Real-Time Discovery and Tracking of Return-Based Anomalies*

Benjamin Marrow†
University of Chicago

Stefan Nagel‡
University of Chicago, NBER, CEPR, and CESIfo

October 2024

We explore the cross-sectional predictability of stock returns with lagged past returns through the lens of a hypothetical Bayesian researcher who begins with an initial prior that is neutral, showing no bias toward momentum, reversal, or other predictable patterns. This researcher considers a wide range of monthly lags as potential return predictors. By applying Gaussian process regression, which flexibly allows expected returns to depend on lagged returns, and using empirical Bayes shrinkage to guard against spurious anomaly discovery due to multiple testing and against mistaking the ex-post visible effects of investor learning as ex-ante expected returns, this researcher would have discovered prominent return-based anomalies—such as momentum and long-term reversal—well before the authors of the published studies analyzed the data. This suggests that these anomalies represented properties of ex-ante expected returns at the time of their academic discovery. However, tracking these anomalies in real-time, with posterior beliefs based on optimally weighted historical data, reveals that the ex-ante expected returns of many prominent anomalies diminish significantly around their publication dates. Based on these findings, there is little justification for viewing momentum, long-term reversal effects, or other return-based anomalies as permanent features of the cross-section of expected stock returns.

*We are grateful for comments to Rob Stambaugh, seminar participants at the Advanced Financial Technologies Laboratory at Stanford University, and participants at the SoFiE conference. We also thank Michael Yip, Zichen Zhao, and Raymon Yue for excellent research assistance.

†University of Chicago, Booth School of Business, 5807 South Woodlawn Avenue, Chicago, IL 60637, e-mail: bmarrow@chicagobooth.edu

‡University of Chicago, Booth School of Business, 5807 South Woodlawn Avenue, Chicago, IL 60637, e-mail: stefan.nagel@chicagobooth.edu

I. Introduction

Many discoveries of cross-sectional return predictability were made in empirical studies that examined only one or a few selected predictor variables at a time, using average returns across the entire sample of historical data available to the researcher as estimates of expected returns. This standard approach implicitly assumes that the expected return premia tied to these predictors are constant, that investors know the data-generating process and do not have to learn about it, and that researchers have not engaged in selective reporting (p-hacking) after having tried other, unreported predictors. Since these assumptions may not hold, the implications of these published findings for asset pricing theories and their predictive relevance for stock returns going forward remain unclear.

One could avoid these unpalatable assumptions by going back in time, rerunning the history of anomaly discovery from the beginning, starting with a clean slate, considering the entire universe of potentially relevant stock characteristics as predictors, without selection based on published results in the literature, applying methods that are robust to multiple testing and that reveal ex-ante predictability rather than in-sample predictability patterns induced by investor learning, while accommodating the possibility that anomalies may disappear and reemerge over time.

In this paper, we present an approach that approximates this ideal. To avoid selection of predictors based on published results, we focus on an entire class of predictors: each stock's history of past returns up to a lag of 120 months. Each monthly lag is a potential predictor of future returns, and we do not impose any ad-hoc combinations of different lags, unlike the classic studies of past-return-based predictors that focus on small subsets of this class, such as combinations of lags of 1-12 months in Jegadeesh (1990) and Jegadeesh and Titman (1993) or long-term returns over multiple years in De Bondt and Thaler (1985). We form 120 portfolios in which stocks in month t are weighted by their cross-sectionally demeaned lagged return in month $t-k$, scaled by the cross-sectional variance of lagged returns. We refer to these portfolios as lag portfolios.

We employ empirical Bayes methods within a Gaussian process regression framework to estimate the expected returns of the lag portfolios.[1] We model the expected return of the lag portfolios as a nonlinear function of their lag, assuming a Gaussian process prior for this expected return function. This prior reflects the beliefs a researcher might hold about expected returns before observing any data. To represent initial uncertainty regarding the nature of possible anomalies, we set the prior mean to zero for all 120 lag portfolios, indicating no predisposition toward momentum or reversal-type patterns. The Gaussian process prior allows the econometrician to entertain the prior belief that portfolios closer in lag have more highly correlated expected returns, which induces a smoothing effect when prior beliefs are combined with returns data to calculate posterior estimates. Each month, we obtain a posterior distribution for the ex-ante expected returns of the 120 portfolios. From these posterior beliefs, we then generate a predictive distribution of future returns, enabling the calculation of predictive Sharpe ratios for linear combinations of the lag portfolios.

The Gaussian process prior involves hyperparameters that control the prior variance of expected returns, which determines how strongly posterior means are shrunk towards the prior means of zero, and correlations, which determine the degree to which portfolios returns at neighboring lags are smoothed to in the calculation of the posterior. We estimate these hyperparameters using an empirical Bayes approach. Every month t, we pick prior hyperparameters that maximize the marginal likelihood of the data until month t. Roughly speaking, we look for hyperparameter values that make the data observed until t most likely under the prior.

This empirical Bayes data-driven shrinkage approach addresses two of the key issues that we outlined earlier. First, as recently highlighted in Chen and Zimmermann (2020) and Jensen, Kelly, and Pedersen (2023), it solves the multiple-testing problem. Second, Martin and Nagel (2022) show that data-driven shrinkage also addresses the investor learning problem. When investors gradually learn about the data-generating process parameters, returns will appear predictable in-sample to an econometrician analyzing the data ex-post, even though they were not predictable ex-ante for a Bayesian investor. However, if the econometrician employs a data-driven shrinkage approach, the posterior estimates accurately capture the ex-ante predictability, avoiding contamination from the

1. Gaussian process regression is used in certain machine learning applications. Some types of neural networks converge to a Gaussian process as the number of nodes approaches infinity. See Williams and Rasmussen (2006).

learning-induced in-sample predictability.[2]

A first question we then address with this method is whether a hypothetical researcher, starting with a clean slate and equipped with this Gaussian process regression approach, would have had discovered well-known anomalies by the time the authors of the published studies examined the data. If the published research were the result of p-hacking, we would not expect our hypothetical researcher's posterior beliefs to indicate evidence of the anomaly. Likewise, if the predictability documented in the published studies was only an ex-post artifact generated by investors' learning about the data-generating process, but not present ex-ante, we would not expect expect the anomaly to appear in the hypothetical researcher's posterior beliefs. On the other hand, if academic researchers, with the technology available at the time, were slow to uncover anomalies, our hypothetical researcher might identify them earlier than the actual researchers did.

For both long-term reversal and momentum effects, our findings clearly support the latter scenario. The posterior beliefs reveal evidence of return reversals for lags between 13 to 71 months as early as the beginning of the 1970s, well before the 1982 end of the sample in De Bondt and Thaler (1985). The fact that long-term reversals only occur after a 12-month gap between portfolio formation and the holding period, as noted by De Bondt and Thaler, is also reflected in our posterior estimates. Similarly, evidence for momentum at lags between 3 and 12 months appears in the posterior as early as the 1960s, well ahead of the 1987 end of the sample period in Jegadeesh and Titman (1993). Additionally, strong evidence for the short-term reversals noted by Jegadeesh (1990) and the seasonality effect identified by Heston and Sadka (2008) emerges very early on as well.

In this analysis, our hypothetical researcher used all available data from the start of the CRSP database in 1926 up to the time of the posterior calculation in month t, with equal weighting across all periods. This approach did not allow the researcher's prior beliefs to consider the possibility that anomalies could disappear or re-emerge over time. To relax this assumption, we introduce exponential weighting, which allows the posterior beliefs to adjust over time as expected returns

2. Martin and Nagel prove their result for hyperparameter choice via cross-validation while here we use a marginal likelihood criterion to estimate hyperparameters. However, as Fong and Holmes (2020) show, cross-validation and marginal likelihood approaches are closely related and equivalent under certain conditions.

evolve. This introduces an additional hyperparameter that controls how quickly data is downweighted as it becomes more distant. Each month t, we select this hyperparameter by minimizing the mean-squared error in predicting returns across 120 portfolios with the posterior mean in past data up to month t. If this method effectively tracks ex-ante expected returns, the posterior mean should be an unbiased predictor of future realized returns. This holds true for most of our sample, except in the years following the 2008 financial crisis, where the posterior means are generally close to zero, indicating minimal predictability based on past returns

With this drifting version of posterior beliefs, we can observe how predictability based on past returns fluctuates over time. We find that many prominent anomalies weaken substantially around their respective publication dates. Most notably, statistical confidence in the long-term reversals identified by De Bondt and Thaler (1985) falls below conventional levels around publication date. Although long-term reversals partially reemerge at certain lags in subsequent years, they never regain the strength observed during De Bondt and Thaler's sample period. The short-term reversals of Jegadeesh (1990) also diminish substantially around the publication date. Meanwhile, the momentum effect of Jegadeesh and Titman (1993) weakens at shorter lags, with the bulk of the effect shifting to lags of 7 to 12 months, a pattern later documented by Novy-Marx (2012).

The decay observed around the publication date echoes the findings of McLean and Pontiff (2016), who document post-publication decay for a much broader set of cross-sectional asset pricing anomalies. However, a key distinction in our analysis is that we do not compare post-publication returns to the in-sample average returns reported in the original studies. Instead, our posterior mean estimates are already adjusted via shrinkage to account for multiple testing and learning effects. The decline in posterior means around the publication date suggests a genuine reduction in ex-ante expected returns.

Behavioral theories of momentum and long-term reversals suggest a connection between the two anomalies. The models of Daniel, Hirshleifer, and Subrahmanyam (1998), Barberis, Shleifer, and Vishny (1998), and Hong and Stein (1999) are grounded in different behavioral biases, yet all predict that strong momentum should be followed by strong long-term reversals in subsequent years. If momentum disappears, long-term reversals should also fade. Our posterior mean estimates support this relationship: although the strength of momentum and reversals fluctuates, they tend to move in tandem to a significant extent.

Finally, as our method provides period-by-period posterior estimates for the expected returns of the lag portfolios, a natural question arises: Do some investors align their portfolio allocations with these posterior beliefs, while others diverge? Investors in the former group tend to mitigate market anomalies, whereas those in the latter group tend to reinforce them. Unlike in earlier work on this question, our analysis takes into account that anomalies change over time, and it captures positioning regarding momentum, reversals, and other patterns in expected returns simultaneously. Analyzing 13F holdings data by institution type, we find that mutual funds and households are the only groups with significant correlations between their portfolio weights and the posterior means. Mutual fund holdings exhibit positive correlations with the posterior means, while household sector holdings show negative correlations.

Several recent studies, like ours, revisit the anomaly discovery process by starting with a comprehensive set of candidate predictors, rather than limiting the analysis to those featured in published research. Most closely related are Chen and Dim (2023) and Li, Rossi, Yan, and Zheng (2022), who use a large set of accounting and past-return signals and also apply a data-driven shrinkage approach. However, unlike our method, they do not incorporate optimal smoothing across related strategies, track the disappearance and (re-)emergence of anomalies via optimal exponential weighting, nor generate the full predictive distribution we obtain through Gaussian process regression. These distinctions also set our work apart from Yan and Zheng (2017), who further do not use shrinkage methods. Martin and Nagel (2022) examine expected returns as a function of lagged returns using ridge regression, but their approach uses ad-hoc rolling windows instead of the optimal weighting of past data to track shifting expected returns that we develop here. Chen and Zimmermann (2020) and Jensen, Kelly, and Pedersen (2023) employ empirical Bayes methods, but they focus solely on published anomalies. Similarly, Filipović and Pasricha (2022) use Gaussian process regression to estimate expected returns based on stock characteristics, but rely on characteristics from published studies.

