{IQR}})/\sigma_{\mathrm{IQR}}$ standardises the IQR within each basket.

6 Methodology

6.1 Initial configuration

The model takes one basket at a time (for example, BTC) for each run. The endogenous block consists of four targets: the median excess return, the 10% conditional value-at-risk (CVaR), the mean of the top quantile of returns (top 25%), and the Sharpe ratio. The macro-financial block includes a set of macro-financial variables. All variables were subjected to unit-root tests prior to model estimation and were pre-labeled according to their stationarity status for later use in the model. Posterior sampling relied on the No-U-Turn Sampler (NUTS) (Hoffman and Gelman, 2014) with four sequential chains. For all eight baskets, we fixed the sampler hyperparameters at `target_accept`=0.95 and `max_treedepth`=12.

6.2 Load basket and horizon panels

We load the merged dataset of sell-side statistics. For each basket, observations are grouped according to the week of selling. Specifically, if a basket is sold within the same calendar week—defined from Monday through Sunday—the selling date is shifted to the start of that week (Monday). This ensures a consistent weekly index across baskets. The endogenous metrics are constructed directly from realized risk–return measures. In other words, for each basket, holding horizon, and selling week, we compute the realized statistics as defined in the initial configuration.The macro-finance variables are aligned to the weekly time index of the endogenous features.

We load each variable's stationarity test results and assess unit-root status with DF-GLS, KPSS, and Zivot–Andrews under two specifications: constant only (c) and constant plus linear trend (ct) at the 5% significance level (Elliott et al., 1996; Kwiatkowski et al., 1992; Zivot and Andrews, 1992). Within each specification, a series is labeled stationary if both DF-GLS and Zivot–Andrews reject the unit-root null ($p < 0.05$) while KPSS fails to reject stationarity ($p > 0.05$). It is labeled unit root if the reverse pattern holds. When the three tests point in different directions, the outcome is labeled ambiguous.

We classify the unit-root test outcomes into three categories: Level, Trend, and RW, using constant-only (c) versus constant-plus-trend (ct) specifications. If the c specification is stationary, we keep the series in its original form (Level). If only the ct specification is stationary, we remove a deterministic trend via a strictly causal rolling linear detrend and use the residuals (Trend). If the outcome indicates a unit root or is ambiguous, we treat the series as a random walk (RW).

For macro variables, there is no horizon dimension, so the label is taken directly from the unit-root test outcome. For endogenous variables, a single series can exhibit different unit-root outcomes across prediction horizons. To ensure consistency, we aggregate per-horizon tags using a majority rule to obtain a single global tag for each variable. This means the category with the highest count becomes the global tag, and we apply the same treatment across all horizons. In the case of ties, the outcome is resolved conservatively with the ordering RW $\succ$ Trend $\succ$ Level. This design ensures horizon-consistent preprocessing for each variable and improves comparability of estimates and impulse responses across horizons.

Fractional differencing for RW **macro inputs.** For macro–finance series labeled RW, we apply strictly causal fractional differencing (Granger and Joyeux, 1980; Hosking, 1981) with fixed order $d = 0.5$ and a truncation at $K = 200$ lags:

$$y_t^{\mathrm{MF}} = \sum_{k=0}^{K} w_k \, x_{t-k}^{\mathrm{MF}}, \qquad w_k = (-1)^k \binom{d}{k}, \qquad \binom{d}{k} = \frac{d(d-1)\cdots(d-k+1)}{k!}, \quad d = 0.5,\ K = 200.$$

Here, x_t^{MF} is the macro–finance series and y_t^{MF} is its fractionally differenced value at time t; w_k weights the k-period lag x_{t-k}^{MF}. The implementation is strictly causal with a finite truncation of the infinite sum. To avoid numerical instability, we implement the weights through the stable recurrence $w_0 = 1$ and $w_k = w_{k-1}\frac{-(d-(k-1))}{k}$ $(k \geq 1)$[11], which is equivalent to the binomial definition.

For endogenous variables labeled as RW, we transform the series into first differences:

$$y_t^{(RW)} \;=\; y_t - y_{t-1}, \qquad t = 2, 3, \ldots, T.$$

This ensures that variables classified as random walks are rendered stationary (I(0)) before entering the model.

The same detrending is applied to endogenous targets and macro finance series when labeled TREND. Let $L = 52$ weeks be the rolling window length. At each time t $(\geq L)$, define the past-only window $\mathcal{W}_t = \{t - L + 1, \ldots, t\}$ and index observations by $j = 0, 1, \ldots, L - 1$ so that the last point is $j = L - 1$. Within $\mathcal{W}_t$, fit an OLS line

$$x_{t-L+1+j} \;=\; \alpha_t + \beta_t\, j + \varepsilon_{t,j}, \qquad j = 0, 1, \ldots, L - 1,$$

using only data in $\mathcal{W}_t$ (strictly causal). Let $\hat{\alpha}_t, \hat{\beta}_t$ be the OLS coefficients. We take the residual at the end of the window as the detrended value:

$$x_t^{(\textsc{Trend})} \;=\; x_t - \big(\hat{\alpha}_t + \hat{\beta}_t\,(L - 1)\big), \qquad t = L, L + 1, \ldots$$

If the available history is shorter than L for the entire series, we conservatively fall back to the first difference,

$$x_t^{(\textsc{Trend})} \;=\; x_t - x_{t-1}.$$

After transforming each macroeconomic series into a strictly causal $I(0)$ series by fractionally differencing the unit-root tags and rolling linear detrending the trend-stationary tags, we construct *EMA* and *VOL* features at multi-week horizons $w \in \{4, 8, 12, 24\}$. This serves two purposes: macro factors typically have lagged effects, and including raw weekly lags would greatly expand the feature amount and lead to basket-specific lag heterogeneity under bootstrap resampling. By contrast, EMA and VOL compress lag structure, retaining long-memory while keeping the predictor set parsimonious.

For each transformed series y_t^{MF}, the exponentially weighted moving average is

$$\mathrm{EMA}_t^{(w)} \;=\; \alpha_w y_t^{\mathrm{MF}} + (1 - \alpha_w)\,\mathrm{EMA}_{t-1}^{(w)}, \qquad \alpha_w = \frac{2}{w + 1}.$$

The recursion starts only once w past observations are available.

We compute realized volatility from the trailing w-week window as

$$\bar{y}_{t,w}^{\mathrm{MF}} = \frac{1}{w}\sum_{j=0}^{w-1} y_{t-j}^{\mathrm{MF}}, \qquad \mathrm{VOL}_t^{(w)} \;=\; \sqrt{\frac{1}{w-1}\sum_{j=0}^{w-1}\big(y_{t-j}^{\mathrm{MF}} - \bar{y}_{t,w}^{\mathrm{MF}}\big)^2}.$$

Together, $\mathrm{EMA}_t^{(w)}$ captures lagged movements and $\mathrm{VOL}_t^{(w)}$ captures rolling-window volatility.

[11] We truncate the recurrence earlier: once a weight's absolute value falls below 10^{-10}.

Tensor alignment. To build horizon-consistent panels, all horizons are harmonized to the same macro-finance feature set and column order. For each horizon h, the feature matrix is reindexed to a common column set $\mathcal{C}$. Next, we restrict the weekly index to the intersection of dates available across all horizons (after transforms such as differencing, fractional differencing, and rolling detrend/EMA/VOL), so that every observation uses the same feature set and the same time index[12].

6.3 Causal standardisation

After the stationarity transforms in §6.2, each horizon h uses an aligned matrix X_h with T_h time points and one column per feature or target. Denote the value in column k at time t by $x_{t,h}^{(k)}$. We apply an expanding-window z-score transformation to avoid look-ahead bias:

$$\mu_{t,h}^{(k)} = \frac{1}{t}\sum_{i=1}^{t} x_{i,h}^{(k)}, \qquad \sigma_{t,h}^{(k)} = \sqrt{\frac{1}{t}\sum_{i=1}^{t}(x_{i,h}^{(k)} - \mu_{t,h}^{(k)})^2}.$$

To ensure numerical stability, the denominator is clipped at $\varepsilon = 10^{-2}$:

$$s_{t,h}^{(k)} = \max\{\sigma_{t,h}^{(k)}, \varepsilon\}, \qquad z_{t,h}^{(k)} = \frac{x_{t,h}^{(k)} - \mu_{t,h}^{(k)}}{s_{t,h}^{(k)}}, \quad t = 1, \ldots, T_h.$$

6.4 Stability selection

We adapt the stability selection framework of Meinshausen and Bühlmann (2010) to our grouped macro-finance predictor structure.

Purged time-series split Our horizons at time t are the *realised* risk–return distribution. Thus, the effective horizon must be fully realized by time t; otherwise trailing windows overlap across folds (for example, a 30-day horizon only becomes fully observed at t+30). To break this overlap we leave a per-fold gap following de Prado (2018). Let τ_k be the first index of the k-th validation block; for horizon h we truncate the training indices to

$$\mathcal{I}_k^{\text{train}}(h) \;=\; \{\, t : t \le \tau_k - g(h) - 1 \,\}, \qquad \mathcal{I}_k^{\text{test}} \;=\; \{\, t : \tau_k \le t \le \tau_k^{\max} \,\},$$

with gap

$$g(h) \;=\; \left\lceil \frac{h}{7} \right\rceil + 1, \qquad h \in \mathcal{H},$$

where $\mathcal{H}$ is the set of horizons in days. Each $g(h)$ converts the daily horizon to weeks through the ceiling and adds one extra week as a buffer. This prevents overlapping trailing windows between train and validation.[13]

[12] On the common time grid $\mathcal{T}$ we assemble $Y \in \mathbb{R}^{T\times H\times n_y}$ and $X \in \mathbb{R}^{T\times H\times n_x}$, aligned along time $T = |\mathcal{T}|$, horizons H, and variables (n_y, n_x).

[13] Folds are generated with scikit-learn's (Pedregosa et al., 2011) `TimeSeriesSplit`.

Hyperparameter selection For each horizon $h \in \mathcal{H}$ we use the purged folds $\mathcal{I}_f^{\text{train}}(h), \mathcal{I}_f^{\text{test}}(h)$ defined above.[14] Let $\mathcal{F}_h$ denote the folds that remain after the purge gap.

Let X_h denote the z-scored predictors at horizon h (with p screened variables), and let Y_h be the corresponding standardized target matrix (with m target series). The multitask Elastic Net with $\rho = 1/2$ (Zou and Hastie, 2005) solves

$$\widehat{B}_h(\alpha) \;\in\; \arg\min_{B \in \mathbb{R}^{p \times m}} \left\{ \frac{1}{2T} \left\| Y_h - X_h B \right\|_F^2 \;+\; \alpha\Big(\tfrac{1}{4} \left\| B \right\|_F^2 + \tfrac{1}{2} \sum_{j=1}^{p} \left\| B_{j\cdot} \right\|_2 \Big) \right\},$$

so no intercept is needed. For each horizon with $|\mathcal{F}_h| \geq 2$ we pick

$$\widehat{\alpha}(h) \;\in\; \arg\min_{\alpha \in \mathcal{A}} \; \frac{1}{|\mathcal{F}_h|} \sum_{f \in \mathcal{F}_h} \frac{1}{|I_f^{\text{test}}(h)|} \left\| Y_{h, I_f^{\text{test}}(h)} - X_{h, I_f^{\text{test}}(h)} \, \widehat{B}_h^{(f)}(\alpha) \right\|_F^2,$$

where each candidate α is run through purged cross-validation: every fold re-fits the Elastic Net on its training slice to obtain $\widehat{B}_h^{(f)}(\alpha)$, applies those coefficients to the unseen observations, and records the mean-squared validation loss for that setting. The α that attains the smallest validation loss becomes $\widehat{\alpha}(h)$. [15] Horizons without two valid folds inherit a shared shrinkage level

$$\alpha_{\text{shared}} = \begin{cases} \text{median}\left\{ \widehat{\alpha}(h) : |\mathcal{F}_h| \geq 2 \right\}, & \text{if at least one horizon has two valid folds,} \\ 0.1, & \text{otherwise,} \end{cases}$$

Bootstrap-based feature selection To preserve short-run temporal dependence in the bootstrap resampling, we use a non-overlapping block bootstrap (NBB) for feature selection (Lahiri, 2003). For each horizon h, each bootstrap sample is constructed from length-b contiguous time blocks drawn with replacement from the aligned in-sample tensor described in §6.2. Block start positions are spaced by b (non-overlapping) and sorted from earliest to latest.[16]

The block size follows the $n^{1/3}$ rule (Hall et al., 1995) and is clipped to a weekly range:

$$b \;=\; \min\{20,\; \max\{4,\; \text{round}(n^{1/3})\}\}.$$

Inside each block the observations remain consecutive, so every resampled block respects the original time order.

Let R denote the number of bootstrap draws.[17] For each bootstrap draw $r = 1, \ldots, R$ we sample an index sequence I_r of length T (as defined in §6.2) by concatenating the length-b blocks above, and refit the multitask Elastic Net on (X_{h,I_r}, Y_{h,I_r}) with the fixed hyperparameter $\alpha_h^\star$ selected in the prior hyperparameter-selection stage, keeping the Elastic Net ℓ_1 ratio at $\rho = 1/2$ as before:

$$\widehat{B}_h^{(r)} \in \arg\min_{B \in \mathbb{R}^{p \times m}} \Big\{ \frac{1}{2|I_r|} \left\| Y_{h,I_r} - X_{h,I_r} B \right\|_F^2 + \alpha_h^\star \big(\tfrac{1}{4} \|B\|_F^2 + \tfrac{1}{2} \sum_{j=1}^{p} \|B_{j\cdot}\|_2 \big) \Big\}.$$

[14] We attempt $K = 3$ purged folds for all horizons; if any horizon would yield fewer than two valid splits we downgrade to $K = 2$. Horizons that still fail the two-fold requirement skip CV and fall back to the shared shrinkage level.

[15] $\mathcal{A}$ denotes the candidate α sequence, which follows scikit-learn's default log-spaced path used by `MultiTaskElasticNetCV`.

[16] More specifically, candidate block starts $\{0, b, 2b, \ldots, n-b\}$ are drawn i.i.d. with replacement, then sorted and concatenated as length-b intervals until the index reaches n (truncating any excess).

[17] We set $R = 1000$ in our bootstrap model.

