Wharton
UNIVERSITY of PENNSYLVANIA

JACOBS LEVY EQUITY MANAGEMENT CENTER
for Quantitative Financial Research

Discussion: The Virtue of Complexity in Return Prediction

Ron Kaniel, Simon Business School, University of Rochester

Virtue of Complexity

- 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:

$$R_{t+1} = S_t^T \beta + \varepsilon_{t+1},$$
$$with\ \varepsilon_{t+1}\ i.i.d.,\ E\left[\varepsilon_{t+1}\right] = E\left[\varepsilon_{t+1}^3\right] = 0,\ E\left[\varepsilon_{t+1}^2\right] = \sigma^2, E\left[\varepsilon_{t+1}^2\right] < \infty$$

State variables:

$$S_t = \Psi^{1/2} X_t,\ X_t \in R^P$$
$$where\ E\left[X_{i,t}\right] = E\left[X_{i,t}^3\right] = 0,\ E\left[X_{i,t}^4\right]\ uniformally\ bounded,$$
$$and\ X_{i,t}\ satisfy\ the\ Linberg\ condition,\ and\ \Psi\ positive\ semi-definite$$

Loadings:

$$\beta \in R^P\ random\ independant\ of\ S\ and\ R,\ rotationally\ symmetric...$$
$$and\ satisfy\ same\ Linberg\ condition\ as\ X$$

Timing Strategy:

$$R_{t+1}^{\pi} = \pi_t R_{t+1},\ \text{centering on the timing strategy}\ \pi_t = \beta^T S_t$$

Main Focus: Evaluation of $\mathrm{SR} = \dfrac{E\left[R_{t+1}^{\pi}\right]}{\sqrt{E\left[\left(R_{t+1}^{\pi}\right)^2\right]}}$

Estimating Method
(Representative Machine Learning Procedure)

- Ridge-regularized least squares

$$\hat{\beta}(z) = \left(zI + T^{-1} \sum_t S_t S_t' \right)^{-1} \frac{1}{T} \sum_t S_t R_{t+1}$$

- Ridgeless regression estimator

$$\hat{\beta}(0^+) = \lim_{z \to 0+} \left(zI + T^{-1} \sum_t S_t S_t' \right)^{-1} \frac{1}{T} \sum_t S_t R_{t+1}.$$

  - Equivalent to the solution

$$H \frac{1}{T} \sum_t S_t R_{t+1}, \textit{ where } H = \text{Moore-Penrose pseudo-inverse of } \mathrm{T}^{-1} \sum_t S_t S_t^T$$

    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

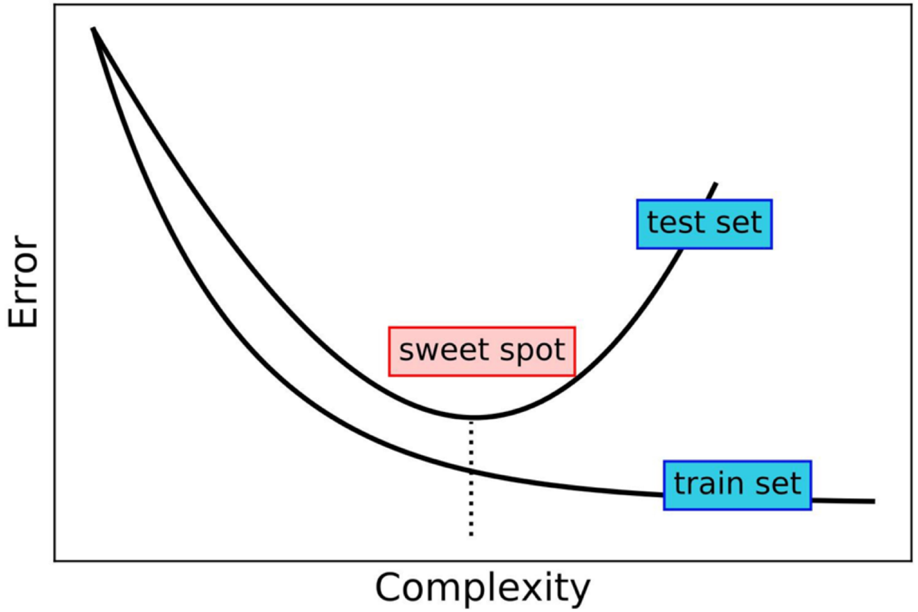
Mean Squared Error (MSE) – classical region