II. Approach

Consider an unbalanced panel of of stocks with returns $R_{i,t}$ across time periods $t = 1, 2, ..., T$ and stocks $i = 1, 2, ..., N_t$. Let $\bar{R}_t$ denote the mean return of all stocks in cross-section t and $X_{i,t} = R_{i,t} - \bar{R}_t$ the return of stock i in excess of the cross-sectional mean. We stack the N_t cross-sectional observations of $X_{i,t}$ into the vector $\boldsymbol{x}_t$. A linear cross-sectional projection of $X_{i,t}$ on its on lags yields

$$X_{i,t} = b_1 X_{i,t-1} + b_2 X_{i,t-2} + ... + b_K X_{i,t-K} + \varepsilon_{i,t}. \tag{1}$$

and we collect $\boldsymbol{b} = (b_1, b_2, ..., b_K)'$. This projection captures the expected returns conditional on past returns and a linear functional form.

II.A. Portfolio aggregation

We now show that we can, approximately, assess the magnitudes of the coefficients $\boldsymbol{b}$ with a portfolio approach. Aggregating to portfolios comes with the advantage of obtaining a balanced panel. All of our empirical analysis will be based on these portfolios.

Specifically, we form portfolios with weight vector

$$\boldsymbol{w}_{k,t} = \boldsymbol{x}_{t-k}(\boldsymbol{x}'_{t-k}\boldsymbol{x}_{t-k})^{-1}, \tag{2}$$

where the weights are formed based on returns with lag k. We refer to the resulting portfolio as the lag k portfolio, and it has portfolio returns

$$F_{k,t} = \boldsymbol{x}'_t \boldsymbol{w}_t(k). \tag{3}$$

From the specification of the weights in (2) one can see that these portfolio returns are also the slope in a cross-sectional regression of $X_{i,t}$ on $X_{i,t-k}$. Since returns have very small autocorrelation, $F_{k,t}$ is therefore also approximately equal to the k-th cross-sectional regression slope factor in a cross-sectional regression of $X_{i,t}$ jointly on its lags for $k = 1, 2, ..., K$. Moreover, substituting (1) into (3) and taking expectations then yields

$$\mathbb{E}[F_{k,t}] \approx b_k. \tag{4}$$

In our empirical analysis, our objective is to estimate the $\mathbb{E}[F_{k,t}]$ for $k = 1, 2, ..., K$. This relationship shows that these expected returns then approximately tell us about the linear projection coefficients in (1).

II.B. Gaussian process regression

We estimate expected returns of the lag portfolios using a Gaussian process regression approach. The Gaussian process regression allows us to specify sensible prior beliefs that, with empirical Bayes methods, induce shrinkage that addresses multiple testing and investor learning, and it allows us to construct a predictive distribution of future returns each period.

The first step is to think of the expected portfolio returns as a function of the lag, $\mu(k) = \mathbb{E}[F_{k,t}]$. For example, if there is momentum at short lags and reversals at long lags, $\mu(k)$ is positive for small k and negative for large k. We can then write the lag portfolio return as

$$F_{k,t} = \mu(k) + \eta_{k,t}, \qquad \eta_{k,t} \sim \mathcal{N}(0, \sigma^2), \tag{5}$$

where $\text{cov}(\eta_{k,t}, \eta_{k',t}) = 0$ for $k \neq k'$. This is an empirically plausible assumption: weighting stocks by lagged returns at different lags k and k' uses two almost uncorrelated weight vectors. There is therefore no reason why a long-short portfolio with weights based on returns at lags k and k' should have correlated returns. Our interest centers on estimating the expected return function $\mu(k)$. Gaussian process regression aims to uncover $\mu(k)$ from observed data on $F_{k,t}$ by smoothing out the contaminating noise.

The second step is to specify prior beliefs. We assume that any pair $\mu(k)$ and $\mu(k')$ are, a priori, jointly Gaussian with mean zero and covariance $\psi(k, k')$. This means that $\mu(k)$ follows a Gaussian process

$$\mu(k) \sim \mathcal{GP}\left(0, \psi(k, k')\right). \tag{6}$$

These prior beliefs express the view that, before looking at the data, it is not known whether the expected return of a portfolio corresponding to lag k is positive or negative. In other words, we consider a Bayesian observer who approaches the realized return data without a prior view whether, say, momentum or reversal patterns are more likely to be present. Prior variances capture, roughly speaking, the range above and below zero in which $\mu(k)$ is likely to be. Prior correlations express views about how expected returns at different lags k and k' are related.

The specification of $\psi(k, k')$ is important for how the posterior will look like. The lower the prior variance, the more shrinkage there will be in the posterior towards the prior mean of zero. The higher the correlation between expected returns at lags k and k', the more the posterior of $\mu(k)$ will also draw on the returns of the lag k' in the calculation of the posterior. Hence, higher correlation means more cross-sectional smoothing in the estimation of the posterior beliefs.

Prior beliefs are supposed to reflect what is plausible before seeing the data. We work with the following assumptions:

(i) At the first lag, $\mu(1)$ may be quite different from the other prior means because unlike portfolios at longer lags, it may partly pick up microstructure-induced return predictability that shows up only at relatively high frequency. The prior variance for $k = 1$ should therefore be allowed to differ from the prior variances for $k > 1$.

(ii) Seasonality may exist. In particular, given the prevalence of annual rebalancing, annual tax years, etc., seasonality in annual frequencies may be particularly likely. To allow for such seasonality, the prior variances at lags 12, 24, ... , and covariances that involve these lags, should be allowed to differ from variances and covariances at other lags.

(iii) Expected returns of portfolios of neighboring lags may be correlated. For example, if there is a momentum effect such that the portfolio at lag k has a positive expected return, then it seems plausible that a portfolio at lag k' close to k is also somewhat likely to have a positive expected return. For example, if momentum arises from underreaction to news, the catch-up of prices to past news may be spread out over a few months and hence several portfolios with neighboring lags will have positive expected returns.

The following specification of the prior covariance function captures all three properties:

$$\psi(k, k') = \sigma_m^2 \times \mathbb{1}_{k=1 \cap k'=k} + \sigma_s^2 \times \mathbb{1}_{k \in \{12,24,\ldots\} \cap k'=k} + \kappa(k, k') \tag{7}$$

Here σ_m^2 is the extra variance for microstructure-induced predictability at the first lag, σ_s^2 is the variance component allowing for seasonal effects, and the component $\kappa(k, k')$ is a kernel that allows covariances to depend on the distance between k and k'. We work with a Matérn 3/2 kernel

$$\kappa(k, k') = \xi^2 (1 + |u|) \exp(-|u|), \qquad u = \frac{k - k'}{\tau}, \tag{8}$$

which is common in machine learning applications of Gaussian process regression (Williams and Rasmussen 2006). The hyperparameter τ controls the bandwidth of the kernel and ξ controls the magnitude of variances and covariances.

Allowing for an extra variance component at the first lag and the seasonal lags is admittedly injecting some ex-post knowledge that expected returns of these lag portfolios could be special relative to those at neighboring lags. However, it is important to keep in mind that we are not imposing that the expected returns at these lags are are different from neighboring lags, as the prior mean is always zero for all of them. We are further not imposing that σ_m and σ_s are actually greater than zero, we are just allowing for the possibility that they are greater than zero. It will be up to the data-driven hyperparameter estimation to find the optimal values for σ_m and σ_s. Zero is a possible value that the data-driven approach could deliver.

Figure I shows 10 draws from the prior distribution. It illustrates the key properties of the prior. First, the lower the distance between lags k and k', the higher the correlation of the portfolio expected returns $\mu(k)$ and $\mu(k')$. Second, there is extra variability at lags one and the seasonal lags that are multiples of 12. At these lags, the prior entertains the possibility that expected returns could deviate more from the prior mean of zero than at the other lags.

Given the Gaussian prior and the Gaussian innovations in (5), the posterior distribution is also Gaussian. Let $\mathbf{\Sigma}$ denote the $K \times K$ covariance matrix of K stacked innovations in (5) of portfolio returns for lags $k = 1, 2, ..., K$. Collecting the average portfolio returns, $\bar{F}_{t,k} = \frac{1}{t}\sum_{j=1}^{t} F_{k,j}$ for lags

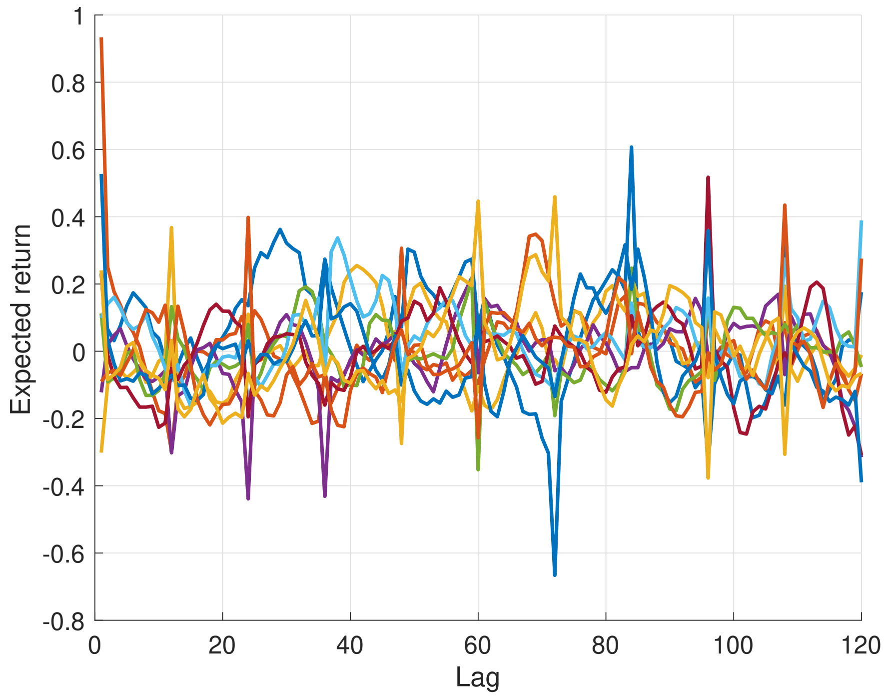


FIGURE I
Prior distribution: 10 draws

$K = 120$, $\xi = 0.1$, $\tau = 2$, $\sigma_m = 0.3$, $\sigma_s = 0.2$.

$k = 1, 2, ..., K$ in the vector $\bar{\boldsymbol{f}}_t$, all corresponding $\mu(k)$ in the vector $\boldsymbol{\mu}$, and all $\psi(k, k')$ in the $K \times K$ covariance matrix $\boldsymbol{\Psi}$, the time-t posterior distribution of $\boldsymbol{\mu}$ is

$$\boldsymbol{\mu} | \bar{\boldsymbol{f}}_t \sim \mathcal{N}(\boldsymbol{m}_t, \boldsymbol{V}_t), \tag{9}$$

with

$$\begin{aligned} \boldsymbol{m}_t &= \boldsymbol{\Psi} \left(\boldsymbol{\Psi} + \frac{1}{t} \boldsymbol{\Sigma} \right)^{-1} \bar{\boldsymbol{f}}_t, \\ &= \left(\boldsymbol{I} + \frac{1}{t} \boldsymbol{\Sigma} \boldsymbol{\Psi}^{-1} \right)^{-1} \bar{\boldsymbol{f}}_t, \end{aligned} \tag{10}$$

and

$$\boldsymbol{V}_t = \boldsymbol{\Psi} - \boldsymbol{\Psi} \left(\boldsymbol{\Psi} + \frac{1}{t} \boldsymbol{\Sigma} \right)^{-1} \boldsymbol{\Psi}. \tag{11}$$

The posterior mean in (10) shrinks the observed average returns $\bar{\boldsymbol{f}}_t$ towards zero. How strong shrinkage is depends on the prior variance hyperparameters ξ, σ_s, and σ_m in (7) and (8). Lower prior variances induce less shrinkage. How much neighboring observations are smoothed is controlled by the hyperparameter τ in (8).