We treat the untransformed macro–finance variables as the base variables, indexed by the set $\mathcal{V}$. For each base variable $v \in \mathcal{V}$, let $\mathcal{G}_v$ collect all EMA and VOL transforms constructed from v. In the bootstrap, for each draw r we declare transform j selected whenever $\|\widehat{B}^{(r)}_{h,j\cdot}\|_2 > 0$. These selections allow us to compute (i) the base–level stability and (ii) the within–base conditional stability for each transform $\gamma \in \mathcal{G}_v$.

$$\hat{\pi}_{\text{base}}(v) = \frac{1}{R_{\text{valid}}} \sum_{r=1}^{R} \mathbf{1}\Big\{\exists\, \gamma \in \mathcal{G}_v :\ \|(\widehat{B}^{(r)}_h)_{[v,\gamma],\cdot}\|_2 > 0\Big\}, \qquad \hat{\pi}_{\text{cond}}(v,\gamma) = \frac{\sum_{r=1}^{R} \mathbf{1}\Big\{\|(\widehat{B}^{(r)}_h)_{[v,\gamma],\cdot}\|_2 > 0\Big\}}{R_{\text{valid}}\, \hat{\pi}_{\text{base}}(v)},$$

If a base variable is never selected, we set its conditional stability to zero. For each horizon h, we retain a base variable v if its base-level stability satisfies $\hat{\pi}_{\text{base}}(v) \geq \tau_g$. For every base that passes this threshold, we then consider its transforms and keep a transform γ only if its conditional stability meets $\hat{\pi}_{\text{cond}}(v,\gamma) \geq \tau_c$.[18] Within each of the EMA and VOL types, we retain the top-1 transform ranked by $\hat{\pi}_{\text{cond}}(v,\gamma)$. The selected feature set at horizon h is denoted $\mathcal{S}_h$; the overall selected feature set is their union $\mathcal{S} = \bigcup_{h\in\mathcal{H}} \mathcal{S}_h$, which preserves horizon-specific signals while avoiding multi-collinearity within transform families.

6.5 Pre-Bayesian Preparation

After the causal standardisation (§6.3) and the stability-selection filter (§6.4), each horizon $h \in \mathcal{H}$ retains the feature set $\mathcal{S}_h$ and their union $\mathcal{S} = \bigcup_h \mathcal{S}_h$. On the common weekly grid $\mathcal{T} = \{1, \ldots, T\}$, we work with the standardized tensors

$$\{(\widetilde{Y}^{(C)}_{t,h},\, \widetilde{X}^{\mathcal{S}}_{t,h}) : t \in \mathcal{T},\ h \in \mathcal{H}\}, \qquad \widetilde{Y}^{(C)} \in \mathbb{R}^{T\times H\times n_y},\ \widetilde{X}^{\mathcal{S}} \in \mathbb{R}^{T\times H\times |\mathcal{S}|},$$

where $\widetilde{Y}^{(C)}_{t,h}$ and $\widetilde{X}^{\mathcal{S}}_{t,h}$ denote the post–expanding-z-score targets and screened features observed at time t.

Training window. Before fitting the Bayesian multi-horizon local projection, we reuse the same gap as the purged time-series split, to ensure every observation has a fully realised future outcome. For each horizon h, define the effective gap in weeks

$$g(h) \ = \ \Big\lceil \frac{h}{7} \Big\rceil + 1, \tag{3}$$

let $g_{\max} = \max_{h\in\mathcal{H}} g(h)$. The estimation sample ends at

$$t^\star \ = \ T - g_{\max},$$

so the training index becomes $\mathcal{T}_{\text{train}} = \{1, \ldots, t^\star\}$.

Future-target tensor. For $t \in \mathcal{T}_{\text{train}}$ and $h \in \mathcal{H}$, the dependent variable corresponds to the realized outcome at week $t + g(h)$. We therefore construct

$$\widetilde{Y}^{(F)}_{t,h} \ \equiv \ \widetilde{Y}^{(C)}_{t+g(h),h}, \qquad \widetilde{Y}^{(F)} \in \mathbb{R}^{t^\star\times H\times n_y},$$

which contains only fully realized future outcomes.

[18] In our implementation we use $\tau_g = 0.55$ and $\tau_c = 0.50$.

Reference scaling. To enable conversion of posterior draws from standardized units back to original scales, we record the expanding standard deviations at the end of the training period $t^\star$. These reference scales—denoted $\sigma_y(h)$ for each target and $\sigma_x^{\mathcal{S}}(h,j)$ for each selected feature—provide a consistent basis for converting z-scored coefficients to native units (see Appendix A.3 for details).

6.6 Bayesian multi-horizon local projection

Building on Ferreira et al. (2025) and related work on smooth Bayesian local projections (e.g., Tanaka, 2020; Huber et al., 2024), we adopt a Bayesian local projection framework for multi-horizon impulse response estimation. We model the standardized tensors from §6.5 using the notation established earlier. We define the horizon-specific design vector $Z_{t,h}$ as

$$Z_{t,h} \;=\; \begin{bmatrix} \widetilde{Y}_{t,h}^{(C)} \\ \widetilde{X}_{t,h}^{(\mathcal{S})} \end{bmatrix} \in \mathbb{R}^P, \qquad P \equiv P_y + P_x, \quad P_y = n_y, \quad P_x = |\mathcal{S}|.$$

By construction, $Z_{t,h}$ only uses information observed at week t, while the aligned future outcome satisfies

$$\widetilde{Y}_{t,h}^{(F)} \;=\; \widetilde{Y}_{t+g(h),h}^{(C)},$$

as defined in §6.5.

Likelihood. For each horizon h we use a Student-t likelihood (Geweke, 1993) with horizon-specific linear predictor:

$$\widetilde{Y}_{t,h}^{(F)} \;\sim\; \text{Student-}t_{\nu=6}\big(\mu_{t,h},\, \Sigma_h\big), \tag{4}$$

$$\mu_{t,h} \;=\; \alpha_h + Z_{t,h}\,\beta_h, \qquad \alpha_h \in \mathbb{R}^{n_y},\;\; \beta_h \in \mathbb{R}^{P\times n_y}. \tag{5}$$

Residual covariance. For estimation stability, we assume independent Student-t innovations across targets at each horizon,

$$\Sigma_h = \operatorname{diag}(\sigma_{h,1}^2, \ldots, \sigma_{h,n_y}^2).$$

Instead of capturing residual correlation across targets, our hierarchical shrinkage priors—with shared global scales τ_y and τ_x across all targets—already induce cross-target information pooling at the coefficient level.

Smoothing across unevenly-spaced horizons. Our six forecast horizons (30, 90, 180, 365, 730, 1095 days) are unevenly spaced in time. If we smooth coefficients uniformly across horizon indices, the prior would treat each horizon distance equally. To impose constant smoothness based on the real time distance, we map each horizon to weeks $w_h \equiv \lceil h/7 \rceil$ and scale the RW1 innovations by the square root of the weekly spacing, so each coefficient path evolves as

$$\Delta_h \;=\; w_h - w_{h-1}, \qquad b_h \;=\; b_{h-1} \;+\; \epsilon_h\,\tau\,\sqrt{\Delta_h}, \quad \epsilon_h \sim \mathcal{N}(0,1). \tag{6}$$

This ensures $\operatorname{Var}(b_h - b_{h-1}) = \tau^2\,\Delta_h$, so that the prior variance scales proportionally with time gap between horizons.

Priors

Hierarchical intercept priors. For each target y, we assign a hierarchical prior on the intercepts across forecast horizons:

$$\begin{aligned} \alpha_{h,y} &= \mu_{\alpha,y} + \sigma_{\alpha,y}\,\tilde{\alpha}_{h,y}, \qquad \tilde{\alpha}_{h,y} \sim \mathcal{N}(0,1), \\ \mu_{\alpha,y} &\sim \mathcal{N}(0,1), \qquad \sigma_{\alpha,y} \sim \mathcal{HN}(0.5), \qquad y = 1,\dots,n_y. \end{aligned} \tag{7}$$

This non-centered parameterization allows intercepts to vary moderately across horizons while pooling information at the target level through shared mean $\mu_{\alpha,y}$ and scale $\sigma_{\alpha,y}$.

Coefficient blocks. The slope coefficients consist of an endogenous block and a macro-finance block:

$$\beta_h = \begin{bmatrix} \beta_h^{(y)} \\ \beta_h^{(x)} \end{bmatrix}, \quad \beta_h^{(y)} \in \mathbb{R}^{P_y \times n_y},\ \beta_h^{(x)} \in \mathbb{R}^{P_x \times n_y}.$$

For both coefficient blocks we retain the RW1 design from (6): for $h \geq 2$ each step is multiplied by $\sqrt{\Delta_h}$. The endogenous coefficients use $p = 1,\dots,P_y$ while the macro-finance block uses $p = 1,\dots,P_x$. The hierarchical shrinkage factors $s_{p,y}^{(y)}$ and $s_{p,y}^{(x)}$ control the innovation variance for each coefficient path.

We adopt a global–local shrinkage hierarchy (e.g., Carvalho et al., 2010; Piironen and Vehtari, 2017b):

$$s_{p,y}^{(y)} = \tau_y\,\lambda_{p,y}^{(y)}, \qquad s_{p,y}^{(x)} = \tau_x\,\lambda_p^{(g)}\,\lambda_{p,y}^{(\ell)},$$

where τ_y, τ_x are global scales and $\lambda_{p,y}^{(y)}$, $\lambda_p^{(g)}$, $\lambda_{p,y}^{(\ell)}$ are local (and group-level) multipliers. The key difference between the endogenous and macro-finance shrinkage factors is that macro-finance predictors employ a three-level hierarchy: the group factor $\lambda_p^{(g)}$ is shared across all targets, so when $\lambda_p^{(g)} \approx 0$, feature p is jointly shrunk to zero across all target variables, reflecting that this macro-finance predictor provides limited explanatory power for the outcomes. The local factor $\lambda_{p,y}^{(\ell)}$ then controls the strength of feature p's effect on each individual target variable. This design reflects that macro shocks jointly affect risk and return outcomes, while allowing target-specific heterogeneity in both magnitude and direction. In contrast, endogenous coefficients use a two-level hierarchy without cross-target group sharing.

The hyperpriors are:

$$\tau_y \sim \mathcal{HN}(\tau_{0,y}), \quad \tau_x \sim \mathcal{HN}(\tau_{0,x}), \qquad \lambda_{p,y}^{(y)},\ \lambda_p^{(g)},\ \lambda_{p,y}^{(\ell)} \overset{\text{i.i.d.}}{\sim} \mathcal{H}t_{\nu=4}(1),$$

where

$$\begin{aligned} \tau_{0,y} &= \frac{p_{0,y}}{\max(P_y - p_{0,y}, 1)} \cdot T_{\text{train}}^{-1/2}, \quad p_{0,y} \equiv \min\{\max(0.25\,P_y, 1),\ 5\}, \\ \tau_{0,x} &= \frac{p_{0,x}}{\max(P_x - p_{0,x}, 1)} \cdot T_{\text{train}}^{-1/2}, \quad p_{0,x} \equiv \min\{\max(0.10\,P_x, 1),\ 5\}. \end{aligned}$$

Here, $\mathcal{HN}(\sigma)$ and $\mathcal{H}t_\nu(\sigma)$ denote half-normal and half-Student-t distributions with mode at zero. We use $\nu = 4$ for moderately heavy tails, allowing important features to escape shrinkage while pushing most coefficients toward zero.

(i) Endogenous block $\beta_h^{(y)}$: global–local shrinkage RW1. For $p = 1, \ldots, P_y$ and $y = 1, \ldots, n_y$,

$$\beta_{1,p,y}^{(y)} \sim \mathcal{N}(0,\, 0.3^2), \tag{8}$$

$$\beta_{h,p,y}^{(y)} = \beta_{h-1,p,y}^{(y)} \;+\; s_{p,y}^{(y)} \sqrt{\Delta_h}\, \varepsilon_{h-1,p,y}^{(y)}, \qquad \varepsilon_{h-1,p,y}^{(y)} \sim \mathcal{N}(0,1), \quad h \geq 2, \tag{9}$$

with $s_{p,y}^{(y)} = \tau_y \, \lambda_{p,y}^{(y)}$ and hyperpriors defined above.

(ii) Macro-finance block $\beta_h^{(x)}$: group–local shrinkage RW1. For $p = 1, \ldots, P_x$ and $y = 1, \ldots, n_y$ we retain the same RW1 recursion, but with a tighter initial variance:

$$\beta_{1,p,y}^{(x)} \sim \mathcal{N}(0,\, 0.2^2), \tag{10}$$

$$\beta_{h,p,y}^{(x)} = \beta_{h-1,p,y}^{(x)} + s_{p,y}^{(x)} \sqrt{\Delta_h}\, \varepsilon_{h-1,p,y}^{(x)}, \qquad \varepsilon_{h-1,p,y}^{(x)} \sim \mathcal{N}(0,1), \quad h \geq 2, \tag{11}$$

where the innovation scale is the three-level factor $s_{p,y}^{(x)} = \tau_x \, \lambda_p^{(g)} \, \lambda_{p,y}^{(\ell)}$ introduced above.

7 Results

7.1 Convergence

Table 2 reports convergence diagnostics for all eight baskets. All baskets achieve convergence ($\hat{R} = 1.00$, ESS > 900, divergences $< 0.05\%$), and no sampler runs into the tree-depth limit of 12. Detailed trace plots and energy diagnostics are available in Appendix B.

Table 2: MCMC convergence diagnostics by basket.

Basket	Max $\hat{R}$	Min ESS (bulk)	Min ESS (tail)	Max MCSE/SD	Min BFMI	Divergence rate	Treedepth saturation
ADA	1.00	1559	1462	0.118	0.76	0.0000	0.00
ALL	1.00	907	1450	0.069	0.71	0.0005	0.00
BNB	1.00	2441	2244	0.111	0.85	0.0000	0.00
BTC	1.00	1637	1484	0.041	0.68	0.0000	0.00
DOGE	1.00	1614	1570	0.143	0.87	0.0000	0.00
ETH	1.00	2501	2601	0.043	0.80	0.0000	0.00
LINK	1.00	1388	1278	0.182	0.81	0.0001	0.00
XRP	1.00	2243	2279	0.060	0.89	0.0001	0.00

7.2 Cross-Basket Stability of Predictors

Figure 4 summarizes the cross-basket stability of top predictors for each target-horizon combination. For each target and horizon, we display the single most stable predictor—defined as the predictor significant in the most baskets.[19] Each bar indicates the number of baskets (out of eight in total) in which the predictor achieves 95% credible-interval significance for the corresponding target and horizon. Bar colors denote the sign of the mean median effect[20] across baskets where the predictor is 95% significant for that target–horizon pair: blue for positive effects and red for negative effects. Annotations display the predictor label and the mean median effect across baskets, measured in the target's native units per one-standard-deviation shock in the predictor.

[19] In the case of ties, we sort by mean absolute median effect, then by maximum absolute median effect. See Appendix A.3.1 for detailed ranking procedures.

[20] "Median effect" denotes the posterior median impulse response produced by the BLP model.

