UNIL | Université de Lausanne
HEC Lausanne

Asset pricing with complexity

Mads Nielsen †

December 31, 2022

† Department of Finance
University of Lausanne (SFI)

Why does machine learning work for return predictability?

... and what does it mean for financial markets?

Why does machine learning work for return predictability?

... and what does it mean for financial markets?

Outline

1. Motivation: Better predictions under complexity.
2. Mechanism: Function approximation as prediction friction.
3. Main result: OOS return predictability.
4. More: Patterns in equity risk premium predictability.

Motivation: Better predictions under complexity.

Machine learning works for return predictability

Empirical literature (Gagliardini and Ma, 2019; Gu, Kelly, and Xiu, 2020, 2021; Ma, 2021)

Machine learning works for return predictability

Empirical literature (Gagliardini and Ma, 2019; Gu et al., 2020, 2021; Ma, 2021)

Table 1: Predicting individual stocks in Gu et al. (2020).

	Curated OLS benchmark	Principal component reg.	Neural net
Predictors	3	900+	900+
Monthly OOS R^2	0.16%	0.26%	0.40%

$\longrightarrow$ better return predictions under complexity (i.e. partially unknown and high dimensional environment).

Machine learning works for return predictability

In markets

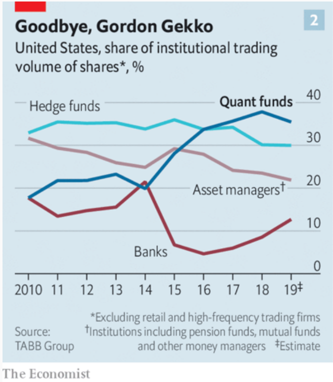

Figure 1: ”The stockmarket is now run by computers, algorithms and passive managers”, Economist, 2019.

Gap in theoretical literature focused on big data rather than ML

Learning in financial markets (Lewellen and Shanken, 2002)

- Parameter uncertainty.
- Return predictability: conditional vs unconditional moments.

Gap in theoretical literature focused on big data rather than ML

Learning in financial markets (Lewellen and Shanken, 2002)

- Parameter uncertainty.
- Return predictability: conditional vs unconditional moments.

Big data in financial markets

- Supply and demand for data (Dessaint et al., 2020; Dugast and Foucault, 2020; Farboodi et al., 2020; Farboodi and Veldkamp, 2020).
- Parameter uncertainty high dimensionality (Martin and Nagel, 2021).

Gap in theoretical literature focused on big data rather than ML

Missing features

(i) "Let the data speak": true model unknown $\rightarrow$ function approximation.

Gap in theoretical literature focused on big data rather than ML

Missing features

(i) "Let the data speak": true model unknown $\rightarrow$ function approximation.

(ii) Over-fitting vs under-fitting: bias-variance trade-off $\rightarrow$ *optimal* bias.

Gap in theoretical literature focused on big data rather than ML

Missing features

(i) "Let the data speak": true model unknown $\rightarrow$ function approximation.

(ii) Over-fitting vs under-fitting: bias-variance trade-off $\rightarrow$ *optimal* bias.

(iii) Technology as cure for curse of dimensionality: Cost of complexity.

Gap in theoretical literature focused on big data rather than ML

Missing features

(i) "Let the data speak": true model unknown $\rightarrow$ function approximation.

(ii) Over-fitting vs under-fitting: bias-variance trade-off $\rightarrow$ *optimal* bias.

(iii) Technology as cure for curse of dimensionality: Cost of complexity.

Related work by Kelly et al. (2022) focuses on the virtue of complex models.

What I do in this paper

1) Formalize function approximation as a prediction friction:

What I do in this paper

1) Formalize function approximation as a prediction friction:
    (ii) Non-zero optimal bias.

What I do in this paper

1) Formalize function approximation as a prediction friction:
   (ii) Non-zero optimal bias.
   (iii) Endogenous cost of complexity.

What I do in this paper

1) Formalize function approximation as a prediction friction:
   (ii) Non-zero optimal bias.
   (iii) Endogenous cost of complexity.
2) Embed in models of trading, impact on measures of market efficiency in equilibrium.

What I do in this paper

1) Formalize function approximation as a prediction friction:
   (ii) Non-zero optimal bias.
   (iii) Endogenous cost of complexity.
2) Embed in models of trading, impact on measures of market efficiency in equilibrium.
3) Find limits to interpretability of OOS return predictability, additional variation required.

Mechanism: Function approximation as prediction friction.

MIRROR STRUCTURE IN EMPERICAL APPLICATIONS OF ML

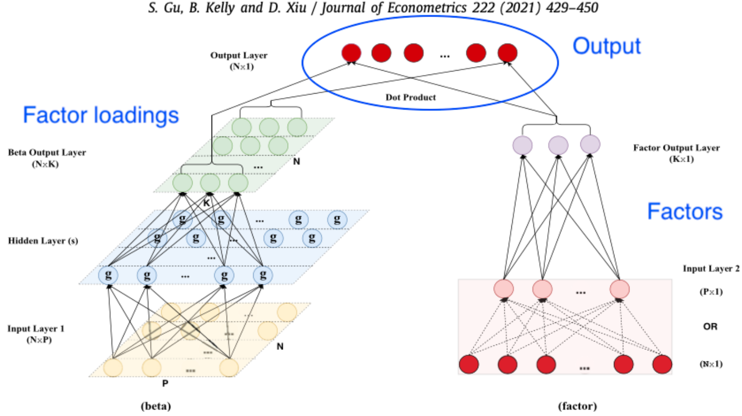

Figure 2: Figure 2 from Gu et al. (2021) with my highlights. Estimation of factors and factors loadings are separated in to two sub-problems connected by the interaction in the dot product.