II.C. Predictive distribution and Sharpe ratio

The time-t predictive distribution for returns in period $t + 1$ is Gaussian with means $\boldsymbol{m}_t$ and covariances

$$\boldsymbol{\Phi}_t = \boldsymbol{V}_t + \boldsymbol{\Sigma}. \tag{12}$$

These covariances capture uncertainty from two sources: the uncertainty due to unpredictable noise in returns given $\mu(k)$ and the uncertainty about $\mu(k)$ as expressed in $\boldsymbol{V}_t$. Based on the predictive distribution, we can calculate the maximum (annualized) Sharpe ratio achievable from the portfolios (according to the predictive distribution):

$$\sqrt{12}\sqrt{\boldsymbol{m}_t'\boldsymbol{\Phi}_t^{-1}\boldsymbol{m}_t} \tag{13}$$

We can further use the means and covariances of the predictive distribution to construct an ex-ante mean-variance efficient portfolio, and then calculate its realized out-of-sample return as

$$r_{mve,t+1} = \boldsymbol{m}_t'\boldsymbol{\Phi}_t^{-1}\boldsymbol{f}_{t+1}. \tag{14}$$

We use 10-year moving averages of the mean and standard deviation of monthly observations of $r_{mve,t+1}$ to estimate an out-of-sample Sharpe ratio (that we then annualize).

II.D. Estimation of hyperparameters

We use an empirical Bayes approach to estimate the four hyperparameters $\boldsymbol{\theta} = (\xi, \tau, \sigma_m, \sigma_s)$ by maximizing the marginal likelihood. The marginal likelihood $p(\bar{\boldsymbol{f}}_t|\boldsymbol{\theta})$ integrates out the latent $\mu(k)$ from the likelihood. Roughly speaking, maximizing the marginal likelihood looks for values of $\boldsymbol{\theta}$ that make the actually observed data $\bar{\boldsymbol{f}}_t$ most likely. Since the per-period portfolio returns $\bar{\boldsymbol{f}}_j$ are Gaussian, the average realized return until time t is also Gaussian and one can express the marginal likelihood based on a multivariate normal distribution with mean zero and covariance matrix $\boldsymbol{\Omega}_t = \boldsymbol{\Psi} + \frac{1}{t}\boldsymbol{\Sigma}$. Hence, the log marginal likelihood of the average realized return vector is

$$\log p(\bar{\boldsymbol{f}}_t|\boldsymbol{\theta}) = -\frac{\bar{\boldsymbol{f}}_t'\boldsymbol{\Omega}_t^{-1}\bar{\boldsymbol{f}}_t}{2} - \frac{\log\det\boldsymbol{\Omega}_t}{2} - K\frac{\log 2\pi}{2}. \tag{15}$$

To compute the log marginal likelihood, we use a plug-in approach, replacing $\boldsymbol{\Sigma}$ in $\boldsymbol{\Omega}_t$ and in $\boldsymbol{V}_t$ with a restricted version of the sample covariance matrix of $\boldsymbol{f}_j$ in the sample $j \leq t$, denoted $\widehat{\boldsymbol{\Sigma}}_t$ and further described below.

Every period t, we repeat the hyperparameter estimation based on updated data $\bar{\boldsymbol{f}}_t$ and $\widehat{\boldsymbol{\Sigma}}_t$ to get a new estimate $\hat{\boldsymbol{\theta}}_t$. We use $\hat{\boldsymbol{\theta}}_t$, again with $\widehat{\boldsymbol{\Sigma}}_t$, to calculate posterior and predictive distribution moments in (10), (11), and (12).

II.E. Estimating the covariance matrix of innovations

In the empirical analysis, we work with $K = 120$ and so the covariance matrix $\boldsymbol{\Sigma}$ is very large. Estimating it with the sample covariance matrix without further restrictions or regularization would not deliver well-behaved estimates. For this reason, we impose the following structure,

$$\boldsymbol{\Sigma} = \boldsymbol{HCH}, \tag{16}$$

where $\boldsymbol{H}$ is a diagonal matrix with standard deviations on the diagonal and $\boldsymbol{C}$ is a correlation matrix with the following banded Toeplitz structure

$$\boldsymbol{C} = \begin{pmatrix} 1 & \rho_1 & \rho_2 & 0 & 0 & .... & 0 \\ \rho_1 & 1 & \rho_1 & \rho_2 & 0 & ... & 0 \\ & ... & & & & & \\ 0 & ... & 0 & \rho_2 & \rho_1 & 1 & \rho_1 \\ 0 & ... & 0 & 0 & \rho_2 & \rho_1 & 1 \end{pmatrix}. \tag{17}$$

Empirically, this banded pattern is evident in the sample covariance matrices of the $\boldsymbol{f}_t$. This structure likely arises because stocks that experienced positive (or negative) shocks at time $t-k$ tend to be similar to those that experienced similar shocks at $t-k-1$. This could reflect a delayed reaction of certain stocks to common factor shocks. For instance, when the market rises, most stocks follow suit, but some may respond with a one-month lag.

To construct the estimate $\widehat{\boldsymbol{\Sigma}}_t$ of $\boldsymbol{\Sigma}$, we plug in the sample standard deviation

$$\hat{\sigma}_{t,k} = \left(\frac{1}{t-1} \sum_{s=1}^{t} (r_{s,k} - \bar{r}_{t,k})^2 \right)^{\frac{1}{2}}, \tag{18}$$

for the k-th standard deviations on the diagonal of $\boldsymbol{H}$. We estimate the correlations ρ_1 and ρ_2 as averages of sample correlations of portfolios one or two lags apart:

$$\hat{\rho}_1 = \frac{1}{(K-1)(t-1)} \sum_{k=1}^{K-1} \frac{1}{\hat{\sigma}_{t,k}\hat{\sigma}_{t,k+1}} \sum_{s=1}^{t} (r_{s,k} - \bar{r}_{t,k})(r_{s,k+1} - \bar{r}_{t,k+1})$$
$$\hat{\rho}_2 = \frac{1}{(K-2)(t-1)} \sum_{k=1}^{K-2} \frac{1}{\hat{\sigma}_{t,k}\hat{\sigma}_{t,k+2}} \sum_{s=1}^{t} (r_{s,k} - \bar{r}_{t,k})(r_{s,k+2} - \bar{r}_{t,k+2}). \tag{19}$$

II.F. Allowing for drift in expected returns

So far the approach is based on the assumption that the expected return function $\mu(k)$ is constant. But if anomalies disappear, e.g. because investors are learning, and possibly reappear, if investors forget, $\mu(k)$ will be drifting over time. To allow such drift, we compute the the sample averages $\bar{f}_{t,k}$ with exponential weighting

$$\bar{f}_{t,k} = \frac{\sum_{s=0}^{t-1} \nu(s) f_{t-s,k}}{\sum_{s=0}^{t-1} \nu(s)}, \qquad \nu(s) = e^{-\delta s}. \tag{20}$$

In this case, we now have an additional hyperparameter δ that controls how strongly data in the past gets downweighted. The sample covariance matrix of returns that we use as plug-in estimate for $\mathbf{\Sigma}$ is still calculated with equal weights, i.e., we focus on drift in means, not covariances.

All other calculations, including the marginal likelihood calculation remain the same, but with average portfolio returns based on exponentially-weighted data. For this reason, in places in the earlier formulas where $\mathbf{\Sigma}$ is pre-multiplied with $\frac{1}{t}$, this pre-multiplication factor must be replaced with

$$\sum_{j=0}^{t-1} \left(\frac{\nu(j)}{\sum_{j=1}^{t-1} \nu(j)} \right)^2 . \tag{21}$$

Given a value for δ, the marginal likelihood in (15) can be used as before to estimate the hyperparameters θ each period. However, we cannot use this marginal likelihood to infer the optimal δ because the drift in $\mu(k)$ that motivates the exponential weighting is not explicitly built into the model.[3] To keep the model parsimonious, we choose a heuristic approach where we

3. To incorporate the drift into the model, we would have to allow for permanent shocks to $\mu(k)$, obtain the resulting likelihood, integrate over the drifting latent $\mu(k)$ to obtain the marginal likelihood, and then maximize with

optimize δ with respect to past prediction error. Every period t, we check which value of δ would have produced the lowest prediction errors in periods leading up to t as measured by the forecast mean-squared error (MSE)

$$\hat{\delta}_t = \arg\min_{\delta} \sum_{j=1}^{t-1} (\boldsymbol{f}_{j+1} - \boldsymbol{m}_j(\delta))'(\boldsymbol{f}_{j+1} - \boldsymbol{m}_j(\delta)). \tag{22}$$

Based on $\hat{\delta}_t$, we then obtain the posterior estimates $\boldsymbol{m}_t(\hat{\delta}_t)$ and $\boldsymbol{V}_t(\hat{\delta}_t)$ and the corresponding predictive distribution and Sharpe ratios following the calculations outlined earlier.

II.G. Data

To build our sample, we start with all NYSE/AMEX/Nasdaq common stocks in the monthly CRSP stock file from the beginning of 1926 to end of 2023, excluding stocks with price less than \$1 and market capitalization lower than the 20th NYSE size percentile at the end of the previous month.

We then form the lag portfolios and calculate their returns as in (3). Since we need 120 months of lagged returns to form lag portfolios, we start at the beginning of 1936. Figure II shows the average returns of the 120 lag portfolios from 1936 to 2023. The figure shows the familiar patterns from the existing literature: short-term reversals at a one-month lag, momentum up to lag 12, and reversals at longer lags, interrupted by positive lag portfolio returns at lags that are multiples of 12.

III. Posterior under the assumption of constant expected returns

In the first part of our analysis, we take the perspective of an econometrician who believes that expected returns of the lag portfolios are constant. At every point in time, the econometrician then uses the Gaussian process regression approach that we laid out above to calculate the posterior distribution of the expected returns of the lag portfolios. As all the inputs to the calculation of

respect to θ and the parameter that controls the variance of the shocks to $\mu(k)$.

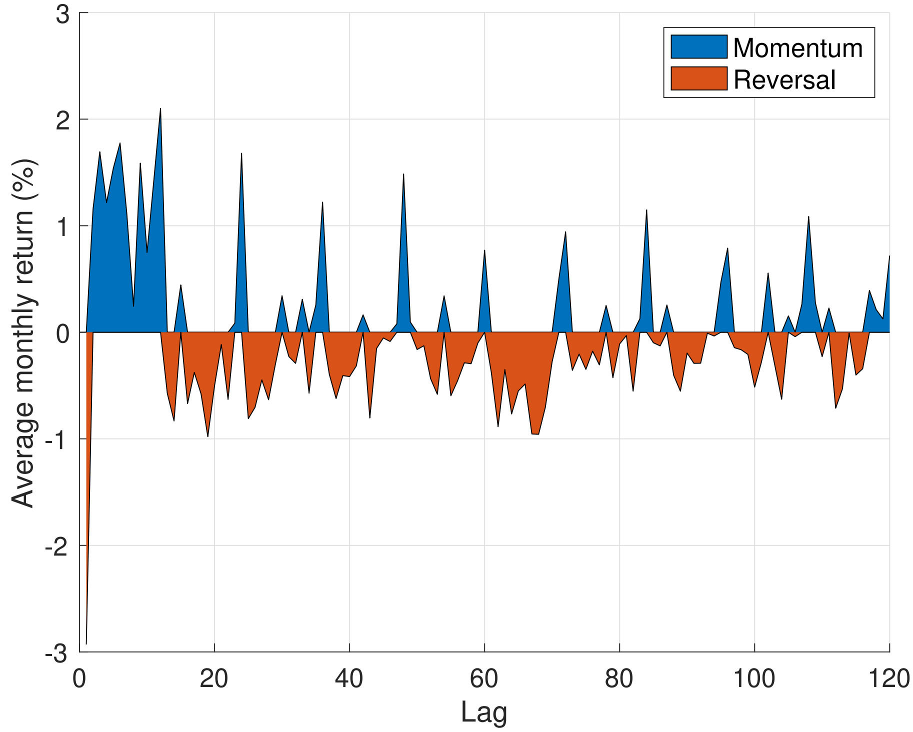


FIGURE II
Full-sample average monthly returns of the 120 lag portfolios

the posterior mean would be available in real time to an investor, and since we did not impose a selection of which lags of past returns may matter, evidence of non-zero expected returns in the posterior distribution represents a real-time discovery of a return-based anomaly.