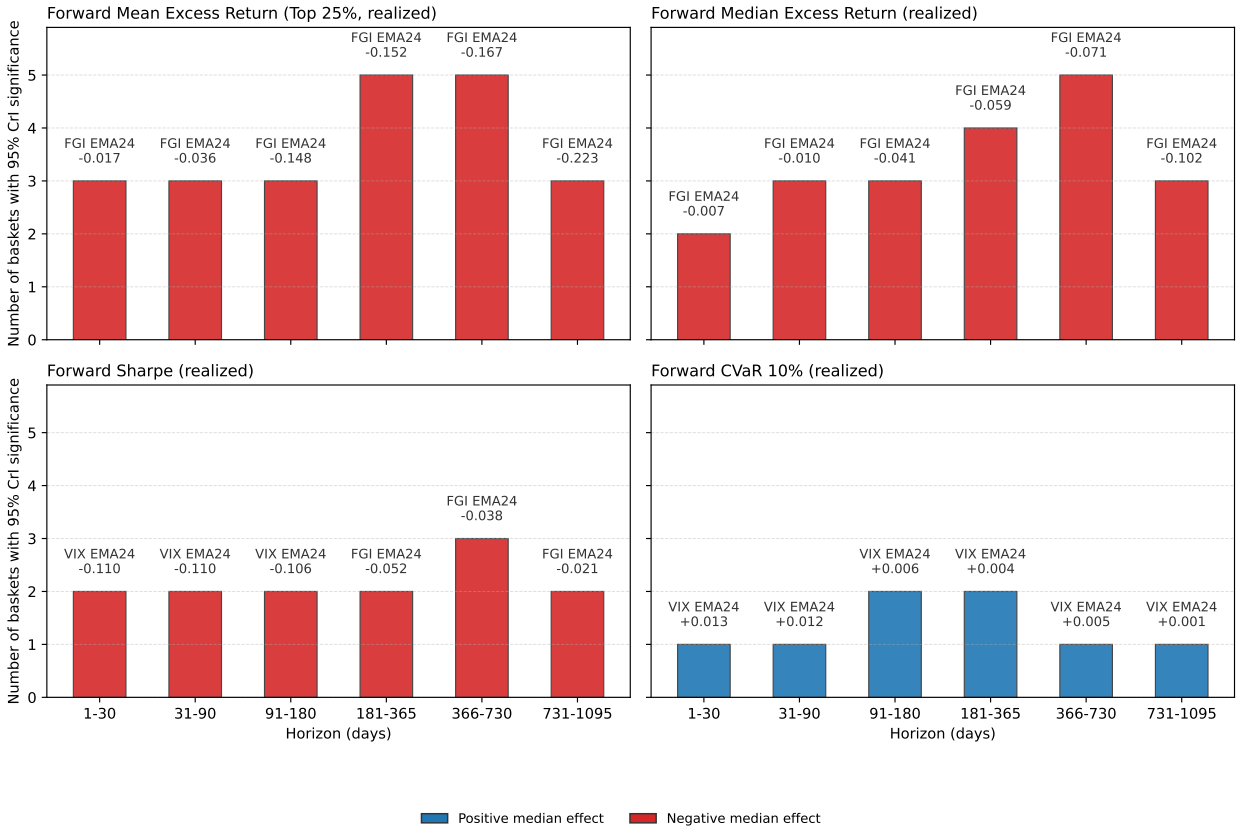


Figure 4: Cross-basket stability of top predictors by target and horizon

Figure 4 reveals several key findings. First, the predictors with the highest cross-basket significance are exclusively macrofinance factors. Second, no single predictor exhibits universal significance: even FGI EMA24, which is the most stable predictor for the forward top-25% and median excess-return targets, does not achieve 95% credible-interval significance across all eight baskets (predictor abbreviations follow Appendix A.3.2).

A one-standard-deviation increase in FGI EMA24 lowers the forward mean excess return of the top 25% quantile by roughly 15–22 percentage points across the 1–3 year horizons we examine. For the forward median excess return at 1–3 year horizons, a one-standard-deviation shock in FGI EMA24 is associated with reductions of 6–10 percentage points.

For the forward Sharpe ratio, VIX EMA24 delivers strong negative effects around -0.1 in two baskets across the 1–180 day horizons, while FGI EMA24 becomes the most stable predictor once horizons extend to 181–365 days and beyond. Tail-risk $\text{CVaR}_{10\%}$ shows little evidence of a systematic response: VIX EMA24 and T10Y2Y RVol24 each achieve significance in at most two baskets per horizon, with economically negligible effect sizes.

Figure 5 lists the eight largest median effects at each horizon, ordered by absolute median impact. Unlike Figure 4, which fixes the target to study cross-basket stability, this panel surfaces the predictor–target combinations that deliver the strongest per-sigma effects regardless of target choice. Each horizontal bar represents the median posterior effect in the target's native units per one-standard-deviation shock to the predictor, with colors indicating the target variable (see the legend); gold borders denote endogenous predictors while dark borders denote macrofinance factors.

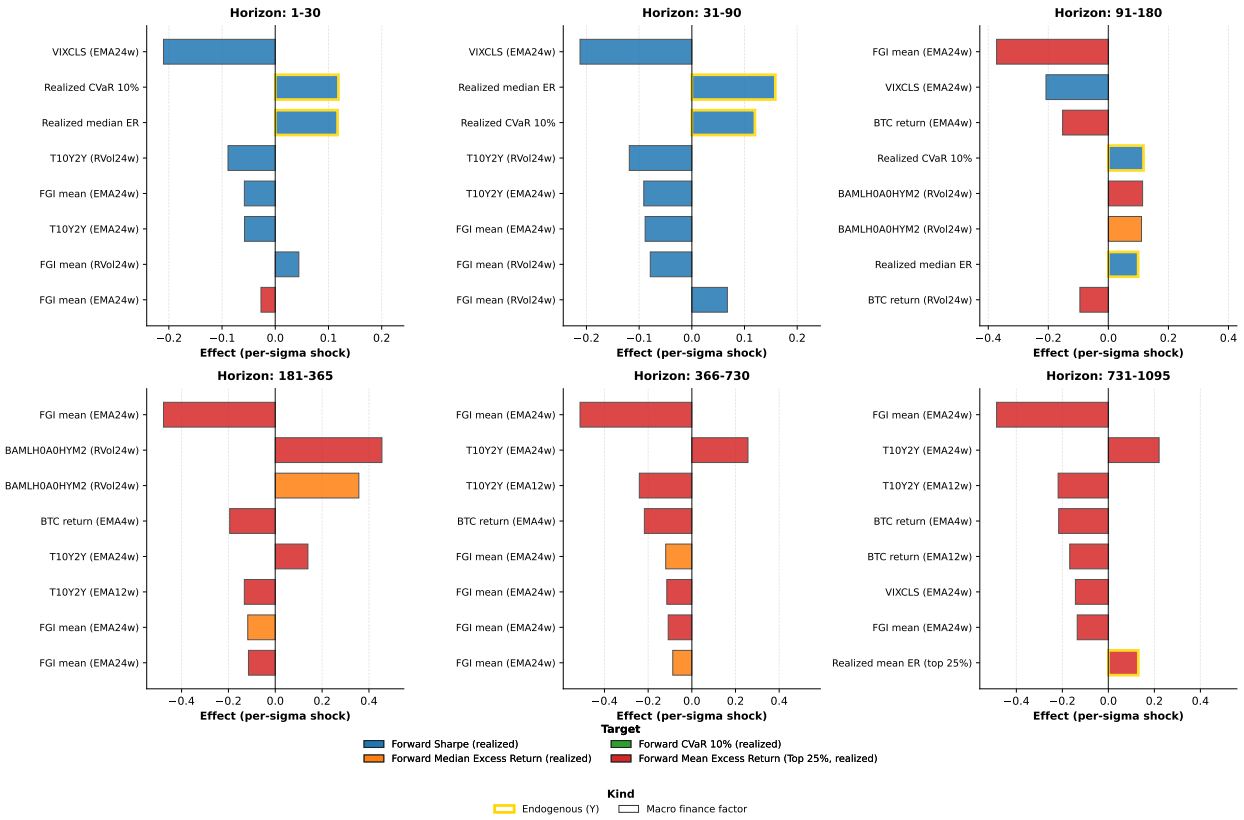


Figure 5: Largest predictor effects by horizon

Each displayed effect is 95% credible-interval significant.

Endogenous predictors, realized risk and return metrics, exhibit large effects primarily at shorter forecast horizons. Among the 24 top-8 entries across horizons spanning 1–180 days, 6 are endogenous predictors. Realized median excess return and realized $\text{CVaR}_{10\%}$ show positive effects on forward Sharpe ratio, while Realized mean ER (top 25%) exhibits negative effects. After 180 days, however, endogenous predictors nearly vanish from the top-8 list—only a single entry (Realized mean ER (top 25%) at the 731–1095 day horizon) remains. At short-to-medium horizons (1–90 days), VIX EMA24w exhibits the largest negative effects on forward Sharpe ratio, while FGI EMA24w also shows negative effects but with smaller magnitudes. Starting at 180 days, FGI EMA24w's largest effects shift to forward mean excess return for the top 25% quantile, where it repeatedly emerges as the single most influential predictor across all three long horizons, aligning with Figure 4.

T10Y2Y exhibits transformation-dependent relationships with forward mean excess return for the top 25% quantile. T10Y2Y EMA24w shows persistent positive effects, while T10Y2Y EMA12w exhibits negative effects. This divergence suggests that different smoothing lengths of the 10-year minus 2-year Treasury spread capture distinct phases of yield-curve dynamics with contrasting implications for forward cryptocurrency returns.

7.3 Endogenous Decay and Cross-Basket Heterogeneity

The findings in the previous section reveal a clear pattern: the most stable predictors across baskets are exclusively macro-finance factors (Figure 4). Endogenous predictors, realized risk and return metrics, appear among the largest effects at horizons spanning 1–180 days (Figure 5), yet only one of them remains in the top-8 effects beyond 180 days. This raises a natural question: do current realized risk and return distributions retain any influence on forward risk and return outcomes at longer horizons, or does their effect decay entirely?

Unlike the macro-finance features, which must pass the stability filter to enter the BLP multi-horizon model, the four endogenous variables enter the model directly across all baskets and horizons. This setup gives us the opportunity to analyze cross-basket effects of the endogenous variables. We conduct a Bayesian random-effects meta-analysis to pool information across baskets and horizons, following the framework detailed in Appendix A.3.3. To ensure comparability with the previous section, we scale the posterior coefficients by the target's reference standard deviation, so that effects are measured in the target's native units per one-standard-deviation increase in the predictor.

Tables 9 and 10 report basket-level effects for predictor-target-horizon combinations exhibiting 95%-credible significance. Out of the four target variables, only forward Mean ER (Top 25%) and forward Sharpe exhibit significant impulse responses for at least one basket-predictor-horizon combination. Forward Median ER and forward CVaR 10% show no significant responses across all baskets and horizons.

For forward Sharpe, only two basket-predictor pairs, LINK's realized CVaR 10% and ETH's realized median ER, exhibit significant positive effects across horizons. Both show responses that decay substantially with horizon: a one-standard-deviation shock increases forward Sharpe by more than 0.11 in Sharpe units at short-to-medium horizons, but decays to approximately 0.01–0.04 in Sharpe units at the longest horizon.

Despite these substantial basket-specific effects, Figure 9 reveals that population-level effects μ_h remain economically negligible. The pooled estimates show realized CVaR 10% generating the strongest population-average response to forward Sharpe, around 0.0027–0.0033 in Sharpe units across horizons. This contrast with the basket-level effects underscores substantial cross-basket heterogeneity: while specific pairs exhibit economically meaningful impulse responses, these patterns are basket-specific rather than generalizable.

This pattern stands in contrast to the behavior of macro-finance predictors. While endogenous risk-return metrics exhibit decaying effects with limited cross-basket consistency at the population level, macro-finance factors, particularly market sentiment indicators, demonstrate more persistent effects across longer horizons. Figure 4 shows that FGI EMA24w emerges as the most stable predictor at all long-run horizons (181–1095 days) for both forward mean excess return (Top 25%) and forward median excess return targets. This raises a natural question: does this long-horizon influence manifest similarly across cryptocurrency baskets, or does it exhibit significant heterogeneity?

Figures 4 and 5 highlighted the consistent negative relationship between FGI EMA24w and both forward mean excess return for the top 25% quantile and forward median excess return at long horizons. Figure 6 decomposes these responses across six baskets[21] and three long-run forecast horizons (181–365, 366–730, and 731–1095 days). Each panel displays posterior median impulse responses to a one-standard-deviation shock in FGI EMA24w. Red markers indicate responses that pass a 95% credible interval excludes zero; gray markers denote non-significant responses. Numbers to the right report the posterior directional probability.

[21] XRP and ALL are excluded because stability selection never retains `FGI mean EMA24w`. In XRP the bootstrap selects `FGI mean RVol24w` instead, while in ALL only `FGI mean RVol12w` exceeds the conditional-stability threshold.

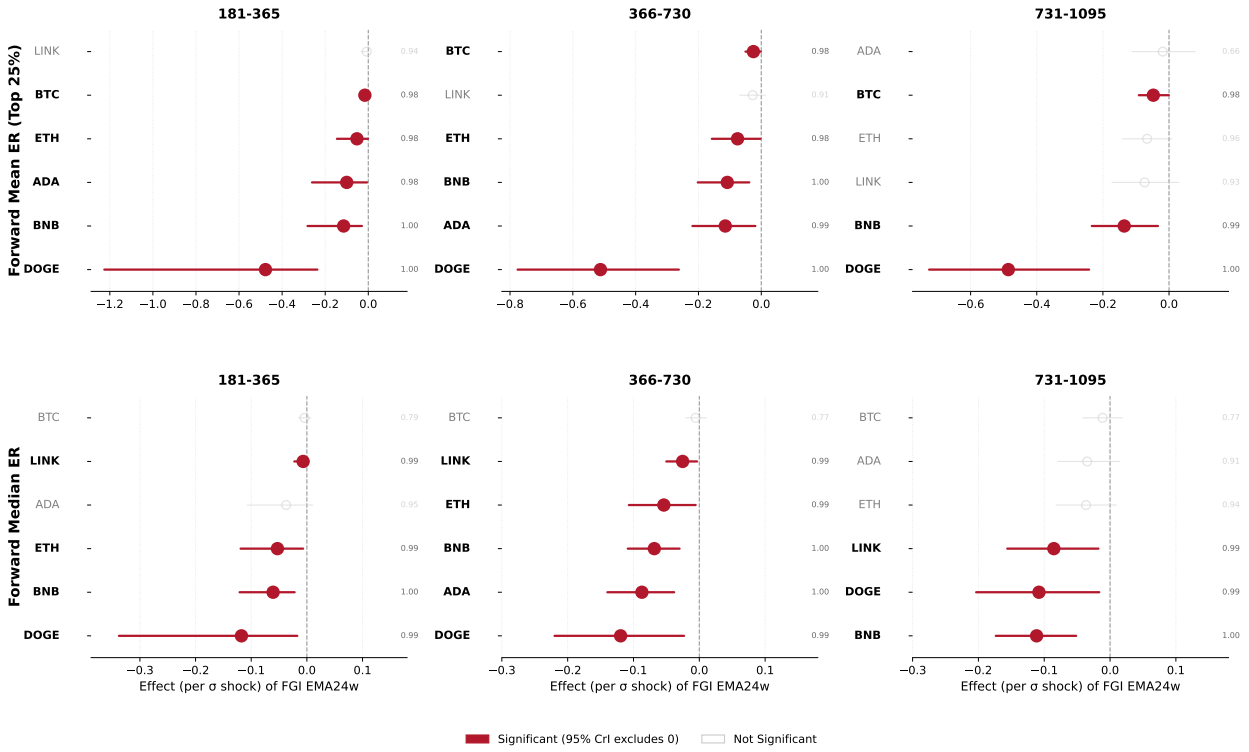


Figure 6: Cross-basket impulse responses to FGI sentiment shocks at long forecast horizons.