DECOMPOSE COMPLEXITY INTO UNKNOWN MODEL AND HIGH-DIMENSIONALITY

Pay-off y, factors q, factor loadings β, and cond. expectation given signals ζ

$$y = \boldsymbol{\beta}^\top \boldsymbol{q}, \quad \boldsymbol{q} \sim \mathcal{N}(\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_q), \quad \boldsymbol{\zeta} = E[\boldsymbol{q}|\boldsymbol{s}] \sim \mathcal{N}(\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_\zeta), \text{ and } \boldsymbol{\Omega}_\zeta = E[\boldsymbol{\zeta}\boldsymbol{\zeta}^\top].$$

Investors must *estimate* $\hat{\beta}$ from noisy data.

DECOMPOSE COMPLEXITY INTO UNKNOWN MODEL AND HIGH-DIMENSIONALITY

Pay-off y, factors q, factor loadings β, and cond. expectation given signals ζ

$$y = \boldsymbol{\beta}^\top \boldsymbol{q}, \quad \boldsymbol{q} \sim \mathcal{N}(\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_q), \quad \boldsymbol{\zeta} = E[\boldsymbol{q}|\boldsymbol{s}] \sim \mathcal{N}(\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_\zeta), \text{ and } \boldsymbol{\Omega}_\zeta = E[\boldsymbol{\zeta}\boldsymbol{\zeta}^\top].$$

Investors must *estimate* $\hat{\beta}$ from noisy data.

Choose controls c: bias $\boldsymbol{\varepsilon}_\beta = f_\varepsilon(\boldsymbol{c})$ and vol $\boldsymbol{\sigma}_\beta = f_\sigma(\boldsymbol{c})$ to min mse of predictor $\hat{y} = \hat{\boldsymbol{\beta}}^\top \boldsymbol{\zeta}$

$$\min_{\boldsymbol{c}} E[\{y - \hat{\boldsymbol{\beta}}(\boldsymbol{c})^\top \boldsymbol{\zeta}\}^2] = \min_{\boldsymbol{c}} \underbrace{\boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\Omega}_\zeta \boldsymbol{\varepsilon}_\beta}_{\text{Bias squared}} + \underbrace{\boldsymbol{\sigma}_\beta^\top (\boldsymbol{R}_\beta \odot \boldsymbol{\Omega}_\zeta) \boldsymbol{\sigma}_\beta}_{\text{Variance}} + \underbrace{Var[y|\boldsymbol{\beta}, \boldsymbol{s}]}_{\text{Irreducible noise}}$$

$$s.t. \quad f'_\varepsilon(c_i) f'_\sigma(c_i) < 0, f_\sigma(c_i) > 0 \; \forall c_i \in \boldsymbol{c}.$$

Decompose complexity into unknown model and high-dimensionality

Pay-off y, factors q, factor loadings β, and cond. expectation given signals ζ

$$y = \boldsymbol{\beta}^\top \boldsymbol{q}, \quad \boldsymbol{q} \sim \mathcal{N}(\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_q), \quad \boldsymbol{\zeta} = E[\boldsymbol{q}|\boldsymbol{s}] \sim \mathcal{N}(\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_\zeta), \text{ and } \boldsymbol{\Omega}_\zeta = E[\boldsymbol{\zeta}\boldsymbol{\zeta}^\top].$$

Investors must *estimate* $\hat{\beta}$ from noisy data.

Choose controls c: bias $\boldsymbol{\varepsilon}_\beta = f_\varepsilon(\boldsymbol{c})$ and vol $\boldsymbol{\sigma}_\beta = f_\sigma(\boldsymbol{c})$ to min mse of predictor $\hat{y} = \hat{\boldsymbol{\beta}}^\top \boldsymbol{\zeta}$

$$\min_{\boldsymbol{c}} E[\{y - \hat{\boldsymbol{\beta}}(\boldsymbol{c})^\top \boldsymbol{\zeta}\}^2] = \min_{\boldsymbol{c}} \underbrace{\boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\Omega}_\zeta \boldsymbol{\varepsilon}_\beta}_{\text{Bias squared}} + \underbrace{\boldsymbol{\sigma}_\beta^\top \boldsymbol{D}_{\boldsymbol{\Omega}_\zeta} \boldsymbol{\sigma}_\beta}_{\text{Variance}} + \underbrace{Var[y|\boldsymbol{\beta}, \boldsymbol{s}]}_{\text{Irreducible noise}}$$

$$s.t. \quad f'_\varepsilon(c_i) f'_\sigma(c_i) < 0, f_\sigma(c_i) > 0 \; \forall c_i \in \boldsymbol{c}.$$

Decompose complexity into unknown model and high-dimensionality

Pay-off y, factors q, factor loadings β, and cond. expectation given signals ζ

$$y = \boldsymbol{\beta}^\top \boldsymbol{q}, \quad \boldsymbol{q} \sim \mathcal{N}(\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_q), \quad \boldsymbol{\zeta} = E[\boldsymbol{q}|\boldsymbol{s}] \sim \mathcal{N}(\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_\zeta), \text{ and } \boldsymbol{\Omega}_\zeta = E[\boldsymbol{\zeta}\boldsymbol{\zeta}^\top].$$

Investors must *estimate* $\hat{\beta}$ from noisy data.

Choose controls c: bias $\boldsymbol{\varepsilon}_\beta = f_\varepsilon(\boldsymbol{c})$ and vol $\boldsymbol{\sigma}_\beta = f_\sigma(\boldsymbol{c})$ to min mse of predictor $\hat{y} = \hat{\boldsymbol{\beta}}^\top \boldsymbol{\zeta}$

$$\min_{\boldsymbol{c}} E[\{y - \hat{\boldsymbol{\beta}}(\boldsymbol{c})^\top \boldsymbol{\zeta}\}^2] = \min_{\boldsymbol{c}} \underbrace{\boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\Omega}_\zeta \boldsymbol{\varepsilon}_\beta}_{\text{Bias squared}} + \underbrace{\boldsymbol{\sigma}_\beta^\top \boldsymbol{D}_{\boldsymbol{\Omega}_\zeta} \boldsymbol{\sigma}_\beta}_{\text{Variance}} + \underbrace{Var[y|\boldsymbol{\beta}, \boldsymbol{s}]}_{\text{Irreducible noise}}$$

$$s.t. \quad f'_\varepsilon(c_i) f'_\sigma(c_i) < 0, f_\sigma(c_i) > 0 \; \forall c_i \in \boldsymbol{c}.$$

Linear-affine functions $f_\varepsilon(c_i) = k_\varepsilon c_i, f_\sigma(c_i) = k_{\sigma 0} + k_\sigma c_i$

$\longrightarrow$ unique solution exists under the feasibility constraint $\boldsymbol{\Omega}_\zeta \mathbf{1} > \mathbf{0}$.

