Rerunning the history of momentum and reversal discoveries

Benjamin Marrow\textsuperscript{1}  Stefan Nagel\textsuperscript{2}

\textsuperscript{1}University of Chicago
\textsuperscript{2}University of Chicago, NBER, CEPR, and CESifo

November 2023
Discoveries of cross-sectional stock return predictability

- **Typical approach**
  - One or a few ad-hoc selected stock characteristics
  - Ex-post analysis of full-sample historical returns

- **Typical conclusion:** Historical average returns ≈ ex-ante expected returns for rational investor

- **Implicit assumptions**
  - No multiple testing (individually, collectively)
  - No investor learning
  - No decay (or emergence, re-emergence) of predictability
Issues that complicate interpretation of published studies

- Multiple testing invalidates conventional significance tests
  - Harvey and Liu 2016, but also Chen 2021

- Investor learning about cash flow process generates in-sample return predictability
  - Martin and Nagel 2022

- Decay (and possibly emergence or re-emergence) of anomalies

- **Goal:** Rerun the history of anomaly discovery in a way that addresses these issues
Solution: Empirical Bayes approach

- **Hierarchical Bayesian** approach removes multiple-testing problem if all potential predictor variables are considered jointly without ex-post selection.

- **Empirical Bayes** data-driven estimation of prior hyperparameters induces appropriate shrinkage to shrink away “excessive” predictability due to learning.

- **Exponential weighting** to allow for decay, emergence, re-emergence of anomalies.
Consider set of potential predictors without ex-post selection

- Stock characteristics that have already appeared in published asset pricing studies are a selected sample, subject to look-ahead bias

- Therefore: Use an entire class of predictor variables without pre-selecting particular variables in this class

- Here: Linear prediction based on lagged monthly past stock returns $r_{it-1}, r_{it-2}, \ldots, r_{it-120}$

- Construct 120 portfolios, each weighted by market-adjusted returns lagged $k$ months: Portfolio return $f_{t,k}$.
  - exclude stocks with lagged price $< 1$ and market cap $< 20$th NYSE percentile
Past-return-based anomalies

- Prior research has selectively focused on ad-hoc (ex-post selected?) subsets and did not adjust for learning effects
Portfolio aggregation: Lagged return portfolios

- Let $X_{i,t} = R_{i,t} - \bar{R}_t$, stacked into $x_t$

- Weights of regression slope portfolio just based on lag $k$ returns
  
  $$w_{k,t} = x_{t-k}(x'_{t-k}x_{t-k})^{-1}$$

- Portfolio return for lag $k$
  
  $$F_{k,t} = x'_t w_{k,t}$$

- In the linear projection
  
  $$X_{i,t} = b_1 X_{i,t-1} + b_2 X_{i,t-2} + ... + b_K X_{i,t-K} + \varepsilon_{i,t}$$

  since returns at different lags have low correlation we have

  $$\mathbb{E}[F_{k,t}] \approx b_k.$$
Full-sample historical average returns

Average returns of 120 portfolios that weight stocks by their market-adjusted returns in month $t - 1$, $t - 2$, ..., $t - 120$

Sample period: 1926 to 2021; first portfolio returns in January 1936.
Hierarchical Bayesian approach with Gaussian process regression

- Let \( \mu(k) = \mathbb{E}[F_{k,t}] \) and assume
  \[
  F_{k,t} = \mu(k) + \eta_{k,t}, \quad \eta_{k,t} \sim \mathcal{N}(0, \sigma^2)
  \]

- Nonlinear function \( \mu(k) \) specified as Gaussian process.

- Prior beliefs
  \[
  \mu(k) \sim \mathcal{GP}(0, \psi(k, k'))
  \]

- Agnostic whether momentum or reversal patterns are more likely
- Expected returns at different lags can be correlated
Prior covariance

- Covariances have three components
  \[ \psi(k, k') = m \times \mathbb{1}_{k=1 \cap k'=k} + s \times \mathbb{1}_{k \in \{12,24,\ldots\} \cap k'=k} + \kappa(k, k') \]

- Microstructure effects at lag 1
- Seasonality at lags that are multiple of 12
- Correlated expected returns at neighboring lags

- Triangular kernel
  \[ \kappa(k, k') = \xi^2 (1 - |u|) \mathbb{1}_{|u|<1}, \quad u = \frac{k - k'}{\tau} \]
Prior distribution: 20 draws

\[ \tau = 10, \sigma_m = 0.3, \sigma_s = 0.2, \xi = 0.1. \]
Posterior

- With $\mu(k)$ for integer lags $k = 1, 2, ..., K$ collected in $\mu$ and all $\psi(k, k')$ in $\Psi$, the time-$t$ posterior is

$$\mu | \bar{f}_t \sim \mathcal{N}(m_t, V_t)$$

with

$$m_t = \left( I + \frac{1}{t} \Sigma \psi^{-1} \right)^{-1} \bar{f}_t$$

$$V_t = \Psi - \Psi \left( \psi + \frac{1}{t} \Sigma \right)^{-1} \psi$$

- Prior exerts two effects in the posterior:
  - Covariance induces cross-sectional smoothing (more if $\tau$ big): raises precision
  - Finite variance induces shrinkage (more if $\xi$ small): lowers magnitude
Empirical Bayes

- Prior hyperparameters $\theta = (\xi, \tau, m, s)$:
  - $\xi$: shrinkage
  - $\tau$: smoothing
  - $m$: microstructure
  - $s$: seasonal

- Estimated each $t$ by max. marginal likelihood

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

with

$$
\Omega_t = \Psi + \frac{1}{t} \Sigma
$$

- Then compute posterior distribution and empirical Bayes $t$-statistics

$$
\text{posterior mean of } \mu(k) \over \sqrt{\text{posterior variance of } \mu(k)}
$$
Predictive distribution and Sharpe ratio