Figure 6 reveals substantial cross-basket heterogeneity in sentiment effects. For forward mean excess return (Top 25%), DOGE exhibits significantly stronger responses than other baskets: at the 365-day horizon, a one-standard-deviation increase in FGI EMA24w reduces forward mean excess return (Top 25%) by approximately 48 percentage points for DOGE, an effect more than 3 times larger than estimates for BNB (-11 percentage points), ADA (-10 percentage points), and other baskets. For forward median excess return, DOGE's response of approximately 12 percentage points at the same horizon is less drastic but still roughly double the effect size of other significant baskets. DOGE's amplified sensitivity persists across all three long horizons with posterior directional probabilities consistently exceeding 0.98, suggesting that meme coins like DOGE display heightened responsiveness to market sentiment across the entire forward return distribution, with particularly strong amplification in the upper tail.

Taken together, these findings answer the question posed at the outset: while endogenous risk-return metrics exhibit limited and decaying influence on forward outcomes at longer horizons, macro-finance sentiment indicators demonstrate more persistent and economically substantial effects. FGI EMA24w emerges as the most stable long-run predictor for forward return metrics across all tested macro-finance factors. However, the effects of sentiment shocks exhibit considerable cross-basket heterogeneity, with certain assets, particularly meme coins like DOGE, displaying disproportionate sensitivity to sentiment shifts.

8 Robustness Checks

8.1 Local Projection Model

To assess whether our main findings are robust to the choice of statistical framework, we implement a classical local projection estimator (Jordà, 2005) as an alternative to the Bayesian multi-horizon local projection model (§6.6). This provides a robustness check that tests whether the impulse response patterns remain qualitatively consistent under different estimation approaches.

We retain the identical data preparation from §6.5, including the same stability-selected features through moving block bootstrap (§6.4). The only change is replacing the Bayesian multi-horizon local projection model (§6.6) with a classical local projection approach. This ensures differences in results reflect the estimator choice rather than data preprocessing.

We use the same standardised objects as defined in §6.5:

$$\{(\widetilde{Y}^{(F)}_{t,h},\, \widetilde{Y}^{(C)}_{t,h},\, \widetilde{X}^{\mathcal{S}}_{t,h}) : t \in \mathcal{T}_{\text{train}},\ h \in \mathcal{H}\}.$$

For each horizon $h \in \mathcal{H}$ and target $k \in \{1, \dots, n_y\}$, we define the design vector

$$\widetilde{Z}_{t,h} \;=\; \begin{bmatrix} \widetilde{Y}^{(C)}_{t,h} \\ \widetilde{X}^{\mathcal{S}}_{t,h} \end{bmatrix} \in \mathbb{R}^P, \qquad P = n_y + |\mathcal{S}|,$$

and estimate the horizon-specific regression

$$\widetilde{Y}^{(F)}_{t,h,k} \;=\; \alpha_{h,k} + \widetilde{Z}^{\top}_{t,h}\beta_{h,k} + \varepsilon_{t,h,k}, \qquad t \in \mathcal{T}_{\text{train}}. \tag{12}$$

We estimate this through ordinary least squares separately for each (h, k) pair, computing heteroskedasticity-robust (HC1) standard errors (MacKinnon and White, 1985). This yields raw estimates $\hat{\beta}_{h,p,k}$ and $\hat{s}_{h,p,k}$ for each predictor $p \in \{1, \dots, P\}$.

Cross-horizon RW1 smoothing. As in the Bayesian model (§6.6), we smooth coefficient paths across horizons using a first-order random-walk structure, but here implemented as a quadratic penalty in GLS rather than as an RW1 prior. For a given predictor $p \in \{1, \dots, P\}$ and target k, define the horizon-specific coefficient path

$$\hat{\beta}_{p,k} = (\hat{\beta}_{h_1,p,k}, \dots, \hat{\beta}_{h_H,p,k})^{\top} \in \mathbb{R}^H,$$

and let $\hat{s}_{h_j,p,k}$ denote the corresponding HC1 standard errors. We map horizons h_j (in days) to weekly indices $\omega_j = \lceil h_j/7 \rceil$ with spacings $\Delta_j = \omega_{j+1} - \omega_j > 0$. The RW1-penalized coefficient path $\hat{\beta}^{\text{rw1}}_{p,k}$ solves the optimization problem

$$\hat{\beta}^{\text{rw1}}_{p,k} = \arg\min_{b \in \mathbb{R}^H} \sum_{j=1}^{H} \frac{(b_j - \hat{\beta}_{h_j,p,k})^2}{\hat{s}^2_{h_j,p,k}} + \lambda_{\text{rw1}} \sum_{j=1}^{H-1} \frac{(b_{j+1} - b_j)^2}{\Delta_j}, \tag{13}$$

with tuning parameter $\lambda_{\text{rw1}} = 1$. We apply this smoothing separately for each predictor–target pair (p, k). Intercepts $\alpha_{h,k}$ are not smoothed across horizons.

Stationary bootstrap and simultaneous bands. The RW1-penalized paths $\hat{\beta}^{\mathrm{rw1}}_{p,k}$ from (13) require standard errors that account for both serial dependence and the smoothing across horizons induced by the RW1 penalization. We construct confidence bands using a stationary bootstrap (Politis and Romano, 1994) combined with studentized k-max simultaneous bands (Romano and Wolf, 2007).

Let $T = |\mathcal{T}_{\mathrm{train}}|$. For each bootstrap replication $b = 1, \ldots, B$, we resample time indices $\{t_1^{(b)}, \ldots, t_T^{(b)}\} \subset \mathcal{T}_{\mathrm{train}}$ using the stationary bootstrap with mean block length

$$\bar{L} = \max\{2,\, \min(T-1,\, \max(1.75\, T^{1/3},\, L_{\mathrm{med}}))\},$$

where L_{med} is the median horizon length in weeks. For each bootstrap sample, we estimate the OLS regressions (12) and apply the RW1 penalization (13), obtaining $\hat{\beta}^{\mathrm{rw1},(b)}_{h_j,p,k}$ and $\hat{s}^{\mathrm{rw1},(b)}_{h_j,p,k}$.

We construct studentized t-statistics

$$t^{(b)}_{h_j,p,k} = \frac{\hat{\beta}^{\mathrm{rw1},(b)}_{h_j,p,k} - \hat{\beta}^{\mathrm{rw1}}_{h_j,p,k}}{\hat{s}^{\mathrm{rw1},(b)}_{h_j,p,k}}.$$

To construct simultaneous bands, for each predictor–target pair (p,k) we compute $M^{(b)}_{p,k}$, the $k_{\max}$-th largest value among $\{|t^{(b)}_{h_1,p,k}|, \ldots, |t^{(b)}_{h_H,p,k}|\}$ with $k_{\max} = \min\{2, H\}$, and let $\hat{c}_{p,k}$ be the 95% quantile of $\{M^{(b)}_{p,k}\}_{b=1}^{B}$. The resulting simultaneous 95% confidence band is

$$\hat{\beta}^{\mathrm{rw1}}_{h_j,p,k} \;\pm\; \hat{c}_{p,k}\, \hat{s}^{\mathrm{rw1}}_{h_j,p,k}, \qquad j = 1, \ldots, H, \tag{14}$$

which controls the family-wise error rate across horizons at 5% for each predictor–target pair (p,k). To match the Bayesian results, we scale by the target-specific reference scale $\sigma_{h_j,k}$ (defined in §6.5) to obtain impulse responses $\widehat{\mathrm{IRF}}^{\mathrm{LP}}_{h_j,p,k} = \hat{\beta}^{\mathrm{rw1}}_{h_j,p,k}\, \sigma_{h_j,k}$ with confidence bands $[\hat{\beta}^{\mathrm{rw1}}_{h_j,p,k} \pm \hat{c}_{p,k}\, \hat{s}^{\mathrm{rw1}}_{h_j,p,k}]\, \sigma_{h_j,k}$.

8.2 Local Projection Robustness Results

Table 3 compares the Bayesian multi-horizon local projection (BLP) and classical local projection (LP) estimates across all $N = 2{,}256$ predictor-target-horizon combinations spanning eight baskets and six horizons. We assess consistency using three metrics: sign match (whether the two models agree on direction), 95% interval overlap (whether credible and confidence intervals intersect), and significance rates (proportion of effects significant at 95% level).

Table 3: BLP vs LP Robustness Summary

Metric	Value
Total Samples	2,256
Sign Match Rate (%)	64.3
95% Interval Overlap (%)	98.7
BLP Significance (%)	9.1
LP Significance (%)	6.2
Sign Match | Both Sig (%)	100.0
Sign Match | BLP Sig (%)	79.1

The results demonstrate strong consistency between the two approaches. The 95% intervals overlap in 98.7% of cases, indicating quantitatively similar uncertainty estimates across both frameworks. The BLP model exhibits a higher significance rate (9.1%) compared to the LP model (6.2%). We attribute this difference primarily to the Bayesian framework's hierarchical shrinkage structure. In contrast, the LP approach estimates each horizon-target pair separately through OLS.

When examining directional consistency, the two models show strong agreement: when BLP signals significance, 79.1% share the same sign with the LP estimate, rising to 100% when both methods flag an effect as significant.

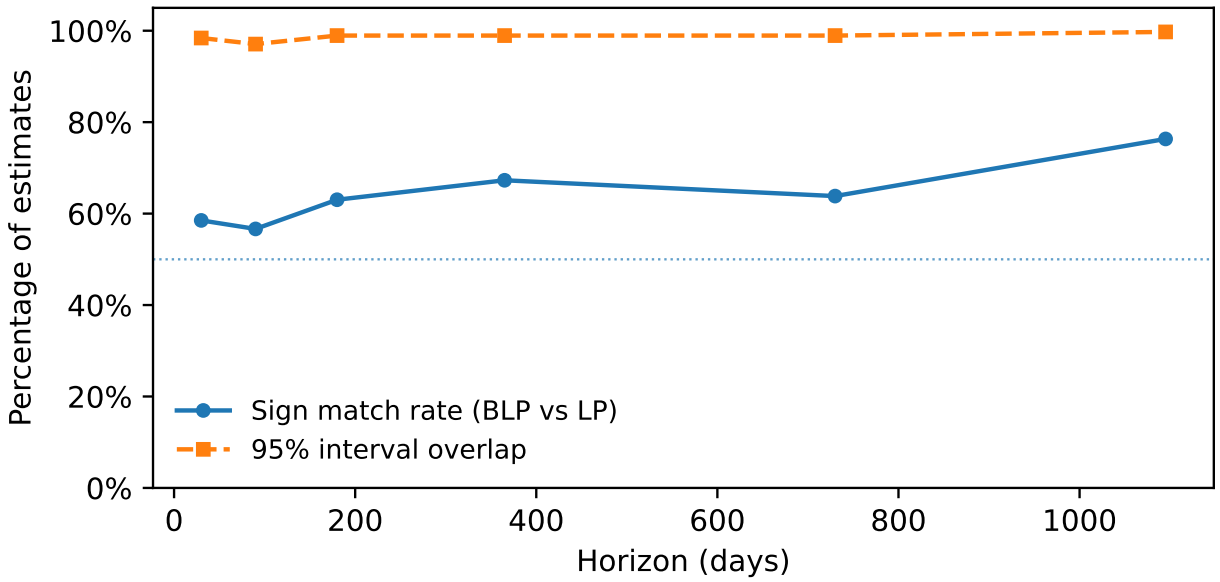


Figure 7: BLP–LP Agreement Across Projection Horizons

Figure 7 shows model agreement varies across forecast horizons, with stronger consistency at longer horizons. The 95% interval overlap remains consistently high across all horizons, while the sign match rate rises from 58.5% at 30 days to 76.3% at 1,095 days. This pattern indicates the two approaches diverge more on short-term dynamics but converge on longer-term directional effects, confirming our main findings are robust to the choice of estimation framework.

9 Conclusion

This study set out to answer a question of growing importance for retail crypto investors: does the popular "HODL" strategy, buying and holding cryptocurrency, truly deliver on its promises? Through a combination of large-scale Monte Carlo simulation (7 individual tokens plus one basket containing 378 non-stablecoin crypto assets, totaling 480 million scenarios) and Bayesian multi-horizon local projection analysis, we tested two related claims central to the crypto investment narrative. First, whether the risk and return distribution across the broad crypto market justifies the HODL strategy when accounting for trading costs, opportunity costs, and six holding-period intervals ranging from 1 to 1095 days. Second, whether past realized risk and return distributions are the dominant factor affecting future outcomes.

Our Monte Carlo simulation reveals two critical insights that challenge the "HODL" narrative.

First, we quantify substantial survivorship bias: while people often cite Bitcoin or Ethereum as proof of the HODL strategy's success, our ALL basket, which randomly samples from 378 non-stablecoin tokens, tells a starkly different story. At the longest holding horizon (731–1095 days), the median excess return is −28.4%, yet the top-quartile mean reaches +1326.7%. This dramatic dispersion exposes how "HODL missionaries" exploit upper-quartile performance to overstate rewards available to the average investors. Moreover, for holding periods beyond 181–365 days, the probability of experiencing losses greater than 10% exceeds 50%, and the $\text{CVaR}_{1\%}$ approaches near-total loss (100%).

Second, we reveal a complex trade-off: as holding periods lengthen, moderate and extreme risks often move in opposite directions. For successful tokens such as BTC and ETH, longer horizons show declining moderate loss probabilities (losses $> 10\%$), dropping from peaks around 90–180 days to much lower levels at the longest horizons, while extreme tail risk (CVaR) rises through mid-range horizons before declining at 2–3 year holding periods. Crucially, however, this pattern is not generalizable when we examine the market-level distribution. The ALL basket demonstrates that across the broader cryptocurrency market, both moderate and extreme risks increase simultaneously with holding length, suggesting no miraculous recovery but rather escalating risk with longer holding periods.