Our aim with this first analysis is to compare how the evidence for momentum, reversals, and other patterns that appeared in published studies compares with the evidence that emerges from the posterior distribution over time. If the results in the published studies were due to p-hacking or if they picked up predictability that appears ex-post due to investor learning, but was not evident to investors in real time ex-ante, the posterior means, based on optimal data-driven shrinkage, would not show evidence that expected returns differ from zero. On the other hand, if ex-ante predictability existed but researchers were slow to recognize the predictability, the posterior distribution may show evidence of non-zero lag portfolio expected returns before the academic studies were published.

Figure III presents the posterior distribution in terms of posterior t-statistics, defined as the

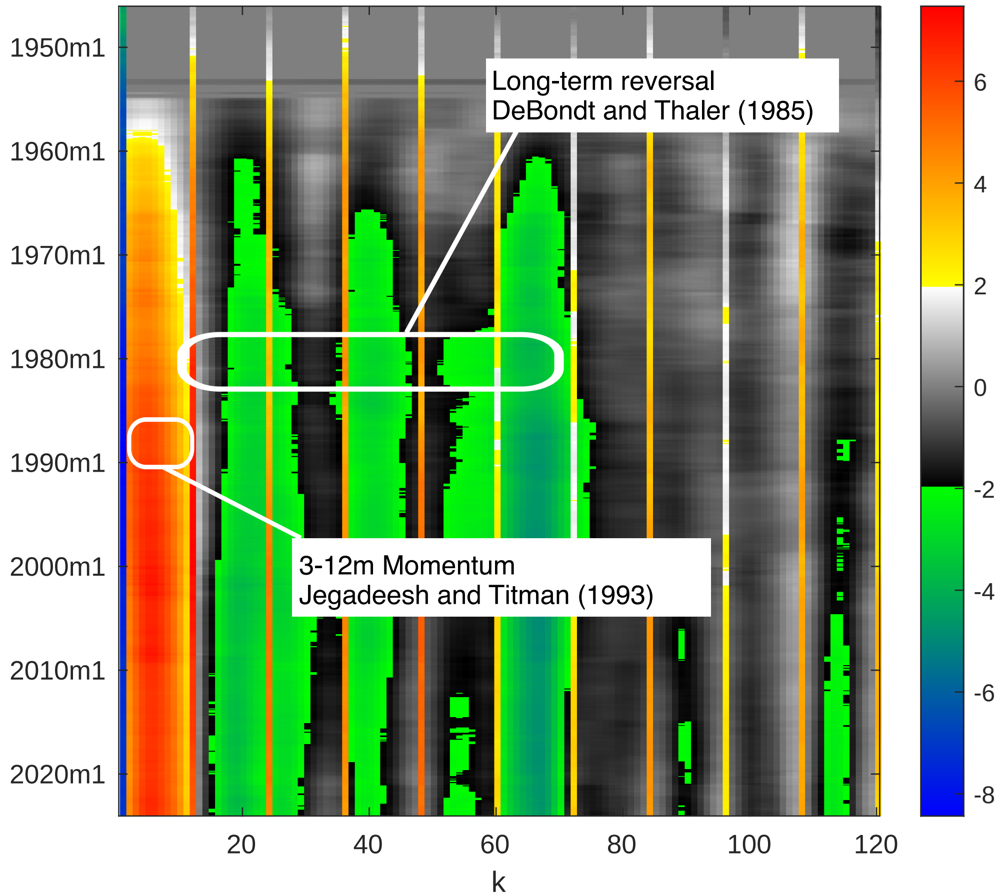


FIGURE III
Posterior t-statistics under the assumption of constant expected returns of lag portfolios: Equally weighted data

ratio of the posterior mean to the posterior standard deviation of $\mu(k)$.[4] Yellow to red color shows positive posterior t-statistics above 1.96, i.e., momentum-type patterns; green to blue color shows negative posterior t-statistics below -1.96, i.e. reversal type patterns. For reference, we also marked the end of the sample periods, and the range of lags of past returns considered in the two perhaps most well-known published studies of return-based anomalies, De Bondt and Thaler (1985) and Jegadeesh and Titman (1993).

4. Appendix B reports the time-series of hyperparameter estimates for the constant expected returns case. We discuss hyperparameter estimates in more detail when we turn to the drifting expected returns case.

The figure shows that for both the 2-12 month lag momentum effect and the long-term reversal effect, the evidence was already apparent in the posterior distribution many years before the end of the sample periods in the seminal academic studies. Evidence of long-term reversals had already started to appear in the late 1960s, and around the same time the momentum effect also becomes apparent in the posterior distribution.

Similarly, the one-month reversal effect of Jegadeesh (1990) is already apparent very early on in the 1950s. The seasonality effect of Heston and Sadka (2008) also shows up by around 1960. The positive posterior means for the portfolios at multiple-of-12 lags represent a sharp break from the reversal pattern that dominate their neighboring lags. If the prior in (7) did not allow for non-zero σ_s, the posterior would smooth over the multiple-of-12 lags and their neighbors, confounding the seasonality effect with the long-term reversals.

Thus, our real-time discovery approach affirms that these anomalies identified in the academic literature are not an artifact of p-hacking, because we start with an entire class of linear past return predictors, without ad-hoc selection of certain subsets, and we use empirical Bayes methods that address the problem of multiple testing. Our results also show that the return predictability patterns identified in the published research are, at least to a large extent, also ex-ante predictability, and not only ex-post predictability that was not exploitable for investors ex-ante as it can arise from investor learning (Martin and Nagel 2022), because our empirical Bayes shrinkage approach takes care of that as well in the construction of the posterior distribution.

IV. Posterior under the assumption of drifting expected returns

Our analysis thus far has not considered the possibility that expected returns may drift over time. There is little economic justification for assuming constant expected returns. For instance, if arbitrageurs are slower than a Bayesian econometrician in identifying return predictability, anomalies may persist temporarily and be captured in our estimated posterior distribution before eventually fading. Similarly, anomalies may emerge or re-emerge and persist for some time after shifts in

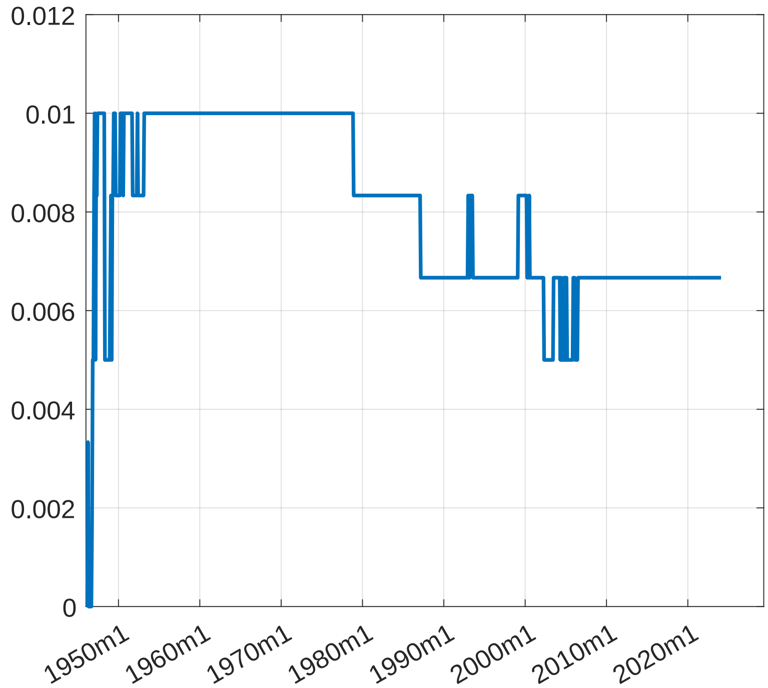
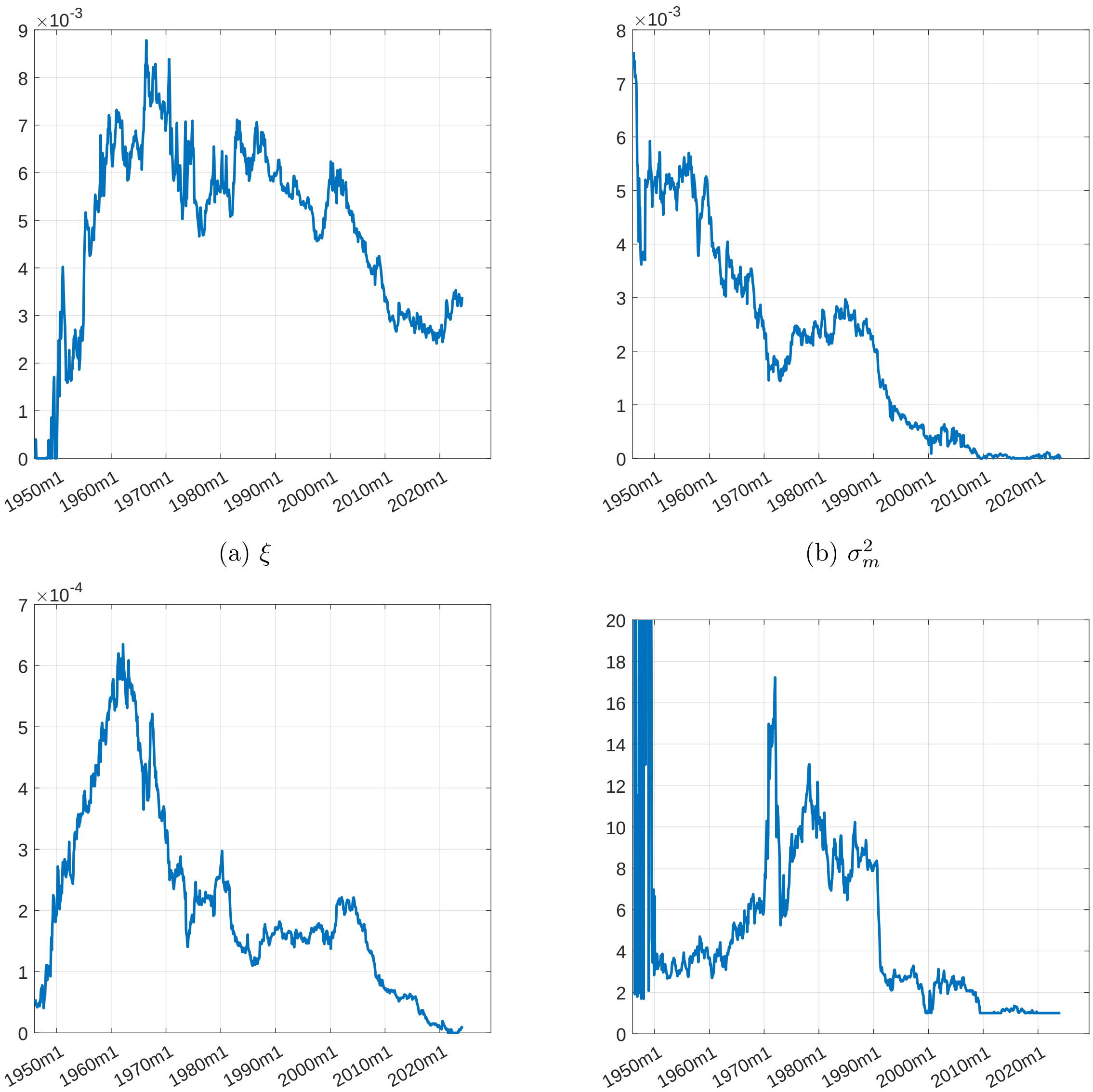


FIGURE IV
Estimates of exponential smoothing hyperparameter δ

uninformed asset demand create new types of mispricing if arbitrageurs take time to adjust.