Explicit solutions for non-zero optimal bias and cost of complexity

Minimized mse as cost of complexity χ vs conditional variance under true model

$$\min_{\boldsymbol{c}} \underbrace{\boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\Omega}_\zeta \boldsymbol{\varepsilon}_\beta}_{\text{Bias squared}} + \underbrace{\boldsymbol{\sigma}_\beta^\top \boldsymbol{D}_{\boldsymbol{\Omega}_\zeta} \boldsymbol{\sigma}_\beta}_{\text{Variance}} + \underbrace{Var[y|\boldsymbol{\beta}, \boldsymbol{s}]}_{\text{Irreducible noise}} := \underbrace{\chi(\boldsymbol{c}^*)}_{\text{cost of complexity}} + \underbrace{Var[y|\boldsymbol{\beta}, \boldsymbol{s}]}_{\text{cond var true model}},$$

$$\chi = \underbrace{k_{\sigma 0}^2}_{\text{est. difficulty}} \mathbf{1}^\top \boldsymbol{X}^{-1} \mathbf{1}, \text{ where } \boldsymbol{X} = \underbrace{k_c^2}_{\text{est. tech quality}} \boldsymbol{\Omega}_\zeta^{-1} + \boldsymbol{D}_{\Omega_\zeta}^{-1} \text{ and } k_c = k_\sigma / k_\varepsilon$$

Explicit solutions for non-zero optimal bias and cost of complexity

Minimized mse as cost of complexity χ vs conditional variance under true model

$$\min_{\boldsymbol{c}} \underbrace{\boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\Omega}_\zeta \boldsymbol{\varepsilon}_\beta}_{\text{Bias squared}} + \underbrace{\boldsymbol{\sigma}_\beta^\top \boldsymbol{D}_{\boldsymbol{\Omega}_\zeta} \boldsymbol{\sigma}_\beta}_{\text{Variance}} + \underbrace{Var[y|\boldsymbol{\beta}, \boldsymbol{s}]}_{\text{Irreducible noise}} := \underbrace{\chi(\boldsymbol{c}^*)}_{\text{cost of complexity}} + \underbrace{Var[y|\boldsymbol{\beta}, \boldsymbol{s}]}_{\text{cond var true model}},$$

$$\chi = \underbrace{k_{\sigma 0}^2}_{\text{est. difficulty}} \mathbf{1}^\top \boldsymbol{X}^{-1} \mathbf{1}, \text{ where } \boldsymbol{X} = \underbrace{k_c^2}_{\text{est. tech quality}} \boldsymbol{\Omega}_\zeta^{-1} + \boldsymbol{D}_{\Omega_\zeta}^{-1} \text{ and } k_c = k_\sigma / k_\varepsilon$$

Optimal bias $\varepsilon_\beta|_{\boldsymbol{c}=\boldsymbol{c}^*} = -k_c^{-1} k_{\sigma 0} \left\{ \boldsymbol{I} - \boldsymbol{D}_{\Omega_\zeta}^{-1} \boldsymbol{X}^{-1} \right\} \mathbf{1} \geq \mathbf{0}$, only approx zero as $k_c \to 0$.

EXPLICIT SOLUTIONS FOR NON-ZERO OPTIMAL BIAS AND COST OF COMPLEXITY

Minimized mse as cost of complexity χ vs conditional variance under true model

$$\min_{\boldsymbol{c}} \underbrace{\boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\Omega}_\zeta \boldsymbol{\varepsilon}_\beta}_{\text{Bias squared}} + \underbrace{\boldsymbol{\sigma}_\beta^\top \boldsymbol{D}_{\boldsymbol{\Omega}_\zeta} \boldsymbol{\sigma}_\beta}_{\text{Variance}} + \underbrace{Var[y|\boldsymbol{\beta}, \boldsymbol{s}]}_{\text{Irreducible noise}} := \underbrace{\chi(\boldsymbol{c}^*)}_{\text{cost of complexity}} + \underbrace{Var[y|\boldsymbol{\beta}, \boldsymbol{s}]}_{\text{cond var true model}},$$

$$\chi = \underbrace{k_{\sigma 0}^2}_{\text{est. difficulty}} \mathbf{1}^\top \boldsymbol{X}^{-1} \mathbf{1}, \text{ where } \boldsymbol{X} = \underbrace{k_c^2}_{\text{est. tech quality}} \boldsymbol{\Omega}_\zeta^{-1} + \boldsymbol{D}_{\Omega_\zeta}^{-1} \text{ and } k_c = k_\sigma / k_\varepsilon$$

Optimal bias $\varepsilon_\beta|_{\boldsymbol{c}=\boldsymbol{c}^*} = -k_c^{-1} k_{\sigma 0} \left\{ \boldsymbol{I} - \boldsymbol{D}_{\Omega_\zeta}^{-1} \boldsymbol{X}^{-1} \right\} \mathbf{1} \geq \mathbf{0}$, only approx zero as $k_c \to 0$.

Cost of complexity increases in the number of signals/data sources n_s

$$\chi_{n_s} \geq \chi_{n_s - 1}, \quad \text{and} \quad Var[y|\boldsymbol{\beta}, \boldsymbol{s}_{n_s}] \leq Var[y|\boldsymbol{\beta}, \boldsymbol{s}_{n_s - 1}].$$

Value of more data depends on relative increase in cost of complexity

New data sources parametrized by k_S in $\Omega_\zeta = \Omega_{\zeta 0} + k_S S$.