Our Bayesian multi-horizon local projection analysis directly addresses the second claim underlying the HODL narrative: that past realized risk and return distributions are the dominant factor affecting future outcomes. We find the opposite. While endogenous predictors (realized risk and return metrics) generate some of the largest short-horizon impulse responses, they exhibit limited cross-basket stability compared with macro-finance predictors. Meta-analysis across all baskets reveals economically negligible population-level effects for endogenous predictors. Instead, macrofinance sentiment indicators, particularly FGI EMA24w, emerge as both the most stable and most influential predictors of long-horizon (181–1095 days) return metrics. Where significant, a one-standard-deviation increase in FGI EMA24w lowers top-quartile returns by roughly 15–22 percentage points and median returns by 6–10 percentage points across these horizons. This dominance of sentiment over historical fundamentals undermines a core premise of the HODL narrative: that past realized distributions drive future outcomes. Moreover, we document striking cross-basket heterogeneity: DOGE exhibits sentiment sensitivity more than three times larger than BNB and ADA (48 versus 10–11 percentage points for top-quartile returns), underscoring the heightened speculative nature of meme coins.

Taken together, our findings fundamentally challenge the viability of the HODL strategy for retail investors. The evidence is unequivocal: relying on top-quartile success stories to forecast personal returns is not optimistic, it is dangerously misleading. The gap between median outcomes (−28.4%) and top-quartile performance (+1326.7%) in the ALL basket represents not a modest discount but a chasm between financial ruin and extraordinary gain. Moreover, the belief that historical risk-return distributions can guide future investment decisions finds no empirical support in our data. Instead, macroeconomic conditions and sentiment shifts, captured most powerfully by FGI EMA24w, emerge as the primary drivers of cryptocurrency returns across holding horizons. Retail investors should prioritize monitoring macro-sentiment indicators rather than extrapolating their expected returns from past return patterns.

Appendix A Additional Tables and Figures

A.1 Monte Carlo Metrics and Definitions

For each *basket* × *horizon*

$$\mathcal{X} = \{X_1, \ldots, X_N\}, \qquad N = |\mathcal{X}|,$$

be the sample of excess-return observations. Let μ be the sample mean, σ the unbiased standard deviation, and $q_p \equiv \text{Quantile}_p(\mathcal{X})$.

A.1.1 Additional notation

- $n_{\text{neg}} = |\{\, i : X_i < 0 \,\}|$ — number of negative returns
- $\mu_{\text{neg}} = \dfrac{1}{n_{\text{neg}}} \displaystyle\sum_{X_i<0} X_i$ — mean of negative returns

A.1.2 Tail-probability thresholds used in the paper

$$\alpha = 10\% \quad \text{for weekly statistics,} \qquad \alpha = 1\% \quad \text{for overall statistics.}$$

A.1.3 Metrics and definitions

Statistic	Formula
Sample size	N
Mean excess return	$\mu = \frac{1}{N}\sum_{i=1}^{N} X_i$
Median excess return	$\tilde{X} = q_{0.50}$
Unbiased standard deviation	$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(X_i-\mu)^2}$
Inter-quartile range	$\mathrm{IQR} = q_{0.75} - q_{0.25}$
Sharpe ratio	$\mathrm{Sharpe} = \frac{\mu}{\sigma}$
Sortino ratio	$\mathrm{Sortino} = \frac{\mu}{\sigma_{\text{neg}}}, \quad \sigma_{\text{neg}}^2 = \frac{1}{n_{\text{neg}}-1}\sum_{X_i<0}\left(X_i-\mu_{\text{neg}}\right)^2$
Value-at-Risk	$\mathrm{VaR}_\alpha = \max\!\left(0,\ -q_\alpha\right)$
Conditional VaR	$\mathrm{CVaR}_\alpha = \max\!\Big(0,\ -\frac{1}{\lvert\{i : X_i \le q_\alpha\}\rvert}\sum_{X_i \le q_\alpha} X_i\Big)$
Probability of profit	$p_{\text{profit}} = \frac{\lvert\{i : X_i > 0\}\rvert}{N}$
Probability of $> 10\%$ loss	$p_{\text{sig loss}} = \frac{\lvert\{i : X_i < -0.10\}\rvert}{N}$
75-th percentile	$q_{0.75}$
Mean of top-quartile returns	$\bar{X}_{\text{top 25}} = \frac{1}{\lvert\{i : X_i \ge q_{0.75}\}\rvert}\sum_{X_i \ge q_{0.75}} X_i$
Top-quartile proportion	$p_{\text{top 25}} = \frac{\lvert\{i : X_i \ge q_{0.75}\}\rvert}{N}$
Skewness	$\gamma_1^{(\text{G1})} = \frac{N\sqrt{N-1}}{N-2}\cdot\frac{\sum_{i=1}^{N}(X_i-\mu)^3}{\left(\sum_{i=1}^{N}(X_i-\mu)^2\right)^{3/2}}$
Excess kurtosis	$\kappa^{(\text{G2})} = \frac{N(N+1)(N-1)\sum_{i=1}^{N}(X_i-\mu)^4}{(N-2)(N-3)\left(\sum_{i=1}^{N}(X_i-\mu)^2\right)^2} - \frac{3(N-1)^2}{(N-2)(N-3)}$

A.1.4 Core aggregated statistics by basket and horizon

Table 4: Key Monte-Carlo statistics by basket and holding horizon

Basket	Horizon (days)	μ	$\tilde{X}$	σ	Sharpe	Sortino	$\mathrm{VaR}_{1\%}$	$\mathrm{CVaR}_{1\%}$	$p_{\mathrm{sig\,loss}}$	$q_{0.75}$	$\overline{X}_{\mathrm{top25}}$
ALL	1–30	0.045	-0.017	1.948	0.023	0.351	0.506	0.591	0.299	0.094	0.473
ALL	31–90	0.216	-0.068	2.847	0.076	1.143	0.725	0.791	0.465	0.248	1.438
ALL	91–180	0.597	-0.112	11.376	0.052	2.710	0.834	0.887	0.509	0.389	3.159
ALL	181–365	1.324	-0.160	28.489	0.046	5.128	0.937	0.964	0.534	0.558	6.256
ALL	366–730	2.876	-0.196	25.525	0.113	10.218	1.003	1.020	0.536	0.960	12.578
ALL	731–1095	3.027	-0.284	15.869	0.191	10.659	1.061	1.079	0.552	1.395	13.267
BTC	1–30	0.031	0.009	0.160	0.192	0.371	0.328	0.386	0.149	0.088	0.234
BTC	31–90	0.144	0.058	0.403	0.356	1.133	0.468	0.506	0.269	0.314	0.676
BTC	91–180	0.396	0.191	0.833	0.475	2.537	0.536	0.568	0.286	0.624	1.494
BTC	181–365	1.043	0.519	1.971	0.529	5.189	0.676	0.711	0.231	1.274	3.451
BTC	366–730	2.949	1.514	4.822	0.612	14.555	0.700	0.742	0.188	3.595	8.967
BTC	731–1095	6.743	2.969	10.311	0.654	38.751	0.544	0.587	0.091	8.309	20.028
ETH	1–30	0.025	0.001	0.204	0.123	0.233	0.423	0.481	0.219	0.104	0.289
ETH	31–90	0.104	0.021	0.458	0.227	0.643	0.605	0.642	0.364	0.316	0.710
ETH	91–180	0.274	0.076	0.893	0.306	1.389	0.735	0.778	0.393	0.506	1.406
ETH	181–365	0.866	0.174	2.258	0.383	3.366	0.873	0.899	0.346	0.808	3.696
ETH	366–730	2.936	0.479	5.737	0.512	12.786	0.870	0.895	0.367	2.475	11.269
ETH	731–1095	5.308	1.643	7.678	0.691	22.996	0.829	0.872	0.252	8.206	16.644
XRP	1–30	0.057	-0.013	0.488	0.116	0.518	0.446	0.539	0.244	0.078	0.459
XRP	31–90	0.175	-0.016	0.889	0.197	1.114	0.600	0.669	0.381	0.217	1.067
XRP	91–180	0.265	-0.010	0.976	0.272	1.481	0.653	0.725	0.401	0.278	1.483
XRP	181–365	0.436	-0.027	1.315	0.331	2.078	0.756	0.826	0.445	0.525	2.283
XRP	366–730	0.684	0.045	1.628	0.420	3.186	0.849	0.898	0.454	1.131	3.172
XRP	731–1095	0.805	0.466	1.529	0.526	3.437	0.903	0.943	0.348	1.411	2.918

A.2 Methodology

A.2.1 Macro unit-root tests

Table 5: Unit-root tests for macro series

Series	DF-GLS p (c)	KPSS p (c)	ZA p (c)	DF-GLS p (ct)	KPSS p (ct)	ZA p (ct)	Decision (c)	Decision (ct)
BTC log return	0.000	0.100	0.000	0.000	0.100	0.001	STATIONARY	STATIONARY
Crypto Fear & Greed Index	0.001	0.078	0.002	0.000	0.098	0.005	STATIONARY	STATIONARY
HY OAS spread (BAMLH0A0HYM2)	0.014	0.076	0.584	0.092	0.023	0.556	AMBIGUOUS	UNIT ROOT
Fed funds rate (DFF)	0.260	0.010	0.024	0.422	0.010	0.697	AMBIGUOUS	UNIT ROOT
US 10Y Treasury yield (DGS10)	0.469	0.010	0.691	0.859	0.010	0.877	UNIT ROOT	UNIT ROOT
Trade-weighted USD (broad) (DTWEXBGS)	0.625	0.010	0.323	0.232	0.022	0.421	UNIT ROOT	UNIT ROOT
Nasdaq Composite (NASDAQCOM)	0.828	0.010	0.232	0.240	0.011	0.327	UNIT ROOT	UNIT ROOT
10Y-2Y Treasury spread (T10Y2Y)	0.266	0.010	0.404	0.742	0.010	0.893	UNIT ROOT	UNIT ROOT
VIX index (VIXCLS)	0.000	0.100	0.000	0.000	0.010	0.001	STATIONARY	AMBIGUOUS

A.3 Posterior Results

A.3.1 Posterior effect metrics

Let $\beta_{h,p,y}$ denote the local projection coefficient (see equation (5)). These coefficients represent standardized impulse responses: $\beta_{h,p,y}$ measures how target y changes in standardized units at horizon h when predictor p increases by one standard deviation at time t, conditional on all other predictors in $Z_{t,h}$. We denote the posterior samples by index $s = 1, \ldots, S$, where S is the total number of MCMC draws.

To analyze effects in target y's native units, we scale the standardized coefficients by $\sigma_{h,y}$, the reference standard deviation of target y at horizon h:

$$\beta^*_{h,p,y} = \beta_{h,p,y} \cdot \sigma_{h,y},$$

Table 6: Posterior Effect Metric Definitions

Symbol	**Definition / Computation**
$\beta^*_{h,p,y}$	Effect on target y (native units) per one-standard-deviation shock to predictor p: $\beta^*_{h,p,y} = \beta_{h,p,y}\,\sigma_{h,y}$.
$\mathrm{med}_{h,p,y}$	Posterior median: $\mathrm{med}_{h,p,y} = Q_{0.50}\{\beta^*_{h,p,y,s}\}_{s=1}^{S}$.
$\mathrm{CrI}_{0.95}(h,p,y) = [\ell_{h,p,y}, u_{h,p,y}]$	95% credible interval: $\ell_{h,p,y} = Q_{0.025}\{\beta^*_{h,p,y,s}\},\ u_{h,p,y} = Q_{0.975}\{\beta^*_{h,p,y,s}\}$.
$\pi_+(h,p,y)$	Posterior probability of positive effect: $\pi_+(h,p,y) = \frac{1}{S}\sum_{s=1}^{S} \mathbb{1}\{\beta^*_{h,p,y,s} > 0\}$.
$\mathcal{S}_{0.95}(h,p,y) \in \{0,1\}$	Significance indicator (two-sided): $\mathcal{S}_{0.95}(h,p,y) = \mathbb{1}\big\{\ell_{h,p,y} > 0 \ \vee\ u_{h,p,y} < 0 \ \vee\ \pi_+(h,p,y) \geq 0.95 \ \vee\ \pi_+(h,p,y) \leq 0.05\big\}$.

Ranking and aggregation procedures. To identify the most important effects, we apply two complementary rankings:

1. **Per-horizon ranking.** At each horizon h, we select all significant predictor-target pairs ($\mathcal{S}_{0.95}(h,p,y) = 1$) and keep the eight with the largest absolute median effects $|\,\mathrm{med}_{h,p,y}\,|$.
2. **Cross-basket stability ranking.** For each horizon-target pair (h,y), we aggregate results across all baskets and rank predictors by the number of baskets where $\mathcal{S}_{0.95}(h,p,y) = 1$; in the case of ties, we then sort by their mean and maximum absolute median effects. We report the four predictors with the highest basket counts for each (h,y) combination.

A.3.2 Posterior label definitions

Label	Series	Transform
VIX EMA24 / EMA24w	VIX index (VIXCLS)	24-week exponential moving average.
FGI EMA24 / EMA24w	Crypto Fear & Greed Index	24-week exponential moving average.
BAMLH0A0HYM2 RVol24w	HY OAS spread (BAMLH0A0HYM2)	24-week rolling volatility.
T10Y2Y EMA24 / EMA24w	10Y-2Y Treasury spread (T10Y2Y)	24-week exponential moving average.
T10Y2Y EMA12 / EMA12w	10Y-2Y Treasury spread (T10Y2Y)	12-week exponential moving average.
T10Y2Y RVol24w	10Y-2Y Treasury spread (T10Y2Y)	24-week rolling volatility.
BTC RVol12w	BTC log return	12-week rolling volatility.
BTC EMA4 / EMA4w	BTC log return	4-week exponential moving average.
BTC EMA8 / EMA8w	BTC log return	8-week exponential moving average.
BTC EMA12 / EMA12w	BTC log return	12-week exponential moving average.
NASDAQCOM EMA4 / EMA4w	Nasdaq Composite (NASDAQCOM)	4-week exponential moving average.

Table 7: Posterior label definitions referenced in Figures 4 and 5.

A.3.3 A Two-Stage Bayesian Meta Framework

We examine how endogenous predictors' effects evolve across forecast horizons by pooling information across baskets through a Bayesian meta-analysis (Sutton and Abrams, 2001; Röver, 2020). For each basket b, horizon h, predictor p, and target y, the local projection model (§6.6) yields posterior samples $\{\beta^{(s)}_{b,h,p,y}\}_{s=1}^{S}$. Following Appendix A.3.1, we scale these standardized coefficients by the target's reference standard deviation to obtain effects in native units:

$$\theta^{(s)}_{b,h,p,y} \;=\; \sigma_{h,y}\,\beta^{(s)}_{b,h,p,y},$$

where θ measures the effect in target-native units per one-standard-deviation increase in predictor p.

Stage 1: Basket-level posterior summaries. For each (b, h, p, y) we compute the posterior mean $\bar{\theta}_{b,h,p,y}$, standard deviation $s_{b,h,p,y}$, equal-tailed credible intervals at 68% and 95%, and the posterior probability of positive effects $\text{ppos}_{b,h,p,y} = \Pr(\theta_{b,h,p,y} > 0)$. We mark an observation as 95%-significant if the credible interval excludes zero.