To allow for drift in $\mu(k)$, we now employ the exponential weighting approach from Section II.F. Figure IV shows the estimates, every period, for the optimal exponential smoothing parameter δ. After the first few initial years, the estimates are quite stable around 0.008, which corresponds to a half-life of the resulting weights on past data of about 7.5 years. In the later part of the sample, the optimal value is slightly lower, which indicates that it is more beneficial to focus on somewhat more recent data in that part of the sample.

Figure V shows estimates for the other four hyperparameters. The upper-left plot shows that ξ first rises strongly in the early part of the sample, before beginning a long decay. This decay means that it has become increasingly beneficial, for the purpose of obtaining good out-of-sample forecasts of returns, to take a skeptical view about the strength of return-based anomalies. The upper-right plot shows that σ_m^2 has declined drastically over the past decades, reaching values very close to zero in the 2000s. This presumably reflects improvements in market liquidity and the emergence of



(a) ξ (b) σ_m^2

(c) σ_s^2 (d) τ

Figure V

Hyperparameter estimates: Exponentially weighted data

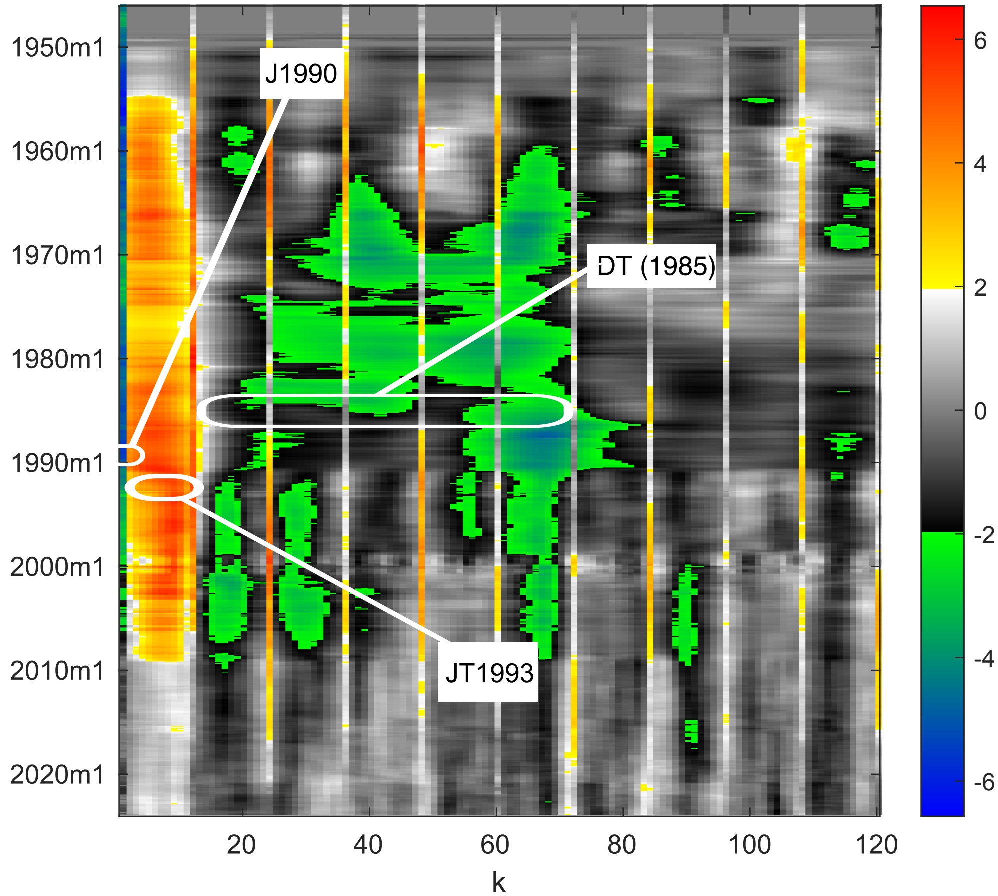


FIGURE VI
Posterior t-statistics under the assumption of drifting expected returns of lag portfolios: Exponentially weighted data

statistical arbitrageurs that eliminated the low-hanging fruit of short-term reversals. Our empirical Bayes approach also suggests that the momentum seasonality effect has basically disappeared in the past couple of decades, as shown by the declining σ_s^2 the lower-left corner. Finally, the cross-sectional smoothing parameter is erratic in the early years of the sample, but then it stabilizes around 4, and then steadily increases before dropping back to values around 2 in the early 1990s. With the kernel in (8), a value of 2 implies that lag portfolio k' in the neighborhood of lag portfolio k gets about half the weight of lag portfolio k when $|k - k'| \approx 3$ in the kernel. So the posterior mean will effectively smooth the lag portfolio returns with a few lags to the right and left.

Figure VI shows the results based on these hyperparameter choices. The picture looks quite different from the earlier one in Figure III. One-month reversals, 2-12 month momentum, long-term reversals, and the seasonality effect also appear with this version of our method that tracks a drifting $\mu(k)$, but this figure is revealing that there are interesting changes over time in the strength of these anomalies.

As a point of reference, the figure also shows the publication dates of the seminal studies of return-based anomalies, as well as the range of lags that the main analysis in these studies emphasizes. As the figure shows, the De Bondt and Thaler (1985) long-term reversal effect weakened to the point that the posterior means in the range of lags from 13 to 59 months dropped below two posterior standard deviations around the time of publication. In the years that followed, reversals re-appeared to some degree at shorter lags from 13 to 35, and they persisted at lags 61 to 71, but the bulk of the effect described by De Bondt and Thaler had disappeared. Interestingly, De Bondt and Thaler's results are often described as showing reversals for returns over the past three to five years,[5] and the re-appearance is concentrated in lags that our outside of this past three- to five-year window.

The one-month reversal effect of Jegadeesh (1990) also weakened around publication date, but even after the weakening, the posterior t-statistic of the lag 1 portfolio still stayed above 1.96. For the 2-12 month momentum effect of Jegadeesh and Titman (1993) the weakening around publication date only happened at shorter lags, making the momentum effect concentrated at lags 7 to 12, as later documented by Novy-Marx (2012). The analysis here shows, though, that the concentration of the effect in lags 7 to 12 is not a general phenomenon that applied in all periods, but a property of relatively recent data after the publication of Jegadeesh and Titman (1993). Around 2008, all anomalies mostly disappeared, including the seasonality effect described in Heston and Sadka (2008) at lags that are multiples of 12.

The result that anomalies become weaker around publication date shares similarities with the

5. See, e.g., the introduction of Jegadeesh and Titman (1993). Regarding the persistence of the reversals at lags > 60 months, it is interesting that the maximum formation period length considered in De Bondt and Thaler's analysis is 60 months. That reversals also existed in their data at longer lags is a bit more subtle to see because this requires examination of the holding period returns over different horizons reported in their paper.

findings in McLean and Pontiff (2016) that the return predictability associated with published return predictors weakens substantially out-of-sample, after the end of the original study's sample period, and then some more after the publication date of these studies, relative to return predictability during the original study's sample period. For comparison with our results, it is important to note that we are looking at the change around publication date in the posterior t-statistics, not the decay relative to the sample average returns in the original studies' sample. The posterior mean is shrunk from sample average returns towards zero due to the influence of prior beliefs. Effects of data mining and Bayesian investor learning are already eliminated by this shrinkage. Therefore, the relevant comparison in McLean and Pontiff's study is not the decay from the original study sample average to post-publication performance, but the decay from out-of-sample pre-publication performance to post-publication performance, which is roughly 50%.[6] This magnitude of decay is roughly also what we see for the posterior t-statistics in Figure VI for one-month reversals and for long-term reversals. In contrast, the decay for momentum around the publication date is much weaker.

Our results here show that the posterior distribution delivered by our method based on optimal exponential weighting can track the decay of individual anomalies. This is not really possible with an analysis of realized returns that McLean and Pontiff (2016) rely on. That said, our papers share the bottom line conclusion that the decay reflects a decline in ex-ante expected returns and is not the artifact of researcher p-hacking or Bayesian learning by investors. The results in their paper and ours are consistent with arbitrageurs learning more slowly than our idealized Bayesian econometrician equipped with the Gaussian process regression method and with unlimited access to historical data would, but once the effect becomes prominent around the publication date of the academic studies, decay sets in.

A remarkable fact emerging from Figure VI is that there is not much evidence for any other return-based anomalies other than the ones that became prominent in the academic literature. For example, there is not much indication that there are any long-term reversals at lags greater than

6. McLean and Pontiff find that post-publication mean returns are about 42% of in-sample mean returns, which is a bit more than half of the out-of-sample pre-publication mean returns, which are 74% of in-sample mean returns.

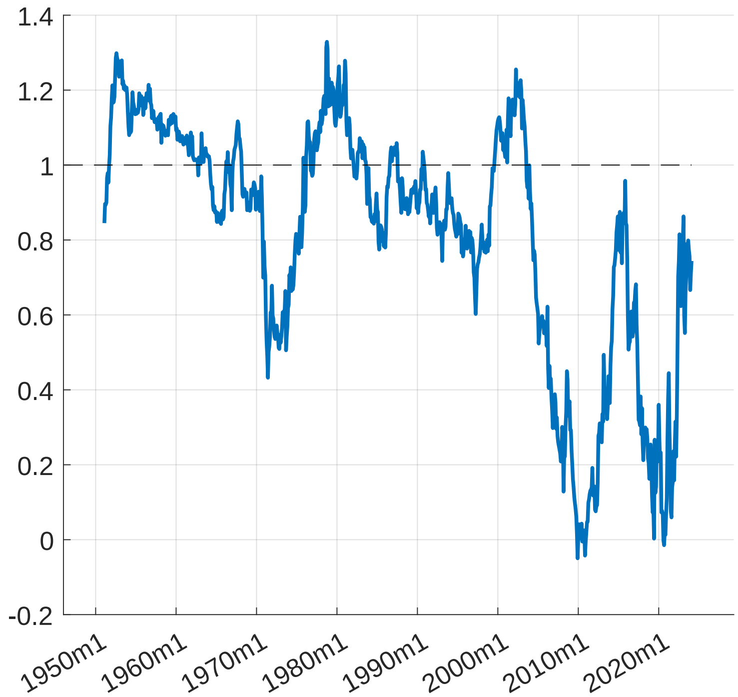


FIGURE VII
Slope in cross-sectional regression of realized returns on posterior means, 5-year moving average

72 months. Similarly, there is no evidence for any momentum effects beyond lag 12. Overall, it looks like researchers studying return-based anomalies have not missed any major predictability patterns that appear in the data.

So far we have treated the posterior distribution as representing the views of ex-ante expected returns that a Bayesian investor would hold. It would be useful to check whether this interpretation is actually correct. One way of doing this is to compare the posterior means with future realized returns of the lag portfolios. Each month, we run cross-sectional regression of month $t+1$ returns of the 120 lag portfolios on the posterior means at the end of month t. If the posterior distribution correctly represents ex-ante expected returns, we should find a slope coefficient equal to unity. If instead the slope coefficient is smaller than unity, this is an indication that shrinkage towards the prior mean of zero was insufficient and posterior means overstate ex-ante expected returns; if the slope coefficient is bigger than unity, this is an indication that shrinkage was too strong and posterior means understate ex-ante expected returns.

