Discussion: The Virtue of Complexity in Return Prediction

Ron Kaniel, Simon Business School, University of Rochester
The magic of machine learning (shown in recent years):

Machine Learning (ML) algorithms utilizing highly complex models can achieve accurate out-of-sample forecast despite fitting the data perfectly.

How does this square with what we have known for a long time that as the number of parameters (P) increase and get closer to the number of observations (T) the out-of-sample forecast quality deteriorates?

Remember: Ordinary Least Squares becomes unworkable when P>=T.

In this paper: show the “virtue of complexity” for market timing portfolio performance:

ML out of sample Sharpe ratio is everywhere positive, despite sometimes having a massively negative $R^2$, even for extreme levels of model complexity.

When the model is mis-specified the out of sample Sharpe ratio generally increases with complexity, even with minimal regularization.
Setting (a market timing problem)

Single risky asset:

\[ R_{t+1} = S_t^T \beta + \varepsilon_{t+1}, \]

with \( \varepsilon_{t+1} \) i.i.d., \( E[\varepsilon_{t+1}] = E[\varepsilon_{t+1}^3] = 0, E[\varepsilon_{t+1}^2] = \sigma^2, E[\varepsilon_{t+1}^2] < \infty \)

State variables:

\( S_t = \Psi^{1/2} X_t, X_t \in R^p \)

where \( E[X_{i,j}] = E[X_{i,j}^3] = 0, E[X_{i,j}^4] \) uniformly bounded,

and \( X_{i,j} \) satisfy the Linberg condition, and \( \Psi \) positive semi–definite

Loadings:

\( \beta \in R^p \) random independant of \( S \) and \( R \), rotationally symmetric...

and satisfy same Linberg condition as \( X \)

Timing Strategy:

\[ R_{\pi}^t = \pi_t R_{t+1}, \] centering on the timing strategy \( \pi_t = \beta^T S_t \)

Main Focus: Evaluation of

\[ SR = \frac{E[R_{\pi}^t]}{\sqrt{E[(R_{t+1}^\pi)^2]}}, \]
Estimating Method
(Representative Machine Learning Procedure)

- Ridge-regularized least squares

\[
\hat{\beta}(z) = \left( zI + T^{-1} \sum_i S_i S_i^T \right)^{-1} \frac{1}{T} \sum_i S_i R_{t+1}
\]

- Ridgeless regression estimator

\[
\hat{\beta}(0^+) = \lim_{z \to 0^+} \left( zI + T^{-1} \sum_i S_i S_i^T \right)^{-1} \frac{1}{T} \sum_i S_i R_{t+1}.
\]

  - Equivalent to the solution

\[
H \frac{1}{T} \sum_i S_i R_{t+1}, \text{ where } H = \text{Moore-Penrose pseudoinverse of } T^{-1} \sum_i S_i S_i^T
\]

  When \( P<T \) OLS is the ridgeless solution

  - \( P=T \) (Interpolation Boundary)
    - OLS breaks down, but still unique least square solution that fits data exactly
Mean Squared Error (MSE) – classical region

- Financial decomposition of MSE

\[
MSE(\hat{\beta}) = E \left[ \left( R_{t+1} - S(\hat{\beta}) \right)^2 \right] = E[R_{t+1}^2] - 2 E[R_{t+1} \pi_t] \beta + E[\pi_t^2] \beta.
\]

- Classical statistical decomposition into bias/variance tradeoff:

\[
MSE(\hat{\beta}) = \left( \text{Bias}(\hat{\beta}) \right)^2 + \text{Var}(\hat{\beta})
\]
Measuring complexity (P- number of Parameters, T- number of periods):

\[ c = \lim_{T,P \to \infty} \frac{P}{T} \]

\[ MSE(z; c) = \lim_{T,P \to \infty, P/T \to c} E \left[ \left( R_{t+1} - S_t \hat{\beta}(z) \right)^2 \left| \hat{\beta}(z) \right| \right] \]
$R^2$ and Leverage

Figure 1: Expected Out-of-sample $R^2$ and Norm of Least Squares Coefficient

$$c = \lim_{T,P \to \infty} P / T$$
Fully Specified Model

Figure 2: Expected Out-of-sample Risk and Return of Market Timing

Figure 3: Expected Out-of-sample Sharpe Ratio of Market Timing
Mis-specified Model (observe only $P_1$ out of $P$ parameters $\lim_{T \to \infty} \frac{P_1}{P} = q$)

- X-axis: holds $c$ fixed and varies $q$

Figure 5: Expected Out-of-sample Timing Strategy Risk and Return From Mis-specified Models

Figure 6: Expected Out-of-sample Timing Strategy Sharpe Ratio From Mis-specified Models
Empirical Exercise

➢ Validates the theory

➢ For high complexity models
  o Produces high out of sample monthly
    • Sharpe ratios (0.4),
    • Alphas (1%-2%),
    • Information ratios (0.25-0.3)
  o Strategy is long-only at heart: almost never bet on market downturns
  o Strategy learns to divest leading up to recessions (14/15 NBER recession dates)
    • Essentially zeros out positions

➢ Latter part of the sample:
  o lower average returns and information ratios, fewer buying recommendations, and smaller position.
Can the analysis be extended to a setting with portfolio constraints.

- The two step procedure of first estimating loadings and then taking them as given to optimize the portfolio is no longer valid in general.
- Directly maximize out of sample Sharpe ratio.

Can the restriction of zero third moments of $\varepsilon_{t+1}$ and $X_{i,t}$ be relaxed.

- Doesn’t seem to be present in some of the other papers that consider estimation in the over-parametrized region.

Consider extending to generalized (weighted) ridge regressions, as in Wu and Xu (2020), where instead of the regularizing term $zI$ one adds a weighted regularization term $z\Sigma_w$ where $\Sigma_w$ is an optimally chosen weighting matrix.
Comments - Analysis

- For mis-specified model consider adding graphs which do not hold \( c \) fixed (\( c = \lim_{T,P \to \infty} P / T, q = \lim_{T \to \infty} P_i / P \))
  - Instead set \( c = q * C_{max} \) so that when varying \( q \) both \( c \) and \( q \) increase

- Add to graphs results for the Sharpe ratio maximizing shrinkage parameter \( z^* \)
  - Consider adding a plot showing optimal shrinkage \( z^* \)

- In empirical analysis:
  - Any way to tease out what leads to the strategy being essentially long only, and what drives divestitures prior to recessions.

- Add confidence intervals to figures 7 and 8.
Comments - Exposition

- Consider shortening Section 4.1, as the key insights in this section have been shown in the recent literature (Heistie et al. (2019), Wu and Xu(2020) and others).
  - Readers will get faster to the novel insights

- A bit less emphasis on the Ridgeless results
  - Seem to always deliver the lowest out of sample Sharpe ratios; both in fully specified and in mis-specified models

- Explain in the text where the restrictions on the third-moments of $\varepsilon_{t+1}$ and $X_{i,t}$ have a bite.
  - At this point, seems central for Proposition 1.
Conclusion

- Very elegant asymptotic theory establishing analytical characterization of expected returns, volatilities and Sharpe ratios, in a setting with a single risky asset.

- Extends recent double-descent literature to portfolio timing
  - Both fully specified and mis-specified settings considered.

- Highlights virtues of complexity and machine learning for portfolio timing problems.

- Empirical exercise is a nice validation of the theory
  - The ML strategy somehow able to identify recessions ex-ante.