(a) Easier estimation (baseline) $k_{\sigma 0} = 0.3$ **(b)** Harder estimation (baseline) $k_{\sigma 0} = 0.6$

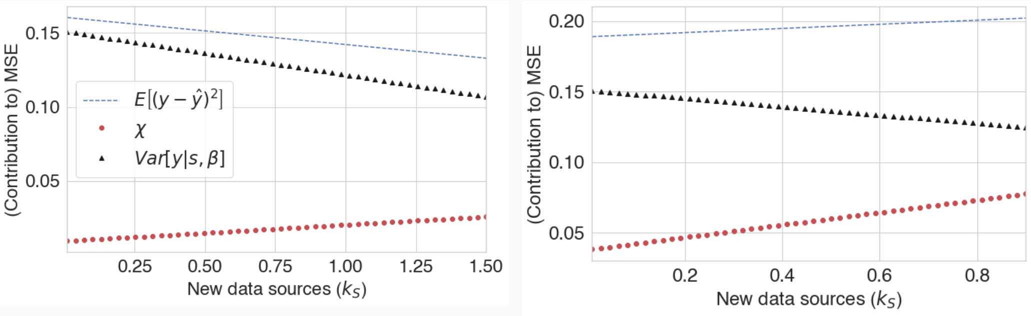

Figure 3: Mean squared error of predictor decreasing or increasing in addition of new data sources.

Main result: OOS return predictability.

Two sources of predictability distinguished through heterogeneity

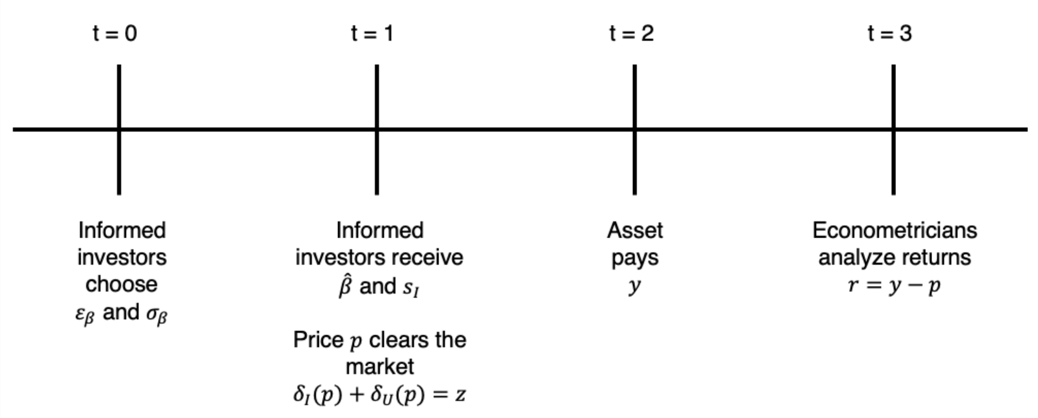

Figure 4: Time-line for predictions of returns generated by adapted Grossman and Stiglitz (1980).

Two sources of predictability distinguished through heterogeneity

Returns in adapted Grossman and Stiglitz (1980)

$$r = y - p = (1-\lambda_p)\underbrace{(y - E[\hat{y}_I])}_{\text{uninformed}} + \lambda_p \underbrace{(y - \hat{y}_I)}_{\text{informed}} + \lambda_p \underbrace{\psi_I^{-1} z}_{\text{stoch. supply}},$$

$\lambda_p \leq 1$: price responsiveness, ψ_I : informed investors' aggressiveness

Assume $|k_c^I| < |k_c^e|$ such that $s_I \subseteq s_e$, only used by econometrician is $\tilde{s}_e$.

TWO SOURCES OF PREDICTABILITY DISTINGUISHED THROUGH HETEROGENEITY

Returns in adapted Grossman and Stiglitz (1980), returns in representative agent model

$$r = y - p = (1-\lambda_p)\underbrace{(y - E[\hat{y}_I])}_{\text{uninformed}} + \lambda_p \underbrace{(y - \hat{y}_I)}_{\text{informed}} + \lambda_p \underbrace{\psi_I^{-1} z}_{\text{stoch. supply}},$$

$$\lambda_p \leq 1: \text{ price responsiveness, } \psi_I: \text{ informed investors' aggressiveness}$$

Assume $|k_c^I| < |k_c^e|$ such that $s_I \subseteq s_e$, only used by econometrician is $\tilde{s}_e$.

Two sources of predictability distinguished through heterogeneity

Returns in adapted Grossman and Stiglitz (1980)

$$r = y - p = (1 - \lambda_p)\underbrace{(y - E[\hat{y}_I])}_{\text{uninformed}} + \lambda_p \underbrace{(y - \hat{y}_I)}_{\text{informed}} + \lambda_p \underbrace{\psi_I^{-1} z}_{\text{stoch. supply}},$$

$\lambda_p \leq 1$: price responsiveness, ψ_I : informed investors' aggressiveness

Assume $|k_c^I| < |k_c^e|$ such that $\boldsymbol{s}_I \subseteq \boldsymbol{s}_e$, only used by econometrician is $\tilde{\boldsymbol{s}}_e$.