TABLE I

Regressions of realized returns on the posterior mean expected returns of the MVE portfolio

Every month t, we calculate the mean-variance efficient portfolio of all lag portfolios based on the weights implied by the predictive distribution at t. We then take the realized returns in future months and regress them on the posterior mean expected return of the MVE portfolio at time t.

	Intercept	Slope	R^2
t+1	-0.00	1.36	0.04
(s.e.)	(0.00)	(0.22)	
t+2	-0.00	1.29	0.04
(s.e.)	(0.00)	(0.22)	
t+3	-0.00	1.30	0.04
(s.e.)	(0.00)	(0.22)	
t+4	-0.00	1.25	0.03
(s.e.)	(0.00)	(0.22)	
t+5	-0.00	1.22	0.03
(s.e.)	(0.00)	(0.22)	
t+6	-0.00	1.23	0.03
(s.e.)	(0.00)	(0.22)	

Figure VII shows the time series of cross-sectional regression slope coefficients. While they fluctuate somewhat, their average value is close to unity until the early 2000s. This shows that throughout most of the sample, the posterior means indeed provide a good characterization of ex-ante expected returns. From the early 2000s onwards, however, the average slope coefficient drops to around 0.5. In this late part of the sample, the explanatory variables in the cross-sectional regressions have very little variation, though. As VI showed earlier, in these years the posterior t-statistics are mostly below 1.96 in absolute value, and there is little cross-sectional variation in posterior t-statistics. This makes it hard to pin down the slope coefficient in the cross-sectional regressions.

The Gaussian process regression approach not only delivers posterior means, but also a whole predictive distribution of future lag portfolio returns. Based on this predictive distribution, we can construct, every month t, a mean-variance efficient (MVE) portfolio that combines the 120 lag portfolios into one.

Table I examines whether the posterior expected return of the MVE portfolio implied by the posterior distribution of the lag portfolio expected returns at the end of month t provides a good forecast of the actually realized returns of this MVE portfolio. We regress the realized return in future months on the posterior expected return of the MVE portfolio. As the table shows, looking up to six months ahead, we obtain a slope coefficient that is a little bigger than unity at all horizons, but it is never as much as two standard errors bigger than unity. The intercept is very close to zero. This means that the posterior mean provides a good characterization of the expected return of the MVE portfolio.

The predictive distribution also implies a Sharpe ratio that this MVE portfolio should earn (the predictive Sharpe ratio), as shown in equation (13). We can compare this predictive Sharpe ratio with the Sharpe ratio of the realized returns that the MVE portfolio earns subsequently in month $t + 1$, as shown in equation (14). We use rolling 10-year windows of these month $t + 1$ returns to calculate the realized Sharpe ratio. To compare with the predictive Sharpe ratio on the same basis, we also form a 10-year moving average of the corresponding month t predictive Sharpe ratios. We leave out the one-month lag portfolio from these calculations, as illiquidity likely plays a much bigger role in these short-run reversals than in the other anomalies.

Figure VIII shows the results. The predictive Sharpe ratio is, on average, of roughly the same magnitude as the subsequently realized Sharpe ratio. This again illustrates that the posterior distribution provides a fairly good characterization of the expected returns. It appears to be a useful to track the evolution of return-based anomalies. However, the predictive Sharpe ratio lags the realized Sharpe ratio by several years. This is not surprising. When the true Sharpe ratio changes, it takes some time, even with optimal exponential weighting, to observe sufficient data such that the posterior can pick up these changes.

The figure also shows the in-sample Sharpe ratio, calculated with exponentially-weighted mean returns until month t, without any shrinkage, and the sample covariance matrix, calculated as described in equation (16) with data up to month t. The in-sample Sharpe ratio is much higher than the predictive Sharpe ratio. Relying on the in-sample Sharpe ratio in predicting future Sharpe ratios of the MVE combination of lag portfolios would provide a far too optimistic view of actually achievable Sharpe ratios for a real-time investor. The gap between the in-sample Sharpe ratio and the predictive Sharpe ratio is quite stable at around 1.5. This gap is a consequence of the shrinkage

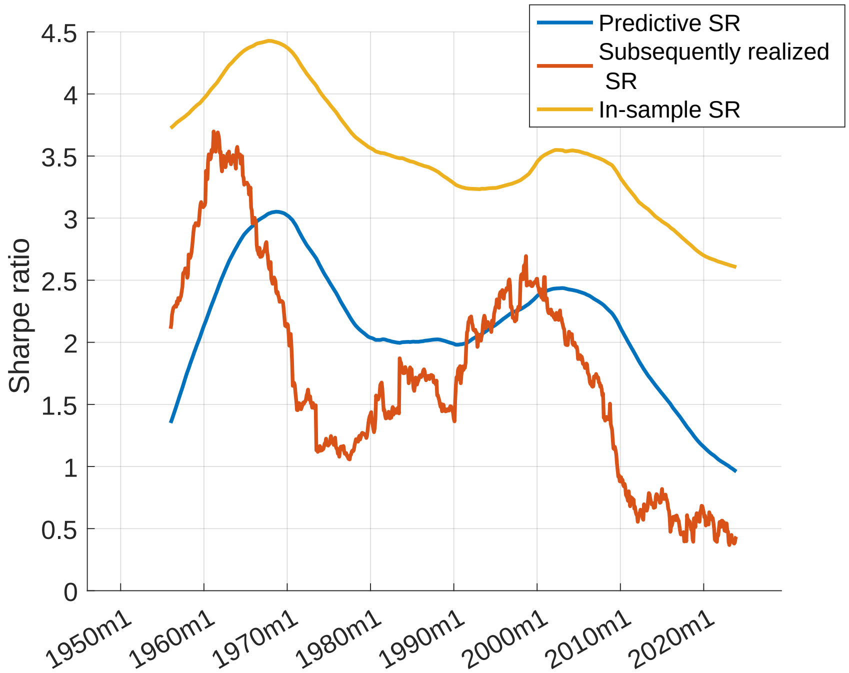


Figure VIII
In-sample, predictive, and out-of-sample Sharpe ratios

Excludes one-month lag portfolio. The out-of-sample Sharpe ratios are based on 10-year moving window means and standard deviations of realized returns of the mean-variance efficient portfolio implied by the predictive distribution of the previous month. The in-sample Sharpe ratios and the Sharpe ratios implied by the predictive distribution are averaged within the same 10-year moving windows. The date shown is the end of the window.

induced by the prior beliefs.

V. Exploring explanations for cross-sectional and time-variation of return-based anomaly expected returns

Having posterior means of lag portfolio expected returns across all portfolios at every point in time gives us an opportunity to shed light on the economic mechanisms that may give rise to these anomalies, as well as their disappearance. We first examine how posterior means of different lag portfolios relate to each other, as this sheds light on behavioral theories of momentum and reversal. When then study how posterior means at every point in time relate to portfolio holdings of different investor types. This provides insights into which investors reinforce and which ones tend to dampen the anomalies with the positioning of their portfolio holdings.

V.A. Relative magnitudes of momentum and reversals

Behavioral theories of momentum and long-term reversals suggest that these anomalies are connected. The models of Daniel, Hirshleifer, and Subrahmanyam (1998), Barberis, Shleifer, and Vishny (1998), and Hong and Stein (1999) feature different behavioral biases, yet in all of them the biases simultaneously generate momentum and reversal effects. However, the theories differ regarding the relative magnitudes of momentum and reversals. In Barberis, Shleifer, and Vishny (1998), and Hong and Stein (1999), initial underreaction is followed by subsequent overreaction and eventual correction. Due to the initial underreaction, the cumulated momentum effects are bigger than the cumulated reversal effects. In contrast, in Daniel, Hirshleifer, and Subrahmanyam (1998), momentum arises from further delayed overreaction that amplifies an initial overreaction, with subsequent reversals as consequence of the eventual price correction. In this model, the cumulated cumulated reversal effects are at least as big as the cumulated momentum effects.

Using our posterior mean estimates, we can evaluate the relative magnitudes of the cumulated momentum and reversal effects, and how the relative magnitudes change over time. For this analysis, we consider the expected cumulative return of a long-term investor who takes a long position in the portfolio for lag $k = 2$ at the end of month t (i.e., we again skip the illiquidity-driven one-month reversals in this analysis). The investor's expected return during month $t + 1$, the first month of the holding period, is $m_{2,t}$, the posterior mean at the end of month t of the portfolio for lag $k = 2$. The investor keeps holding this portfolio and at the end of month $t + 1$ this portfolio becomes the lag $k = 3$ portfolio, which, by that time, has expected return $m_{3,t+1}$, then it becomes the lag $k = 4$ portfolio, and so on. To calculate the cumulative expected return over a 10-year holding period, we sum up the posterior means as follows, separating the part due

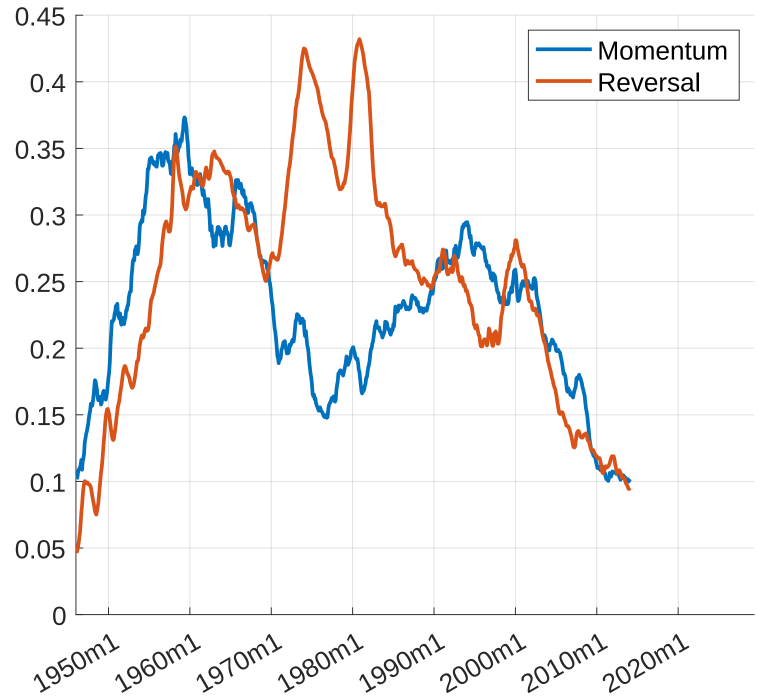


FIGURE IX
Cumulated positive (momentum) and negative (reversal) posterior means of lag portfolio expected returns over a 10-year holding period.

Excludes one-month lag portfolio. The date shown is start of the holding period.

to momentum (positive posterior mean) and reversal (negative posterior mean):

$$\begin{aligned} \text{mom}_t &= \sum_{k=2}^{120} \max(m_{k,t+k-2}, 0) \\ \text{rev}_t &= -\sum_{k=2}^{120} \min(m_{k,t+k-2}, 0) \end{aligned} \tag{23}$$

From Figure VI we know that momentum, to the extent it exists, mostly arises for lags 2-12 plus at some seasonal lags, while reversals occur at longer lags. Hence, the two summations above largely capture posterior means in different regions of lags.

Figure IX shows the result. In most periods, cumulative momentum and reversals are remarkably close in magnitude. The exception are the 1970s and early 1980s when cumulative reversal effects are substantially bigger than cumulative momentum effects. Overall, in most periods, momentum effects are largely reversed subsequently, and more than fully reversed in some periods. This does not leave much room for underreaction effects as an explanation of momentum, suggesting that it is mostly a delayed overreaction phenomenon.

V.B. Posterior expected returns of investor types

As the previous analyses showed, the nature and strength of ex-ante predictability of stock returns based on past returns has undergone major changes during the post-WW II decades. Posterior means have risen and fallen in different regions of the lag space. While we do not attempt a full investigation of which investors caused these changes in expected returns over time, we pursue here the more limited objective of evaluating which types of investors were positioned, ex-ante, to benefit from the investment opportunities that a Bayesian observer would have perceived.