- Financial decomposition of MSE

$$MSE(\hat{\beta}) = E\left[\left(R_{t+1} - S_t'\hat{\beta}\right)^2 |\hat{\beta}\right] = E[R_{t+1}^2] - 2\underbrace{E[\hat{\pi}_t R_{t+1}|\hat{\beta}]}_{\substack{\text{Timing} \\ \text{Expected Return}}} + \underbrace{E[\hat{\pi}_t^2|\hat{\beta}]}_{\substack{\text{Timing} \\ \text{Leverage}}}.$$

- Classical statistical decomposition into bias/variance tradeoff:

$$MSE\left(\hat{\beta}\right) = \left(Bias\left(\hat{\beta}\right)\right)^2 + Var\left(\hat{\beta}\right)$$

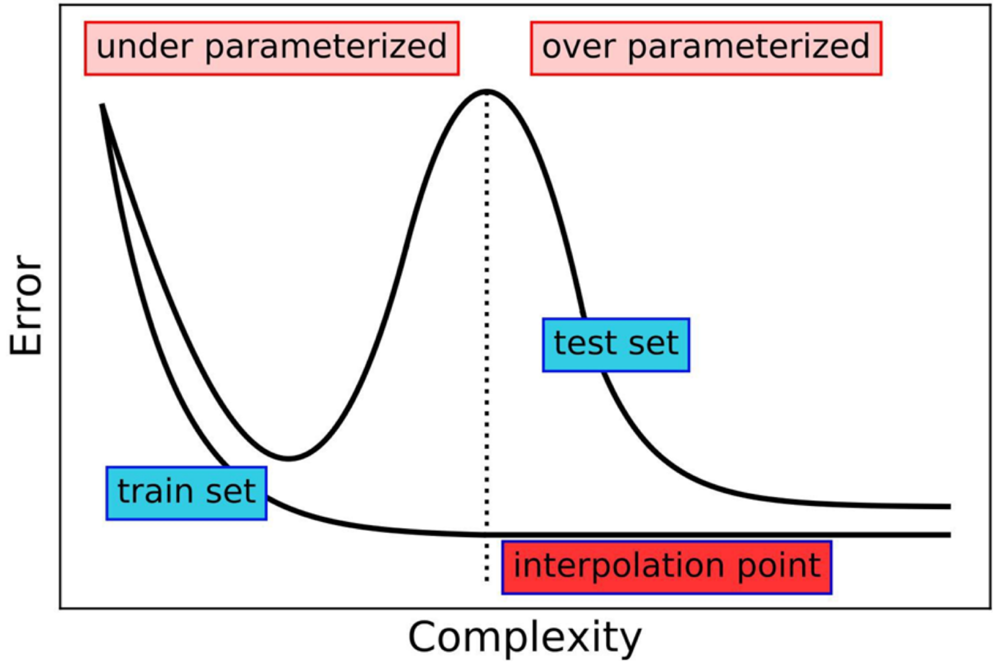


Machine Learning and Double Descent

- Measuring complexity (P- number of Parameters, T- number of periods):

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

$$MSE(z;c) = \lim_{T,P\to\infty,\ P/T\to c} E\left[\left(R_{t+1} - S_t'\hat{\beta}(z)\right)^2 |\hat{\beta}(z)\right]$$