Econometricians' expected projection

$$E\left[\tilde{E}[r|k_c^e, \boldsymbol{s}_e]\middle| \boldsymbol{s}_e\right] = (\boldsymbol{\varepsilon}_\beta - \boldsymbol{\varepsilon}_{\beta e})^\top \boldsymbol{\mu}_q + \left\{\lambda_p \boldsymbol{\varepsilon}_\beta - \boldsymbol{\varepsilon}_{\beta e} + (1 - \lambda_p)\boldsymbol{\beta}\right\}^\top \boldsymbol{\Lambda}_I(\boldsymbol{s}_I - E[\boldsymbol{s}_I])$$
$$+ (\boldsymbol{\beta} - \boldsymbol{\varepsilon}_{\beta e})^\top \boldsymbol{\Lambda}_{\tilde{e}}(\boldsymbol{s}_{\tilde{e}} - E[\boldsymbol{s}_{\tilde{e}}]) \quad \text{where} \quad \boldsymbol{\Lambda}_\ell = \partial E[\boldsymbol{q}|\boldsymbol{s}_\ell]/\partial \boldsymbol{s}_\ell.$$

Two sources of predictability distinguished through heterogeneity

Returns in adapted Grossman and Stiglitz (1980)

$$r = y - p = (1-\lambda_p)\underbrace{(y - E[\hat{y}_I])}_{\text{uninformed}} + \lambda_p\underbrace{(y - \hat{y}_I)}_{\text{informed}} + \lambda_p \underbrace{\psi_I^{-1} z}_{\text{stoch. supply}},$$

$$\lambda_p \leq 1: \text{ price responsiveness}, \psi_I: \text{ informed investors' aggressiveness}$$

Assume $|k_c^I| < |k_c^e|$ such that $\boldsymbol{s}_I \subseteq \boldsymbol{s}_e$, only used by econometrician is $\tilde{\boldsymbol{s}}_e$.

Econometricians' expected projection

$$E\left[\tilde{E}[r|k_c^e, \boldsymbol{s}_e]\middle| \boldsymbol{s}_e\right] = (\boldsymbol{\varepsilon}_\beta - \boldsymbol{\varepsilon}_{\beta e})^\top \boldsymbol{\mu}_q + \left\{\lambda_p \boldsymbol{\varepsilon}_\beta - \boldsymbol{\varepsilon}_{\beta e} + (1-\lambda_p)\boldsymbol{\beta}\right\}^\top \boldsymbol{\Lambda}_I(\boldsymbol{s}_I - E[\boldsymbol{s}_I])$$
$$+ (\boldsymbol{\beta} - \boldsymbol{\varepsilon}_{\beta e})^\top \boldsymbol{\Lambda}_{\tilde{e}}(\boldsymbol{s}_{\tilde{e}} - E[\boldsymbol{s}_{\tilde{e}}]) \quad \text{where} \quad \boldsymbol{\Lambda}_\ell = \partial E[\boldsymbol{q}|\boldsymbol{s}_\ell]/\partial \boldsymbol{s}_\ell.$$

Differences in non-zero optimal bias,

Two sources of predictability distinguished through heterogeneity

Returns in adapted Grossman and Stiglitz (1980)

$$r = y - p = (1-\lambda_p)\underbrace{(y - E[\hat{y}_I])}_{\text{uninformed}} + \lambda_p \underbrace{(y - \hat{y}_I)}_{\text{informed}} + \lambda_p \underbrace{\psi_I^{-1} z}_{\text{stoch. supply}},$$

$\lambda_p \leq 1$: price responsiveness, ψ_I : informed investors' aggressiveness

Assume $|k_c^I| < |k_c^e|$ such that $s_I \subseteq s_e$, only used by econometrician is $\tilde{s}_e$.

Econometricians' expected projection

$$E\left[\tilde{E}[r|k_c^e, s_e]\middle| s_e\right] = (\varepsilon_\beta - \varepsilon_{\beta e})^\top \mu_q + \left\{\lambda_p \varepsilon_\beta - \varepsilon_{\beta e} + (1-\lambda_p)\beta\right\}^\top \Lambda_I (s_I - E[s_I])$$
$$+ (\beta - \varepsilon_{\beta e})^\top \Lambda_{\tilde{e}}(s_{\tilde{e}} - E[s_{\tilde{e}}]) \quad \text{where} \quad \Lambda_\ell = \partial E[q|s_\ell]/\partial s_\ell.$$

Differences in non-zero optimal bias, lower cost of complexity.

Two sources of predictability distinguished through heterogeneity

Returns in adapted Grossman and Stiglitz (1980)

$$r = y - p = (1-\lambda_p)\underbrace{(y - E[\hat{y}_I])}_{\text{uninformed}} + \lambda_p \underbrace{(y - \hat{y}_I)}_{\text{informed}} + \lambda_p \underbrace{\psi_I^{-1} z}_{\text{stoch. supply}},$$

$\lambda_p \leq 1:$ price responsiveness, $\psi_I:$ informed investors' aggressiveness

Assume $|k_c^I| < |k_c^e|$ such that $s_I \subseteq s_e$, only used by econometrician is $\tilde{s}_e$.

Econometricians' expected projection

$$E\left[\tilde{E}[r|k_c^e, s_e]\Big| s_e\right] = (\varepsilon_\beta - \varepsilon_{\beta e})^\top \mu_q + \left\{\lambda_p \varepsilon_\beta - \varepsilon_{\beta e} + (1-\lambda_p)\beta\right\}^\top \Lambda_I (s_I - E[s_I]) + (\beta - \varepsilon_{\beta e})^\top \Lambda_{\tilde{e}}(s_{\tilde{e}} - E[s_{\tilde{e}}]) \quad \text{where} \quad \Lambda_\ell = \partial E[q|s_\ell]/\partial s_\ell.$$

Differences in non-zero optimal bias, lower cost of complexity. Variation in λ_p.

Improving est. tech. lowers cost of complexity → signal inclusion

Set-up: 2 factors, 4 signals: 2 used by investors (s_{I1}, s_{I2}), 2 ignored $(s_{\tilde{e}1}, s_{\tilde{e}2})$.