Investors who are positioned such that their portfolio weights are positively related to the posterior means of the lag portfolios contribute, with their asset demands, towards suppressing the return-based anomalies. For example, in times when posterior means indicate strong long-term reversals, an investor who underweights stocks with high past returns at long lags contributes to an easing of the price pressure on these stocks, relative to the counterfactual of not underweighting these stocks, raising the expected return.

Existing literature has focused on momentum trading at horizons of a few quarters, finding that mutual funds and investment advisors have a particularly pronounced tendency to momentum trade (Grinblatt, Titman, and Wermers (1995); Badrinath and Wahal (2002)). Here we look not only at momentum, but, simultaneously, also at investors' positioning with respect to reversals. Moreover, and most importantly, we examine, period-by-period, how investors' positioning aligns with time-varying ex-ante expected returns of the lag portfolios.

The data for this exercise is from the Thomson Reuters Institutional Holdings Database 13F institutional holdings database. We follow Koijen and Yogo (2019) and classify institutions into banks, insurance companies, investment advisors, mutual funds, pension funds, and other 13F institutions. We label the residual outstanding shares not captured by holdings of these six types of institutions as household sector holdings. For every type n, we calculate holdings as a proportion of stock i's market capitalization in month t as

$$H_{int} = \frac{P_{it} S_{int}}{\sum_{n \in \mathcal{N}} P_{it} S_{int}}, \tag{24}$$

where S_{int} are the aggregate holdings, in terms of number of shares, of type n in stock i and P_{it} is the stock price.

Before computing the posterior mean for each investor type's expected returns, we first illustrate the differences in investor types' positioning with respect to stocks' past returns by calculating a past-return weighted holdings share for each lag k:

$$M_{ntk} = \frac{\sum_{i \in \mathcal{I}_t} H_{int}(R_{it-k} - \bar{R}_{t-k})}{\sum_{i \in \mathcal{I}_t} |R_{it-k} - \bar{R}_{t-k}|} \tag{25}$$

This calculation is similar to the lag portfolio return calculation in 3, but with month t stock returns replaced with the investor type's holdings in these stocks and with scaling in the denominator by the sum of stocks' absolute past returns instead of the sum of squared returns. This measure M_{ntk} will be high if investor type n has an aggregate portfolio that is tilted towards stocks with high R_{it-k}. We then smooth this measure cross-sectionally across lags k with a simple moving average of each lags with its two neighboring lags, and in time-series with exponential weighting with weights that have a half-life of 5 years. These degrees of cross-sectional and time-series smoothing are roughly in line with the smoothing that produced the posterior means in Figure VI.

Figure X shows the smoothed past-return weighted holdings shares across time and across lags. There are two groups that stand out: households and mutual funds. Households have a contrarian stance throughout all lag ranges, but with the strongest contrarian positioning in lags 2 to 15 in the 2000s. As Figure VI showed earlier, this was a time when momentum between lags 2 to 12 was still signifiant, and hence households are positioned wrongly with regards to momentum. However, the overall effect on households' expected returns is unclear because households also have contrarian positioning at much longer lags, and hence they may benefit in some periods from the presence of long-term reversals. Mutual funds' positioning is basically the mirror image of households'. They are positioned to benefit from momentum, but they also have momentum-type exposure at much

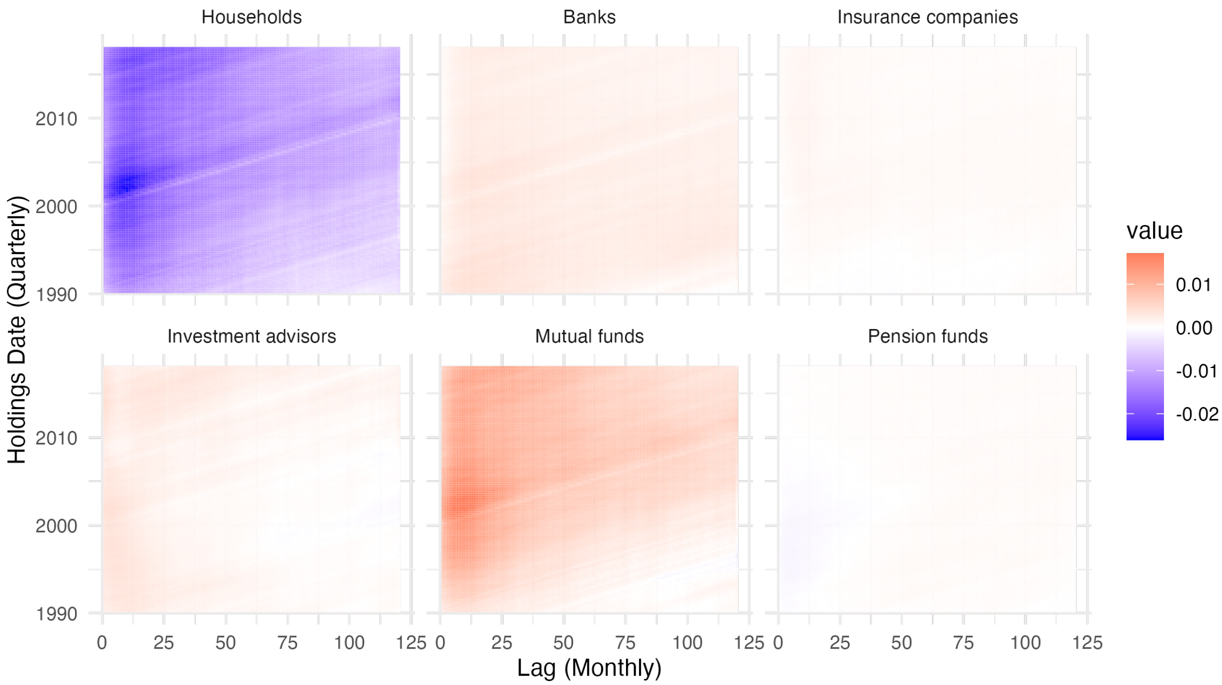


FIGURE X
Past-return weighted holdings shares by investor type
Positive numbers indicate momentum positioning, negative numbers indicate contrarian positioning

longer lags, which means they may be hurt by long-term reversals.

To analyze the overall effect on the investor types' posterior mean of expected returns, we aggregate the posterior mean estimates from Figure VI as follows. We first calculate the posterior mean for each stock i based on its own history of past returns,

$$\mu_{i,t} = \sum_{k=2}^{120} m_t(k)(r_{i,t-k} - \bar{r}_{t-k}). \tag{26}$$

This calculation relies on the approximation in (4) that the expected lag k portfolio return is approximately equal to the slope coefficient in a linear projection of future returns on past returns. Next, we aggregate these stock-level posterior means for every 13F institution j using their portfolio

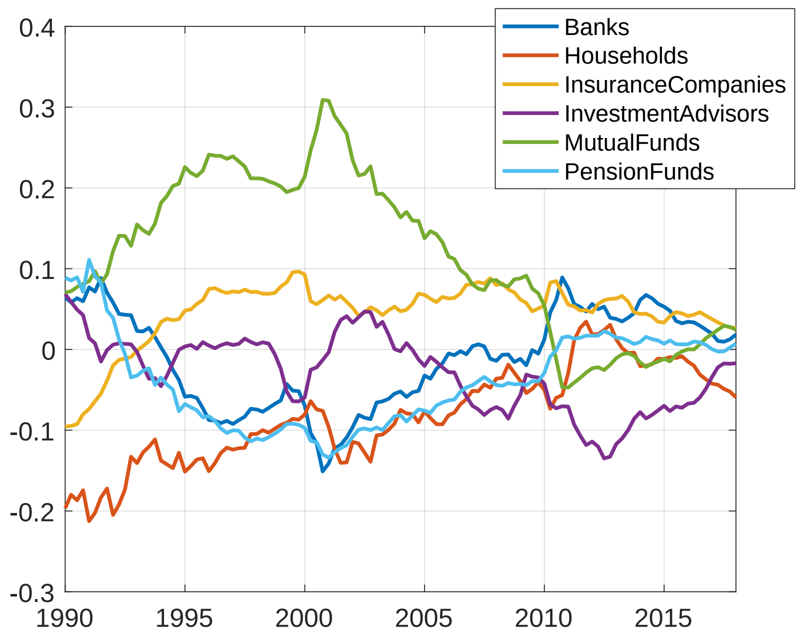


FIGURE XI
Posterior expected returns of investors by type

Annualized in percent, 10-year moving averages, relative to cross-sectional mean of all types. Excludes one-month lag portfolio.

weights $w_{j,i,t}$

$$\mu_{j,t} = \sum_{i \in \mathcal{I}_t} w_{j,i,t} \mu_{i,t}. \tag{27}$$

Finally, we aggregate at the level of institution type n using each institution's total 13F holdings of stocks ($AUM_{j,t}$) as weight

$$\mu_{s,t} = \sum_{j \in \mathcal{S}_t} \mu_{j,t} \times \frac{AUM_{j,t}}{\sum_{j \in \mathcal{S}_t} AUM_{j,t}}. \tag{28}$$

Figure XI shows 10-year moving averages of the posterior expected returns by type, relative to the cross-sectional mean of all types each period. When posterior expected returns of many lag portfolios were still significantly different from zero in the years leading up to around 2010, as seen earlier in Figure VI, mutual funds and households stand out most. Mutual funds had positive posterior expected returns, more positive than other types most of the time, and households had

TABLE II

Posterior expected returns for institutional managers by type

This table uses the same data as in Figure XI and it shows the full-sample averages and t-statistics.

	Mean	(t-statistic)
Households	-0.10	(-3.46)
Banks	0.00	(0.15)
Insurance Companies	0.02	(1.00)
Investment Advisors	-0.01	(0.32)
Mutual funds	0.10	(3.59)
Pension funds	-0.01	(-0.37)

negative posterior expected returns, more negative than other types most of the time. This is consistent with the fact from Figure X that the holdings of mutual funds were tilted towards momentum during these years, while households took a contrarian position, which means that the presence of momentum hurt their relative performance. At the same time, neither type had their holdings perfectly aligned, or perfectly misaligned with posterior expected returns. Mutual funds' momentum positioning in their holdings extends lags beyond lag 12 where reversals were dominant, hurting their expected returns, while household contrarian positioning beyond lag 12 benefited their expected returns. This is one reason why the magnitude of the differences in posterior expected returns between households and mutual funds is relatively small. The maximum spread between them in Figure XI is around 0.45% per year.

Table II presents full-sample averages of the cross-sectionally demeaned posterior expected returns by type, as well as the associated standard errors. The only types with substantial positive and negative average of the posterior means are mutual funds and households, respectively.

The bottom line is that there is some evidence that more sophisticated investors (mutual funds) positioned their portfolios more in line with the prescriptions of the posterior distribution of expected returns than less sophisticated investors (households) did, consistent with the notion that the posterior mean expected returns are perceived as exploitable mispricing by the sophisticated investors, but the alignment is far from perfect.

VI. Conclusion

Empirical asset pricing research is often retrospective, looking at in-sample moments ex-post to uncover what investors are presumed to have known ex-ante. In contrast, this paper considers a hypothetical researcher focused on discovering and tracking return-based anomalies in real time. The researcher starts in 1926, with no prior knowledge of return-based anomalies, using Gaussian process regression methods with empirical Bayes shrinkage that guards against spurious anomaly discoveries due to multiple testing and against picking up the ex-post visible effects of investor learning. We find that this researcher would have discovered prominent return-based anomalies—such as momentum and long-term reversal—well before the authors of the published studies analyzed the data. This suggests that these anomalies represented, at the time of their discovery by academics, genuine ex-ante expected returns.