- Predictive distribution has means $m_t$ and covariances

$$\Phi_t = V_t + \Sigma$$

- Sharpe ratio implied by predictive distribution

$$\sqrt{12} \sqrt{m_t' \Phi_t^{-1} m_t}$$

- Out-of-sample MVE portfolio return

$$r_{mve,t+1} = m_t' \Phi_t^{-1} f_{t+1}$$
Estimating the covariance matrix of innovations

► Decompose into diagonal standard deviation matrix $H$ and correlation matrix $C$

$$\Sigma = HCH$$

► Assume banded structure

$$C = \begin{pmatrix}
1 & \rho_1 & \rho_2 & 0 & 0 & \ldots & 0 \\
\rho_1 & 1 & \rho_1 & \rho_2 & 0 & \ldots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & \ldots & 0 & \rho_2 & \rho_1 & 1 & \rho_1 \\
0 & \ldots & 0 & 0 & \rho_2 & \rho_1 & 1
\end{pmatrix}.$$

► Estimate elements of $H$ and $C$ with sample standard deviations and correlations in data up to $t$. 
Posterior $t$-statistics

Estimated over expanding windows since December 1935.
Estimated over expanding windows since December 1935.
Posterior $t$-statistics

Estimated over expanding windows since December 1935.
Posterior $t$-statistics

Estimated over expanding windows since December 1935.
Allow for decay or emergence of anomalies

- Make posterior beliefs adapt to **structural change**: Exponential weighting

\[
\tilde{f}_{t,k} = \frac{\sum_{s=0}^{t-1} \nu(s)f_{t-s,k}}{\sum_{s=0}^{t-1} \nu(s)}, \quad \nu(s) = e^{-\delta s}
\]

- Additional hyperparameter $\delta$ that controls weight decay

- Estimate this hyperparameter by minimizing mean log loss

\[
- \log p(f_{j+1}|\delta, \theta) = \frac{\left( f_{j+1} - m_j(\delta) \right)'^\top \Phi_j^{-1}(f_{j+1} + m_j(\delta))}{2} + \frac{\log \det(\Phi_j)}{2} + \frac{\log(2\pi)}{2}.
\]

in past data up to $t$. 
Posterior $t$-statistics

![Posterior t-statistics](image-url)
Temporary Disappearance of LT reversals

DeBondt-Thaler 3-5yr reversals
Temporary Disappearance of Momentum

Jegadeesh-Titman 3-12m momentum
Momentum and LT reversals appear, disappear, reappear together
Sharpe ratios
Who creates/destroys anomalies? [Preliminary]

- Look at descriptive patterns in 13F institutional holdings data
  - Next steps: demand system estimation to get counterfactual posteriors without effect of demand of investor type \( n \)

- Investor type \( n \) holdings of stock \( i \) in month \( t \) holds shares \( S_{int} \). As proportion of outstanding market cap

\[
H_{int} = \frac{P_{it}S_{int}}{\sum_{n \in N} P_{it}S_{int}}
\]

- Measure of correlation of holdings with weights of lag-\( k \) portfolio

\[
M_{nt} = \frac{\sum_{i \in I} H_{int}(R_{it-k} - \bar{R}_{t-k})}{\sum_{i \in I} |R_{it-k} - \bar{R}_{t-k}|}
\]

Smoothed cross-sectionally and in time (one-sided, exponentially).
Institutional holdings

2.3 Stock Ownership (Levels)

I define ownership as:

\[ S_{int} := p_{it} \times \text{Shares}_{int} \]

P_{m2N} p_{it} \times \text{Shares}_{imt} \]

i.e., the percent of market cap held by institution \( i \) at date \( t \).

I plot the smoothed measures of \( M_{ntk} \) with \( S_{int} \) in place of \( \checkmark \).

(a) Households

(b) Banks
Institutional holdings

2.3 Stock Ownership (Levels)

I define ownership as:

$$S_{int} := \frac{p_{it}}{\text{Shares}_{int}}P,$$

i.e., the percent of market cap held by institution $i$ at date $t$. I plot the smoothed measures of $M_{ntk}$ with $S_{int}$ in place of $\times_{int}$.

(a) Households  (b) Banks
(c) Insurance Companies  (d) Investment Advisers
(e) Mutual Funds  (f) Pension Funds

Figure 3: Stock Ownership and Lag Returns (Monthly)
Institutional holdings

2.3 Stock Ownership (Levels)

I define ownership as:

\[ S_{int} := p_{it} \times \text{Shares}_{int} \]

\[ P_{m2N} \times p_{it} \times \text{Shares}_{imt} \]

i.e., the percent of market cap held by institution \( i \) at date \( t \). I plot the smoothed measures of \( M_{ntk} \) with \( S_{int} \) in place of the

- (a) Households
- (b) Banks
- (c) Insurance Companies
- (d) Investment Advisers
- (e) Mutual Funds
- (f) Pension Funds

Figure 3: Stock Ownership and Lag Returns (Monthly)
Conclusion

- Traditional approach of examining cross-sectional return predictability
  - with ad-hoc selected subset of stock characteristics
  - without robustness to multiple testing
  - without shrinkage to adjust for learning effects
  - without allowing for structural change

has little justification.

- Major past-return based anomalies would have been discovered by Bayesian, non-data-mining researcher around the time of the original studies

- Dynamic decay, (re)emergence patterns
Future work

- In progress: Contribution of different investor types to creation/destruction of anomalies

- Interesting potential extensions: How can one investigate other anomalies without pre-selecting stock characteristics.
  - Example: Use all variables one can construct from COMPUSTAT as potential predictors?