Stage 2: Meta-analysis across horizons. For each predictor-target pair (p, y), we pool the Stage-1 summaries across basket-horizon combinations. Let $i = 1, \ldots, N$ index these combinations, with $\bar{\theta}_i$ and s_i denoting the posterior mean and standard deviation, and $h(i) \in \{30, 60, 90, 180, 365, 730, 1095\}$ the horizon in days. We fit a horizon-specific random-effects model

$$\bar{\theta}_i \mid \mu_{h(i)}, \tau \;\sim\; \mathcal{N}\Big(\mu_{h(i)},\; s_i^2 + \tau^2\Big)\,,$$

where μ_h represents the population mean effect at horizon h and $\tau > 0$ quantifies between-basket heterogeneity (assumed constant across horizons). To smooth temporal variation in effects, we impose a first-order random walk prior over horizons measured in days:

$$\mu_{h_1} \sim \mathcal{N}(0, \sigma_\mu^2), \qquad \mu_{h_{j+1}} = \mu_{h_j} + \delta_j, \qquad \delta_j \sim \mathcal{N}(0,\ \sigma_{\text{rw}}^2\, \Delta_j),$$

with $\Delta_j = (h_{j+1} - h_j)/30$ to handle unequally spaced horizons, $\sigma_\mu = 1.0$, $\sigma_{\text{rw}} \sim \text{HalfNormal}(0.05)$, and $\tau \sim \text{HalfStudentT}(\nu = 3, \sigma = 0.3)$.

We perform inference using PyMC's NUTS with 4 chains, 3,000 tuning iterations and 2,000 posterior draws per chain (target acceptance 0.99).

Panel A: pooled effect $\mu(h)$. For each horizon, we plot the posterior median of μ_h with 68% and 95% credible bands.

Panel B: retention and half-life. We plot the retention ratio $R(h) = |\mu_h|/|\mu_{30}|$ (posterior median with 68%, 95% credible bands). The half-life is the horizon where $R(h)$ falls to 0.5. If fewer than 50% of draws cross this threshold by 1095 days, we report the half-life as right-censored.

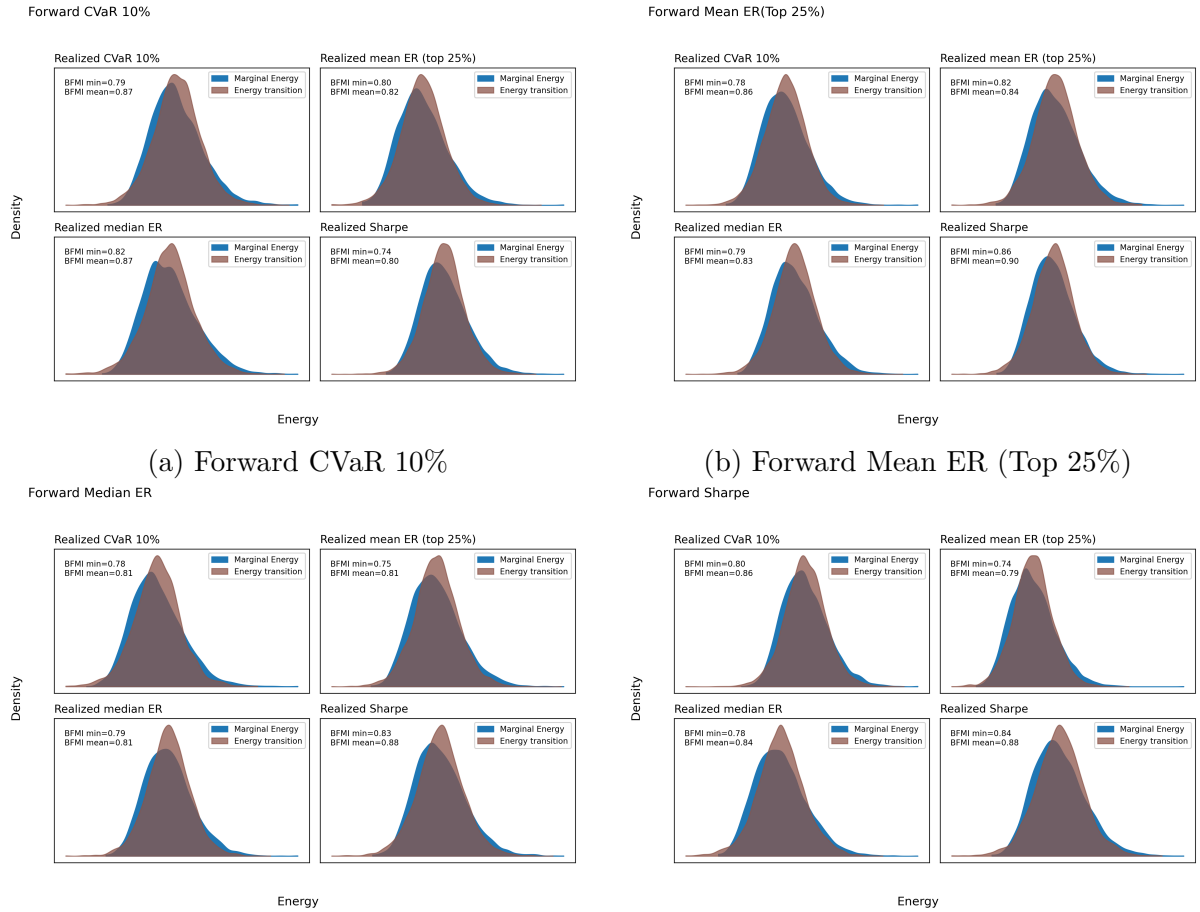
Panel C: cross-basket significance. For each horizon, we count how many baskets exhibit 95%-significant effects (credible interval excludes zero).

A.3.4 Meta-Analysis Model Convergence

Table 8: Convergence diagnostics for horizon-specific random-effects meta-analysis models

Basket	Predictor	Max $\hat{R}$	Min ESS (bulk)	Min ESS (tail)	Max MCSE/SD	Min BFMI	Divergence rate
Forward CVaR 10%	Realized CVaR 10%	1.00	2713	2895	0.015	0.79	0.00025
Forward CVaR 10%	Realized mean ER (top 25%)	1.00	2595	2885	0.015	0.80	0.00025
Forward CVaR 10%	Realized median ER	1.00	2368	1925	0.013	0.82	0.00038
Forward CVaR 10%	Realized Sharpe	1.00	2273	1779	0.015	0.74	0.00038
Forward Mean ER (Top 25%)	Realized CVaR 10%	1.00	3026	3129	0.013	0.78	0.00013
Forward Mean ER (Top 25%)	Realized mean ER (top 25%)	1.00	3036	2672	0.013	0.82	0.00000
Forward Mean ER (Top 25%)	Realized median ER	1.00	3111	1689	0.014	0.79	0.00013
Forward Mean ER (Top 25%)	Realized Sharpe	1.00	3513	2702	0.013	0.86	0.00013
Forward Median ER	Realized CVaR 10%	1.00	2906	3054	0.013	0.78	0.00000
Forward Median ER	Realized mean ER (top 25%)	1.00	3219	2858	0.013	0.75	0.00013
Forward Median ER	Realized median ER	1.00	2371	2984	0.013	0.79	0.00000
Forward Median ER	Realized Sharpe	1.00	3722	2857	0.014	0.83	0.00025
Forward Sharpe	Realized CVaR 10%	1.00	2295	2276	0.014	0.80	0.00075
Forward Sharpe	Realized mean ER (top 25%)	1.00	2373	2032	0.015	0.74	0.00063
Forward Sharpe	Realized median ER	1.00	2964	2450	0.015	0.78	0.00038
Forward Sharpe	Realized Sharpe	1.00	2340	2557	0.016	0.84	0.00100

Table 8 reports convergence diagnostics for all 16 target-predictor pairs. All models achieve convergence with $\hat{R} \leq 1.00$, ESS ≥ 1689, and divergence rates $\leq 0.1\%$.



(a) Forward CVaR 10%

(b) Forward Mean ER (Top 25%)

(c) Forward Median ER

(d) Forward Sharpe

Figure 8: Energy diagnostics for all target–predictor pairs.

Figure 8 shows energy diagnostics for each target–predictor pair. With minimum BFMI of 0.74, the overlapping energy transition and marginal energy distributions indicate the models did not encounter geometric pathologies.

A.3.5 Meta-Analysis Results

Table 9: Meta-analysis results: forward mean ER (Top 25%) significant horizons

Basket	Horizon (days)	Predictor	Target	Mean	SD	Pr(>0)	95% Credible Interval (lower)	95% Credible Interval (upper)	95% Credible Interval
BNB	366–730	Realized mean ER (top 25%)	Forward Mean ER (Top 25%)	0.0557	0.0267	0.9824	0.0041	0.1087	✓
BNB	731–1095	Realized mean ER (top 25%)	Forward Mean ER (Top 25%)	0.1302	0.0606	0.9841	0.0133	0.2480	✓
ETH	1–30	Realized median ER	Forward Mean ER (Top 25%)	0.0057	0.0028	0.9764	0.0001	0.0112	✓
ETH	31–90	Realized median ER	Forward Mean ER (Top 25%)	0.0080	0.0040	0.9775	0.0002	0.0159	✓
ETH	91–180	Realized median ER	Forward Mean ER (Top 25%)	0.0150	0.0072	0.9800	0.0009	0.0294	✓
ETH	181–365	Realized median ER	Forward Mean ER (Top 25%)	0.0358	0.0168	0.9828	0.0029	0.0693	✓
ETH	366–730	Realized median ER	Forward Mean ER (Top 25%)	0.0640	0.0288	0.9862	0.0072	0.1218	✓
ETH	731–1095	Realized median ER	Forward Mean ER (Top 25%)	0.0777	0.0343	0.9879	0.0101	0.1461	✓

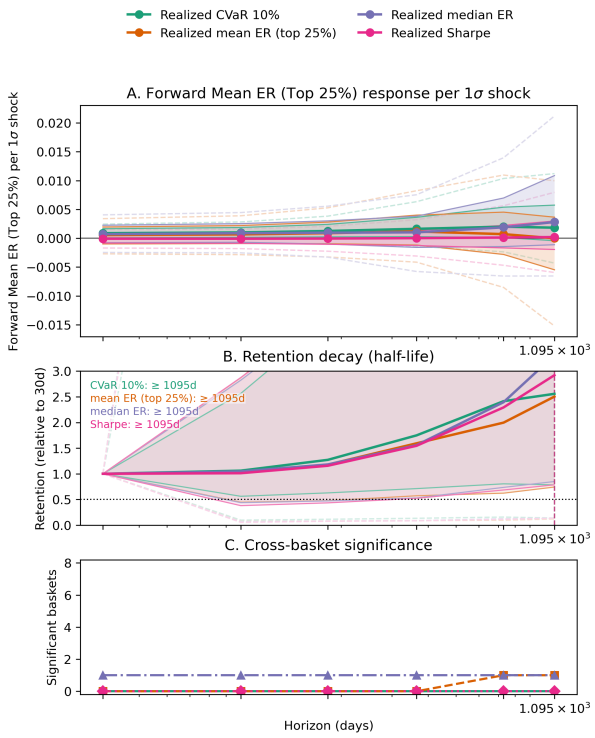
Table 10: Meta-analysis results: forward Sharpe significant horizons

Basket	Horizon (days)	Predictor	Target	Mean	SD	Pr(>0)	95% Credible Interval (lower)	95% Credible Interval (upper)	95% Credible Interval
LINK	1–30	Realized CVaR 10%	Forward Sharpe	0.1183	0.0520	0.9908	0.0167	0.2227	✓
LINK	31–90	Realized CVaR 10%	Forward Sharpe	0.1194	0.0526	0.9908	0.0165	0.2248	✓
LINK	91–180	Realized CVaR 10%	Forward Sharpe	0.1167	0.0513	0.9908	0.0167	0.2189	✓
LINK	181–365	Realized CVaR 10%	Forward Sharpe	0.0430	0.0190	0.9902	0.0059	0.0809	✓
LINK	366–730	Realized CVaR 10%	Forward Sharpe	0.0138	0.0061	0.9900	0.0018	0.0260	✓
LINK	731–1095	Realized CVaR 10%	Forward Sharpe	0.0134	0.0060	0.9904	0.0017	0.0252	✓
ETH	1–30	Realized median ER	Forward Sharpe	0.1163	0.0584	0.9760	0.0009	0.2320	✓
ETH	31–90	Realized median ER	Forward Sharpe	0.1575	0.0789	0.9759	0.0013	0.3128	✓
ETH	91–180	Realized median ER	Forward Sharpe	0.0986	0.0494	0.9756	0.0006	0.1955	✓
ETH	181–365	Realized median ER	Forward Sharpe	0.0374	0.0187	0.9755	0.0003	0.0747	✓
ETH	366–730	Realized median ER	Forward Sharpe	0.0250	0.0124	0.9776	0.0008	0.0496	✓
ETH	731–1095	Realized median ER	Forward Sharpe	0.0366	0.0179	0.9781	0.0012	0.0718	✓

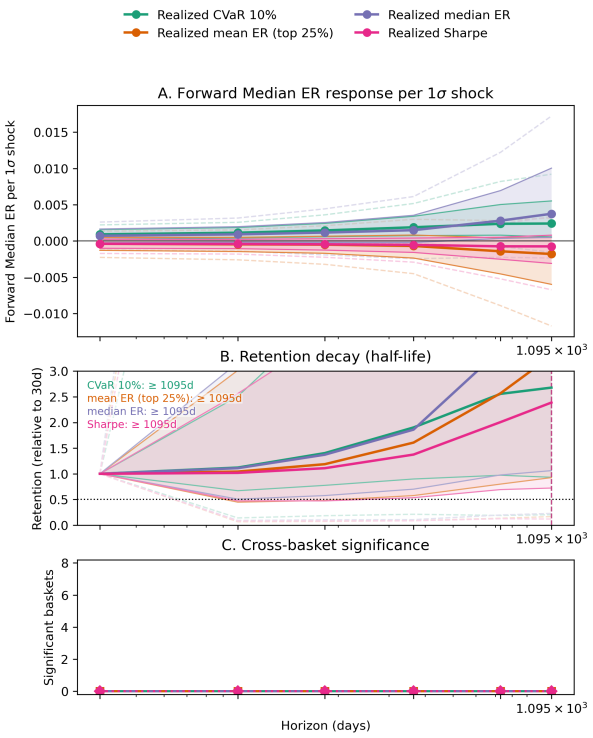
Tables 9 and 10 report basket-level posterior summaries $\bar{\theta}_{b,h,p,y}$ for combinations exhibiting 95%-significant effects, as defined in Appendix A.3.3. Out of the four target variables, only forward Mean ER (Top 25%) and forward Sharpe exhibit significant effects for at least one basket-predictor-horizon combination. For forward Sharpe, only two basket-predictor pairs exhibit significant effects. LINK's realized CVaR 10% shows positive effects that remain substantial over horizons between 1–180 days; one standard deviation would increase forward Sharpe by more than 0.11 in Sharpe units, but later decay to approximately 0.013 in Sharpe units at longer horizons. For ETH, realized median ER exhibits strong impact over horizons between 1–90 days (one standard deviation of realized median ER would increase forward Sharpe by more than 0.11 in Sharpe units), and then later decay with horizon; for the horizon 731–1095 days the effects decay to 0.0366 in Sharpe units.

