Asset Pricing with Complexity

Mads Nielsen
University of Lausanne, Swiss Finance Institute

UNIL | Université de Lausanne
HEC Lausanne

s:fi

Abstract

Machine learning methods for big data trade off bias for precision in prediction. To understand the implications for financial markets, I formulate a trading model with a prediction technology where investors optimally choose a biased estimator. The model identifies a novel cost of complexity that arises endogenously. This effect makes it optimal to ignore costless signals and introduces in- and out-of-sample return predictability that is not driven by priced risk or behavioral biases. Empirically, the model can explain patterns of vanishing predictability of the equity risk premium. The model calibration is consistent with a technological shift following the rise of private computers and the invention of the internet. When allowing for heterogeneity in information between agents, complexity drives a wedge between the private and social value of data and lowers price informativeness. Estimation errors generate short-term price reversals similar to liquidity demand.

Prediction friction: Data complexity

Data complexity means data are:

- High dimensional
- From (partially) unknown data generating process

→ prediction function must be approximated and estimated

Main implication: generates sources of OOS predictability

- Optimal bias
- Cost of complexity

Predict pay-off of risky asset with structure

$$y = \boldsymbol{\beta}'\boldsymbol{q}$$

Choose estimator for loadings $\boldsymbol{\beta}$ to min MSE of predictor $\hat{y} = \widehat{\boldsymbol{\beta}}'\boldsymbol{\zeta}$

Equiv to trade-off between

- bias $\boldsymbol{\varepsilon_\beta} = \boldsymbol{f_\varepsilon(c)}$ and
- std. deviation $\boldsymbol{\sigma_\beta} = \boldsymbol{f_\sigma(c)}$

High-dim fundamentals $\boldsymbol{q}$ and signals $\boldsymbol{s}$ are gaussian with expectations

$$\boldsymbol{\zeta} = \boldsymbol{E}[\boldsymbol{q}|\boldsymbol{s}]$$

with $\boldsymbol{\Omega_\zeta} = \mathbf{E}[\boldsymbol{\zeta}\boldsymbol{\zeta}']$

and $\boldsymbol{D_{\Omega_\zeta}} = \mathrm{diag}(\boldsymbol{\Omega_\zeta})$

Minimization

$$\min_{\boldsymbol{c}} E[(y-\hat{y})^2] = \min_{\boldsymbol{c}} \boldsymbol{\varepsilon_\beta}'\boldsymbol{\Omega_\zeta}\boldsymbol{\varepsilon_\beta} + \boldsymbol{\sigma_\beta}'\boldsymbol{D_{\Omega_\zeta}}\boldsymbol{\sigma_\beta} + Var[y|\boldsymbol{\beta},\boldsymbol{s}]$$

Conditional variance under the true model is unaffected by estimator choice (irreducible noise).

Parametrization:

$\sigma_{\beta i} = f_\sigma(c_i) = k_{\sigma 0} + k_\sigma c_i$

$\varepsilon_{\beta i} = f_\varepsilon(c_i) = k_\varepsilon c_i$

$k_c = k_\sigma / k_\varepsilon$ is **est. tech. quality**

$k_{\sigma 0}$ is **base est. difficulty**

Let $\boldsymbol{X} = k_c^2\boldsymbol{\Omega_\zeta^{-1}} + \boldsymbol{D_{\Omega_\zeta}^{-1}}$

Optimal bias: $\boldsymbol{\varepsilon_\beta} = -\frac{k_{\sigma 0}}{k_c}\left\{\boldsymbol{I} - \boldsymbol{D_{\Omega_\zeta}^{-1}}\boldsymbol{X^{-1}}\right\}\mathbf{1} \geq \mathbf{0}$

Cost of complexity: $\chi = k_{\sigma 0}^2\mathbf{1}'\boldsymbol{X^{-1}}\mathbf{1}$

Equilibrium

Predictor represents investors' model of the world and enter portfolio optimization as beliefs, i.e. they are taken as given. Linear demand derived from robust profit maximization objective or CARA-utility with ambiguity aversion with uncertainty aversion α_i

$$\delta_i = \psi_i(\hat{y}_i - p), \text{ where } \psi_i = \{\alpha_i E[(y-\hat{y}_i)^2]\}^{-1}$$

and market clearing yields price:

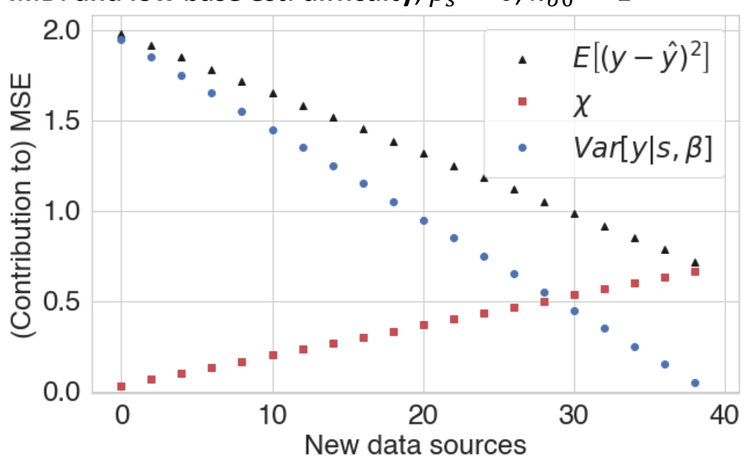
- Representative agent model
$$p = \hat{y}$$
- Adapted Grossman & Stiglitz (1980) where only informed investors I solve prediction problem and uninformed investors U mimic by reacting to price facing stochastic supply z
$$p = (1-\lambda_p)E[\hat{y}_I] + \lambda_p(\hat{y}_I - \psi_I^{-1}z)$$

where $\lambda_p = \frac{\psi_I + \lambda_U\psi_U}{\psi_I + \psi_U} \leq 1$ since $\lambda_U = \frac{Var[\hat{y}_I]}{Var[\hat{y}_I] + \psi_I^{-2}Var[z]} \leq 1$