R^2 and Leverage

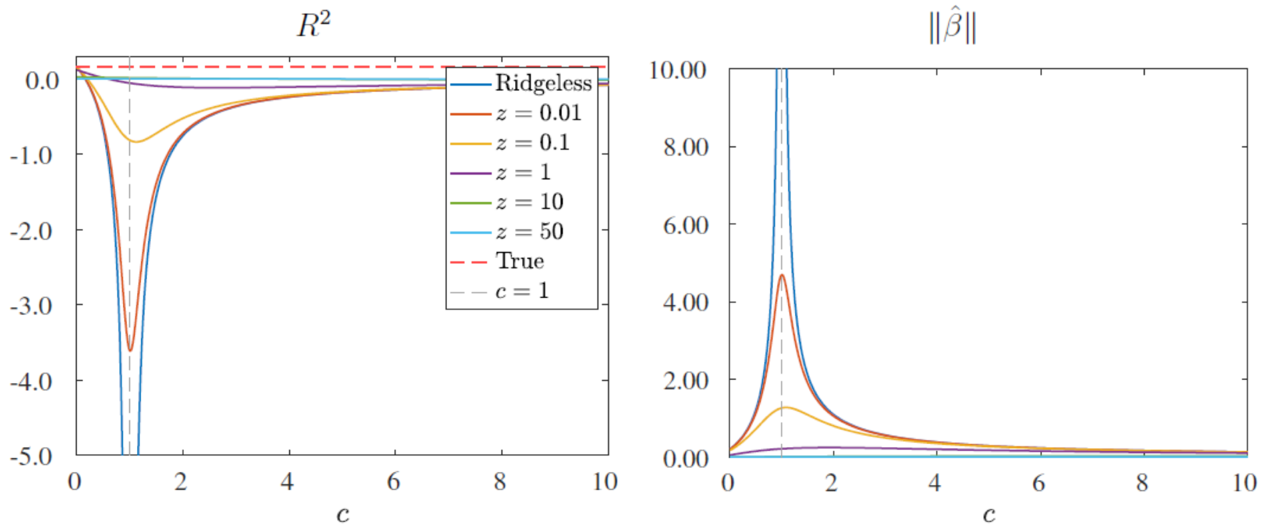


Figure 1: Expected Out-of-sample R^2 and Norm of Least Squares Coefficient

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

Fully Specified Model

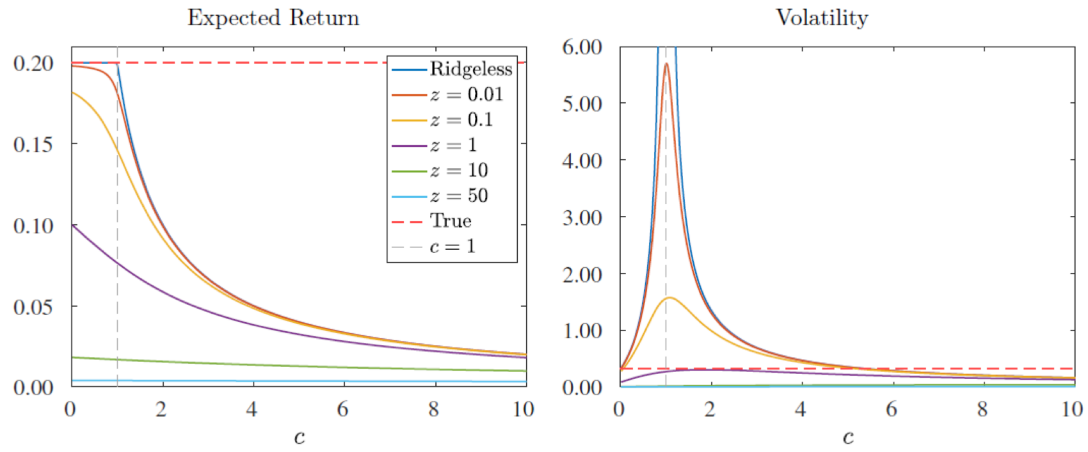


Figure 2: Expected Out-of-sample Risk and Return of Market Timing

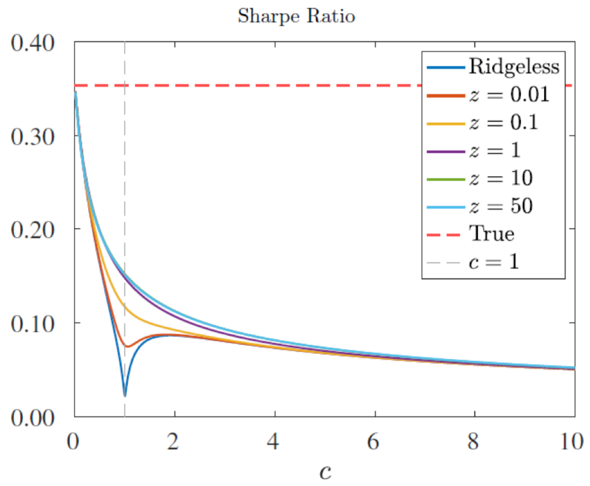


Figure 3: Expected Out-of-sample Sharpe Ratio of Market Timing

Mis-specified Model (observe only P_1 out of P parameters $\lim_{T\to\infty} P_1/P = q$)

- X-axis: holds c fixed and varies q

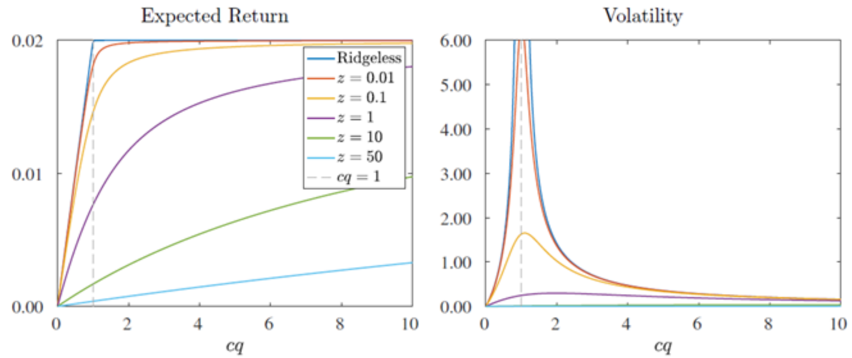


Figure 5: Expected Out-of-sample Timing Strategy Risk and Return From Mis-specified Models

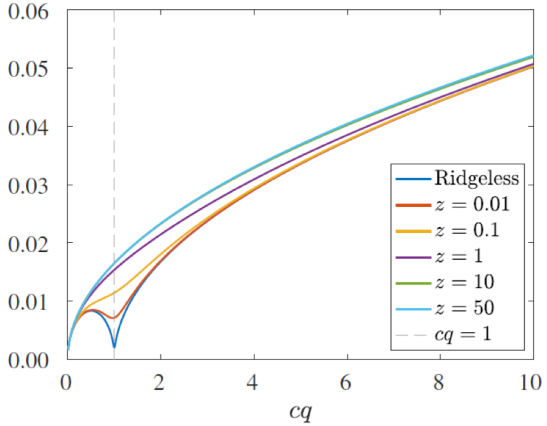


Figure 6: Expected Out-of-sample Timing Strategy Sharpe Ratio From Mis-specified Models

Empirical Exercise

- Validates the theory
- For high complexity models
  - Produces high out of sample monthly
    - Sharpe ratios (0.4),
    - Alphas (1%-2%),
    - Information ratios (0.25-0.3)
  - Strategy is long-only at heart: almost never bet on market downturns
  - Strategy learns to divest leading up to recessions (14/15 NBER recession dates)
    - Essentially zeros out positions
- Latter part of the sample:
  - lower average returns and information ratios, fewer buying recommendations, and smaller position.

Comments - Theory

- 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 ε_{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 Σ_w is an optimally chosen weighting matrix.

Comments - Analysis

- For mis-specified model consider adding graphs which do not hold *c* fixed ($c = \lim_{T,P\to\infty} P/T$, $q = \lim_{T\to\infty} P_1/P$)
  - Instead set *c=q*Cmax* 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 ε_{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.