The Virtue of Complexity in Return Prediction

B. Kelly, S. Malamud, and K. Zhou

Discussion:

Eben Lazarus
MIT Sloan

CICF
July 2022
Motivation

It’s hard to make predictions, especially about the future \textit{out of sample} \cite{WelchGoyal2008}.

Intuition and theory tell us that returns \textit{should} be predictable:

- Valuations vary over time
- Essentially all quantitatively viable models (reduced-form, consumption-based, . . .) feature time-varying price and/or quantity of risk

Lack of predictability — i.e., negative out-of-sample $R^2$ relative to prediction based on historical mean — could reflect either:

1. Unpredictable returns
2. Predictable returns, but with a high-dimensional or time-varying DGP

This paper’s starting point
Outline

1. Recap
2. Comments
3. Portfolio Interpretation
Overview

What the paper does (a lot!):

1. A theoretical characterization of returns to market-timing strategies in a machine-learning context
   - Assume that the true DGP is high-dimensional, with many relevant predictors. Asymptotic embedding:
     \[ E_t[R_{t+1}^{\text{market}}] = \beta' S_t, \quad \beta \in \mathbb{R}^P, \]
     \[ P/T \to c > 0 \quad \text{as} \quad T \to \infty, \]
     \[ E[\beta\beta'] \to P^{-1}b_\ast I_P \quad \text{for some constant} \quad b_\ast \]
   - Stein (1955): OLS is inadmissible (even if \( c < 1 \)). Practically speaking, estimation error blows up MSE, leading to (very) negative OOS \( R^2 \).
   - Instead, consider ridge estimator that shrinks \( \hat{\beta} \) toward 0
   - Characterize \( R^2 \) and Sharpe ratio of timing strategy that takes positions equal to cond. expected return \( (\pi_t = \hat{\beta}' S_t) \), both for (i) correctly specified and (ii) incorrectly specified (only using \( P_1 < P \) predictors) models
Overview

What the paper does (a lot!):

2. Empirical investigation of these ridge-regularized market-timing strategies
   ▶ Take random linear combinations of 15 RHS variables from Welch and Goyal (2008) to generate between $P = 2$ and $P = 10,000$ predictors
   ▶ Estimate $\beta$ using rolling 12-month window of observations
   ▶ Invest $\hat{\beta}'S_t$ in the market and assess one-month returns

Results:

▶ **Theory:** Uncover **virtue of complexity** in high-dimensional setting
   ▶ Best-case scenario: $c = 0 \implies$ easy prediction. Obviously not the case!
   ▶ Second-best: $c \gg 1!$ Better to have lots of signals (each contributing a little predictability) than highly ill-conditioned problem where $c \approx 1$ ($P \approx T$), especially under misspecification

▶ **Data:** Aligns well with theory
   ▶ Performance improves with model complexity & shrinkage, increasing Sharpe ratio by 0.3 relative to static strategy ($t \approx 2.7$)
Outline

1. Recap

2. Comments

3. Portfolio Interpretation
Initial Comments

Extremely useful and promising set of results.

- Much of the high-dimensional return prediction literature has consisted of measurement without theory
- Without theoretical grounding, methods are exciting but black-boxy
- This paper provides a sound theoretical basis for high-dimensional return prediction, and shows that there are in fact extractable predictive signals in the data
- Useful additional point: Positive $R^2$ is not necessary for Sharpe ratio improvements
  - I think this is a version of Clark & West (2007): $R^2$ should be negative under the null of no predictability (because of noise in estimating coefficients that are 0 in population), so $R^2$ needs to be adjusted in order to be useful
  - Related to Campbell & Thompson (2008) point that mean-variance investor can improve utility by timing even with small $R^2$
  - Tangentially ties into Lazarus, Lewis, Stock (2021): MSE is often the wrong decision metric, either for testing (there) or prediction (here)
Some Questions

Paper is a very good proof of concept, but some issues related to implementation and interpretation remain unanswered:

▶ Guidance for shrinkage parameter $z$: Estimate via cross-validation or empirical Bayes? Does this affect theoretical results (which are currently conditional on pointwise $z$)?
  ▶ Is OOS $R^2$ positive under optimal (or your preferred) $(c, z)$ pair?

▶ Back to motivation: Return predictability DGP likely time-varying, but dealt with here just by using rolling 12-month estimation
  ▶ This seems quantitatively important: $t$-stat on market timing $\alpha$ shrinks to below 2 when using 120-month training window
  ▶ “Pockets” of predictability are thus quite narrow [Farmer, Schmidt, & Timmermann (2022)]

▶ Conceptually mildly worrisome. But presumably time variation can be dealt with by considering lagged data; does considering random Fourier transforms in time domain (in addition to cross-section of signals) help with longer windows?
Outline

1. Recap

2. Comments

3. Portfolio Interpretation
Interpretation of Optimal Portfolios

Optimal timing positions are interesting: long-mostly & shrink before recessions.

Figure 10: Market Timing Positions
Interpretation of Optimal Portfolios

Optimal timing positions are interesting: long-mostly & shrink before recessions.

Why does this happen? My interpretation:

- Returns and predictors are normalized, so smaller $|\hat{\pi}|$ arises from smaller $||\hat{\beta}||$

- In ridge regression, smaller $||\hat{\beta}|| \iff$ first few principal components less informative
  - Principal component decomposition of covariance matrix of signals $S_t$: $\Psi = \Gamma \Lambda \Gamma'$, with ordered squared eigenvalues $\lambda_1^2, \lambda_2^2, \ldots$
  - With transformed parameter vector $\hat{\alpha} = \Gamma' \hat{\beta}$, ridge regression can be interpreted as $\hat{\alpha}_i = \frac{\lambda_i^2}{\lambda_i^2 + \hat{\alpha}_i^{OLS}} \implies$ lower eigenvalues lead to more shrinkage

- In order for $\hat{\pi}$ to be asymmetrically distributed around 0 and smaller before recessions, it must be the case that signal eigenvalues are smaller at those times (and then increase during/after recessions)

- This is intuitively plausible! The world is much closer to one-factor during a recession (high eigenvalues), while expansions feature variation with much weaker factor structure (decreasing eigenvalues as the expansion proceeds, which ends up predicting low returns well)

  - And holds up in the data: Using normalized Welch & Goyal signals and 12-month rolling windows, I estimate $\text{Corr}(\lambda_1^2, R_{t\rightarrow t+12}^{\text{mkt}}) = 0.27$
Final Notes

- Very cool paper
- Meaningful progress on longstanding question: are returns predictable in the time series?
- Paper’s job is difficult, especially empirically: Return prediction is likely easier for longer horizons, while this considers just the one-month horizon
- Lots of follow-up work opened up by this proof of concept
- Excited to see future work using this as a starting point