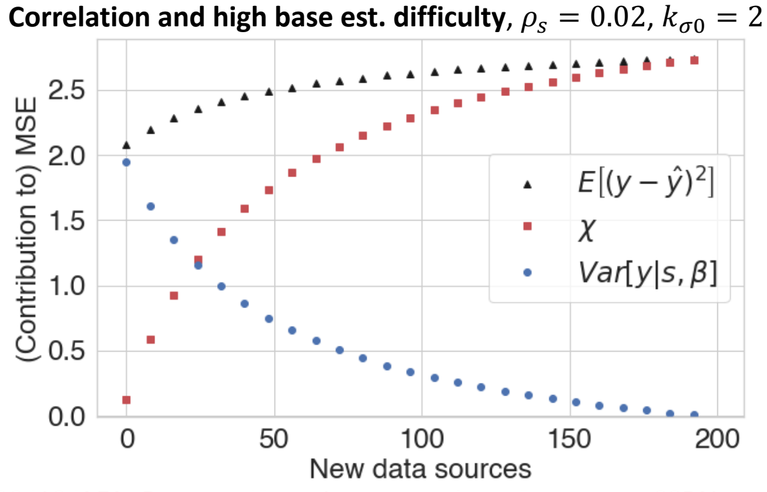
Cost of complexity

The cost of complexity can be so high that it is better to ignore a data source and avoid the extra degree of complexity. In addition to estimation technology quality, cost of complexity depends on fundamentals and the base estimation difficulty and as such vary across assets. Illustrated below by two examples of adding data sources (I.I.D. and shared correlation parameter) and different base estimation difficulty.

I.I.D. and low base est. difficulty, $\rho_s = 0$, $k_{\sigma 0} = 1$



Correlation and high base est. difficulty, $\rho_s = 0.02$, $k_{\sigma 0} = 2$



Return predictability

Returns are $r = y - p$. Consider econometricians analyzing returns after the fact with better estimation technology

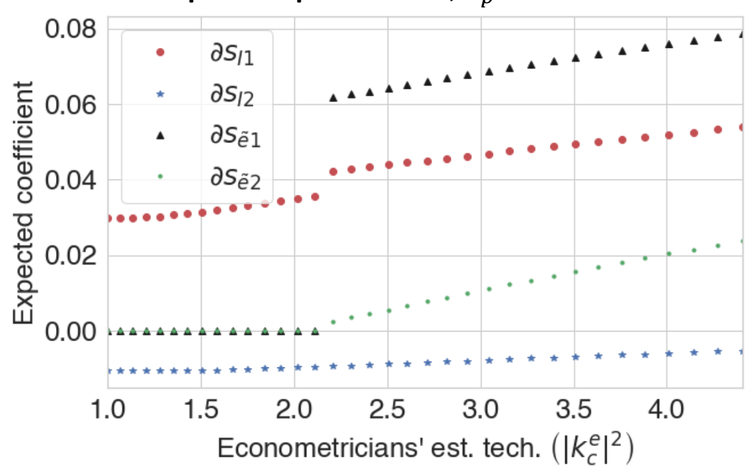
$$|k_c^e| > |k_c^I|$$

There is a gap between optimal biases and investors might ignore data sources due to cost of complexity that are feasible for econometricians to include in their projection.

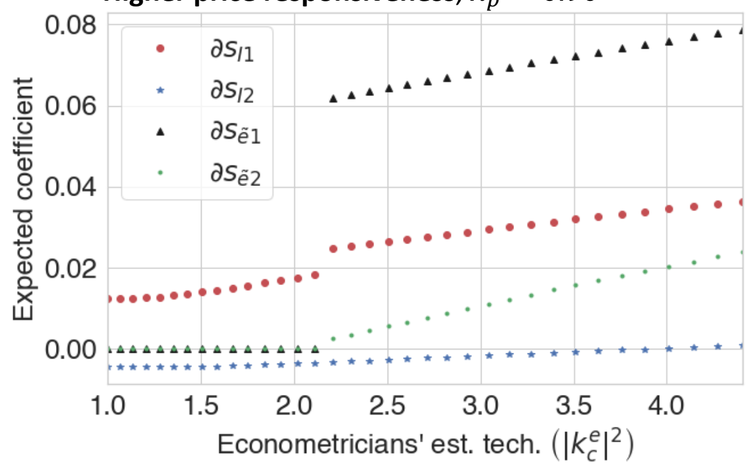
The two sources of predictability can be distinguished through variation in price responsiveness.

Predictability and variability are illustrated below through the expected coefficients of econometricians projection in a simple set-up: 2 factors, 4 signals where 2 are used by investors (s_{I1}, s_{I2}) and 2 are ignored ($s_{\tilde{e}1}$, $s_{\tilde{e}2}$).

Lower price responsiveness, $\lambda_p \approx 0.75$



Higher price responsiveness, $\lambda_p \approx 0.90$



Perspective

Can machine learning explain the factor zoo?

Yes, to the extent factors really reflect differences in estimation methods. However, more generally, no, or at least not on its own. This is because OOS predictability is insufficient to draw conclusions about asset pricing models.

References

Grossman, S. J., Stiglitz, J. E., 1980. On the Impossibility of Informationally Efficient Markets. The American Economic Review 70, 393–408.

Latest version of poster on: https://bibow.dk/wp-content/uploads/2022/12/AFA_Poster.pdf Latest version of paper on SSRN