(a) Lower price responsiveness, $\lambda_p \approx 0.75$ **(b)** Higher price responsiveness, $\lambda_p \approx 0.90$

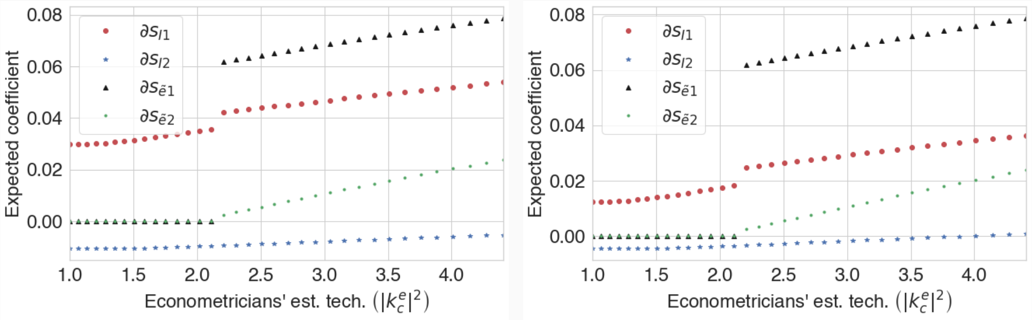

Figure 4: Expected coefficients of econometricians' projection.

UNCONDITIONAL EXP. RETURNS INCREASING IN DIFFERENCE IN OPTIMAL BIAS

(a) Econometricians' bias

(b) Projection constant

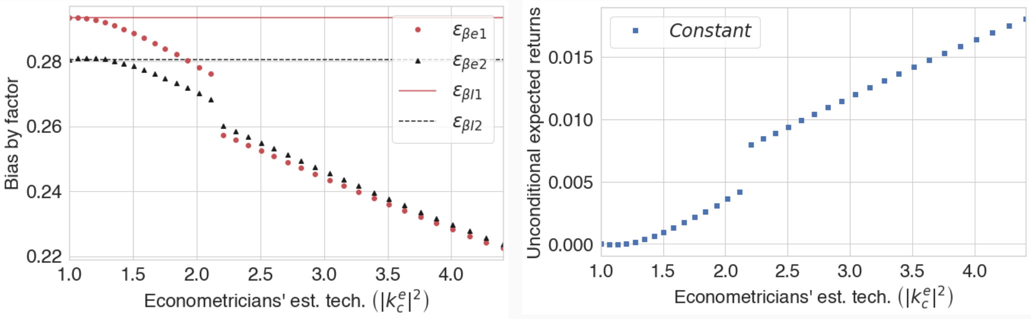

Figure 5: Bias and unconditional expected returns over econometricians' estimation technology.

More: Patterns in equity risk premium predictability.

Predictive out-performance followed by under-performance

Match pattern by calibrating change in $|k_c^I|$ between the two periods.

Result: $|k_{c2}^I|/|k_{c1}^I| - 1 \approx 233\%$ and $\varepsilon_{\beta i,2}/\varepsilon_{\beta i,1} - 1 \approx 82\%\ \forall i$, increasing bias.

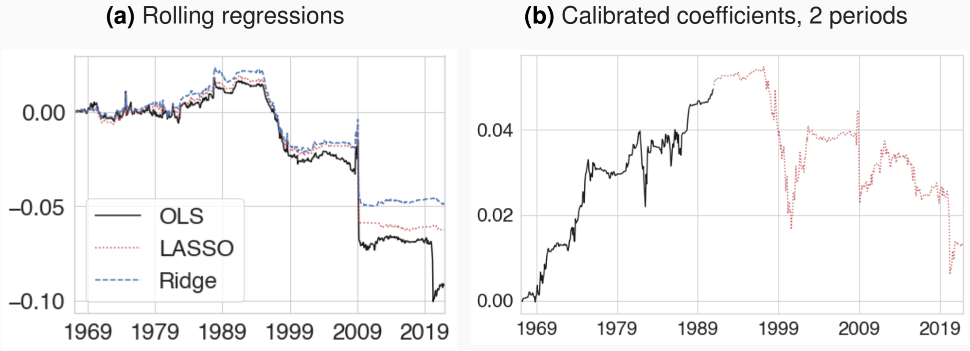

Figure 6: Ten predictors from Welch and Goyal (2008), updated data.

Conclusion

Complexity is missing in standard framework of learning in financial markets.

Function approximation as a prediction friction generates missing features:

- Optimal bias.
- Cost of complexity.

OOS return predictability is not sufficient to draw conclusions about asset pricing models.

Other measures of market efficiency

Prediction friction covers missing features

(i) Pay-off as dot-product with unknown factor loadings on well-behaved factors

$$\underbrace{\hat{y}}_{\text{prediction}} = \underbrace{\hat{\boldsymbol{\beta}}^\top}_{\text{estimate}} \zeta \quad \text{where} \quad \zeta = E[\boldsymbol{q}|\boldsymbol{s}] \sim \mathcal{N}(\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_\zeta),$$

Prediction friction covers missing features

(i) Pay-off as dot-product with unknown factor loadings on well-behaved factors

$$\underbrace{\hat{y}}_{\text{prediction}} = \underbrace{\hat{\boldsymbol{\beta}}^\top}_{\text{estimate}} \boldsymbol{\zeta} \quad \text{where} \quad \boldsymbol{\zeta} = E[\boldsymbol{q}|\boldsymbol{s}] \sim \mathcal{N}(\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_\zeta),$$

minimize mean squared error $\min_{\boldsymbol{c}} E\left[\{y - \hat{\boldsymbol{\beta}}(\boldsymbol{c})^\top \boldsymbol{\zeta}\}^2\right],$

for bias $\boldsymbol{\varepsilon}_\beta(\boldsymbol{c}) = E\left[\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}(\boldsymbol{c})\right] \neq 0$ and variance $Var[\hat{\boldsymbol{\beta}}(\boldsymbol{c})] = \boldsymbol{\sigma}_\beta(\boldsymbol{c})^\top \boldsymbol{R}_\beta \boldsymbol{\sigma}_\beta(\boldsymbol{c})$.