For forward Mean ER (Top 25%), a one-standard-deviation shock from realized median ER produces positive and significant effects on ETH's forward Mean ER (Top 25%) across all horizons, with the effect at 731–1095 days reaching 7.77 percentage points. BNB's realized mean ER (top 25%) also shows positive and significant effects on forward Mean ER (Top 25%) at longer horizons (366–730 and 731–1095 days), with the longest-horizon effect reaching 13.02 percentage points.

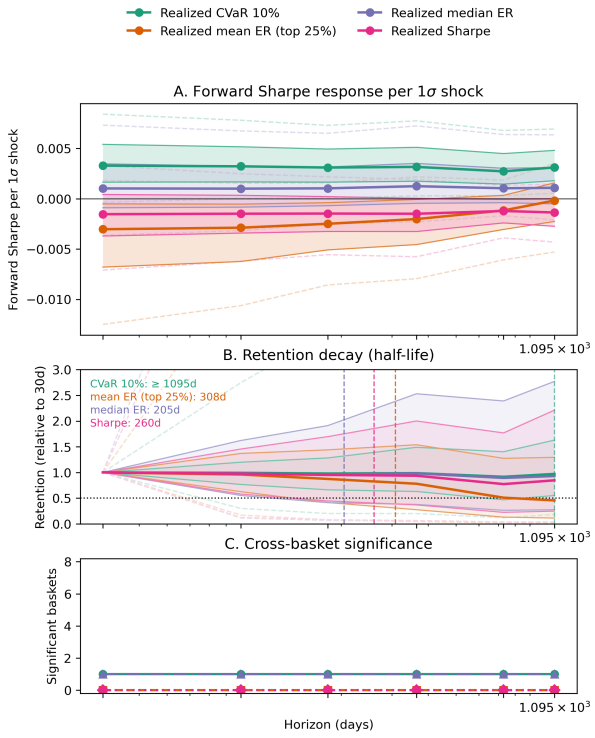
Figure 9 reports the pooled (population-average) effect $\mu(h)$ under the random-effects model as defined in Appendix A.3.3. At each horizon h, we plot the posterior median (solid line with markers), the 68% credible band (shaded region), and the 95% credible interval (outer dashed lines), with each predictor distinguished by color. The pooled estimates reveal generally weak population-level effects across all targets. For forward Sharpe, realized CVaR 10% shows the strongest pooled effect (one standard deviation would increase forward Sharpe by roughly 0.0027–0.0033 in Sharpe units across horizons); however, most endogenous predictor effects remain economically negligible. Overall, the section reveals substantial cross-basket heterogeneity for the endogenous predictors. While specific pairs exhibit large effects (e.g., LINK's CVaR 10% driving more than 11 percentage points increases in forward Sharpe), these patterns are basket-specific rather than generalizable.



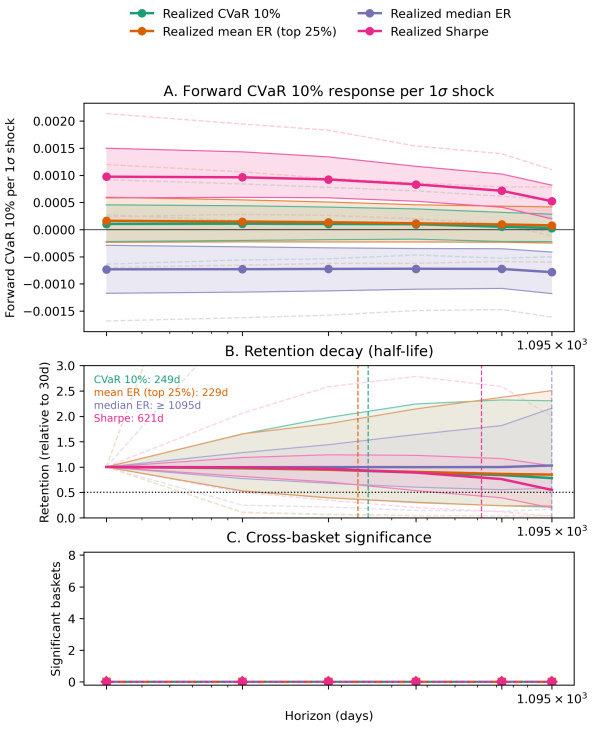
(a) Forward Mean ER (Top 25%)



(b) Forward Median ER



(c) Forward Sharpe



(d) Forward CVaR 10%

Figure 9: Endogenous predictor effects across forecast horizons by target.

A.3.6 Long-horizon top-quartile impulse responses to FGI EMA24w shocks (Figure 6 data)

Basket	Target	Horizon	Median	Lower	Upper	p_{pos}	Sig. (95%)
DOGE	Forward mean ER (Top 25%)	181–365	-0.477	-1.226	-0.236	0.000	✓
BNB	Forward mean ER (Top 25%)	181–365	-0.114	-0.282	-0.030	0.001	✓
ADA	Forward mean ER (Top 25%)	181–365	-0.100	-0.261	-0.006	0.018	✓
ETH	Forward mean ER (Top 25%)	181–365	-0.053	-0.146	-0.001	0.022	✓
BTC	Forward mean ER (Top 25%)	181–365	-0.016	-0.033	-0.001	0.018	✓
LINK	Forward mean ER (Top 25%)	181–365	-0.007	-0.033	0.002	0.065	
DOGE	Forward mean ER (Top 25%)	366–730	-0.512	-0.776	-0.262	0.000	✓
ADA	Forward mean ER (Top 25%)	366–730	-0.115	-0.219	-0.019	0.009	✓
BNB	Forward mean ER (Top 25%)	366–730	-0.108	-0.202	-0.038	0.000	✓
ETH	Forward mean ER (Top 25%)	366–730	-0.076	-0.158	-0.002	0.021	✓
LINK	Forward mean ER (Top 25%)	366–730	-0.027	-0.068	0.013	0.087	
BTC	Forward mean ER (Top 25%)	366–730	-0.025	-0.050	-0.001	0.021	✓
DOGE	Forward mean ER (Top 25%)	731–1095	-0.486	-0.726	-0.242	0.000	✓
BNB	Forward mean ER (Top 25%)	731–1095	-0.135	-0.234	-0.033	0.005	✓
LINK	Forward mean ER (Top 25%)	731–1095	-0.074	-0.173	0.029	0.073	
ETH	Forward mean ER (Top 25%)	731–1095	-0.066	-0.141	0.008	0.041	
BTC	Forward mean ER (Top 25%)	731–1095	-0.047	-0.091	-0.001	0.025	✓
ADA	Forward mean ER (Top 25%)	731–1095	-0.019	-0.112	0.079	0.343	

A.3.7 Long-horizon median impulse responses to FGI EMA24w shocks (Figure 6 data)

Basket	Target	Horizon	Median	Lower	Upper	p_{pos}	Sig. (95%)
DOGE	Forward median ER	181–365	-0.118	-0.338	-0.017	0.013	✓
BNB	Forward median ER	181–365	-0.061	-0.121	-0.022	0.004	✓
ETH	Forward median ER	181–365	-0.053	-0.119	-0.007	0.007	✓
ADA	Forward median ER	181–365	-0.037	-0.107	0.010	0.051	
LINK	Forward median ER	181–365	-0.007	-0.023	-0.001	0.012	✓
BTC	Forward median ER	181–365	-0.004	-0.015	0.007	0.211	
DOGE	Forward median ER	366–730	-0.120	-0.220	-0.023	0.007	✓
ADA	Forward median ER	366–730	-0.087	-0.140	-0.038	0.000	✓
BNB	Forward median ER	366–730	-0.068	-0.109	-0.030	0.000	✓
ETH	Forward median ER	366–730	-0.054	-0.107	-0.006	0.014	✓
LINK	Forward median ER	366–730	-0.025	-0.050	-0.003	0.013	✓
BTC	Forward median ER	366–730	-0.006	-0.021	0.011	0.228	
BNB	Forward median ER	731–1095	-0.112	-0.174	-0.051	0.000	✓
DOGE	Forward median ER	731–1095	-0.108	-0.204	-0.016	0.011	✓
LINK	Forward median ER	731–1095	-0.085	-0.156	-0.018	0.006	✓
ETH	Forward median ER	731–1095	-0.037	-0.082	0.009	0.061	
ADA	Forward median ER	731–1095	-0.035	-0.079	0.015	0.085	
BTC	Forward median ER	731–1095	-0.011	-0.042	0.019	0.233	

Appendix B Additional Figures

B.1 Additional figures for Monte Carlo simulation results

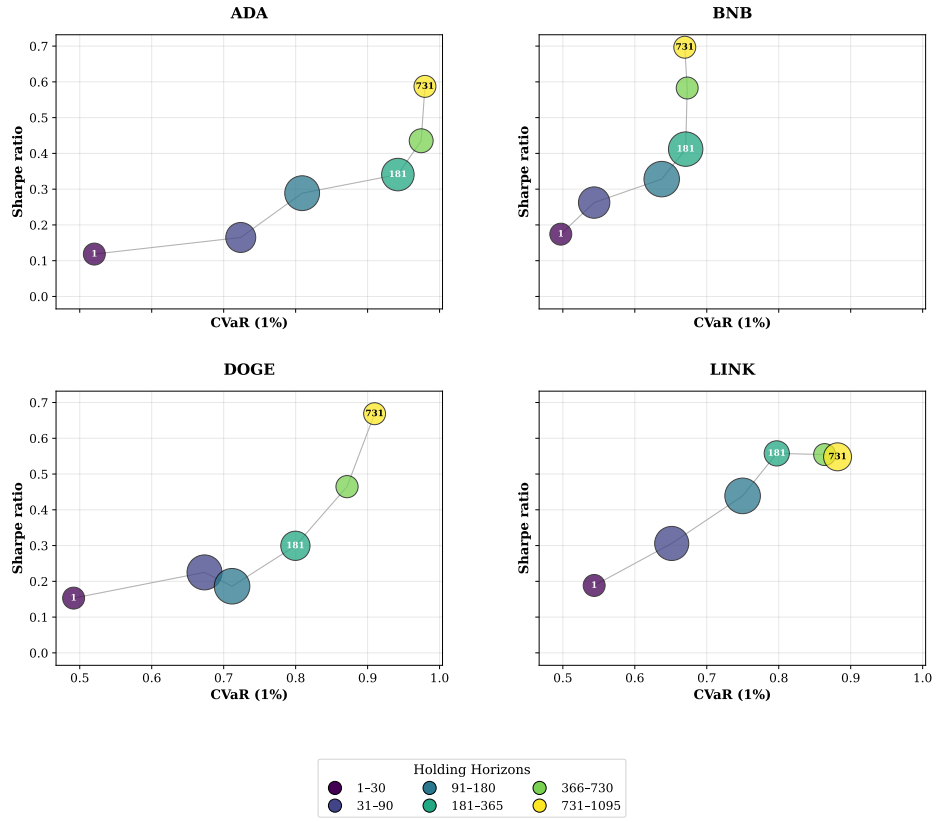


Figure 10: Risk–return trade-off across holding horizons for the remaining four baskets

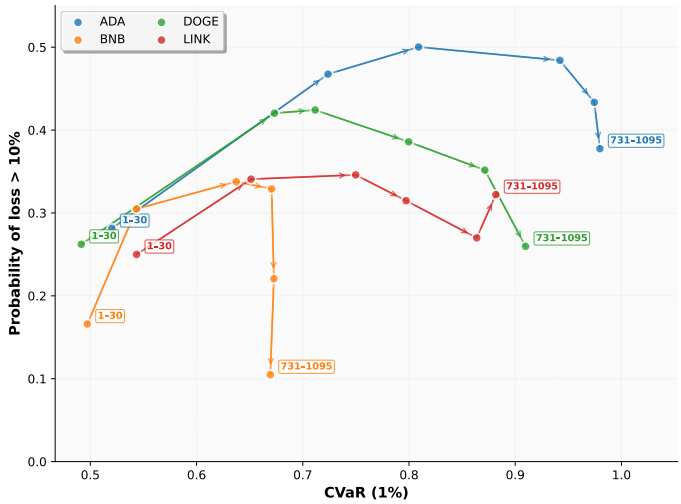


Figure 11: Trade-off between moderate and extreme risks across holding horizons for the remaining four baskets

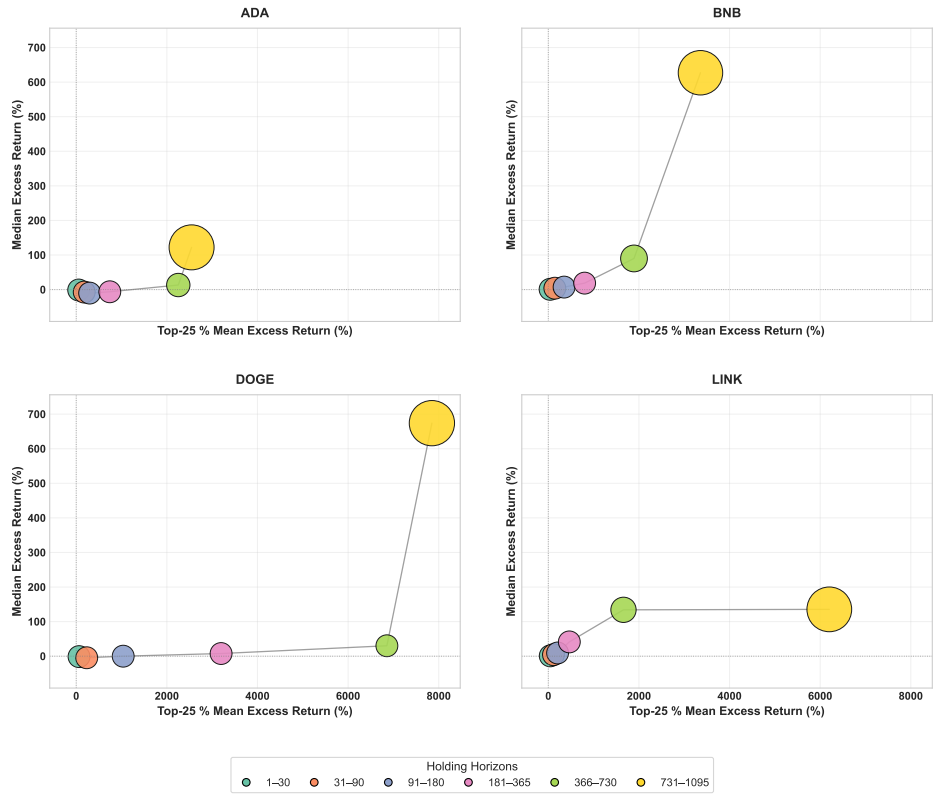


Figure 12: Median vs. top-25% mean excess returns across holding horizons for the remaining four baskets

B.2 Convergence diagnostics

B.2.1 Trace and autocorrelation diagnostics

Figures 13 and 14 present trace and autocorrelation plots for the parameters with the highest $\hat{R}$ and MCSE/SD ratios in each basket. All chains exhibit stable mixing with rapidly decaying autocorrelation.