However, tracking these anomalies in real-time with optimal weighting of past data reveals that many prominent anomalies weaken substantially around their respective publication dates. For example, the long-term reversal effect identified by De Bondt and Thaler (1985) largely disappears in the mid-1980s. Since our posterior mean estimates are already shrunk towards zero to account for the effects of multiple testing and investor learning, this is a decay in ex-ante expected returns. This decay is consistent with investors eventually learning about the anomaly, but slower than our hypothetical researcher with the Gaussian process regression method.

These findings underscore the dynamic and evolving nature of return-based anomalies. Empirical analyses that assume constant expected returns over time overlook the patterns of emergence, decay, and occasional reappearance of anomalies. Based on the evidence presented in this paper, there is little justification for regarding momentum, long-term reversal effects, or any other return-based anomalies as permanent attributes of the cross-section of expected stock returns.

The Gaussian process regression approach we develop in this paper is well-suited for tracking the evolution of return-based anomalies through time. Extending this approach to include other cross-sectional return predictors could be useful, but would require some further methodological advances. For strategies based on lagged returns, the distance in terms of lag is a natural measure of the likely similarity of expected returns, and hence for the covariance in the Gaussian Process prior. To apply the approach to cross-sectional return predictors more generally would require development of alternative similarity metrics, and methods to deal with a much more high-dimensional covariance matrix in the Gaussian Process prior. This is a promising area for further research.

References

Badrinath, Swaminathan G, and Sunil Wahal, 2002, "Momentum Trading by Institutions," *Journal of Finance* 57, 2449–2478.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny, 1998, "A Model of Investor Sentiment," *Journal of Financial Economics* 49, 307–343.

Chen, Andrew Y, and Chukwuma Dim, 2023, "High-Throughput Asset Pricing," Working paper, arXiv.

Chen, Andrew Y, and Tom Zimmermann, 2020, "Publication Bias and the Cross-Section of Stock Returns," *Review of Asset Pricing Studies* 10, 249–289.

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam, 1998, "Investor Psychology and Security Market Under-and Overreactions," *Journal of Finance* 53, 1839–1885.

De Bondt, Werner F.M., and Richard Thaler, 1985, "Does the Stock Market Overreact?," *Journal of Finance* 40, 793–805.

Filipović, Damir, and Puneet Pasricha, 2022, "Empirical Asset Pricing via Ensemble Gaussian Process Regression," Working paper, arXiv.

Fong, Edwin, and Chris C Holmes, 2020, "On the Marginal Likelihood and Cross-Validation," *Biometrika* 107, 489–496.

Grinblatt, Mark, Sheridan Titman, and Russ Wermers, 1995, "Momentum Investment Strategies, Portfolio Performance, and Herding: A Study of Mutual Fund Behavior," *American Economic Review* 85, 1088–1105.

Heston, Steven L, and Ronnie Sadka, 2008, "Seasonality in the Cross-Section of Stock Returns," *Journal of Financial Economics* 87, 418–445.

Hong, Harrison, and Jeremy C Stein, 1999, "A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets," *Journal of Finance* 54, 2143–2184.

Jegadeesh, Narasimhan, 1990, "Evidence of Predictable Behavior of Security Returns," *Journal of Finance* 45, 881–898.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, "Returns to Buying Winners and Selling Losers: Implications for Market Efficiency," *Journal of Finance* 48, 65–91.

Jensen, Theis Ingerslev, Bryan Kelly, and Lasse Heje Pedersen, 2023, "Is There A Replication Crisis in Finance?," *Journal of Finance* 78, 2465–2518.

Koijen, Ralph S. J., and Motohiro Yogo, 2019, "A Demand System Approach to Asset Pricing," *Journal of Political Economy* 127, 1475–1515.

Li, B, A Rossi, S Yan, and L Zheng, 2022, "Real-time Machine Learning in the Cross-Section of Stock Returns," Working paper, Georgetown University.

Martin, Ian, and Stefan Nagel, 2022, "Market Efficiency in the Age of Big Data," *Journal of Financial Economics* 145, 154–177.

McLean, David R., and Jeffrey Pontiff, 2016, "Does Academic Research Destroy Stock Return Predictability?," *Journal of Finance* 71, 5–32.

Novy-Marx, Robert, 2012, "Is Momentum Really Momentum?," *Journal of Financial Economics* 103, 429–453.

Williams, Christopher KI, and Carl Edward Rasmussen. *Gaussian Processes for Machine Learning*, 2006).

Yan, Xuemin, and Lingling Zheng, 2017, "Fundamental Analysis and the Cross-Section of Stock returns: A Data-Mining Approach," *Review of Financial Studies* 30, 1382–1423.

Appendix

A. Example to illustrate connection between marginal likelihood estimation of prior hyperparameters and cross-validation

Assume that the return on a single asset follows

$$r_t = \mu + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma^2) \tag{A.1}$$

with prior beliefs

$$\mu \sim \mathcal{N}(0, \sigma_0^2). \tag{A.2}$$

The variance of $\bar{r}_t = \frac{1}{t}\sum_{s=1}^{t} r_s$ under the prior distribution is

$$\omega^2 = \sigma_0^2 + \frac{\sigma^2}{t} \tag{A.3}$$

and the marginal likelihood, expressed in terms of the marginal likelihood of the average realized return, is

$$\log p(\bar{r}_t|\sigma_0^2) = \text{const.} - \frac{\bar{r}_t^2}{\omega^2} - \log(\omega^2). \tag{A.4}$$

This yields the first-order condition w.r.t σ_0^2,

$$\bar{r}_t^2(\omega^2)^{-2} - (\omega^2)^{-1} = 0 \tag{A.5}$$

which we can solve for

$$\sigma_0^2 = \bar{r}_t^2 - \frac{\sigma^2}{t}. \tag{A.6}$$

To estimate σ_0^2 we effectively have one observation, the average return $\bar{r}_t$. With mean equal to zero under the prior beliefs, we estimate σ_0^2 with $\bar{r}_t^2$ corrected for the expected estimation error component $\frac{\sigma^2}{t}$. With this estimate of σ_0^2, the factor by which $\bar{r}_t$ is shrunk towards the prior mean of zero is

$$\gamma(t) = \frac{\bar{r}_t^2 - \frac{\sigma^2}{t}}{\bar{r}_t^2} \tag{A.7}$$

Now consider leave-one-out cross-validation criterion when using the optimal shrinkage factor from (A.7)

$$\sum_{j=1}^{t}[r_j - \bar{r}_{-j}\gamma(t-1)]^2, \tag{A.8}$$

where $\bar{r}_{-j}$ denotes the average return in the sample up to t with the j-th observation excluded.

The first-order condition with respect to $\gamma(t-1)$ yields,

$$\sum_{j=1}^{t} r_j \bar{r}_{-j} + \bar{r}_{-j}^2 \gamma(t-1)]^2 = 0. \tag{A.9}$$

Substituting from (A.7), the left-hand side becomes

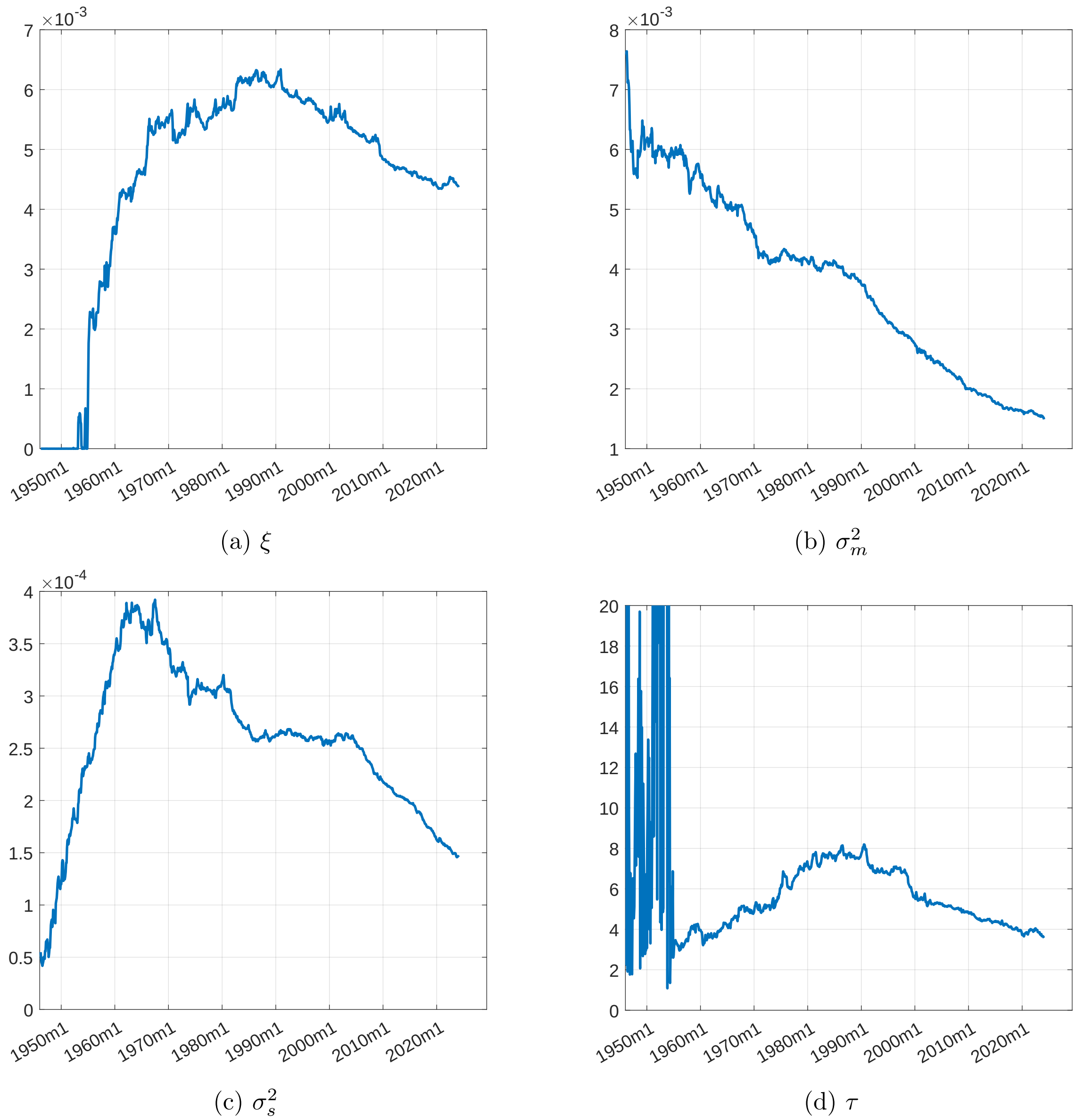
$$\sum_{j=1}^{t} r_j \bar{r}_{-j} + \bar{r}_{-j}^2 - \frac{\sigma^2}{t-1} \tag{A.10}$$

$$= \sum_{j=1}^{t} \bar{r}_{-j}(\bar{r}_{-j} - r_j) - \frac{\sigma^2}{t-1} \tag{A.11}$$

$$= \sum_{j=1}^{t} (\mu + \bar{\varepsilon}_{-j})(\bar{\varepsilon}_{-j} - \varepsilon_j) - \frac{\sigma^2}{t-1} \tag{A.12}$$

The first term in this summation has expected value $\frac{\sigma^2}{t-1}$ and it is an estimate of $\frac{\sigma^2}{t-1}$. Therefore, the cross-validation criterion is satisfied approximately by the shrinkage factor (and underlying hyperparameter σ_0^2) chosen by marginal likelihood.

B. Hyperparameter estimates in constant expected returns case



(a) ξ (b) σ_m^2

(c) σ_s^2 (d) τ

FIGURE A.1
Hyperparameter estimates: Exponentially weighted data