Prediction friction covers missing features

(i) Pay-off as dot-product with unknown factor loadings on well-behaved factors

$$\underbrace{\hat{y}}_{\text{prediction}} = \underbrace{\hat{\boldsymbol{\beta}}}_{\text{estimate}}^{\top} \zeta \quad \text{where} \quad \zeta = E[\boldsymbol{q}|\boldsymbol{s}] \sim \mathcal{N}(\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_\zeta),$$

minimize mean squared error $\min_{\boldsymbol{c}} E\left[\{y - \hat{\boldsymbol{\beta}}(\boldsymbol{c})^\top \zeta\}^2\right]$,

for bias $\boldsymbol{\varepsilon}_\beta(\boldsymbol{c}) = E\left[\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}(\boldsymbol{c})\right] \neq 0$ and variance $Var[\hat{\boldsymbol{\beta}}(\boldsymbol{c})] = \boldsymbol{\sigma}_\beta(\boldsymbol{c})^\top \boldsymbol{R}_\beta \boldsymbol{\sigma}_\beta(\boldsymbol{c})$.

(ii) Non-zero optimal bias.

Prediction friction covers missing features

(i) Pay-off as dot-product with unknown factor loadings on well-behaved factors

$$\underbrace{\hat{y}}_{\text{prediction}} = \underbrace{\hat{\boldsymbol{\beta}}^\top}_{\text{estimate}} \zeta \quad \text{where} \quad \zeta = E[\boldsymbol{q}|\boldsymbol{s}] \sim \mathcal{N}(\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_\zeta),$$

minimize mean squared error $\min_{\boldsymbol{c}} E\left[\{y - \hat{\boldsymbol{\beta}}(\boldsymbol{c})^\top \zeta\}^2\right] = \underbrace{\chi}_{\text{cost of complexity}} + Var[y|\boldsymbol{\beta}, \boldsymbol{s}]$

for bias $\boldsymbol{\varepsilon}_\beta(\boldsymbol{c}) = E\left[\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}(\boldsymbol{c})\right] \neq 0$ and variance $Var[\hat{\boldsymbol{\beta}}(\boldsymbol{c})] = \boldsymbol{\sigma}_\beta(\boldsymbol{c})^\top \boldsymbol{R}_\beta \boldsymbol{\sigma}_\beta(\boldsymbol{c})$.

(ii) Non-zero optimal bias.

(iii) Endogenous cost of complexity decreasing in weakness of trade-off (technology).

Price volatility

Excess price variance

Representative agent

$$Var[p] - Var[y] = Var[\hat{y}_I] - Var[y] = \underbrace{\chi}_{\text{cost of complexity}} - Var\left[y|\boldsymbol{\beta}, \boldsymbol{s}_I\right] - 2\boldsymbol{\beta}^\top \boldsymbol{\Sigma}_\zeta \boldsymbol{\varepsilon}_\beta,$$

Heterogeneous agents

$$Var[p] = \lambda_p^2 Var[s_U] = \lambda_p^2 \left\{Var[\hat{y}_I] + \psi_I^{-2}\sigma_z^2\right\},$$

where $\psi_I^{-2} = \alpha_I^2 \left\{\chi + Var[y|\boldsymbol{\beta}, \boldsymbol{s}_I]\right\}^2$

Price informativeness I

Planner's maximization of price informativeness heterogeneous agents

$$\min_{\boldsymbol{c}} E\left[(y - E[y|p,\boldsymbol{\beta}])^2\right]^{-1} = \min_{\boldsymbol{c}} \left\{ \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_q \boldsymbol{\beta} - \frac{\left\{\boldsymbol{\beta}^\top \boldsymbol{\Sigma}_\zeta \boldsymbol{\mu}_\beta(\boldsymbol{c})\right\}^2}{Var[\hat{y}_I(\boldsymbol{c})] + \{\psi_I(\boldsymbol{c})\}^{-2}\sigma_z^2} \right\}^{-1},$$

where

$$Var[\hat{y}_I] - Var\left[E[y|\boldsymbol{\beta}, \boldsymbol{s}_I]\right] = \chi - 2\boldsymbol{\beta}^\top \boldsymbol{\Sigma}_\zeta \boldsymbol{\varepsilon}_\beta, \quad Cov[y, \hat{y}_I]^2 = \left(Var\left[E[y|\boldsymbol{\beta}, \boldsymbol{s}_I]\right] - \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_\zeta \boldsymbol{\varepsilon}_\beta\right)^2,$$

$$\psi_I^{-2} = \alpha_I^2 \left\{\chi + Var[y|\boldsymbol{\beta}, \boldsymbol{s}_I]\right\}^2.$$

Price informativeness II

Convergence with better technology k_c^2 not stronger new signal (data-source) k_S under hard estimation scenario.

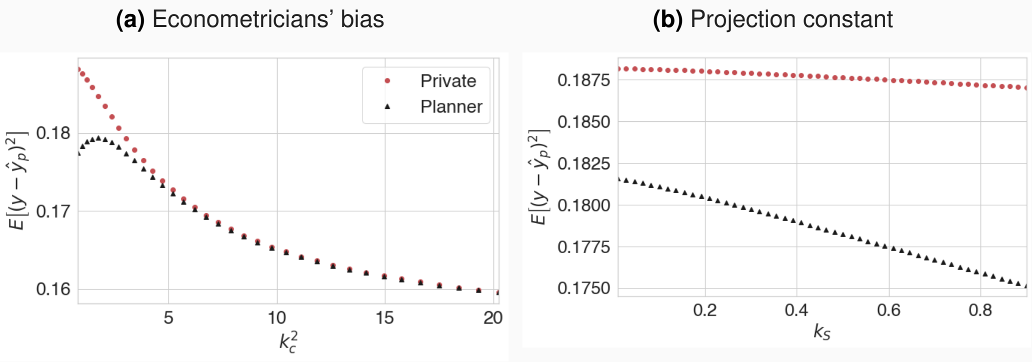

Figure 7: Comparative statics for price informativeness optimized by investors (Private) or Planner.