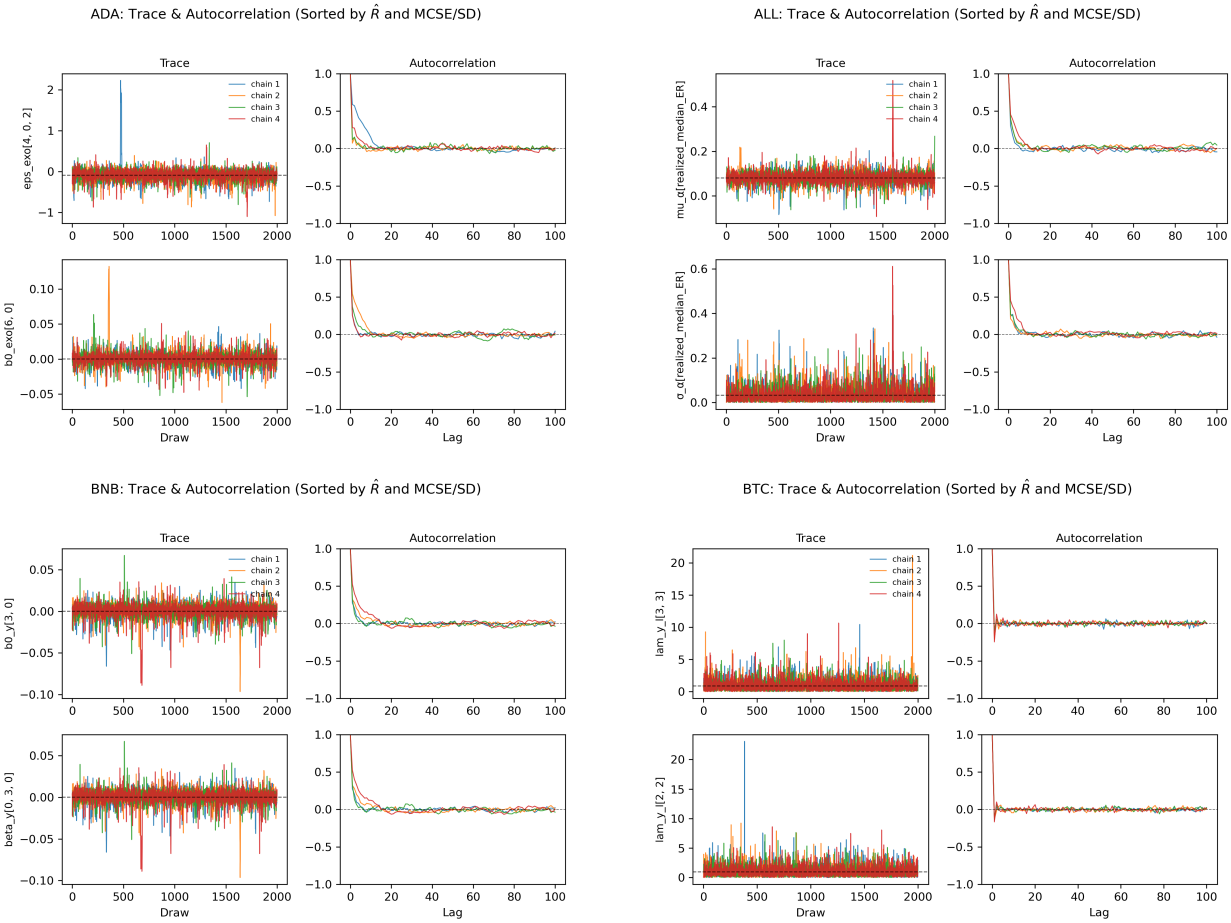


Figure 13: Trace and autocorrelation diagnostics for ADA, ALL, BNB, and BTC baskets.

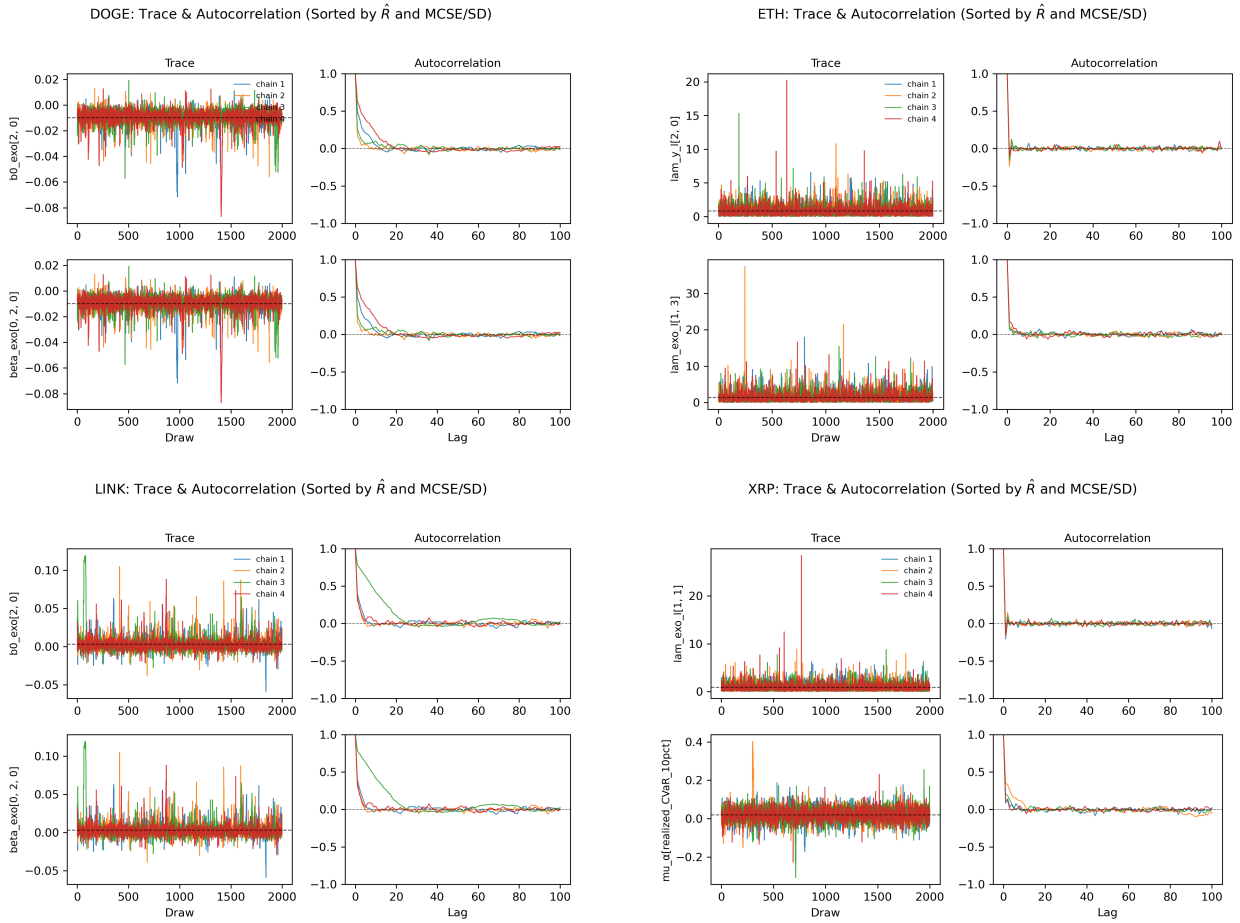


Figure 14: Trace and autocorrelation diagnostics for DOGE, ETH, LINK, and XRP baskets.

B.2.2 Energy diagnostics

Figure 15 displays the energy diagnostics across all baskets. With minimum BFMI of 0.68, the overlapping energy transition and marginal energy distributions indicate the model did not encounter geometric pathologies.

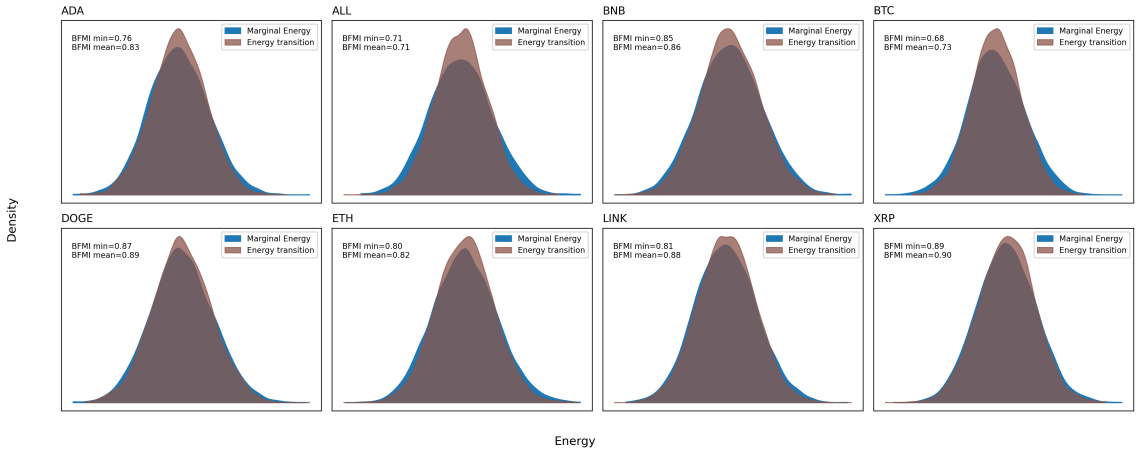


Figure 15: Energy diagnostics for all baskets.

References

Anarkulova, A., Cederburg, S., and O'Doherty, M. (2022). Stocks for the long run? evidence from a broad sample of developed markets. *Journal of Financial Economics*, 143:409–433.

Barnichon, R. and Brownlees, C. (2019). Impulse response estimation by smooth local projections. *Review of Economics and Statistics*, 101(3):522–530.

Breiman, L. (1996). Heuristics of instability and stabilization in model selection. *Annals of Statistics*, 24(6):2350–2383.

Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480.

Conlon, T., Corbet, S., and McGee, R. (2024). Enduring relief or fleeting respite? bitcoin as a hedge and safe haven for the us dollar. *Annals of Operations Research*, 337:45–73.

de Prado, M. L. (2018). *Advances in Financial Machine Learning*. John Wiley & Sons.

Elliott, G., Rothenberg, T. J., and Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica*, 64(4):813–836.

Ferreira, L. N., Miranda-Agrippino, S., and Ricco, G. (2025). Bayesian local projections. *Review of Economics and Statistics*, 107(5):1424–1438.

Geweke, J. (1993). Bayesian treatment of the independent student-t linear model. *Journal of Applied Econometrics*, 8:S19–S40.

Gkillas, K. and Katsiampa, P. (2018). An application of extreme value theory to cryptocurrencies. *Economics Letters*, 164:109–111.

Granger, C. W. J. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1(1):15–29.

Güler, D. (2023). The impact of investor sentiment on bitcoin returns and conditional volatilities during the era of covid-19. *Journal of Behavioral Finance*, 24(3):276–289.

Hall, P., Horowitz, J. L., and Jing, B.-Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika*, 82(3):561–574.

Hoffman, M. D. and Gelman, A. (2014). The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15(47):1593–1623.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, 68(1):165–176.

Huber, F., Matthes, C., and Pfarrhofer, M. (2024). General seemingly unrelated local projections. *arXiv preprint arXiv:2410.17105*.

Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review*, 95(1):161–182.

Karau, S. (2023). Monetary policy and bitcoin. *Journal of International Money and Finance*, 137:102880.

Kilian, L. and Kim, Y. J. (2011). How reliable are local projection estimators of impulse responses? *Review of Economics and Statistics*, 93(4):1460–1466.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54(1-3):159–178.

Lahiri, S. N. (2003). *Resampling Methods for Dependent Data*. Springer Series in Statistics. Springer.

Li, Y., Urquhart, A., Wang, P., and Zhang, W. (2021). Max momentum in cryptocurrency markets. *International Review of Financial Analysis*, 77:101829.

Li, Y., Zhang, W., Xiong, X., and Wang, P. (2020). Does size matter in the cryptocurrency market? *Applied Economics Letters*, 27(14):1141–1149.

Likitratcharoen, D., Kronprasert, N., Wiwattanalamphong, K., and Pinmanee, C. (2021). The accuracy of risk measurement models on bitcoin market during covid-19 pandemic. *Risks*, 9(12):222.

Ma, C., Tian, Y., Hsiao, S., and Deng, L. (2022). Monetary policy shocks and bitcoin prices. *Research in International Business and Finance*, 62:101711.

MacKinnon, J. G. and White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29(3):305–325.

Maghyereh, A. and Abdoh, H. (2021). Time–frequency quantile dependence between bitcoin and global equity markets. *The North American Journal of Economics and Finance*, 56:101355.

Meinshausen, N. and Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4):417–473.

Nouira, A. and Azencott, C.-A. (2022). Multitask group lasso for genome wide association studies in diverse populations. *Pacific Symposium on Biocomputing*, 27:163–174.

Nouira, A. and Azencott, C.-A. (2025). Sparse multitask group lasso for genome-wide association studies. *PLOS Computational Biology*, 21(9):1–27.

Nzokem, A. and Maposa, D. (2024). Bitcoin versus s&p 500 index: Return and risk analysis. *Mathematical and Computational Applications*, 29(3):44.

Osterrieder, J. and Lorenz, J. (2017). A statistical risk assessment of bitcoin and its extreme tail behavior. *Annals of Financial Economics*, 12(01):1750003.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Piironen, J. and Vehtari, A. (2017a). On the hyperprior choice for the global shrinkage parameter in the horseshoe prior. In Singh, A. and Zhu, J., editors, *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pages 905–913. PMLR.

Piironen, J. and Vehtari, A. (2017b). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics*, 11(2):5018 – 5051.

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89(428):1303–1313.

Romano, J. P. and Wolf, M. (2007). Control of generalized error rates in multiple testing. *The Annals of Statistics*, 35(4):1378 – 1408.

Röver, C. (2020). Bayesian random-effects meta-analysis using the bayesmeta R package. *Journal of Statistical Software*, 93(6):1–51.

Shah, R. D. and Samworth, R. J. (2013). Variable selection with error control: Another look at stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(1):55–80.

Sutton, A. J. and Abrams, K. R. (2001). Bayesian methods in meta-analysis and evidence synthesis. *Statistical methods in medical research*, 10(4):277–303.

Tanaka, M. (2020). Bayesian inference of local projections with roughness penalty priors. *Computational Economics*, 55(2):629–651.

Zivot, E. and Andrews, D. W. (1992). Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of Business & Economic Statistics*, 10(3):251–270.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320.