Estimation parameters

Minimized mse as cost of complexity χ vs cost of simplicity

$$\min_{\boldsymbol{c}} \underbrace{\boldsymbol{\varepsilon}_\beta^\top \boldsymbol{\Omega}_\zeta \boldsymbol{\varepsilon}_\beta}_{\text{Bias squared}} + \underbrace{\boldsymbol{\sigma}_\beta^\top \boldsymbol{D}_{\boldsymbol{\Omega}_\zeta} \boldsymbol{\sigma}_\beta}_{\text{Variance}} + \underbrace{Var[y|\boldsymbol{\beta}, \boldsymbol{s}]}_{\text{Irreducible noise}} := \underbrace{\chi}_{\text{cost of complexity}} + \underbrace{Var[y|\boldsymbol{\beta}, \boldsymbol{s}]}_{\text{cost of simplicity}},$$

$$\chi = k_{\sigma 0}^2 \mathbf{1}^\top \boldsymbol{X}^{-1} \mathbf{1}, \text{ where } \boldsymbol{X} = k_c^2 \boldsymbol{\Omega}_\zeta^{-1} + \boldsymbol{D}_{\Omega_\zeta}^{-1} \text{ and } k_c = k_\sigma / k_\varepsilon$$

Interpretation of parameters

$k_{\sigma 0}$: baseline estimation difficulty

$$E[(y - \hat{y})^2]_{c=0} = k_{\sigma 0}^2 \mathbf{1}^\top \boldsymbol{D}_{\boldsymbol{\Omega}_\zeta} \mathbf{1} + Var[y|\boldsymbol{\beta}, \boldsymbol{s}]$$

k_c^2: estimation technology quality (‘machine learning parameter’)

$$\partial\chi / \partial k_c^2 < 0, \qquad \lim_{k_c^2 \to \infty} \chi = 0$$

Robust linear demand

Demand is linear in the difference between prediction and price and derived from maximizing the expectation of the scaled profit function $\tilde{\pi}_i(y) := \alpha_i(y - p)$ applied to the prediction $\hat{y}_i$ with an uncertainty adjustment for the fact that investors optimize estimated rather than true profits.

$$\delta_i = \arg\max \ \tilde{\pi}_i(\hat{y}_i) - \frac{1}{2} E\left[(\tilde{\pi}_i(y) - \tilde{\pi}_i(\hat{y}))^2 \right] = \psi_i \left(\hat{y}_i - p \right),$$

$$\text{where} \ \ \psi_i = \left\{ \alpha_i E\left[(y - \hat{y})^2 \right] \right\}^{-1}.$$

For simplicity, assume that investors know the true mean squared error.

ASSET PRICING WITH COMPLEXITY

Return predictability OOS: Improving technology $\rightarrow$ different optimal bias and lower cost of complexity $\rightarrow$ (potentially) larger information set.

Price volatility: Noise in estimation drives excess, bias is ambiguous with high dimensionality.

Price informativeness: wedge between socially and privately optimal estimator.

Heterogeneous agents (Grossman and Stiglitz, 1980)

Value of information: Informed predictions are not always better.

Price reversals (price pressure): Estimation errors similar to liquidity demand but differ in relation to price volatility (not trading volume).

Fund performance: Under-performance of informed investors ‘predicted’ ex-post by over-optimism.

Machine learning in asset pricing with complexity

Optimal bias: Best prediction vs unbiasedness $\rightarrow$ contrasting views under the model:

- Investors' inference is well-modelled as an unbiased (potentially inefficient) estimator, econometricians' machine learning 'predicts' its own bias.
- Investors' inference is optimally biased and any technology faces the challenge of 'predicting' differences in bias.

Cost of complexity: Technological developments leads to discovery of ignored information.

Empirical implication: OOS predictability might be necessary but is not sufficient to draw conclusions about asset pricing models. Time-series and cross-sectional analysis of predictability. Prediction of non-market data.

Example: Extension to heterogeneous agents, distinguish ignored information from bias through variation in market digestion (in model: liquidity demand/noise trading).

What I do

1) Close the gap with new mechanism
2) Derive implications for measures of market efficiency:
    - return predictability (IS and OOS),
    - price volatility,
    - price informativeness,

    and market health

    - value of data,
    - price reversal (price pressure),
    - fund performance.
3) Calibrate the

Bibliography I

Dessaint, O., Foucault, T., Frésard, L., 2020. Does Big Data Improve Financial Forecasting? The Horizon Effect .

Dugast, J., Foucault, T., 2020. Equilibrium Data Mining and Data Abundance. SSRN Electronic Journal .

Farboodi, M., Matray, A., Veldkamp, L., Venkateswaran, V., 2020. Where Has All the Data Gone? NBER .

Farboodi, M., Veldkamp, L., 2020. Long-Run Growth of Financial Data Technology. American Economic Review 110, 2485–2523.

Gagliardini, P., Ma, H., 2019. Extracting Statistical Factors When Betas are Time-Varying. SSRN Electronic Journal .

Bibliography II

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

Gu, S., Kelly, B., Xiu, D., 2020. Empirical Asset Pricing via Machine Learning. The Review of Financial Studies 33, 2223–2273.

Gu, S., Kelly, B., Xiu, D., 2021. Autoencoder asset pricing models. Journal of Econometrics 222, 429–450.

Kelly, B. T., Malamud, S., Zhou, K., 2022. The Virtue of Complexity in Return Prediction. NBER .

Lewellen, J., Shanken, J., 2002. Learning, Asset-Pricing Tests, and Market Efficiency. The Journal of Finance 57, 1113–1145.

Ma, H., 2021. Conditional Latent Factor Models Via Econometrics-Based Neural Networks. SSRN Electronic Journal .

Bibliography III

Martin, I. W., Nagel, S., 2021. Market efficiency in the age of big data. Journal of Financial Economics .

Welch, I., Goyal, A., 2008. A Comprehensive Look at The Empirical Performance of Equity Premium Prediction. Review of Financial Studies 21, 1455–1508.